NEURAL AND OCULAR CORRELATES OF CONCEPTUAL GROUNDING IN VERBAL INTERACTION: A MULTIMODAL HYPERSCANNING APPROACH

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF INFORMATICS OF
THE MIDDLE EAST TECHNICAL UNIVERSITY
BY

EFECAN YILMAZ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE
OF DOCTOR OF PHILOSOPHY
IN
THE DEPARTMENT OF COGNITIVE SCIENCE

SEPTEMBER 2024

NEURAL AND OCULAR CORRELATES OF CONCEPTUAL GROUNDING IN VERBAL INTERACTION: A MULTIMODAL HYPERSCANNING APPROACH

Submitted by Efecan Yılmaz in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Cognitive Science Department, Middle East Technical University** by,

Prof. Dr. Banu Günel Kılıç
Dean, Graduate School of Informatics _____

Assoc. Prof. Dr. Barbaros Yet
Head of Department, Cognitive Science _____

Assoc. Prof. Dr. Murat Perit Çakır
Supervisor, Cognitive Science Dept., METU _____

Examining Committee Members:

Assoc. Prof. Dr. Barbaros Yet
Cognitive Science Dept., METU _____

Assoc. Prof. Dr. Murat Perit Çakır
Cognitive Science Dept., METU _____

Assoc. Prof. Dr. Cengiz Acartürk
Cognitive Science Dept., Jagiellonian University _____

Assoc. Prof. Dr. Hacı Hulusi Kafaligönül
Faculty of Medicine, Gazi University _____

Asst. Prof. Dr. Umut Özge
Cognitive Science Dept., METU _____

Date: *05.09.2024*

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name : Efecan Yılmaz

Signature :

ABSTRACT

NEURAL AND OCULAR CORRELATES OF CONCEPTUAL GROUNDING IN VERBAL INTERACTION: A MULTIMODAL HYPERSCANNING APPROACH

Yılmaz, Efecan

Ph.D, Department of Cognitive Sciences

Supervisor: Assoc. Prof. Dr. Murat Perit Çakır

September 2024, 117 pages

Effective communication depends on interlocutors agreeing on ambiguity resolution and constructing conceptual representations into shared meanings. To achieve this, continuous adaptation by all participants to establish and to maintain alignments is required, which is known as grounding in communication. Previous work on grounding has largely explored linguistic and ocular phenomena, whereas neural correlates have not been explicitly studied. In this thesis, hyperscanning and verbal interaction analysis approaches are combined to investigate the process of establishing a common ground –a social aspect of communication– using a multimodal approach that includes EEG, eye tracking, and fNIRS devices.

Eye-tracking results indicated that with mutual alignment, shared gaze becomes less central with explicit verbal communication taking precedence. Gaze entropy reduces as interlocutors move from high workload visual feature extraction to utilizing common meanings. EEG results showed that increased inter-brain synchrony in the frontal, midline, temporal, parietal, and occipital electrode sites in delta, theta, and gamma bands, which aligns with sustained attention, memory processes, and conscious integration of perceptions into actions that all correlate with shared linguistic contexts in communication. The synchrony observed across frequency bands highlights the collaborative nature of communication and the role of pragmatic alignment in facilitating effective communication. fNIRS hyperscanning results were inconclusive due to device malfunctions; yet single device hemodynamic response analysis revealed significant effects of repeated trials and experiment conditions on PFC oxygenation, supporting the role of this region in interaction and communication in relation to working memory and communicative intent. The findings reveal that dyadic communication involves multiplex synchrony, relying on both verbal and non-verbal cues as dyads established a common ground, with neural and ocular underpinnings.

Keywords: hyperscanning investigation, neuroergonomics considerations for linguistic grounding, grounding in communication, multi-modal fusion

ÖZ

SÖZEL ETKILEŞIMDE KAVRAMSAL TEMEL EDİNMENİN SİNİRSEL VE OKÜLER ÖRÜNTÜLERİ: BİR ÇOK KİPLİ HIPER TARAMA YAKLAŞIMI

Yılmaz, Efecan

Doktora, Bilişsel Bilimler Bölümü

Tez Yöneticisi: Doç. Dr. Murat Perit Çakır

Eylül 2024, 117 sayfa

Etkili iletişim, muhatapların belirsizlikleri nasıl çözecekleri konusunda anlaşmalarına ve kavramsal temsilleri paylaşılan anlamlara dönüştürmelerine bağlıdır. Bunu başarmak için, tüm katılımcıların anlam uyumluğu kurması, sürdürmesi ve sürekli adaptasyonu gereklidir, bu süreç iletişimde ortak temel edinme olarak bilinir. Ortak temel edinme üzerine yapılan önceki çalışmalar büyük ölçüde dilbilimsel ve göz hareketleri ile ilgili olguları araştırmışken, sinirsel örüntüler açıkça incelenmemiştir. Bu tezde, iletişimin sosyal yön –ortak temel edinme süreci– hipertarama ve sözel etkileşim analizi yaklaşımları birleştirilerek; EEG, göz izleme ve fNIRS cihazlarıyla çok kipli bir yaklaşımda araştırılmıştır.

Göz izleme sonuçları, ortak temel edinilmesiyle birlikte, paylaşılan gözbakışın önceliğini kaybettiğini ve açık sözlü iletişimin öncelik kazandığını gösterir. Muhataplar, yüksek iş yükü gerektiren görsel özellik çıkarmadan stabil ortak temelleri kullanmaya geçtikçe gözbakış entropisi azalır. EEG sonuçları, ön, orta çizgide, temporal, paryetal ve oksipital elektrotlarda delta, teta ve gama bantlarında artan beyinler arası senkronizasyonun, sürdürülen dikkat, hafıza süreçleri ve algıların bilinçli olarak aksiyonlara entegrasyonuyla uyumlu olduğunu ve bu süreçlerin iletişimde paylaşılan dilbilimsel bağlamlarla tutarlı olduğunu göstermiştir. Farklı frekans bantları üzerinde gözlenen senkronizasyon, işbirlikçi iletişimin doğasını ve pragmatik hizalamanın etkili iletişimdeki rolünü vurgular. fNIRS hipertarama sonuçları, cihaz arızası nedeniyle sonuçlanamamıştır. Ancak, tek cihazla yapılan hemodinamik yanıt analizi, deney koşullarının önlob oksijenlenmesi üzerindeki önemli etkilerini ortaya koyarak, bu bölgenin etkileşim ve iletişimdeki, çalışan bellek ve iletişim niyetiyle ilgili rollerini destekler. Bulgular, ikili iletişimin sözel ve sözel olmayan ipuçlarına ve çoklu senkronizasyona dayandığını, ortak zeminler kurulduğunda sinirsel ve göz hareketlerine dayalı temellerin de rol oynadığını ortaya koymaktadır.

Anahtar Sözcükler: hipertarama araştırması, dilsel ortak temel edinmede sinir-ergonomi değerlendirmeleri, iletişimde ortak temel edinme, çok-kipli füzyon

Home, at last;

breathe deep,

breathe safe,

breathe joy,

breathe, to a new adventure…

ACKNOWLEDGEMENTS

I would not be where I am without the touch of so many of whom I cannot count. Yet I must name only a few fit to this page.

So, thank you:

For your unwavering faith and efforts in, as well as patience for me; my bellowed wife Begüm, my mother Filiz, and my brother Eşref.

For your protective wings, endless support, and faith in me; my advisors and colleagues Dr. Murat Perit Çakır and Dr. Cengiz Acartürk.

For giving me my first chance in academia, for your support, guidance, and friendship; my past advisors, colleagues, and friends Erkan, Hatice, Serpil, Serkan, and Vural.

For your patience and endurance, being there when I needed the most from personal support to brain storming, and philosophical discussions; my friends, Buğra, Çağatay, Alaz, Bartu, Berrin, Hakan, and Hüseyin.

Thank you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABREVIATIONS

CCorr	- Circular correlation coefficient
CRQA	- Cross-Recurrence Quantification Analysis
DDRoIA	- Dynamically detected regions of interest annotator
DFA	- Discriminant function analysis
EEG	- Electroencephalography
ERP	- Event-related potentials
FDR	- False Discovery Rate
fNIRS	- Functional Near-Infrared Spectroscopy
G-Causality	- Granger-Causality
KMI	- Kraskov's mutual information estimator
MFC	- Medial frontal cortex
mPFC	- Medial prefrontal cortex
NE	- Neuroelectrics
PDC	- Partial direct coherence
PFC	- Prefrontal cortex
PLV	- Phase-locking value
PSTS	- Posterior superior temporal sulcus
SRP	- Saccade-related potentials
TME	- Transition Matrix Entropy
WTC	- Wavelet-transform coherence

CHAPTER 1

1. INTRODUCTION

Humans seek to transfer information using both explicit forms, such as verbal or literary methods, and implicit forms, such as gaze interaction (Hessels, 2020; Jording et al., 2018; Mendoza et al., 2021). Conversation is the primary method of linguistic interaction, occurring within a framework of various constraints. These constraints range from simpler ones, like the perception of time or the conditions of the communication medium, to more complex ones, such as past experiences, communicative tasks, or intent. Research in cognitive science and linguistics consistently highlights that human communication is inherently a lossy process, where not all information communicated is perceived as *intended* or as *meant*, as theorized by Grice (1957, 1975, 1978) and evidenced by Clark and Wilkes-Gibbs (1986).

Due to inherent constraints in verbal natural language communication (Clark & Brennan, 1991) what a speaker utters and what a listener perceives and understands can be similar, but never exactly the same. This is because ambiguities in natural language arise from the complex interplay of syntax and semantics, which are also influenced by the broader structures of cognition, perception, and action. This also defines that in human-human interaction, the interlocutors of a verbal communication setting may allow meanings previously unincluded with the initial conceptual representations of stimuli in the process of resolving local ambiguations. This may be due to the fact that, for any joint and synchronous communication (e.g., verbal communication), individuals may have different worldviews or goals shaped by their communicative intent and past experiences. These initial frames are updated recursively as new information, such as the expanded affordances of visual stimuli in a feature extraction task is integrated into existing concepts. These differences can affect the conceptual alignment between what is said and what is implied, whether for conventional or unconventional meanings (Grice, 1978). This conceptual alignment among interlocutors is crucial for successful communication (Grice, 1975). As without a sufficiently broad and deep conceptual alignment, participants may struggle to ground shared meanings and achieve successful joint action (Sebanz et al., 2006) or a conceptual pact (Brennan & Clark, 1996), hindering both the transfer and perception of information.

In natural linguistic interaction, the actions of oneself and others (Sebanz & Knoblich, 2009), the objects in the environment (Hanna & Brennan, 2007), and the medium of communication may contribute to the manipulation of how initial perceptions (global meanings (Grice, 1975, 1978) are shaped and how representations local to the conversation evolve (Clark & Brennan, 1991). However, because these factors are subjective, only an accepted level of alignment between the representations of interlocutors allows for successful communication. Particularly in a temporally constrained communication setting, such as a non-literary real-time conversation a critical point may arise, in which all participants involved in the joint setting must adapt dynamically to these manipulations (Cooke et al., 2012). Clark and Brennan (1991) described such a setting an interaction where each speaker and listener were jointly tasked with attending to one another, as listener(s) signal to their speakers that they could hear, comprehend, and follow what was being said by gaze, gestures, or verbal cues. They suggested that the interlocutors' ability to engage in audience design (also visited in, Horton & Gerrig, 2005) would be constrained by the communication medium as well as the joint goal, and that the interlocutors would be entrained during the communication process to refine their interaction to collectively expend the least collaborative effort. The authors stated that this process had a significant workload (Bard et al., 2009; Dehais et al., 2020), indicated by the constraints, and the complexity of the setting interacting with the dynamicity of accepted level of alignment. Clark and Wilkes-Gibbs (1986), as well as Grice (1975, 1978), provided discussions to that this acceptable level of alignment was not always matched among interlocutors and that the extent to which there is a mismatch can reflect engagement and comprehension of the interlocutors – two key aspects of communication (Barr & Keysar, 2002; Sebanz et al., 2006; Sebanz & Knoblich, 2009). The process of establishing this alignment among interlocutors for what is said and perceived with their intended meaning was called common grounding in communication (Clark & Brennan, 1991). This process may take place both intentionally and automatically. While Garrod and Pickering (2009) argue that some aspects of dialogue are automatic and do not require intentional effort from the speaker, Baldwin (1995) suggests that intention might still be necessary for certain aspects of this process.

In prior research, the establishment of common ground has been characterized as a gradual process (Clark & Wilkes-Gibbs, 1986), potentially extending beyond individual cognition to function as a distributed process in multi-party interaction settings (Hutchins & Tove, 1999). The social dimension of grounding has been subject to debate, with Barr and Keysar (2002) positing that grounding was a process independent of specific interlocutors, whereas Brennan & Clark (1996), as well as Metzing & Brennan (2003), documented the emergence of partner-specific effects dependent on specific interlocutors. From a joint action perspective, this renders grounding a distributed process in which interlocutors may complement one another's actions (Sebanz et al., 2006). The process of monitoring others during interaction (Clark & Krych, 2004; Ninomiya et al., 2018; Vesper et al., 2010), alongside perceptual constraints (Blakemore & Decety, 2001; Knoblich & Flach, 2003; Repp & Knoblich, 2004), and the cognitive load associated with grounding, has been shown to correlate with both oculomotor and neural measures.

Due to this and on the precedent that joint attention was likely a pre-linguistic process (Hopkins & Taglialatela, 2011); investigations focusing on interaction by employing eye tracking (e.g. Gilchrist & Harvey, 2000; Wilms et al., 2010), and electroencephalogram (EEG) (e.g. Berka et al., 2007; Ceh et al., 2020) have long sought to identify the neural correlates and underlying mechanisms of communication, including in studies employing gaze-contingent agents in virtual environments (Pfeiffer et al., 2013; Schilbach et al., 2013). Recent studies (Dehais et al., 2020; Omurtag et al., 2019) have highlighted the prospects of elucidating the interplay between mental workload and the representational states of interlocutors, particularly under the assumption that conceptual grounding during communication may proceed automatically (Garrod & Pickering, 2009). From a neuroergonomics perspective, these findings are particularly relevant for understanding linguistic interaction, where structural complexity (working memory and spatial processing), temporal constraints (for instance, the recall time of an old grounding struct or the nature in which sustained attention can be challenged), and lexical factors (referent word frequency or the process of refinement into common grounding) all contribute into the cognitive workload and may be reflected neural and ocular correlates associated with the grounding process.

Frith and Frith (2003) highlighted that classifying the attentional states of interlocutors is crucial for advancing research on linguistic interaction in social contexts. Baldwin (1995), along with Frith and Frith (2010) argued that a shared locus of attention was a necessity for coordination of informative communication. This view was also supported by eye-tracking studies, such as (D'Angelo & Gergle, 2016; Pfeiffer et al., 2013; Richardson & Dale, 2005) which identified a linear relationship between shared gaze (i.e., focusing on the same object as an interlocutor) and complexity of deictic references, as well as the neural correlates of gaze in social interactions. These studies also examined the impact of co-location versus non-co-location on the connection between gaze and discourse comprehension. In line with Richardson and Dale (2005), and D'Angelo and Gergle (2016), Bard et al. (2009) explored the role of shared visual space and the complexity of references.

Overall, there is an in depth and very wide research focused on participants in joint settings establishing a common ground as they conversed towards a common goal. Yet there is a gap exists in the literature for hyperscanning or second-person neuroscience approaches employed to investigate conversational dynamics and their neuro-electric, ocular, and hemodynamic correlates, as well as how these correlates interact with the linguistic process of establishing a common ground. This gap is the focus for the present thesis.

Montague et al. (2002) and Cui et al. (2012) pioneered the use of hyperscanning for multi-subject monitoring of neural activity in interactive settings, employing EEG and functional near-infrared spectroscopy (fNIRS), respectively. Subsequent studies utilizing fNIRS, such as those by Holper et al. (2012) and Nozawa et al. (2016), further elucidated functional connectivity during imitation tasks and workload dynamics in verbal games. These findings underscore the significance and validity of hyperscanning techniques, particularly EEG and fNIRS, in exploring brain-to-brain connectivity and cognitive workload during social interactions.

In the light of the preceding literature, and in support of the remarks on the assumed gradual nature of the pragmatic effect that is the process of establishing a common ground, we hypothesize that the process of grounding will entail measurable neuroelectric, ocular, and hemodynamic correlates. And further explicate the inner dynamics of these correlates and how they interact with this gradual process. For instance, in neuroelectric measures we might see increased inter-brain connectivity. In ocular measures dynamics of shared visual space in recurrent gaze or in gaze entropy might reveal an interplay between grounding and ocular dynamics. While in frontal cortex hemodynamics, through the process of grounding might be affected as neuroergonomics of the grounding process might take place, such as sustained attention and workload, and conversational roles.

In the present thesis, firstly, a literature review on the neural correlates of verbal communication, joint attention, action coordination, and action identity are presented to draw a conceptual framework for our investigation. Then, the methodology section is presented, wherein the visual matching and spatial sorting paradigm employed by Clark & Wilkes-Gibbs (1986) as a task to study grounding in communication is restructured in a hyperscanning experiment setting to accommodate two fNIRS, two eye tracking devices, and two EEG devices. This is followed by a section on the analysis methodology where behavioral, eye tracking, EEG, and fNIRS analysis pipelines are described in detail. The results section presents the behavioral evidence for the achievement of common ground in terms of linguistic indicators, and the corresponding neural (neuroelectric with EEG and hemodynamic with fNIRS) and eye tracking outcomes. Next, a discussion of the identified neural and ocular correlates is presented in relation to the preceding related literature. Finally, a short summary on how the findings of the present thesis address our main research question together with future prospects are given in the conclusion section.

CHAPTER 2

2. LITERATURE REVIEW

2.1. Grounding in Communication, Joint Action, and Joint Attention

Herbert Clark and Deanna Wilkes-Gibbs (1986) investigated how interlocutors of a conversation would coordinate to a mutual meaning to, for instance, resolve ambiguous references. The authors aimed to develop a framework for how or whether the utterances and the referring expressions their participants produced developed throughout the experiment. This development could be in form of refinement or subjective entrainment, which Clark and Brennan (1991) later called the grounding process. To this end, Clark and Wilkes-Gibbs employed a verbal communication setting, in which dyadic interaction naturally consisted of a speaker and a listener, acting in the roles of a director and a matcher. The interaction task was a joint task for arranging complex figures to obtain a matching arrangement between the matcher and the director participants. In their investigation, the authors focused on the roles of interlocutors during conversation; a speaker, who was tasked with planning and revising their utterances and a listener, who was tasked with hearing and understanding without delay. In the proposed framework, the speaker presented a noun phrase about one of the tangram figures as the participants repaired, expanded, or replaced this noun phrase until an acceptable level of alignment on the meaning of the referring expression was achieved among the pair. From the perspective of Grice (1975, 1978), the alignment process would not only comply with but also highlight the existence of a conversational implicature among the interlocutors, with which they were tasked to resolve by means of establishing a ground for their conceptual differences. Grice discussed that there might be a distinction between the implicated and said meanings within utterances, and Clark & Wilkes-Gibbs and Clark & Brennan account for both within-participant (i.e., what is global meaning for a participant and what is local) and between participant (i.e., mutual agreement and the grounded referents) distinction of conceptual meanings. Additionally, Clark & Wilkes-Gibbs stated that the acceptable level of understanding between interlocutors would not always match, which highlighted a further possible distinction where the participants might not have engaged in the joint task equally; an account they later investigated from the perspective of historic accounts in Brennan & Clark (1996) and Metzing & Brennan (2003).

Clark further investigated the grounding process with his collaboration with Susan Brennan from a Gricean perspective in (1991) for communication to be a joint action (Sebanz et al., 2006). According to Sebanz et al. joint actions must involve joint attention, action observation, task sharing, and action coordination. Clark & Brennan's research and discussion focused on the joint nature of linguistic communication, which involved the shared task of communication and coordinating one's conceptual representations with the other to update these to match that of the others through observations, both of which necessitated joint attention. In a natural interaction setting, however, the utterances may afford new features to visual stimuli (Gibson, 1979), and therefore result in, altering manipulations to the concepts each participant had prior to their interacting. In accordance with this, Clark and Brennan argued that meanings on which grounding can take place must be contextually localized to the present discourse and also must be updated by the participants of the communication medium continuously. The authors concluded their discussion that grounding as a joint action was most affected by communicative intent and medium of the communication. To which end, the content, such as references or verbatim information, as well as the medium, co or non-co location settings, and literary or verbal interaction settings would affect the process of grounding to always result in the least collective joint effort by all participants; the phenomenon was named the principle of least effort by the authors.

The joint aspect of communication was later discussed by Baldwin (1995) for whether joint attention might itself be a mental construct, for the forming of which interlocutors must beware of the joint nature of their interaction setting. Baldwin, firstly argued that for joint attention to be formed, each joining party need not be co-located and/or observant of the others, and secondly, that the forming of joint attention would not require intentionality. As a consequence of attending without intention, interlocutors of such a setting might lack intersubjective understanding, and, therefore, local representation of another interlocutor's intentional stance might not be initially perceived. In which case, Baldwin argued that these participants of the joint attention setting would also play no role in the coordination of joint attention without intention. Clark and Krych (2004), on the other hand, argued shared spatial representations and co-location of interlocutors were important for the forming of joint attention. Yet from the perspective of intentionality, the authors assumed existence of a common goal to communicate with a stable conversational organization. To this end, they argued that for success in communication, a speaker would be tasked with getting their addressee's attention, for their addressee (i.e., listener) to conceptually identify, understand, to, finally, consider responding to the speaker.

The term "intentional stances" originates from the work by Dennett, who in (1981) discussed how we might attribute an intentional stance to entities based on their behavior or design, as a way to conceptualize beliefs. A belief, according to Dennett, is not an objective structure but is subjective to the system in question and the observer's perspective. This means that while two identical systems exposed to identical sensory inputs might conceptualize similar beliefs, variations in internal processes and context can lead to different outcomes, in line with the theories by Grice (1957, 1975, 1978) from a linguistic perspective or Gibson (1979) from the perspective of visual cognition. Thus, the intentional stance is a tool for understanding and

predicting behavior, rather than a strict assertion that the system possesses actual beliefs or intentions. This might signify the distinction between an attending (Reddy, 2011) participant to the discourse present in the medium, as in Clark and Krych (2004), and a non-attending observer, as in Baldwin (1995). This distinction predominantly highlights the necessity of a joint attention setting for a successful communication situation, an example of which was studied by Garrod and Pickering (2009) on the account that the joint alignment between dyads could be at many linguistic levels, such as the lexicon, the grammar, the intonation, as well as at a non-linguistic level. Regardless, they argued that despite their multiple systems view, the alignment would always be positively affected by the existence of a joint goal; which highlights the multi-modal nature to their interactive alignment model.

The positive effect of a common goal as one of the pillars of joint action (Sebanz et al., 2006) and on the success of a communication setting was the focus on the roles of interlocutors in a dialogue setting by Bard et al., (2009). They employed a Tangram experiment with a joint goal similar to that of (Clark & Wilkes-Gibbs, 1986); however, Bard et al. also introduced eye tracking and mouse tracking measures alongside their behavioural and linguistic measures. Additionally, the authors also introduced new experiment conditions; where the participants would change or share the two roles (directing, matching), and the experimenters would manipulate whether participants knew of the other's attention and actions by showing or not showing the screen-space actions or a gaze cursor. Their analysis results showed that only speakers' actions and their roles affected the use of deictic references and that players who declared strategies and accessibility to referents were less affected by local facts. Furthermore, the elaboration of referring expressions was deemed relevant to the players' ability to coordinate joint attention by the authors. The authors' discussion reflected that linguistic complexity may be significantly affected by the social setting and the accessibility of the interlocutors to the self-representations of one another. They claimed that more access interlocutors had to each other's representations then this would entail more efficient referents. As the joint goal from the perspectives of the differing roles as well as the actions of each participant affected the affordances of their implicatures, it may be possible to claim that both the level of collaborative effort and the affordances of the utterances of each participant were subject to change during the entrainment process. The pivotal discussion on the subject of accessibility was by Marr and Nishihara (1978), wherein along with the scope and the uniqueness of a visual search phenomenon, accessibility was marked as one of the pillars of higher-order processing.

Under the assumption that interlocutors of joint action have no pre-existing concepts for one another particularly, the entrainment process to reach an acceptable level of alignment during a temporally constrained joint task was investigated for partner-specific effects by Brennan and Clark (1996). The authors claimed that there may exist a historical account of the production of referring expressions to explicate the role of an entrainment effect on both the individual and distributed processes of language use in joint attention settings. To this end, they conducted three experiments to evaluate their proposed model for referring; and the outcome of the experiments was that a lexical entrainment effect could be observed, as suggested by the informativeness, recency, and refinement of an utterance, and further resulting in that the grounding

established for one part of the task may carry over to another, and to new partners as well. In other words, the participants of their dyadic setting were likely to employ their previously achieved and successful groundings with new partners; which the authors deemed the conceptual pact. Furthermore, a follow-up study, which employed eye-tracking and a measure of gaze dispersion as an indicative of visual search space by Metzing and Brennan (2003) on the effects of intentional manipulation of a conceptual pact, showed that a previous grounding being broken – for instance, a previously agreed referent to the object being replaced with a novel one – resulted in a larger search space and, as a result, increased workload of disambiguation by the matcher.

Theory of conceptual pacts and partner-specific entrainment in interaction might be action prediction and successfully constructing a concept of others, which is a topic explored by Knoblich and Flach (2003) as well as Sebanz and Knoblich (2009). In the first of these studies, Knoblich and Flach investigated action identity under the assumption that the action system contributed to perception, providing a theoretical basis for action identity. In the discussion of the common coding theory, highlight how the perception of one's own actions and the resulting outcomes may be more accurately predicted compared to the perceptions and predictions of others' actions. This asymmetry suggests that self-observation offers a more reliable basis for predicting action outcomes than observing others. This idea aligns with Gibson's theory of affordances (1979), which proposes that the environment offers opportunities for action based on an individual's capabilities, and with the concept of joint action (Sebanz et al., 2006), where coordination with others relies on these predictions. This connection underscores the notion that self-observation might provide a deeper understanding of potential actions compared to observing others, thereby influencing how interlocutors engage in collaborative tasks. This perspective might be related to how in a conceptual pact, where the predicted outcome of action might result in a misidentification when the utterance was unexpectedly changed to no longer fit the localized (self) concept of the matcher; as previously explicated by Metzing and Brennan (2003). Sebanz and Knoblich's (2009) work on how joint action was coordinated serves as an important baseline discussion as they approached the outcomes of prediction in joint action as an extension of the common ground (Clark & Brennan, 1991; Clark & Wilkes-Gibbs, 1986). For this purpose, Sebanz and Knoblich evaluated the what (the inferred action), the when (a modal of joint performance in executing the predicted action), and the where (the concept of the shared perceptual space and joint attention) properties of action prediction. As a result of their discussion, the authors remarked that both spatial and temporal coordination was necessary for action-prediction to serve as a self-supervision method towards entrainment in joint action.

This necessary temporal and spatial coordination highlight a similar notion to that of Garrod and Pickering (2009) that a common ground may requisite processes outside of the linguistic domain. Another study similar to the work by Garrod and Pickering had previously been conducted by Horton and Gerrig (2005), who discussed a conversational common ground as a memory-related process in addition to a representational construct emergent in interlocutors. In that, the authors evaluated strategic methods, such as the presentation, acceptance or repair suggested by Clark & Brennan (1991) as well as automated executions, such as a recall from a previous

occurrence of a predicted action and, thus, a previous common ground. Horton and Gerrig specifically remarked lexical and/or syntactic priming as important contributors to message formation, and also hinted at the possibility of conceptual pacts (Brennan & Clark, 1996). Consequently, the authors proposed and demonstrated that conversational common ground was not always the result of a novel process; in that, a common ground dependent on the interlocutors and the environment of the discourse could be recalled from memory, albeit that the recalled commonality would still be subject to the grounding process if an error was made in the recall, or the original representation of the commonality, or if there was a false action prediction. The topic of lexical priming was also the subject of an investigation by Cleland & Pickering (2003), wherein the authors argued that lexical retrieval was a two-stage process consisting of activation of lexical information and the production of complex noun phrases. The authors' suggested a theoretical existence for anchoring points interlocutors provide and/or propose to one another during a conversation, which may extend to a perceptual basis and a projection of this process might correlate with action coordination (Sebanz et al., 2006).

In the linguistic neuroscience literature, study by Hopkins and Taglialatela (2011) examined whether joint attention precedes linguistic processes during cognition. They explored this question by leveraging clinical evidence on early reciprocal communication skills, while addressing ethical and experimental challenges associated with neuroimaging in children. They proposed that chimpanzees could provide valuable insights into the neurological correlates of pre-linguistic communication. Their study outlined behavioral and neuro-anatomical evidence that suggested similarity in joint attention skills between humans and chimpanzees, with findings indicating that the posterior temporal cortex and gesture asymmetries in chimpanzees highlight a homologous function to Broca's region in humans, emphasizing its significance in communication, as the authors emphasized the foundational role of joint attention in effective communication.

2.2. Hyperscanning and Distributed Cognition

The joint action perspective of conversational common ground and its correlation with action coordination was notably explicated by the work of Hutchins & Tove (1999), wherein they highlighted that cognition in an interaction setting must not be solely evaluated from the perspective of a single participant. To this end, the authors investigated the distributed cognition model, consisting of an interaction setting evaluated as a holistic system. In that, the distributed cognition system might be capable of carrying a higher mental capacity than each member individual of the holistic system could. The authors investigated their hypothesis in an aeroplane cockpit interaction setting, where the flight crew were tasked with flying the aeroplane as each of them had shared roles due to the amount of workload exceeding what a single human participant could achieve due to spatial and temporal constraints. Hutchins and Tove's theory suggest that when interlocutors share the workload of memory, perceptual tasks (such as visual search, where participants' visual spaces overlap but do not fully converge), and linguistic processes (like grounding as a shared process), they

constitute a more effective system. This insight into the complex interactions between interlocutors hints at the possible application of gestalt theories in understanding cognition, where the whole system performs better than the sum of its parts. Similarly, Cooke et al. (2012) drew a parallel line to the work by Hutchins and Tove (1999), as the researchers investigated their proposed term interactive team cognition as an emergent feature for a co-located group of individuals involved in a joint action situation. Cooke et al. primarily evaluated the alignment and/or coordination in a shared task and/or joint goal situation as a distributed process, with the detail in which the joint representation of the task in itself must align recursively in between the team members in a notion similar to that of Baldwin (1995). However, from a methodology perspective, the distributed cognition theory bares the importance that – both methodologically and theoretically – only an investigation of all participants of such a joint and shared interaction setting may explicate the underlying neural and/or cognitive aspects in a methodology deemed as a hyperscanning study by Montague et al. (2002), two-body neuroscience by (Dumas, 2011), and the second-person approach by Schilbach et al. (2013).

In their pioneering work, Montague et al. (2002) addressed the gap caused by investigating only one side of multi-participant interaction settings to propose the simultaneous recording of neural signals synchronized on a temporal domain, which offered the advantage of correlating interlocutors' neuro-signals to allow the possibility of revealing a dynamical relationship between the activations of brain tissues between interlocutors. To explicate their approach for a novel investigation setting, Montague et al. conducted a dual fMRI experiment, wherein participants attended the experiment synchronously, but without co-location, in a competitive social interaction scenario of playing a game. In their methodology, the authors particularly listed inter-device variabilities, such as calibration and analysis related sampling noise as possible issues. Schilbach et al. (2013) drew a parallel line with Montague et al., as they expanded the hyperscanning notion to that not only must the methodology be constructed to observe neural, behavioural, or ocular signals from all parties but the participants must also be engaged in the encounter, rather than employing a setting where some participants may solely observe others. Thus, they outlined the theoretical framework of the second-person approach that addresses the neural correlates of detachment versus engagement state in dynamic interaction settings.

Dumas et al. (2010) described social interaction as a temporally continuous adaptation process the result of which was contributed by all attending participants, and, as such, the interaction setting necessitated a hyperscanning investigation for this interactional synchrony to be explored. In their investigation, dual-EEG recordings were collected on an interacting dyad, which imitated each other's hand movements over a video stream. The authors reported that the interactional (task related) synchrony correlated with an increase in alpha mu band activity in the right centroparietal regions among the dyad. Dumas (2011) in their follow up discussion on the cognitive aspects of their previous work focused on the importance of a well-structured experiment setting for a hyperscanning topology to function. In that, the author drew a frame in which ecological validity and a successful joint attention setting may not alone be sufficient to explore possible connections among an interacting dyad; wherein, cognition,

perception, and action processes for a perceptibly synchronous task shared by the interlocutors might still not establish a strong inter-brain connectivity due to the phase shift the participants go through. The authors utilised the analysis method phase locking value (PLV) to deal with this inherent "delay" of synchronizing neural activity and explore the neuro-cognitive outcomes successfully. PLV is one of the most commonly used methods of analysis in connectivity, particularly with EEG (Burgess, 2013) and has been first demonstrated in (Lachaux et al., 1999) as a method allowing for investigation of quantifying the consistency of phase relationships between the EEG signals of different individuals, for instance in a hyperscanning setting, at specific wave frequency bands, such as alpha mu band or delta band.

However, the interpretation of these EEG signals in hyperscanning investigations have been a hot topic of discussion in the past decade, just as much as analysis methodologies have been. The work by Marriott Haresign et al. (2022), attended to the issue of temporally constrained analysis methods in hyperscanning, particularly to that of analyzing interaction data regarding the process of inter-brain-entrainment (IBE), which is the particular focus of the present thesis as well. To this end, the authors evaluated PLV (phased connectivity), Granger Causality (G-Causality, temporally consistently shifted connectivity), and correlation analyses (for instance, to evaluate zero lag connectivity). Marriott Haresign et al., also focused on the importance of ecologically valid experiment design and how experiment design, particularly in terms of IBE might interact with artifact cleaning during signal processing.

Another work, particularly highlighting the importance of both the experiment setting employed and how the analysis was conducted, as well as interpreted, was published by (Burgess, 2013). Burgess firstly introduces the most commonly utilised three methods of analysis with covariance in amplitude or power, partial direct coherence (PDC), and phase synchrony. The author, then compares particularly PDC and phase-locking value (PLV, a method of phase synchrony analysis) with alternative metrics, such as coherence, circular correlation coefficient (CCor), and Kraskov's mutual information estimator (KMI) to highlight the strengths and short comings of these widely utilised methods of analysis in hyperscanning literature. For instance, Burgess defines the necessity of temporal consistency for a high PLV value as a short coming where a covariance analysis might have been stronger against coincidental effects. In a modelling approach, Burgess utilised a hyperscanning EEG data of human participants to empirically discuss emergence of spurious hyper-connectivity in a controlled analysis setting and how this phenomenon interacted negatively with the biases of PLV or PDC; wherein, they conclude that CCor and KMI are valid alternative methods of analysis to support PLV and PDC in prevention of outcome errors.

Hyperscanning investigations may be well represented by another pioneering work by Cui, Bryant, and Reiss (2012), where the authors offered near-infrared spectroscopy (NIRS) as a solution with cost, mobility, and high ecological validity in conducting hyperscanning experiments on interaction settings. To this end, they devised a button pressing experiment, in which dyads either cooperated, competed, or attended individually with the non-attending participant in an observation role. The results of their analysis utilizing the wavelet transform coherence (WTC) method showed that the pairs had increased incoherent activity in the superior frontal cortex when cooperating, and further that the effect increased as the participants cooperated more

accurately in secondary trials, as demonstrated by a significant decrease in inter-button press times. Another dual fNIRS study utilizing WTC and Granger-causality (G-causality) analysis methods was conducted by Holper, Scholkmann, and Wolf (2012) in an investigation of both the modal and the imitator in an imitation task from a neuro-imaging brain-to-brain coherence perspective. To this end, the authors utilized a finger-tapping experiment task for the variable of self-paced versus stimuli paced action coordination, brain-to-brain coherence, and the physiological measure of heart rate. Other than coherence, moreover, the authors validated inter-participant effects on synchronicity; in that, the participants of the dyadic setting either had or had not a shared visual space. The results of their study demonstrated that a WTC analysis was successful in distinguishing task-related (period of finger tapping) and physiological (heart rate synchronicity); which showed, firstly, that self-paced imitation resulted in higher coherence, and secondly that heart rate coherence was only significant when the participant pairs did not have their visual spaces restricted (i.e. when the participants could see each other in the main task, in comparison with the control condition when they could not see each other) during the task. G-causality further demonstrated results validating WTC that visual space restriction resulted in a similar separation between main and control conditions, in addition to a similar separation between self and stimuli paced action.

Another work on hyperscanning has been conducted by Funane et al. (2011) for an investigation of the possible correlation between cooperating participants' hemodynamic responses during a shared task. Their experimental procedure consisted of the two participants counting to ten on the onset of an auditory cue as an internally directed cognition task, which was followed by a press of a button in temporal synchronization together with a collaborator. It was demonstrated in their results that the Spatio-temporal activation in the prefrontal cortex was inversely correlated with the button press time intervals between participants; in that, a higher correlation of neural activity between participants and shorter inter-participant intervals were correlated. A recent review by Redcay and Schilbach (2019) on the other hand, focused on the underlying neural mechanisms correlated with social interaction specifically. In that, the authors investigated whether interaction always entails a social component and how the ecological setting within which the interaction takes place may affect the interaction. The authors outlined that the components of a successful social engagement must include mutual engagement, behavioural alignment, joint attention, and an ecologically valid, naturalistic setting; which highlight the utilization of EEG, fNIRS, and eye-tracking alike high-mobility and non-intrusive neural or ocular acquisition techniques.

Eye-tracking as one of these high-mobility sensors has been a frequent tool of investigating interaction among researchers for this reason. Richardson and Dale (2005) evaluated the relationship between speaker and listener eye movements to correlate the temporal coupling with linguistic interaction on a time domain, and with additional perspectives of coordination and comprehension. As a result of their empirical investigation, the authors successfully explicated that the alignment rate and delay of addressees' gaze in an interaction setting with a non-co-located interlocutor were indicative of linguistic alignment among the interpretations. An investigation that employed an interaction setting where the interlocutors were co-located was conducted

by Hanna and Brennan (2007), wherein the authors evaluated eye gaze as a constraint in the interpretation of referring expressions as well as the disambiguation of these expressions in a joint attention setting. Building upon the ambiguity in spoken utterances due to the nature of temporality in speech, the researchers focused on (1) the utilization of gaze in the resolution of temporary ambiguities, (2) when in the linguistic process the utilization of eye gaze is integrated, (3) whether the integration of eye gaze is an orientating cue (i.e., a directing cue for an interlocutor in the visual space), or a context-dependent cue. The researchers' results from their experiments showed that the existence of non-target competitors in the visual search space influenced the point in time when gaze was integrated into the linguistic disambiguation process. Additionally, the presentation of congruent or incongruent displays in the experiment setting significantly affected the total time participants spent fixating on the target, therefore validating that gaze was an auto-orienting factor in the modulation of attention in the visual space.

The gaze of an interlocutor may also serve as a non-linguistic referring expression to others. As investigated by Schilbach et al. (2010), where the authors investigated the neural correlates of joint attention. Using utilizing a director, a matcher, and an intermediary avatar, Schilbach et al. were able to manipulate gaze contingency, and the joint (the avatar follows director's gaze) and non-joint (the avatar gazes elsewhere). Their results showed that engagement in joint attention showed differential activations in, for instance, both the dorsal and ventral portions of the medial prefrontal cortex. Schilbach et al.'s work on fMRI and eye-tracking was but one example of neural and eye-tracking studies. Another investigation in the same domain was conducted by Pfeiffer, Vogeley, and Schilbach (2013) to explicate the neural correlates of gaze as an attention orienting tool. To deal with this, the authors evaluated joint attention for its neural correlates that medial prefrontal cortex (mPFC) and posterior superior temporal sulcus (pSTS) were attributed as brain areas most active during detection of gaze direction, orientation and/or re-orientation of attention and mentalizing to propose a novel neurofunctional model of social gaze in interaction. Visual search paradigms in joint attention experiments have often been utilized along with eye-tracking, as exemplified by Brouwer et al. (2018) in an EEG and eye-tracking study where the researchers investigated the event-related potential (ERP) and ocular features to provide an overview of saccade-related potentials (SRP) for possible interactions between these features in a visual search paradigm to propose SRPs as a novel classifier of target and non-target stimuli. The authors successfully demonstrated that the interaction between an ERP analysis (P3 region) and other ocular features in a multi-modal approach were more robust classifiers of target versus non-target detection during high mental load conditions (employed as a dual-task paradigm in the present paper) than sole utilization of ocular features.

The medial frontal cortex (MFC) region was also the focused neural region of Ninomiya et al. (2018) in their investigation of the social aspect of action observation with entrainment, defined by the authors as the result of a feedback process due to observing and structuralizing the consequences of actions and mentalizing the intentionality as well as the intended desired outcomes. They reported on the results of their empirical work that the neural correlates of performance observation of others in an fMRI setting using macaque monkeys that MFC was a crucial brain structure in

the differentiation of the monitoring process for self and/or other's task performance, additional with the superior temporal sulcus (STS) region. The authors, then, distinguished posterior MFC as an area specifically active during both social and non-social task performance monitoring and a gradually increasing activation of this area as well as the overlapping dorsomedial prefrontal cortex were deemed indicatives of entrainment to an interlocutor's actions (i.e., action prediction), as well as possible indicatives of workload for the neural tissue.

The neural tissue activations considered under the term workload was initially a task difficulty measure, utilized by Izzetoglu et al. (2005), wherein the authors exemplified a modelling approach driven by the hemodynamic response in the prefrontal cortex. They utilized a minimally intrusive method of monitoring brain tissue with functional imaging via fNIRS. The authors suggested that the continuous representation of trials summate the hemodynamic response; therefore, a block of trials set for a single type of variable was their recommendation as to the experimental setting. To this end, they devised an experiment that consisted of an anagram task with a discrete difficulty variable distinguishing block of trials. Izzetoglu et al. validated their methodology for the outcomes of time to peak and amplitude for the hemodynamic response in relation to the task difficulty variable and participants' response times to successfully model the hemodynamic response to each type of stimulus. Another work in workload research was by Berka et al. (2007) where the authors utilized an EEG device as they provided an overview for the correlates of vigilance, memory recall, learning, and visual search in a series of psychophysical experiments (in other words, an experiment battery). Following this, they proposed a modelling approach to classify their two variables (1, task engagement and 2, workload) and correlation analysis for the behavioural and subjective measures obtained utilizing their experiments. The results by Berka et al. showed that their approach of utilizing discriminant function analysis (DFA) for modelling baseline states of participants to, then, classify high-level states, such as task engagement, or a differentiating factor for low and high workload conditions was validated.

