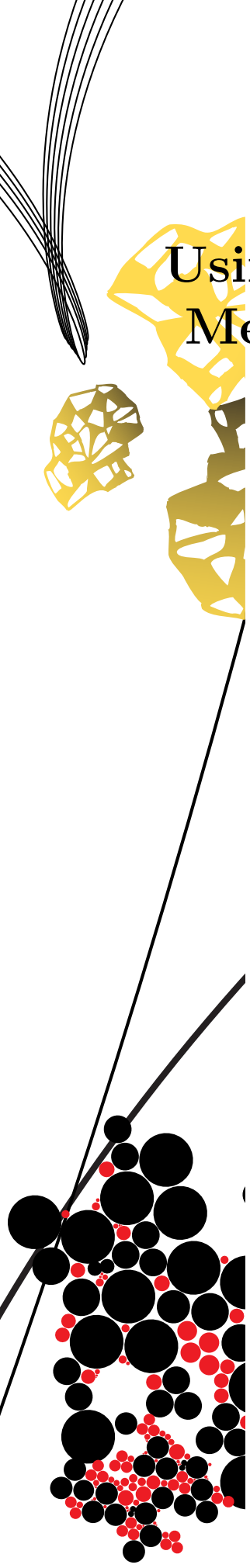


Using Neurophysiological Signals to Measure Social Exclusion Induced by a Language Barrier

Kim Gouweleeuw (s2204789)

Master Thesis
Credits: 30 ECTS

Supervisors
dr. Mannes Poel
dr. Nattapong Thammasan
prof. dr. Dirk K.J. Heylen
dr. Anne-Marie Brouwer

January 21, 2021

UNIVERSITY OF TWENTE.

Abstract

Universities and other workplaces are becoming more and more international. This calls for good communication and teamwork between international teams. However, people often fall back on their mother tongue when they cannot express themselves well enough in another language, thus speaking a language that most people around them cannot understand. This linguistic ostracism could cause feelings of social exclusion, anger and sadness, and even trust issues, instead of the social inclusion which should be aimed for when working together as a team. This paper will endeavour to determine how, and to what extent, this social exclusion induced by a language barrier reflects in neurophysiological signals (Electroencephalogram (EEG), Heart Rate (HR), Galvanic Skin Response (GSR)). To this end, an experiment has been conducted in which three participants worked together as a team to solve seven small riddles. During this experiment, two participants communicated with each other in a language the third participant did not understand, thereby ignoring the third participant and causing feelings of social exclusion. Based on the existing literature, it was expected that the two participants that could understand each other and worked together as a team would show a higher level of synchronisation in the measured neurophysiological signals (EEG, HR, GSR) compared to the socially excluded participant with either of the other two participants. The level of synchrony for the EEG modality was computed using the Phase-Locking Value (PLV) method and the synchrony for the HR and GSR modality was computed with the Pearson Correlation Coefficient (PCC). The results showed that, out of the three measured modalities, the EEG modality was best suited for measuring this synchrony and social exclusion. The data was compared both per brain region (frontal, central, parietal, temporal, occipital) and per channel (32 channels). For the regional comparison, the theta and alpha frequency bands had the strongest result, while the gamma band achieved the strongest results for the channel comparison. The statistical analysis indicates that the central brain region (and channel C4 in particular) looks the most promising, with statistically significant results before False Discovery Rate (FDR) correction. Both the HR and GSR analyses indicate that there is a difference between the socially excluded participant and the other two participants although there is a large variability between the results. However, these differences are only significant for a few individual experiments. Therefore, even though the EEG, HR, and GSR results all indicate that there is indeed a difference in the synchrony between the socially excluded participant and the other two participants, and that social exclusion can thus, to some extent, be measured using these three modalities, further investigation is needed to draw definitive conclusions.

Contents

Acronyms

ACC Anterior Cingulated Cortex. 24

ANOVA Analysis of Variance. 26, 49, 81

ASR Artifact Subspace Reconstruction. 40, 43

BOLD Blood-Oxygenation-Level-Dependent. 12, 15, 20

CCorr Circular Correlation. 24

CMS Common Mode Sense. 39

DFA Discriminant Function Analysis. 26

DRL Driven Right Leg. 39

ECG Electrocardiogram. 3, 26, 33, 39, 43–45, 48

EEG Electroencephalogram. 2–4, 8–17, 19, 21–26, 29, 32, 33, 35, 36, 39, 40, 45, 46, 48, 50, 60, 65, 76–79, 81, 84–86

ERP Event-Related Potential. 11

FDR False Discovery Rate. 2, 49, 55, 61, 63, 65, 77, 78, 85, 86

FFT Fast Fourier Transform. 22

fMRI Functional Magnetic Resonance Imaging. 3, 11, 12, 14, 15, 20, 23

fNIRS Functional Near-Infrared Spectroscopy. 3, 11, 12, 14, 15, 19, 23, 84

GSR Galvanic Skin Response. 2–4, 8–10, 16, 26, 27, 32, 33, 35, 39, 44, 45, 47–50, 71–76, 80, 81, 83–86

HR Heart Rate. 2–4, 8–10, 16, 26, 27, 32, 33, 35, 43, 47, 49, 50, 66, 76, 79–81, 83–86

HRV Heart Rate Variability. 80, 84, 86

IBI Inter-Beat Interval. 4, 47, 48, 66–71, 80, 84, 86

ICA Independent Component Analysis. 20, 40–42

IFG Inferior Frontal Gyrus. 17

IPC Interbrain Phase Coherence. 24

IPL Inferior Parietal Lobule. 17

LSL Lab Streaming Layer. 39, 45, 48

MANOVA Multivariate Analysis of Variance. 26

MEG Magnetoencephalogram. 3, 11, 12, 14, 15, 23

MNS Mirror Neuron System. 17, 18, 23, 78

MS Mentalizing System. 17, 18, 23, 78

MVAR Multivariate Autoregressive Model. 21

PCC Pearson Correlation Coefficient. 2, 21, 32, 45, 47–49, 66, 80, 85

PDC Partial Directed Coherence. 21, 22, 24, 84, 86

PFC Prefrontal Cortex. 17, 18, 23

PLI Phase Locking Index. 24

PLI Phase Lag Index. 19, 20, 78

PLV Phase-Locking Value. 2, 3, 19, 20, 22, 23, 32, 33, 45–47, 49–62, 64, 65, 76–79, 83–86, 88–90, 96, 102, 104, 105, 115, 116

RTPJ Right Temporal-Parietal Junction. 15

SCR Skin Conductance Response. 84, 86

STG Superior Temporal Gyrus. 17

TPJ Temporal-Parietal Junction. 17, 18, 23

WTC Wavelet Transform Coherence. 19, 20

Acknowledgements

I would like to thank my supervisor, Mannes Poel, for his constant support and guidance in addition to all his useful suggestions and feedback during this whole process.

Furthermore, my immense gratitude also goes to Nattapong Thammasan for assisting me during my experiments as well as helping me and providing valuable comments whenever I was stuck with my code or some of the pre-processing steps.

I am also grateful to Michele Romani for assisting me during my experiments; not only for preparing the participants but also for helping me find more participants for my experiments.

Of course, I would also like to express my thanks to all participants who participated in my study. Thanks to your voluntary participation, you made it possible for me to conduct this research.

Additionally, my gratitude also goes to the BMS lab at the University of Twente for allowing me to conduct my experiments in their lab facilities.

Finally, I would like to thank Anne-Marie Brouwer for letting me borrow equipment from TNO. Without that equipment, the research would not have been possible.

1 Introduction

The world is becoming more and more international because it is a lot easier to travel in this almost borderless world. This encourages many students to decide to study abroad. Students want to learn more about a new country, escape their normal day-to-day lives, explore new career opportunities, and doing that all while earning academic credits (Nyaupane et al., 2010). This is also a trend which can be seen in the Netherlands. The University of Twente even has an international student handbook to help international students settle in their new temporary home (UTwente, 2018).

Since universities are so internationally oriented nowadays, it becomes very important that communication occurs in a language everyone can understand. Luckily almost everyone in the Netherlands can speak English. However, language differences can still form a problem since people often switch back to their mother tongue when they, for example, want to streamline communication or want to better express their emotions (Tenzer et al., 2014). Nearly everyone has, at some point in their lives, experienced that people around them talk to each other in a language they cannot understand (Williams, 2007).

How would it make students feel if they could not understand a word the people around them were saying? This could make them feel socially excluded, ignored and less trusting of those they cannot understand, instead of the social inclusion which should be aimed for. Especially if students are supposed to work together as a team. With the increase in the number of languages spoken at universities and other workplaces, there is a heightened chance of this form of social exclusion, otherwise known as linguistic ostracism (Oshri et al., 2008; Fiset and Bhave, 2019). This will cause increased feelings of sadness and anger as well as a feeling of desynchronisation from their peers, making these students less likely to engage in helping behaviours (Thau et al., 2007; Williams, 2007).

Even though social exclusion is, thus, arising more and more often in universities, it has only been researched in workplace settings (Oshri et al., 2008; Fiset and Bhave, 2019; Thau et al., 2007; Williams, 2007). Furthermore, the emotional impact of excluding students in social interactions such as this has not been investigated using neurophysiological signals (Electroencephalogram (EEG), Heart Rate (HR), and Galvanic Skin Response (GSR)) before. The fact that social exclusion, thus, has not been researched with regards to EEG, HR, and GSR, makes it very interesting to investigate this psychological construct to determine how social exclusion is reflected in these neurophysiological signals.

Previous research has shown that neural synchronisation in EEG is a good measurement for social interactions such as the level of team coordination between individuals. This neural synchronisation increases when communication between people is smooth and they work together as a team (Cha and Lee, 2019). Consequently, when someone is ignored and excluded during this teamwork (e.g. because they cannot understand the language that is spoken) they should have a lower neural synchronisation with the people excluding them compared to the people who are actually working together and understanding each other.

People who are constantly being ignored by their peers will thus likely experience negative feelings and (neural) desynchronisation from their supposed peers (Thau et al., 2007; Williams, 2007; Cha and Lee, 2019). Since feelings and emotions can be reflected in the HR and GSR of a person and the neural synchronisation is reflected in the brain, investigating

EEG, HR and GSR while someone is linguistically ostracised will give an understanding about how and to what extent social exclusion is reflected in these neurophysiological signals (Ferdinando et al., 2014). Results could, therefore, show that there is an actual neurophysiological reaction to being socially excluded.

The research above yields the following research question for this thesis:

How does social inclusion/exclusion induced by a language barrier reflect in neurophysiological signals?

This research question can be answered by first answering the following sub-questions:

1. To what extent can a language barrier cause feelings of social exclusion?
2. To what extent is social inclusion/exclusion reflected in the synchrony between individuals?
3. To what extent can social inclusion/exclusion be measured using EEG, HR, and GSR?

So far, most EEG and other brain-imaging studies record the brain activity for one single participant at a time. However, to study brain activity in participants during social interactions such as this, it becomes necessary to simultaneously record the brain activity of all interacting participants. This simultaneous brain-imaging of two or more participants is called hyperscanning (Xie et al., 2019). Most previous hyperscanning studies have been done with no more than two participants (Xie et al., 2019; Burgess, 2013). However, to linguistically ostracise someone, two or more people are required to interact in a language unfamiliar to the socially excluded person (Dotan-Eliaz et al., 2009). This means that, to make it possible to study social exclusion, a many-to-one setting of at least three participants is needed. Given the complexity of the recording and the synchronisation of the three modalities (EEG, HR, GSR), this hyperscanning research will be conducted with the minimum requirement of three participants, two of which will socially exclude the third participant.

It is hypothesised that the two participants speaking the same language will have a high brain-to-brain synchrony, while the socially excluded participant will have a lower level of synchronisation with the other two participants. Furthermore, it is hypothesised that heart rate and galvanic skin response will also reflect this synchrony.

Several contributions to the research field are made in this thesis. First of all, while this form of linguistic ostracism has previously only been researched in workplace settings, this study will investigate this form of social exclusion among students at a university (Oshri et al., 2008; Fiset and Bhave, 2019; Thau et al., 2007; Williams, 2007). Furthermore, this study will pioneer in using three neurophysiological signals (EEG, HR, and GSR) to measure the psychological construct of feeling socially excluded by a language barrier. Conversely, earlier research has only ever used surveys to measure social exclusion. Lastly, previous research has mostly only conducted hyperscanning studies using two participants while this research will be one of the first to simultaneously measure the data from three participants during an experiment (Xie et al., 2019; Burgess, 2013).

This thesis describes the work that has been done to answer the research question posed in this chapter. First, several studies related to hyperscanning, social exclusion, and team coordination are reviewed. This literature review is divided into two parts: background

(Section 2) and related work (Section 3). The pros and cons of several brain-imaging techniques are discussed in Section 2.1.1. Furthermore, the differences and similarities between hyperscanning and synchrony are explained in Section 2.1.2 while Section 2.1.3 will go into more detail about hyperscanning itself. In addition, neurophysiological signals such as Heart Rate (HR) and Galvanic Skin Response (GSR) are examined in Section 2.2 for their potential added value in measuring synchrony. Moreover, Section 2.3 will explain multiple ways to analyse hyperscanning data. Next, several hyperscanning studies using EEG will be discussed in further detail in Section 3.1. Additionally, research about combining EEG, HR, and GSR will be investigated in Section 3.2. Section 3.3 will then correlate social inclusion/exclusion and language barriers to each other. Furthermore, Section 3.4 will explain team coordination itself, as well as different paradigms for measuring team coordination. The methodology for executing the research will be discussed in Section 4 while Section 5 will show the results. Lastly, Section 6 will discuss the results as well as the limitations and strengths of this research in addition to ideas for future work, and Section 7 will conclude the findings of the research.

2 Background

2.1 Brain-Imaging Techniques and Hyperscanning

In most brain-imaging studies, the activity of the brain during either a simple or complex task has been recorded for one single participant at a time. However, to study brain activity in participants who are interacting in cooperation or other social activities, it becomes necessary to simultaneously record the brain activity of all participants interacting in these activities. Thus, hyperscanning becomes necessary for an experiment in which someone is socially excluded from the activity.

This chapter will first briefly introduce EEG and other non-invasive brain-imaging techniques, which is partly based on an overview by Czeszumski et al. (2020). Then, the differences and similarities of hyperscanning and synchrony will be discussed. Afterwards, hyperscanning will be explained in further detail by discussing several hyperscanning studies.

2.1.1 EEG, MEG, fMRI & fNIRS: Pros and Cons

Electroencephalogram (EEG)
EEG is one of the oldest and perhaps one of the most widely used brain-imaging techniques (Czeszumski et al., 2020; Burgess, 2013). It is an electrophysiological measurement technique which measures neural activity directly with the use of electrodes which are placed on the scalp. These electrodes detect the variations of electrical potentials. An EEG response which is aligned with a certain stimulus is called an Event-Related Potential (ERP). The temporal resolution of these ERPs is higher than other methods (Michel and Brunet, 2019). However, it is difficult to determine the exact location of the neural activity which is causing the electric potentials because the electrodes are placed on the scalp (Anderson, 2009; Anwar et al., 2016). This characteristic makes EEG best suited for investigating the cerebral cortex rather than deep brain structures (Czeszumski et al., 2020). While restricted movement used to be an issue for EEG, the development of new technologies has greatly improved the mobility of EEG systems (Melnik et al., 2017). This mobility, in addition to its great temporal resolution, makes EEG a great tool for studying social interactions. After all, social interactions unfold at a fast scale and, thus, require a method that is sensitive to it, allowing for a more precise type of brain-to-brain analysis (Czeszumski et al., 2020; Dikker et al., 2017). Moreover, using EEG makes it easier to measure more than two individuals at a time, which is very useful in hyperscanning studies (Dikker et al., 2017). Lastly, perhaps the most key advantages of using EEG are the availability of the (mobile) equipment, its relatively low price, and the fact that it can be used in naturalistic experiment settings (Burgess, 2013; Czeszumski et al., 2020).

Magnetoencephalogram (MEG)
Similar to EEG, MEG is an electrophysiological measurement, but MEG measures magnetic fields produced by the electrical activity instead of electrical potentials. This variation of ERP offers a better spatial resolution than EEG. MEG is best at detecting activity in the sulci of the cortex and is not very well suited for measuring activity in the gyri or other places deep within the brain (Anderson, 2009; Anwar et al., 2016). Despite its similar characteristics to EEG, MEG has much lower mobility which makes it less suited to be used in hyperscanning studies which aim to research social interactions (Czeszumski et al., 2020).

Functional Magnetic Resonance Imaging (fMRI)
fMRI is a hemodynamic measurement technique and measures brain activity indirectly through changes associated with blood flow. To achieve this, the Blood-Oxygenation-Level-Dependent (BOLD) contrast is used which depicts changes in deoxyhemoglobin concentration (Glover, 2011). Perhaps its most important advantage is the spatial resolution (usually around 3mm). Because of its great spatial resolution, fMRI is the best method for localising neural activity in the brain (Czeszumski et al., 2020). However, it is difficult to trace the time course of that localised activity (Anderson, 2009; Anwar et al., 2016). Furthermore, because it uses blood flow for its estimation of neural activity, its temporal resolution is not nearly as good as that of EEG and MEG (Glover, 2011). Another disadvantage of fMRI is its very low mobility. Participants are required to stay still and stable in a laying position during the experiment. This makes fMRI unsuitable for studying social interactions in naturalistic settings (Czeszumski et al., 2020).

Functional Near-Infrared Spectroscopy (fNIRS)
Just like fMRI, fNIRS is a hemodynamic measurement technique as well. This technique indirectly measures brain activity through changes in the contrast between oxygenated (O_2Hb) and deoxygenated (HHb) haemoglobin concentrations (Anwar et al., 2016; Czeszumski et al., 2020). Similar to EEG, fNIRS is best suited for measuring superficial brain areas with a low spatial resolution (1 cm) (Scholkmann et al., 2013). Furthermore, its temporal resolution is lower than that of EEG. Nevertheless, fNIRS is widely used in research studies because of its mobility and particularly because of its resistance to motion artefacts. fNIRS is the best-suited method for experiments with a lot of movement since fNIRS is not strongly influenced by the movements of participants. This allows for even more naturalistic settings than the previous methods, especially when studying social interactions which require actions from participants (Czeszumski et al., 2020).

Figure 1 shows the different measuring equipment for (a) EEG, (b) MEG, (c) fMRI, and (d) fNIRS.

(a) EEG

(b) MEG

(c) fMRI

(d) fNIRS

Figure 1: *Brain-imaging measuring equipment*

2.1.2 Hyperscanning and Synchrony

Since hyperscanning and synchrony research are sometimes confused with each other and both terms will be used in this thesis, it is important to first note the similarities and differences between these terms.

Research by Cha and Lee (2019) has shown that neural synchronisation in Electroencephalogram (EEG) increases when communication between people is smooth and they work together as a team. This neural synchronisation can thus be used as a measure of how well people work together as a team. It can be stood to reason that if someone is ignored and socially excluded during teamwork (e.g. because they cannot understand the language that the other people in their team are speaking), the neural synchrony should be lower for this individual who is being ignored than the people who are working together.