Baseline methods for resting-state neural activity may be of utmost importance for the outcomes of an empirical investigation as a key effect to whether levels of workload may be classified or not; as explicated by Liu, Ayaz, and Shewokis (2017) as the authors proposed a novel model for investigations into neural tissue workload by introducing improvements to the pre-experiment baseline procedures. To validate their model, they employed a multi-modal EEG, fNIRS, and physiological measures, such as heart rate or respiratory monitoring of their participants. To this end, their participants attended an n-back experiment as the researchers utilized their proposed classifier algorithm, in which the baseline was calculated by factoring in the multi-modal inter participant data (i.e., all participants' baselines pooled and factored into one baseline equation) rather than the precedented method of factoring in per participant data discretely. Liu et al. presented their results that their algorithm was more accurate in distinguishing and/or classifying differing levels of workload. Moreover, the term workload under this context has been recently evaluated in the proposal of a dynamic approach of neuro-ergonomics by Dehais et al. (2020), in which case the authors claimed that the elucidation of the interaction between mental workload (neural tissue activation) and the representation of perceived information

may provide a significant step in the neurological understanding of both linguistic and perceptual interaction states of interlocutors. They also argued that previous approaches could not explain the discrepancies of non-linear neuromodulation. For instance, the engagement of automated schemes or well-learned behaviours causing disengagement from a challenging task that continues to require sustained attention to the result of effort withdrawal or preservation.

A dynamical approach to investigating representations of perceived information may be achieved by employing linguistic evaluation of a phenomenon, such as the conversational common ground or the linguistic entrainment studies on, for instance, lexical priming might suggest. Shockley, Richardson, and Dale (2009) discussed behavioural coordination in the context of a conversation setting to ultimately investigate the possibility that interpersonal coordination in conversation was an emergent coordinative structure. The authors reviewed previous literature in their discussion to the conclusion that previous investigations of coherent activation of the motor system during action observation and the communicative common ground based on conceptual representations suggested shared neural resources between behavioural and representational coordination. As such, they proposed that interpersonal gaze coherence was a function of the achievement of common ground, both perceptually and from a linguistic point of view, and in support of the interactive alignment account (Garrod & Pickering, 2009). Similarly, Stephens, Silbert, and Hasson (2010) investigated the conversational common ground for its neural correlates in a speaker and listener setting by employing an fMRI neuroimaging experiment to explicate both temporal and spatial activations among the dyad. Significant results for speaker-listener coupling were reported by the authors that the Wernicke's and the Broca's areas, as well as the precuneus, dorsolateral prefrontal, orbitofrontal, and medial prefrontal cortex all showed high coherence per the present model. Furthermore, the authors conducted a between modality analysis for their speaker-listener and listener-listener results to that the neural activations showed significant overlap, which also highlighted a delay of 1 to 3 seconds in which the listeners' neural activations indicated correlation. However, this delay observed was not constant across neural tissues, and instead, it varied depending on the role and the spatial location within the brains. The authors also reported behavioural measures that highlighted the neural coupling as indicative of comprehension. Stephens et al. finally, evaluated whether the perception of language was enough with a second experiment, in which the listener participants (non-Russian speaking) listened to a Russian recording. The results provided further evidence that communication was a requisite of their proposal for neural coupling, that only the early auditory cortices showed indications of coupling whence the linguistic setting was not shared, for instance when the listeners listened to a language that they did not understand.

From a speech production perspective of linguistic communication, Nozawa et al. (2016) conducted a hyperscanning investigation for interpersonal neural synchronization in a cooperative verbal communication setting to demonstrate inter-interlocutor correlates of social communication. The authors specifically utilized an unstructured verbal communication setting, where a group of four interlocutors shared a common goal in the cooperative task. Their results demonstrated a successful communication setting, wherein the interpersonal neural alignment was increased

during speech, but not during non-speech. Nozawa et al. reported that their multi-party cooperation setting might have been the cause of the present novel significance, assisted by their detailed artefact removal process and the fine-grained temporal resolution utilized for their analyses. Another study on particularly the complexity of utterances, such as deictic references, and alignment was conducted by D'angelo and Gergle (2016) where the authors investigated the utilization of shared gaze (i.e., mutual gaze between interlocutors) for its ability to orientate shared attention (similar with (Hanna & Brennan, 2007), and its effects on linguistic interaction in a collaborative task-sharing setting with a dual eye-tracking study. They also explicated the interaction between the availability of gaze with efficiency in referring expressions through assessing deictic references uttered by participants. The assessment was coded with a measure of linguistic complexity and feature-wise similarity of objects. The results were that the existence of social gaze significantly interacted with the complexity of deictic references and that a higher linguistic complexity in utterances also facilitated more gaze interaction between the interlocutors. The same interaction effect was also significant for the efficiency in deictic references; in that increased efficiency reduced the amount of gaze interaction among the participants.

2.3. Summary

Despite the recent emphasis on two-person neuroscience and the proliferation of hyperscanning studies focusing on various forms of social interaction, there remains a significant gap in understanding how speakers establish a linguistic common ground in ecologically valid, naturalistic conversational settings. Prior linguistic studies on grounding in communication provide a perspicuous setting to explore the neural and ocular correlates of such key aspects of communication, given their robustness in making the emergence of shared concepts explicit in naturalistic dialogue. Although some studies have partly explored the ocular correlates of grounding, and the workload induced by communication problems in terms of hemodynamic responses, there is still room to expand our understanding of inter- and intra-brain responses, as well as ocular activity associated with the establishment and maintenance of common ground in conversation. This gap is especially notable in hyperscanning studies, where the focus on simple and replicable neural and ocular correlates of the gradual process by which interlocutors establish and maintain common ground remain limited. Although some studies have explored the ocular correlates of grounding, providing valuable insights into the dynamics of gaze interaction, and others explored the workload of communication with hemodynamics, there is still room to expand our understanding of both linear and non-linear relationships that characterize distinct conversational roles and processes, such as grounding, in natural dialogue.

In this thesis, we investigate communication as a joint action by employing a multimodal hyperscanning approach with dual-EEG, dual-fNIRS, and dual-eye-tracking to capture reliable and easily replicable neural and ocular measures. By analyzing these neural and ocular dynamics in relation to grounding and assessing linguistic complexity through methods like behavioral analyses and gaze interaction annotation, we aim to correlate cortical activity, frontal hemodynamics, and ocular

activity during the process of establishing common ground in dyadic interactions. Our focus is on the hyperscanning approach taken to explore the intricate dynamics of establishing and maintaining linguistic common ground and forming conceptual pacts. We hypothesize that this process involves measurable neuroelectric, ocular, and hemodynamic correlates, as well as interplay with conversational dynamics, sustained attention, and neuroergonomic factors, such as workload. Measuring these effects through a hyperscanning approach offers the prospect of expanding our understanding of how this process takes place from a multimodal perspective, including both speaker and listener roles. Consequently, this thesis seeks to expand our understanding of how meaning and perception interplay in the joint action of communication.

CHAPTER 3

3. METHODOLOGY

3.1. Experiment Methodology

3.1.1. Experimental Background

For the present thesis, we employed a multi-modal approach in our investigation towards the neural and ocular correlates of grounding in communication. Firstly, we adapted the visual matching task experiment devised in Clark and Wilkes-Gibbs (1986), where participants were originally tasked with ordering Tangram figures resembling human characters (Figure 1) on cards and on a table. We employed a similar experiment setting, again with twelve Tangram figures broadly similar in design to the original work but in a computerized approach and in a co-located dyadic interaction environment for our participants.

This approach was aimed to observe pragmatic effects, such as conceptual grounding, partner specificity, convergence towards expending the least collaborative effort, and entrainment into performing better in a temporally constrained, dyadic, verbal interaction setting; so that we could correlate the hyperscanning results with these a priori established phenomena. To this end, we introduced neuroelectric (e.g., electroencephalography or EEG), ocular (e.g., eye-tracking), and hemodynamic (e.g., functional near-infrared spectroscopy or fNIRS) measures that we recorded in an inter-participant synchronized time-series data format in a hyperscanning setting on top of the behavioral pragmatic metrics, such as accuracy in the matching task, turn taking behaviour, number of utterances per figure, and number of words per utterance, similar to the original work by Clark and Wilkes-Gibbs (1986). The computerized topology of our hyperscanning approach including two EEG devices, two eye trackers, and two fNIRS devices for two participants attending and being monitored at the same time is provided in Figure 4.

We paired our participants into groups of two or a dyad to attend the computerized interaction experiment with two distinct roles within the conversation. To this end, a participant from each dyad was given the director role and the other the matcher role. Directors of all dyads were given the Tangram figures (Figures 2 and 5) in the correct

order as they were tasked with describing these figures in the order in which they were given to their matcher partner. The matcher participants were tasked with listening to these descriptions and converse towards the given common goal of ordering the twelve figures (Figures 3 and 6). These roles were defined as free-form to the participants; in that, the dyad could converse openly and not only the directors but the matchers would also propose descriptions from time to time.

The participants repeated the visual matching task for a total of thirteen trials in two blocks of seven and six trials respectively. The first trial was considered a practice run where the participants were shown the figures for the first time and where they also conversed towards their common goal for the first time. Following this, six more repeating trials were presented to them where they continued their director and matching roles. The number of trials was decided using two pilot experiment runs, wherein the results showed that the practice trial was of utmost importance to the interaction task, and that it could take more than half the total of the experiment run period for dyads that effectively worked towards their common goal. Furthermore, it was revealed that by the third or the fourth trial of the first block, dyads would establish a common ground. As a result, we decided to include this so-called practice run to the analyses, resulting in that the first block of the experiment was seven repeated trials in total. The experiment design captured the entirety of the entrainment process during the first block and also allowed for data recording for at least two more trials following what seemed to be the average amount of trials most dyads would need to establish a linguistic common ground with aim to provide a more stable analysis structure.

In previous work by (Brennan & Clark, 1996; Metzing & Brennan, 2003) the authors reported on partner specific effects in which the grounding established for one part of the task may exhibit carry over effects across experiment conditions, such as role or partner swap conditions. In order to both control and evaluate the social neuroscience factors for this effect, including the possibility for personal performance where one of the participants might be better at this task (i.e., better at describing visuals), we introduced a role-swap condition into our experiment setting. Due to this condition, the dyad swapped their director and matcher roles at the half-way of the experiment, introducing a second block of six trials to the experiment. During the pilot experiments, we observed the carry over effects for the previously established common ground as per the previous work in the literature by Brennan & Clark (1996). As a result, we opted not to extend this second block to a seventh trial similarly with the first block.

This role-swap condition was presented to the participants after a thirty second rest following the first seven trials and as part of it, the first matcher participant became the second director, and vice versa for the first director participant. As part of the role-swap condition, no other conditions are introduced into the experiment flow, such as confederate participants similar with (Metzing & Brennan, 2003) or new Tangram figures; in that, the twelve Tangram figures remained the same and were even shown in the same pseudo-randomized order to the second director as the first director saw them. This was a decision taken during experiment design to not introduce any confounding factor which might have arisen during the first block, specific to the order of the figures. No participants reported their noticing of this pseudo-randomized order being repeated.

We also utilised video and audio recordings from the experiment setting (Figure 12) to observe whether, despite a separator, the participants utilized any other means of communication than verbal interaction, such as with gestures, or account for attempts to cheat by looking at the other's screen. Finally, the participants were swapped seats for every other dyad to prevent device effects due to laboratory equipment constraints where the EEG (one thirty-two and one twenty channel device were used) and fNIRS (a 2Hz and a 4Hz device were used) devices were different models. As a result, if for one dyad the first director sat at the left-hand side, then for the next dyad the first director sat at the right-hand side, and vice versa for the matcher participants.

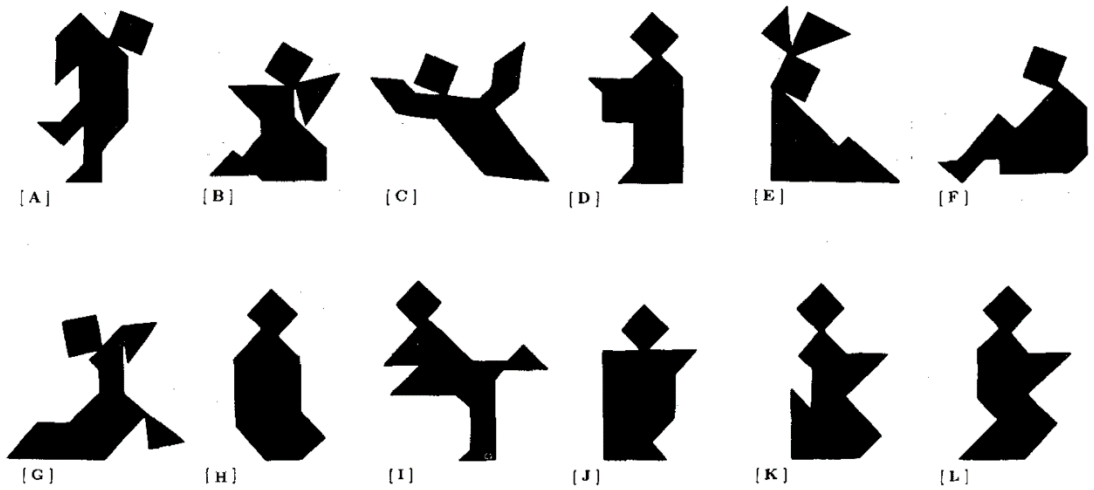

Figure 1: Example stimuli of 12 Tangram figures (Clark & Wilkes-gibbs, 1986).

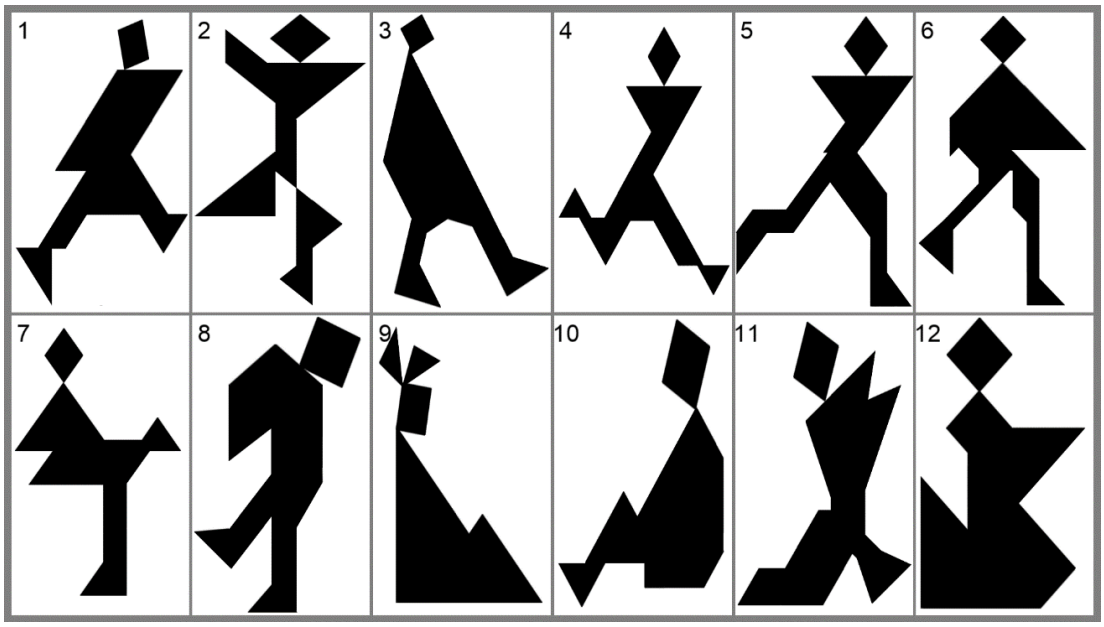

Figure 2: The set of 12 Tangrams utilised in the present thesis, as shown in the director participants' screen.

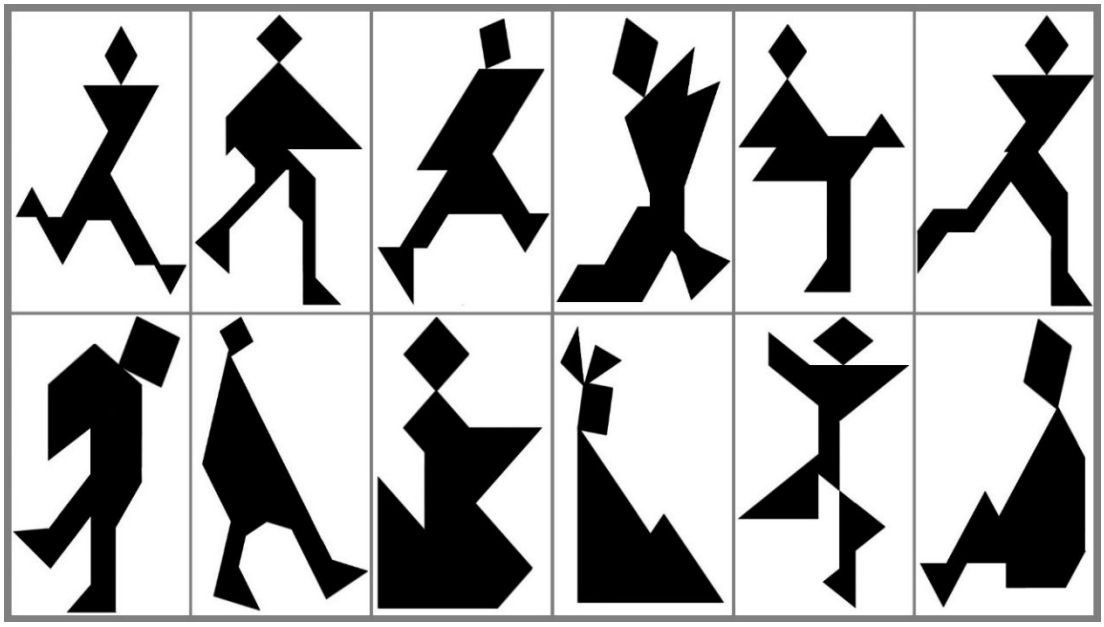

Figure 3: The set of 12 Tangrams utilised in the present thesis, as shown in the matcher participants' screen.

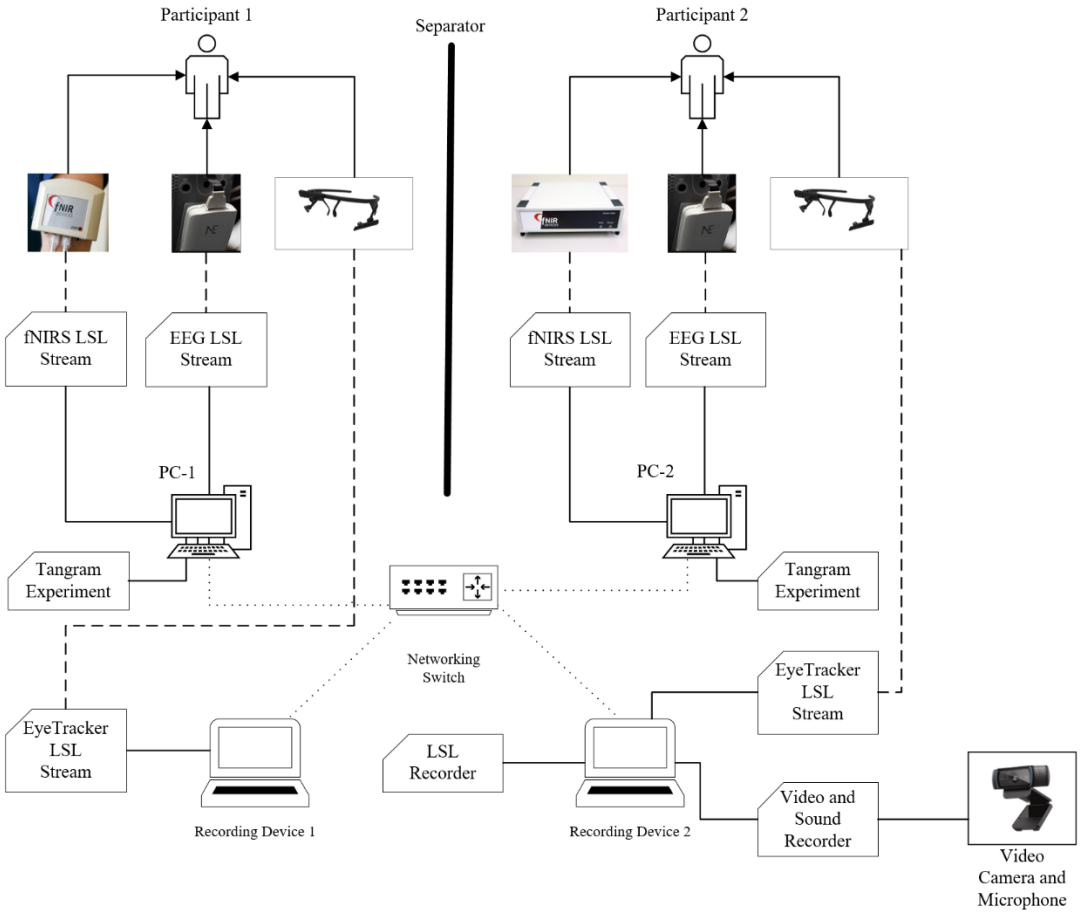

Figure 4: Computerized experiment topology.

3.1.2. Participants

During the data recording period, there were 25 experiment sessions, each of which with 2 participants to the total of 50 participants (M_{age} = 24.22, SD = 3.43). All participants were right-hand dominant, undergraduate or graduate students in the university and native speakers of Turkish. The participants attended the experiment voluntarily and were compensated for the at-the-time equivalent of 5 USD in the local currency. The participants were put in pairs (24 females and 26 males in 14 mixed sex and 11 same sex pairs) and were given the Edinburgh Handedness Inventory (Oldfield, 1971) and a consent form prior to their participation. The handedness results revealed that all participants were either right-handed or ambidextrous (N_{RH}: 44, M_{RH}: 79.12%); this revealed that the brain-to-brain coherence investigation utilizing EEG and fNIRS in the present thesis was not confounded by the variation of lateralization which might have caused by the variation in motor dominance which might interact with neural activations, for instance during a linguistic task such as ours (Gao et al., 2015; Sainburg, 2014).

3.1.3. Inter-Participant Time-Series Synchronization

The experiment software as well as the sensory apparatus (EEG, eye tracker, fNIRS) devices utilised a time-series synchronization framework, the Lab Streaming Layer (Lab Streaming Layer (LSL), n.d.), and fit the experiment setting shown in Figure 12. LSL is a network application layer-based time-series synchronization framework, which allows for centralized, one-click time-series alignment between all multicast streams on a local area network. These steams can be of any data type defined in the framework implementation library file, and in our experiment particularly, all device and experiment marker data were streamed on these multicast network streams for time-series alignment.

Through LSL, the computerized experiment software streamed TTL markers for landmark events, such as trial beginning, ending, mouse click events, as well as directors' described and matchers' selected Tangram order as participants clicked on figures during trials. Finally, for the synchronization of the sound recording and the experiment data (EEG, fNIRS, eye tracker, experiment), the experiment software broadcasted a specific TTL marker, flashed a white fullscreen image, and played a beep noise to be captured by all means of recordings (the video camera, the microphone, LSL Lab Recorder software) for purposes of synchronizing all experiment markers, the video, and sound recordings.

3.1.4. Computerized Experiment Software Suite

The computerized experiment software was prepared in PsychoPy 2020.2.5 (Peirce et al., 2019) experiment builder. We created two versions of the experiment, one for each participant, with one version acted as the experiment server and the other as the experiment client. This also allowed us to control the participants' roles for the role-swap condition; in that, the server side always started with the director participant trials (Figure 2, 5), whereas the client side always started with the matcher participant trials (Figure 3, 6) and each of them swapped the roles after the seventh trial, and following a 30 second rest period (Figures 10 and 11).

For custom code implementation, we took Python script outputs from PsychoPy and edited them in a Python editor to implement inter-participant synchronization features for both experiment behavioral data synchronization with our multi-modal devices and for experiment flow. To this end, the experiment software on each participant station streamed experiment markers for all experiment stages as well as interaction events, such as block beginning and ending point-in-time markers (Table 1). This also included displaying the trial screen of twelve Tangram figures at the same time for both computers and both participants, as well as implementing a mouse click event for each of the figures on the display so that the participants could mark their figures and markers for each click would be streamed to the LSL multicast network. For instance, the Figures 5 and 6 below are taken from the same trial with Figure 5 being the screen displayed to the director and Figure 6, the screen of the matcher. Here, the director participant, having started from the first figure, click on the figure they have finished describing and that their matching partner has selected, turning the figure number blue (Figure 5, leftmost three figures). Similarly, the button click event implemented for the matcher participants allowed their originally unnumbered figures to be ordered (Figure 6, figures which are numbered were the ones that had been clicked). At the point of this example, the dyad has finished matching the first three figures and are in the process of conversing towards matching the fourth one.

The EEG, fNIRS, and eye-tracking devices all used their respective proprietary software suite where possible. This meant that the EEG device (brand and model specified in the hardware back-end section) was connected, configured, and broadcasted the data with their own software, Neuroelectrics NIC2 (Neuroelectrics, n.d.). The 20 (Figure 13) and the 32 (Figure 14) channel EEG devices had a total of 16 electrodes in common, with which we proceeded to the analysis. These channels were P7, P4, CZ, PZ, P3, P8, O1, O2, T8, OZ, C4, F4, FZ, C3, F3, and T7 channels, as shown in Figure 15.

The eye-tracker device also could be connected and configured with using their own software, PupilLabs Capture (Pupil Labs, n.d.-b). However, the Capture software did not come with integration for our time-series synchronization framework LSL so it had to be imported into the Capture software as a Python plug-in obtained open-source through LSL GitHub page (Lab Streaming Layer (LSL), n.d.) and customized for the present experiment setting to provide better stability with the right-eye only Pupil Core devices.

The fNIR devices also came with their own software named COBI Studio. However, COBI studio did not support LSL integration. As a result of this, a custom program was developed which read COBI studio TCP transmissions through the local area network and broadcasted this data with an LSL multicast encapsulation. Another issue with the COBI software was that due to device differences between the 2Hz and the 4Hz versions (Figure 12, 4Hz device on the left-hand side station, to the lower left-hand side corner, and the 2Hz device on the right-hand side station, gray box next to the separator), we were forced to use two different versions of COBI studio, namely COBI Studio and COBI Studio Modern. This cased further fragmentation for the fNIRS data recording settings due to the resulting software difference that the newer 4Hz device and its software supported better adjustment, supported by the newer device's digitally addressable near-infrared LEDs and optodes or light receptors; whereas the old device supported analogue, single channel LEDs and optodes. Following the per participant adjustments, the experimenter launched the custom fNIR LSL broadcaster.

The final software utilised was Open-Broadcaster Software OBS Studio for audio and video recording from the experiment setting. For instance, image in Figure 12 is taken from a video recording of in this setting.

3.1.5. Experiment Procedure

The participants were initially given briefings into the non-invasive characteristics of the EEG with the use of drythrodes or dry electrodes developed by Neuroelectrics for the Enobio EEG devices we used. Then the use and working principle of fNIRS was explained to them that the device emitted photons energy levels were harmless. Finally, the wearable and video based ocular measurement devices were explained. As a result, all measures were taken to ensure that any discomfort participants might have had regarding experiment apparatus was addressed a priori to the experiment. The briefing also provided details into how the devices would be put on the participants by the experimenter and then set up. Additionally, due to the Covid19 pandemic at the time of the data recording, initial information regarding ozone disinfectation conducted before each recording session was provided. By the end of the data recording sessions, no participants reported any discomfort with the devices or concerns regarding laboratory and/or recording conditions.

Following their pre-recording briefings, the participants were randomly seated to the leftwards or the rightwards experiment station (Figure 12, where the participants sat as they wished without experimenter intervention) and were assisted by the experimenter with putting on the EEG cap. The neoprene EEG cap also contained the fNIRS forehead band and its cables, with the cables protruding through the F7 and F8 electrode holes in the cap (Figures 18, 19). The eye tracker assembly was put on top of the EEG cap without touching the electrodes or pushing down on the fNIRS band as well. After all the devices were put on and connected to their respective streaming computers and software suites; for instance, the leftwards participant's EEG and fNIRS devices were connected via 802.11g WiFi and Bluetooth LE, respectively, to data recording computers, with the eye trackers connected to a discrete computer

running Pupil Capture (Pupil Labs, n.d.-b) software with cables. The experimenter, then, evaluated and made the necessary adjustments for the placement and contact for all EEG electrodes. The EEG devices in use (details given in section 3.1.6 Experiment, Computer and Networking Hardware Utilized section) did not support impedance measurement on the electrodes so visual evaluation by the experimenter on the signal data, as well as color coded representation of signal quality algorithmically measured by the proprietary Neuroelectrics NIC2 software were utilised during these adjustments and placement checks.

The experimenter, then, configured the channels of the fNIRS device for appropriate ranges for the measured levels of photons in the reflection from the prefrontal cortex. Due to the unique characteristic of infrared light that light in this wavelength does not penetrate through keratin – the primary protein for hair, and because fNIRS devices always utilize these wavelengths of light emission for under-tissue oxygenation measurement; the experimenter firstly tried to block the most amount of outer light, then adjusts the LED emission power as well as optode reception sensitivity and gain settings, in order to ensure that the data recorded is physiological data and not environmental, nor noise. Particularly blocking of the environmental light sources is assisted strongly by our neoprene EEG cap placement over the fNIRS headband (Figure 18).

Finally, the eye tracking parameters, such as dark pupil detection parameters (Morimoto et al., 2000; Pupil Labs, n.d.-a) were adjusted. However, the eye tracking calibration took place after the data recording session had begun to make sure all possible corrections could be made in post-hoc analyses, if deemed necessary. The eye tracking post-hoc tests and corrections are explained in detail in the 3.2.2.1 Post-Hoc Gaze Data Correction section.

Following this final step, the centralized recording software LSL Lab Recorder, which is distributed under open-source licensing together with LSL was setup for the data recording session. A YYMMDD_Dyad Number_Session Number format was used by the experimenter to organize the recordings (e.g., 560424_24_1 would have stood for 24th of April 1956, dyad 24, session 1). Following this step each LSL broadcast was checked for status; in that four for each EEG device (EEG data, accelerometer data, quality data, and marker data from NIC2 software), one for each eye tracker from the Pupil Capture LSL plugin that was customized, one for each fNIR device, and one for each experiment software was controlled for their status. The experimenter then started recording on each Pupil Capture instance (two in total), started recording for video and audio streams on OBS Studio, started recording on the centralized LSL Lab Recorder software, then notified the participants that they could click to initiate their experiments, which in the background started the network handshake between the server experiment and the client experiment.

This handshake was done through custom network programming in Python so that the two experiments conducted a two-way handshake where the server side (experiment software executed on the first directors' stations) proceeded to wait for a connection from the client side (experiment software executed on the first matchers' stations) to proceed. Following the experiment software connection, the experimenter proceeded with eye-tracking calibration where a nine-point calibration screen (e.g., Figure 7,

Left) was used with moving markers to the four corners, four sides at their center points, and for the center of the screen. Following this, a neuro-physiological resting baseline was recorded for the fNIRS and EEG devices; in that, the participants were told to close their eyes until a beep which came after twenty seconds (Figure 7, Center), then another eyes-open baseline was recorded again for twenty seconds but with the participants staring at a center point fixation cross on their screen (Figure 7, Right). Following the eye-tracking calibration and the resting state neuro-physiological baseline procedures, both participants were given written descriptions of what their shared task and their common goals were in the screen; in other words, the director (Figure 8) participant was given instructions that there would be an array of Tangram numbered figures on their screen and that they should describe these figures in the given, ascending order from one to twelve to their partner. Similarly, the matcher (Figure 9) participant was given instructions that their partner would be describing figures to them in a certain order but that they would see these figures in a shuffled order, which they would sort by clicking to the figure of which they thought was being described. These different parts of the experiment from trial screens for either participant, pre-experiment calibration and rest period baseline screens, as well as the descriptions given initially in the experiments as well as for the role swap condition are as exemplified with screen shots in Figures 2, 3, and 5-11.

Table 1: Experiment LSL markers for time-series synchronization both with devices and inter-experiment stations.

Experiment Marker	Marker Description	Marker Status
0	Experiment LSL Initialize	-
100	Video/Audio/Marker Sync.	start
101	Eyes closed resting baseline	end
102	Eyes closed resting baseline	start
103	Eyes open resting baseline	end
104	Eyes open resting baseline	start
110	Practice Trial	end
111	Practice Trial	start
112	First Block	start
113	First Block	end
114	Second Block	start
115	Second Block	end
121	Stimuli Display	start
[201-212]	Click event for Tangram Number	end
122	Stimuli Display	end
129	Experiment	start

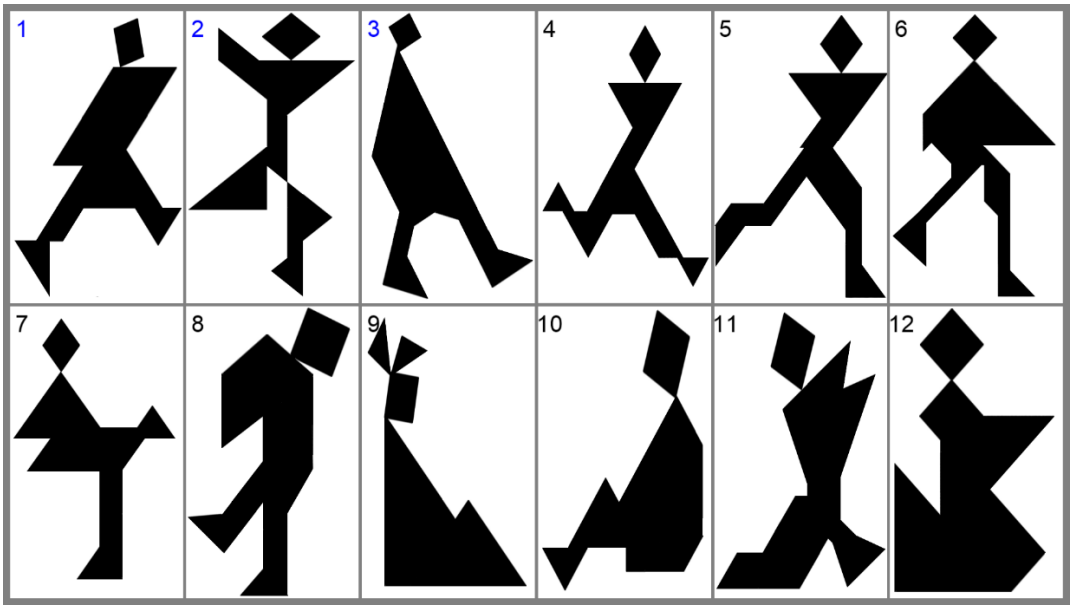

Figure 5: Director participant's screen from a trial, as they have finished describing the third figure and are, then, in the process of describing the fourth.

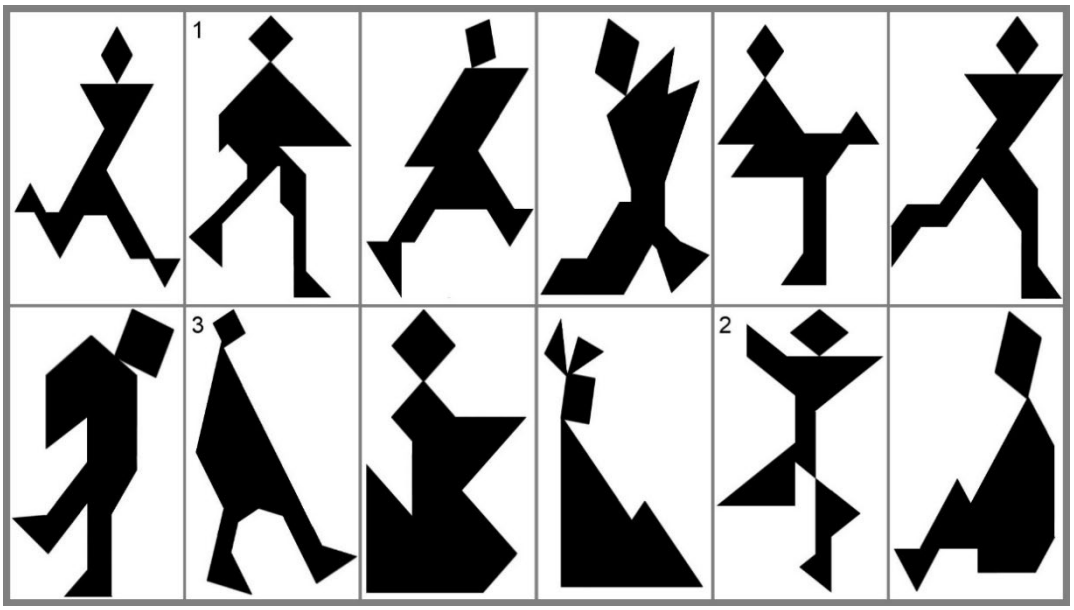

Figure 6: Matcher participant's screen from a trial, as they have matched the third figure thus far.

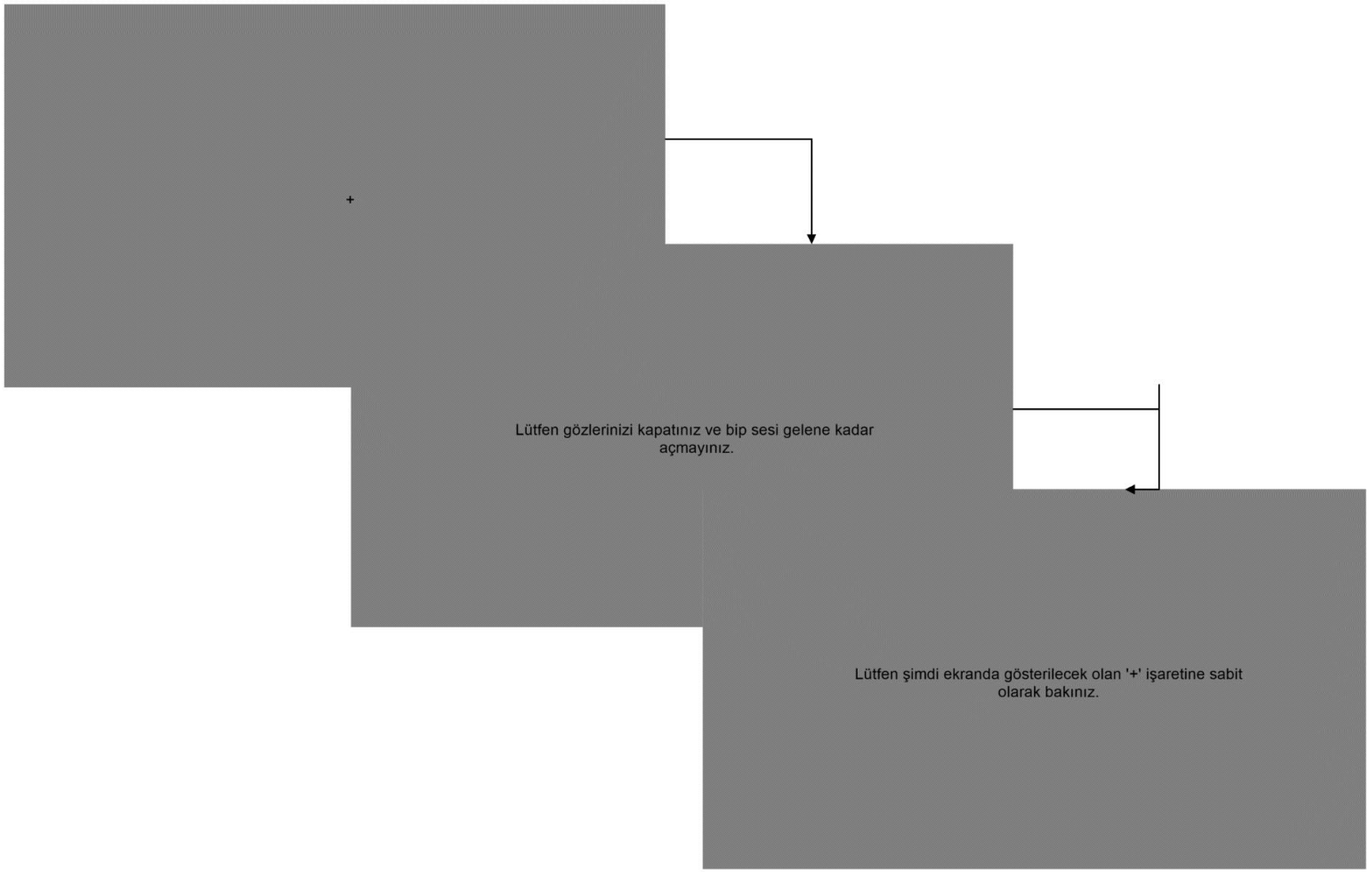

Figure 7: Initial screens of the computerized experiment. Left: Center fixation cross out of the nine total shown for eye-tracking device calibration. Center: Beginning notice for the eyes-closed neuro-physiological baseline (translates to: Please close your eyes and do not open them until there is a beep tone). Right: Beginning notice for the eyes-open neuro-physiological baseline (translates to: Please fixate and hold your gaze at the '+' symbol on your screen).

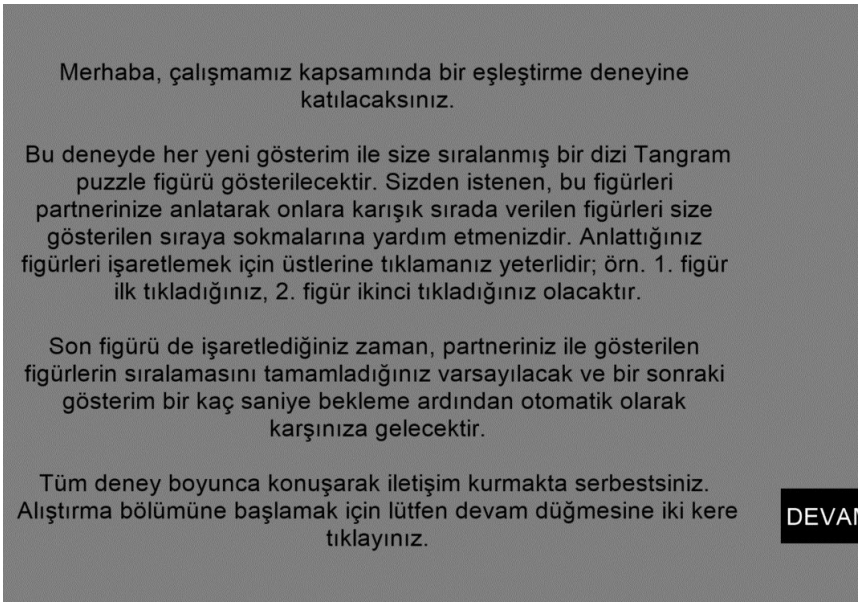

Figure 8: Experiment task and goal descriptions for the director participant in the beginning of the experiment procedure. Summary translation: In this experiment, you will be shown some Tangram figures, ordered with ascending numbers. Your goal is to describe these to your partner in the given order and help them match their shuffled figures with yours in the correct order. You are free to communicate verbally. Please double-click on the "Continue" button to proceed to the practice round.