This neural synchrony refers to the similarities in the physiological responses of two or more individuals (Stuldreher et al., 2019). While this can be measured by recording the brain activity of individuals one at a time and seeing if they respond similarly based on the same stimulus, individuals can also be measured simultaneously. When the brain activity of two or more individuals is measured simultaneously, it is called hyperscanning (Burgess, 2013).

One example of a study which investigated synchrony but did not use hyperscanning is the study by Cohen and Parra (2016). In their study, they presented 72 participants individually with simultaneous auditory narratives and visual animations while using EEG to record the neural activity. Memory retention was tested three weeks later. Their results showed that individuals with better memory retention had higher brain-to-brain synchrony with their peers since their neural responses were more correlated.

Thus, even though hyperscanning can always be performed to discover synchrony between individuals, synchrony does not always need to be measured by performing a hyperscanning study. However, if the objective of a study is to measure synchrony during social interaction, the use of hyperscanning becomes essential. After all, social interactions always involve two or more individuals interacting with each other and, thus, have to be measured at the same time (Xie et al., 2019).

2.1.3 Hyperscanning Studies

Hyperscanning is the simultaneous brain-imaging of two or more participants. The objective of hyperscanning studies is to determine how co-variations in the neural activity between individuals are influenced by their behavioural and social interactions. These co-variations in neural activity are also called hyperconnectivity. In most previous studies, hyperscanning has been done with no more than two participants, otherwise known as dyadic hyperscanning. These studies have been used to examine the social interaction between two participants (Xie et al., 2019; Burgess, 2013).

Several different methodologies (fMRI, fNIRS, MEG, EEG) can be used in order to demonstrate brain-to-brain synchrony. This synchrony can then, in turn, be used to predict a range of social and cognitive outcomes, such as communication quality (Dikker et al., 2014; Stephens et al., 2010), social closeness (Bevilacqua et al., 2019; Dikker et al., 2017; Parkinson et al., 2018), the degree of engagement in a task (Bevilacqua et al., 2019; Cohen et al., 2018; Dikker et al., 2017; Cohen et al., 2017) and memory retention (Cohen et al., 2018; Hasson et al., 2008; Cohen and Parra, 2016).

The first actual hyperscanning study was done by Montague et al. (2002). In this study, Montague et al. (2002) used two linked Functional Magnetic Resonance Imaging (fMRI) scanners in which they had two participants playing a children's guessing game. To further tackle the mobility problem, King-Casas et al. (2005) conducted a study using two scanners (one in Texas and one in California) and connected them via the internet. Afterwards, studies were performed in facilities containing two scanners (Koike et al., 2016). For example, Krueger et al. (2007) investigated the neural correlates of trust between two people and discovered that trust is an essential social process which is involved in all human interaction. Nevertheless, the lack of mobility and restriction of movement and communication between participants is still a huge problem because they result in very different experiment situations compared to real life. Furthermore, the complexity of fMRI data needs a new type of analysis to answer questions about brain-to-brain relations (Czeszumski et al., 2020). Besides, the cost of having multiple fMRI setups is very high, making it not accessible and available for everyone (Wang et al., 2018). These restrictions have resulted in the fact that fMRI hyperscanning studies have not been conducted that often to investigate social interactions (Czeszumski et al., 2020).

Other studies, such as the one by Funane et al. (2011) have used Functional Near-Infrared Spectroscopy (fNIRS) to investigate the relationship between brain-activation signals of two participants and the performance of these participants when doing a cooperative-task. Their results suggest that synchronised activation patterns of the participants cause better performance when they have to cooperate during a task. Since then, there have been many types of research that have adopted the fNIRS hyperscanning method (Czeszumski et al., 2020). In a recent study by Reindl et al. (2018), they even developed a fNIRS system for babies to study the brain functions that are related to parent-child interaction. fNIRS is particularly useful when studying infants or children's brain activation because it is relatively more tolerant of movement artefacts than other methods. For this reason, fNIRS is also a good method to apply to more naturalistic settings. Although fNIRS has a relatively good temporal resolution, its spatial resolution is low which means that it has a limited capability for detecting deep brain structures (Wang et al., 2018).

Additionally, Baess et al. (2012) even demonstrated that it is feasible to perform hyperscanning with Magnetoencephalogram (MEG) by connecting and synchronising two faraway neuromagnetometers. Furthermore, MEG hyperscanning has also been used to research the interaction between mothers and their children (Hirata et al., 2014). In a study by Mandel et al. (2016), the researchers used MEG to research speaker-listener roles during natural conversation. In other research by Ahn et al. (2018) MEG was combined with EEG in order to study verbal, inter-brain turn-taking. Boto et al. (2018) have even developed a mobile MEG system, making MEG a more attractive choice for future hyperscanning studies. However, these mobile systems are very new and more expensive than regular MEG systems. Thus, MEG is still not very suitable for researching social interaction in naturalistic settings, even though it does have better spatial resolution (Czeszumski et al., 2020).

While previously highlighted research has shown that hyperscanning is possible with two participants, the study by Xie et al. (2019) is one of the first and only studies which aimed at revealing neural correlates of social interactions using a triadic (3-person) hyperscanning technique. Specifically, they used a Functional Magnetic Resonance Imaging (fMRI) hyperscanning paradigm to measure the Blood-Oxygenation-Level-Dependent (BOLD) signal of the participants while they were doing a joint drawing task. Their results showed increased synchrony of the Right Temporal-Parietal Junction (RTPJ) while the participants were collaborating as a team. Furthermore, this increased synchrony in the RTPJ was shown to be positively associated with the overall performance. This shows that hyperscanning is also useful for revealing brain-to-brain synchrony of more than two participants. Just like Xie et al. (2019), this research will also endeavour to use hyperscanning for revealing brain-to-brain synchrony between three participants.

Even though fMRI, fNIRS and even MEG have been proven to be used successfully in hyperscanning studies, most studies have relied upon EEG hyperscanning. After all, equipment for EEG measurements is not only cheaper and more readily available in research facilities than the equipment for these other methods, but it is also better suited in settings which are more naturalistic because of its mobility (Burgess, 2013). This mobility, in addition to its great temporal resolution and usefulness for studying more than two individuals at a time, makes EEG hyperscanning a great tool for studying social interaction (Czeszumski et al., 2020). The EEG brain-imaging technique has, therefore, been chosen for the experiment in this thesis. Section 3.1 will go into more depth about previous EEG hyperscanning studies.

2.2 Heart Rate and Galvanic Skin Response

Humans are social creatures. A social connection or sense of belonging is even seen as a fundamental human need. When people feel like there is a threat to their sense of belonging (e.g. by being ignored and socially excluded) it could evoke powerful emotions (Molden et al., 2009). The occurrence of social exclusion by feeling ignored because of a language barrier could evoke a variety of negative emotions, such as anger, sadness, disappointment, and fear (Williams, 2002). These emotions are the exact opposite (e.g. happy, satisfied etc.) of what should be achieved in good team coordination situations in which people feel socially included.

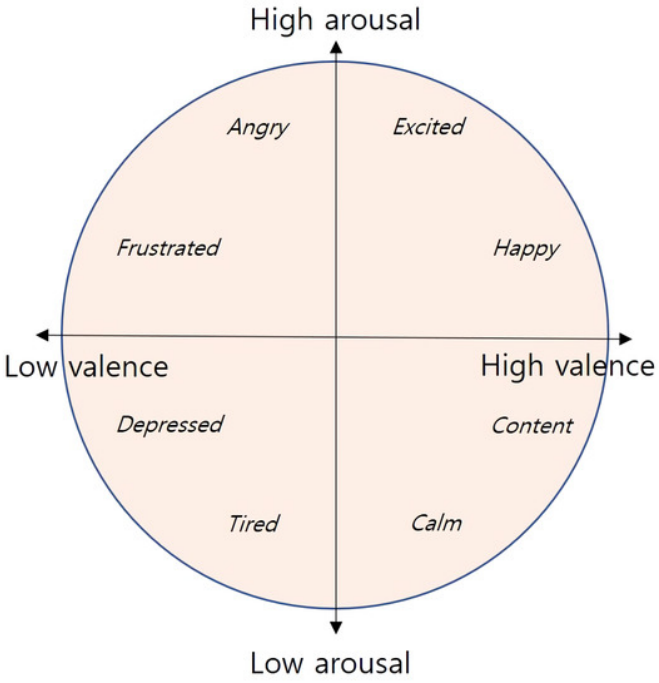


Figure 2: *Arousal-Valence Model*

It has been established that certain physiological signals have a strong correlation with emotions (Ferdinando et al., 2014). Heart Rate (HR) and Galvanic Skin Response (GSR) are two of these physiological signals which can be used to differentiate between emotions. According to the Arousal-Valence model by Russell (1980), emotions usually have different levels of arousal and valence (see Figure 2). Arousal is especially interesting since that can be reflected in HR and GSR. After all, high HR and GSR mean a high level of arousal which corresponds to the negative emotions associated with social exclusion, such as anger. Conversely, low HR and GSR correspond to a low level of arousal and thus to the positive emotions associated with social inclusion, such as feeling content (see Figure 2). In addition to EEG, it would, therefore, be very interesting to use these two physiological signals determine whether these signals are (de)synchronised in a similar way to the EEG signal (e.g. the emotions between the two participants that can understand each other will be synchronised while the emotions with the other participant will not be).

2.3 Analysing Hyperscanning Data

The following sections will go into detail about the neural systems which are involved in social interaction and which types of measurements there are for analysing hyperscanning data and finding synchrony, particularly with regards to EEG hyperscanning data.

2.3.1 Neural Systems Involved in Social Interactions

In general, there are two main neural systems involved in brain-to-brain connections which are made visible through hyperscanning: the Mirror Neuron System (MNS) and the Mentalizing System (MS). The MNS includes the primary motor cortex as well as the posterior parietal cortex while the MS consists of the Temporal-Parietal Junction (TPJ), the precuneus and the Prefrontal Cortex (PFC). Figure 3 gives a visual representation of these neural systems. Both systems play a vital role in social interactions. Therefore, this section will now go into more detail about each of these systems (Wang et al., 2018).

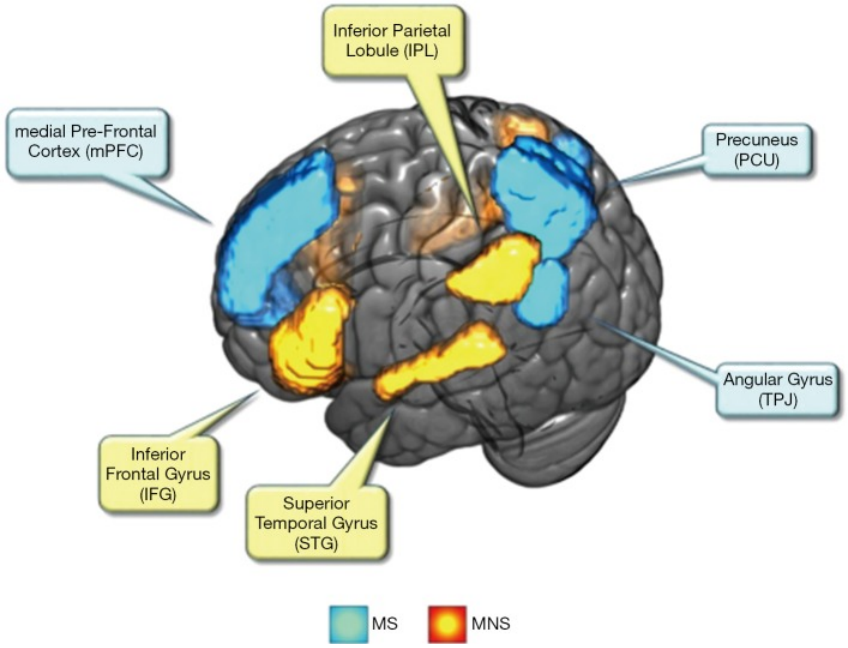


Figure 3: *Neural Systems Involved in Hyperscanning: MNS and MS*

Mirror Neuron System (MNS)
When people imitate or mirror other people's actions or movements, neurons in the MNS are fired. This even happens when people are just watching actions or movements performed by other people. The MNS consists of the Inferior Frontal Gyrus (IFG), the Inferior Parietal Lobule (IPL), and the Superior Temporal Gyrus (STG). Both the IFG and the IPL are related to language, motor and sensory detection (Wang et al., 2018). The STG is able to provide additional and more sophisticated visual information. This information can then be delivered to the IPL (Iacoboni and Dapretto, 2006).

Potential movements can then be executed once the IPL is activated. Additionally, the IFG is also activated. This activation allows for manipulation of potential action which can provide additional information. Such information could, for example, be the goal of the action (Wang et al., 2018).

One hyperscanning study which showed that MNS is involved in imitation is the study by Dumas et al. (2010). This study showed that when two participants were synchronised in their behaviours because they were imitating each other, their brains would also be tuned to the same frequency. This meant that there existed a brain-to-brain synchronisation in the alpha-mu band between the right centroparietal regions of the two participants.

Mentalizing System (MS)

Mentalizing is known as trying to understand other people's intentions or emotions by looking at their gestures, behaviours, and facial expressions. The two main brain regions which are associated with this mentalizing process are the TPJ and the (dorsomedial) PFC (Saxe, 2006).

In the study by Carter and Huettel (2013), it has become clear that the mentalizing process consists of two steps. In the first step of this process, static social images will be coded as a neural representation. In the second step, these encoded representations are then used to generate moving social entities, which can be used for interpreting the intention.

The TPJ is located between the temporal and parietal cortex and can be seen as a boundary region between them (Wang et al., 2018). In a hyperscanning study by Tang et al. (2016) it was shown that the interpersonal brain coherence of the right TPJ was higher when participants were face-to-face than when they were blocked from each other. This result demonstrated that the right TPJ is important for social interactions.

The PFC is responsible for regulation, planning and integration of information, as well as other high cognitive functions. One study by Jiang et al. (2012) demonstrated evidence that PFC is related to brain-to-brain synchronisation. They showed that inter-brain coherence (for the left inferior frontal cortex) was higher when participants were having face-to-face conversations than when they were back-to-back.

Both the MNS and MS neural systems are, thus, of particular interest in this thesis, since these brain areas are most likely to show the brain-to-brain synchrony between the participants.

2.3.2 Synchrony Measures

There are several different types of measurements which can be used to analyse hyperscanning data. These measurements can be divided into the following categories: coupling/connectivity measures, correlation and dependence analysis, and the analysis of information flow. These different types of measurements will now be explained in further detail.

Coupling/Connectivity Measures
Methods in this category estimate the strength of coupling/connectivity between brains. While they generally make calculations in the Fourier domain, they all differ in how they combine different frequencies and the kind of normalisation they use. Examples of coupling/connectivity measures are Phase-Locking Value (PLV), Phase Lag Index (PLI), and phase coherence (Czeszumski et al., 2020).

PLV measures how two signals from two brains are phase-locked in a specific time window. It was first introduced by Lachaux et al. (1999) and looks at whether the phase difference varies across trials. This means that multiple trials are needed to compute the PLV. An example of a study in which this method was used is the one by Dumas et al. (2010). In this study, the PLV was computed for each pair (i, k) of electrodes between two EEG caps. Electrode i and k represent the caps 1 and 2, respectively. This computation was done for each frequency band according to the equation:

$$PLV_{i,k} = \frac{1}{N} \left| \sum_{t=1}^{N} exp^{j(\phi_i(t) - \phi_k(t))} \right| \tag{1}$$

where N is the number of samples considered in each time window, ϕ is the phase and $||$ the complex modulus. Thus, the value of PLV is 1 if the two signals are in perfect sync (if the phase difference does not vary across trials) and 0 if the signals are unsynchronised.

PLI is a very similar measure to PLV and was used in the study by Ahn et al. (2018) for instance. Just like PLV, the value of the PLI is also 1 if two signals are in perfect synchronisation and 0 if there is no coupling between the signals. Nevertheless, the way it is computed is slightly different:

$$PLI = |\langle sign[\Delta\phi(t_k)]\rangle| \tag{2}$$

where $\Delta\phi(t_k)$ is a time series of phase differences between two signals with $k = 1...N$ and N being the number of samples (Stam et al., 2007).

However, while PLV suffers from the common source problem, PLI does not. Nevertheless, because the sources are separated between brains in the case of hyperscanning studies, both PLV and PLI should give the same results (Aydore et al., 2013).

Additionally, phase coherence is another measure related to phase or neural oscillations synchronisation between brains. It measures the similarity between two signals. Thatcher (2012) provides a good overview of the different phase coherence measures. Related to phase coherence is the Wavelet Transform Coherence (WTC) which also measures the coherence of two signals. This method is mostly used in fNIRS hyperscanning studies such as the one by Cui et al. (2012). However, this method is also pretty common for

studying brain-to-brain synchrony in other hyperscanning studies, although it is a more complex method than PLV and PLI (Czeszumski et al., 2020).

The WTC of a time series x_n of length N which is sampled from an underlying waveform at equal time steps of size Δt is computed using the following equation (Chang and Glover, 2010):

$$WTC^X(n,s) = \sqrt{\frac{\Delta t}{s}} \sum_{n'=1}^{N} x_n \psi_0[(n'-n)(\frac{\Delta t}{s})] \tag{3}$$

where n is a time index and s denotes the wavelet scale. ψ_0 is a function which then also needs to be chosen. The complex Morlet wavelet is often chosen (Chang and Glover, 2010):

$$\psi_0(\eta) = \pi^{-1/4} e^{i\omega_0\eta} e^{-\eta^2/2} \tag{4}$$

where ω_0 represents the relative time and frequency resolution. $WTC^X(n,s)$ can then be used to express the amount of power in x as a function of time and frequency, whose angle represents the local phase. For hyperscanning, the cross-wavelet transform is needed between two time series X and Y, which is defined as follows (Chang and Glover, 2010):

$$WTC^{XY}(n,s) = WTC^X(n,s) WTC^Y(n,s) \tag{5}$$

All mentioned coupling/connectivity measures are measures of the similarity between two neural signals from different brains. Such similarity is then interpreted as brain-to-brain synchrony, otherwise known as inter-brain synchrony (Czeszumski et al., 2020).

Correlation and Dependence Analysis

The second category consists of measures which estimate the correlation between signals from two brains to measure synchronisation. Different types of correlation measures are applied to different brain-imaging techniques. These measures are mostly used for analysing fMRI hyperscanning data. For fMRI, it is not the BOLD signal itself which is used for the correlation analysis, but the regression model coefficients which represent activations (Czeszumski et al., 2020). Thus, linear dependence is used to estimate the relation between two measured brains. Koike et al. (2016), for example, have used this measure for their analysis.

Cross-correlation in combination with Independent Component Analysis (ICA) decomposition of the BOLD signal can also be used to further extend the dependence analysis of fMRI data. For instance, the paper by Bilek et al. (2017) applied this analysis on joint attention and interpreted the cross-correlation between two brain signals as information flow.