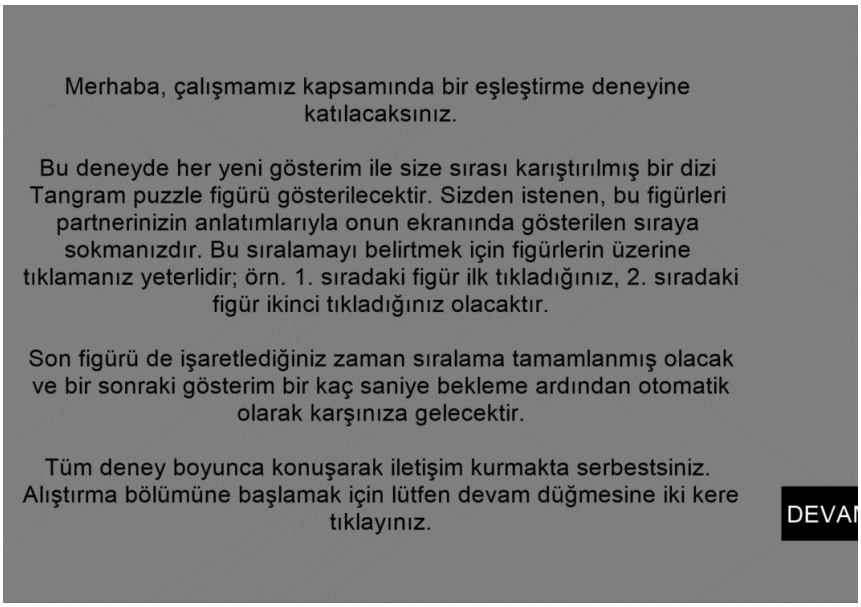

Figure 9: Experiment and goal descriptions for the matcher participant in the beginning of the experiment procedure. Summary translation: In this experiment, you will be shown some Tangram figures in a shuffled order compared with your partner's screen. Your goal is to listen to their descriptions and put the figures to the correct order, as shown on their screen by means of clicking on the figures in order. You are free to communicate verbally. Please double-click on the "Continue" button to proceed to the practice round.

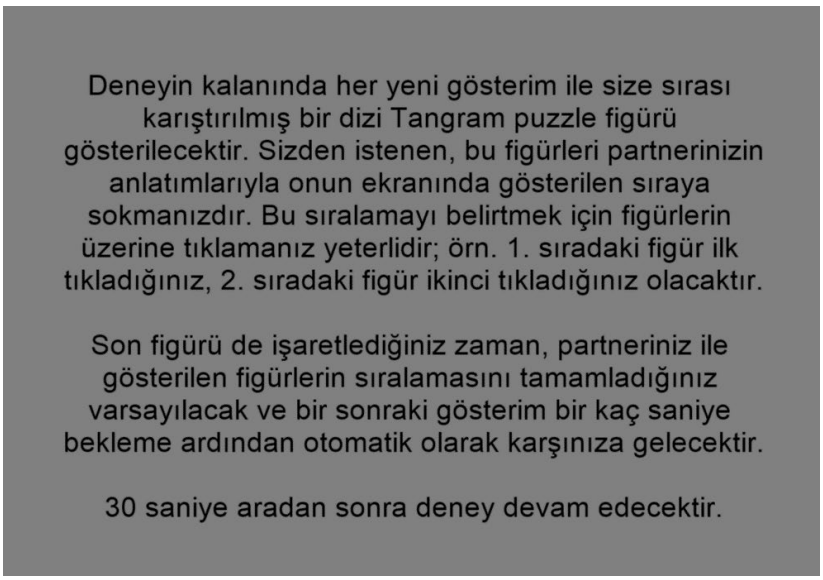

Figure 10: Descriptions for the role-swap condition and for the second block of the experiment shown to the participant who was the first director at the end of the first block (after the seventh trial). For the remainder of the experiment, you will be shown the tangram figures in shuffled order. Your goal is to listen to their descriptions and put the figures to the correct order, as shown on their screen by means of clicking on the figures in order. The experiment will proceed after 30 seconds.

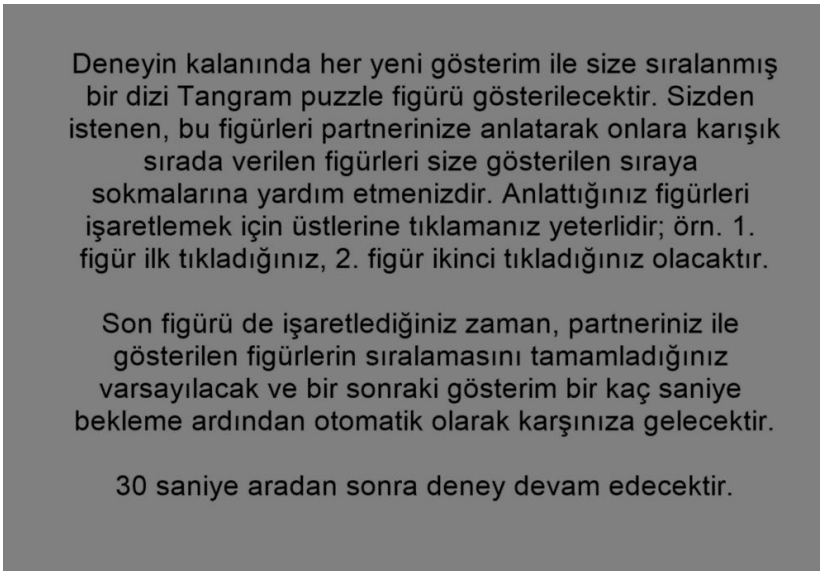

Figure 11: Descriptions for the role-swap condition and for the second block of the experiment shown to the participant who was the first matcher at the end of the first block (after the seventh trial). Summary translation: For the remainder of the experiment, you will be shown the ordered and numbered Tangram figures. Your goal is to describe these to your partner in the given order and help them match their shuffled figures with yours in the correct order. The experiment will proceed after 30 seconds.

Figure 12: The experiment setting with the left-hand side participant holding the director role and the right-hand side participant.

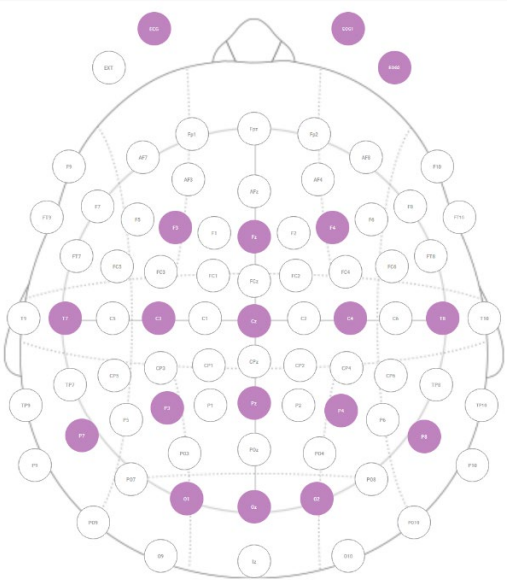

Figure 13: NE Enobio 20 channel EEG recording protocol, electrode sites.

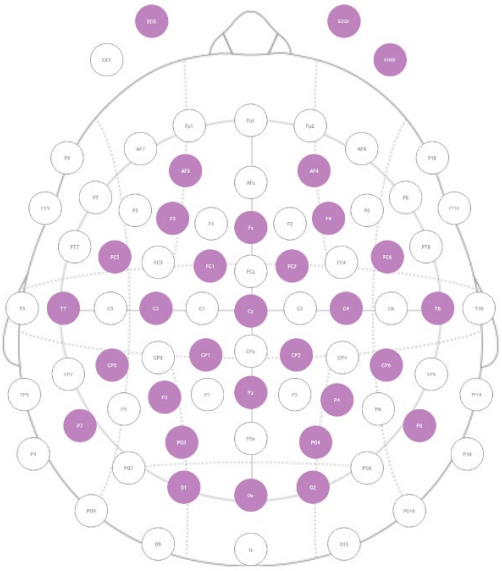

Figure 14: NE Enobio 32 channel EEG recording protocol, electrode sites.

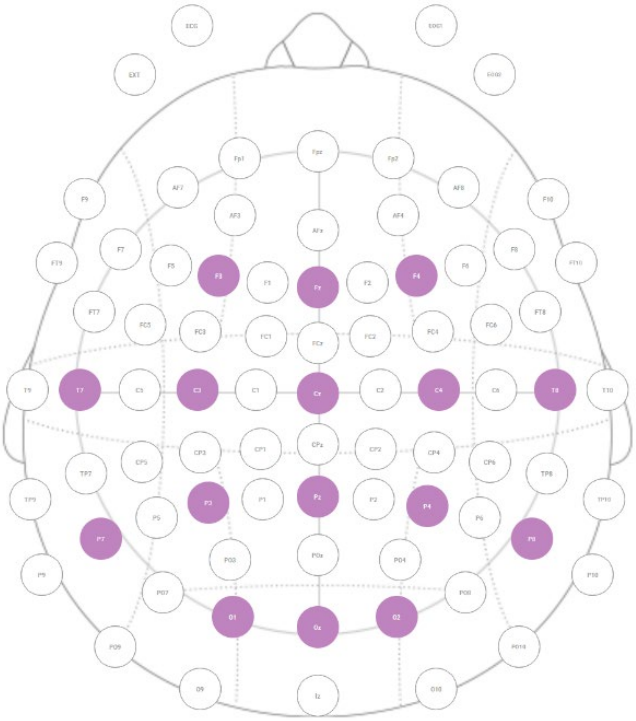

Figure 15: Selected EEG electrode sites for analysis.

Figure 16: fNIR Devices Model 1200/2000 headband and forehead sensor band.

Figure 17: Neoprene EEG cap, fNIRS headband, and wearable eye tracker Pupil Labs Pupil Core worn.

Figure 18: Neoprene EEG cap cabled with dry-electrodes and fNIR forehead band protruding from the cap through F7, F8 channels. Right-hand side view.

Figure 19: Neoprene EEG cap cabled with dry-electrodes and fNIR forehead band protruding from the cap through F7, F8 channels. Left-hand side view.

3.1.6. Experiment, Computer and Networking Hardware Utilized

In the experiment setting topology (Figure 4) it can be seen that we used four computers in total for the experiment. A switch connected all computers to a private local area network via cable, whereas 802.11g WiFi for EEG and Bluetooth for connecting the fNIR devices were employed. The EEG, fNIR, and the computerized experiment software all ran on the two experiment stations. However, due to performance concerns due to high processor utilizations on these PCs, the eye tracker devices were connected using their USB type-C cables to two different computers in the same room as the experiment stations. One of these computers (Recording Device 1 in the Figure 4 topology) ran only the Pupil Labs Pupil Capture software and was responsible with that one eye tracker and its LSL multicast stream to the network. The other PC (Recording Device 2 in the Figure 4 topology) was a slightly higher performance device than Recording Device 1 and it ran the other eye tracker and its multicast LSL stream, as well as the LSL Lab Recorder centralized recording software for all LSL streams, and OBS Studio for video and audio recordings. Below in Table 2 are the hardware specifications of these devices utilised.

Table 2: Experiment topology computer hardware devices.

Hardware	Technical Specifications	
Recording Device 1	Brand / Model	Dell G15 5587
	CPU	Intel Core i7-8750H
	Memory	16GB DDR4
	Storage	240GB SSD Sata
	Monitor	15.6”, 1920x1080 resolution
	GPU	NVIDIA GeForce 1060 Max-Q (6GB GDDR5)
	I/O	USB 3.0 (x3), USB-C (Thunderbolt 3.0 x1), Killer-E2400 NIC (RJ45, x1), Bluetooth 5.1, Intel Wireless-AC 9560, WiFi 5 (802.11ac)
	OS	Windows 10 Pro x64
Recording Device 2	Brand / Model	HP Zbook 15 G5
	CPU	Intel Core i7-8750H
	Memory	32GB DDR4
	Storage	240GB SSD Sata, 512GB SSD PCIe NVMe, 1TB HDD 7200 rpm
	Monitor	15.6”, 1920x1080 resolution
	GPU	NVIDIA Quadro P2000 (4GB GDDR5)
	I/O	USB 3.0 (x3), USB-C (Thunderbolt, x2), Intel I219-LM (RJ45, x1), Bluetooth 5.1, Intel Wireless-AC 9560, WiFi 5 (802.11ac)
	OS	Windows 10 Pro

Router	Brand / Model	Xiaomi AX1800
	Supported Wireless Standards	IEEE 802.11 a/b/g/n/ac/ax
	Transmission Rates	574mbps over 2.4 GHz, 1201mbps over 5.0GHz bands
	Memory	256 MB
	CPU	IPQ6000 quad-core A53 1.2 GHz CPU
	Encryption	WPA-PSK/WPA2-PSK/WPA3-SAE encryption
	I/O	RJ45 for 1000 BaseT for LAN/WAN x 1, RJ45 for 1000 BaseT for LAN x 3

Below in Table 3 are the technical specifications for the Neuroelectrics Enobio 32 and 20 channel EEG devices, the Pupil Labs Core eye tracker, and the fNIR Devices fNIR Model 1200 and 2000M devices.

Table 3: Experiment topology neurophysiological and ocular measurement devices.

Hardware	Technical Specifications	
EEG System	Brand / Model	Neuroelectrics Enobio
	Number of Channels	20 and 32
	Sampling Frequency	500 Hz
	Bandwidth	0-125 Hz (DC coupling)
	Sampling Resolution	24 bits at 0.05 microVolts
	Noise Floor	<1 microVolt
	Input impedance	>1 gigaohm
	Connection	Wi-Fi IEEE 802.11g or USB
	Output	EDF+, ASCII, NEDF standard data formats TCP/IP raw data transmission LSL Server
	Accelerometer	3 DoF accelerometer with 100 samples per second
	Storage	MicroSD Card slot
	Battery	Li-ion chargeable battery
	Operating Time	5.5 hours (WiFi), 16.5 hours (MikroSD), 19 hours (USB)
	Dimensions	89 x 61 x 24 mm
	Weight	97 gr
	Types of Electrodes	Dry*, wet, hybrid electrodes * Utilised in the experiment
	Cap	Neoprene Cap, international 10/20 system
fNIRS System (left-hand side station)	Brand / Model	fNIR Devices 2000 Mobile
	Light Source	LED
	Receiver	Silicon photodiode optodes with integrated trans-impedance preamp

	Wavelengths of measurement	730nm, 805nm (ambient), 850nm
	Number of optodes	18: 4 IR-LED sources 12 long throw (fNIRS standard) optodes pad with 2 short-throw surface optodes pad (Figure 20, bottom)
	Inter-optode distance	25 mm long throw optodes 10mm short throw optodes
	Sampling Frequency	4 Hz
	Interface	Bluetooth LE
	Power Supply	90-264 VAC, 50/60Hz 10W adapter or Standard 10440 Li-Ion batteries
	Dimensions	96 mm (w), 85mm (h), 32mm (d)
fNIRS System (right-hand side station)	Brand / Model	fNIR Devices 1200
	Light Source	LED
	Receiver	Silicon photodiode optodes with integrated trans-impedance preamp
	Wavelengths of measurement	730nm, 805nm (ambient), 850nm
	Number of optodes	16: 4 IR-LED sources 12 long throw (fNIRS standard) optodes pad (Figure 20, top)
	Inter-optode distance	25 mm
	Sampling Frequency	2 Hz
	Interface	USB
	Power Supply	100-240 VAC, 50/60Hz, 15V 1.5A DC adapter
Eye tracker	Brand / Model	Pupil Labs Core
	Tracking Type	Video based dark pupil detection eye tracker
	Gaze sampling freq.	120 Hz
	World video sampling freq.	60 Hz @ 720p 120 Hz @ 480p
	World camera latency	~3 ms
	Eye camera field of view	100°
	Eye camera latency	4.5 ms
	Resolution error	Within 0.02°
	Accuracy error	Within 0.60°
	Recording method	Mobile: Locally via an Android smartphone and Pupil Core application WiFi: Streaming via an Android smartphone and Pupil Core application to Pupil Capture software on PC for wireless recording USB: USB Type-C connection to Pupil Capture software on PC for wired recording

3.2. Analysis Methodology

As a result of our multi-modal experiment design in investigating the correlates of establishing a common ground, a multi-modal analysis methodology is defined in this section. To begin with, we selected the data from twenty-one hyperscanning recording sessions for further analysis. Four recording sessions were eliminated due to networking errors which prevented the two computerized experiments from conducting the "show trial handshake" (more details in the section 3.1.5) designed so that both participants are shown their trial screens of twelve Tangram figures at the same time. This occurrence was limited to these four dyads only.

3.2.1. Behavioral and Linguistic Analysis Methodology

Firstly, the sound recordings for all 21 dyads were transcribed by a transcriber for the selected behavioral statistics, similarly of Clark & Wilkes-Gibbs (1986). The transcription format was structured as number of trial or number of the figure, role, and transcription making a single line of data. In this structure, each trial (1-7 in the first block and 8-13 in the second block) was first denoted and each turn the participants took on the speech floor was transcribed with a marker for their role (D for director, M for matcher) for all thirteen trials. In Table 4, an example of this transcription is given from a dyads' data, where the director explained the 8th figure on their screen (Figure 2 and 5, Tangram number 8) for the first time to their partner.

For the present hyperscanning investigation, the entire transcription was completed by a single expert transcriber with training on annotating Turkish discourse, whose native language was also Turkish. This decision to include only one transcriber was taken to not introduce further variability to the already challenging free-form conversation data that might confound the analyses by introducing inter-transcriber variances in to the transcribed corpora. The transcription conventions are as follows:

Table 4: Example transcription in the given structure. Figure number, participant role, data.

Figure #	Role	Transcription
Tangram #8	D	Sekizinci figür, kafası tamamıyla bir, hani, kare seklinde. Bir ayağı, bir ayağının tabanı yerde, obur ayağının tabanı sol çapraza doğru bakıyor, yani, yukarıda, fakat sol çapraza doğru bakıyor. Sanki dizini çekmiş vaziyette. Kollarını açmış fakat tam bir kol, şeyi, gözükmüyor yani, obur figürlerde olduğu gibi kollar daha fazla acili gibi değil. Üçgen veya ters üçgen veya dik üçgen tarzı bir figür de değil bu, kolları veya gövdesi. Fakat, kafasının sanki düşecekmiş gibi
	M	hmm, evet
	D	yani, gövdesinden. Yani kare seklinde fakat kare seklinde azıcık dokunuyor gövdesine
	M	hmm, anladım, evet bu az önce işaretlediğim o zaman, keşke sorsaydım az önce kafası düşecek gibi mi diye soracaktım, neyse

- The transcriber punctuated with commas as long as the speaker, for instance, expanded the description of the same sub-property of the described figure. For instance, in the utterance:

In Turkish: “Bir ayağı, bir ayağının tabanı yerde, obur ayağının tabanı sol çapraza doğru bakıyor, yani, yukarıda, fakat sol çapraza doğru bakıyor”

English translation: “One foot, the sole of one foot is on the ground, the sole of the other foot is facing towards the left diagonal, i.e., above, but facing towards the left diagonal”

Here, the director starts to talk about the feet of the figure and builds the following representation:

1) This figure has both its feet visible.
2) One foot with its sole on the floor.
3) One foot with its sole towards the left-hand side diagonal but it is slightly higher than the other foot

The initial, lengthy description of the director in this specific example might not fit what is afforded to us as the stature of the 8th Tangram figure but given the remaining figures at the time, along with their discourse up until the point in time at which this description is given appears enough for the matcher participant that the matcher is able to identify which figure this description fit. As such, it is unimportant to the transcriber what or how the dyad communicated; to this end, the transcriber used commas, where the description of the feet (1) of the figure was transcribed verbatim but with the addition of the second (2) and the third (3) properties to the initial first inference (1) that the figure had both its feet visible were considered additions that expanded a description.

- The transcriber punctuated with a full-stop if a speaker switched to a new property entirely for the described figure. For instance, following the previously exemplified utterance, the director said:

In Turkish: “Sanki dizini çekmiş vaziyette. Kollarını açmış fakat tam bir kol, şeyi, gözükmüyor yani, obur figürlerde olduğu gibi kollar daha fazla acili gibi değil. Üçgen veya ters üçgen veya dik üçgen tarzı bir figür de değil bu, kolları veya gövdesi. Fakat, kafasının sanki düşecekmiş gibi”

English translation: “It's like he's pulling his knee. He has his arms outstretched but it's not a full arm, well, you can't see it, it's not like the arms are more urgent like in the other figures. It's not a triangular figure or an inverted triangle or a right triangle, either, the arms or the torso. But you can see that the head looks like it's about to fall off.”

Here, the director expands the total representation of the figure with information regarding the perceived knees, arms, torso, and head descriptions.

These expansions were considered new properties of the total figure than the properties of sub-properties; as a result, they were separated by full-stops.

- The transcriber marked a question mark instead of a full-stop if the semantic nature or the intonation of the last utterance was deemed a query.
- Finally, the transcriber moved to a new "turn" on the speech floor for every time the participant who had the floor (e.g., the director in our two examples so far) stopped their speech following a full-stop punctuation long enough for the speech floor to be freed and that a new turn can be taken. This did not always result in that a new turn would be taken by the other participant of the dyad. Often, instead, if, for instance, the director finished their descriptions, stopped, and received no accepting nor rejecting feedback from their matcher partner, then the director would take the floor again to try to repeat, expand, or clarify their descriptions as per theorized by the likes of (Clark & Brennan, 1991; Grice, 1975, 1978; Horton & Gerrig, 2005).

As a result of these four rules, we obtained twelve Tangram descriptions for each trial, a number of turns taken during the description of each figure, a number of words and a number of utterances (i.e., each new sentence) within a turn. In addition to these, the experiment LSL markers (Table 1) were utilised for behavioral measures of matching accuracy as well as the time it took for the dyad to complete all twelve Tangram figures in each trial. These data driven statistics were obtained using the following methods for each measure.

- **Number of words:** This metric was obtained by counting the number of blanks, in each turn taken as a simplistic measure of complexity within utterances. The final statistical analysis was then conducted by calculating a mean word count value for each trial that consisted of the dyads conversing towards matching and sorting all 12 of their Tangram figures, and analysing this metric using a repeated measures ANOVA test.
- **Number of utterances:** This metric was obtained by counting each full-stop and question mark in each turn the participants took on the conversational floor. It was taken as a measure of complexity of interaction across each trial. Therefore, in the final statistical analysis, the per turn taken utterance counts were summed up to a total for each of the 13 trials.
- **Number of turns taken:** This statistic was obtained by counting each turn taken by participants for each of their trials and calculating a mean value over their conversation to obtain a mean number of turns taken for each figure and for each trial. Whereas the mean word count and mean number of utterances for each trial were complexity within the conversation, the number of turns taken metric aimed to better evaluate the organizational complexity of the dyadic interaction overall. For instance, in the example given above turns taken were as D – M – D – M, therefore from the director to the matcher participant and back, twice, which constituted three turn changes for the 8^{th} figure described. If, instead, the turns taken were D – D – M then this would constitute a single (1) turn change. The decision to not count turns taken and, instead, count the change of turns was taken due to the free-form nature of the conversation given to the participants, which resulted in directors not leaving the speech floor across a number of figures or even for entire trials on occasion,

in later stages of the experiment. This is an occurrence that correlates with dyad's interaction strategies (Horton & Gerrig, 2005) and whether the dyad has established and consolidated their aligned representations into a common ground on the descriptions of figures, as well as whether they have utilised definite descriptions, such as naming the Tangram figures.

Table 5: Example transcription for descriptions of three Tangram figures from an early stage of the experiment. Note that "hıhı" is an affirmation in Turkish.

Figure #	Role	Transcription	Transcription Translation
Tangram #2	D	on birinci figür, yani, şey, bir ayağı yerde bir ayağı havada, şey hani	the eleventh figure, I mean, well, one foot on the ground and one foot in the air, you know.
	M	dans eden dedik biz, bu, böyle koluna, iki kolu da sağ çapraza bakıyor sağ üst çapraza bakan	We called it dancing, it's like this, with its arm like this, both arms facing the right diagonal, facing the upper right diagonal.
	D	yok iki kolu da sağ üst çapraza bakmıyor bunun	No, neither arm is facing the upper right diagonal.
	M	sol kolu bulunan, sol üst çapraza bakan	With left arm, facing the upper left diagonal
	D	ik, iki iki, iki kolu da farklı acı, farklı şeye bakıyor pozisyonda	Tw- two, two, two, both arms in a position of different pain, looking at
	M	tamam	Ok
Tangram #10	D	dört, sağa doğru oturmuş dizlerini çekmiş kendisine doğru	four, sitting to the right, pulling his knees towards him
	M	sırtını sağa vermiş değil mi	He's got his back to the right, right?
	D	aynen, sırtını sağa vermiş, sola doğru bakıyor	Exactly, with his back to the right, looking to the left.
	M	hıhı, tamam	Uh-huh, okay
Tangram #6	D	ikincisi, sağ ayağı tabanda, sağ ayağı tabanda olanları şey yap	Second, right foot on the sole, make those with right foot on the sole a thing
	M	hıhı	Uh-huh
	D	sol ayağı, ssss asa, sol çapraza doğru bakıyor	left foot, d- downwards, looking towards the left diagonal
	M	hıhı	Uh-huh
	D	figür sağa doğru bakıyor gibi	the figure seems to be looking to the right
	M	hıhı	Uh-huh
	D	kafası kare seklinde fakat tam, sim et-, hani elmasımsı bir kare hani yine kare, fakat yamuk yani yamuk dediğim de hani karenin yerleştirilme pozisyonu yamuk değil örneğin	the head is square, but it's a complete, sim-, you know, diamond-like square, you know, square again, but it's crooked, I mean it's crooked, I mean the position of the square is not crooked, for example
	M	evet?	Yes?
	D	simet- olabiliyor	Symmet- can be
	D	üst gövdesi, ters üst gövdesi normal üçgen seklinde fakat üçgenin sol alt kısmi bir yerde kesilmiş gibi	the upper body, the inverted upper body is in the shape of a normal

			triangle, but the lower left part of the triangle seems to be cut somewhere
	M	Tamam	Okay
	D	doğru şeyi mi işaretledin?	Did you mark the right thing?
	M	do? Doğru sanırım	Righ- I think that it is right
	D	Tamam	okay
	M	Anladığım, bu	In my understanding is for *that*

Table 6: Example transcription for descriptions of the same Tangram figures as in Table 5 from a later stage of the experiment. Note that "efe" is a concept to which a Northern Mediterranean person dancing a certain way might fit in Turkish culture; signifying that the dyad has converged into a definite description entailed by a globally available meaning.

Figure #	Role	Transcription	Transcription Translation
Tangram #2	D	üç, efe	The third, efe
Tangram #10	D	dört, sağa oturan sola bakan	The fourth, sitting towards the right side, looking towards the left
Tangram #6	D	beş üçgen gövde solu kesik	The fifth, triangle body with its left side cut off

- **Matching accuracy:** This statistic was obtained using the LSL experiment markers in Table 1. Specifically, the 201 to 212 Tangram click event markers which were programmed into the experiment software such that if both participants clicked on the same figure, the same marker would be transmitted from their software. The LSL time-series synchronized data was parsed for each recording session to obtain one output file per participant station, which were merged into a single file. The time-series data obtained were, then, matched to see how many of the click event markers, then, matched among the dyad for each trial, wherein 12 out of 12 matched would deliver 100% accuracy in the respective trial of the experiment.
- **Time taken to complete trials:** This statistic was also obtained using the LSL experiment markers. Specifically, the 121 and 122 markers (Table 1) where the experiment data obtained through the LSL data parser and was merged into a single file was utilised to calculate the time passed between the beginning (121) and the ending (122) marker for each trial. The decision to use stimuli display begin and end marker was taken, in order to limit participant errors not related with their experiment performance. For instance, sometimes the director participants forgot to click at their already ordered figures as they described and clicked on them yet in the correct order but at the end of the trial at once. Other strategical differences between different dyads also prevented us from conducting a higher granularity analysis with respect to participant matching response times despite the very high time resolution allowed us by the LSL framework. For instance, some participants verbally synchronized their clicks as well, whereas others were, on occasion, too excited to click quickly and made mistakes, affecting their matching accuracy incorrectly, which we wanted to avoid.

In the latter stages of the behavioural annotation process, the transcriber reported subjective effects they could observe that could become further distinguishing factors for the dyads' organizational behaviour. In that, the transcriber reportedly observed subjective differences regarding the dyadic organization of participants for whether they were mixed or same sexes, a factor we analysed post-hoc to ensure we controlled for possible variances in the behavioral data, which is compatible with this type of analysis. To this end, in the post-hoc analysis we investigated possible organizational variances introduced to the interaction setting due to the composition of our dyads, we evaluated each of our metrics with the data organized into three groups; such as, mixed sex dyads, same sex female-female dyads, and same sex male-male dyads.

3.2.2. Eye-Tracking Analysis Methodology

As aforementioned in the methodology section, simultaneous eye tracking data was collected from the two participants during the experiment. Each participant wore a Pupil Core eye tracker with their right eyes tracked. Each eye tracker was calibrated using a nine-point grid of fixation crosses using Pupil Capture software natural features calibration mode; where the participants are instructed to gaze at a cross. For validation and to be utilised in the post-processing for corrections, if necessary, we also utilised salient markers in the surround of the experiment displays. Specifically, we put little post-it papers with numbers from one to four in each corner of the display and also utilised the branding name of the display as a point-of-interest at times.

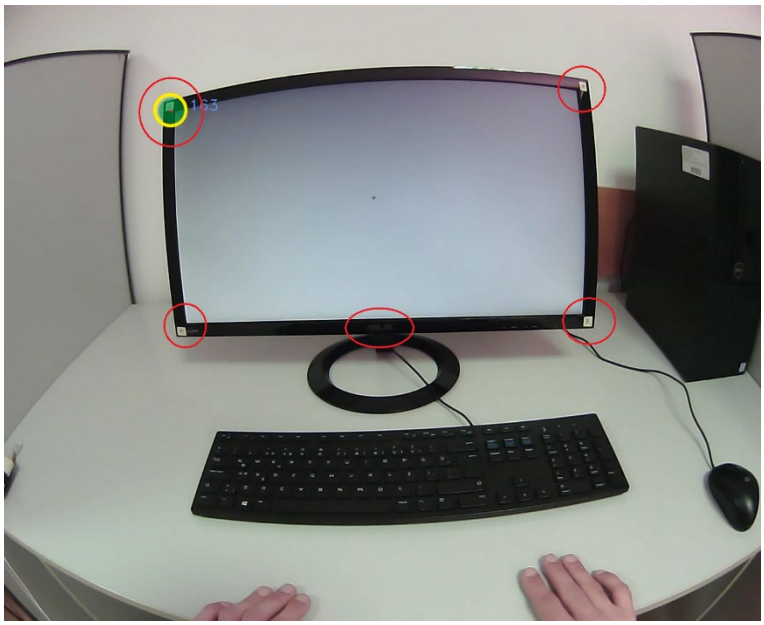

Figure 20: Physical calibration validation markers utilised in the experiment setting for gaze calibration validation and possible corrections in post-processing of gaze data.

3.2.2.1. Post-Hoc Processing of Gaze Data

In the post processing pipeline, the calibration profiles were the first bits of data under scrutiny. In that, firstly, data from the eye tracker was analysed using the Pupil Player software from Pupil Labs for fixation analysis.

In eye tracking research, a fixation is defined as a result of an algorithmic classification process; wherein, the time series data for coordinates recorded from an eye tracker, as well as timestamps of each datum are analysed. This analysis is dependent on the algorithm utilised and method dependent parameters to distinguish when tracked eye movements fall beyond these detection parameters (for instance, in a saccadic movement) or not (for instance, in a fixation).

The most widely utilised examples of these algorithms are I-VT (identification by velocity threshold), I-DT (identification by dispersion threshold), and I-VVT (identification by velocity and velocity threshold), In particular, Pupil Player utilizes I-DT algorithm by default. This is a dispersion-based algorithm that requires a degree of the field of view as a reference, a minimum duration, and a maximum duration parameter. The algorithm works with these parameters to analyse raw eye tracking data to classify when the eye remains static enough to provide gaze samples for at least the minimum duration (ex. 100ms in total gaze duration) and within the degrees of the field of view (ex. 1.5°), and up until a maximum duration. From a perspective of eye tracking and gaze analyses a maximum duration is not necessitated by our oculo-motor mechanisms nor from a visual cognition perspective, but instead, the maximum duration exists more so that the algorithm is computationally tractable. Summarized to that I-DT defines a period and the correlating data which matches the set parameters as a fixation, and all other periods in between fixations as non-fixation (Holmqvist et al., 2011, p. 219).

To this end, in our analysis of participants gaze, we executed the fixation classifier I-DT in Pupil Player with the following parameters:

- 1.5° degrees of field of view
- 100 ms minimum duration
- 4000 ms maximum duration

Following the fixation detection, data from each participant was evaluated on the Pupil Player interface for subjective calibration validation as well as pupil and gaze detection quality by a trained annotator. Three different classification results were possible as the trained annotator decision; firstly, the recorded pupil detection and calibration profiles were deemed reliable and consistent. To check for this the annotator watched the videos for whether the initial calibration, as validated with physical world markers was a good fit for the accuracy and resolution of the Pupil Core device (Table 3). The annotator then watched for whether the participants moved their wearable, glasses type eye tracker after the calibration, which would break the calibration profile. No data was rejected in this manner as a result of post-hoc analysis.

However, the second outcome that was possible as the subjective observations by the trained annotator was when it was deemed that the pupil detection and/or the calibration driven gaze detection and mapping could be improved. In this case, the

annotator carried an offline calibration process, utilizing world video of the calibration process, as well as the physical validation markers in the experiment setting.

In this method, the annotator watches the calibration process as recorded into the world video output of the Pupil Core device in the Pupil Player to conduct point-in-time marking for when the participant gazed at the calibration markers. This is possible due to the Pupil Capture software recorded both the eye video and the world video at the time of recording, allowing the annotator to see and hear when exactly the participant gazed at a calibration marker.

In rare cases of the third outcome, wherein the dark pupil detection parameters of Pupil Capture software at the time of recording appeared problematic, then, the entire detection of dark pupil would also be reinitialized. This process necessitated the Player software to parse the eye video from the recording time for re-detection with corrected dark pupil threshold parameters, again in Pupil Player software. In these cases, the annotator had to repeat offline calibration and gaze mapping processed as well.

Following this, all data including fixation timestamp, coordinates (x, y), and detected fixation durations defined by the I-DT algorithm were exported, along with world videos with fixations mapped for each participant's data. As a result of the Pupil Player side post-processing, data from a total of 21 pair of participants were accepted into the next step of annotating the classified fixations on the twelve Tangram figures.

3.2.2.2. Dynamically Detected Regions of Interest for Annotating Gaze Data for Wearable Eye-Tracking Devices (DDRoIA)

Annotating gaze data for fixation analysis in interaction research poses an analysis challenge. The best practice in this field of research when using eye trackers is to utilise desktop eye trackers and head restraints to ensure that both the participants of the experiment and the experiment stimuli, which they are presented with are stabilized, making automated analysis possible. However, particularly when using wearable eye trackers this process can become an immensely time-consuming task, which is also prone to human-error, due to the fact that both the wearable tracker orientation and the environment in the field of view changing in time, as well as issues with annotation reliability. The manual annotation process, the inconsistencies it can lead to and its challenges have been addressed in existing literature as an issue that can reduce the impact of research conducted and its reliability of results (Ahlström et al., 2021; Barz & Sonntag, 2021; Macinnes et al., 2018). As a result, solutions to this issue have been investigated by the event driven gaze researchers for a while with methods, such as computer vision (Holmqvist et al., 2011, p. 228; Macinnes et al., 2018), head tracking (Holmqvist et al., 2011, p. 228), and deep-learning (Barz & Sonntag, 2021).

In the laboratory where the experiments for this thesis were conducted, we addressed this challenge using a solution grounded in computer vision and machine learning. These methods were selected as the most suitable for our experimental setup, which involved training models to detect areas of interest with participants' postures as well as stimuli display moving between frames of detection in the world video recording

along with gaze-video overlays provided by Pupil Player. This approach was chosen over deep learning methods because it allowed for a more controlled analysis process, particularly enabling a detailed examination of the model training and execution processes.

To this end, we developed our tentatively named Dynamically Detected Regions of Interest Annotator (DDRoIA). This software in its current beta stage utilizes computer vision models for each of the twelve Tangram figures in order to detect each of the figures and match them for whether a fixation collided with any or more of them. We employed OpenCV (Bradski, 2000) and supervised learning to train for the twelve Tangram figures. In that, a HAAR (Viola & Jones, 2004) cascade classifier model for each figure was trained and then executed in serial processing for each frame that contained a fixation for each participants' gaze-video overlay video. Example training images are as shown in Figures 21 and 22, and the training process as well as OpenCV parameters are detailed in the next section 3.2.2.2.2 Training AOI Classification Models. In its current stage, the DDRoiA software was able to annotate about 1.5 fixations a second of data from the present thesis including twelve models in unparalleled execution. While this is already twice to thrice faster than a trained expert human annotator can achieve; we measured that with a high-performance computer dedicated to it, the software was able to parallelize up to 10-12 fixations a second while conducting twelve object detections.

The eye tracking data of the present thesis after 4 dyads were rejected due to problematic data consisted of a total of 21 dyads' data, which resulted in 42 individuals, and a total of 256654 fixations recorded and annotated. DDRoiA allowed us to conduct the annotation process at a fraction of human annotation speed, even when we include the man hours spent for supervised modelling, training and optimization, the total execution period, and the validation process.

Figure 21: Example negative image from the training dataset of one of the 12 Tangram figures. Here, the grey box blocks the image we want the model to detect. Many types of negative images were used, this is one example to one type of them.

Figure 22: Example positive image. In contrast with the negative image, this one contains only the Tangram figure we wanted the model to detect.

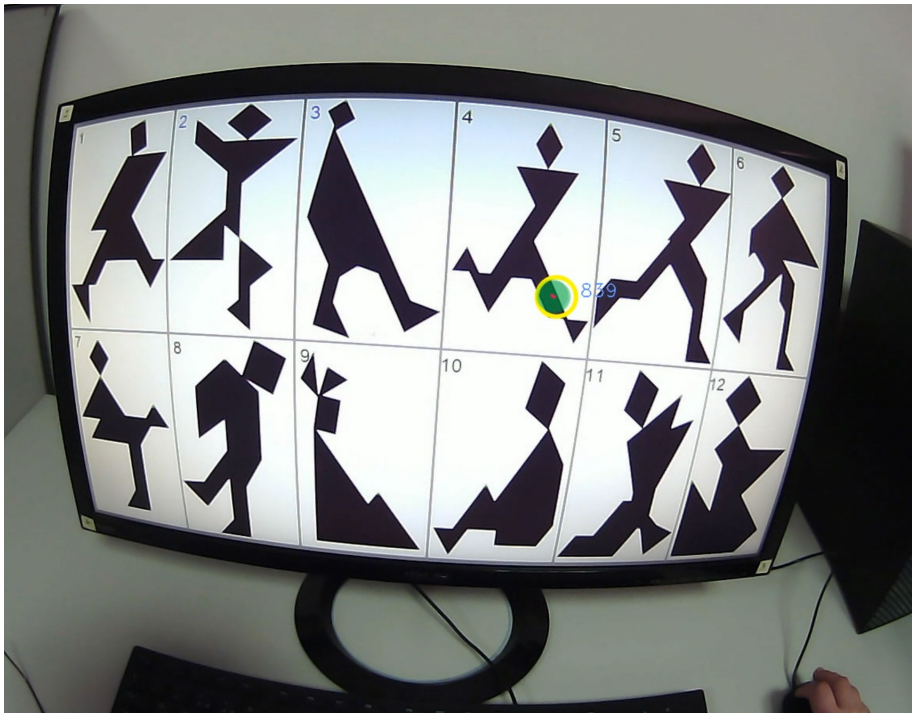

Figure 23: An example still image from the video feed extracted from Pupil Player exported world video file. This image is provided to DDRoiA for detection of the 12 Tangram figures.

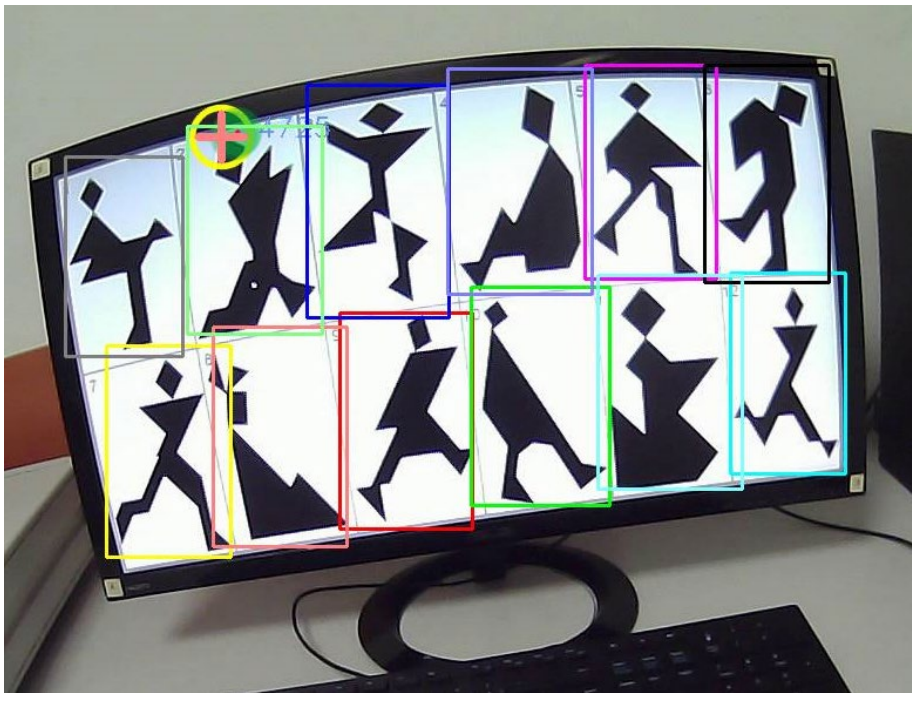

Figure 24: DDRoiA detected models display from an example debug image. Each detection is distinctly coloured and placed in relation with where the detected Tangram figure is located in the video feed.

3.2.2.2.1. Application Methodology for DDRoIA

DDRoiA opens each world video feed and fixations.csv file exported from Pupil Player to compute the first frame refresh of the video feed following the onset of a fixation. This is achievable because the fixations.csv file as exported from Pupil Player includes video framing information. DDRoiA, then, executes each of the twelve Tangram detection models for what each detects on the first frame of each fixation to provide a visual mark up of these frames (as shown in Figure 24 for a base image of Figure 23).

OpenCV library allows for detection using cascade classifier models using three models in its latest iteration version 4.10.0. We employed the detectmultiscale3 function out of the three, which gives an output for confidence levels when multiple detections for a given classifier had occurred. This allowed us to always go for the highest confidence coordinates when drawing the detected Tangram frames.

Once the frames were drawn, we utilised the fixations data and if the coordinates from the specific fixation under analysis collided with any of the frames drawn, then it was recorded in an output file. For collisions with more than one Tangram figure framed, then DDRoIA algorithmically selected the highest ratio of collision among all collisions.

During execution, each cascade classifier was given optimized detection parameters, as required by the OpenCV detectmultiscale3 function (Table 7) as well as the performance requirements set by our analysis schedule. Approximately 90% positive sample detection was targeted during testing to adjust these parameters.

3.2.2.2.2. Training AOI Classification Models

In training the HAAR models utilizing OpenCV back-end for DDRoIA, a trained annotator created 26676 negative (Figure 21) and 4955 positive (Figure 22) for the total sample set containing 12 Tangram figures as they appeared on world videos to provide to the supervised learning method. The negative image samples were divided into two groups; the first, about 1200 random images from other eye tracking video recordings done in the laboratory archives were selected. Secondly, positive samples for the eleven Tangram figures other than the one in the current positive sample set were also provided as negatives. Finally, for the positive images, the annotator cut sub-sampled frames from the world view camera of randomly selected participants to ensure different viewing and/or sitting angles as well as sitting distances from the screen were included in the positive data sets. On average, 400 positive samples per figure were needed to reach positive detection hit rates targeted (Table 8) in OpenCV parameters in tractable number of training stages (set at 20 for the present approach). The tractable number of stages was not hit by any of our training sessions for any of the twelve figures. However, there were a few occasions of non-convergence in the models, resulting in errors where the number of positive figures required increasing.

Table 7: Twelve HAAR Model Testing Parameters that were used in annotation. Minimum neighbor parameter defines the least number of detections, whereas the scale factor specifies the rate in which the image size is reduced at each image scale (1.05 defines a 5% reduction at each scale). For these parameters, minimum neighbor parameter affects the precision of detection, while scale factor has more to do with performance where lower scale factor means lower performance during detection in terms of processing time.

Tangram Number	Minimum Neighbors	Scale Factor
1	15	1.07
2	15	1.07
3	15	1.07
4	15	1.07
5	15	1.1
6	15	1.05
7	15	1.05
8	15	1.05
9	14	1.1
10	15	1.07
11	15	1.03
12	13	1.05

Table 8: OpenCV training parameters for the twelve HAAR models, one each for the twelve Tangram figures. This table denotes that we used 90% of all positive samples for training and left 10% for testing purposes with the positive samples scaled down to 32x48 images, and training continued until the minimal hit, as well as maximal false rates were achieved or until 20 stages of training had passed.

Parameter Value	Parameter Value
Positive images used in training	90%
Positive images used in testing	10%
Negative images used in training	100%
Sample width	32
Sample heigh	48
Feature type	HAAR : ALL
Boost type	GAB
Minimal Hit Rate	99.95%
Maximal False Alarm Rate	50%
Weight Trim Rate	95%
Maximal Depth Weak Tree	1.0
Maximal Weak Trees	100

3.2.2.2.3. Validation Testing and Inter-Annotator Reliability Analysis Results

To determine the quality of annotation conducted by DDRoIA software, we employed a reliability analysis to assess the reliability of annotation agreement among two annotators. The first annotator was an expert eye tracking analyst, conducting the annotation with the twelve Tangram figures for a selection of fixations sampled randomly from the 21 dyads data. To this end, 6 participants (roughly 14% of the total attendance) data were selected and 2849 (roughly 1% of the total) fixations were annotated in total by both the human annotator and the DDRoIA software.

The sole advantage of a human annotator from our perspective is that a human annotator would be able to entrain themselves on which figure the participant last looked at to anticipate which figure they might look next throughout the annotation of the entire recorded data. However, in a complex, free-form experiment setting as is the one in the present thesis, even a human annotator cannot perform better than a single fixation by single fixation basis via this entrainment process, due to the randomness which exists in the nature of verbal linguistic interaction a priori to the establishment of a common ground. We observed that even when also listening to the conversation of the dyad this was the case and therefore would increase the already high risk for annotator inference errors, as well as the required man hours to annotate a given fixation. It is also possible to implement dynamic memory effects at a software level, however, we again opted not to at the current stage of development of DDRoIA due to the possible confounding factors it might introduce. As a result of this thought process, we decided that it was more relevant to validate and assess the performance of DDRoIA in annotating by comparing the annotation it does with an expert human annotator who conducted no inference, and solely annotated on a single fixation basis, just as with the software.

We employed a Krippendorf's Alpha coefficient reliability analysis (De, 2012; Krippendorff, 2011) to compare the annotated samples in IBM SPSS 28 in a coding scheme that solely included human annotator and machine annotator in two columns. The calculated Krippendorff's alpha coefficient for the two raters was 0.8117, the 95% confidence interval for which ranging from 0.7956 to 0.8273 for 1000 bootstrapped analysis, 0.7968 to 8270 for 2000 bootstrapped analysis, and 0.7964 to 0.8266 for 10000 bootstrapped analysis. The alpha value as well as the proximity of increasing bootstrapped analyses indicated a high level of agreement among the raters, suggesting that the annotation was applied consistently by DDRoIA most importantly.

3.2.2.3. Gaze Data Organization

The DDRoiA software was configured to name the 12 Tangram figures with letters from A to L and mark all other regions than the model detected Tangram figures with X. As a result, the software populates a CSV (comma separated values) output file per eye tracker recording the analysing experimenter targets in the following format and with a CSV file name "scanpathOutput.csv":

- A fixation index number the matches the original output fixations.csv file from Pupil Player.
- Fixation onset and offset timestamps from the fixations.csv file.
- Total duration in between onset and offset from the fixations.csv file.
- The area of interest (AOI) identifier that collided with the X, Y coordinates of the indexed fixation.
- The X, Y coordinates of the indexed fixation for debugging when necessary.
- The X, Y coordinates that correspond to the top left hand side corner of the box marking the detected Tangram figure (coloured boxes as shown in Figure xy).

The resulting scanpath output file for each participant were moved along with the experiment data that matched it. Resulting in four files in total per dyad as follows – here, recall that the participants were sat side by side:

- Left hand side participant experiment marker data
- Right hand side participant experiment marker data
- Left hand side participant scanpath data
- Right hand side participant scanpath data

Due to utilization of LSL for time series alignment between all streams and the continuation of this method in Pupil Player and DDRoiA outputs as well, the participants' experiment marker data and the scanpath data were all within the same time series period. The experiment markers were utilized to slice the scanpath data for further analysis at this point; in that, each scanpath file were sliced into 13 slices the parts of which are as follows (recall that the experiment included a role swap condition wherein the dyad swapped their director and matcher roles after the first seven trials):

- First block trials 1 to 7
- Second block trials 8 to 13

Following the slicing operation of the overall data into experiment block slices, the block slices needed to be prepared for analysis. The first step was to normalize the scanpaths for fixation durations so that the resulting strings of characters representing what each participant gazed at in the experiment setting were comparable. This is necessary in a recurrence analysis since the participants spent the exact same time in every trial (due to the experiments being synchronized over the network), their gaze patterns might differ due to many confounding factors we wanted to eliminate. One of these factors was due to the distinction of participants' roles in the interaction setting we employed in the linguistic joint attention situation, where one participant always directed and the other matched. To this end, we defined the following post-processing pipeline and executed in MATLAB.

- The participants fixation durations were adjusted based on the total duration of fixations between the dyad by means of scaling down the durations of the participant with the longer mean duration of fixations by the ratio in which the two differed.
    - This step reduces the error introduced by the two roles condition as well as individual differences significantly. Resulting in strings of scanpaths

comparable with higher accuracy as it allows for an optimal map to be established among the dyad (Holmqvist et al., 2011, p. 347) so that statistical approaches based on the length of scanpath strings; for instance, cross-recurrence quantification analysis (Coco & Dale, 2014; Richardson & Dale, 2005) may be conducted.

- Each AOI identifier was repeated for every 100 milliseconds of fixation duration.
- AOI identifiers were concatenated to result in a single column vector for each experiment block slice.

As a result of this pipeline, the organized data set consisted of 13 trials data for 21 dyads.

3.2.2.4. Gaze Cross-Recurrence Analysis

Cross-recurrence quantification analysis (CRQA) is a method that specifically focuses on the co-visitation principle (Coco & Dale, 2014) to provide a quantitative metric as a result of analysing for complex relationships which might exist within synchronized time-series vectors. In our experiment, this provides an excellent method of exploring the interaction behaviour our dyads might have established as part of their linguistic entrainment into establishing a common ground based on descriptions of Tangram figures.

The CRQA analysis was conducted with high sensitivity for time variance of when the dyad was most in synchrony. To this end, each participant's data for a given experiment block in the order of experiment flow were shifted back and forth four seconds in total, with 100ms steps to result in an analysis that provided 40 recurrence ratio values (Bard et al., 2009; Fındık-Coşkunçay & Çakır, 2022) that ranged from director ahead by four seconds until matcher ahead by four seconds. This allowed us to investigate how the gaze recurrence ratio was affected by the repeated presentation of the Tangram trials by evaluating the the resulting ±4 seconds windows between participants in five distinct sized windows of 8: from director 4 behind to matcher 4 behind, 6: from director 3 behind to matcher 3 behind, 4: from director 2 behind to matcher 2 behind, 2: from director 1 behind to matcher 1 behind, and 0: window for the no lag condition. In addition to this and to provide a control case, similar to (Coco & Dale, 2014; Richardson & Dale, 2005) we utilized a randomized data as a third comparison. To this end, one of the participants' data was shuffled utilizing the same random seed for every experiment block slice of every dyad and included in the statistical analysis. Finally, CRQA results were also observed for which delay and in which direction resulted in the highest rate of recurrence in a global scale.

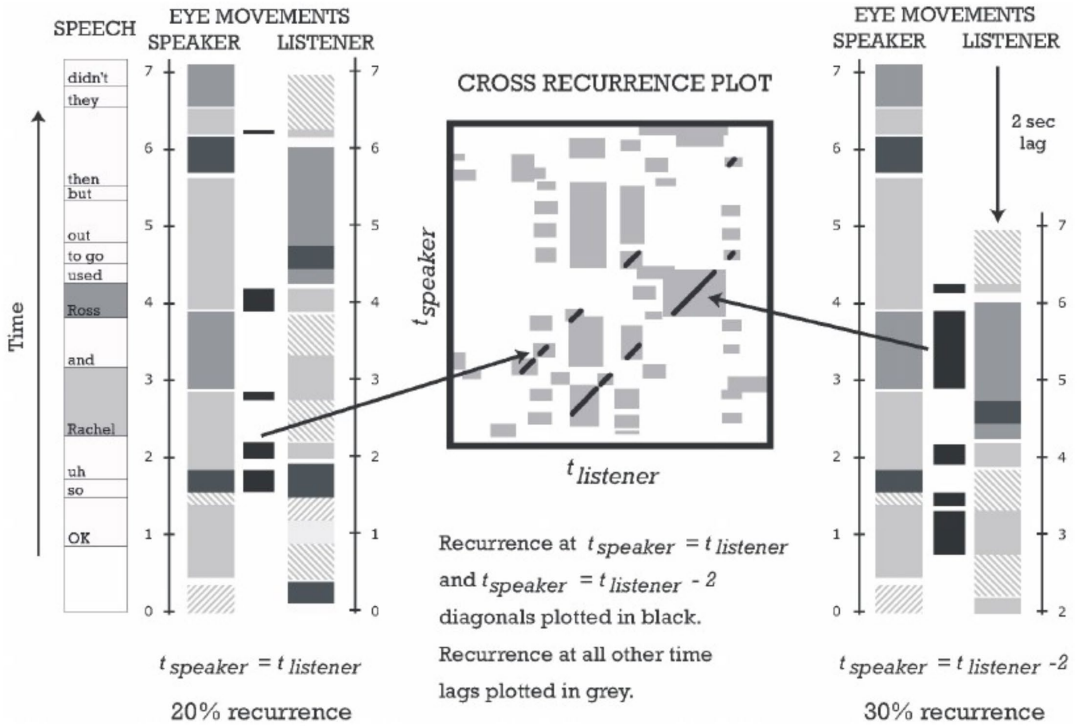

Figure 25: Scarf plot and explanation of cross-recurrence analysis for the exemplified 2 second delay condition (Coco & Dale, 2014; Richardson & Dale, 2005).

3.2.2.5. Gaze Transition Matrix Entropy Analysis

The DDRoiA generated and then organized data was also analysed for transition matrix entropy (TME), which is a measure to analyse for uncertainty in the gaze patterns of each participant. By its nature, as defined by (Holmqvist et al., 2011, p. 477; Shannon, 1948) and in contrast with CRQA, this is a measure to focus on the distribution of data, in this case – a matrix, of a single participant at a time to further evaluate the gaze pattern for each individual for uncertainty or randomness.

To conduct TME analysis, firstly a visitation matrix is constructed for each origin AOI to each destination AOI. In the present thesis, this results in a 12x12 matrix (twelve Tangram figures as AOIs defined in DDRoiA). The matrix is then processed by taking the summation of the matrix, normalizing each transition cell probability, then calculating the entropy with the following formula:

$$H = -\sum_i p_i \log_2 p_i$$

Equation 1: Shannon Entropy formula used to calculate transition matrix entropy for gaze and AOI interaction.

3.2.3. Electroencephalography Analysis Methodology

To approach the EEG data recorded utilizing Neuroelectrics (hereafter, NE) Enobio 32 and 20 channel devices from both participants (layout configurations in Figures 13 and 14 respectively) for a phase and signal synchrony analysis, firstly a pipeline of filtering as well as three methods for analysis were defined. The signal filtering and data organization pipeline consisted of mapping the output data to electrode locations. This is necessary due to proprietary closed source software NE NIC2 that hosted the built in LSL streamers without options for configuration; as such, NIC2 named channels with electrode numbers and not the standardized site name locations per the configured electrode sites in the protocol. We selected matching channels that fit both devices, defined frequency bands of analysis, to then filter and process the raw EEG signals. For the investigation, the immediate overlapping electrode sites were selected with the remaining preserved for future expansion in case of significant amounts of missing data, for instance in the case of from the 32-channel device.

A total of 16 electrodes (Figure 15) from each participant's raw data were selected to match the electrode sites among the dyad as accurately as possible. This resulted in the electrode sites in the international 10-20 system, despite that the 32-channel device handled more data, with a few exceptions as the focus of our present analysis. Due to the forehead band type fNIRS sensor utilised as part of data gathering and experimentation processes the electrode sites Fp1, Fp2, F7 and F8 could not be utilised for EEG electrodes; while the central occipital site OZ, a single channel ECG electrode, and a two channel electrooculogram (EOG) electrodes were added in for signal filtering and/or analyses, such as heart rate variability if deemed necessary in the future following conclusion of the present thesis work. Resulting in the overall list of sites for EEG as shown in Figure 15 and Table 9. We utilised MATLAB 2023a and the Fieldtrip Toolbox (Oostenveld et al., 2011) for initial signal filtering, however, having studied their connectivity functions we decided to implement function for phase locking value (PLV) with custom programming (formula given in Equation 2 below) in order to have better control over the parameters. In the process, we utilised the fieldtrip filter method for delta (0.5Hz – 3Hz), theta (4Hz – 7Hz), alpha (8Hz – 12Hz), beta (13Hz – 30Hz), and gamma (31Hz – 48Hz) band frequencies. Due to the high ecological validity aim of our experiment setting, it was deemed imperative and supported by the literature (Baldwin, 1995; Hanna & Brennan, 2007; Hutchins & Tove, 1999) that the participants sit in a co-located setting for the joint action setting to be at its most possible optimal level; as a result, we were only able to conduct this experiment in a non-isolated room, disallowing for higher band frequency analyses at around and beyond 50Hz line noise.

Table 9: Channels and electrode sites mapping utilised in the present analysis.

C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16
P7	P4	CZ	PZ	P3	P8	O1	O2	T8	OZ	C4	F4	FZ	C3	F3	T7

Consequently, our four blocks defined by an initiation of the linguistic entrainment process or a near zero convergence (warm-up block) period, a high convergence period (the first blocks' latter three trials, as detailed in the behavioural and linguistic results section), a post-role-swap period right after the high-convergence period (the second blocks' former three trials), and a high convergence - control period (the second blocks' latter three trials); resulting in a 4 (experiment blocks) x 5 (EEG band frequency) x EEG electrode sites (16 in total) as our analysis design.

3.2.3.1. Exploratory Investigation

To better understand the spatial distribution of oscillations of both roles in the dyadic interaction setting, mean power values for each electrode and all included dyads were calculated for each EEG band frequency (5). This data was then organized for each experiment role (director, matcher) and each experiment block (4) to be visualized using Fieldtrip Toolbox Topoplots (Oostenveld et al., 2011). This resulted in the Figures 26-35 below. Note that role swap correction has not been applied to these data. As a result, the participant roles are defined as server and client, with server claiming the role of director for the practice and first blocks (the left-most two graphs in Figures 26-35) and the role of matcher for the second block (the right most two graphs); whereas, for the client side participant it is vice versa, with practice and first blocks as matcher participants, and directing in the second blocks of the experiment. The exploratory view of the raw data allows observation of how cortical oscillations shifted spatially for the experiment conditions, experiment blocks and roles.

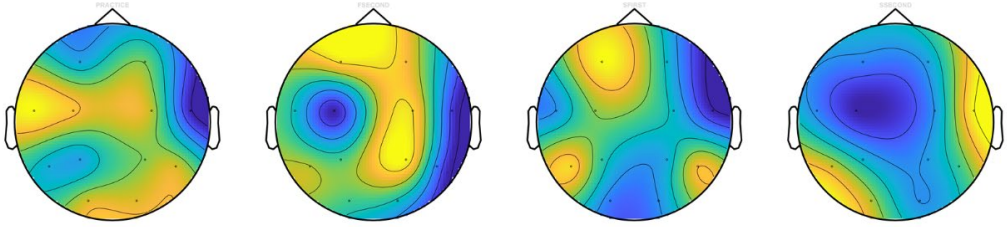

Figure 26: Mean time-domain oscillation powers visualized for the Delta band of 21 dyads with practice, first block second half, and second block first and second halves order from left to right, for the server side participant.

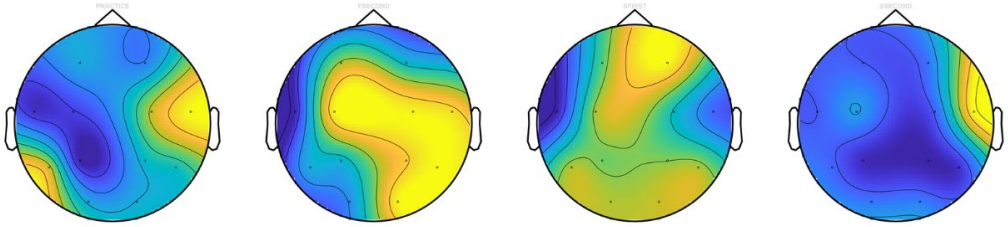

Figure 27: Mean time-domain oscillation powers visualized for the Delta band of 21 dyads with practice, first block second half, and second block first and second halves order from left to right, for the client side participant.

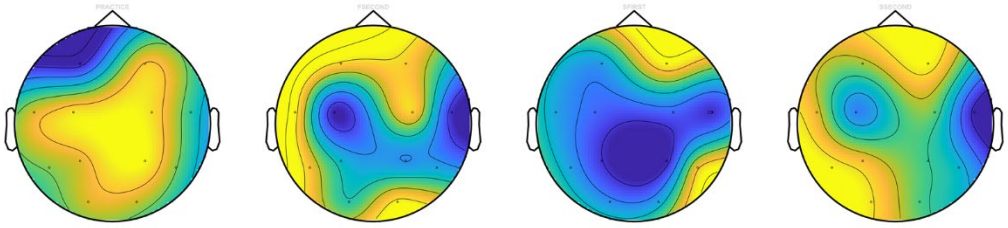

Figure 28: Mean time-domain oscillation powers visualized for the Theta band of 21 dyads with practice, first block second half, and second block first and second halves order from left to right, for the server side participant.

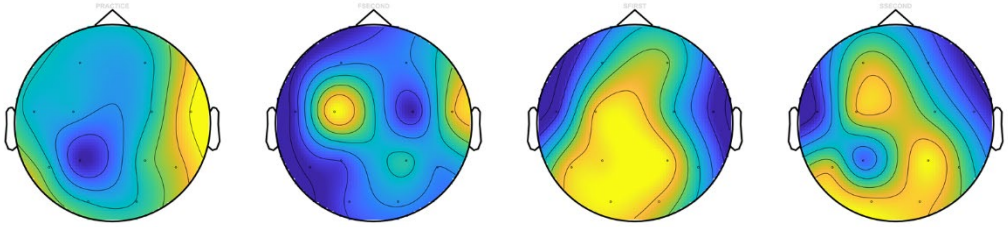

Figure 29: Mean time-domain oscillation powers visualized for the Theta band of 21 dyads with practice, first block second half, and second block first and second halves order from left to right, for the client side participant.

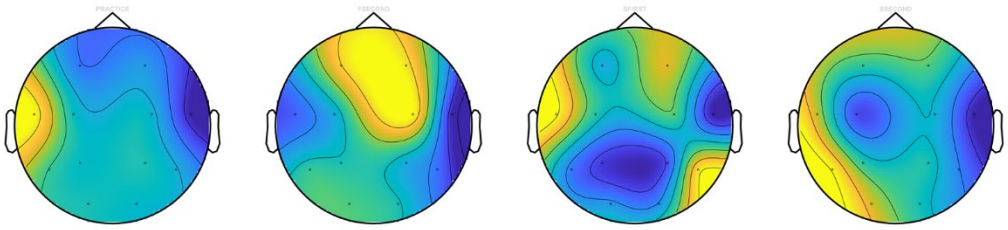

Figure 30: Mean time-domain oscillation powers visualized for the Alpha band of 21 dyads with practice, first block second half, and second block first and second halves order from left to right, for the server side participant.

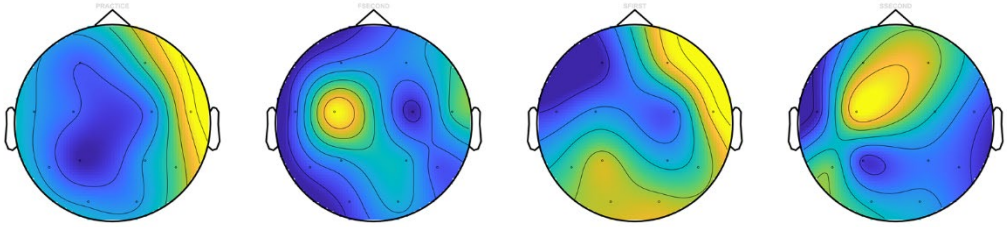

Figure 31: Mean time-domain oscillation powers visualized for the Alpha band of 21 dyads with practice, first block second half, and second block first and second halves order from left to right, for the client side participant.

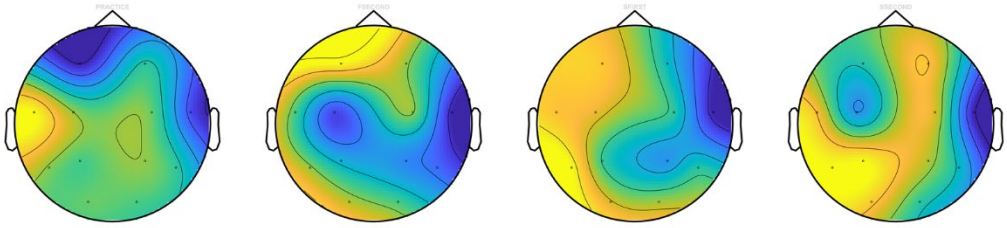

Figure 32: Mean time-domain oscillation powers visualized for the Beta band of 21 dyads with practice, first block second half, and second block first and second halves order from left to right, for the server side participant.

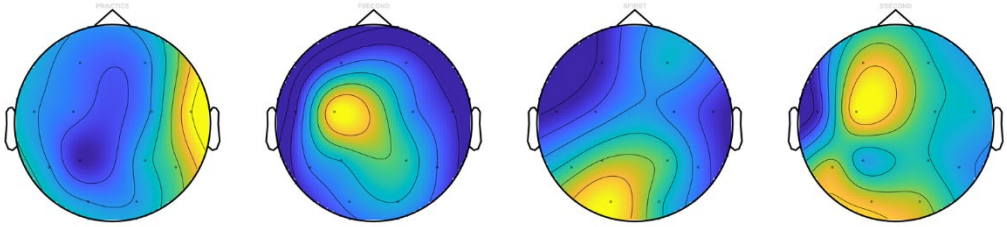

Figure 33: Mean time-domain oscillation powers visualized for the Beta band of 21 dyads with practice, first block second half, and second block first and second halves order from left to right, for the client side participant.

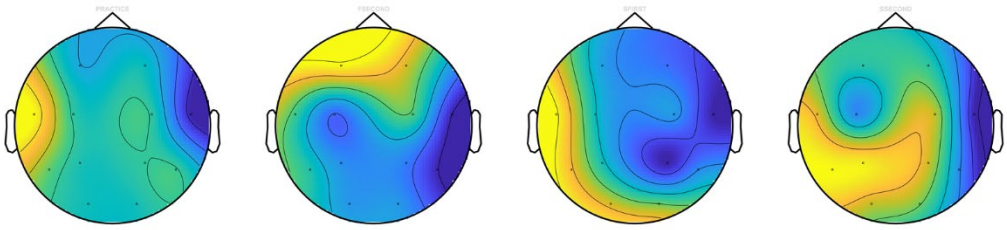

Figure 34: Mean time-domain oscillation powers visualized for the Gamma band of 21 dyads with practice, first block second half, and second block first and second halves order from left to right, for the server side participant.

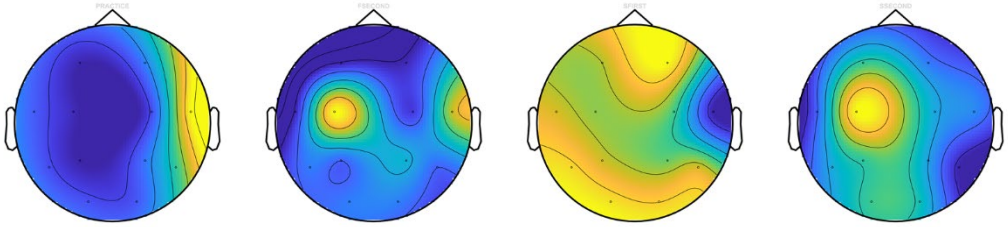

Figure 35: Mean time-domain oscillation powers visualized for the Gamma band of 21 dyads with practice, first block second half, and second block first and second halves order from left to right, for the client side participant.

3.2.3.2. Phase-Locking Value Analysis

Phase-locking value (PLV) is one of the most common methods of analysis for hyperscanning connectivity analysis of EEG neuroelectric signals (Burgess, 2013; Dumas, 2011; Dumas et al., 2010a; Lachaux et al., 1999; Marriott Haresign et al., 2022). It is a method which measures consistency of signals within phase of an absolute delta value for the real and the imaginary signals using the equation given in Equation 1. To execute this equation, the raw EEG filters were again analysed in the 5x5 analysis design, with five blocks and five band frequencies. Each block of data for each participant and for each electrode site were then processed with Hilbert transformation to separate the amplitude and phase of the signals, which then were processed for their instantaneous phase value for each signal at a given time point so that they can be statistically analysed. As a result, a PLV value for each electrode site, for each block, and for each band frequency were obtained and analysed statistically using a repeated measures generalized linear model analysis to investigate experimental effects on the focused electrode sites. The effect sizes obtained for all ANOVA results were then analysed for type 1 errors utilizing False-Discovery Rate (FDR) correction (Benjamini & Hochberg, 1995; Benjamini & Yekutieli, 2001a, 2001b; Lage-Castellanos et al., 2010; Singh & Dan, 2006).

$$PLV = \left| \frac{1}{N} \sum_{t=1}^{N} e^{\mathrm{i}\Delta\phi(t)} \right|$$

Equation 2: Phase-locking value equation, where: N is the total number of samples, $\Delta\phi(t)$ is difference in instantaneous phase between the two individuals at each point-in-time for their EEG data t, i is the imaginary unit.

3.2.4. fNIRS Analysis Methodology

We investigated hemodynamic correlates of the joint task by employing an FNIR Devices Model 2000m and the same brand's Model 1200 functional NIR imaging devices, worn by each participant of the experiment. The data recorded was, firstly put through a post-processing pipeline to ensure that the individual data from two different

generations of fNIR devices were compatible the analysis. In fNIRS data, the raw data is called the light data as it is the direct output from each of the optodes that were adjusted per participant. For our next step in the post-processing pipeline, the light data gathered by each device were loaded into the proprietary FNIR Devices software, FNIRSOFT (Ayaz, 2010). In the FNIRSOFT software firstly a low-pass filter fit for the 4Hz (2000m device) and 2Hz (1200 device) light data recording devices were applied to remove other physiological data from the light signals than blood oxygenation related data, such as respiration and/or heart beat noise (Ayaz et al., 2007).

Due to imperfect recording sessions caused by morphological issues with participants and/or devices, such as participants' hairline in their forehead preventing infrared light from passing through or an uneven forehead morphology preventing necessary contact for either or both the optode and the light source, not all fNIRS data recorded for our experiment had all 16 channels of optodes within the healthy working range recommended for the optodes. In such cases, a trained analyst rejected these channels the light values of which were not reliable functional neural data. In order to circumvent the data loss caused by these missing signals, oxygenation data from 16 channels were grouped into three groups regarding the prefrontal cortex; as left dorsal (Figure 36, optodes 1-4), right dorsal (Figure 36, optodes 13-16), left and right medial frontal cortical (Figure 36 optodes 5-8 and 9-12 respectively) regions.

Using FNIRSOFT, the built-in finite-impulse response filter (FIR) was applied to remove noise artifacts. Following this, a baseline period for each participant's resting state baseline activity were defined as necessitated by the Modified Beer-Lambert Law (MBLL, Ayaz et al., 2007; Chance et al., 1998; Cope, 1991) in order to convert light data into oxygenation data through the formulas outcome of converting light sourced data into oxygenation utilizing this given baseline. This conversion process was again conducted in the FNIR Devices proprietary FNIRSOFT software.

The following analyses of hemodynamic response (oxygenation dynamics) and WTC among the dyad were conducted by employing the frontal four optodes in two prefrontal regions of interest; in that, the medial prefrontal cortex is the locus of attention, due to its known importance in Theory of Mind, and particularly mental representation workloads (Frith & Frith, 1999; Hartwright et al., 2014; Krause et al., 2012; Stephens et al., 2010).

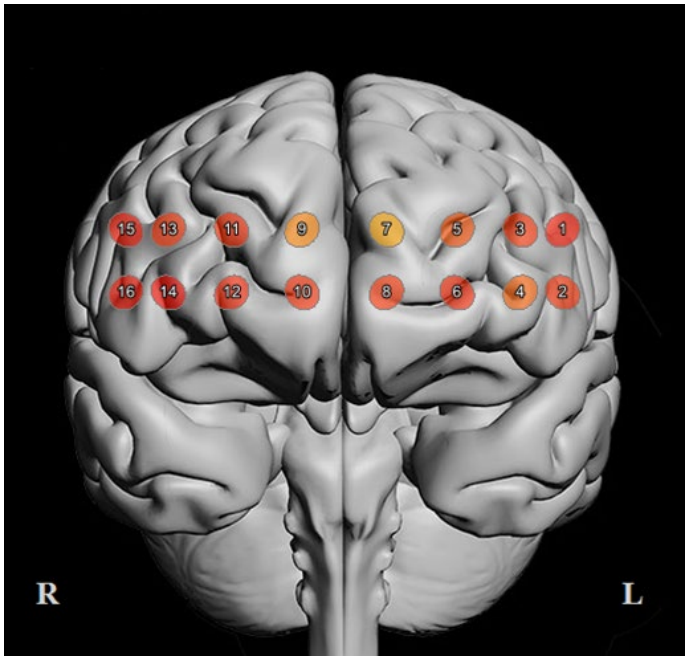

Figure 36: fNIR Devices Model 2000m and Model 1200 optode headbands sensor locations.

3.2.4.1. Wavelet-Transform Coherence Analysis

Wavelet transform coherence (WTC) is a method often utilized to analyse hyperscanning fNIRS data in investigations of cortical neuro-activations from a functional perspective, for instance the hemodynamic correlates in a dyadic interaction setting, such as the one in our experiment setting with the Tangram matching task. To conduct WTC analysis, firstly, oxygenation data from FNIR Devices 2000m model device which had a sampling frequency of 4Hz was down sampled to 2Hz to match data from Model 1200 for the both frequency and time domain related analysis method WTC namely the Wavelet Transform Coherence (WTC) analysis (Cui et al., 2012; Grinsted et al., 2004).

In the implementation of this analysis, we utilised MATLAB 2023a and the built in WTC function with the two participants' data for the x and y axes, Monte Carlo count function enabled for determining a reliable significance threshold for the data, and 128 for maximum scale setting. The function was called such that we obtained an output file for each of the four focus blocks (left and right regions for both dorsomedial and dorsolateral regions) as well as each of the 16 optodes. An output file was given regardless of whether there was data or not as in the case with no data from an optode due to rejection in the post-processing pipeline, then the output became an empty file due to the functional programming implementation.

For analysing the WTC results, we organized the results such that for each of our experiment blocks (resting baseline, practice, first block second half trials, second block first and second half trials) x each available optode (number varied from one dyad to another) + the four regions of interest (left and right, dorsal and medial frontal

cortices) constituted to at least a 5x4 analysis design to be evaluated with a repeated measures ANOVA analysis in IBM SPSS 28.

3.2.4.2. Hemodynamic Response Analysis

Analysing for single-participant hemodynamics is crucial in investigating two aspects to the present thesis. Firstly, the main experiment condition of continuously matching and sorting the twelve Tangram figures in repeated trials might entail a metric for physiological workload reflected in the oxyhemoglobin (HbO) concentration of our interacting dyads. Secondly, this effect might again interact with the distinct roles assigned to dyads as well as with the role-swap condition.

To this end, the oxygenation files resulting from the FNIRSOFT software MBLL formula were organized into matching the present experiment design similarly with the WTC analysis. The analysis design for HbO metric was a block based analysis. As is the case recommended with the hemodynamic response, a wide period metric. To obtain blocks with similar durations, the focused periods were: A low convergence period – the first half excluding the warm-up trial of the first block before the role swap condition, a high convergence period – the second half of the first block, a control period right after the role swap – the first half of the second block, and a re-convergence period – the second half of the second block periods.

Due to the device issue resulting in inconsistent WTC results, we conducted a single participant analysis for the hemodynamics, focusing solely on the functioning device without artifacts. Furthermore, similar with the WTC analyses, to counter the missing optodes that can result in inconsistent or loss data due to improper mounting, low hairline, or uneven forehead skull morphology of the wearable headband on participants foreheads; we utilised a grouping technique again to utilize left and right dorsal and medial regions respectively along with all the available optodes in the hemodynamic response analyses. In which, data from 17 dyads was evaluated with a repeated measures ANOVA analysis in IBM SPSS 28 and with an analysis design of dyad, role condition (1 – director or 2 – matcher), and experiment block condition, for all available optodes, as well as cortical regions of interest. Consequently, resulting in a 2 (roles) x 4 blocks x (16 optodes where available + 4 cortical regions) investigation.

CHAPTER 4

4. RESULTS

4.1. Behavioral and Linguistics Analyses Results

The behavioral data collected through the synchronized matching task experiments consisted of five measures in total. Mean number of words and mean number of utterances for each trial, number of speech-turns taken for each Tangram figure within trials, dyad's accuracy in matching the twelve Tangram figures as well as the time in seconds that it took the dyad to complete their trial were these five measures.

4.1.1. Results of the Analysis for Accuracy of Dyads in the Matching Task

The 21 dyads' accuracy in matching 12 Tangram figures in each of their trials constituted the data selected during the initial post-hoc analyses period for this analysis and a repeated measures ANOVA analysis was conducted. The results showed that there was a significant improvement in accuracy across repeated trials, with Greenhouse-Geisser corrected value F (2.807, 47.724) = 6.801, $p < .001$, $\eta^2 = .286$, and with the mean values shown in Figure 37.

The post-hoc grouped analysis of accuracy results showed that significant effects remained for the mixed dyad pairs, while for the male-male and the female-female pairings the effects turned into non-significant values for the main effect of repeated trials: F_{mixed} (12, 108) = 5.331, $p < .001$, partial $\eta^2 = .372$, $F_{M\text{-}M}$ (12, 48) = 1.492, $p = .160$, partial $\eta^2 = .272$, $F_{F\text{-}F}$ (12, 24) = 1.000, $p = .478$, partial $\eta^2 = .333$.

4.1.2. Results for the Analysis of Time Taken to Complete Each Trial

The total time the 21 dyads spent to complete a trial were analysed by means of a repeated measures ANOVA test. Firstly, the Mauchly's test of sphericity showed that the assumption for sphericity was violated $\chi^2(77) = 438.621$, $p < .001$; therefore Greenhouse-Geisser results were reported. The results of the analysis showed that

there was a significant decrease in time taken to complete each trial F (2.613, 44.421) = 105.462, $p < .001$, $\eta^2 = .861$, and with the mean values shown in Figure 38.

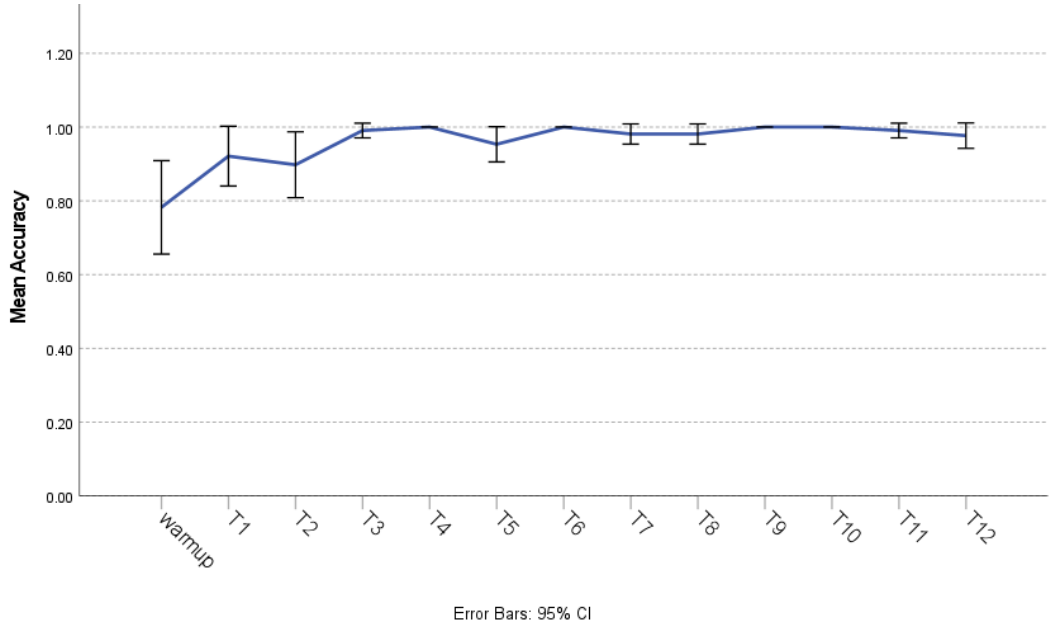

Figure 37: Mean accuracy of 21 dyads in completing 13 trials of matching and sorting 12 Tangram figures.

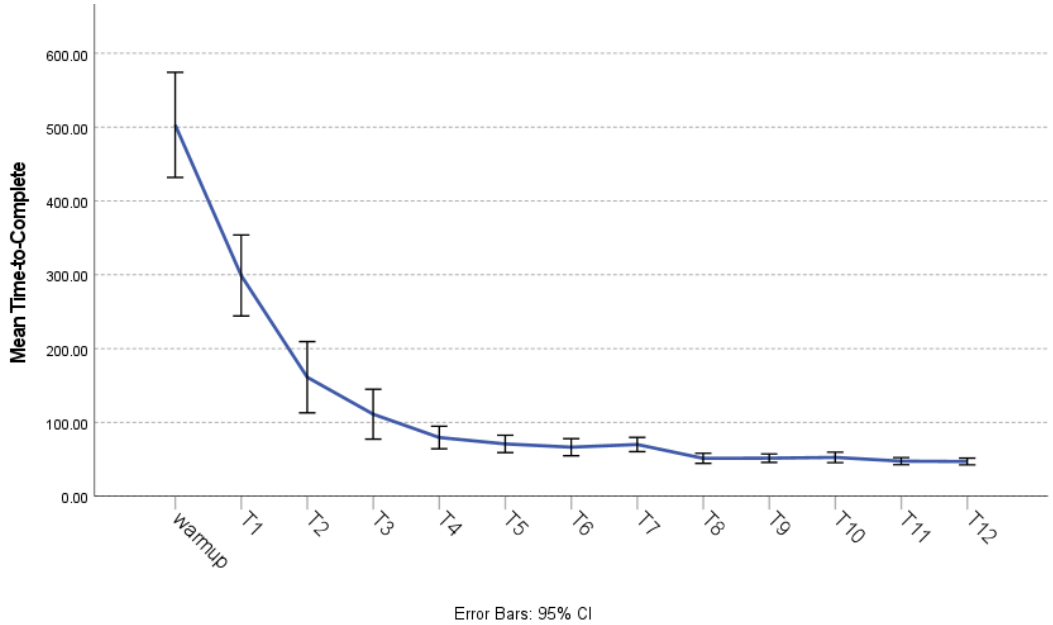

Figure 38: Total time taken to complete of 21 dyads in completing 13 trials of matching and sorting 12 Tangram figures.

The post-hoc analysis for organizational variance, the results all remained significant for the dyads with distinct participant organizations: F_{mixed} (12, 108) = 72.838, $p <$

.001, partial $\eta^2 = .890$, $F_{M\text{-}M}$ (12, 48) = 19.877, $p < .001$, partial $\eta^2 = .832$, $F_{F\text{-}F}$ (12, 24) = 18.315, $p < .001$, partial $\eta^2 = .902$.

4.1.3. Results for the Mean Number of Turns Taken for the Conversational Speech Floor

The mean number of times the dyad changed who had the conversational speech floor is a metric of complexity in organizational characteristics for the dyad. We investigated this metric by utilizing a repeated measures ANOVA test. Firstly, the Mauchly's test of sphericity showed that the assumption was violated $\chi^2(77) = 361.094$, $p < .001$; therefore, the Greenhouse-Geisser corrected values were reported for the analysis.

The results showed that as trials progressed, there was a significant decrease in number of turns taken to complete each trial F (2.680, 45.558) = 71.156, $p < .001$, $\eta^2 = .807$. The mean values for the turn taking count are given in Figure 39.