According to Asuero et al. (2006), the co-variance between two signals x and y is a measure of the correlation of the fluctuation. It can be computed using the following equation:

$$cov(x,y) = \frac{1}{n-1} \sum (x_i - \bar{x})(y_i - \bar{y}) \tag{6}$$

However, the co-variance itself is often not a useful measure of correlation because its value depends on the scales in which x and y are measured. It must be standardised before it can be applied as a measure of correlation. If the co-variance is divided by the product of the standard deviation of x and y, the correlation coefficient r_{xy} is obtained. A standard equation for the computation of the correlation coefficient r_{xy} for two signals x and y with a linear relationship is then as follows:

$$r_{xy} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2(y_i - \bar{y})^2}} \tag{7}$$

Using the deviations of both signals x and y from the mean, the above part of the equation measures the degree to which x and y vary together. The lower part of the equation measures the degree to which x and y vary separately. Thus, a correlation coefficient is obtained (Asuero et al., 2006). This correlation coefficient is known as the Pearson Correlation Coefficient (PCC) (Benesty et al., 2009). The value of the Pearson Correlation can range between -1 and 1, with 1 being the most correlated.

In EEG hyperscanning studies, different aspects of the EEG signal were used for the correlation analysis. Thus, different values of an EEG signal can be used as values of x and y in the equation above. Kawasaki et al. (2013), for example, used the correlation between the theta and alpha frequencies while Kinreich et al. (2017) used alpha, beta and gamma frequency bands. The correlations that are found in hyperscanning studies using these kinds of measurements are interpreted as neural synchronisation between brains.

Information Flow

Previous categories focused on the analysis of synchrony and similarity. Besides these two options, hyperscanning analysis can also be focused on the information flow from one brain to another. Measurements which can be used are, for instance, Granger Causality and Partial Directed Coherence (PDC) (Czeszumski et al., 2020). The study by Astolfi et al. (2011) is one of the studies which applied these methods to estimate links between brains of pilots who were cooperating with each other and found that causal links are stronger during increased cooperation.

Baccalá and Sameshima (2001) defined PDC as:

$$PDC_{ij}(f) = \frac{A_{ij}(f)}{\sqrt{a_j^*(f)a_j(f)}} \tag{8}$$

where $A_{ij}(f)$ is an element of $A(f)$. $A(f)$ is a Fourier transform of Multivariate Autoregressive Model (MVAR) model coefficients $A(t)$ in which $a_j(f)$ is the j-th column of A(f). The * in the equation denotes the transpose operation.

Even though the causal links between brains can be estimated with these information flow methods, it is important to understand the difference between information flow and actual brain-to-brain synchrony. After all, the sensory input in both cases is identical (Czeszumski et al., 2020).

2.3.3 Most Common Measurements for EEG Synchrony

According to Burgess (2013), there are three main methods which have been used in EEG hyperscanning studies to determine brain-to-brain synchrony between socially interacting people: co-variance in amplitude or power, Partial Directed Coherence (PDC), and phase synchrony (mostly by using Phase-Locking Value (PLV)). However, some studies have used variations of these three main methods.

The most frequently used method is to show that there are adjacent or almost adjacent changes/co-variances in EEG amplitude or power (see equations 6 and 7). This amplitude or power is mostly estimated from event-related changes or Fast Fourier Transform (FFT). However, showing that there are co-variances in the EEG amplitude or power is a weak form of showing brain-to-brain synchrony since it is by no means conclusive (Burgess, 2013).

The second most commonly used method is the use of PDC (see equation 8). This method was used in the first EEG hyperscanning study by Babiloni et al. (2006). PDC is based on multivariate autoregressive modelling and Granger Causality and can show the linear direction flow of information between two different systems. This makes PDC ideally suited for determining inter-brain synchrony in a hyperscanning study in which one person's behaviour drives the behaviour of the other individual (Burgess, 2013). However, PDC does have some limitations since the results of the use of PDC in hyperscanning studies could not be replicated well (Konvalinka and Roepstorff, 2012).

The last most frequently used method involves measures of phase synchrony, with PLV being the most common one (see equation 1). PLV is a statistical measure, introduced by Lachaux et al. (1999), which can be used to research task-induced changes in synchronisations of neural activity. The value of PLV is close to 1 if the variability of the phase is very small, otherwise, it is close to 0. This measure is well suited for capturing the rapid flow of information that exists between people who are interacting in social situations (Burgess, 2013). For this reason, PLV is the most suitable brain-to-brain synchrony measurement for measuring the level of social exclusion.

The Challenge of Interpreting Synchrony Correctly

Even though both PDC and PLV have been widely used for measuring brain-to-brain synchrony between two or more people by measuring coupling between cortical oscillations in the EEG, they actually measure different things depending on each case. The reason why these measurements do not always measure the same thing is that they might be measuring different kinds of synchronisation (Burgess, 2013). Additionally, this also makes it difficult to relate the results from different studies (Czeszumski et al., 2020).

As can be seen in Figure 4, there are four different types of synchrony. Depending on the context, each of these types of synchrony (except for (D)) might be of interest in a hyperscanning study. Figure (A) shows *reciprocal* synchrony in which the pendulum of the clocks are swinging in phase because there is a mutual influence between the two. Figure (B) shows *induced* synchrony in which the phase of both pendulums are influenced by a common external driver. In hyperscanning studies, this form of synchronisation might occur if the participants simultaneously experience the same stimulus even though they are not actually interacting (e.g. watching a movie together). Figure (C) shows *driven* synchrony in which the pendulum of one clock influences the second pendulum without there being any mutual influence. This is the type of coupling PDC is designed to identify.

Figure (D) shows *coincidental* synchrony in which there is no coupling between the two clocks but the pendulums are swinging in the same phase because they coincidentally have the same frequency. Thus, simply observing a consistent phase does not mean that there is any synchronisation or information exchange. Most hyperscanning studies, therefore, do not simply measure phase coupling between individuals but also compare the degree of coupling between different experimental conditions (Burgess, 2013). This is what this thesis research will do as well since the coupling between the two participants who are able to understand the same language will be compared with their coupling with the excluded participant.

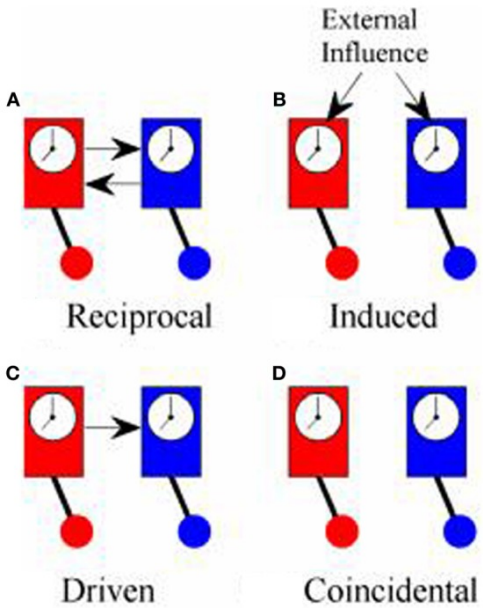


Figure 4: *Four different types of synchrony*

2.4 Preliminary Conclusions

The literature described throughout this chapter shows that while hyperscanning research can be successfully performed using either MEG, fMRI, or fNIRS, the usage of EEG has the most advantages. After all, its equipment is not only cheaper and more readily available, it is also better suited for more naturalistic settings because of its mobility. Moreover, this makes EEG ideal for studying more than two individuals at the same time, which is necessary for studying social interactions in which someone is linguistically ostracised. Therefore, the EEG brain-imaging technique has been chosen for the current thesis research.

Furthermore, the research in this chapter indicates that the MNS and MS neural systems are most likely to show the brain-to-brain synchrony between the participants. The specific brain areas included in these neural systems are: the primary motor cortex, the posterior parietal cortex, the TPJ, the precuneus and the PFC.

Lastly, the literature suggests that the PLV synchrony measure is best suited for measuring EEG synchrony in the current study because of its ability to capture the rapid flow of information which exists between socially interacting people. The PLV measure will, therefore, be used to determine the level of neural synchronisation between participants.

3 Related Work

3.1 Hyperscanning Studies Using EEG

To measure social exclusion during interactions, hyperscanning becomes necessary. After all, social exclusion only occurs in many-to-one interactions and the simultaneous brain imaging of two or more participants can only be done using hyperscanning (Xie et al., 2019; Burgess, 2013). There are several brain-imaging techniques which can be used in hyperscanning studies. However, as has been explained before, this chapter will focus on hyperscanning studies using EEG in particular. There are several reasons for this; the most important one being that EEG measurement equipment is more easily available than for the other brain-imaging techniques. Therefore, this section will go into more detail about the methodology and data analysis used in these studies. Part of these insights are based on the overview by Czeszumski et al. (2020).

The first EEG hyperscanning study was done by Babiloni et al. (2006). It involved five groups of four participants playing a cooperative card game. Cortical activity from the EEG recordings was estimated by solving the associated linear inverse problem in the following areas: frontal, prefrontal, and parietal cortical area. The functional connectivity between the waveforms of the brain areas of the same participants has been computed by using the Partial Directed Coherence (PDC) algorithm. In their analysis Babiloni et al. (2006) used the beta frequency band, but the results were representative of the results obtained in the other frequency bands as well. This study has revealed functional links between prefrontal areas (especially with regards to the Anterior Cingulated Cortex (ACC)) of the different participants while they are participating in the cooperative game.

While this research by Babiloni et al. (2006) was conducted in a lab with full control over the environments, further developments of the EEG equipment made it possible to conduct more naturalistic paradigms (Czeszumski et al., 2020; Melnik et al., 2017). Lindenberger et al. (2009), for example, made it possible to perform EEG hyperscanning on eight pairs of participants while they were playing a short melody on their guitars together. Brain areas of interest in this study were the frontal, central, and temporal areas. The fronto-central areas are associated with control and coordination of motor activity for controlling the guitars. Furthermore, the temporal area has been shown to be involved in music production and coordinated behaviour. Their analysis was restricted to frequency bands below 20 Hz since this frequency range is important for interpersonal action coordination. Brain-to-brain synchrony was analysed by the Phase Locking Index (PLI) and Interbrain Phase Coherence (IPC). Their results showed that brain-to-brain synchrony primarily involved fronto-central regions of the brain and was strongest in the frequency range between 0.5 Hz and 7.5 Hz. Additionally, Mueller et al. (2013) performed a similar study on musical improvisation with pairs of guitarists. Their results also showed that inter-brain connections primarily involved lower frequency bands (e.g. delta and theta).

Davidesco et al. (2019a) have performed an EEG hyperscanning study to record brain activity from four students and a teacher at the same time to investigate whether brain-to-brain synchrony can predict learning outcomes. To identify this brain-to-brain synchrony, they looked at the alpha band (8-12Hz) because it has been shown to be the most robust frequency range for brain-to-brain synchrony (Dumas et al., 2010). Davidesco et al. (2019a) computed the Circular Correlation (CCorr) values for all combinations of

electrodes across pairs of participants. Their results showed that synchrony in the alpha band predicted delayed memory retention while brain-to-brain variation in the alpha band can be used to discriminate between retained or forgotten content. Davidesco et al. (2019b) also conducted a very similar study using EEG hyperscanning in which they discovered that brain-to-brain synchrony is a more accurate predictor for long-term memory retention than individual measures.

Additionally, Kawasaki et al. (2013) conducted an EEG hyperscanning study with twenty pairs of participants. These participants had to perform a speech task in which each of the individuals in a pair alternatively pronounced letters of the alphabet. All individual participants also performed the task with a machine that pronounced the letters (both before and after the duo task). They conducted a wavelet analysis on the EEG data to characterise the brain oscillatory activity. Their results showed that in human-human tasks, speech rhythms were more likely to become synchronised than in the human-machine tasks. Furthermore, the theta/alpha (6-12 Hz) amplitudes in the central and parietal regions were significantly higher during the human-human tasks than the human-machine tasks. Lastly, they discovered that brain-to-brain synchronisation is tightly linked to speech synchronisation.

All these EEG hyperscanning studies have shown that EEG is a good tool which can be used to map moment-to-moment interactions between individuals simultaneously. Furthermore, the co-variations of the brain activations of these individuals were shown to be correlated with social interaction and brain-to-brain synchronisation. Its high temporal resolution is a big reason for this (Wang et al., 2018). Moreover, this previous EEG hyperscanning research showed that the most important brain regions for social interactions are the frontal, parietal, central and temporal regions. Lastly, these studies have shown that frequency bands below 12 Hz (e.g. alpha/theta/delta) are most important for studying synchrony during social interactions.

Nevertheless, even though previous research has, thus, shown that the most important brain regions are the frontal, parietal, central and temporal regions, and that the most important frequency bands are below 12 Hz, it does not necessarily mean that it will be the same for this thesis research. After all, the beta frequency band is relevant to active thinking and the gamma frequency band for attention (Baumeister et al., 2008; Tatum, 2014). All brain regions and frequency bands will therefore be compared to each other to determine if the current study will have similar findings as the previously mentioned related work.

3.2 Combining EEG, Heart Rate and Galvanic Skin Response

In this thesis, the approach is to measure social exclusion, not just by using Electroencephalogram (EEG), but to combine it with Heart Rate (HR) and Galvanic Skin Response (GSR) to determine the level of synchronisation. However, there have not been many studies which combined EEG, HR and GSR data into a multi-modal system. Let alone with regard to using this multi-model system to measure synchrony between individuals. Nevertheless, this section will review some studies that have combined these modalities.

The research by Vecchiato et al. (2010) is one of these studies which did combine EEG, HR and GSR. They used these modalities to measure changes in brain activity during the observation of TV commercials. However, they did not actually combine these three modalities into one single multi-modal system. Instead, they analysed EEG separately from HR and GSR. Nevertheless, Vecchiato et al. (2010) did use the same measure for all three modalities since they computed the z-score based on indices derived from the three measurements. The obtained z-scores were divided into two groups: the group that remembered the TV commercials they watched during the experiment and the group that did not. Statistical analysis was then performed using the Analysis of Variance (ANOVA) to determine whether there was a significant difference between the two groups for four different frequency bands.

Das et al. (2018) combined the three modalities (although they used photoplethysmogram instead of ECG) to improve stress detection. While this research did combine the features of all three modalities, it was used in a classification system. While such a classification system can indeed be used to detect differences between two groups, only a statistical analysis can be used to prove or disprove a hypothesis. Thus, since the goal is to test the hypothesis that there is a difference in synchrony between the social inclusion and social exclusion condition, a statistical analysis is needed for now instead of a classifier.

Furthermore, Reinerman-Jones et al. (2011) compared different statistical analysis methods for the combination of neurophysiological signals (EEG, HR, and GSR) to determine which method is best suited for each situation. In their research, they compared the results of five different statistical analysis methods: correlation, ANOVA, Multivariate Analysis of Variance (MANOVA), regression, and Discriminant Function Analysis (DFA). The analyses used data which is computed from the change in power measured during the task compared to the baseline. Reinerman-Jones et al. (2011) conducted the analysis with the different measures (e.g. ANOVA etc.) for each of the different modalities (e.g. EEG, heart rate etc.) as the dependent variable and the elicited emotion as the independent variable. Their results show that, although they are limited in their ability for differentiating between the results of the different modalities, the ANOVA and Correlation provide direct methods for analysing neurophysiological data (Reinerman-Jones et al., 2011).

This last research by Reinerman-Jones et al. (2011) is probably the most useful to use as reference for what is needed for measuring social inclusion/exclusion using these three neurophysiological signals (EEG, HR, and GSR). However, it is still unclear how exactly these three modalities can be combined to measure synchrony. One possible solution would be to calculate the correlation/synchrony for the data from each of the three modalities separately. High correlation/synchrony would then be expected in the social inclusion condition for the two individuals who are working together. Low correlation/synchrony with the other two individuals for the third individual who is socially excluded would be expected in the social exclusion condition. Then, statistical analysis can be used

to determine whether there is a significant difference in correlation/synchrony between the conditions. Although Reinerman-Jones et al. (2011) concluded that computing the correlation for their neurophysiological is a relatively weak method, it seems, nevertheless, to be the most promising option for the HR and GSR data.

3.3 Social Inclusion/Exclusion and Language Barriers

To measure social exclusion it is essential to understand what exactly it means to be socially excluded as well as what it means to be socially included. Social inclusion and social exclusion are relatively new terms and are the inverse of each other. They first started to be used in the 1990s. In these times, social exclusion was used in reference to those excluded from the Social Contract (e.g. through lack of payment for their work). The terms were later used in the European Union's Lisbon Strategy of 2000 which made them gain prominence (Piller et al., 2012). Social exclusion sometimes refers to the absence of economic well-being (e.g. un- and underemployment) and the absence of civil and social rights (e.g. healthcare and education) (Burchardt and Le Grand, 2002). In the paper by Rawal (2008), social exclusion has been defined as 'the process through which individuals or groups are wholly or partially excluded from full participation in the society within which they live'.

As can be seen, social exclusion (and thus social inclusion) is defined in a relatively extreme, broadly societal and intercultural way. However, there are also more mellow forms of social exclusion such as ignoring, excluding, and rejecting someone. As has been argued by Williams (2007), nearly everyone has experienced this form of social exclusion or isolation at some point in their lives. Another word for social exclusion in this context is ostracism. When people are ostracised they can feel an increase in sadness and anger because their needs of belonging, self-esteem, control and a meaningful existence are thwarted (Williams, 2007).

One way of feeling excluded could arise because a non-mutually understood language is spoken. This linguistic ostracism could arise at work, at school or any other place with social interaction. Employees or students could perceive that others at work or school have rejected and/or excluded them by using a language that they do not comprehend (however unintentional it may be) (Fiset and Bhave, 2019). Moreover, social exclusion caused by a language barrier occurs in many-to-one settings (Dotan-Eliaz et al., 2009). This means that to linguistically ostracise someone, two or more people are required to interact in a language unfamiliar to that person. Reasons for members of such multinational teams to switch to their mother tongue when they are conversing with colleagues from their home country are, for example, to streamline communication and to make it easier to express emotions (Tenzer et al., 2014).

As discussed previously, with the increase in the number of languages spoken in workplaces such as universities, there is a heightened prospect of linguistic ostracism. Employees, students or peers who are linguistically ostracised will view themselves as members of a linguistic outgroup. This linguistic outgroup is the exact inverse of the so-called linguistic ingroup in which people converse in a language that everyone in the group understands. Being part of such a linguistic ingroup is also called linguistic inclusion (Dotan-Eliaz et al., 2009). Perceiving themselves as belonging to a linguistic outgroup will cause people to perceive disidentification from their peers (Kulkarni, 2015; Voss et al., 2014). After all, language acts as a primary method for communicating information. When you're

unable to understand what is communicated, it puts a significant strain on interpersonal relationships (Kulkarni, 2015). Consequently, people feel a stronger connection to those peers who are part of their linguistic ingroup than those of their linguistic outgroup. This is in line with the ethnolinguistic identity theory (Fiset and Bhave, 2019). While previous research has been mostly focused on linguistic ostracism in the workplace it can be inferred that this would also cause problems for students or other people who need to work as a team.