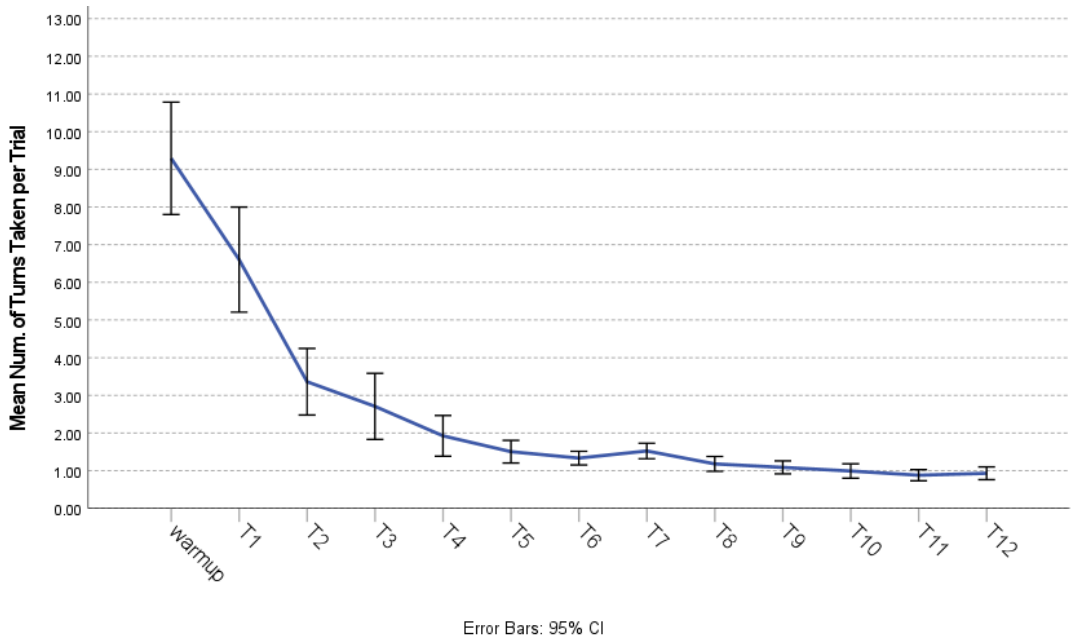

Figure 39: Mean number of conversational speech floor turns taken by 21 dyads in completing 13 trials of matching and sorting 12 Tangram figures.

The post-hoc analysis for organizational variance, the results all remained significant for the dyads with distinct participant organizations: F_{mixed} (12, 108) = 54.093, $p < .001$, partial $\eta^2 = .857$, $F_{M\text{-}M}$ (12, 48) = 15.479, $p < .001$, partial $\eta^2 = .795$, $F_{F\text{-}F}$ (12, 24) = 6.168, $p < .001$, partial $\eta^2 = .755$.

4.1.4. Results for the Total Number of Utterances

We employed a repeated measures ANOVA analysis to investigate the change in the total number of utterances the dyad conversed in matching 12 Tangram figures that constituted a single trial. The Mauchly's test of sphericity showed that the assumption was violated $\chi^2(77) = 361.094$, $p < .001$; therefore, the Greenhouse-Geisser corrected values were reported for the analysis. The results of the analysis showed that dyads' total number of utterances was significantly decreased as repeated trials continued F (2.650, 45.051) = 86.626, $p < .001$, $\eta^2 = .836$, with the mean values for the metric shown in Figure 40.

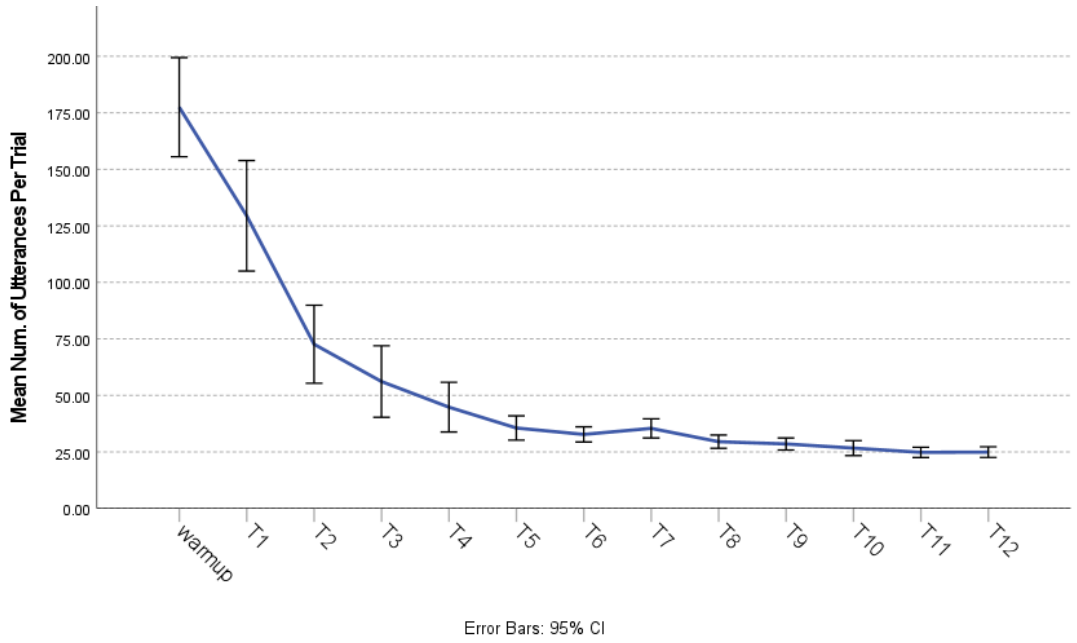

Figure 40: Mean count of utterances 21 dyads uttered in completing their 13 trials of matching and sorting 12 Tangram figures.

The post-hoc analysis for organizational variance, the results all remained significant for the dyads with distinct participant organizations: F_{mixed} (12, 108) = 71.348, $p <$ .001, partial $\eta^2 = .888$, $F_{M\text{-}M}$ (12, 48) = 15.812, $p < .001$, partial $\eta^2 = .798$, $F_{F\text{-}F}$ (12, 24) = 9.527, $p < .001$, partial $\eta^2 = .826$.

4.1.5. Results for the Mean Number of Words in Each Utterance

Analysing for the mean number of words in each utterance is the top-level metric we are investigating for the behavioral interaction perspective to our verbal conversation setting. To this end, we conducted a repeated measures ANOVA analysis. The Mauchly's test of sphericity showed that the assumption of sphericity was violated for the present data set $\chi^2(77) = 171.667$, $p < .001$; therefore, the Greenhouse-Geisser

corrected values were reported for the repeated measure ANOVA results. The results of the analysis showed that during the repeated trials the mean word length for utterances in the experiment was significantly decreased F (2.967, 50.432) = 52.604, $p < .001$, $\eta^2 = .756$, with the mean values for the metric shown in Figure 41.

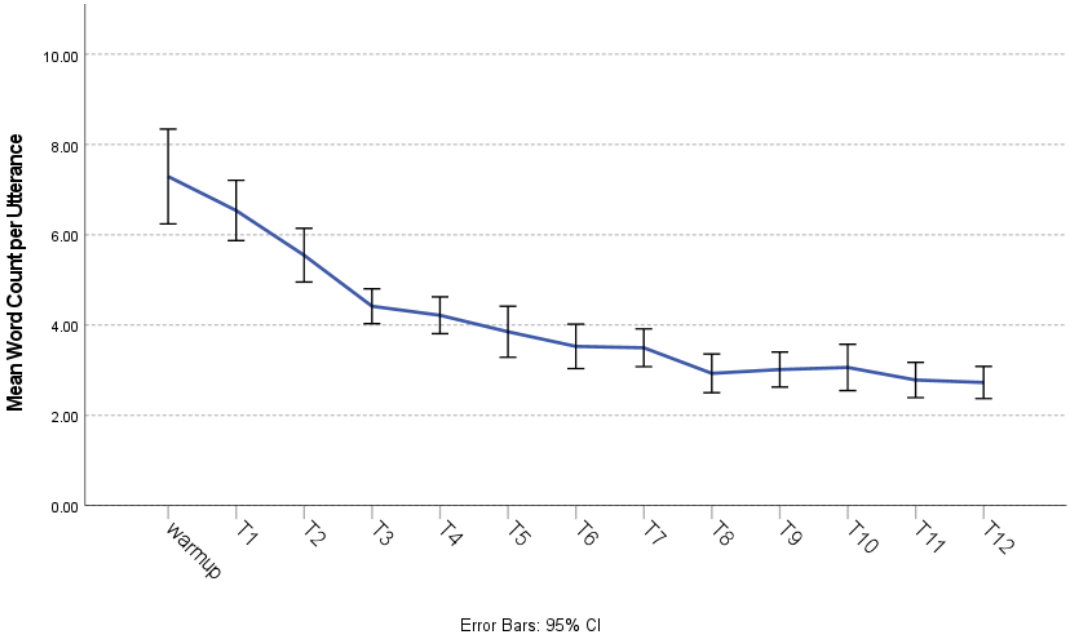

Figure 41: Mean number of word length of utterances for the 21 dyads in completing their 13 trials of matching and sorting 12 Tangram figures.

The post-hoc analysis for organizational variance, the results all remained significant for the dyads with distinct participant organizations: F_{mixed} (12, 108) = 32.012, $p <$ .001, partial $\eta^2 = .781$, $F_{M\text{-}M}$ (12, 48) = 15.140, $p < .001$, partial $\eta^2 = .791$, $F_{F\text{-}F}$ (12, 24) = 7.763, $p < .001$, partial $\eta^2 = .795$.

4.2. Eye-Tracking Analyses Results

4.2.1. Cross-Recurrence Quantification Analysis (CRQA) Analysis Results

To analyse the results of recurrence rates obtained, we employed a 5x13x2 design in a repeated measures ANOVA analysis; wherein, each dyad data was organized for the no delay and one, two, three, and four second time delay periods of evaluation in each direction for all 13 trials repeated measures as exemplified in (Coco & Dale, 2014; Fındık-Coşkunçay & Çakır, 2022; Richardson & Dale, 2005) and for both natural and shuffled scanpath strings. In short, shuffled random data and natural data were compared for between trial and within trial effects, and for five distinct levels of time window windows in both directions (for matcher and for director delays combined)

introduced to the participants' gaze data. Finally, descriptive data achieved by correcting for the role-swap condition to see the point in which the recurrence was at its highest is shown in Figure 43.

Firstly, the Mauchly's test of sphericity for the five distinct time window conditions showed that the sphericity assumption was violated for all conditions, thus the Greenhouse-Geisser corrected results are reported for all conditions. Mean values for the five distinct time window conditions are shown in Figure 42, where the natural gaze recurrence data of no time window and one, two, three, and four seconds time window conditions are presented in comparison with their respective shuffled gaze, control data.

- The results of the repeated measures ANOVA (Mauchly's test results of χ^2 (77) = 246.300, $p < .001$) revealed a significant effect of repeated trials on natural gaze data for the recurrence rates of the dyad for the no time window condition, F_{noLag} (4.261, 170.450) = 3.440, $p = .008$, partial $\eta^2 = .079$. The interaction effect for repeated trials and the data type (i.e., natural gaze and shuffled gaze control data) was also significant on the recurrence rate F (4.261, 170.450) = 2.754, $p = .027$, $\eta^2 = .064$.
- For the two second time window condition, the results of the repeated measures ANOVA (Mauchly's test results of χ^2 (77) = 263.046, $p < .001$) again revealed a significant effect of repeated trials on the gaze recurrence rates of the dyad, F_{lag1} (4.186, 167.424) = 3.583, $p = .008$, partial $\eta^2 = .082$. The interaction effect between repeated trial condition and shuffled gaze control condition was also significant on the recurrence rate F (4.186, 167.424) = 2.762, $p = .027$, $\eta^2 = .065$.
- The results of the repeated measures ANOVA (Mauchly's test results of χ^2 (77) = 315.924, $p < .001$) revealed a significant effect of repeated trials on the recurrence rates of the dyad for the four seconds time window condition, F_{lag2} (3.954, 158.144) = 4.621, $p = .002$, partial $\eta^2 = .104$. The interaction effect for trial condition and shuffled gaze was also significant on the recurrence rate F (3.954, 158.144) = 3.816, $p = .006$, $\eta^2 = .087$.
- The results of the repeated measures ANOVA (Mauchly's test results of χ^2 (77) = 353.129, $p < .001$) showed a significant effect of repeated trials on the recurrence rates of the dyad for the six seconds time window condition, F_{lag3} (3.663, 145.357) = 5.624, $p < .001$, partial $\eta^2 = .123$. The interaction effect between repeated trials and control condition was also significant on the recurrence rate F (3.663, 145.357) = 4.981, $p = .001$, $\eta^2 = .111$.
- The results of the repeated measures ANOVA (Mauchly's test results of χ^2 (77) = 388.010, $p < .001$) revealed a significant effect of the trial condition on the recurrence rates of the dyad for the eight seconds time window condition, F_{lag4} (3.374, 134.953) = 6.419, $p < .001$, partial $\eta^2 = .138$. The interaction effect between shuffled gaze control condition and trial condition was also significant on the recurrence rate F (3.374, 134.953) = 6.279, $p < .001$, $\eta^2 = .136$.

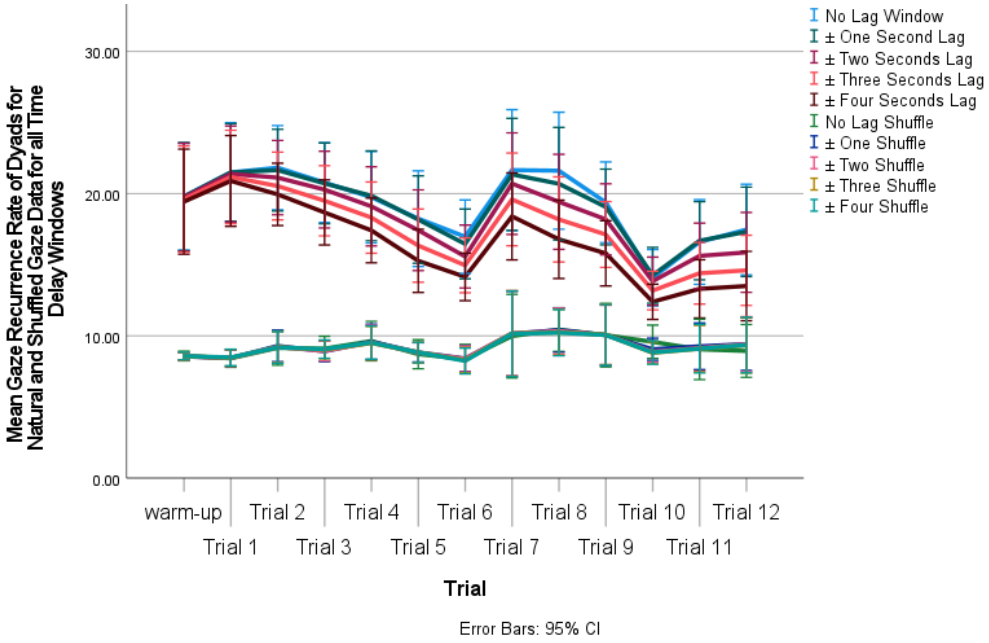

Figure 42: Mean percentage of gaze cross recurrence of 21 dyads (Natural Data) as they are presented with repeated trials of matching and sorting 12 Tangram figures. Warmup-T6 first block, T7-T12 second block after the role-swap condition. Shuffled data was constructed by randomizing the order of the client experiment side (first matcher, second director) participant as a control condition.

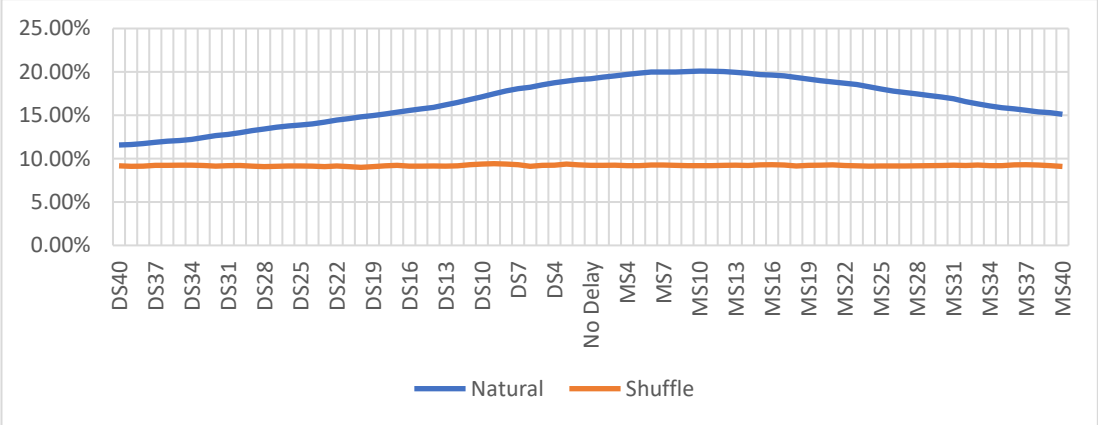

Figure 43: CRQA results for rate of recurrence at differing time lag points demonstrating that at MS10 was the highest rate of recurrence among the communicating dyad was achieved, where the matcher participants on average trailed directors' gazes by 1 second. Note that the X axis ranges from DS40: where the directors' gaze were shifted ahead 4 seconds, until MS40: where the matchers' gaze were shifted ahead 4 seconds.

4.2.2. Transition Matrix Entropy (TME) Analysis Results

The DDRoiA generated and then organized data was also analysed for transition matrix entropy, which is a measure to analyse for uncertainty in the gaze patterns of each participant. By its nature, as defined by (Holmqvist et al., 2011, p. 477; Shannon, 1948) and in contrast with CRQA, this is a measure to focus on the distribution of data, in this case – a matrix, of a single participant at a time to further evaluate the gaze pattern for each individual for randomness. We utilised this to evaluate the role condition, wherein the dyad attended to their joint task with the distinct roles of a director who had ordered Tangram figures and a matcher who had the figures shuffled into a pseudo-random order. The experiment setting that we employed, therefore, can be analysed using this measure in a 13 (number of trials) x 2 (number of roles) again by utilizing a repeated measures ANOVA analysis. Mean values for the 13x2 design are shown in Figure 44. In this graph, it is helpful to recall that the experiment role swap condition was after the seventh trial of the first block, which consisted of a warm-up trial and then six more trials.

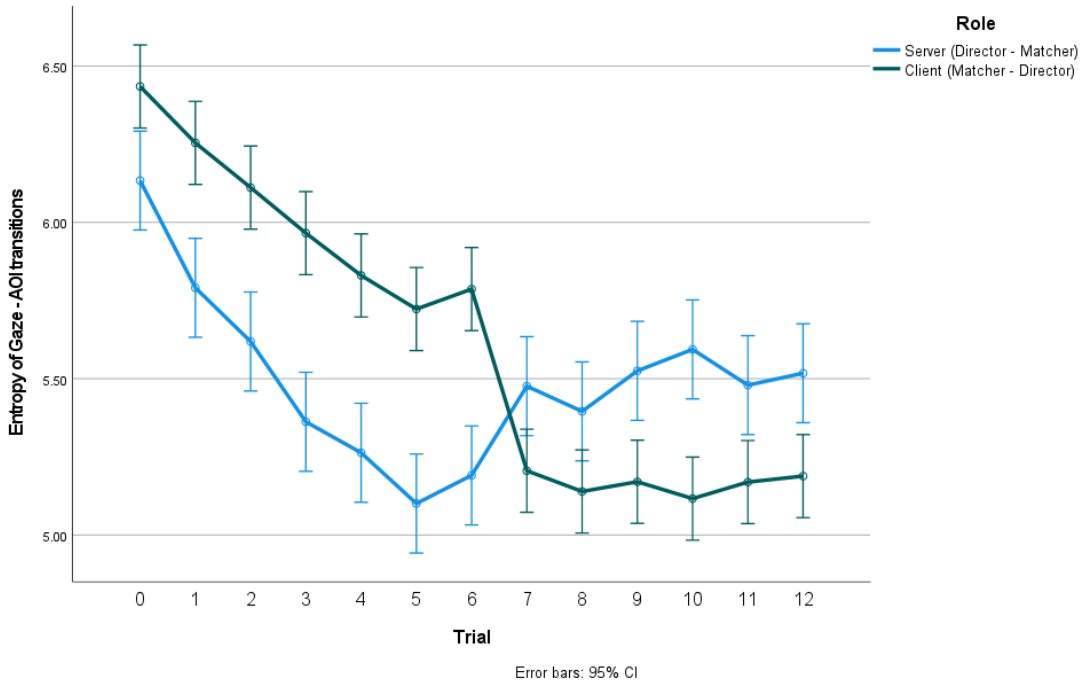

Figure 44: Mean rates of entropy for dyads in the joint matching task condition for each trial. Recall that after the first block (warm-up and 6 trials) there is a role-swap condition, hence the notation for server and client instead of director and matcher, where server denotes the participant who is the first director and the second matcher, as the client denotes the participant who is the first matcher and the second director.

The results of the repeated measures ANOVA showed that there was a significant effect of role F (1, 260) = 19.723, $p < .001$, partial $\eta^2 = .071$, as well as for the interaction effect of role and trial F (12, 260) = 19.397, $p < .001$, partial $\eta^2 = .472$.

4.3. EEG PLV Analyses Results

The phase-locking value of all dyads was analysed using a one-way ANOVA for the four experiment blocks (warm-up, first block second half, as well as second block first and second halves) across 5 band frequencies for each 16x16 electrode site combination. False Discovery Rate (FDR) correction was employed to control for Type 1 error inflation due to multiple comparisons (Benjamini & Hochberg, 1995).

The results of the analysis showed that the effect of establishing a common ground during the joint task of matching and sorting the Tangram figures resulted in significantly different phase-locking value changes in the Delta (0.5 to 3Hz), Theta (4 – 7Hz), and lower Gamma (31 – 48Hz) band frequencies, as shown in the phase connectivity of differing electrode sites for each of these band d frequencies.

4.3.1. Significant Electrode Site Connectivity in the Delta Frequency Band

In the Delta frequency band electrode site connectivity between electrode pairs C3-FZ, C4-P4, F3-F4, F4-P4, O2-FZ, OZ-P4, P7-FZ, P7-P3, PZ-F4, PZ-FZ, T8-FZ, C3-P3, CZ-F4, P7-F4, T7-F3, F3-FZ, F4-P3, and F4-T8 remained significant after the FDR correction. Partial η^2 effect sizes for all electrode pairs are presented in Table 10. Example mean PLV results are shown for the electrode site pairs of the hyperscanning PLV for the five highest effect sizes in Figures 45-49. The increasing PLV trend visible in these examples were present in all pairs. Overall, the results showed that there is a significant increasing trend in PLV connectivity strength in the delta band between the server (presented as rows in PLV, recall that server was the director before the role swap) and the client (presented as columns in PLV analysis, the matcher in the first block before the role swap) as the experiment progressed.

Table 10: Delta PLV connectivity matrix, shown as RM ANOVA effect sizes for the experiment block effect on PLV across all dyads. Warmer (yellow) colors mean higher effect sizes.

	FZ	CZ	PZ	OZ	F3	F4	C3	C4	P3	P4	P7	P8	O1	O2	T7	T8
FZ	0.162	0.051	0.063	0.024	0.075	0.206	0.120	0.127	0.105	0.158	0.145	0.120	0.123	0.091	0.092	0.113
CZ	0.151	0.028	0.053	0.005	0.022	0.259	0.035	0.159	0.076	0.072	0.065	0.063	0.016	0.076	0.038	0.100
PZ	0.323	0.025	0.033	0.124	0.069	0.261	0.050	0.163	0.048	0.066	0.123	0.111	0.126	0.112	0.085	0.129
OZ	0.177	0.044	0.098	0.018	0.146	0.264	0.011	0.002	0.148	0.438	0.225	0.090	0.212	0.140	0.140	0.096
F3	0.245	0.095	0.033	0.030	0.177	0.296	0.036	0.017	0.098	0.038	0.066	0.011	0.020	0.028	0.070	0.068
F4	0.190	0.092	0.152	0.039	0.116	0.201	0.162	0.215	0.217	0.307	0.075	0.139	0.156	0.111	0.057	0.226
C3	0.314	0.039	0.045	0.023	0.004	0.219	0.035	0.196	0.239	0.042	0.109	0.037	0.024	0.022	0.030	0.044
C4	0.190	0.147	0.087	0.040	0.081	0.198	0.100	0.070	0.093	0.289	0.129	0.152	0.135	0.177	0.018	0.082
P3	0.207	0.043	0.032	0.107	0.018	0.174	0.077	0.074	0.105	0.118	0.174	0.119	0.149	0.114	0.079	0.041
P4	0.078	0.080	0.064	0.075	0.064	0.086	0.153	0.116	0.066	0.069	0.179	0.074	0.172	0.134	0.075	0.096
P7	0.288	0.143	0.181	0.041	0.104	0.232	0.118	0.004	0.291	0.141	0.194	0.134	0.176	0.195	0.154	0.062
P8	0.153	0.125	0.040	0.036	0.108	0.200	0.045	0.139	0.175	0.091	0.206	0.027	0.090	0.063	0.006	0.119
T7	0.185	0.212	0.149	0.085	0.260	0.129	0.150	0.131	0.162	0.135	0.189	0.120	0.136	0.131	0.160	0.186
T8	0.239	0.027	0.053	0.006	0.034	0.189	0.063	0.167	0.110	0.096	0.224	0.026	0.123	0.037	0.081	0.081
O1	0.175	0.040	0.051	0.108	0.087	0.155	0.007	0.197	0.176	0.044	0.250	0.071	0.072	0.064	0.155	0.102
O2	0.281	0.073	0.166	0.061	0.021	0.162	0.011	0.013	0.126	0.170	0.142	0.139	0.167	0.136	0.070	0.157

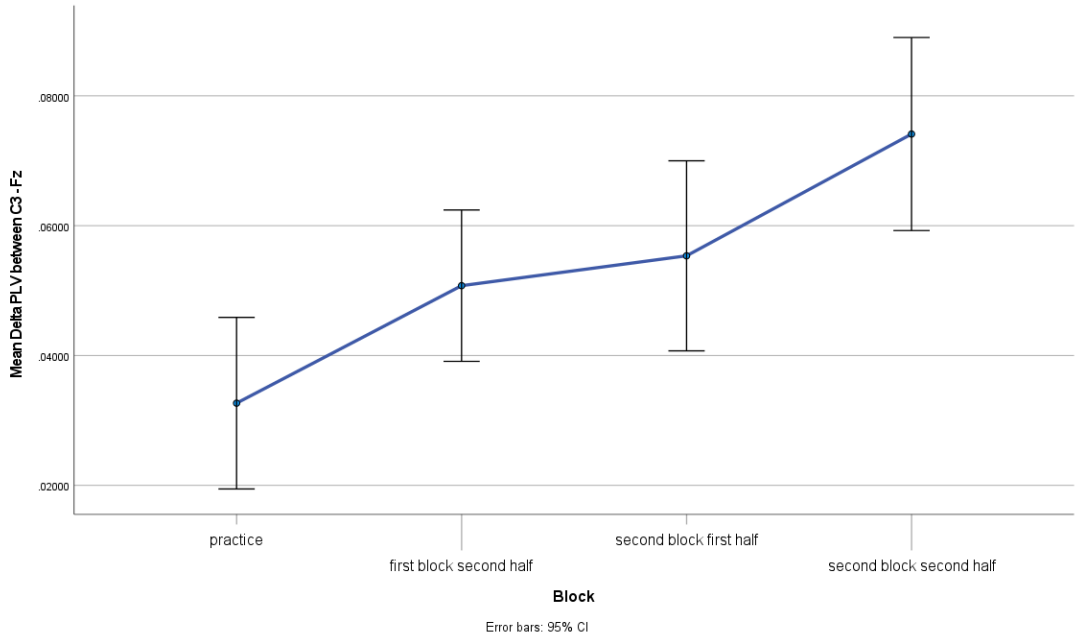

Figure 45: Mean Delta PLV between CZ and Fz electrode sites for the interacting dyad for the four experiment blocks.

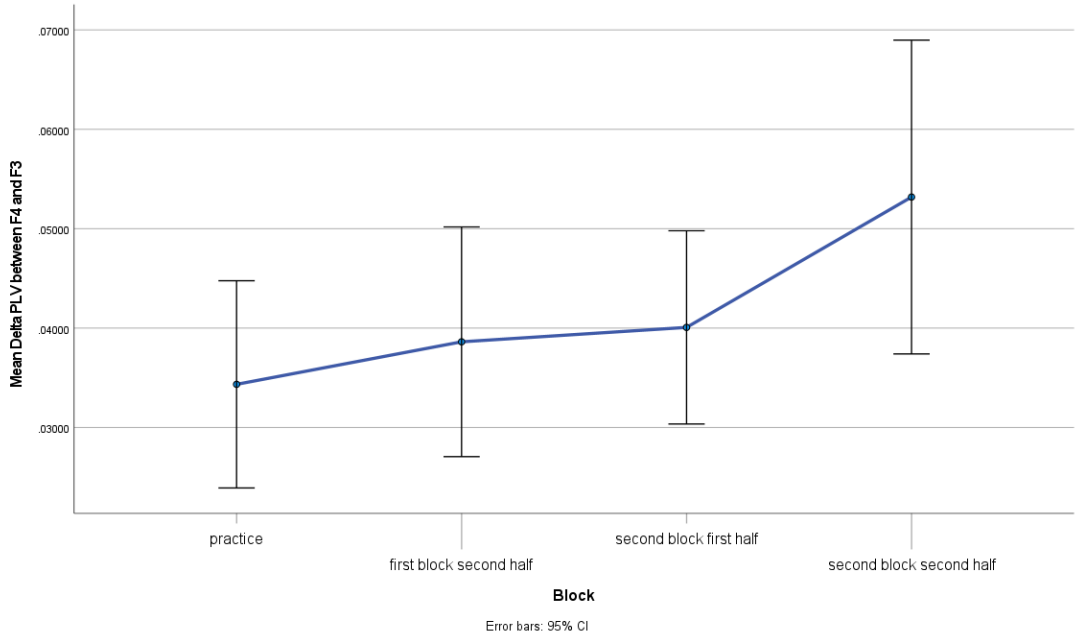

Figure 46: Mean Delta PLV between F4 and F3 electrode sites for the interacting dyad for the four experiment blocks.

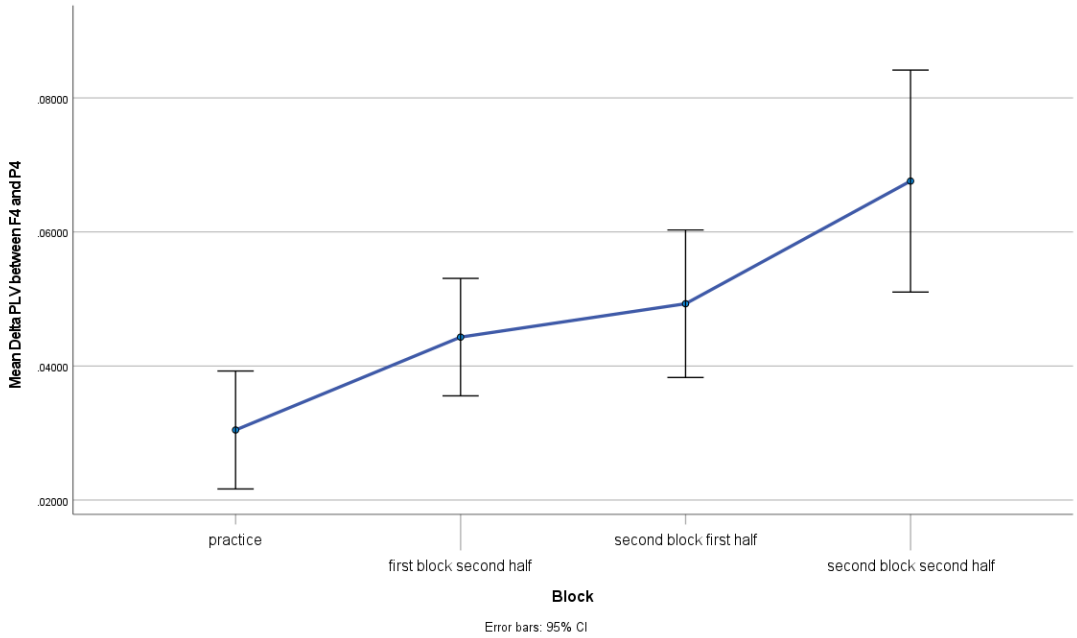

Figure 47: Mean Delta PLV between F4 and P4 electrode sites for the interacting dyad for the four experiment blocks.

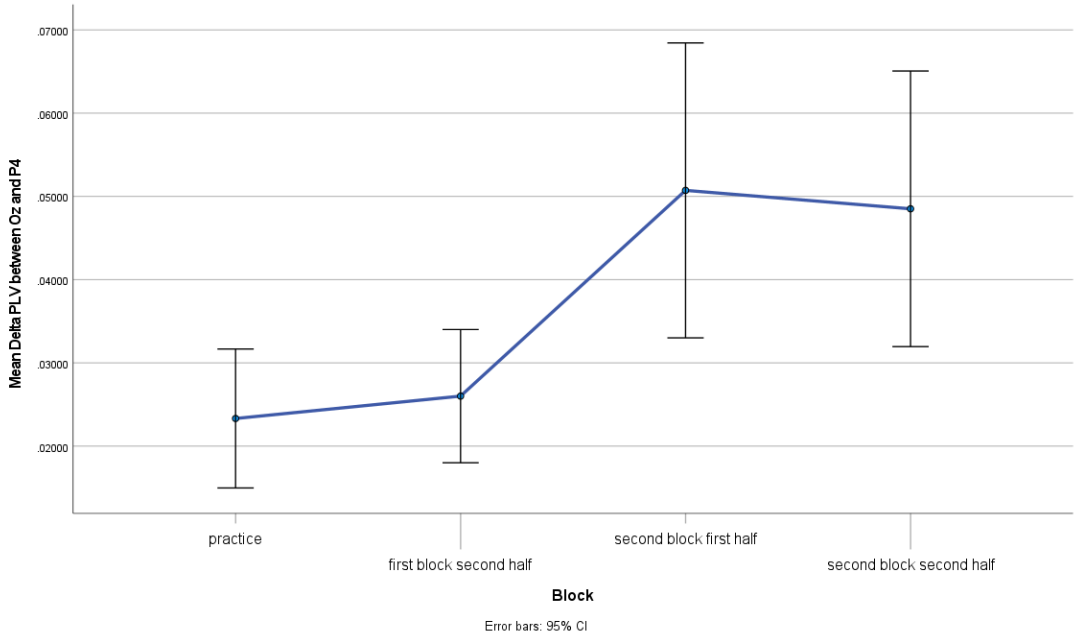

Figure 48: Mean Delta PLV between Oz and P4 electrode sites for the interacting dyad for the four experiment blocks.

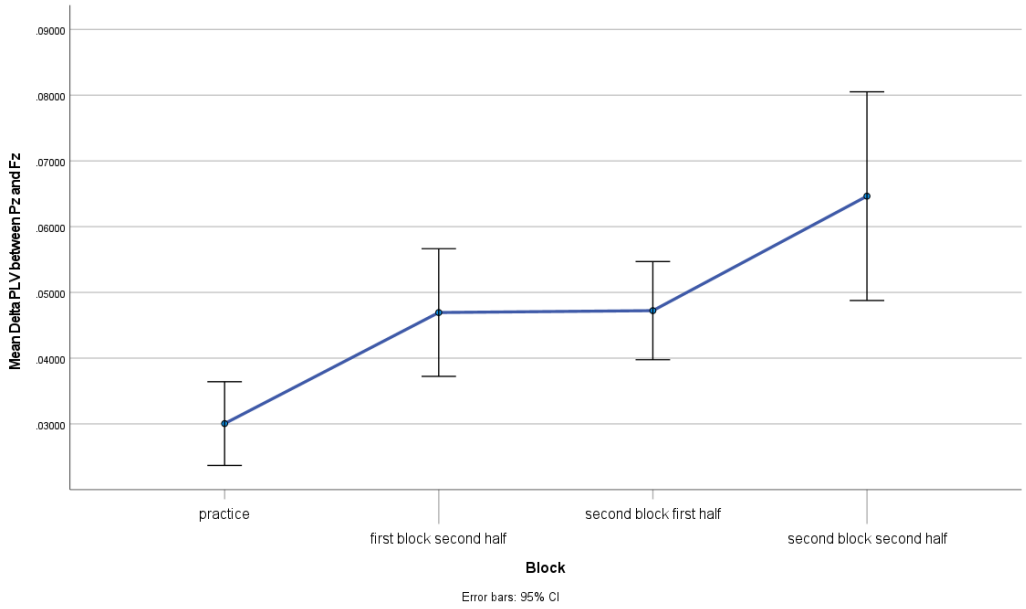

Figure 49: Mean Delta PLV between Pz and Fz electrode sites for the interacting dyad for the four experiment blocks.

4.3.2. Significant Electrode Site Connectivity in the Theta Frequency Band

The PLV connectivity difference in electrode pairs C3-P3, CZ-P7, O2-C3, O2-C4, O2-FZ, O2-O1, O2-P7, O2-T7, OZ-T7, P4-C3, P4-T8, P8-C3, P8-C4, P8-F4, P8-PZ, CZ-C4, O2-O2, O2-P3, O2-P4, P4-F4, P7-P8, OZ-F4, P4-F3, P7-O2, P7-P7, F4-P8, P8-CZ, PZ-P7, T8-C4, C3-P7, C4-P7, P4-C4, P7-P4, P8-P7, CZ-PZ, T8-P3, C4-C3, C4-P4, C4-PZ, CZ-F3, CZ-F4, CZ-P3, CZ-T7, O2-CZ, P4-CZ, P7-C3, and P7-F4 remained significant after FDR correction.

Table 11: Theta PLV connectivity matrix, shown as RM ANOVA effect sizes for experiment block effect on PLV across all dyads. Warmer (yellow) colors mean higher effect sizes.

	FZ	CZ	PZ	OZ	F3	F4	C3	C4	P3	P4	P7	P8	O1	O2	T7	T8
FZ	0.048	0.147	0.084	0.069	0.077	0.123	0.126	0.167	0.063	0.110	0.157	0.180	0.102	0.116	0.078	0.144
CZ	0.089	0.184	0.210	0.060	0.220	0.200	0.176	0.256	0.209	0.113	0.322	0.171	0.209	0.167	0.200	0.116
PZ	0.016	0.035	0.021	0.021	0.044	0.044	0.084	0.192	0.021	0.006	0.252	0.093	0.017	0.025	0.036	0.017
OZ	0.203	0.205	0.116	0.163	0.192	0.291	0.206	0.257	0.183	0.090	0.109	0.146	0.180	0.141	0.366	0.170
F3	0.032	0.032	0.056	0.005	0.039	0.055	0.027	0.049	0.051	0.050	0.015	0.052	0.016	0.009	0.004	0.191
F4	0.017	0.119	0.144	0.069	0.071	0.059	0.069	0.131	0.053	0.113	0.087	0.213	0.086	0.124	0.032	0.139
C3	0.062	0.093	0.139	0.006	0.186	0.173	0.142	0.191	0.257	0.158	0.250	0.144	0.076	0.061	0.008	0.206
C4	0.036	0.160	0.200	0.076	0.110	0.186	0.199	0.144	0.118	0.197	0.233	0.169	0.157	0.156	0.079	0.080
P3	0.064	0.136	0.110	0.008	0.099	0.094	0.124	0.198	0.152	0.074	0.140	0.093	0.108	0.092	0.108	0.028
P4	0.145	0.188	0.169	0.112	0.240	0.249	0.308	0.228	0.141	0.114	0.190	0.130	0.185	0.136	0.169	0.248
P7	0.124	0.194	0.141	0.072	0.133	0.192	0.218	0.185	0.103	0.207	0.264	0.252	0.191	0.225	0.154	0.148
P8	0.185	0.209	0.290	0.030	0.116	0.327	0.253	0.255	0.111	0.140	0.239	0.145	0.138	0.094	0.106	0.126
T7	0.027	0.058	0.080	0.023	0.022	0.015	0.052	0.091	0.121	0.089	0.171	0.140	0.097	0.063	0.127	0.083
T8	0.004	0.047	0.061	0.081	0.004	0.026	0.072	0.222	0.189	0.067	0.172	0.051	0.088	0.049	0.024	0.034
O1	0.042	0.051	0.051	0.017	0.057	0.086	0.028	0.173	0.073	0.035	0.146	0.075	0.037	0.068	0.045	0.038
O2	0.250	0.221	0.166	0.091	0.170	0.167	0.274	0.323	0.240	0.243	0.281	0.160	0.293	0.224	0.313	0.155

The effect sizes of the RM ANOVA results are shown in Table 11, with the mean Theta PLV for the five electrode site pairs with the highest effect sizes presented in Figures 50-54.

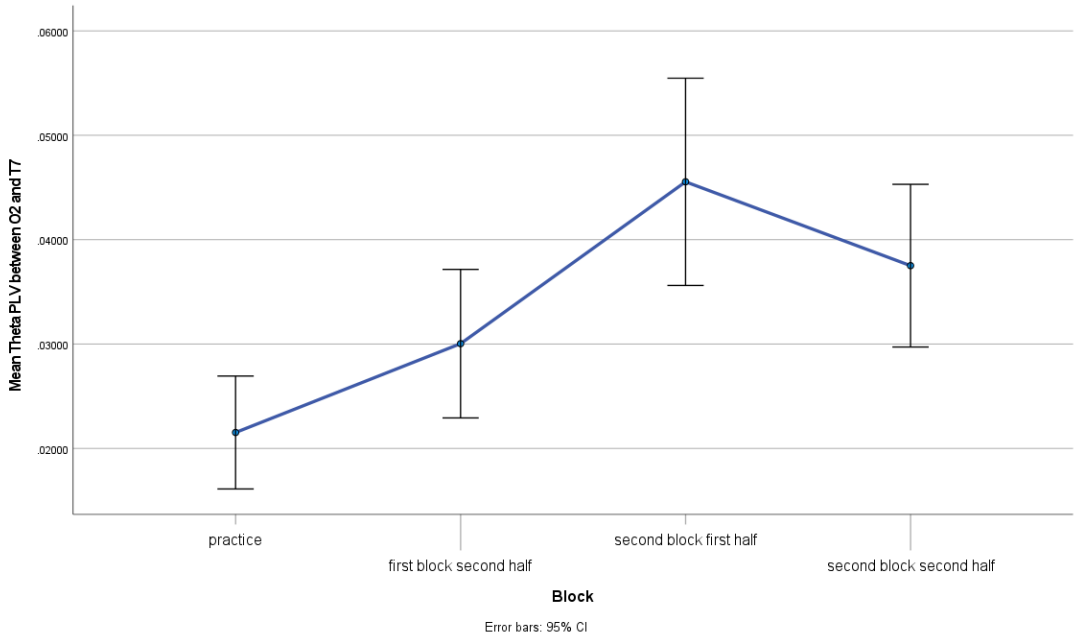

Figure 50: Mean Theta PLV between O2 and T7 electrode sites for the interacting dyad for the four experiment blocks.

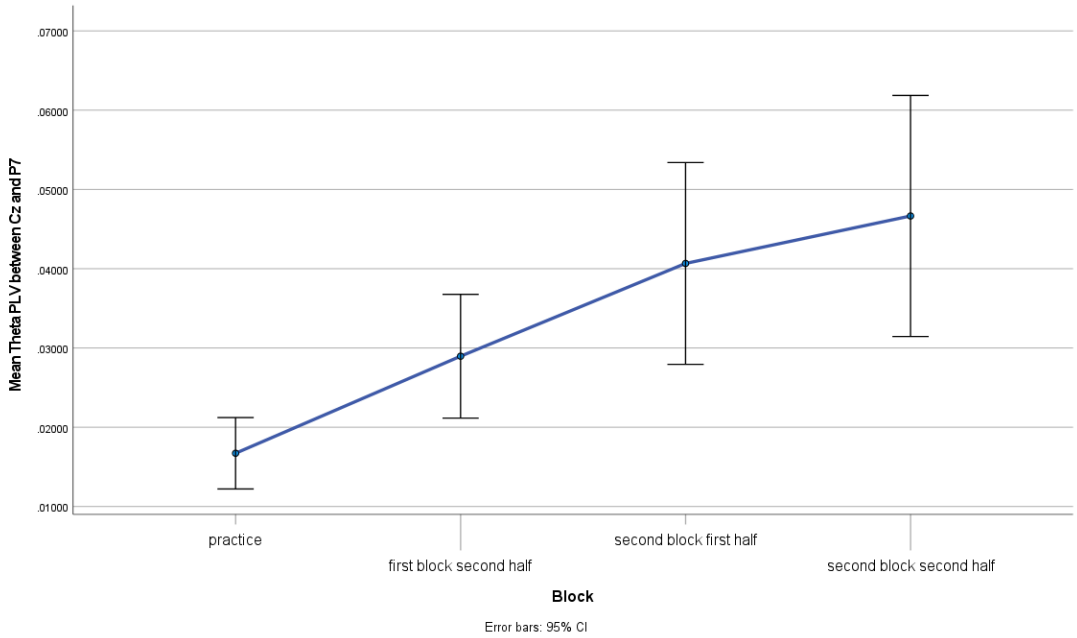

Figure 51: Mean Theta PLV between Cz and P7 electrode sites for the interacting dyad for the four experiment blocks.

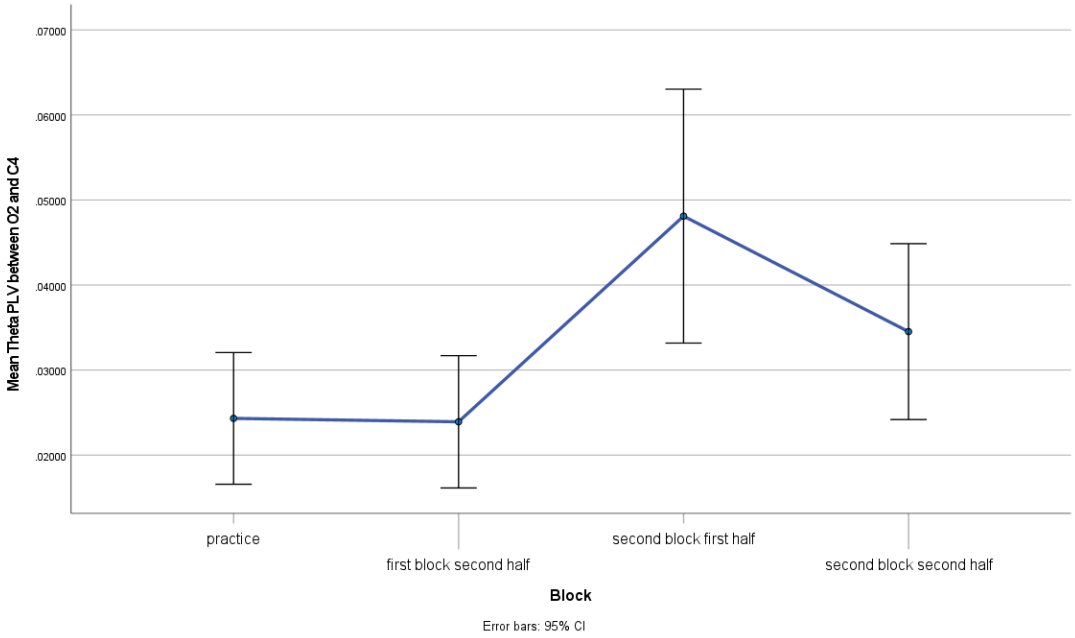

Figure 52: Mean Theta PLV between O2 and C4 electrode sites for the interacting dyad for the four experiment blocks.

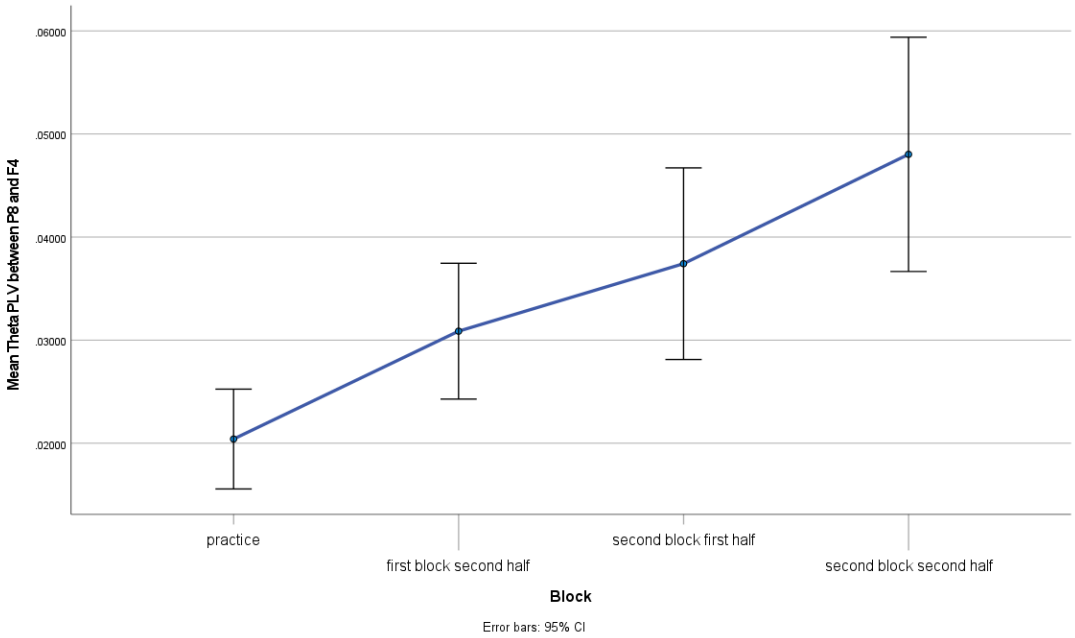

Figure 53: Mean Theta PLV between P8 and F4 electrode sites for the interacting dyad for the four experiment blocks.

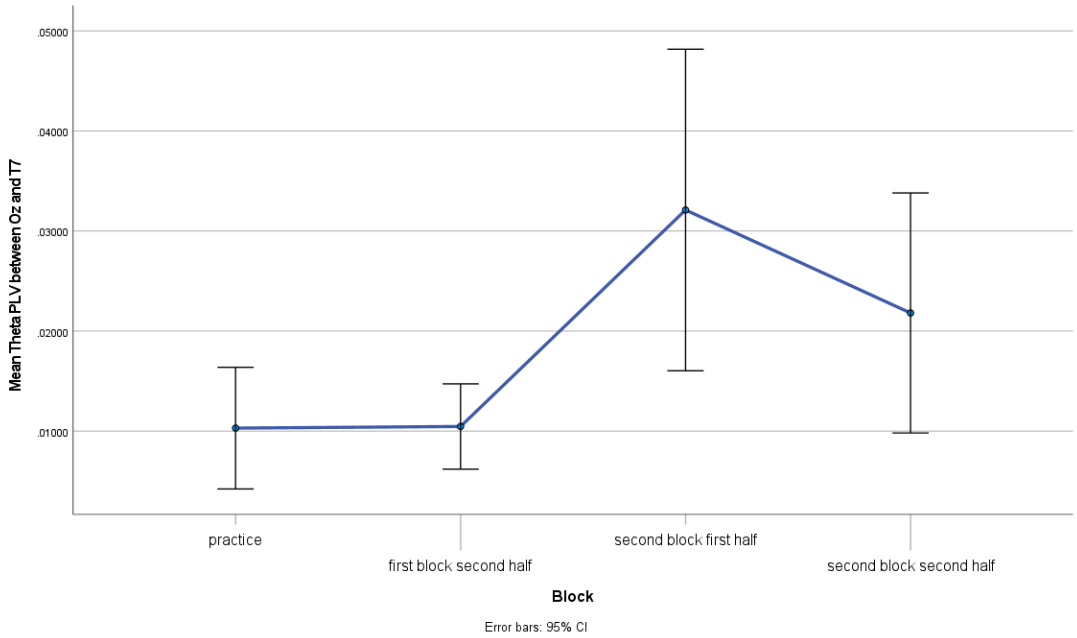

Figure 54: Mean Theta PLV between Oz and T7 electrode sites for the interacting dyad for the four experiment blocks.

4.3.3. Significant Electrode Site Connectivity in the Gamma Frequency Band

The Gamma frequency band was the frequency range in which the highest inter-brain synchrony was observed. With the 16x16 matrix, we calculated a total of 256 PLV values. After FDR correction, 250 out of the 256 values remained significant. In Table 12, the effect sizes of RM ANOVA results are presented, with the five highest values shown in Figures 55-59.

Table 12: Gamma PLV connectivity matrix, shown as RM ANOVA effect sizes for experiment block effect on PLV across all dyads. Warmer (yellow) colors mean higher effect sizes.

	FZ	CZ	PZ	OZ	F3	F4	C3	C4	P3	P4	P7	P8	T7	T8	O1	O2
FZ	0.514	0.480	0.216	0.294	0.439	0.455	0.506	0.449	0.554	0.468	0.573	0.395	0.400	0.428	0.447	0.486
CZ	0.513	0.485	0.231	0.287	0.351	0.433	0.503	0.460	0.560	0.470	0.547	0.366	0.381	0.440	0.450	0.494
PZ	0.250	0.202	0.200	0.127	0.292	0.263	0.242	0.212	0.328	0.252	0.348	0.254	0.177	0.226	0.359	0.360
OZ	0.393	0.351	0.214	0.040	0.243	0.306	0.376	0.255	0.384	0.360	0.427	0.287	0.321	0.406	0.141	0.238
F3	0.476	0.369	0.145	0.160	0.412	0.457	0.384	0.332	0.410	0.369	0.377	0.335	0.321	0.450	0.312	0.333
F4	0.415	0.385	0.180	0.225	0.341	0.407	0.401	0.351	0.439	0.398	0.443	0.352	0.376	0.404	0.383	0.442
C3	0.474	0.420	0.252	0.292	0.298	0.416	0.457	0.414	0.537	0.444	0.520	0.337	0.397	0.420	0.433	0.501
C4	0.530	0.488	0.403	0.301	0.387	0.485	0.513	0.428	0.547	0.561	0.562	0.417	0.440	0.503	0.477	0.530
P3	0.522	0.493	0.410	0.323	0.435	0.469	0.533	0.463	0.588	0.521	0.631	0.418	0.533	0.513	0.452	0.509
P4	0.527	0.487	0.290	0.295	0.431	0.487	0.507	0.447	0.573	0.501	0.546	0.420	0.486	0.513	0.451	0.478
P7	0.546	0.514	0.387	0.282	0.499	0.516	0.529	0.460	0.568	0.516	0.614	0.455	0.535	0.537	0.450	0.507
P8	0.390	0.369	0.183	0.272	0.420	0.467	0.405	0.451	0.508	0.414	0.465	0.384	0.352	0.338	0.435	0.446
T7	0.490	0.468	0.309	0.298	0.434	0.484	0.488	0.457	0.532	0.527	0.576	0.442	0.456	0.489	0.458	0.532
T8	0.582	0.565	0.262	0.355	0.532	0.586	0.575	0.553	0.594	0.524	0.556	0.484	0.530	0.508	0.503	0.514
O1	0.513	0.464	0.489	0.172	0.386	0.451	0.482	0.381	0.513	0.415	0.466	0.378	0.433	0.483	0.359	0.349
O2	0.657	0.630	0.491	0.344	0.553	0.614	0.628	0.529	0.609	0.610	0.566	0.610	0.598	0.711	0.498	0.483

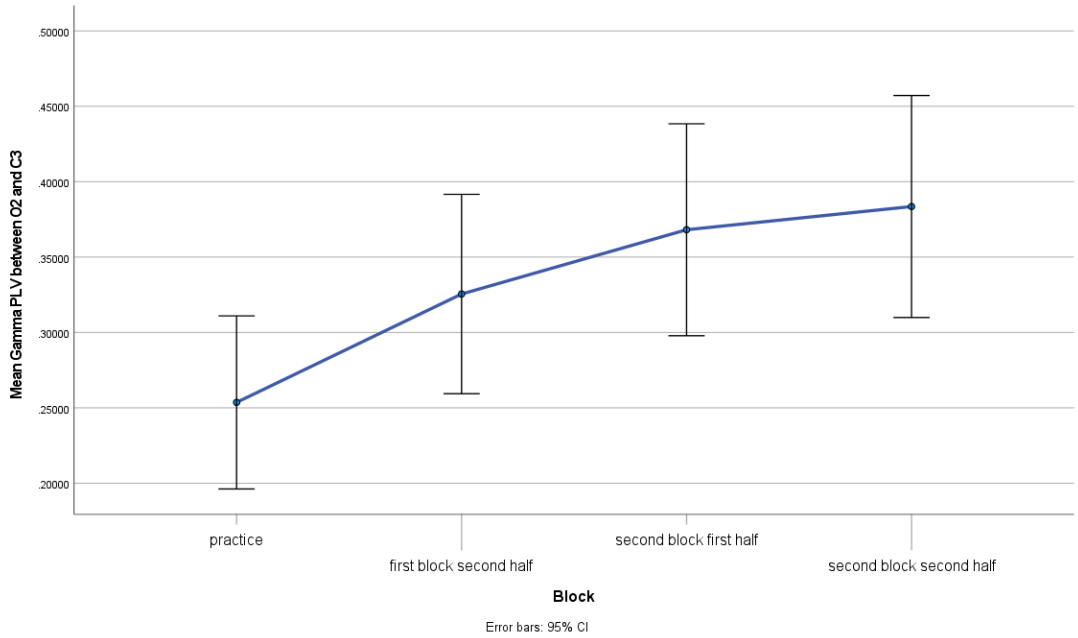

Figure 55: Mean Gamma PLV between O2 and C3 electrode sites for the interacting dyad for the four experiment blocks.

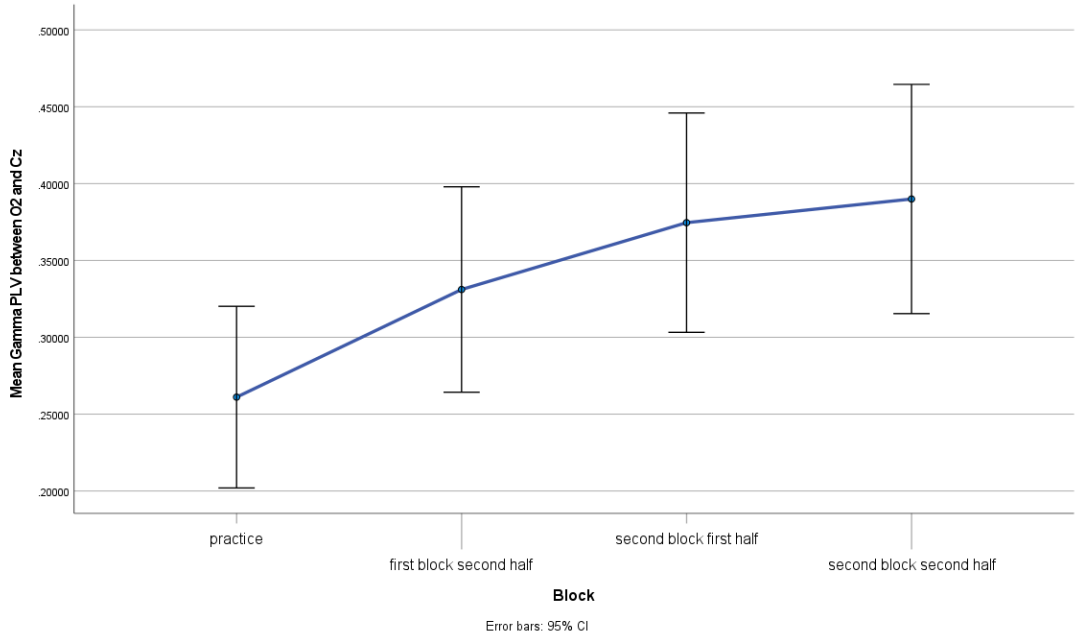

Figure 56: Mean Gamma PLV between O2 and Cz electrode sites for the interacting dyad for the four experiment blocks.

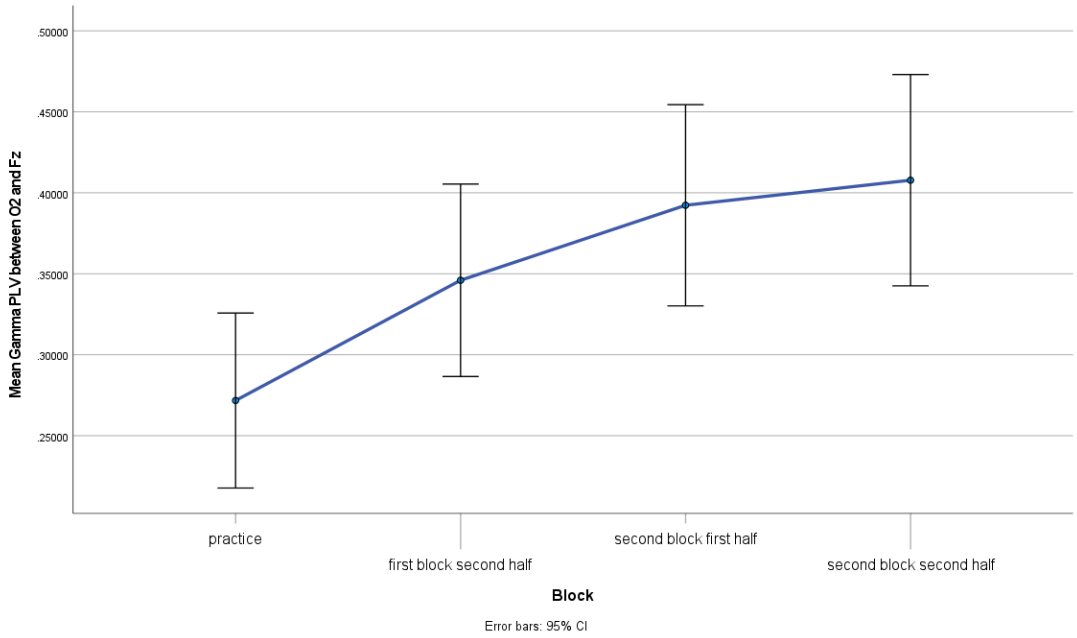

Figure 57: Mean Gamma PLV between O2 and Fz electrode sites for the interacting dyad for the four experiment blocks.

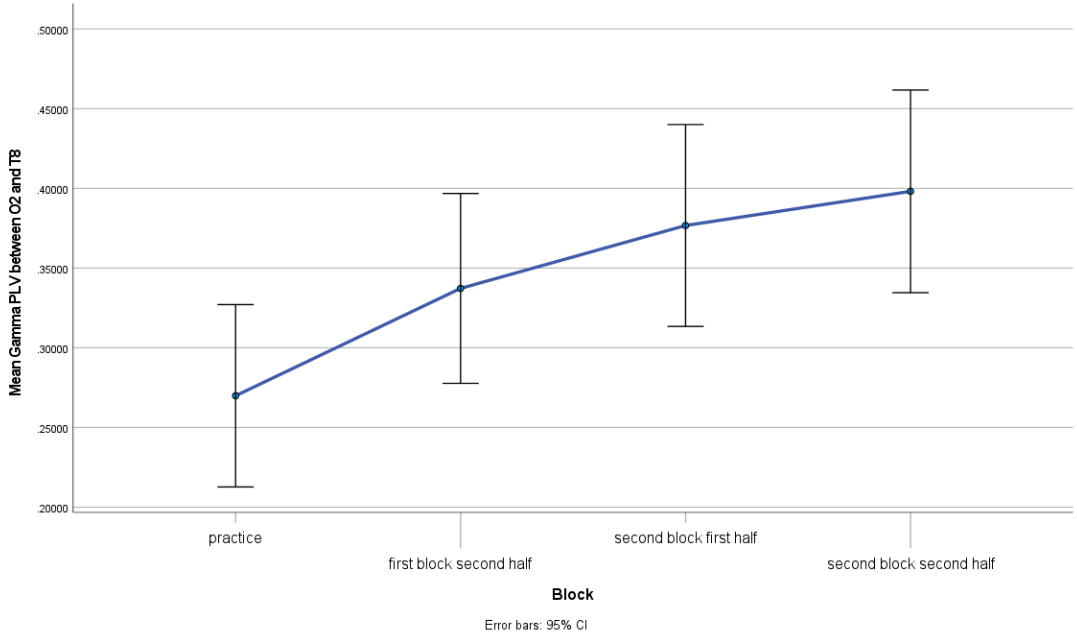

Figure 58: Mean Gamma PLV between O2 and T8 electrode sites for the interacting dyad for the four experiment blocks.

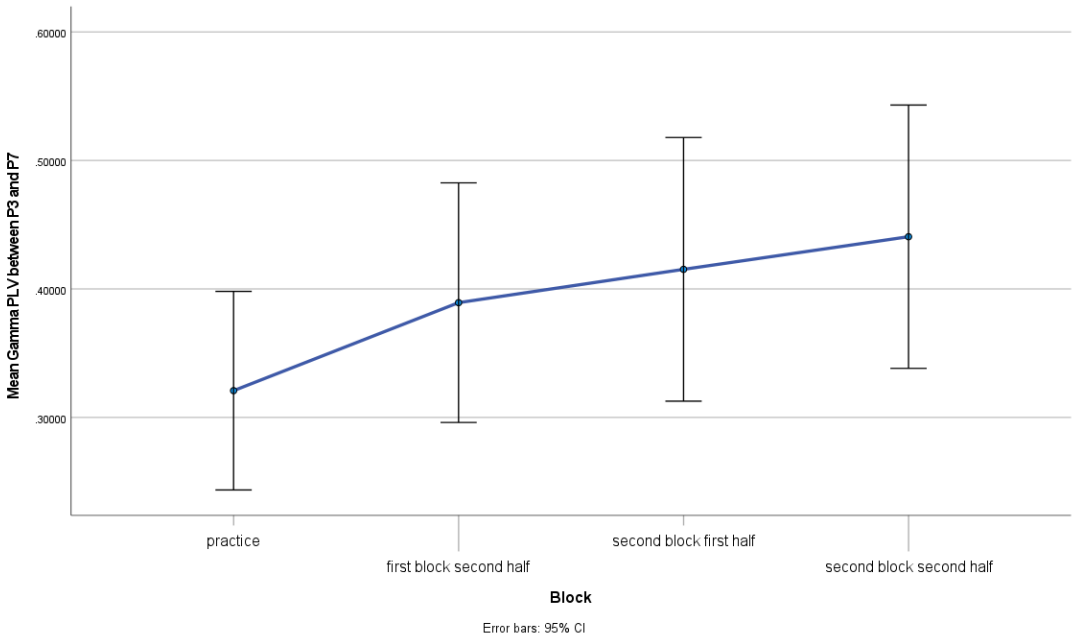

Figure 59: Mean Gamma PLV between P3 and P7 electrode sites for the interacting dyad for the four experiment blocks.

4.4. fNIRS Analyses Results

4.4.1. WTC Analysis Results

The post-processed oxygenation data was investigated for the effects of the dyadic interaction setting entailed with the process of linguistic entrainment by means of Wavelet Transform Coherence (WTC) analysis (Cui et al., 2012; Grinsted et al., 2004). Exemplified WTC increase results in Figure 60 for each block (bottom right-hand side is coherence during the experiment trials, a delta value relative to the trial before) for given pairs of participants and their raw oxygenation signals (at top right-hand side) are given below. In the WTC figures, the left-hand side column provides the resulting plot of a WTC analysis result, in which the 32 to 128 sampling period were focused for the results of coherence among our 2Hz signals, which means a 16 to 64 second time window was the average coherence results extraction window from the matrix. In that, if there is an increase in coherence during a period of samples (x-axis in Figure 59), and at 128 period, then this is a coherence increase within a 64 second time window. Before the WTC results are presented, however; during the preliminary WTC analyses, findings of inconsistent noise profiles were detected, which interfered with biological data particularly for the newer, 5Hz fNIR 2000m device recordings. The device manufacturer later communicated to us that this noise profile was due to their

device's electronical design, wherein the early production sample available to the experiment setting in the laboratory was not shielded against radiofrequency interference between the data sources (NIR receiver sides of each optode) and the amplifier circuitry on the fNIR headband. This resulted in an interference pattern that was not distinguishable from the biological data due to raw data corruption at source. The older fNIR 1200 device recordings were not affected by this due to its electronics design. This limited the present investigation to only 6 pairs of participants, where the interference was observably at its lowest that were subjectively selected for best evaluation at the present state of investigation by an expert analyst for exploratory purposes only.

A baseline WTC correction may be applied to the WTC analysis in certain experimental cases in order to better visualize and/or output the results for evaluation or further analyses. During our investigation of WTC results from fNIRS oxygenation data, both baseline corrected and raw forms of the WTC increase per block metric to best explore the novel data at hand were explored. The baseline correction was applied by taking the first 20 samples of data (10 seconds at 2Hz sampling rate) and subtracting it from the other, trial block-wise WTC averages. Both

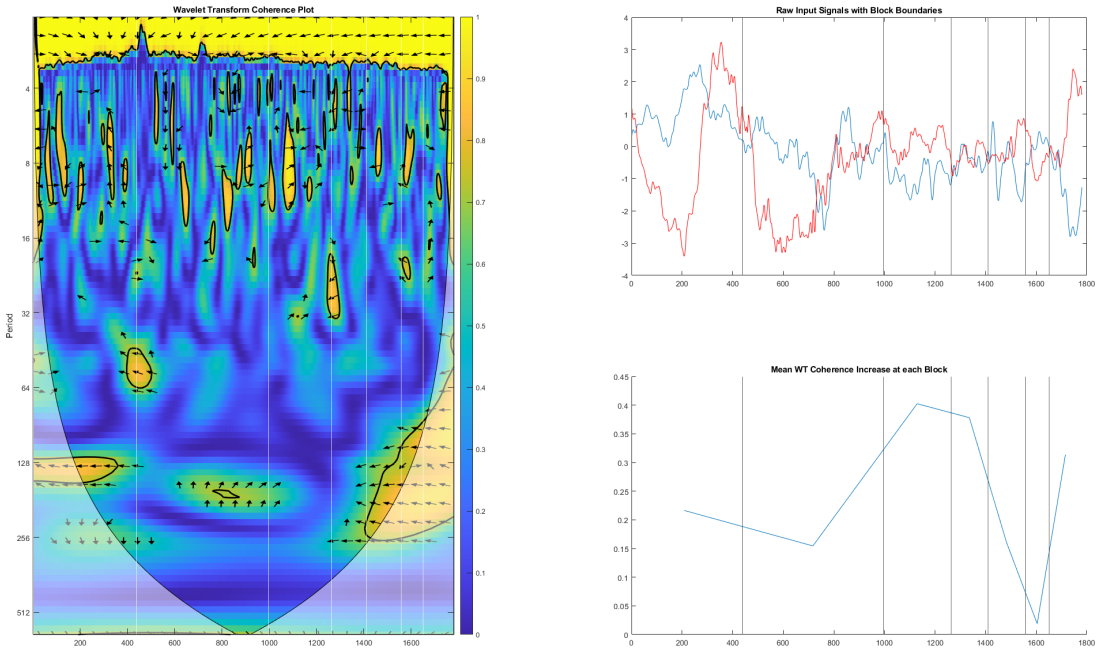

Figure 60: Dyad 5, optode 14, block 2 (after role-swap), natural data.

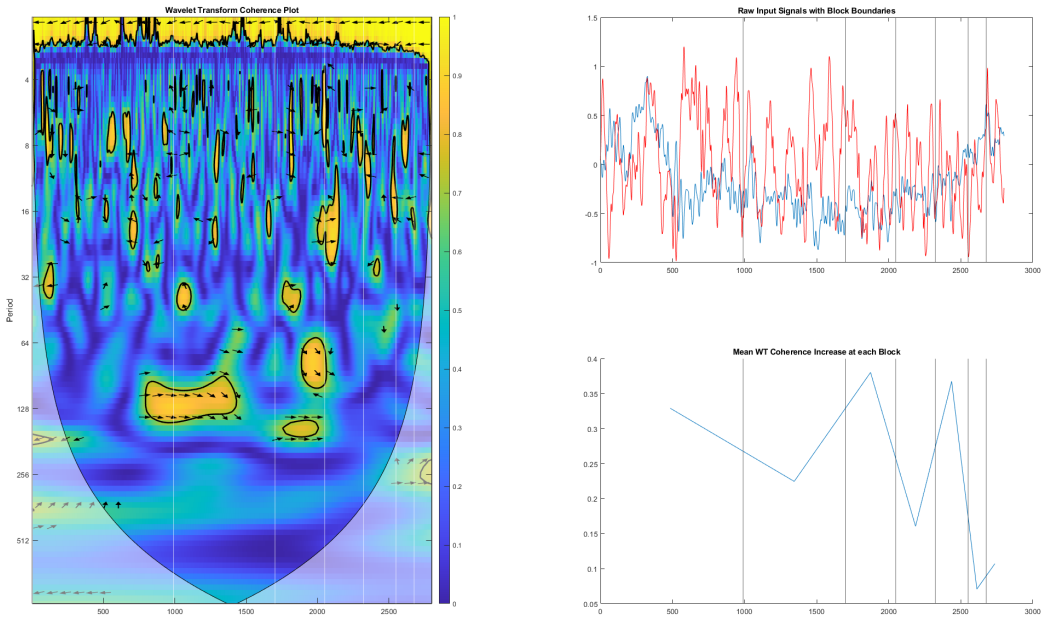

Figure 61: Dyad 9, right dorsomedial prefrontal cortex region, block 2 (after role-swap) with baseline shift correction.

the baseline corrected and the raw data were analysed; however, only the raw, non-baseline-corrected results are reported here as the ecologically valid, free-form linguistic interaction experiment setting did not control for baseline coherence specifically, for instance by means of using a secondary experiment task, and as a result the baseline correction results were not conclusive; thus, rendering the baseline shift a solely visual effect. The differences the baseline correction application caused on the WTC increase per block metric were visualized in Figures 61 and 62.

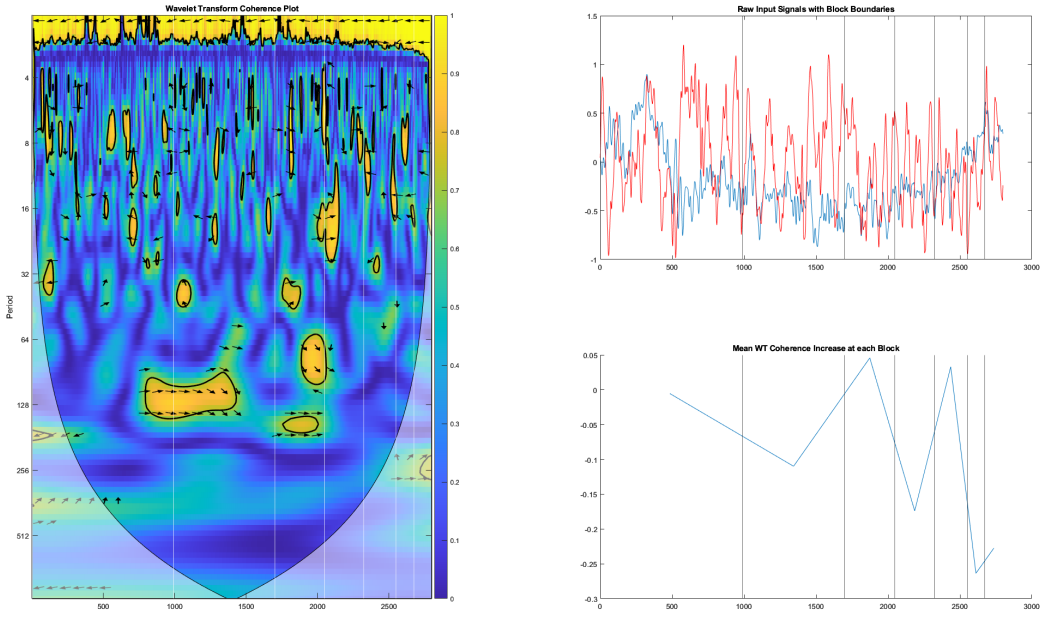

Figure 62: Dyad 9, right dorsomedial prefrontal cortex region, block 2 (after role-swap) natural data without baseline shift correction.

WTC increase based on oxygenation data for locally available optodes as well as the four focused analysis regions were analysed in a repeated measures ANOVA analysis, which focused the effect of experiment trials with relation to the experiment block (before and after the role-swap condition) rendering this a 6x2 analysis.

The results of the repeated measures ANOVA analysis for relative WTC coherence increase in none of the optodes or regions showed significant effects, with the mean levels of coherence shown in Figures 63 – 66 for the four frontal cortical regions.

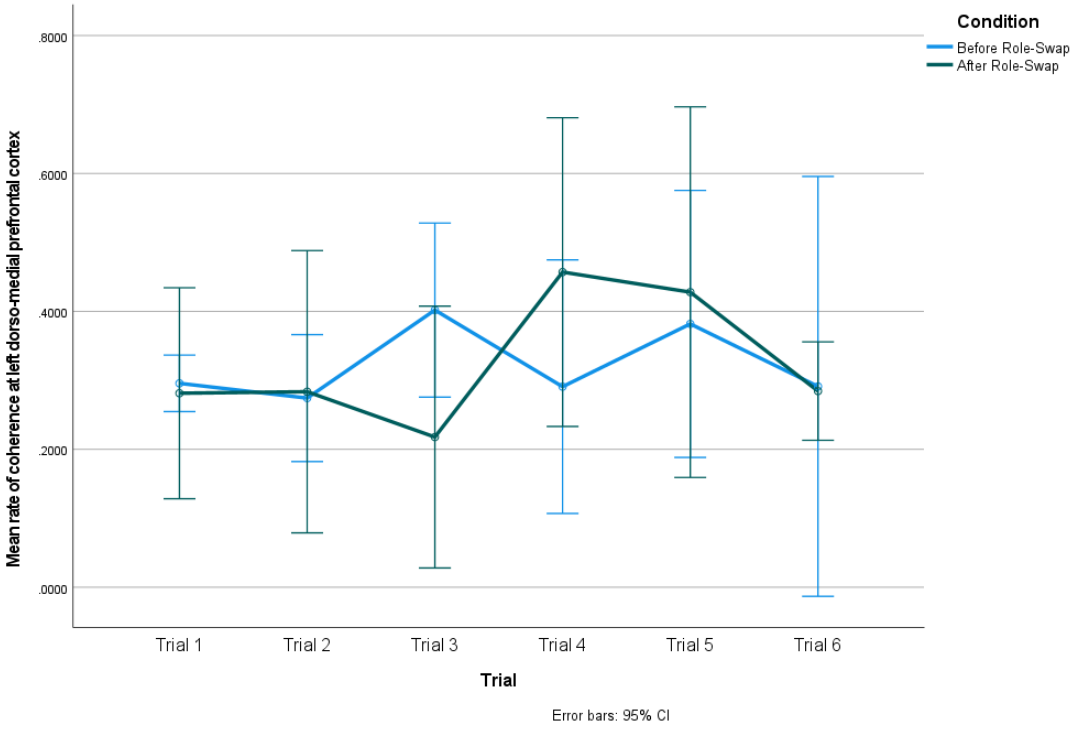

Figure 63: Mean rates of coherence based on the left dorso-medial prefrontal cortex hemodynamic response.

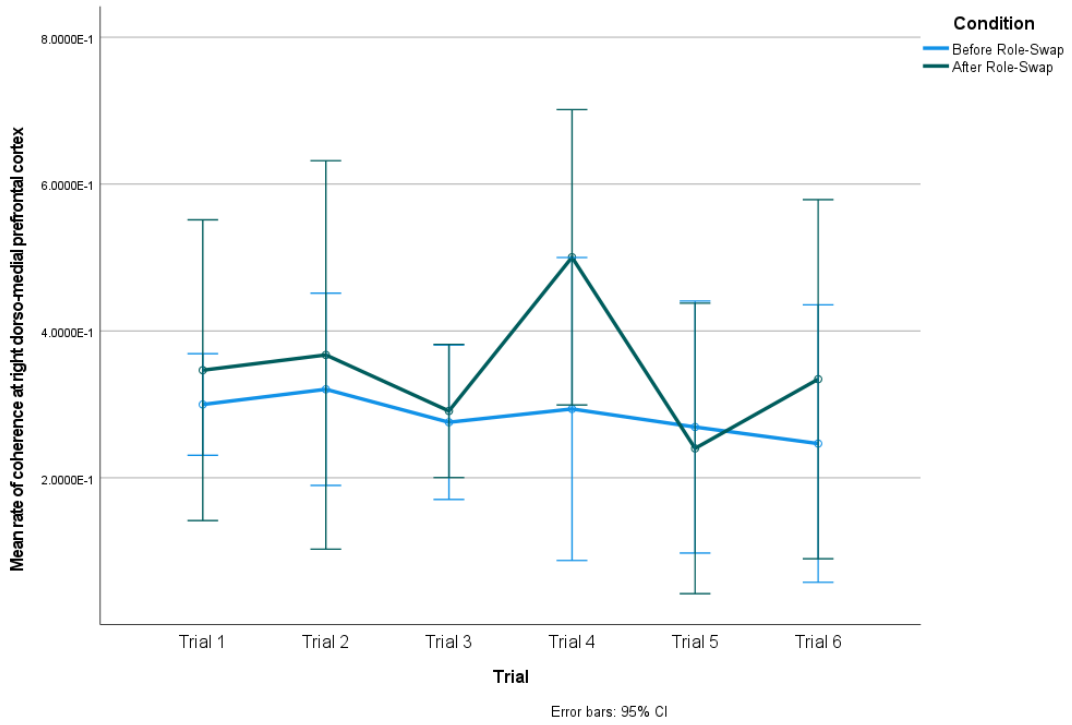

Figure 64: Mean rates of coherence based on the right dorso-medial prefrontal cortex hemodynamic response.

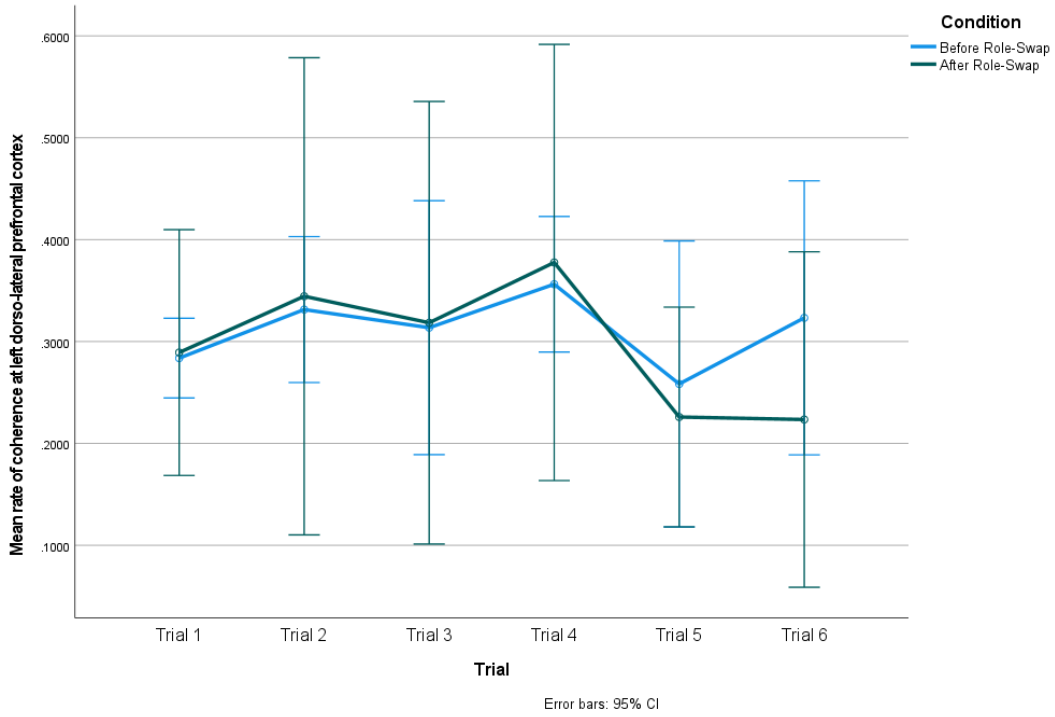

Figure 65: Mean rates of coherence based on the left dorso-lateral prefrontal cortex hemodynamic response.

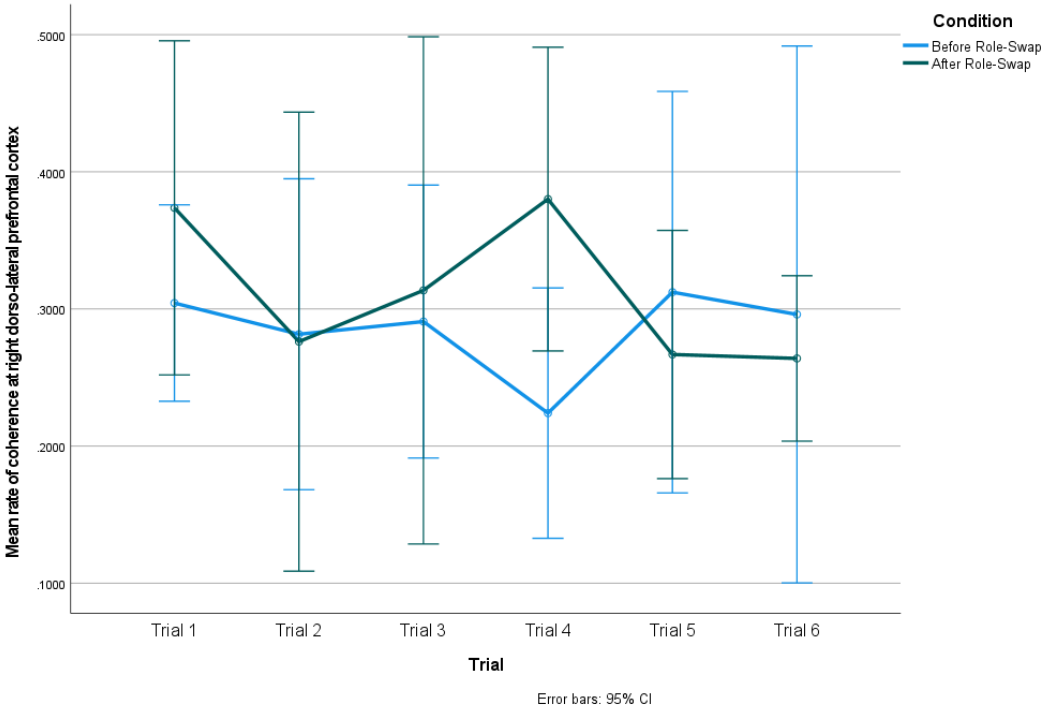

Figure 66: Mean rates of coherence based on the right dorso-lateral prefrontal cortex hemodynamic response.

4.4.2. Hemodynamic Response Analysis Results

The hemodynamic response analysis focusing on the oxygenated hemoglobin ratio in the frontal cortex of one of our dyads resulted in a 17 participant analysis for four blocks and all available optodes, and four cortical regions.