Furthermore, linguistic ostracism also affects trust formation. Henderson (2005) was the first to have made this connection between language and trust formation, especially with regards to teamwork. Cognitive and emotional reactions that people have to these language barriers influence the perceived trustworthiness of their peers as well as their intention to trust them (Tenzer et al., 2014). This is not at all favourable when teamwork is required since trust is the most basic ingredient of team collaboration and coordination (Kasper-Fuehrera and Ashkanasy, 2001).

Linguistic ostracism removes verbal social contact by using a non-mutually understood language (Neeley et al., 2009). According to Robinson et al. (2013), ostracism can thus be seen as an act of omission which violates social norms. However, linguistic ostracism is unique in the sense that it is a usually non-purposeful form of ostracism (Robinson et al., 2013). Therefore, linguistic ostracism may not be attributed to ill will (Ferris et al., 2017).

Nevertheless, this does not negate the fact that, when people are linguistically ostracised, they will no longer experience the benefits of belonging to the ingroup. Such a benefit is, for example, that they view their personal success and that of their peers as being inevitably linked, which causes them to work harder (Fiset and Bhave, 2019). Thus, when people feel part of the outgroup they will disidentify from their peers and make fewer beneficial contributions to the team (Lauring, 2008). Furthermore, they will also put a lower value on identifying themselves as being a member of their team (Tenzer et al., 2014). They feel affronted and angry by their treatment and will likely engage in fewer helping behaviours (Thau et al., 2007). It will be interesting to confirm these results and find out whether this desynchronisation from their peers is also visible in the brain.

3.4 Team Coordination

As has been said by Thau et al. (2007) and Lauring (2008), people who are socially excluded will engage in fewer helping behaviours when working as a team. Conversely, people who are socially included will work harder to reach success with their team (Fiset and Bhave, 2019). This means that studying team coordination could reveal social exclusion and inclusion during teamwork.

According to Gorman et al. (2010) team coordination consists of the dynamics of team member interactions as well as the environmental dynamics (such as being in the same team twice (intact teams) or in a different team the second time (mixed teams)) to which a team is subjected. Their research showed that teams who did not have a history together (mixed teams), were more adaptive to changes in the task. By coordinating as a team, people can accomplish more than they would when working alone. Thus, the quote by Aristotle "the whole is greater than the sum of its parts" fits perfectly in the context of team coordination (Gorman, 2014).

Unlike normal coordination, which naturally happens when doing daily activities, team coordination can be studied in a lab environment as well as in naturalistic settings. It is related to the degree of common or complementary knowledge of team members. This is called the shared-knowledge approach or shared mental model (DeChurch and Mesmer-Magnus, 2010). Thus, a shared mental model of a team is a model in which the individual mental models overlap or complement each other in terms of knowledge. This shared mental model ensures that team members can describe and explain knowledge to each other as well as predict each other. Therefore, it facilitates a team's ability to coordinate activities, which is directly related to team performance. Better team coordination means better team performance (Gorman, 2014).

3.4.1 Brain-Imaging and Team Coordination

Team coordination has, of course, also been studied using brain-imaging techniques. Likens et al. (2014), for example, have studied the neural signatures of team coordination by using multifractal analysis. They explored the behaviour of a team consisting of six people while they were doing a Submarine Piloting and Navigation task. The data showed the distribution of activity across all team members and was recorded by a nine-channel EEG. The multifractal analysis was able to identify social patterns from the brain activity of the individuals participating in this social interaction as a team (Likens et al., 2014).

Additionally, Szymanski et al. (2017) researched how local and inter-brain phase synchronisation correlates with better teamwork. In their study, they asked participants to perform a visual search task while they were simultaneously measuring the participants EEG. This task was to either be performed alone or in a duo. Both local phase synchronisation, as well as inter-brain synchronisation, were found to be higher when the participants worked together on the visual search task as opposed to when they did this task individually. However, not all duos benefited from working together. This performance gain (or lack thereof) was positively correlated with inter-team differences in local and inter-brain phase synchronisation. The higher the phase synchronisation, the better the team coordination and performance (Szymanski et al., 2017).

3.4.2 Team Coordination Hyperscanning Paradigms

Several different paradigms have been used in hyperscanning studies about team coordination and other social interactions. Overall, six different categories can be identified: cooperation and competition tasks, imitation tasks, coordination tasks, eye contact/gaze tasks, game theory/exchange tasks, and a natural scenario (Wang et al., 2018). While this thesis research will focus on a cooperation task, this section will briefly explain all the different kind of tasks that have been used in previous hyperscanning studies.

Cooperation and Competition Tasks
The first category (and category of choice) are cooperation and competition tasks. In these kinds of tasks, participants need to achieve a certain goal cooperatively or competitively (Wang et al., 2018). An example of a study in which such a task was used was in the study by Cui et al. (2012). This study consisted of three conditions: cooperate, competitive, and control. In the cooperate condition, both participants needed to press a button as soon as possible after a certain cue. Their response time needed to be below a threshold for both of them to score a point. If this was not the case they would both receive nothing. In the competition condition, the participant that responded the fastest received the point. In the control condition, just one of the participants needed to press the button. Their results showed that coherence between signals (brain synchrony) in the right superior frontal cortices increased during cooperation, but they did not increase during competition.

Imitation Tasks
The second category consists of imitation tasks. In these tasks, participants need to imitate the others' movements and/or behaviours (Wang et al., 2018). For example, Dumas et al. (2010) asked their participants to imitate the other participant's meaningless hand movements. With their research, they showed that brain-to-brain synchronisation of the right centroparietal regions (at the alpha-mu band) and the interactional synchrony were strongly correlated.

Coordination Tasks
The third category is the coordination tasks. In this category participants (two or more) need to act in a synchronised way (Wang et al., 2018). Sometimes, people even do this unconsciously, such as when footsteps are synchronised with the footsteps of a friend while you are walking together (Yun et al., 2012). An example of a study in which a coordination task is used is the research by Mu et al. (2016). In their study pairs of participants were instructed to synchronise with each other by rhythmically counting in their head. Their results showed greater brain-to-brain synchrony during the coordination task (vs the control task).

Eye Contact/Gaze Tasks
The fourth category consists of eye contact or gaze tasks. In these kinds of tasks pairs of participants are asked to look into each other's eyes or look towards a certain object (Wang et al., 2018). Eye contact is very important during non-verbal communication and for inferring other's intentions (Hirsch et al., 2017; Koike et al., 2016). One example of research done with such a task was the study by Hirsch et al. (2017). In their research, pairs of participants were asked to look into either each other's eyes or into the eyes of people in portraits. Their results showed that inter-brain coherence was significantly greater in the eye-to-eye gaze condition compared to the eye-to-portrait gaze condition.

Game Theory/Exchange Tasks

The fifth category is economic games involving game theory/exchange tasks. In such tasks, one participant is given an economic offer while the other participant of the duo needs to decide whether they want to take the offer or not. Exchange tasks mostly involve a really basic type of social interaction in which social behaviour is exchanged for some sort of reward (Wang et al., 2018). A good example of a game theory/exchange task is the trust game. In this game, one of the participants needs to decide how much money will be returned to their opponent. It was shown that for building a trustworthy relationship, synchrony in the paracingulate cortex is critically involved (King-Casas et al., 2005; Krueger et al., 2007).

Natural Scenario

The last category is a natural scenario. While the previously mentioned tasks can offer good opportunities to research brain-to-brain synchrony during social interactions such as team coordination, only natural scenario's can offer a reflection of real-life situations (Wang et al., 2018). An example of such a natural scenario is researched by Dikker et al. (2017). Their study showed that students' brain-to-brain synchrony was increased when they were highly engaged in the teaching.

While all these team coordination paradigms can be used to measure brain-to-brain synchrony, not all of them can be used to measure social exclusion that is caused by a language barrier. After all, to socially exclude someone based on language, the task needs to be one with many opportunities for conversations. This only leaves cooperation tasks and the natural scenario as the best options. However, the natural scenario is extremely difficult to research, making the cooperation task the chosen option for this thesis research.

3.5 Preliminary Conclusions

The findings of the related work allow for some more preliminary conclusions to take into account for this thesis research in addition to the preliminary conclusions from the background research. First of all, from the hyperscanning studies using EEG, it can be concluded that the most important brain regions for brain-to-brain synchrony are the frontal, parietal, central, and temporal regions. This is somewhat in line with the findings of Wang et al. (2018) about the two main neural systems that are involved in this brain-to-brain synchrony. Furthermore, the literature suggests that the most important frequency bands for showing this synchrony are below 12 Hz. All brain regions and frequency bands will be compared to each other to determine if this thesis research will have similar findings.

Furthermore, previous research shows that there have not been many studies which combined EEG, HR, and GSR data into a multi-modal system. Nevertheless, the literature indicates that computing the correlation between each of the three participants for the data from the HR and GSR modalities (EEG synchrony will already be computed using the PLV measure as explained in Section 2.3.3) will be the best solution. Thus, the PCC, as defined in equation 7 will be used as the synchrony measure for the HR and GSR data.

Additionally, the literature about social exclusion induced by language barriers indicates that people who are linguistically ostracised will feel affronted and angry by their treatment and will likely engage in fewer helping behaviours. This thesis research will endeavour to confirm whether participants indeed feel socially excluded when they cannot understand the language other participants are speaking and whether they appear to engage in fewer helping behaviours. It will also be determined whether this desynchronisation from their peers is visible in the measured neurophysiological signals (EEG, HR, and GSR) as well.

Lastly, previous work suggests that studying team coordination could reveal social exclusion and inclusion during teamwork. Thus, a team coordination hyperscanning paradigm is needed to measure the social exclusion that is of interest in this thesis research. From the six different task categories, it can be concluded that a cooperation task is the best suited for studying social exclusion since such a task will give many opportunities for verbal communications. A language-oriented cooperation task is thus chosen for the thesis research.

The next chapter will describe how these conclusions are reflected in the experimental setup and the following (pre-)processing steps that are needed for finding an answer to the research question which has been proposed in the Introduction.

4 Methodology

A global graphical overview of the steps needed to (statistically) compare the synchrony between the participants (for each of the modalities: EEG, HR, GSR) in the social inclusion condition and the social exclusion condition is given in Figure 5.

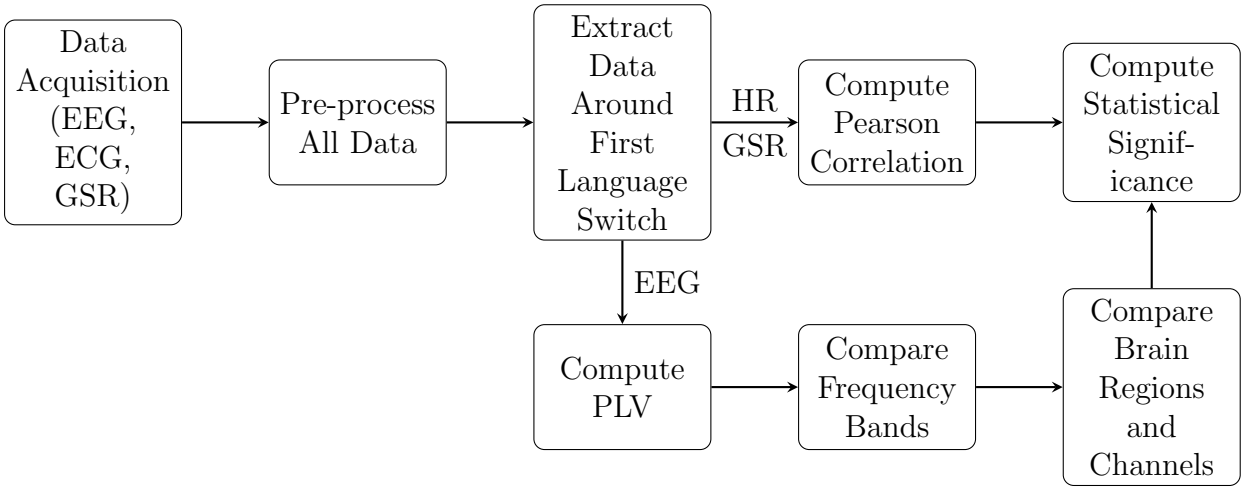


Figure 5: *Global graphical overview of methodology*

The remainder of this chapter will go into detail about the experimental setup, the participants, the recording and synchronisation of data, and the (pre-)processing and analysis of the acquired data as specified in the graphical overview (see Figure 5).

4.1 Experimental Setup

In order to determine how social exclusion/inclusion induced by a language barrier reflects in EEG, HR, and GSR data, a cooperation task based experiment has been conducted. For this experiment, the research by Szymanski et al. (2017) has been extended to groups of three participants instead of two participants. However, some additional changes were made to be able to answer the research question proposed in the Introduction. These changes will be explained in more detail.

The basic setup of the experiment is as follows: three participants have formed a group and worked together as a team (team coordination) to solve seven small riddles (the reasoning behind this will be explained shortly). Two of the participants in the group were able to speak a language that the third participant did not understand. These two participants received an information brochure with all the details about the experiment, while the third participant (who was socially excluded during the experiment) received an information brochure which stated that the experiment was merely about measuring team coordination. The two participants who spoke a different language were instructed that when they communicate with each other, they only do so in their mother tongue language that the other participant does not understand. They were tasked with trying to ignore the other participant as much as possible while solving the riddles to induce a feeling of social exclusion. If the socially excluded participant called them out on their behaviour, the two other participants were instructed to acknowledge that, say sorry, switch back to English for a little while and then slowly continue as they have been doing before. This, thus, created a language barrier between the participants.

Therefore, instead of having one condition in which the task is performed together in a team of two with joined attention and another condition in which the task is performed alone (as in Szymanski et al. (2017)), the first condition is now the part until the first language switch during which the three participants jointly attend to the task, and the second condition is the part in which the third participant was being excluded (as much as possible) and working alone. This new first condition is thus the social inclusion, and the second condition is when the third participant is socially excluded.

An extra control group in which the third participant is not ignored was not needed, because the data before the first language switch can be used as data of a control group. This control condition is used as a baseline and will be subtracted from the data after the first language switch (the moment the two participants had switched to their mother tongue) to get rid of individual bias and make the data more meaningful. The researcher has manually pushed a button to insert a marker (thus splitting the data for the control condition and experiment condition) at the time of the switch. The participants were also video recorded during the experiment. These recordings are used to confirm when exactly this switch happens.

Solving riddles has been chosen as the cooperation task (instead of the visual search task in Szymanski et al. (2017)) for two main reasons: (1) they do not cause much movement, thus lowering the chance of artefacts, and (2) it is a language-oriented task, forcing the participants to actually communicate and making it more justifiable why two of the participants of the team talk in a different language than the third. During the experiment, the participants had a time limit (of 15 minutes) to solve these riddles to make sure that the experiment did not go on indefinitely. After all, participants might have become frustrated at some point if they could not figure out the riddles which could have taken away from the actual purpose of the experiment: measuring social exclusion. However, the most important reason for upholding this time limit is that the excluded participant might feel uncomfortable during the experiment because he/she is ignored. A time limit of 15 minutes has been chosen because the conducted pilot experiment showed that the participants took around 10 minutes to solve the riddles. To account for the fact that some participants might be a little slower than the pilot group, the time limit of the experiment was set to 15 minutes. This time limit ensured that the discomfort participants might have felt did not become too much. To further ensure the comfort of the participants, other measures had also been taken (see Appendix A for details).

Participants were asked to solve the following riddles[1]:

1. What comes down, but never goes up?
2. I'm tall when I'm young and I'm short when I'm old. What am I?
3. What starts with the letter "t", is filled with "t" and ends in "t"?
4. What occurs once in a minute, twice in a moment and never in one thousand years?
5. What is so delicate that saying its name breaks it?
6. What tastes better than it smells?
7. What goes up and never comes down?

[1]adapted from `https://grouptravelleader.com/articles/group-game-10-riddle-challenges/`

Seating formations also play an important role in whether someone feels included or excluded during a conversation. F-formations, for example, are defined as a spatial organisation of people gathered for conversation in which each member has an equal ability to sense all other members (Zhang and Hung, 2016). Thus, for someone to feel excluded/ignored, a group formation was needed in which the socially excluded participant could sense all other members, but the other members were not able to sense the excluded participant. The participants were, therefore, seated in a U-formation at 1.5m distance from each other (as per the Covid-19 protocol). In this U-formation, the two participants who could speak the same language were seated at the ends of the U, facing each other. The third participant who was verbally ignored was seated at the belly of the U, facing the other two, while not being faced themselves. In practice, the U-formation is very similar to a circle. However, the two participants that speak the same language faced each other while they were discussing the riddles and, thus, turned their side toward the socially excluded participant (see Figure 6).

Figure 6: *Seating formation during the experiment. In this image, the socially excluded participant is seated in the chair to the far left side (near the window).*

Similar to Szymanski et al. (2017), the participants were wearing an EEG cap to measure neural activity while they were solving the riddles. All channels have been recorded because this study wants to compare the level of synchronisation across the different regions of the brain (see Figure 7). This data can then be used to determine the level of brain-to-brain synchrony that occurred (see Section 4.4). Furthermore, in addition to the research by Szymanski et al. (2017), the participants were also wearing a wearable HR and GSR sensor. This data will be used to determine whether the emotions the socially excluded participant felt also differed from the other two (see Section 5).

After the experiment, a debriefing took place in which it was explained that the two participants were tasked with ignoring the third participant and why. Furthermore, during this debriefing, the socially excluded participant was asked to give post-experiment approval for the usage of their recorded data. After all, the socially excluded participants were slightly deceived about the whole context of the experiment, making it necessary to ask for approval to use their data again. Their anonymised data can be used in this study as well as further research into this topic. All participants were also asked to fill

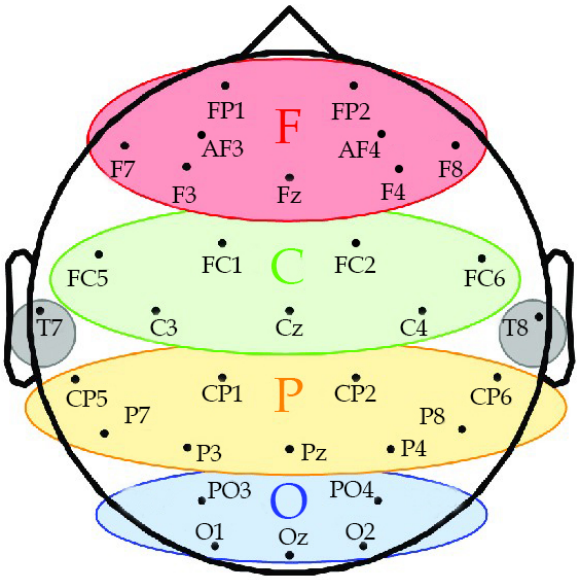


Figure 7: *Schematic of the five brain regions for a 32-channel EEG cap. The nine channels in the red ellipse constitute the frontal region (F). The seven channels in the green ellipse are divided into the central region (C). The nine channels in the yellow ellipse represent the parietal region (P). The seven channels in the blue ellipse comprise the occipital region (O). T3 and T4 in the grey circles are classified as the temporal region (adapted from Liang et al. (2018)).*

out a small questionnaire. The first two participants were given a questionnaire about whether they felt a better connection to their partner in crime compared to the socially excluded participant and about how they thought that participant was feeling[2]. The socially excluded participant was given a questionnaire about whether they indeed felt ignored/excluded[3]. These participants have also been asked when exactly they started feeling excluded to confirm the assumption that the feeling of exclusion starts after the first language switch. This has been done to form a baseline (social inclusion condition) for the comparison of the EEG and other neurophysiological signals. Lastly, the socially excluded participant was asked if they experienced something similar in their life before in which they were excluded because of a language barrier.

This experiment has been approved by the Ethics Committee of the faculty of Electrical Engineering, Mathematics and Computer Science (EEMCS) at the University of Twente. During the experiments, all Covid-19 related measures were taken into account by both the researchers and the participants.

[2] `https://forms.gle/TMnvwSKtxk4boGuUA`
[3] `https://forms.gle/AeqZm3UwwvsoJF3Q6`

4.2 Participants

To form the groups of three participants in which two of the participants speak another language than the third (which the third participant also does not understand by hearing it), a pre-experiment questionnaire[4] about the language(s) people can speak and understand has been conducted before making the actual experiment groups. After all, someone can only feel linguistically ostracised (socially excluded) if they do not understand the language the other two participants are speaking. Thus, to ensure that one of the participants would not understand the language the other two participants could speak, the pre-experiment questionnaire was needed to create experiment groups in which social exclusion could actually occur. A total of 16 experiments have been conducted, comprising of 48 participants in total. From these 48 participants, 21 were male and 27 were female. All participants were either students of the University of Twente or PhD students. Unfortunately, the data from one participant from each of the first three experiments was lost because of a firewall issue. Thus, a total of 13 experiments, consisting of 39 participants (16 male, 23 female) was left (see Table 1).

In this experiment, the researchers were, of course, bound by the available Nationalities. Since the University of Twente is located in the Netherlands, near the border with Germany, more than half of the participants are either Dutch or German (see Table 1). Nevertheless, the researchers have striven to compose teams of as many different combinations of nationalities as possible (see Table 2). This way the results of this research will be applicable across many nationalities, and not just here in the Netherlands. Participation was completely voluntary and all participants signed a consent form describing the detailed experimental procedure. However, the participants which were to be socially excluded received a slightly different consent form in which they were not informed that the experiment was about measuring social exclusion induced by a language barrier. Instead, they signed a consent form thinking that the experiment was just about measuring team coordination. After a debriefing of the actual goal of the experiment, these socially excluded participants signed a post-experiment approval form, allowing the use of their data.

[4] `https://forms.gle/iREE5e5LZ8FSDduP7`

Nationality	**Male**	**Female**	**Total**
Dutch	4	7	11
German	2	8	10
Italian	2	1	3
Romanian	2	1	3
Indian	2	0	2
Spanish	1	1	2
French	1	0	1
Irish	1	0	1
Zimbabwean	1	0	1
Finnish	0	1	1
Ukrainian	0	1	1
Chinese	0	1	1
Cameroonian	0	1	1
Singaporean	0	1	1
Total	**16**	**23**	**39**

Table 1: *Nationalities of the participants*

Excluded Participant	**Included Participants**	**Language Spoken**
German	Dutch	Dutch
German	Dutch, Dutch+German	Dutch
Romanian	Dutch	Dutch
Irish	Dutch, German	Dutch
Cameroonian	Dutch	Dutch
French	German	German
Finnish	German	German
Singaporean	German	German
Italian	German, Dutch	German
Dutch	Spanish	Spanish
Ukrainian	Italian	Italian
Chinese	Romanian	Romanian
Zimbabwean	Indian	Indian

Table 2: *Nationality combinations and languages spoken during experiments (Note that one participant had a double nationality and that two participants spoke in a another language then expected based on their nationality. Nevertheless, these two participants spoke the language they used in the experiment fluently).*

4.3 Recording and Synchronisation of Data

The Biosemi EEG system was used to record electrical activities for each participant with a sampling rate of 2,048 Hz. 32 EEG electrodes were attached over the whole scalp (see Figure 7). The Common Mode Sense (CMS) and Driven Right Leg (DRL) electrodes were used for reference which drove the average potential of the participant. Furthermore, two ECG electrodes were attached to the participant, one on the left collarbone and the other on the right lower leg.

The EdaMove 4 system was used to record GSR data with a sampling rate of 32 Hz. Two electrodes were placed on the palm of the non-dominant hand for each of the participants and the device was placed on the wristband. The non-dominant hand was chosen to minimise movement artefacts.

All EEG and ECG data were synchronised by the Lab Streaming Layer (LSL)[5]. The three Biosemi devices were first connected to their own PC which then streamed the data to the main PC which held the Lab Recorder. A keyboard was also connected to this PC to synchronise pressed markers as well. These markers were used to indicate the start of each new trial (new riddle) as well as each time the participants switched to their mother tongue and back to English. This was done based on the observations from the experimenter. Since the Edamove system is not compatible with the LSL system, synchronisation of the GSR data with the EEG and ECG data has been done offline.

4.4 Pre-processing the Data

Before the analysis of the data was possible, pre-processing of the data was needed, especially for the EEG data. To do that, the recorded data needed to be distributed to the correct participant. Prior to the experiments, it had been decided to distribute the data to the participants according to the seat arrangement. The socially excluded participant was always directed to be seated at the same spot (with the same equipment), making it easier to recognise which data belonged to them after extracting it from the LSL recording (far left in Figure 6). Participant 1 and 2 were also always seated in the same spot for each experiment (far right and front for participant 1 and 2, respectively in Figure 6). However, they were not directed to their seats by the researcher and could choose which of those two chairs they preferred. This section will go into detail about the pre-processing steps that have been taken for the EEG, ECG, and GSR data.

[5] `https://labstreaminglayer.readthedocs.io/index.html`

4.4.1 Electroencephalogram (EEG) data

Pre-processing of the EEG data has been done with MATLAB's EEGLAB (Delorme and Makeig, 2004) toolbox and according to an adaptation of Miyakoshi's Pre-processing Pipeline (Miyakoshi, 2020). The following overview shows the pre-processing steps that have been taken. Each of these steps will be explained in further detail below:

1. Remove baseline
2. High- and low-pass filter the data
3. Remove bad channels
4. Interpolate all removed channels
5. Re-reference to the average
6. Run Independent Component Analysis (ICA) and remove noisy components
7. Correct noisy data using Artifact Subspace Reconstruction (ASR)

Step 1: Remove baseline
The DC offset can introduce large filter artefacts at the beginning and end of the signal. Therefore, it is important that, prior to applying any filtering to the signal, the DC offset is removed (Delorme, 2020). This is done by removing the baseline; the mean of the complete trial.

Step 2: High- and low-pass filter the data
The next step is to high- and low-pass filter the data. Using a high-pass filter removes the 'baseline-drift' in the data. A 1 Hz edge for the high-pass filter was used for two main reasons:

- If the data is finite (as is the case), ICA is biased toward high amplitude.
- The EEG signal below 1 Hz could be contaminated by sweating etc. which affects ICA.

At the same time, a low-pass butterworth filter with a cutoff frequency at 45 Hz was also applied. This cutoff frequency was chosen because all important EEG frequency bands (delta, theta, alpha, beta, gamma) fall within this 45 Hz range. Any frequency component above this can, thus, be removed without any loss of information (Hasan et al., 2014).

Step 3: Remove bad channels
Removing the bad channels is a very important step when re-referencing to the average (as was done in step 5). After all, the average reference is an average of all signals. Thus, if you have noisy channels, including these channels in the average will introduce noise to all channels (Miyakoshi, 2020).

For this step, automatic channel rejection in EEGLAB has been used. In total, 2-4 channels have been removed from the data of each of the participants. This was based on either the Kurtosis (24 participants) or the Spectrum (15 participants) method. The Kurtosis method is well suited for flat-line channels or channels with strong power-line noise, while the Spectrum method is well suited for data with strong movement. The threshold of the Kurtosis method ranged between a cutoff standard deviation of 2 and 5 (with one outlier at 9) and the threshold of the Spectrum method ranged between a

cutoff standard deviation of 0.8 and 2. Channels were thus rejected if they had a higher standard deviation than the specified threshold.

Step 4: Interpolate all removed channels
The next step is to interpolate all removed channels using the build-in function in EEGLAB with the Spherical interpolation method. The adjacent channels (e.g. the two neighbouring channels of the removed channels) are used to interpolate these missing channels. This is done to minimise a potential bias in the next step in which the channels will be re-referenced to the average. After all, if, for example, all rejected bad channels are only from the left hemisphere, the average will be biased toward the right hemisphere. Interpolation of the channels ensured that there will be an equal number of channels in both hemispheres again and, thus, prevented this bias from happening (Miyakoshi, 2020).

Step 5: Re-reference to the average
There are multiple ways to re-reference the data, but in this step, the average reference is used. This referencing method is an approximation of the scalp potentials that is independent of the reference location. It assumes that the positive and negative potential changes balance each other out, which means that the scalp topography should sum to zero. Average referencing enforces this. The average reference is also helpful for suppressing line noise (Miyakoshi, 2020).

Step 6: Run ICA and remove noisy components
Running ICA on the 32-channel data resulted in 32 noisy components. Around 8 to 18 of these components were rejected for each participant. This component rejection was done based on the ICLabel (Pion-Tonachini et al., 2019) toolbox from MATLAB.

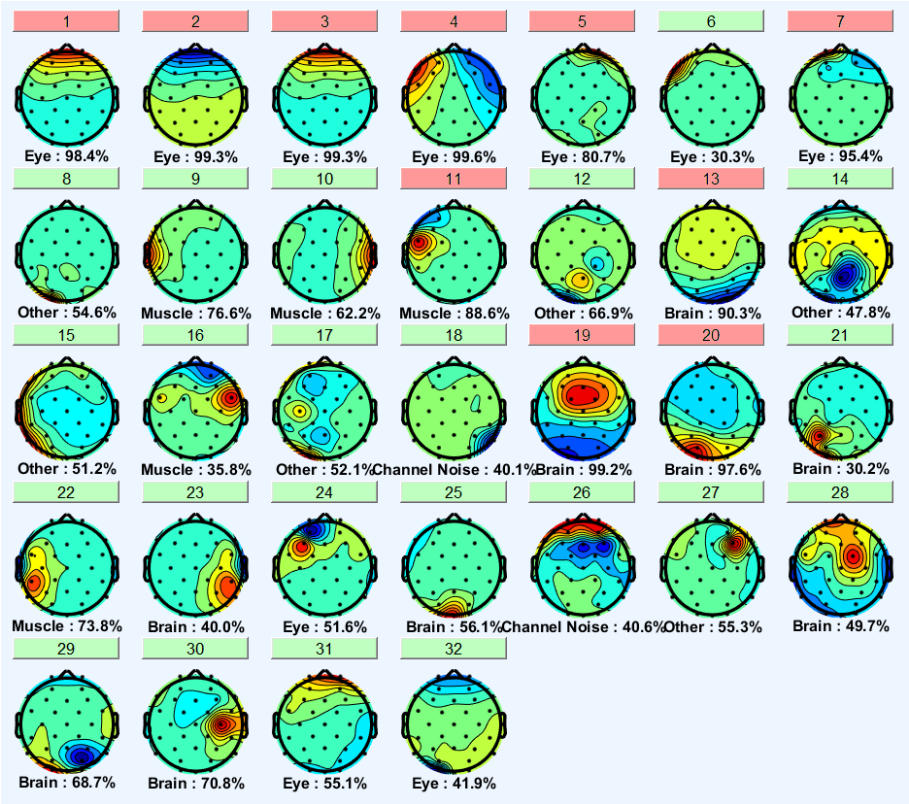


Figure 8: *Example of data after running ICA and letting ICLabel determine which components to reject (red components)*

There are seven different types of noisy components: brain ICs, muscle ICs, eye ICs, heart ICs, line noise ICs, channel noise ICs, and other ICs. ICLabel rejected the components if they were classified to contain more than 80% noise. Most components which were removed were either muscle or eye ICs since other types of noise were already mostly removed in the previous pre-processing steps. Figure 8 shows an example of components which have been rejected by ICLabel. Furthermore, Figure 9 shows example data before ICA component rejection while 10 shows a much cleaner signal after ICA component rejection.

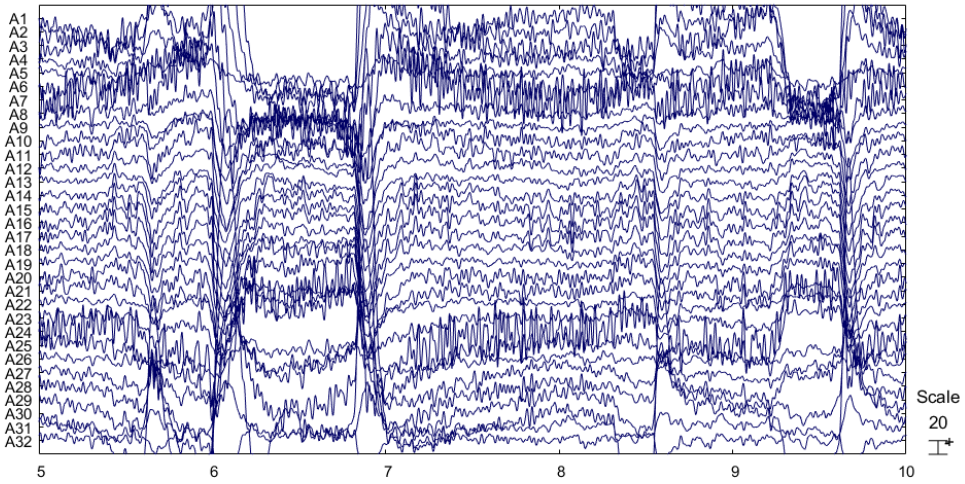


Figure 9: *Example of data before ICA component rejection*

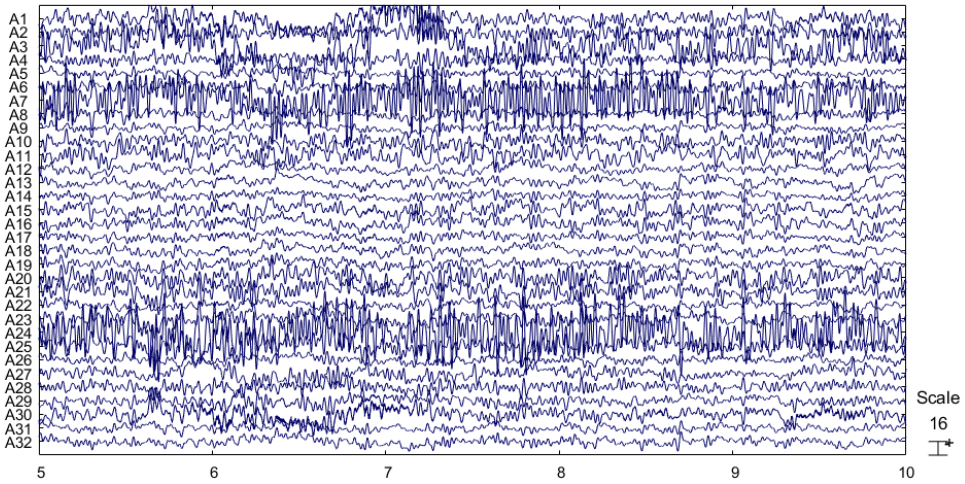


Figure 10: *Example of data after ICA component rejection*

Step 7: Correct noisy continuous data using ASR
This last step is done because the data was still too noisy in some portions of the data after the previous six steps. There were some short bursts of artefacts across the data which needed to be removed to improve the analysis results. These noisy bursts have likely been induced because of the re-referencing step (step 5). After all, small bursts of noise in one single channel will now have been introduced in all other channels as well because the average of all channels is used. One way get rid of this noise is by correcting these noisy bursts using Artifact Subspace Reconstruction (ASR).

Miyakoshi (2020) suggested to use a standard deviation of 10-20 based on unpublished research done by Nima (from the PREP pipeline (Bigdely-Shamlo et al., 2015)). After testing, a standard deviation of 20 has been used. The bad portions of data (which were above the threshold of the standard deviation of 20) were then corrected instead of removed to make it easier to keep the data alignment between the participants (this is simply a feature which can be selected when using the ASR function in EEGLAB). Figure 11 shows an example of how ASR corrected noisy data.

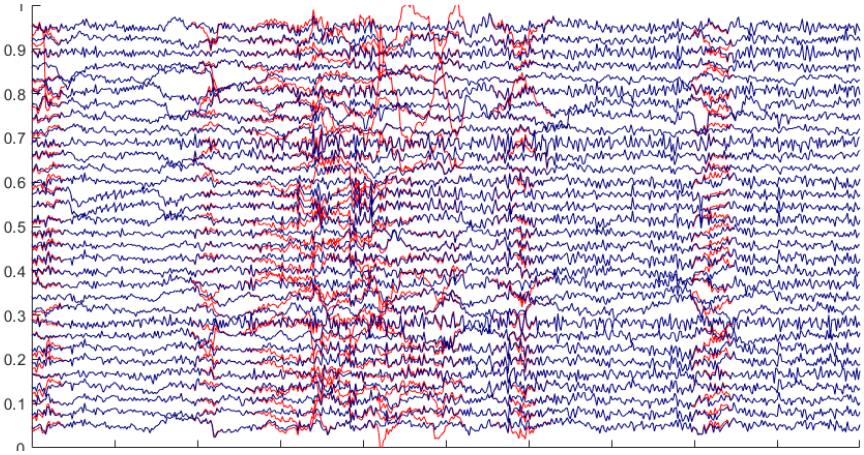


Figure 11: *Example of data after data correction using ASR. Red shows the noisy data while blue shows the corrected data.*

4.4.2 Electrocardiogram (ECG) data

The first step for pre-processing the ECG data, is to acquire the HR data. This is done by subtracting the two recorded ECG channels from each other.

Then, the HR data was pre-processed by detecting the noisy data based on a maximum slope threshold (10000). The indices of this maximum slope and the signal portions surrounding this high slope (1 s before and 1 s after the slope) were then replaced with a NaN value. Lastly, these NaN values were then replaced again by linear interpolation of the neighbouring data (the first data point before and after the missing value). If an endpoint was missing, it was filled by the same value as the nearest non-missing value (Cao et al., 2020).

4.4.3 Galvanic Skin Response (GSR) data

The GSR data is first pre-processed in a similar way to the ECG data. Just like the ECG pre-processing, noisy data is detected based on a maximum threshold (1 uS). The bad signal indices which are below that threshold (and the surrounding area of 2 s before and 2 s after) are replaced with NaN values, and these values were then replaced by linearly interpolating the nearest data. The noisy data is removed because these noisy portions mean a drop of the GSR signal to an ultra-low level, which can be an indicator of connection loss of the electrode (Bakker et al., 2011).