Evaluating for the overall effect of repeated trials in form of the four experiment blocks and the possible interaction effect of distinct conversational roles; firstly, the results of the Mauchly's test of sphericity on the overall data containing all optodes for both roles showed that the assumption of sphericity was violated $\chi^2(5) = 242.936$, $p < .001$, therefore values for Greenhouse-Geisser corrected results are reported. The repeated measures ANOVA showed that there was a significant effect for the within-subject measure of the experiment blocks F (2.038, 464.694) = 115.582, $p < .001$, partial η^2 = .337. Furthermore, for the interaction effect of conversational roles, again there was a significant effect F (2.038, 464.694) = 4.865, p = .008, partial η^2 = .021. The mean values for all optodes across the four experiment blocks and for both conversational roles are shown in Figure 67.

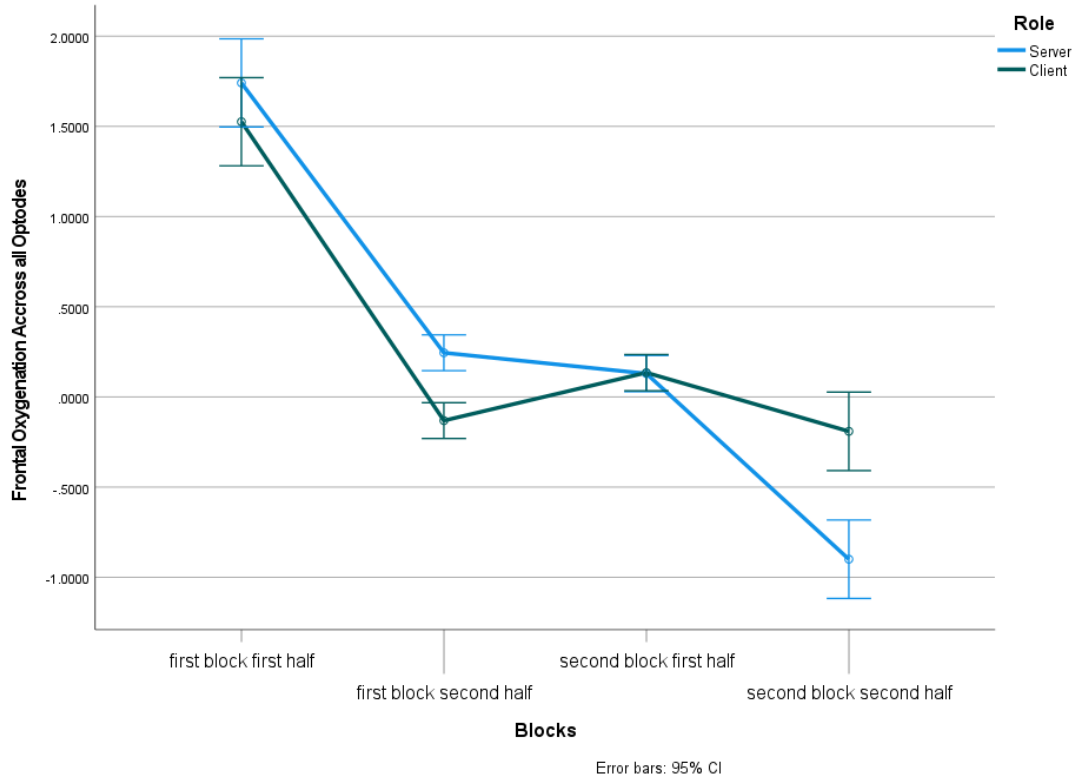

Figure 67: Mean rates of HbO concentration for participants in the joint matching task condition for each trial. Recall that after the first block (warm-up and 6 trials) there is a role-swap condition, hence the notation for server and client instead of director and matcher, where server denotes the participant who is the first director and the second matcher, as the client denotes the participant who is the first matcher and the second director.

Next, each optode was investigated individually for their contributions to this effect and to elucidate where in the frontal cortical regions the effects of experiment blocks was significant specifically.

In optodes 1, 2, 3, 5, 7, 9, 10, 11, 13, 15, and 16 there were significant effects of experiment blocks, demonstrating that during the process of establishing a common ground, participants' rates of oxygenated hemoglobin were significantly affected in both dorsal (1, 2, 3 for left dorsal, 13, 15, 16 for right dorsal) as well as in fronto-polar regions (left medial frontal regions of 5, 7, and right medial frontal regions of 9, 10, 11). Mean values for each of the 11 optodes the analysis for which showed significant experiment block effects are shown in Figures 68 – 78, with the ANOVA results shown in Table 10. For optodes 5, 7, 10, and 16 Mauchly's test of sphericity showed that the assumption for sphericity was violated, so Greenhouse-Geisser corrected values are reported for these optodes with the test results shown next to optodes in Table 10.

Table 13: Results of repeated measures ANOVA analysis of experiment blocks for HbO concentration on each fNIRS optode.

Optode 1	*F (3, 21) = 7.085*	p = .002	η^2 = .503
Optode 2	*F (3, 39) = 10.016*	p < .001	η^2 = .435
Optode 3	*F (3, 24) = 13.427*	p < .001	η^2 = .627
Optode 5 ($\chi^2(5)$ = 12.394, p = .031)	*F (1.456, 11.647) = 18.787*	p < .001	η^2 = .701
Optode 7 ($\chi^2(5)$ = 11.588, p = .044)	*F (1.506, 10.540) = 13.159*	p = .002	η^2 = .653
Optode 9	*F (3, 24) = 10.888*	p < .001	η^2 = .576
Optode 10 ($\chi^2(5)$ = 12.381, p = .032)	*F (1.919, 15.352) = 5.711*	p = .015	η^2 = .417
Optode 11	*F (3, 27) = 10.808*	p < .001	η^2 = .546
Optode 13	*F (3, 18) = 4.120*	p = .022	η^2 = .407
Optode 15	*F (3, 9) = 5.954*	p = .016	η^2 = .665
Optode 16 ($\chi^2(5)$ = 12.651, p = .027)	*F (1.928, 25.070) = 5.770*	p = .009	η^2 = .307

The between subjects effect of conversational role (director or matcher) was not significant for any optodes, which demonstrates that the workload of communicating towards a common goal, establishing a common ground, and communicating more efficiently as a result were homogeneous workloads among these roles.

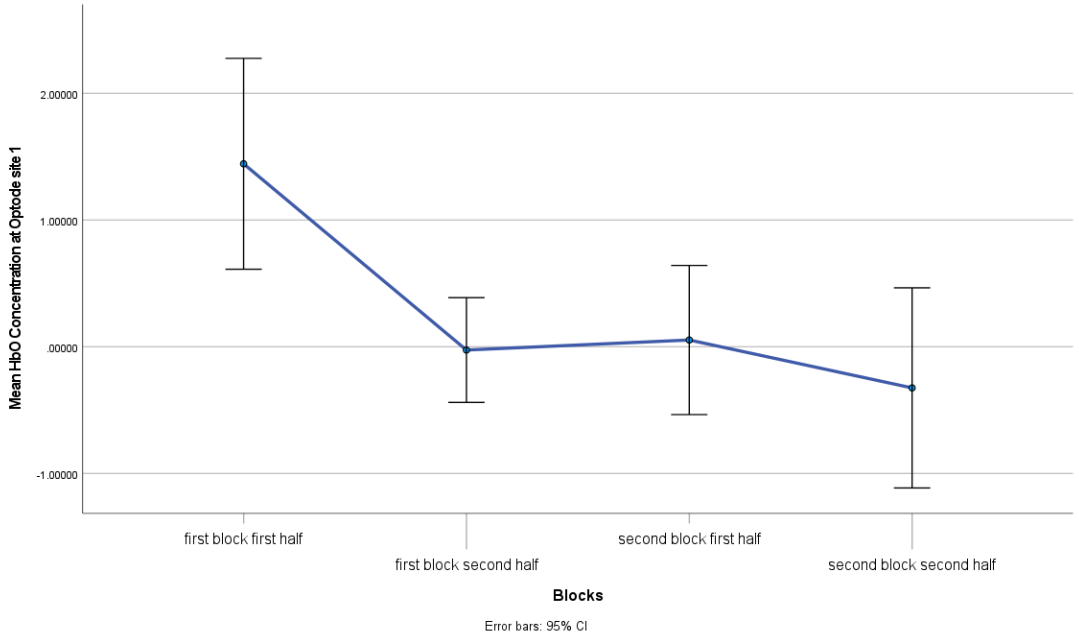

Figure 68: Mean delta oxygenated hemoglobin (HbO) at Optode site 1, left dorsolateral prefrontal cortex.

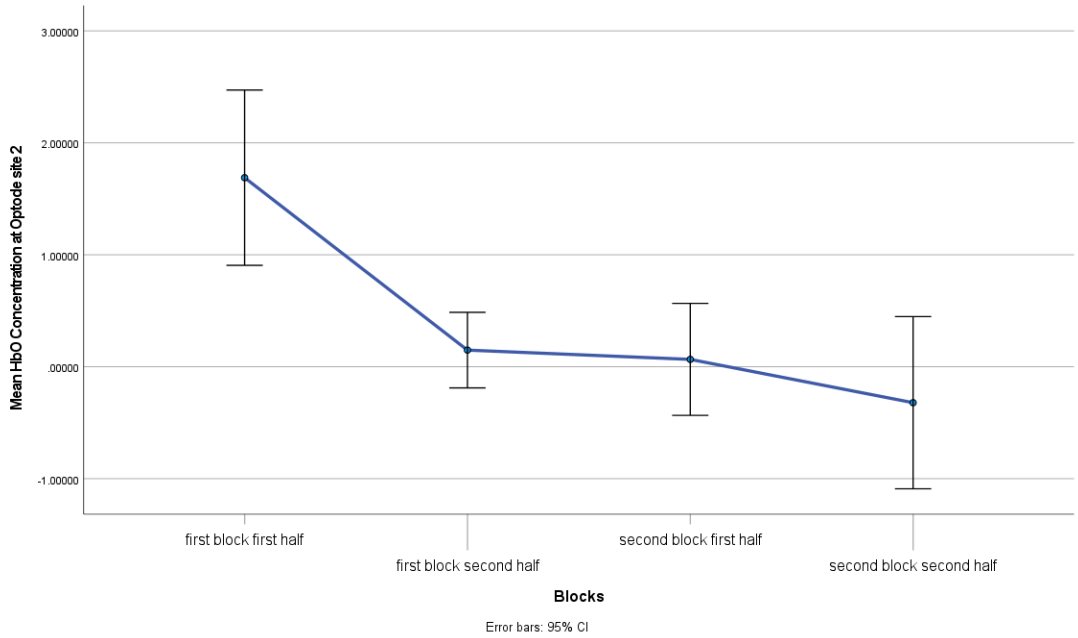

Figure 69: Mean oxygenated hemoglobin concentration (HbO) at Optode site 2, left dorsolateral prefrontal cortex, relative.

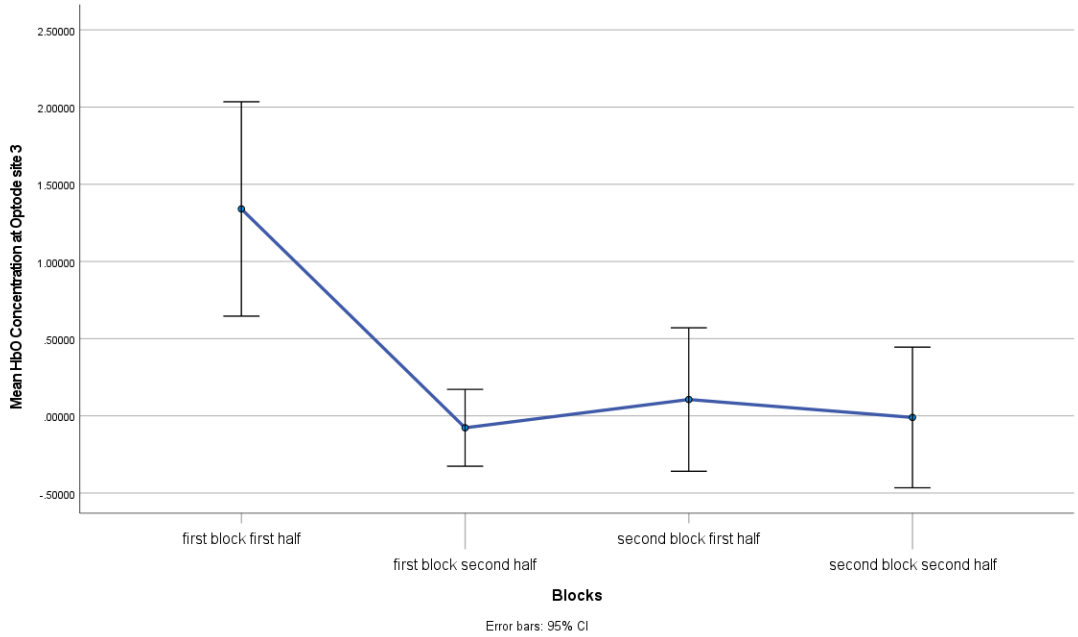

Figure 70: Mean oxygenated hemoglobin (HbO) at Optode site 3, left dorsolateral prefrontal cortex.

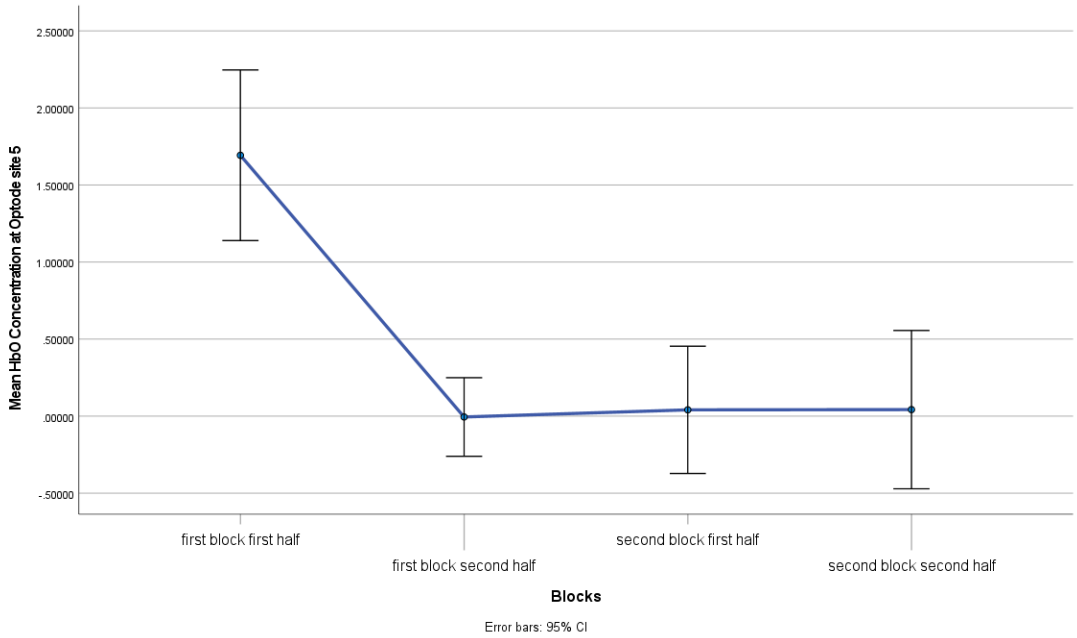

Figure 71: Mean oxygenated hemoglobin (HbO) at Optode site 5, left dorso-medial prefrontal cortex.

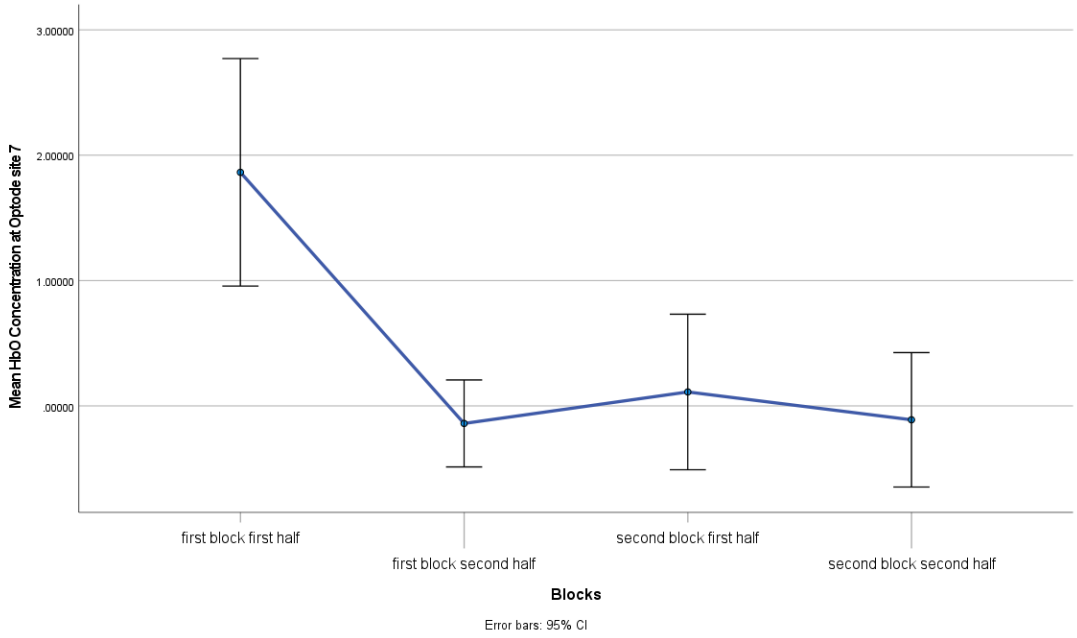

Figure 72: Mean oxygenated hemoglobin (HbO) at Optode site 7, left medial prefrontal cortex.

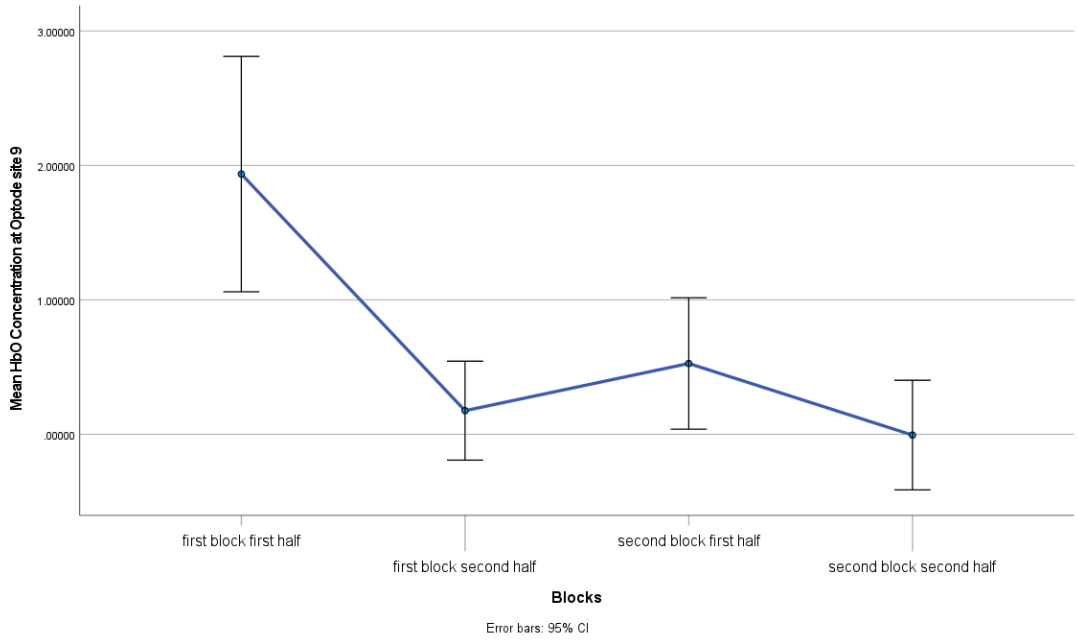

Figure 73: Mean oxygenated hemoglobin (HbO) at Optode site 9, right medial prefrontal cortex.

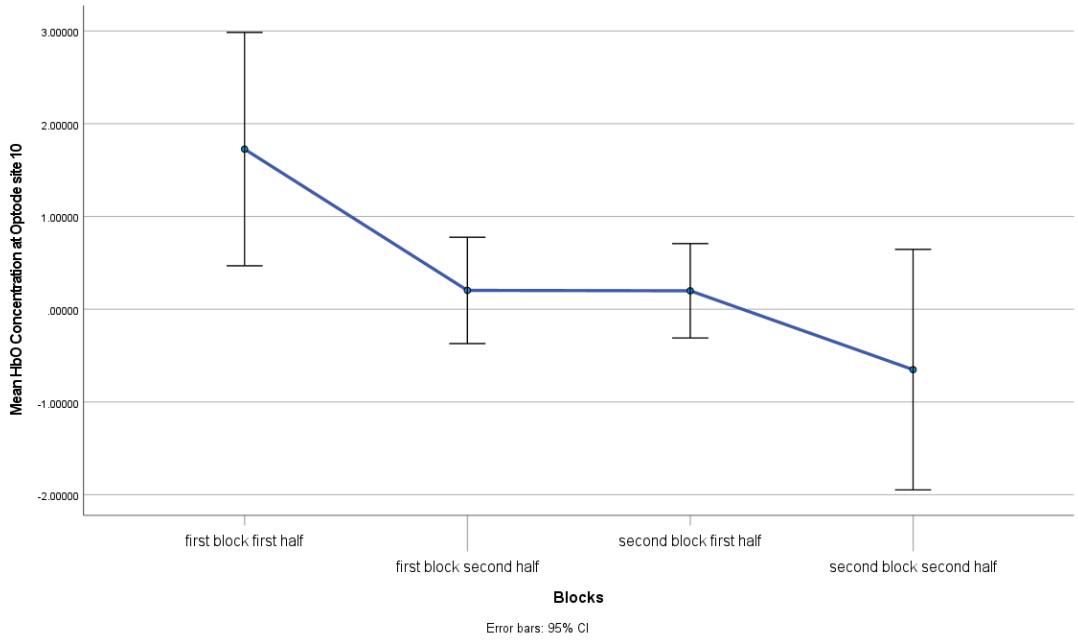

Figure 74: Mean oxygenated hemoglobin (HbO) at Optode site 10, right medial prefrontal cortex.

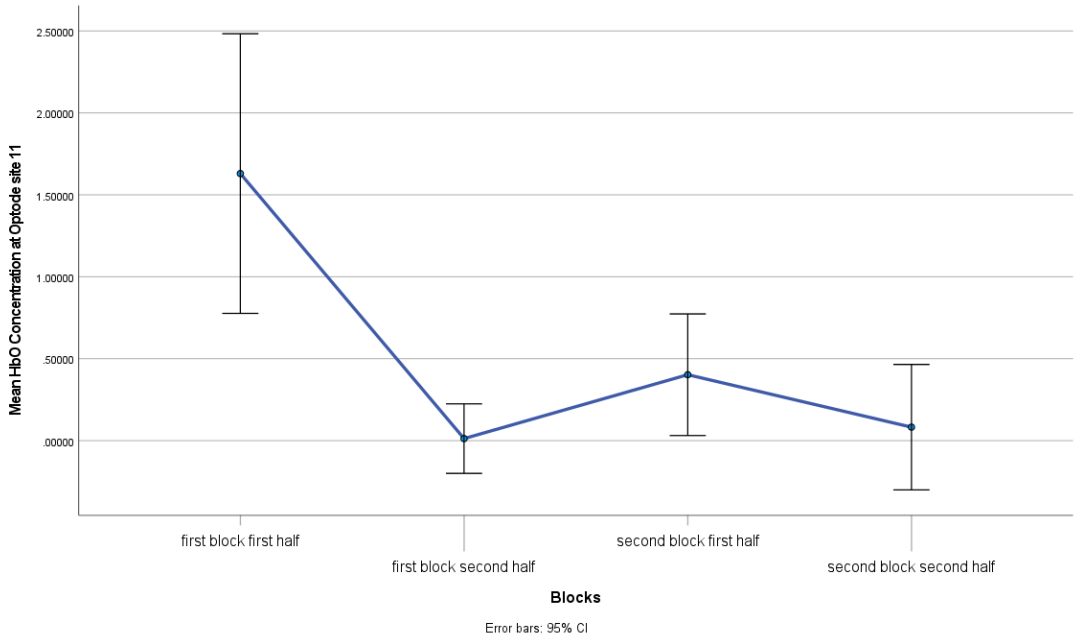

Figure 75: Mean oxygenated hemoglobin (HbO) at Optode site 11, right dorsomedial prefrontal cortex.

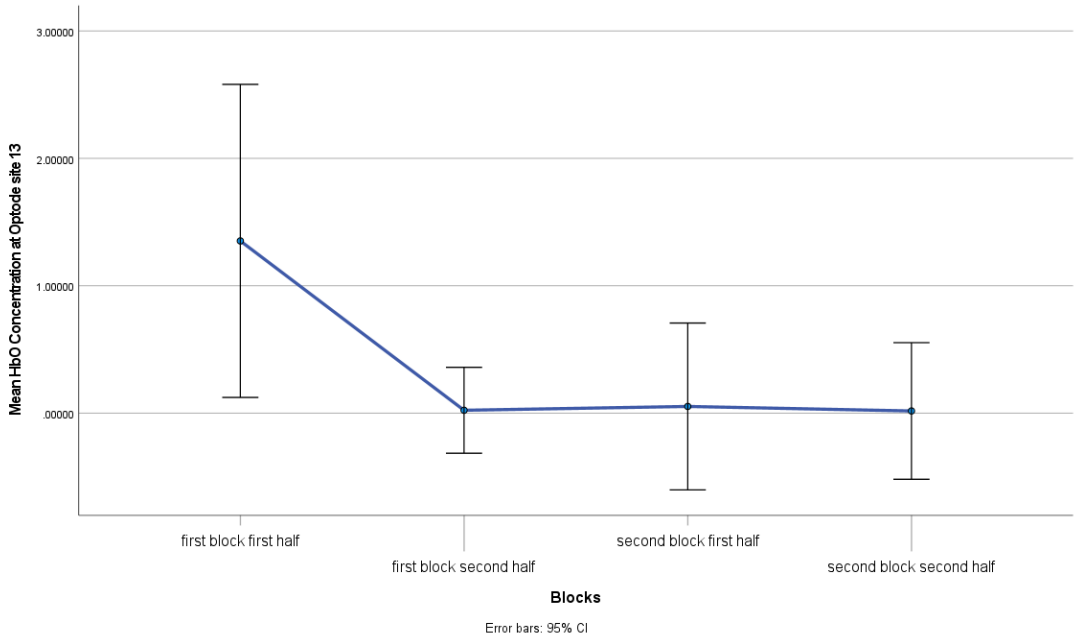

Figure 76: Mean oxygenated hemoglobin (HbO) at Optode site 13, right dorsolateral prefrontal cortex.

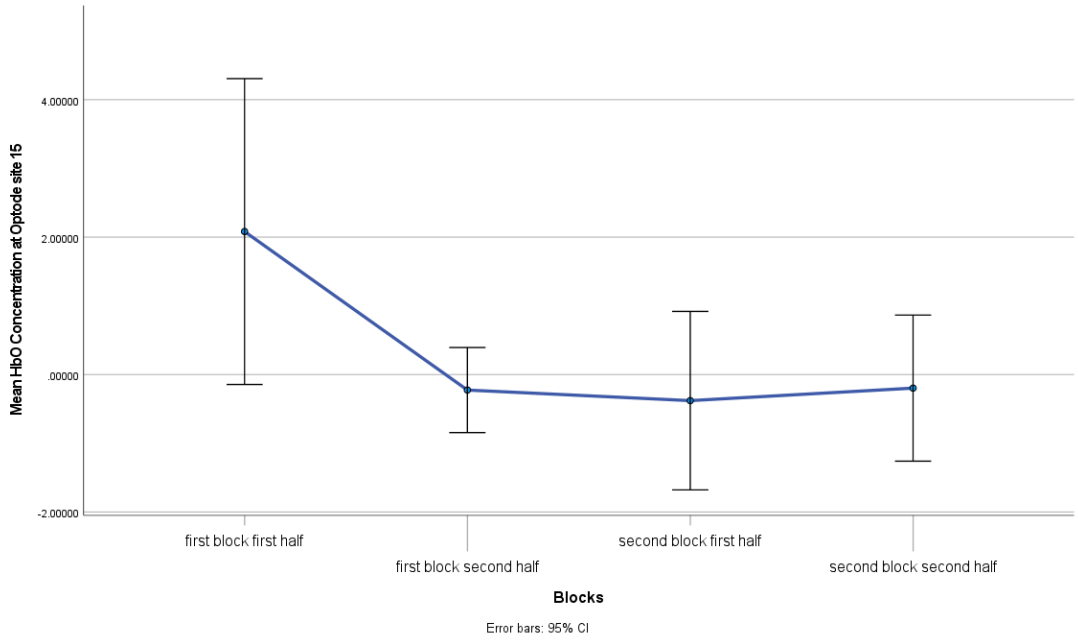

Figure 77: Mean oxygenated hemoglobin (HbO) at Optode site 15, right dorsolateral prefrontal cortex.

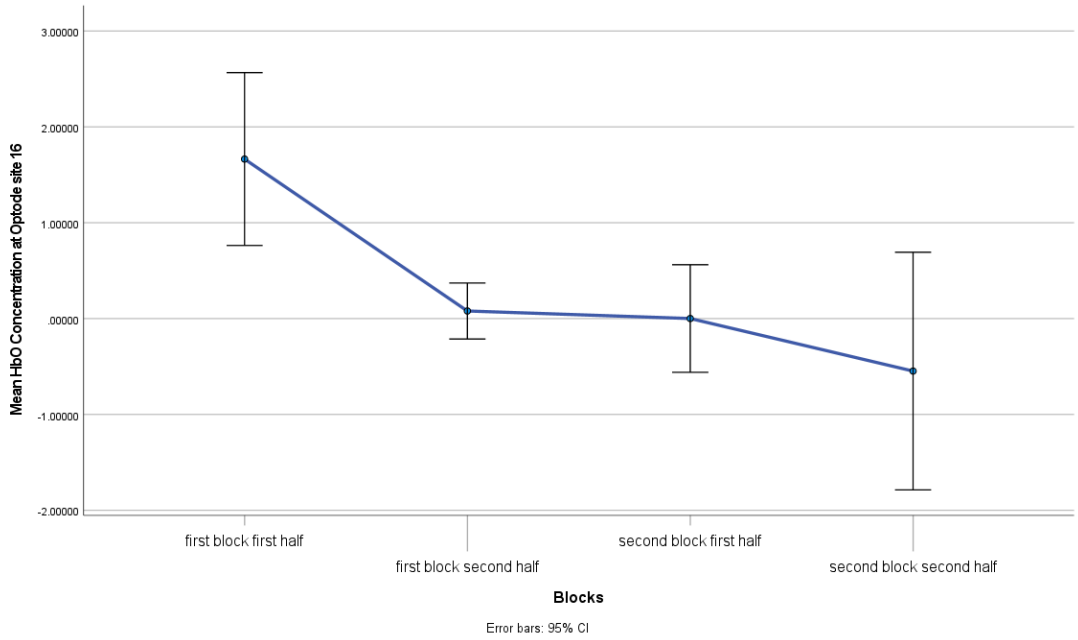

Figure 78: Mean oxygenated hemoglobin (HbO) at Optode site 16, right dorsolateral prefrontal cortex.

CHAPTER 5

5. DISCUSSION

In the present thesis, the pragmatic process that might take place in a human-human interaction setting was evaluated for its neural and ocular correlates, which were simple to reproduce in ecologically valid interaction settings. The employed experiment setting constituted of a dyadic, verbal interaction setting for participants who sat in a co-located manner. With experiments on both participants' screens as well as all experiment apparatus (EEG, eye-tracker, fNIRS) including environmental monitoring devices (microphone, video camera) synchronized on the same time domain, this setting allowed the participants to be presented a shared goal and for their shared task to be repeated in a controlled manner. The task was to match and sort twelve Tangram figures per each trial by communicating verbally. The setting employed consequently allowed for an investigation in which a common goal was always present to the dyad as they sustained joint attention and cooperated (Grice, 1975, 1978) to communicate better (Brennan & Clark, 1996; Clark & Brennan, 1991; Clark & Wilkes-Gibbs, 1986), constituting to an entrainment process that might have been an automatic process (Garrod & Pickering, 2009). Finally, a role-swap condition in half point of the experiment allowed controlling for partner specific effects, for when a partner might be better at one of the two roles (director and matcher) as well as for whether a presently established common ground would be sustained, to replicate partner specificity effects in (Metzing & Brennan, 2003) in Turkish, the native language of all participants.

Clark & Wilkes-Gibbs (1986) defined referring as a collaborative process, wherein the globally conventional and globally unconventional meanings afforded to each dyad at the beginning of their task would converge into local meanings agreed upon by interlocutors as they communicated; at times, even converging to the form of definite descriptions that are available to only the two interacting participants. While, Sebanz, et al. (2006) claimed that a collaborative process such as referring would necessitate more than action imitation, instead these are joint actions where interlocutors collaborate by complementing the actions of the other; an important perspective of joint action preceded by (Clark & Brennan, 1991), while others claimed that such actions might achieve more than they could on their own, rendering the manner in which some joint actions take place distributed cognitive processes (Hutchins & Tove, 1999).

In the behavioral basis of the present thesis, firstly, the results from the pivotal study by Clark & Wilkes-Gibbs (1986) were replicated successfully in Turkish; as the results of the behavioral linguistics analyses showed that dyads not only got more accurate, but they took less time, spoke in fewer utterances, with fewer words in each utterance, and while taking fewer turns on the conversational floor. From the perspective of organizational behavior, as indicated by turn-taking, linguistic measures such as the number of sentences and words per sentence, and the total effort, measured by the time taken by dyads to match and sort their twelve Tangram figures, the joint task provided a sustained, sound, and valid basis for pragmatic entrainment of aligned concepts. This process, characterized as a gradual improvement in communication efficiency, aligns with the theories of Clark & Wilkes-Gibbs (1986), and further elaborated by Clark & Brennan (1991) and Brennan & Clark (1996), which emphasize the joint and historical nature of communication. These results showed that the experiment setting was successful; the participants were able to commit into a joint action, coordinate their actions towards their common goal and establish a common ground. This enables the present thesis to correlate neural and ocular correlates with the emergence of this shared mental construct, which necessitated intersubjective understanding (Baldwin, 1995) and a total cognitive ability including perceiving affordances of figures, their actions, and shapes to extract identifying and distinguishing features to communicate that perhaps spans beyond a single mind (Cooke et al., 2012; Hutchins & Tove, 1999) as demonstrated by conversational dynamics among our dyads.

Furthermore, with the analyses, it can be observed that there is more to entrainment in communication than the approaches already established in the previous literature, in which the consensus is that a novel description might be first presented, then it can be accepted by the other(s) or there can be a rejection as well as a request for clarifying the original description, for repairing or repeating it. Our behavioural data together with our correlates would suggest that the complex mechanism of grounding in communication, in interaction with the principle of least effort, an enunciated automatic process by Garrod & Pickering (2009) to communicate better can result in once established "accepted" descriptions to be over-simplified, for once accepted referents to go through the same process yet again or even go through semantic changes without breaking the referent mapping among the participants. This underpins the existence of a reiteration process that expands the previous consensus, perhaps elucidating an interplay with historic accounts (Brennan & Clark, 1996) or memory processes (Horton & Gerrig, 2005). Additionally, this process shows how strong a common ground can become without losing its dynamicity in natural language that as long as the subjective meaning (locally available to our dyad and us in this case) doesn't change; then, the syntax and the semantics can yet change, possibly posing a challenge for modelling approaches of future studies. Below, in Table 11-16, the first set of examples are presented for such dynamic linguistic behavior.

Table 14: Example description 1, Dyad 10, warm-up trial. Turkish original above, English translation below. D for director speech, M for matcher speech.

Warm-up	D	üçüncüsü aslında diyorlar ya sırtını şeye dayamış
	M	şey, sırtında çok uzun bir palto varmış da kamburu varmış gibi böyle geriye vermiş kendini, o mu
	D	evet o

	M	mu, kulakları yok ama bak küçük kare bir kafası var
	D	küçük kare bir kafası var, aynen
	M	minicik
Warm-up	D	The third one is actually, as they say, he's got his back to something.
	M	Well, he's got a very long coat on his back, and he's backwards like he's got a hunchback, is that it?
	D	The third one is actually, as they say, he's got his back to something.
	M	Yes, it is.
	D	mu, he doesn't have ears, but look, he's got a little square head.
	M	he's got a little square head, just like that.
	D	tiny

Table 15: Example description 2, Dyad 10, First block first trial. Turkish original above, English translation below.

Trial 1	D	birincisi sırtını duvara dayamış var ya. İlk
	M	sırtını sağımıza mi dayamış
	D	solumuza, bizim solumuza
	M	tavşan kafası olan mi
	D	yok o değil ya. İlk basta hani, başladığımızda şey yapmıştık ya. Yürüyor ama sırtını duvara dayamış, palto varmış gibi
	M	tamam küçük kare kafası olan
	D	küçük kare kafası olan
	M	bir
Trial 1	D	the first one with his back against the wall. The first
	M	with his back to our right?
	D	to our left, our left
	M	the one with the rabbit head
	D	No, not that. You know what we did at first, when we started. He's walking but with his back against the wall, like he's wearing a coat.
	M	Okay, the one with the little square head
	D	with a little square head
	M	one

Table 16: Example description 3, Dyad 10, First block second trial. Turkish original left, English translation right.

Trial 2	D	on bir, palto giyip sırtını sola veren	eleven, wearing a coat and giving his back to the left
	M	sola	left
	D	evet sola doğru veren	Yes, the one that gives to the left

The descriptions given in Tables 11, 12, and 13 belong to the same Tangram figure, Tangram number 3, which is shown in Figure 79. By focusing on the approach of the first director here, because this is the first block of the experiment before the role-swap condition, it can be seen that initially in the warm-up trial that the figure's posture (that the figure is subjectively leaning backwards to our left as observers), its body shape (that it has a hunchback), its long coat, and its square head being a few distinguishing factors for the dyad in focus. For the warm-up trial this was the case, whereas in Trial 1 the affordance that the figure is perceivably conducting the action of walking comes into play, yet in both of these trials the feature "square head" appears to be the decision point. It must be remarked at this point that some features, for in the present case, the hunchback feature, is presented first by the matcher, rather than the director. This is a very common occurrence throughout the corpus of interaction that was generated during the present thesis; in which, some features that are identifying or distinguishing are not always afforded or perhaps even perceived by the director whose task was to direct, as the matcher listened. These occurrences provide insight into the distributed nature of how this entrainment process can take place, even with distinct hierarchical roles established.

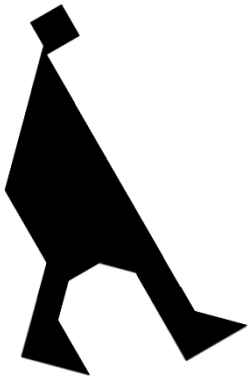

Figure 79: Tangram # 3.

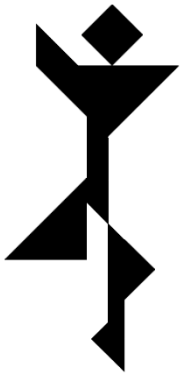

Figure 80: Tangram # 2.

Moving onto Trial 2, it can be seen that in between the first figure of Trial 1 and the eleventh figure in Trial 2, which are where Tangram 3 were located as our focus figure in this example, the dyad might have decided that a square head was not a distinguishing feature. As shown in Figures 2 and 3 in the Experimental Background section under the Methodology section, this is the correct strategy; as almost all figures have a "square head". It must be noted at this point in the example that the dyad took 7 minutes and 28 seconds to get from warm-up, 3rd figure to Trial 2, 11th figure, yet they were able to recall their "coat" feature.

Table 17: Example description 4, Dyad 10, First block sixth (last) trial. Turkish original left, English translation right.

Trial 6	D	paltolu olan	The one with the coat
	M	palto	coat

In Table 14 is shown the total interaction of the dyad #10 for Tangram #3 in their last trial before the role-swap condition. At this point, it is valid to observe that the feature "coat" was sufficient in distinguishing this figure for this dyad. However, in Table 15, and following the role swap, the description for the same figure is given where the new director chooses to reintroduce the feature "long coat" for the first time since the warm-up trial in the first block, this recall action spanned a period of 12 minutes and 11 seconds, and a total of 91 matching and sorting actions.

Table 18: Example description 5, Dyad 10, Second block, second trial. Turkish original left, English translation right.

Trial 8	D	on birincisi paltolu adam. Uzun paltolu	The eleventh is the guy with the long coat. Long coated man
	M	tamam	okay

Following the example in Table 15, the experiment proceeds (scoped down specifically for Tangram number 3) without any change until the director commits to the following error in Table 16, in trial 12.

Table 19: Example description 6, Dyad 10, Second block, fifth trial. Turkish original left, English translation right.

Trial 12	D	on paltolu uzun insan	The tenth is the tall guy with the coat
	M	tamam	okay

The error that the director committed in this trial caused no observable differences in the actions of the matcher participant, possibly as the "coat" feature still stands, and that the "long" property of coat being mistakenly transferred to the character itself that it is a "tall" man was perhaps a secondary feature that was not part of the common ground. In the present thesis, the linguistic aspect is observed solely as foundations for neural and ocular correlates to be grounded upon, and a more focused, further research is necessary to elucidate how this observed durability of an established common ground might interact with neural correlates, for example with an EEG ERP study similar to (Costanzo et al., 2013).

In the second example for how fluid the referent common ground can become, data from dyad 23 as they conversed to match and sort Tangram figure 2, Figure 80, starting in Table 17 with description 7. This example was chosen specifically as Tangram number 2 was this dyad's most difficult figure, with their establishment of a stable common ground occurring at the fourth trial, compared with their other eleven figures, which occurred at the third trial.

Table 20: Example description 7, Dyad 23, warm-up trial. Turkish original above, English translation below.

warm-up	D	Ee, kolları açık gibi düşünebiliriz. Kafası şey ee, iskambil kartlarında var ya, karo muydu
	M	tamam
	D	onun gibi. ee, işte s. Su an benim baktığım sağ tarafta kalan ayağı
	M	hi hi
	D	ee, alt kenarın sağ ortasında. Sol ayağı da sol kenarın ortalarına yakın
	M	sol kenarın
	D	bir coşku halinde gibi bu
	M	ortalarına yakın. Ee, kafası karo seklinde
	D	evet, biraz yayvan
	M	sağ ve solları var
	D	kolları açık gibi. Sag kolu, ya onun sağ kolu daha yukarı gibi, benim solumda kalıyor.
	M	hmm
	D	beli ince, düz
	M	beli ince ve düz. Peki, ince düz, uzun mu, ince düz, kısa mi
	D	ee
	M	iki tane benzer sekil var çünkü
	D	ince kısmı ee, kısa aslında. Tam ortalara doğru denk geliyor
	M	kolları açık ee, kafası karo seklinde
	D	sağ bacağı böyle ee, bir dik üçgen seklinde geniş
	M	dik üçgen seklinde geniş. Peki dik üçgen tam mi yoksa bacağına mi bağlanıyor
	D	bağlanıyor. Buldun mu
	M	tamam, buldum
	D	okey
warm-up	D	Uh, you could think of it as having open arms. His head is like, uh, you know in playing cards, diamonds?
	M	OK.
	D	Like him. Uh, here's the s. The one I'm looking at is the foot on the right side.
	M	hi hi
	D	uh, right center of the bottom edge. His left foot is near the center of the left edge.
	M	left edge
	D	it's like in a state of ecstasy.
	M	near the middle. Well, his head is shaped like a diamond.
	D	Yeah, it's a little splayed.
	M	they have left and right
	D	like his arms are open. His right arm, like his right arm is higher, my left arm is lower.
	M	Hmm.
	D	waist thin, straight
	M	her waist is thin and straight. So, slim straight, tall or slim straight, short?
	D	uhm

	M	because there are two similar shapes
	D	the thin part is, uh, short, actually. It's right in the middle.
	M	his arms are open, uh, his head is shaped like a diamond.
	D	the right leg is like, uh, wide in the shape of a right triangle.
	M	It's wide in the shape of a right triangle. So is the right triangle complete or does it connect to the leg?
	D	connecting. Did you find it?
	M	Okay, I got it.
	D	Okay

Table 21: Example description 8, Dyad 23, first block first trial. Turkish original left, English translation right.