Additionally, Savitzky-Golay filtering was applied with a cubic polynomial order and a window length of 1 s to smooth the data (see Figure 12) (Savitzky and Golay, 1964). The Savitzky-Golay filter was chosen because it is appropriate for preserving the amplitude of the original curve while still smoothing the small noisy signals caused by the quantisation error of the EdaMove 4 (Thammasan et al., 2020).

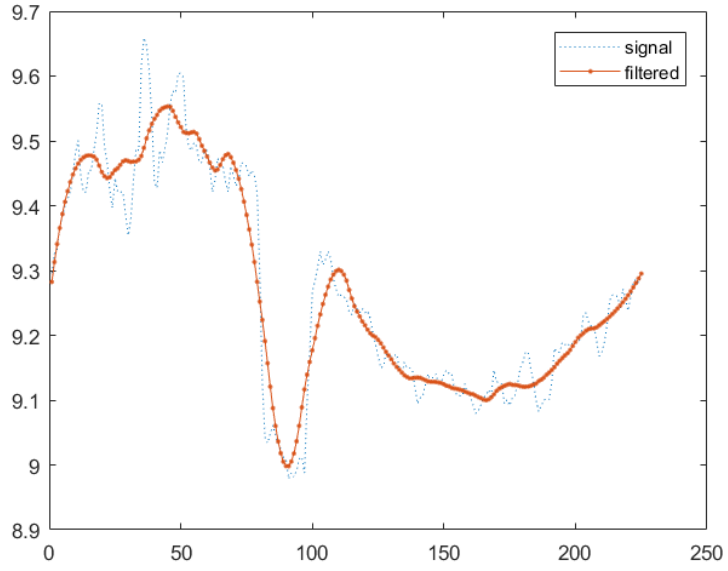


Figure 12: *Example of GSR data before (blue) and after applying Savitzky-Golay filter (red)*

Furthermore, the GSR signal is composed of two components: the 'tonic' (fast-changing) component and the 'phasic' (slow changing) component. The 'phasic' component is the most relevant to the psychological factors (feeling exclusion) of interest, while the 'tonic' component might include noise which influences the signal (Thammasan et al., 2020). After all, the 'tonic' component can change from 0 uS to 20 uS in a short time, but the 'phasic' component is often in the range of 0-5 uS. (Dehzangi et al., 2018). Therefore, the 'phasic' component has been extracted from the raw GSR signal using the Ledalab (version 3.4.9) toolbox (Benedek and Kaernbach, 2010). See Figure 13. As can be seen, the data is smoother and much less noisy compared to the filtered data in Figure 12 which included both the 'phasic' and 'tonic' component.

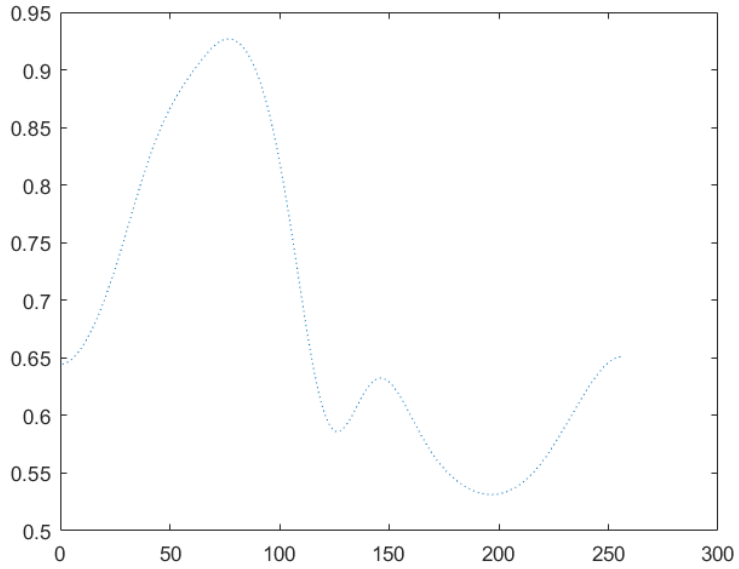


Figure 13: *Example of only the 'phasic' component of the GSR data*

4.5 Analysis

To analyse the data, the level of synchronisation between the participants first needs to be computed. For the EEG data, this will be done by calculating the Phase-Locking Value (PLV) and for both the ECG and GSR data, this will be done based on calculating the Pearson Correlation Coefficient (PCC). Both these measures and the reasoning for choosing them will be explained in further detail below. Afterwards, the statistical analysis measure which has been used to evaluate the results will be explained as well.

4.5.1 Compute level of Synchronisation

4.5.1.1 Phase-Locking Value (EEG data)

To analyse the EEG data, the Phase-Locking Value (PLV) (see equation 1) was used since this measure is the best suited for capturing the rapid flow of information which exists during social interactions such as solving riddles with a team (Burgess, 2013). It is expected that the socially excluded participant will have a lower synchronisation with both other participants (lower PLV value) and these two participants will have a higher level of synchronisation with each other (higher PLV value). Thus, to test this hypothesis, the PLV values need to be computed. This has been done in MATLAB in several steps which will be explained below.

Step 1: Extract data
The data was extracted based on the first language switch. This was done based on the first time the "M" was pressed on the keyboard, since the keyboard was synchronised with the EEG (and ECG) data through the LSL system.

The 3 seconds before the language switch were used as a baseline and the 5 seconds after the language switch was seen as the actual experiment. The first language switch has been chosen because most socially excluded participants indicated in the post-questionnaire that they started to feel excluded after the other two participants switched to their mother

tongue. It is assumed that this feeling is strongest right after the first time this happens so this should give the greatest difference in PLV value.

Step 2: Create pseudo-trials
The PLV computation only works if there are multiple trials. After all, the PLV measure looks at whether the phase difference varies across trials (if it varies little the PLV is close to 1) (Lachaux et al., 1999). Therefore, this cannot be done if the data only consists of one single trial. Since the extracted data is very short and only part of one trial, this calls for the usage of pseudo-trials. These pseudo-trials were created with a window of 1 second and an overlap of 0.5 seconds.

Step 3: Compute PLV
Previous EEG hyperscanning research showed that the most important brain regions for social interactions are the frontal, parietal, central and temporal regions. Furthermore, these studies have shown that frequency bands below 12 Hz (e.g. alpha/theta) are most important for studying synchrony during social interactions (Babiloni et al., 2006; Lindenberger et al., 2009; Mueller et al., 2013; Davidesco et al., 2019a,b; Dumas et al., 2010; Kawasaki et al., 2013).

Therefore, the channels in the 5 different brain regions have been compared to each other to determine if these results can be confirmed. Moreover, the delta, theta, alpha, beta and gamma frequency bands have also been compared to each other to determine if it is indeed the case that the frequency bands below 12 Hz (e.g. delta, theta, alpha) will show the synchrony most clearly. Thus, the PLV has been computed for each of the five brain regions and the different frequency bands (delta, theta, alpha, beta, gamma) by using a band-pass FIR filter for each of the frequency ranges.

The PLV computation[6] has been done with three different methods:

M1. by first calculating the PLV value per channel for each participant combination and then averaging the PLV value for each of the 5 brain regions (see Figure 7).

M2. by first averaging the EEG signal for each of the 5 brain regions and then calculating the PLV per brain region for each participant combination.

M3. by calculating the PLV value per channel for each participant combination.

Step 4: Use only a subset of experiments
Additionally, 3 subsets of experiments have also been investigated to determine if those subsets might improve the results:

1. All experiments in which the socially excluded participant indicated in the questionnaire to have felt "very excluded"
2. All experiments in which the excluded participant indicated in the questionnaire to have felt either "excluded" or "very excluded"
3. All experiments in which the excluded participant indicated in the questionnaire to have felt either "excluded" or "very excluded" AND the other two participants indicated that they felt more connected to each other than to the socially excluded participant (higher team coordination)

[6]code has been adapted from `https://nl.mathworks.com/matlabcentral/fileexchange/31600-phase-locking-value`

These subsets are chosen because when participants feel more excluded, it is assumed that this will also be reflected in the lack of neural synchronisation to a greater degree. Furthermore, when the non-excluded participants indicated that they felt more connected to each other than to the socially excluded participant, it is assumed that they will have a higher level of team coordination and will, thus, show a higher level of neural synchronisation.

Step 5: Correct for baseline before language switch
While in the previous steps, only the PLV value from after the language switch has been investigated, this step will be used to get rid of individual baselines. This should give a clearer view of how the PLV value has changed after the language switch. It is expected that the difference in PLV is greatest for the socially excluded participant and that the PLV values of the other two participants should remain mostly the same before and after the language switch.

Thus, to correct for the baseline, the mean PLV value from the 3 seconds before the language switch has been subtracted from the PLV values from the 5 seconds after the language switch. This should give a greater indication of the differences in synchronisation between the participants.

4.5.1.2 Pearson Correlation (HR and GSR data)

The Pearson Correlation metric (see equation 7) was used to compute the level of synchrony for the HR and GSR data since Correlation provides the most direct method for analysing the synchrony between these two modalities (Reinerman-Jones et al., 2011). The correlation for the data from both HR and GSR has been calculated for each pair within the three participants. High correlation is expected for the two participants who are working together while low correlation with the other two participants is expected for the socially excluded participant.

Heart Rate (HR)
Before the PCC can be computed for the HR, the desired data window first needs to be extracted. Again, this window is from 3 seconds before the first language switch until 5 seconds after the language switch.

Then, the Pan-Tompkins algorithm is applied to the extracted data to detect the peaks and Inter-Beat Interval (IBI) in the data (Pan and Tompkins, 1985). Because time windows of 3 and 5 seconds were chosen, the actual HR was not used, because the time windows are simply too short (usually at least several minutes are needed). Thus, as an alternative, the IBI is used instead. While these methods are not the same, they can both be used to indicate emotions (Ravaja et al., 2006).

Afterwards, this IBI is re-subsampled to 2 Hz, resulting in an interpolated IBI. The idea is that connecting all IBIs in the time series will result in time points that have values of IBIs. This means that the intervals between the time points are unequal. A new time series where time points are equally spaced (in this case every 0.5 seconds) are then generated to calculate the interpolated IBI for a newly-sampled time point. This way, the IBI time-series can be coupled point-wise across participants because they now have the same sampling rate. It is now possible to calculate the PCC across participants.

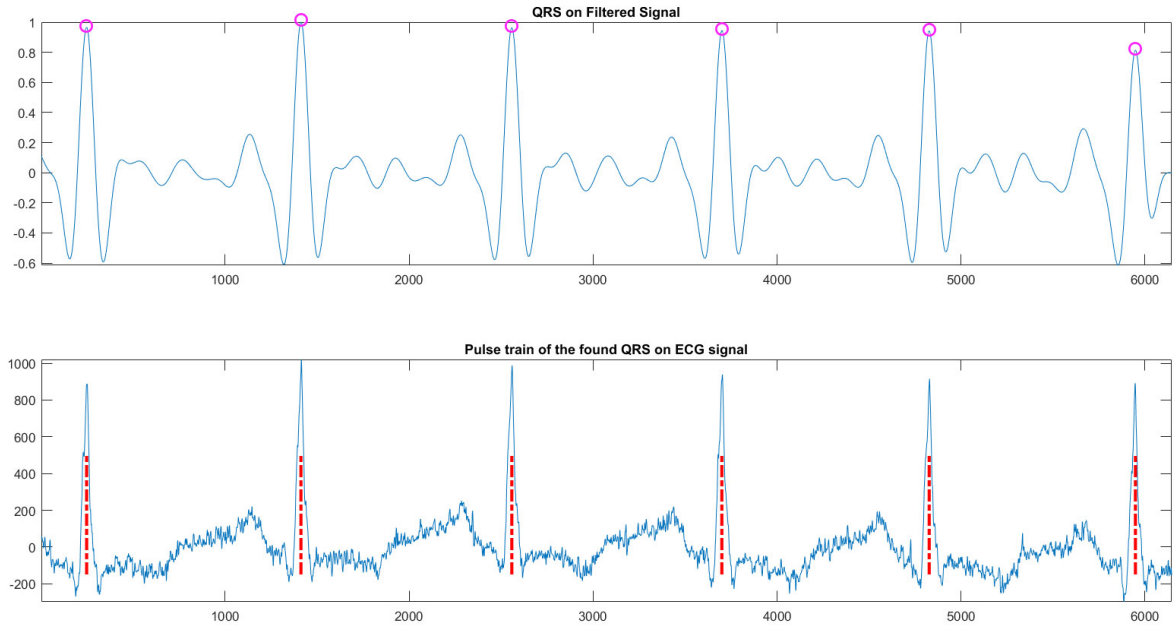


Figure 14: *Example of heart rate peak detection depicted by the purple/pink circles (above) and Inter-Beat Interval detection (below). The IBI is the part between the peaks (red vertical stripes).*

Furthermore, since the number of datapoints is still quite low in a time window of 5 seconds after the language switch, the results of this time window will be compared to a larger time window of 10 seconds as well.

Galvanic Skin Response (GSR)

The only processing that needs to be done before the PCC can be calculated for the GSR data is the extraction of the right part of the data. As has been said before, it was not possible to synchronise the GSR data with the other data through the LSL. Therefore, this step is a little more elaborate for the GSR data. Another problem is that the LSL and GSR data timestamped their data in a completely different manner. The GSR data used a datetime timestamp, while the LSL used something similar to UNIX time, but started counting the time from the moment the PC turned on.

To synchronise the data and find the correct GSR timepoint for the first language switch, the timestamp on which the LSL file was last modified was used. The number of seconds between this last timestamp of the EEG data and the timestamp of the first language switch was then calculated. By subtracting this number of seconds from the datetime timestamp of the last time the LSL file was modified, the timestamp from the language switch was acquired. The correct GSR data was then extracted by finding the closest matching GSR timestamp to the timestamp that was previously found for the language switch and getting the data from 3 seconds before the language switch and 5 seconds after it. However, since the sampling rate of the Edamove device was 32 Hz (for which all data points have an equal timestamp) it is possible that the datapoint indicating the language switch could have been anywhere within this range of 32 data points. This means that the synchronisation between the EEG/ECG data and the GSR data could have a small error in accuracy (with a maximum difference of 1 second).

4.5.2 Evaluation of Results

To evaluate the results and compute the statistical significance, the Permutation Test was used. The Permutation Test is a non-parametric test that makes minimal assumptions about the distribution of the data, unlike commonly used tests like the t-test or ANOVA (Collingridge, 2013). Since the data is not normally distributed, the Permutation Test is, thus, better suited for computing the statistical significance of the differences in synchrony. Another advantage of the Permutation Test is that it can also be used on small sample sizes such as this and that they usually have low false positive rates and high true positive rates (Collingridge, 2013; Good, 2006).

The Permutation Test is a data-driven approach that uses all possible values of the test statistic under random permutations of the data. This is done to obtain the distribution of the test statistic under the null hypothesis (Legendre and Legendre, 1998). The resulting p-value of the test is calculated by computing the proportion of the permutations where the permutation test statistic is larger than the observed test statistic. If this p-value is below the significance level of 0.05, the null hypothesis is rejected. The p-value is calculated with a one-sided Permutation Test since it is assumed that the synchronisation/correlation of the first two participants (sample 1) is higher than the synchronisation/correlation of the excluded participant with both other participants (sample 2). Thus, the p-value is only calculated based on the hypothesis that the mean of sample 1 is larger than the mean of sample 2.

Lastly, when multiple hypothesis tests are executed simultaneously (as is the case, since multiple comparisons of the five different brain regions as well as all 32 channels are conducted), the number of p-values that are below the significance level due to sheer chance increases. Even though the Permutation Test is shown to have a low false-positive rate, it is still possible to have wrongful rejections of the null hypothesis (type I errors) because of the multiple comparisons (Cohen et al., 1997). The False Discovery Rate (FDR) correction is a way to correct for these types of errors, thus ensuring that the potentially significant differences are actually significant and not just by chance (Genovese et al., 2002).

In the next chapter, the PLV values of the different brain regions and frequency bands will be compared to each other for the different methods (M1, M2, M3) which were used. The Permutation Test (with 10000 permutations) was used to compute the p-values of the best performing brain regions, channels and frequency band combinations. The results are statistically significant ($p < 0.05$) if the PLV value from participant 1 and 2 (non-excluded participants) is significantly higher than the PLV values from either of them with the socially excluded participant. As for the HR and GSR data, the PCC was computed and compared per experiment as well as for the dataset overall. The Permutation Test (again with 10000) was used to see if there is a statistically significant ($p < 0.05$) difference in the correlations between participant 1 and 2 (non-excluded participants) compared to each of their correlations with the socially excluded participant.

5 Results

5.1 Post-Experiment Questionnaire

In the post-experiment questionnaire, the socially excluded participants were asked if they actually felt excluded (to determine if the experiment had the desired effect). Figure 15 shows that 4 out of 13 participants (31%) felt very excluded, 7 out of 13 participants felt excluded (54%), and 2 out of 13 felt slightly excluded (15%). None of them indicated that they felt neutral, or not excluded at all, which were the other two options.

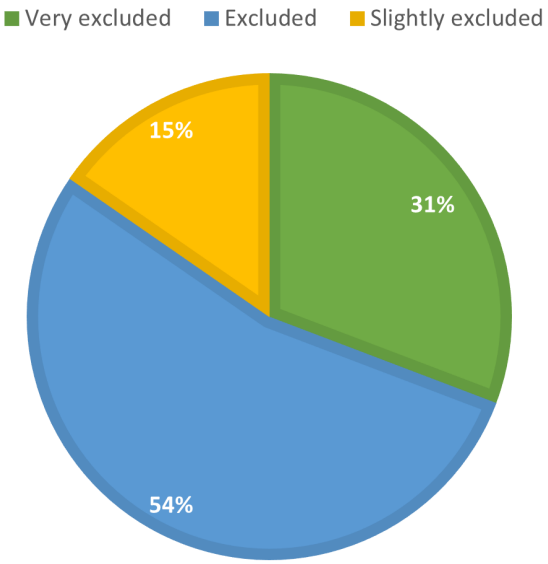


Figure 15: *Pie chart of post-experiment questionnaire results: Did you feel excluded?*

To analyse the results from the EEG, HR, and GSR data, the level of synchronisation between participant 1 and 2 (non-excluded participants) will be compared to the level of synchronisation of both these participants with participant 3 (socially excluded participant).

5.2 Electroencephalogram (EEG) Analysis

The following sections will describe the results from the EEG analysis, both with and without baseline correction. In the next section (without baseline correction), only the PLV values in the 5 seconds after the first language switch are compared. Afterwards, the analysis results when applying baseline correction will be described. To apply baseline correction, the mean PLV of the 3-second window before the first language switch was subtracted from the PLV values in the 5 seconds after the language switch.

5.2.1 Without Baseline Correction

As has been specified in Section 4.5.1.1, the EEG results are analysed both with and without baseline correction. This section will give the results of the analysis on the PLV values without baseline correction below.

The PLV values have been calculated with three different methods (which have been defined in step 3 from Section 4.5.1.1):

- The PLV value is first calculated per channel for each participant combination and then an average PLV value is calculated for each of the five brain regions (according to the schematic in Figure 7) → **Method M1**
- The data is first averaged for each of the five brain regions (according to the schematic in Figure 7) and then the PLV value is calculated for each region → **Method M2**
- The PLV value is calculated per channel for each participant combination → **Method M3**

There are, thus, two methods for the PLV calculation on a regional level and one method for the PLV calculation on a channel level. The two regional methods are discussed and compared first.

5.2.1.1 Regional Level

		Frontal	**Central**
D	1+2	0.3196 ± 0.0185	***0.3230*** ± 0.0108
	1+3	0.3222 ± 0.0193	0.3210 ± 0.0106
	2+3	0.3196 ± 0.0158	0.3178 ± 0.0127
T	1+2	0.3196 ± 0.0185	***0.3230*** ± 0.0108
	1+3	0.3222 ± 0.0193	0.3210 ± 0.0106
	2+3	0.3196 ± 0.0158	0.3178 ± 0.0127
A	1+2	0.3196 ± 0.0185	***0.3230*** ± 0.0108
	1+3	0.3222 ± 0.0193	0.3210 ± 0.0106
	2+3	0.3196 ± 0.0158	0.3178 ± 0.0127
B	1+2	0.3199 ± 0.0176	***0.3222*** ± 0.0103
	1+3	0.3216 ± 0.0186	0.3200 ± 0.0099
	2+3	0.3193 ± 0.0145	0.3170 ± 0.0114
G	1+2	***0.3187*** ± 0.0077	***0.3187*** ± 0.0083
	1+3	0.3167 ± 0.0113	0.3146 ± 0.0078
	2+3	0.3173 ± 0.0074	0.3142 ± 0.0071

Table 3: *Mean and Standard Deviation of PLV values for each frequency band (D=delta, T=theta, A=alpha, B=beta, G=gamma) and participant combination (3=excluded participant). The bold/italic values indicate the frequency band and brain region combinations for which the mean PLV value is higher for participant combination 1+2 (non-excluded participants) than both other combinations.* → ***Method M1***

Tables 3 and 4 show the mean and standard deviations of the PLV values for all frequency bands and participant combinations when using the two regional methods. Table 3 shows the PLV values after using Method M1, while Table 4 shows the results from Method M2. Only the results from the brain regions for which at least one frequency band showed that the participant combination from participant 1 and 2 (non-excluded participants) had a higher PLV value than both other combinations are included. The complete version of Tables 3 and 4 with mean PLV values and standard deviations for all brain regions can be found in B.1.

		Frontal	**Central**	**Temporal**
D	1+2	0.3178 ± 0.0301	0.3251 ± 0.0348	***0.3162*** ± 0.0185
	1+3	0.3218 ± 0.0456	0.3325 ± 0.0340	0.3087 ± 0.0241
	2+3	0.3198 ± 0.0442	0.3133 ± 0.0249	0.3052 ± 0.0255
T	1+2	0.3178 ± 0.0300	0.3250 ± 0.0346	***0.3162*** ± 0.0184
	1+3	0.3218 ± 0.0454	0.3323 ± 0.0337	0.3087 ± 0.0239
	2+3	0.3197 ± 0.0438	0.3135 ± 0.0250	0.3052 ± 0.0256
A	1+2	0.3178 ± 0.0298	0.3248 ± 0.0341	***0.3162*** ± 0.0181
	1+3	0.3218 ± 0.0450	0.3317 ± 0.0330	0.3087 ± 0.0233
	2+3	0.3194 ± 0.0430	0.3141 ± 0.0250	0.3050 ± 0.0256
B	1+2	0.3176 ± 0.0286	0.3238 ± 0.0316	***0.3159*** ± 0.0171
	1+3	0.3221 ± 0.0426	0.3293 ± 0.0304	0.3094 ± 0.0199
	2+3	0.3181 ± 0.0389	0.3168 ± 0.0250	0.3045 ± 0.0261
G	1+2	***0.3244*** ± 0.0160	***0.3223*** ± 0.0226	***0.3184*** ± 0.0207
	1+3	0.3139 ± 0.0210	0.3215 ± 0.0205	0.3161 ± 0.0234
	2+3	0.3161 ± 0.0224	0.3204 ± 0.0252	0.3012 ± 0.0163

Table 4: *Mean and Standard Deviation of PLV values for each frequency band (D=delta, T=theta, A=alpha, B=beta, G=gamma) and participant combination (3=excluded participant). The bold/italic values indicate the frequency band and brain region combinations for which the mean PLV value is higher for participant combination 1+2 (non-excluded participants) than both other combinations.* → ***Method M2***

From Table 3 it can be seen that the PLV value from participant combination 1 and 2 (non-excluded participants) is higher than both other participant combinations for the central brain region for all frequency bands (indicated in bold/italic). In the gamma band, the PLV value from participant combination 1 and 2 is higher than both other participant combinations for the frontal region as well.

Table 4 shows that the PLV value from participant combination 1 and 2 is higher than both other participant combinations for the temporal region for all frequency bands. In the gamma band, this is also the case for the frontal and central region. With Method M1, the central region, thus, seems to be the most important region, while Method M2 indicates that the temporal region is the most relevant.

Furthermore, from both Tables 3 and 4 it can be seen that Method M1 has a much lower standard deviation than Method M2, suggesting the robustness of M1 compared to M2. Since Method M1 appears to be the best-suited method for the calculation of the PLV value on a regional level (because of its low standard deviation), all following analyses results on the regional level will be given using this method.

Another thing which should be noted from Tables 3 and 4 is that the overall mean PLV value of all frequency bands and brain regions lies around +/- 0.32. Given that the PLV value can range from 0 to 1, this value is quite low.

Freq band	Combination	Region	p-value
D	1+2/1+3	Central	0.322
	1+2/2+3		0.135
T	1+2/1+3	Central	0.312
	1+2/2+3		0.132
A	1+2/1+3	Central	0.321
	1+2/2+3		0.131
B	1+2/1+3	Central	0.300
	1+2/2+3		0.121
G	1+2/1+3	***Central***	***0.101***
	1+2/2+3		***0.080***
	1+2/1+3	Frontal	0.299
	1+2/2+3		0.310

Table 5: *p-values for the frequency bands (D=delta, T=theta, A=alpha, B=beta, G=gamma) and regions for which the mean PLV value from participant combination 1 and 2 (non-excluded participants) is higher than both other combinations (3=excluded participant) using Method M1 (according to Table 3). The bold/italic value indicates the frequency band and brain region combination for which the p-value is closest to being significant for both participant combination comparisons.*

A Permutation Test (with 10000 permutations) was performed on the PLV values acquired with Method M1. With this Permutation Test, the PLV values from participant combination 1 and 2 (non-excluded participants) were compared with the PLV values from both other participant combinations. Table 5 shows the p-values for all frequency bands and brain regions for which the mean PLV value from participant combination 1 and 2 was higher than both other combinations.

As can be seen, none of the combinations of frequency bands and brain regions have resulted in statistically significant differences ($p < 0.05$). However, the frequency band and region for which the p-value is closest to being significant for both comparisons are for the central region with the gamma frequency band (indicated in bold/italic). These comparisons give p-values of $p = 0.101$ and $p = 0.080$ for the comparison of participant combination 1 and 2 with combination 1 and 3, and the combination of 1 and 2 with 2 and 3, respectively. See Appendix B.1 for an overview of the p-values from both Method M1 and Method M2.

A boxplot of this combination of Method M1 with the gamma frequency band is shown in Figure 16. Note that the central region shows that the median PLV value of participant combination 1 and 2 (non-excluded participants) is higher than both other combinations and that this region appears to have just a small variation for all three combinations as well. See Appendix B.1 for all other combinations of method (M1 and M2) and frequency bands.

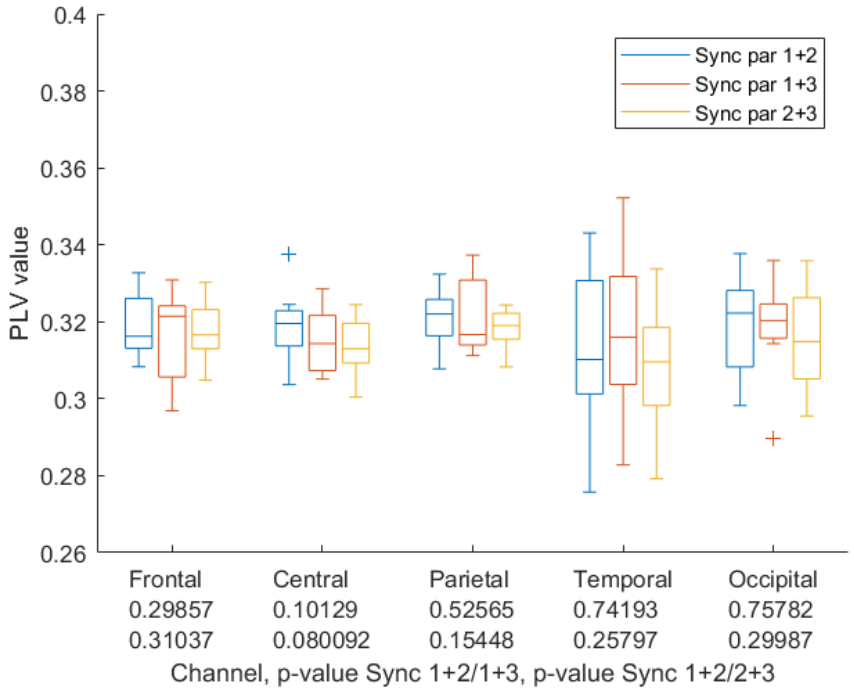


Figure 16: *Method M1: Gamma frequency band. Best performing combination for central brain region.*

5.2.1.2 Channel Level

The PLV values at channel level (Method M3) were also investigated and compared for each of the five frequency bands. For these comparisons, the Permutation Test was once again used to determine whether the differences between the participant combinations were statistically significant.

Table 6 shows the channels which had at least one significant value ($p < 0.05$) after comparing the PLV value from participant combination 1 and 2 (non-excluded participants) with either participant combination 1 and 3, or 2 and 3. Only those channels for which the p-value from the other comparison was at least below 0.25 have been included in the table (see Appendix B.2 for the table with the complete list of all channels for which at least one p-value was significant).

In Table 6, it can be seen that only channel 23 (C4) is statistically significant when comparing participant combination 1 and 2 with both other combinations. This is the case for all frequency bands except the gamma frequency band. Both the delta and theta frequency bands have the lowest p-values ($p = 0.017$, $p = 0.022$ and $p = 0.019$, $p = 0.020$, respectively) and have, thus, performed best. These results are in contrast with the results obtained at the regional level, where the best performing frequency band was the gamma band. Nevertheless, both analyses at the regional level as well as at channel level show that the central region has the best results. After all, channel C4 is also located in the central region (see Figure 7).

Freq band	Combination	Channel	p-value
D	1+2/1+3	21 (CP6)	0.038
	1+2/2+3		0.215
	1+2/1+3	***23 (C4)***	***0.017***
	1+2/2+3		***0.022***
T	1+2/1+3	21 (CP6)	0.035
	1+2/2+3		0.21728
	1+2/1+3	***23 (C4)***	***0.019***
	1+2/2+3		***0.020***
A	1+2/1+3	21 (CP6)	0.039
	1+2/2+3		0.212
	1+2/1+3	***23 (C4)***	***0.020***
	1+2/2+3		***0.021***
B	1+2/1+3	21 (CP6)	0.046
	1+2/2+3		0.191
	1+2/1+3	***23 (C4)***	***0.020***
	1+2/2+3		***0.022***
G	1+2/1+3	19 (P4)	0.127
	1+2/2+3		0.017
	1+2/1+3	26 (FC2)	0.038
	1+2/2+3		0.058

Table 6: *p-values for the frequency bands (D=delta, T=theta, A=alpha, B=beta, G=gamma) and channels for which the mean PLV value from participant combination 1 and 2 (non-excluded participants) is higher than both other combinations and the other p-value was below 0.25 (3=excluded participant). The bold/italic values indicate the channels for which both comparisons are statistically significant.*

However, after False Discovery Rate (FDR) correction, none of these results remain significant. This is even the case after using just a subset of the cleanest (least noisy) channels to reduce the number of multiple comparisons.

See Appendix B.2 for boxplots from the PLV values from all channels for each of the frequency bands. Each of these figures also contains the p-values per participant combination comparison for each channel.

5.2.1.3 Regional Level: Subsets

As has been explained in Section 4.5.1.1, the effect of using 3 different subsets of experiments have also been investigated to determine if those subsets might improve the results:

1. All experiments in which the excluded participant indicated in the questionnaire to have felt "very excluded": 4 experiments in total (see Figure 15). → **Very Excluded (VE)**
2. All experiments in which the excluded participant indicated in the questionnaire to have felt either "excluded" or "very excluded": 11 experiments in total (see Figure 15). → **Excluded (E)**
3. All experiments in which the excluded participant indicated in the questionnaire to have felt either "excluded" or "very excluded" AND the other two participants indicated that they felt more connected to each other than to the excluded participant (higher team coordination): 6 experiments in total. → **Excluded+Connected (EC)**

For these results, only the best performing combinations from the analysis at regional level (Method M1 with gamma frequency band) were used, as has been shown in Section 5.2.1.1. Once again, only the results are given for the brain regions in which there was at least one subset (or for all experiments) where the PLV value from participant combination 1 and 2 (non-excluded participants) was higher than the PLV value from both other combinations.

Subset		**Frontal**	**Central**	**Parietal**
all	1+2	***0.3187***	***0.3187***	0.3208
	1+3	0.3167	0.3146	0.3210
	2+3	0.3173	0.3142	0.3182
VE	1+2	***0.3239***	***0.3187***	***0.3191***
	1+3	0.3178	0.3179	0.3161
	2+3	0.3187	0.3113	0.3167
E	1+2	***0.3198***	***0.3197***	***0.3219***
	1+3	0.3192	0.3154	0.3219
	2+3	0.3162	0.3143	0.3174
EC	1+2	***0.3233***	0.3161	0.3177
	1+3	0.3216	0.3179	0.3183
	2+3	0.3189	0.3146	0.3180

Table 7: *Mean of PLV values per brain regions for each subset (VE=Very Excluded, E=Excluded, EC=Excluded+Connected) and each participant combination (3=excluded participant). Results are only shown from the gamma frequency band with Method M1. The bold/italic values indicate the frequency band and brain region combinations for which the mean PLV value is higher for participant combination 1+2 (non-excluded participants) than both other combinations.*

Table 7 shows the mean PLV values for the three subsets as well as when all experiments are used. It is shown that for the 'VE' subset (only experiments with participants who felt very excluded), there are more regions for which the mean PLV value of participant combination 1 and 2 (non-excluded participants) is higher than both other combinations:

frontal, central and parietal (instead of just frontal and central). The same holds for the 'E' subset with all experiments in which the socially excluded participant indicated to either feel excluded or very excluded. For the 'EC' subset (experiments in which the socially excluded participant felt (very) excluded and the other two participants felt more connected to each other) the results have not improved, with only the frontal region having a higher mean PLV.

Subset	**Combination**	**Region**	**p-value**
all	1+2/1+3	Frontal	0.299
	1+2/2+3		0.310
	1+2/1+3	Central	0.101
	1+2/2+3		0.080
VE	1+2/1+3	Frontal	0.248
	1+2/2+3		0.156
	1+2/1+3	Central	0.452
	1+2/2+3		0.054
	1+2/1+3	Parietal	0.272
	1+2/2+3		0.324
E	1+2/1+3	Frontal	0.445
	1+2/2+3		0.134
	1+2/1+3	***Central***	***0.109***
	1+2/2+3		***0.052***
	1+2/1+3	Parietal	0.491
	1+2/2+3		0.055
EC	1+2/1+3	Frontal	0.391
	1+2/2+3		0.133

Table 8: *p-values for the gamma frequency band and regions for which the mean PLV value from participant combination 1 and 2 (non-excluded participants) is higher than both other combinations (3=excluded participant). Significance values are shown for each subset (VE=Very Excluded, E=Excluded, EC=Excluded+Connected). The bold/italic value indicate the best performing subset and brain region combination.*

The Permutation Test was performed once more to determine if these differences are significant. Table 8 shows the p-values for the subsets and regions for which participant combination 1 and 2 (non-excluded participants) had a higher mean PLV value compared to the other two combinations. From this table, it can be seen that subset 'E' (all experiments in which the socially excluded participant felt (very) excluded) performed best. In the central region, the results were slightly better than when all experiments were used. However, the p-values still are not statistically significant, although the comparison between participant combination 1 and 2 with 2 and 3 comes very close ($p = 0.109$ and $p = 0.052$). Figure 17 shows the boxplot of subset 'E' with the gamma frequency in which the PLV values were calculated with Method M1. Boxplot figures for the other two subsets can be found in Appendix B.3.

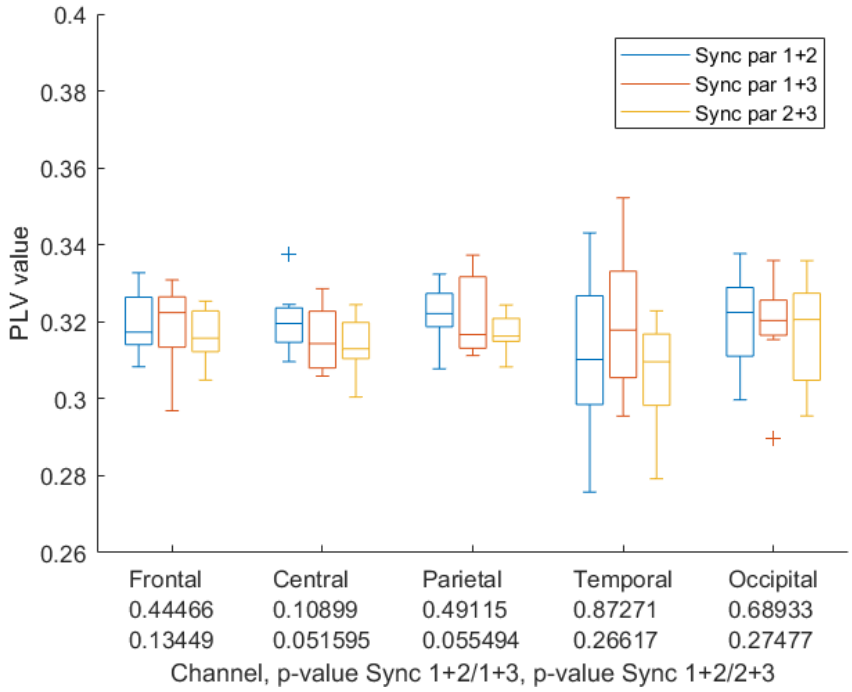


Figure 17: *Subset 'E' (Excluded) with gamma frequency band: all experiments in which the excluded participant actually felt (very) excluded. Best performing combination for central brain region.*

5.2.1.4 Channel Level: Subsets

With these three subsets, the PLV values at channel level were also investigated and compared for the delta and theta frequency bands (the best performing frequency bands at channel level as shown in Section 5.2.1.2). Tables 9 and 10 show the channels which had at least one significant value after comparison with either participant combination 1 and 3, or 2 and 3 for frequency band delta and theta, respectively. Only those channels for which the p-value from the other comparison was at least below 0.25 have been included in the tables (see Appendix B.4 for the table with the complete list of all channels for which at least one p-value was significant).