Trial 1	D	bu böyle hani kol. Yine soyut bir tip. Kolları açık	you know, that's the arm. Again, an abstract type. Arms outstretched
	M	bana makas gibi dediğin	what you call me like scissors
	D	ne	what
	M	makas gibi dediğin	what you call scissors
	D	yok o değil, o değil bu. Ee, yani bir bacağı dümdüz üçgen, dik üçgen	No, not that one, not that one, this one. So, so one leg is a straight triangle, one leg is a straight triangle
	M	dik üçgen	right triangle
	D	kafası yayvan bir karo. Düz ama. Yayvan bir karo, sağ kolu da havada gibi	a tile with a splayed head. It's flat. It's a splayed diamond, and his right arm seems to be in the air
	M	tamam	OK
	D	buldun mu	Did you find it?
	M	buldum	I found it

Table 22: Example description 9, Dyad 23, first block second trial. Turkish original left, English translation right.

Trial 2	D	ee, bunu anlatmakta zorlanmıştım. Bir bacağı dümdüz üçgen, iste bir	Well, I had a hard time explaining that
	M	tamam	OK
	D	sağ kolu havada gibi. Okay	like his right arm is in the air. Okay
	M	hi hi	hi hi

Table 23: Example description 10, Dyad 23, first block third trial. Turkish original left, English translation right.

Trial 3	D	okay ee, bacağı üçgen. Sag kolu havada. Anlatmakta zorlandığım var ya. Geniş karo kafası var	Okay, uh, his leg's a triangle. His right arm is in the air. You know what I'm having trouble explaining? He's got a large diamond head.
	M	geniş karo. Tamam	wide diamond. Okay

	D	okay	okay
	M	bir dakika. Bir şey soracağım, sağ kolunun havada olduğuna emin misin	Just a minute. Let me ask you something, are you sure your right arm is in the air?
	D	yani ben bana bakıyormuş gibi düşünüyorum. O zaman sağ kolu havada ama başka yöne bakıyorsa sol kolu da havada olabilir	I mean, I think he's looking at me. Then his right arm is in the air, but if he's looking the other way, his left arm could also be in the air
	M	bence sol kolu	I think his left arm
	D	tamam okay, fark etmiyor	okay okay, it doesn't matter

With the example description 10 in Table 20, the first example of how the order of figures might have contributed into the grounding process is presented; wherein the director always explains them in the given order and with the example in Table 19, the figure 2 was the 8th figure in the order, whereas with the example in Table 20, it was the first one. So, despite that there was only 36 seconds in between these two figures in the adjacent trials 2 and 3 respectively, the figure that the dyad had the most difficulty in distinguishing from Tangram number 2 (as can be observed in Trial 1 – Table 18 when the dyad had a moment of confusion for the feature "scissor like arms") was still available, and therefore necessitating an expansion in Trial 3 – Table 20, after the matching task was easier in Trial 2 – Table 19. To remark, following Trial 3, the dyad establishes their common ground with sufficient strength that their next trials until the role-swap condition as well as a couple of trials after the role-swap condition without issue with the following features "right hand up" and "triangle leg", as shown in Table 21.

However, with trials 10, 11, and 12 both an error similar with that exemplified in Tables 15 and 16, where the director confuses between "leg" and "foot" features can be observed; yet, again the grounding was strong enough not to be affected (error between Trials 11 and 12), and another where the matcher gets confused and the grounding is affected (Trial 10), as evidenced by the matcher's request for clarification as they observably thought the two features presented by the director meant two different figures for a moment until the reassurance came. The dyad then complete trial 13, the final trial, without issues.

Table 24: Example description 11, Dyad 23, first block last three trials. Turkish original left, English translation right.

Trial 5	D	ee sağ kolu havada dediğim, bir ayağı komple üçgen falan	er right arm in the air I said, one leg is completely in a triangle or something
	M	tamam	okay
Trial 6	D	sağ kolu havada, üçgen bacak	right arm in the air, triangle leg
	M	tamam	okay
Trial 7	D	sağ kolu havada	right arm in the air
	M	tamam	okay

Table 25: Example description 12, Dyad 23, second block third trial (trial 10). Turkish original left, English translation right.

Trial 10	D	üçgen bacak, sağ kolu havada	triangular leg, right arm in the air
	M	dur yavaş, yavaş. Üçgen bacak sağ kolu havada, tek kişi değil mi o	Stop, slow down, slow down. The triangular leg with the right arm in the air, isn't that one person?
	D	nasıl	how do you mean?
	M	sağ kolu ve üçgen bacak ayni kişi değil mi o	Isn't that the same person with the right arm and the triangular leg?
	D	evet ayni kişi	Yes, it's the same person.
	M	tamam	Okay.

Table 26: Example description 13, Dyad 23, second block fourth and fifth trials (trials 11 and 12). Turkish original left, English translation right.

Trial 11	D	sağ kolu havada, üçgen ayak	right arm in the air, triangle foot
	M	evet	yes
Trial 12	D	sağ kolu havada üçgen bacak	right arm in the air, triangle leg
	M	evet	yes

In dyadic linguistic communication, observations from the behavioral data are that partners working towards a stable common goal do indeed go through the process of establishing a common ground. This gradual process takes place by the participants first retrieving their globally available meanings, for the features they are afforded by their own perceptions of actions and/or shapes of the Tangram figures, to map their local meanings. However, the process both spans beyond a single mind, providing foundations for our ocular and neural correlates, as well as allows us to explicate that their locally common referent the dyads agree upon might not be that they alter their local meanings wholly, but instead form referents to those local meanings themselves. This is evidenced by the interaction of directors and matchers as they communicate, wherein both parties could present new referents, which would sometimes be even definite descriptions, such as pronouns. These findings both support the theory that for a joint action to take place, interlocutors need to both observe others' actions continuously, incorporate them into their own, if necessary (Sebanz et al., 2006; Sebanz & Knoblich, 2009), and keep their conceptual representations up to date at all times (Brennan & Clark, 1996; Clark & Brennan, 1991). The evidence that all dyads successfully established common grounds, in line with prior theories, not only validates the current thesis but also replicates significant pragmatic findings from past literature in Turkish. This study contributes a robust corpus of free-form conversation of dyadic interaction for establishment of a common ground in this language, providing a valuable resource for future research.

As a last point regarding the outcomes of our behavioral results, the possible distinction observed by the linguistic transcriber due to the organizational setup of our dyads; in that, there were mixed sex dyads and same sex dyads (both female-female and male-male groups existed) shown only a change in significance for the entrainment

effects in the accuracy of our dyads. With the other metrics we evaluated the significance did not change, while effect sizes were indeed objectively different for all organizational setups. For the present thesis, it was not hypothesized that there would be organizational differences between groups as demonstrated, and therefore the number of dyads and their organization were not controlled for this; resulting in that there were only 4 female – female, and 5 male – male groups, with the remaining 12 that were evaluated being mixed, and therefore these effects necessitate a more focused, further research with controlled groups to evaluate such distinctions.

In the present thesis, a hypothesis was presented in which the gradual process of establishing a common ground and the linguistic entrainment process that must take place for this establishment would entail neural and ocular correlates. The results demonstrate that this is indeed the case.

Firstly, with the ocular correlates, which was evaluated with the use of an eye tracker the gaze interaction of our participants by employing CRQA and TME analyses to investigate, respectively, how the synchronized gaze vectors of each of our participants aligned for their distributions on the AOIs defined by the twelve Tangram figures as the experiment progressed as well as how their matching and sorting task interacted with the entropy of their eye movements. The cross-recurrence analysis results showed a significant distinction from control condition in the evaluation of natural gaze data and shuffled data comparisons; in that, for all trials, the gaze interaction showed increased gaze cross-recurrence. However, the nature in which the repeated trials condition affected gaze cross-recurrence was not the same through all experiment stages. For instance, in the beginning parts of the experiment setting, where the dyads were presented with the twelve Tangram figures for the first time, their distinct (from a behavioral linguistics perspective and both in between subject evaluations and sometimes even interacting significantly with dyadic organization) strategies resulted in what can be considered a baseline gaze cross-recurrence rate (Figure 42, warm-up trial). Dyads, furthermore, first experienced a global increase in the ratio of recurrent gaze as they gazed upon the same figure, for a couple trials and they progressed from their first presentations for descriptions to match each figure in their warm-up and Trial 1 trials but afterwards the recurrence began to drop as the repeated trials progressed, with dyads showing less gaze cross-recurrence with each new trial until the role-swap condition where the effect recovered, to again drop until a plateau around the 10^{th} trial (Figure 42) by which time the dyad had again strengthened their common ground sufficiently. The present thesis is that this is a direct effect resulting from an increase to the level of entrainment, and in line with the findings from (D'Angelo & Gergle, 2016), resulting in dyads that did not need to gaze at the same figures for as long as it was required initially, following establishment of a common ground. For instance, when going through the feature extraction phase where dyads first accessed to their globally available concepts for what actions or shapes drawn into the figures might afford to them, to then agree upon local meanings with their partners to establish these as their common ground. Finally, the time window at which the gaze cross-recurrence was at its highest was when matcher participants' gaze was shifted by one second, a period one second shorter than in the work by Richardson & Dale (2005) and Coco & Dale (2014) but in line with theoretical discussion regarding ocular correlates of establishing a common ground from a neural perspective by (Stephens et al., 2010).

The effect of repeatedly matching and sorting the 12 Tangram figures on gaze cross-recurrence rates in the light of our behavioral results reflect that as per Richardson & Dale (2005, 2009), conversational partners' gaze patterns do indeed become more aligned as they converse towards a common goal, however, this effect interacts significantly with the linguistic foundations in the present thesis that mutual gaze as a non-verbal form of coordination is only secondary to verbal, linguistic formation of a common ground: As the linguistic foundation strengthens, the non-verbal domain appears to lose its importance.

This is further supported by the stage for the gaze correlate of linguistic entrainment that the entropy of gaze movements is lowered with the linguistic foundation for an established common ground is strengthened among the dyad. Furthermore, the correlation between transition matrix entropy and participants' role in the sorting and matching task was also a distinguishing factor; allowing for organizational roles among the dyads to be categorized (i.e., between roles of director and matcher or speaker and listener).

These results summarize to that an established verbal common ground allows for gaze interaction to become secondary, allowing for a more explicit and less ambiguous communication local to the interlocutors that share the common ground and with a diminished need for shared gaze interaction, as well as for less ocular activity per the TME results, both of which demonstrate the efficiency entrainment in communication can bring. Specifically, TME that was also explored in a similar context by (Metzing & Brennan, 2003), is a metric that can be generalized to entropy of gaze interaction overall, allows for role and task distinctions very reliably for future modelling approaches to human – human, and perhaps human – others interaction research. Overall, CRQA and TME serve to function together as correlates for workload, task, and organizational distinctions in pragmatics; allowing for a three stage distinction of the interactional state of interlocutors. The first stage, wherein, high entropy and low recurrence maps to a higher workload state as evidenced both by lower performance and lower efficiency in the behavioral metrics and experiment design, where initially a visual feature extraction process and a linguistic process is combined to describe visual cues and features, which directors initially perceived. By the second stage, as most dyads established a shared set of referents into a common ground for the set of 12 Tangram figures, which is evidenced by increasing behavioral performance, correlated with an increase in gaze recurrence and decreasing gaze entropy. The reverse trends in the two metrics highlights a non-linear relationship inherent in the gradual process of aligning representations into a common ground, wherein the correlates are as dynamic as the grounding process. Finally, in the third stage and once dyads have successfully achieved a common ground by agreeing upon and strengthening their agreed-upon local meanings, then gaze entropy decreases further, this time together with ratio of recurrent gaze. This highlights that dyads no longer needed to gaze at the same figure as much or as often, after providing their agreed upon referents. The role-swap introduced a disturbance to these trends (Figures 42 and 44), however the fact that dyads globally recovered their common ground and did so quickly reaffirm the existence of historic effects (Brennan & Clark, 1996).

In the EEG hyperscanning results of the present thesis, the linguistic process of entrainment into communicating more effectively by establishing shared common

ground resulted in significant PLV increases across delta, theta, and gamma frequency bands, but with differing topological distributions in inter-brain connectivity. This showcases effectiveness of hyperscanning in a pragmatic experiment setting, consistent with our behavioral and eye tracking correlates.

Firstly, the enhanced inter-brain synchrony observed particularly in frontal, midline, left parietal and left temporal regions in the slow-waves delta and theta are consistent with previous literature (Dehais et al., 2020; Horton & Gerrig, 2005; Pérez et al., 2017; Schilbach et al., 2010; Viswanathan et al., 2019) that constitute in the present thesis as possible correlates of sustained attention and memory processes that support shared linguistic contexts during verbal communication tasks, consistent with the present linguistic findings as well. This is further supported by previous literature that considered communication a joint action (Clark & Brennan, 1991) and with implications that left parietal cortex contains neural networks correlated with the theory of mind (Bzdok et al., 2016; Coslett & Schwartz, 2018) that enables the existence of spatio-temporal self-supervision in social settings (Ninomiya et al., 2018; Sebanz & Knoblich, 2009).

The gamma was the frequency band in which the biggest effects were observed, and observed across the entire cortex. In the existing literature, gamma frequency band has been associated with higher-order cognitive functions, such as attention (Doesburg et al., 2008), memory encoding and retrieval (Tallon-Baudry, 2009), and the integration of sensory information, as well as with conscious integration of perceptions into actions (Engel & Singer, 2001; Tallon-Baudry et al., 1997). The extensive increase in gamma-band PLV across nearly all electrode pairs suggests a global enhancement of neural synchronization during the matching and sorting task and the block based experiment condition of the present thesis. Particularly from the perspective of social aspect of communication, the present findings relate with the existing literature (Dumas et al., 2010); while the existence of a common goal and the necessity of task coordination to increase the chance of success in communication can be related with another motivator, such as the existence of a threat (Mu et al., 2017). Overall, the increasing trend in gamma oscillation power and the resulting inter-brain synchrony might be a correlate of a distributed decision making processes, across the dyad, where the complexity of activity across the cortex (Sohal, 2016) reflects the complexity of pragmatics in a free-form conversation setting with a common goal, such as the one in the present thesis.

Summarizing these results allows us to observe that inter-brain synchrony in the time-frequency domain analysis of PLV in frontal, midline, occipital, temporal, and parietal electrode sites were consistent across the acoustic range, perhaps underpinning joint attention and memory processes, as well as conscious effort of communicating dyads as rate and effectiveness of information transfer increased. An increasing rate of coherence might be facilitated with the alignment of conceptual representations as dyads established a common ground through which to improve their communication.

The nature in which these activations took place also persist through the experimental condition of role-swap after the seventh trials for each dyad. In that, PLV increases were at least sustained through the role-swap, and in certain cases even increased, further demonstrating both the reliability of PLV increase as a correlate of effective

communication as a function of an established and a sustained common ground, as well as the consistency in the hyperscanning analysis setting employed in the present thesis.

Finally, with our fNIRS results, the lack of hyperscanning data due malfunctioning on one of the devices might have contributed into an inconclusive result for the WTC analysis, where only a handful of dyads' data were sufficiently clean of RF interferences for analysis. The lack of sample size as well as the existence of disruptive noise in the data resulted in insignificant effects of the repeated trials in the experiment as dyads established a common ground, otherwise successfully demonstrated in the behavioral, eye tracking, and EEG results. In that, the relative increase in coherence was hypothesized to correlate with the brain-to-brain coherence the interacting dyad may have achieved towards establishing a common ground could not be validated for hemodynamic hyperscanning. In contrast, the working fNIRS device provided us with insights into that there was a significant effect to prefrontal cortex oxygenation by the repeated trials of matching and sorting the twelve Tangram figures, which is consistent with the role of the prefrontal cortex particularly during social interaction (Izzetoglu et al., 2005), as well as communication, two higher-order cognitive functions (Roberts et al., 1998; Valencia & Froese, 2020).

Dorso-lateral prefrontal cortex specifically has been associated with working memory (Balconi, 2013; Levy & Goldman-Rakic, 2000), whereas medial prefrontal regions have been associated with attribution of communicative intent (Rushworth, 2008), as well as linguistic comprehension and attention (Smallwood et al., 2013). The significant effect of trials in this region therefore, support the notion that these physiological effects are results of the repeated trials as dyads established a common ground, a pragmatic process, perhaps emphasizing that there is a social nature to learning (Vygotsky & Michael Cole, 1978).

CHAPTER 6

6. CONCLUSION

The hypotheses that the gradual process of establishing a linguistic common ground (Brennan & Clark, 1996; Clark & Brennan, 1991; Clark & Wilkes-Gibbs, 1986) through the automatic process of interactive alignment (Garrod & Pickering, 2009; Pickering & Garrod, 2004) and striving to expend the least collaborative effort (Clark & Brennan, 1991) to communicate would entail reliably measurable neuroelectric, ocular, and hemodynamic correlates. This hypothesis was validated through an efficient but complex restructuring of the experiment setting utilised in Clark & Wilkes-Gibbs (1986) in a multi-modal hyperscanning investigation. In the present thesis, the effects both seen in preceding literature for all modalities (eye-tracking, fNIRS, and EEG) and newly founded were correlated with significant effects seen in behavioral linguistics. Furthermore, some of these effects demonstrated the existence of non-linear relationships; in that, the establishment of a common ground allow for more efficient communication, but the linearity of neural and ocular correlates significantly interact with the conversational roles as well as the gradual nature of pragmatics.

However, subjective evaluations have explicated possible organizational dynamics that underpin how this gradual process might take place despite that entrainment into effective communication, regardless of intent – as seen within the complex dynamics among different organizational constructs of our dyads – always occurs, and dyads always converged on locally agreed upon, aligned descriptions to make their communication more efficient. Such effects of organizational dynamics were not hypothesized in the present thesis; therefore, the experiment methodology did not factor them in to allow for an objective evaluation of these effects, which necessitate further research to be explored.

Another significant contribution of this thesis is the adaptation of Clark & Wilkes-Gibbs' (1986) experimental framework to record and transcribe 21 sessions of dyadic interactions in Turkish. The resulting transcribed corpora provide a well-structured discourse for each trial and each of the twelve Tangram figures, documenting the gradual process of establishing a common ground as dyads adhered to the principle of least effort, becoming increasingly efficient in their communication.

Finally, in the present thesis the gradual process of establishing a linguistic common ground in a dyadic interaction setting was investigated only in a human-human interaction construct. However, in the present day, humans interact with machines and computers, even AI agents that are all capable of optimizing their abstract concepts in memories to more efficiently communicate with their attending humans. To this end, further research evaluating how an AI agent would fit this complex communication construct must be investigated through, for instance, training on the corpus of dyadic communication within the frame defined by the experimental methodology of the present thesis might contribute significantly into expanding our understandings of how perception of AI generated language might interact with what is natural to humans from the perspectives beyond behavioral data, with neural and ocular correlates of grounding in verbal interaction.

REFERENCES

Ahlström, C., Kircher, K., Nyström, M., & Wolfe, B. (2021). Eye Tracking in Driver Attention Research—How Gaze Data Interpretations Influence What We Learn. *Frontiers in Neuroergonomics*, *2*. https://doi.org/10.3389/fnrgo.2021.778043

Ayaz, H. (2010). Functional Near Infrared Spectroscopy based Brain Computer Interface. *School of Biomedical Engineering Science & Health Systems*, 214. https://doi.org/10.1016/j.applanim.2006.09.023

Ayaz, H., Izzetoglu, M., Bunce, S., Heiman-Patterson, T., & Onaral, B. (2007). Detecting cognitive activity related hemodynamic signal for brain computer interface using functional near infrared spectroscopy. *Proceedings of the 3rd International IEEE EMBS Conference on Neural Engineering*, 342–345. https://doi.org/10.1109/CNE.2007.369680

Balconi, M. (2013). Dorsolateral prefrontal cortex, working memory and episodic memory processes: Insight through transcranial magnetic stimulation techniques. In *Neuroscience Bulletin* (Vol. 29, Issue 3, pp. 381–389). https://doi.org/10.1007/s12264-013-1309-z

Baldwin, D. A. (1995). Understanding the Link Between Joint Attention and Language. In C. Moore & P. J. Dunham (Eds.), *Joint Attention: Its Origins and Role in Development* (pp. 131–158). Lawrence Erlbaum Associates, Inc. https://doi.org/10.2307/302397

Bard, E. G., Hill, R., Arai, M., & Foster, M. E. (2009). Accessibility and attention in situated dialogue: Roles and regulations. *Proceedings of the Workshop on Production of Referring Expressions Pre-CogSci*.

Barr, D. J., & Keysar, B. (2002). Anchoring Comprehension in Linguistic Precedents. *Journal of Memory and Language*, *46*(2), 391–418. https://doi.org/10.1006/jmla.2001.2815

Barz, M., & Sonntag, D. (2021). *Automatic Visual Attention Detection for Mobile Eye Tracking Using Pre-Trained Computer Vision Models and Human Gaze*. https://doi.org/10.3390/s21124143

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal*

Statistical Society Series B: Statistical Methodology, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Benjamini, Y., & Yekutieli, D. (2001). The Control of the False Discovery Rate in Multiple Testing under Dependency. In *Source: The Annals of Statistics* (Vol. 29, Issue 4).

Berka, C., Levendowski, D. J., Lumicao, M. N., Alan Yau, G. D., Zivkovic, V. T., Olmstead, R. E., Tremoulet, P. D., & Craven, P. L. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, Space, and Environmental Medicine*, *78*(5), B231–B244.

Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(6), 1482–1493. https://doi.org/10.1037/0278-7393.22.6.1482

Brouwer, A.-M., Hogervorst, M. A. J., Oudejans, B., Ries, A. J., & Touryan, J. (2018). Electroencephalography and Eye Tracking Signatures of Target Encoding During Guided Search. In *Neuroergonomics* (pp. 307–308). Elsevier. https://doi.org/10.1016/B978-0-12-811926-6.00088-9

Burgess, A. P. (2013). On the interpretation of synchronization in EEG hyperscanning studies: A cautionary note. *Frontiers in Human Neuroscience*, *7*(DEC), 1–17. https://doi.org/10.3389/fnhum.2013.00881

Bzdok, D., Hartwigsen, G., Reid, A., Laird, A. R., Fox, P. T., & Eickhoff, S. B. (2016). Left inferior parietal lobe engagement in social cognition and language. In *Neuroscience and Biobehavioral Reviews* (Vol. 68, pp. 319–334). Elsevier Ltd. https://doi.org/10.1016/j.neubiorev.2016.02.024

Ceh, S. M., Annerer-Walcher, S., Körner, C., Rominger, C., Kober, S. E., Fink, A., & Benedek, M. (2020). Neurophysiological indicators of internal attention: An electroencephalography–eye-tracking coregistration study. *Brain and Behavior*, *10*(10). https://doi.org/10.1002/brb3.1790

Chance, B., Anday, E., Nioka, S., Zhou, S., Hong, L., Worden, K., Li, C., Murray, T., Ovetsky, Y., Pidikiti, D., & Thomas, R. (1998). A novel method for fast imaging of brain function, non-invasively, with light. *Optics Express*, *2*(10), 411. https://doi.org/10.1364/OE.2.000411

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In *Perspectives on socially shared cognition.* (pp. 127–149). American Psychological Association. https://doi.org/10.1037/10096-006

Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, *50*(1), 62–81. https://doi.org/10.1016/j.jml.2003.08.004

Clark, H. H., & Wilkes-gibbs, D. (1986). Referring as a collaborative process. *Cognition*, *22*(1), 1–39. https://doi.org/10.1016/0010-0277(86)90010-7

Cleland, A. A., & Pickering, M. J. (2003). The use of lexical and syntactic information in language production: Evidence from the priming of noun-phrase structure. *Journal of Memory and Language*, *49*(2), 214–230. https://doi.org/10.1016/S0749-596X(03)00060-3

Coco, M. I., & Dale, R. (2014). Cross-recurrence quantification analysis of categorical and continuous time series: An R package. *Frontiers in Psychology*, *5*(JUN). https://doi.org/10.3389/fpsyg.2014.00510

Cooke, N. J., Gorman, J. C., Myers, C. W., & Duran, J. L. (2012). Interactive Team Cognition. *Cognitive Science*, *37*(2), 255–285. https://doi.org/https://doi.org/10.1111/cogs.12009

Cope, M. (1991). The application of near infrared spectroscopy to non invasive monitoring of cerebral oxygenation in the newborn infant. *Department of Medical Physics and Bioengineering*, *342*.

Coslett, H. B., & Schwartz, M. F. (2018). The parietal lobe and language. In *Handbook of Clinical Neurology* (Vol. 151, pp. 365–375). Elsevier B.V. https://doi.org/10.1016/B978-0-444-63622-5.00018-8

Costanzo, M. E., McArdle, J. J., Swett, B., Nechaev, V., Kemeny, S., Xu, J., & Braun, A. R. (2013). Spatial and temporal features of superordinate semantic processing studied with fMRI and EEG. *Frontiers in Human Neuroscience*, *JUN*. https://doi.org/10.3389/fnhum.2013.00293

Cui, X., Bryant, D. M., & Reiss, A. L. (2012). NIRS-based hyperscanning reveals increased interpersonal coherence in superior frontal cortex during cooperation. *NeuroImage*, *59*(3), 2430–2437. https://doi.org/10.1016/j.neuroimage.2011.09.003

D'Angelo, S., & Gergle, D. (2016). Gazed and confused: Understanding and designing shared gaze for remote collaboration. *Conference on Human Factors in Computing Systems - Proceedings*, 2492–2496. https://doi.org/10.1145/2858036.2858499

De, K. (2012). *Calculating inter-coder reliability in media content analysis using Krippendorff's Alpha*. http://repository.upenn.edu/asc_papers/43

Dehais, F., Lafont, A., Roy, R., & Fairclough, S. (2020). A Neuroergonomics Approach to Mental Workload, Engagement and Human Performance. *Frontiers in Neuroscience*, *14*(April), 1–17. https://doi.org/10.3389/fnins.2020.00268

Dennett, D. C. (1981). True Believers: The Intentional Strategy and Why It Works. In A. F. Heath (Ed.), *Scientific Explanation: Papers Based on Herbert Spencer Lectures Given in the University of Oxford* (pp. 150–167).

Doesburg, S. M., Roggeveen, A. B., Kitajo, K., & Ward, L. M. (2008). Large-scale gamma-band phase synchronization and selective attention. *Cerebral Cortex*, *18*(2), 386–396. https://doi.org/10.1093/cercor/bhm073

Dumas, G. (2011). Towards a two-body neuroscience. *Communicative & Integrative Biology*, *4*(3), 349–352. https://doi.org/10.4161/cib.4.3.15110

Dumas, G., Nadel, J., Soussignan, R., Martinerie, J., & Garnero, L. (2010). Inter-Brain Synchronization during Social Interaction. *PLoS ONE*, *5*(8), e12166. https://doi.org/10.1371/journal.pone.0012166

Engel, A. K., & Singer, W. (2001). Temporal binding and the neural correlates of sensory awareness. *Trends in Cognitive Sciences*, *5*(1), 16–25. https://doi.org/10.1016/S1364-6613(00)01568-0

Fındık-Coşkunçay, D., & Çakır, M. P. (2022). An investigation of the relationship between joint visual attention and product quality in collaborative business process modeling: a dual eye-tracking study. *Software and Systems Modeling*, *21*(6), 2429–2460. https://doi.org/10.1007/s10270-022-00974-6

Frith, C. D., & Frith, U. (1999). Interacting Minds--A Biological Basis. *Science*, *286*(5445), 1692–1695. https://doi.org/10.1126/science.286.5445.1692

Funane, T., Kiguchi, M., Atsumori, H., Sato, H., Kubota, K., & Koizumi, H. (2011). Synchronous activity of two people's prefrontal cortices during a cooperative task measured by simultaneous near-infrared spectroscopy. *Journal of Biomedical Optics*, *16*(7), 077011. https://doi.org/10.1117/1.3602853

Gao, Q., Wang, J., Yu, C., & Chen, H. (2015). Effect of handedness on brain activity patterns and effective connectivity network during the semantic task of Chinese characters. *Nature Publishing Group*, *June*, 1–11. https://doi.org/10.1038/srep18262

Garrod, S., & Pickering, M. J. (2009). Joint Action, Interactive Alignment, and Dialog. *Topics in Cognitive Science*, *1*(2), 292–304. https://doi.org/10.1111/j.1756-8765.2009.01020.x

Gibson, J. J. (1979). The Theory of Affordances. In *The Ecological Approach to Visual Perception* (pp. 119–135). Houghton Mifflin.

Gilchrist, I. D., & Harvey, M. (2000). Refixation frequency and memory mechanisms in visual search. *Current Biology*, *10*(19), 1209–1212. https://doi.org/10.1016/S0960-9822(00)00729-6

Grice, H. P. (1957). Meaning. *The Philosophical Review*, *66*(3), 377. https://doi.org/10.2307/2182440

Grice, H. P. (1975). Logic and conversation. In *Syntax and semantics 3: Speech arts* (pp. 41–58). Brill.

Grice, H. P. (1978). Further Notes on Logic and conversation. *Syntax and Semantics*, *9*, 113–127.

Grinsted, A., Moore, J. C., & Jevrejeva, S. (2004). Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Processes in Geophysics*, *11*(5/6), 561–566. https://doi.org/10.5194/npg-11-561-2004

Hanna, J. E., & Brennan, S. E. (2007). Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language*, *57*(4), 596–615. https://doi.org/10.1016/j.jml.2007.01.008

Hartwright, C. E., Apperly, I. A., & Hansen, P. C. (2014). Representation, Control, or Reasoning? Distinct Functions for Theory of Mind within the Medial Prefrontal Cortex. *Journal of Cognitive Neuroscience*, *26*(4), 683–698. https://doi.org/10.1162/jocn_a_00520

Holmqvist, K., Nystrom, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van De Weijer, J. (2011). Eye Tracking: A Comprehensive Guide to Methods and Measures. In *OUP, Oxford*. Oxford University Press.

Holper, L., Scholkmann, F., & Wolf, M. (2012). Between-brain connectivity during imitation measured by fNIRS. *NeuroImage*, *63*(1), 212–222. https://doi.org/10.1016/j.neuroimage.2012.06.028

Hopkins, W. D., & Taglialatela, J. P. (2011). Some Preliminary Observations on The Neural Correlates of Joint Attention in Chimpanzees. In A. Seeman (Ed.), *Joint Attention: New Developments in Psychology* (pp. 242–263). MIT Press.

Horton, W. S., & Gerrig, R. J. (2005). Conversational Common Ground and Memory Processes in Language Production. *Discourse Processes*, *40*(1), 1–35. https://doi.org/10.1207/s15326950dp4001_1

Hutchins, E., & Tove, K. (1999). Distributed cognition in an airline cockpit. In *Cognition and communication at work* (pp. 27–46).

Izzetoglu, M., Nioka, S., Chance, B., & Onaral, B. (2005). Single Trial Hemodynamic Response Estimation In A Block Anagram Solution Study Using fNIR Spectroscopy. *Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, *5*, 633–636. https://doi.org/10.1109/ICASSP.2005.1416383

Knoblich, G., & Flach, R. (2003). Action identity: Evidence from self-recognition, prediction, and coordination. *Consciousness and Cognition*, *12*(4), 620–632. https://doi.org/10.1016/S1053-8100(03)00070-9

Krause, L., Enticott, P. G., Zangen, A., & Fitzgerald, P. B. (2012). The role of medial prefrontal cortex in theory of mind: A deep rTMS study. *Behavioural Brain Research*, *228*(1), 87–90. https://doi.org/10.1016/j.bbr.2011.11.037

Krippendorff, K. (2011). *Reliability Data Reliability Data Observed Difference a-Agreement*.

Lab Streaming Layer (LSL). (n.d.). Swartz Center for Computational Neuroscience. https://github.com/sccn/labstreaminglayer

Lachaux, J. P., Rodriguez, E., Martinerie, J., & Varela, F. J. (1999). Measuring phase synchrony in brain signals. *Human Brain Mapping*, *8*(4), 194–208. https://doi.org/10.1002/(SICI)1097-0193(1999)8:4<194::AID-HBM4>3.0.CO;2-C

Lage-Castellanos, A., Martínez-Montes, E., Hernández-Cabrera, J. A., & Galán, L. (2010). False discovery rate and permutation test: An evaluation in ERP data analysis. *Statistics in Medicine*, *29*(1), 63–74. https://doi.org/10.1002/sim.3784

Levy, R., & Goldman-Rakic, P. S. (2000). Segregation of working memory functions within the dorsolateral prefrontal cortex. In *Experimental Brain Research* (Vol. 133, Issue 1, pp. 23–32). Springer Verlag. https://doi.org/10.1007/s002210000397

Liu, Y., Ayaz, H., & Shewokis, P. A. (2017). Multisubject "Learning" for Mental Workload Classification Using Concurrent EEG, fNIRS, and Physiological Measures. *Frontiers in Human Neuroscience*, *11*(July). https://doi.org/10.3389/fnhum.2017.00389

Macinnes, J. J., Iqbal, S., Pearson, J., & Johnson, E. N. (2018). Wearable Eye-tracking for Research: Automated dynamic gaze mapping and accuracy/precision comparisons across devices. *BioRxiv*. https://doi.org/10.1101/299925

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *200*(1140), 269–294. https://doi.org/10.1098/rspb.1978.0020

Marriott Haresign, I., Phillips, E. A. M., Whitehorn, M., Goupil, L., Noreika, V., Leong, V., & Wass, S. V. (2022). Measuring the temporal dynamics of inter-personal neural entrainment in continuous child-adult EEG hyperscanning data. *Developmental Cognitive Neuroscience*, *54*(February), 101093. https://doi.org/10.1016/j.dcn.2022.101093

Metzing, C., & Brennan, S. E. (2003). When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language*, *49*(2), 201–213. https://doi.org/10.1016/S0749-596X(03)00028-7

Montague, P. R., Berns S., G., Cohen, J., McClure M., S., Pagnoni, G., Dhamala, M., Wiest C., M., Karpov, I., King D., R., Apple, N., & Fisher E., R. (2002). Hyperscanning: Simultaneous fMRI during Linked Social Interactions. *NeuroImage*, *16*(4), 1159–1164. https://doi.org/10.1006/nimg.2002.1150

Morimoto, C. H., Koons, D., Amir, A., & Flickner, M. (2000). Pupil detection and tracking using multiple light sources. *Image and Vision Computing*, *18*(4), 331–335. https://doi.org/10.1016/S0262-8856(99)00053-0

Mu, Y., Han, S., & Gelfand, M. J. (2017). The role of gamma interbrain synchrony in social coordination when humans face territorial threats. *Social Cognitive and Affective Neuroscience*, *12*(10), 1614–1623. https://doi.org/10.1093/scan/nsx093

Neuroelectrics. (n.d.). *NIC2.0*. https://www.neuroelectrics.com/wiki/index.php/NIC2.0

Ninomiya, T., Noritake, A., Ullsperger, M., & Isoda, M. (2018). Performance monitoring in the medial frontal cortex and related neural networks: From monitoring self actions to understanding others' actions. *Neuroscience Research*, *137*, 1–10. https://doi.org/10.1016/j.neures.2018.04.004

Nozawa, T., Sasaki, Y., Sakaki, K., Yokoyama, R., & Kawashima, R. (2016). Interpersonal frontopolar neural synchronization in group communication: An exploration toward fNIRS hyperscanning of natural interactions. *NeuroImage*, *133*, 484–497. https://doi.org/10.1016/j.neuroimage.2016.03.059

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia*. https://doi.org/10.1016/0028-3932(71)90067-4

Omurtag, A., Roy, R. N., Dehais, F., Chatty, L., & Garbey, M. (2019). Tracking Mental Workload by Multimodal Measurements in the Operating Room. In *Neuroergonomics* (pp. 99–103). Elsevier. https://doi.org/10.1016/B978-0-12-811926-6.00016-6

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, *2011*, 1–9. https://doi.org/10.1155/2011/156869

Pérez, A., Carreiras, M., & Duñabeitia, J. A. (2017). Brain-To-brain entrainment: EEG interbrain synchronization while speaking and listening. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-04464-4

Pfeiffer, U. J., Vogeley, K., & Schilbach, L. (2013). Neuroscience and Biobehavioral Reviews From gaze cueing to dual eye-tracking : Novel approaches to investigate the neural correlates of gaze in social interaction. *Neuroscience and Biobehavioral Reviews*, *37*(10), 2516–2528. https://doi.org/10.1016/j.neubiorev.2013.07.017

Pickering, M. J., & Garrod, S. (2004). The interactive-alignment model: Developments and refinements. *Behavioral and Brain Sciences*, *27*(02), 212–225. https://doi.org/10.1017/S0140525X04450055

Pupil Labs. (n.d.-a). *Pupil Labs Capture, Pupil Detection*. https://docs.pupil-labs.com/core/software/pupil-capture/#pupil-detection

Pupil Labs. (n.d.-b). *Pupil Labs Core, Capture, and Player*. https://docs.pupil-labs.com/core/

Redcay, E., & Schilbach, L. (2019). Using second-person neuroscience to elucidate the mechanisms of social interaction. *Nature Reviews Neuroscience*, *20*(8), 495–505. https://doi.org/10.1038/s41583-019-0179-4

Reddy, V. (2011). A Gaze at Grips With Me. In A. Seeman (Ed.), *Joint Attention: New Developments in Psychology* (pp. 137–157). MIT Press.

Richardson, D. C., & Dale, R. (2005). Looking To Understand: The Coupling Between Speakers' and Listeners' Eye Movements and Its Relationship to Discourse Comprehension. *Cognitive Science*, *29*(6), 1045–1060. https://doi.org/10.1207/s15516709cog0000_29

Roberts, A. C., Robbins, T. W., & Weiskrantz, L. (1998). *The Prefrontal CortexExecutive and Cognitive Functions* (A. C. Roberts, T. W. Robbins, & L. Weiskrantz, Eds.). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198524410.001.0001

Rushworth, M. F. S. (2008). Intention, choice, and the medial frontal cortex. In *Annals of the New York Academy of Sciences* (Vol. 1124, pp. 181–207). Blackwell Publishing Inc. https://doi.org/10.1196/annals.1440.014

Sainburg, R. L. (2014). Convergent models of handedness and brain lateralization. *Frontiers in Psychology*, *5*(SEP), 1–14. https://doi.org/10.3389/fpsyg.2014.01092

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral and Brain Sciences*, *36*(4), 393–414. https://doi.org/10.1017/S0140525X12000660

Schilbach, L., Wilms, M., Eickhoff, S. B., Romanzetti, S., Tepest, R., Bente, G., Shah, N. J., Fink, G. R., & Vogeley, K. (2010). Minds Made for Sharing: Initiating Joint Attention Recruits Reward-related Neurocircuitry. *Journal of Cognitive Neuroscience*, *22*(12), 2702–2715. https://doi.org/10.1162/jocn.2009.21401

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends in Cognitive Sciences*, *10*(2), 70–76. https://doi.org/10.1016/j.tics.2005.12.009

Sebanz, N., & Knoblich, G. (2009). Prediction in Joint Action: What, When, and Where. *Topics in Cognitive Science*, *1*(2), 353–367. https://doi.org/10.1111/j.1756-8765.2009.01024.x

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shockley, K., Richardson, D. C., & Dale, R. (2009). Conversation and Coordinative Structures. *Topics in Cognitive Science*, *1*(2), 305–319. https://doi.org/10.1111/j.1756-8765.2009.01021.x

Singh, A. K., & Dan, I. (2006). Exploring the false discovery rate in multichannel NIRS. *NeuroImage*, *33*(2), 542–549. https://doi.org/10.1016/j.neuroimage.2006.06.047

Smallwood, J., Gorgolewski, K. J., Golchert, J., Ruby, F. J. M., Engen, H., Baird, B., Vinski, M. T., Schooler, J. W., & Margulies, D. S. (2013). The default modes of reading: Modulation of posterior cingulate and medial prefrontal cortex connectivity associated with comprehension and task focus while reading. *Frontiers in Human Neuroscience*, *NOV*. https://doi.org/10.3389/fnhum.2013.00734

Sohal, V. S. (2016). How close are we to understanding what (If anything) γ oscillations do in cortical circuits? *Journal of Neuroscience*, *36*(41), 10489–10495. https://doi.org/10.1523/JNEUROSCI.0990-16.2016

Stephens, G. J., Silbert, L. J., & Hasson, U. (2010). Speaker-listener neural coupling underlies successful communication. *Proceedings of the National Academy of Sciences*, *107*(32), 14425–14430. https://doi.org/10.1073/pnas.1008662107

Tallon-Baudry, C. (2009). The roles of gamma-band oscillatory synchrony in human visual cognition. In *Frontiers in Bioscience* (Vol. 14).

Tallon-Baudry, C., Bertrand, O., Delpuech, C., & Pernier, J. (1997). *Oscillatory-Band (30-70 Hz) Activity Induced by a Visual Search Task in Humans*.

Valencia, A. L., & Froese, T. (2020). What binds us? Inter-brain neural synchronization and its implications for theories of human consciousness. *Neuroscience of Consciousness*, *2020*(1). https://doi.org/10.1093/NC/NIAA010

Viola, P., & Jones, M. (2004). Rapid object detection using a boosted cascade of simple features. *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, *1*, I-511-I–518. https://doi.org/10.1109/CVPR.2001.990517

Viswanathan, V., Bharadwaj, H. M., & Shinn-Cunningham, B. G. (2019). Electroencephalographic signatures of the neural representation of speech during selective attention. *ENeuro*, *6*(5). https://doi.org/10.1523/ENEURO.0057-19.2019

Vygotsky, L. S., & Michael Cole. (1978). *Mind in society: Development of higher psychological processes*. Harvard university press.

Wilms, M., Schilbach, L., Pfeiffer, U., Bente, G., Fink, G. R., & Vogeley, K. (2010). It's in your eyes—using gaze-contingent stimuli to create truly interactive paradigms for social cognitive and affective neuroscience. *Social Cognitive and Affective Neuroscience*, *5*(1), 98–107. https://doi.org/10.1093/scan/nsq024