Tables 9 and 10 show the p-values for the three subsets using the delta and theta frequency band, respectively. It is shown that none of the subsets have a channel with a significant statistical difference for both comparisons, while channel 23 (C4) is significant for when all experiments are used. Nevertheless, channel C4 is still important when looking at the three subsets as well (with channel C4 being almost significant for subset 'E' and 'EC').

Subset	Combination	Channel	p-value
all	1+2/1+3	21 (CP6)	0.038
	1+2/2+3		0.215
	1+2/1+3	***23 (C4)***	***0.017***
	1+2/2+3		***0.022***
VE	1+2/1+3	16 (Oz)	0.204
	1+2/2+3		0.035
	1+2/1+3	17 (O2)	0.015
	1+2/2+3		0.073
	1+2/1+3	23 (C4)	0.216
	1+2/2+3		0.043
E	1+2/1+3	23 (C4)	0.068
	1+2/2+3		0.035
EC	1+2/1+3	23 (C4)	0.077
	1+2/2+3		0.006

Table 9: *p-values for the delta frequency band and channels for which the mean PLV value from participant combination 1 and 2 (non-excluded participants) is higher than both other combinations (3=excluded participant). Only those channels for which the p-value from both comparisons was at least below 0.25 are included. Significance values are shown for each subset (VE=Very Excluded, E=Excluded, EC=Excluded+Connected). The bold/italic values indicate the channels for which both comparisons are statistically significant.*

Subset	Combination	Channel	p-value
all	1+2/1+3	21 (CP6)	0.035
	1+2/2+3		0.217
	1+2/1+3	***23 (C4)***	***0.019***
	1+2/2+3		***0.020***
VE	1+2/1+3	16 (Oz)	0.195
	1+2/2+3		0.034
	1+2/1+3	17 (O2)	0.016
	1+2/2+3		0.070
	1+2/1+3	23 (C4)	0.219
	1+2/2+3		0.044
E	1+2/1+3	23 (C4)	0.070
	1+2/2+3		0.041
EC	1+2/1+3	23 (C4)	0.076
	1+2/2+3		0.006

Table 10: *p-values for the theta frequency band and channels for which the mean PLV value from participant combination 1 and 2 (non-excluded participants) is higher than both other combinations (3=excluded participant). Only those channels for which the p-value from both comparisons was at least below 0.25 are included. Significance values are shown for each subset (VE=Very Excluded, E=Excluded, EC=Excluded+Connected). The bold/italic values indicate the channel for which both comparisons are statistically significant.*

When looking at the number of channels which meet the requirements of one comparison being significant and the other one at least being below 0.25, it can be observed that subset 'VE' has more channels in total for which the requirements are met. Conversely, subset 'E' and 'EC' have fewer channels for which one of the comparisons is statistically different (and $p < 0.25$ for both comparisons). Another thing to note is that, for subset 'VE', the other two channels besides channel C4 which meet the set requirements (channel 16 (Oz) and channel 17 (O2)) are both from the occipital region instead of the central region.

All boxplot figures of these subsets can be found in Appendix B.4.

5.2.2 With Baseline Correction

Now that all EEG results without correcting for the baseline from before the language switch have been explored, the results after baseline correction will be described.

5.2.2.1 Regional Level

Since it was shown that Method M1 has a much lower standard deviation than Method M2 (see Section 5.2.1.1), only the first method was once again used to compare the five brain regions and frequency bands after baseline correction. Table 11 shows the baseline-corrected mean PLV values for the brain regions and frequency bands computed with M1.

		Frontal	**Central**	**Parietal**	**Temporal**	**Occipital**
D	1+2	***-0.1255***	***-0.1221***	***-0.1267***	-0.1313	-0.1223
	1+3	-0.1321	-0.1332	-0.1293	-0.1311	-0.1173
	2+3	-0.1300	-0.1319	-0.1323	-0.1397	-0.1356
T	1+2	***-0.1255***	***-0.1221***	***-0.1267***	-0.1313	-0.1223
	1+3	-0.1320	-0.1332	-0.1293	-0.1311	-0.1173
	2+3	-0.1300	-0.1319	-0.1322	-0.1397	-0.1356
A	1+2	***-0.1255***	***-0.1220***	***-0.1266***	-0.1313	-0.1223
	1+3	-0.1320	-0.1331	-0.1293	-0.1310	-0.1173
	2+3	-0.1300	-0.1318	-0.1322	-0.1397	-0.1355
B	1+2	***-0.1250***	***-0.1227***	***-0.1260***	***-0.1321***	-0.1220
	1+3	-0.1325	-0.1341	-0.1297	-0.1328	-0.1185
	2+3	-0.1302	-0.1325	-0.1323	-0.1391	-0.1359
G	1+2	***-0.1270***	***-0.1271***	***-0.1249***	***-0.1335***	***-0.1272***
	1+3	-0.1353	-0.1374	-0.1310	-0.1347	-0.1283
	2+3	-0.1329	-0.1360	-0.1320	-0.1423	-0.1342

Table 11: *Mean of baseline corrected PLV values for each frequency band (D=delta, T=theta, A=alpha, B=beta, G=gamma) and participant combination (3=excluded participant). The bold/italic values indicate the frequency band and brain region combinations for which the mean PLV value is higher for participant combination 1+2 (non-excluded participants) than both other combinations.* → ***Method M1****, with baseline correction.*

As can be seen in Table 11, all mean PLV values are slightly negative. This shows that the PLV value was higher before the language switch, for all participant combinations. Furthermore, the PLV value from participant combination 1 and 2 (non-excluded participants) is higher than both other participant combinations for the frontal, central, and parietal brain regions for all frequency bands. In the beta and gamma band, the PLV value from participant combination 1 and 2 is higher than both other participant combinations for the temporal region as well. Lastly, for the gamma band, this is also the case for the occipital region.

Table 12 shows the p-values for all frequency band and brain region combinations that had a higher PLV value for participant combination 1 and 2 (non-excluded participants) compared to the other two combinations. These p-values were again computed using the Permutation test with 10000 permutations.

From Table 12 it can be seen that the comparison for all participant combinations is significantly different ($p < 0.05$) in the central region for both the beta and gamma frequency bands. For the gamma band, this statistical significance is the greatest. These results are in line with the results from the regional analysis without baseline correction (see Section 5.2.1.1). However, in contrast to the results achieved without baseline correction, these results are statistically significant. In the central region, these comparisons give p-values of $p = 0.006$ and $p = 0.010$ for the comparison of participant combination 1 and 2 (non-excluded participants) with 1 and 3 and the combination 1 and 2 with 2 and 3 (each of the non-excluded participants with the socially excluded participant), respectively. Even though the delta, theta and alpha frequency bands are not statistically significant for both comparisons in the central region, they are almost significant (one of the p-values even is significant). This is in great contrast for the p-values for the other regions, showing that the central region has the best results overall.

However, after applying the False Discovery Rate (FDR) correction, none of these results remain significant. This is even the case after using a subset of the regions (the occipital region is discarded since it is the least important according to the literature) to reduce the number of multiple comparisons.

Figure 18 shows the boxplot of the gamma frequency band with all brain regions. See Appendix B.5 for boxplot figures of all other combinations as well.

Freq band	**Combination**	**Region**	**p-value**
D	1+2/1+3	Frontal	0.267
	1+2/2+3		0.322
	1+2/1+3	Central	0.044
	1+2/2+3		0.058
	1+2/1+3	Parietal	0.306
	1+2/2+3		0.175
T	1+2/1+3	Frontal	0.271
	1+2/2+3		0.320
	1+2/1+3	Central	0.049
	1+2/2+3		0.057
	1+2/1+3	Parietal	0.306
	1+2/2+3		0.167
A	1+2/1+3	Frontal	0.261
	1+2/2+3		0.319
	1+2/1+3	Central	0.047
	1+2/2+3		0.059
	1+2/1+3	Parietal	0.313
	1+2/2+3		0.171
B	1+2/1+3	Frontal	0.221
	1+2/2+3		0.278
	1+2/1+3	***Central***	***0.034***
	1+2/2+3		***0.047***
	1+2/1+3	Parietal	0.250
	1+2/2+3		0.128
	1+2/1+3	Temporal	0.480
	1+2/2+3		0.141
G	1+2/1+3	Frontal	0.069
	1+2/2+3		0.101
	1+2/1+3	***Central***	***0.006***
	1+2/2+3		***0.010***
	1+2/1+3	Parietal	0.102
	1+2/2+3		0.041
	1+2/1+3	Temporal	0.434
	1+2/2+3		0.105
	1+2/1+3	Occipital	0.439
	1+2/2+3		0.082

Table 12: *p-values for the frequency bands (D=delta, T=theta, A=alpha, B=beta, G=gamma) and regions for which the mean PLV value from participant combination 1 and 2 (non-excluded participants) is higher than both other combinations (3=excluded participant). The bold/italic values indicate the channels for which both comparisons are statistically significant. With baseline correction.*

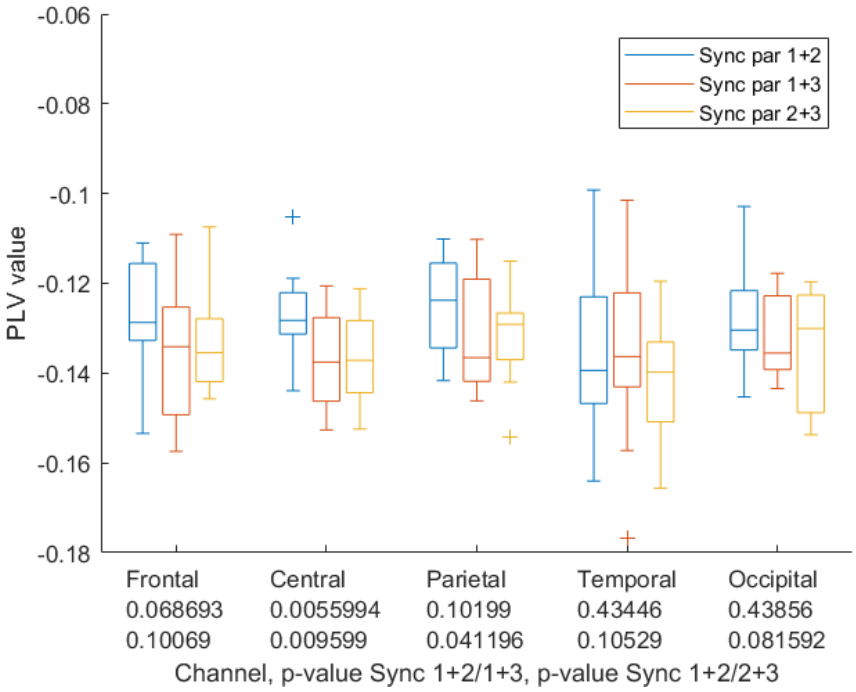


Figure 18: *Method M1: Gamma frequency band. Best performing combination for central brain region after baseline correction.*

5.2.2.2 Channel Level

The results at channel level were also investigated. Table 13 shows all channels which had at least one significant value after comparison with either participant combination 1 and 3, or 2 and 3. Only those channels for which the other p-value was at least below 0.25 were included in the table. The complete table (divided over two tables) can be found in Appendix B.6.

Table 13 shows that channel 23 (C4) is statistically significant when comparing participant combination 1 and 2 (non-excluded participants) with both other combinations. This is the case for all frequency bands except the gamma frequency band. For the gamma band, channel 26 (FC2) is the only channel which is statistically significant. Both these significant channels are located in the central region of the brain. For channel C4, both the theta and alpha band show the greatest significance. This is slightly different compared to the analysis of channel level results without baseline correction since the delta and theta band were the frequency bands that showed the greatest significance there (see Section 5.2.1.2).

Even though these results are better than the analysis results without baseline correction, none of these results remain significant after applying FDR correction. This is even the case after using a subset of the cleanest (least noisy) channels to reduce the number of multiple comparisons.

The boxplot figures with an overview of the channel results of all frequency bands can be found in Appendix B.6.

Freq band	Combination	Channel	p-value
D	1+2/1+3	6 (FC5)	0.043
	1+2/2+3		0.182
	1+2/1+3	21 (CP6)	0.003
	1+2/2+3		0.140
	1+2/1+3	22 (CP2)	0.049
	1+2/2+3		0.201
	1+2/1+3	***23 (C4)***	***0.010***
	1+2/2+3		***0.007***
T	1+2/1+3	6 (FC5)	0.044
	1+2/2+3		0.180
	1+2/1+3	21 (CP6)	0.004
	1+2/2+3		0.142
	1+2/1+3	22 (CP2)	0.043
	1+2/2+3		0.207
	1+2/1+3	***23 (C4)***	***0.008***
	1+2/2+3		***0.007***
A	1+2/1+3	6 (FC5)	0.044
	1+2/2+3		0.177
	1+2/1+3	21 (CP6)	0.004
	1+2/2+3		0.139
	1+2/1+3	22 (CP2)	0.047
	1+2/2+3		0.206
	1+2/1+3	***23 (C4)***	***0.009***
	1+2/2+3		***0.006***
B	1+2/1+3	6 (FC5)	0.040
	1+2/2+3		0.195
	1+2/1+3	21 (CP6)	0.005
	1+2/2+3		0.125
	1+2/1+3	***23 (C4)***	***0.009***
	1+2/2+3		***0.009***
G	1+2/1+3	19 (P4)	0.069
	1+2/2+3		0.008
	1+2/1+3	20 (P8)	0.036
	1+2/2+3		0.063
	1+2/1+3	***26 (FC2)***	***0.015***
	1+2/2+3		***0.035***
	1+2/1+3	29 (AF4)	0.239
	1+2/2+3		0.017

Table 13: *p-values for the frequency bands (D=delta, T=theta, A=alpha, B=beta, G=gamma) and channels for which the mean PLV value from participant combination 1 and 2 (non-excluded participants) is higher than both other combinations and the other p-value was below 0.25 (3=excluded participant). The bold/italic values indicate the channel for which both comparisons are statistically significant. With baseline correction.*

5.2.3 Summary of Main EEG Results

The multiple EEG analyses above were quite extensive. Therefore, the main results of these analyses will now be provided in a short summary.

All EEG data analysis results indicate that the central brain region has the most promising results. This is shown both at the regional level as well as at channel level since the most significant channel (C4) is part of the central region as well. Furthermore, the frequency band with the strongest result is not the same for the regional level and channel level comparison. For the regional level analysis, the best results are achieved with the gamma frequency band, and for the channel level, the best results are achieved with either the theta or alpha frequency bands.

Lastly, while the usage of subsets of experiments only slightly improved the results, a clear improvement was found when analysing baseline-corrected PLV values. The central region with the gamma frequency band now showed a significant difference between the PLV values from participant combination 1 and 2 (non-excluded participants) with both other combinations ($p = 0.006$, $p = 0.010$). This improvement was also found at channel level since channel C4 in the theta frequency band now showed a significance level of $p = 0.008$ and $p = 0.007$ compared to the significance level of $p = 0.019$ and $p = 0.020$ from before the baseline correction. However, none of these results remain significant after FDR correction.

5.3 Heart Rate (HR) Analysis

For the heart rate IBI analysis, the Pearson Correlation Coefficient (PCC) results for a 5-second and 10-second time window were compared. Then, the trend of the curve of the data was examined as well.

5.3.1 Correlation 5-Second Time Window

First, the results for the 5-second window will be described. Table 14 shows the mean correlation values for the 5-second time window for each of the 13 experiments, as well as the overall mean.

Experiment	**Correlation 1+2**	**Correlation 1+3**	**Correlation 2+3**
1	-0.6430	-0.0989	0.3287
2	***0.5019***	-0.1231	0.1563
3	***0.2639***	-0.9554	-0.0762
4	-0.8882	-0.9005	0.7923
5	0.9605	0.9651	0.9753
6	0.2714	-0.7897	0.2871
7	-0.4627	0.1040	-0.4971
8	-0.4917	0.6895	0.1698
9	-0.9661	-0.5925	0.7164
10	0.1449	-0.8110	0.4079
11	-0.7969	-0.8087	0.8152
12	***0.8210***	-0.8302	-0.9357
13	0.7125	0.8566	0.6596
Mean Correlation	-0.0440	-0.2535	0.2923

Table 14: *IBI correlation values for each participant combination in the 13 experiments (time window: 5 sec). The bold/italic values indicate the experiments for which the mean correlation is higher for participant combination 1+2 (non-excluded participants) than both other combinations.*

As can be seen in Table 14, the correlation from participant 1 and 2 (non-excluded participants) is only higher than both other combinations for experiment 2, 3, and 12. In fact, the mean correlation between participant 1 and 2 is even negative. Overall, there is quite a lot of variance between the experiments. The p-values (once again computed using the Permutation Test with 10000 permutations) which have been calculated for the comparison between these mean correlation values are not significant at all with $p = 0.2191$ and $p = 0.9133$ for the comparison between participant 1 and 2 with 1 and 3, and 1 and 2 with 2 and 3, respectively.

Thus, it was also investigated whether significance could be found at an experiment level. Table 15 shows the p-values for each experiment and participant combination. It can be seen that the experiments for which the correlation from participant 1 and 2 (non-excluded participants) is higher than the other two combinations (experiment 2, 3, 12), this difference is significant ($p < 0.05$).

Experiment	Combination 1+2/1+3	Combination 1+2/2+3
1	0.4985	0.4968
2	***0.0001***	***0.0001***
3	***0.0001***	***0.0001***
4	0.0001	0.5041
5	0.4910	0.4969
6	0.0001	0.4992
7	0.4968	0.0001
8	0.4988	0.5011
9	0.5007	0.5032
10	0.0001	0.5106
11	0.0001	0.4889
12	***0.0001***	***0.0001***
13	0.5005	0.001

Table 15: *IBI p-values for each participant combination in the 12 experiments (time window: 5 sec). The bold/italic values indicate the experiments for which the mean correlation is significantly higher for participant combination 1+2 (non-excluded participants) compared to both other combinations.*

5.3.2 Correlation 10-Second Time Window

Experiment	Correlation 1+2	Correlation 1+3	Correlation 2+3
1	-0.6188	-0.3637	0.3227
2	0.0820	0.1280	0.2624
3	-0.2476	-0.6312	-0.0920
4	-0.5875	-0.4954	0.6891
5	***0.8147***	0.4524	0.2907
6	-0.4016	-0.5325	0.6997
7	***0.5911***	0.4575	0.2056
8	-0.8281	0.7956	-0.4460
9	-0.8032	0.2826	0.2892
10	***0.1894***	-0.6250	0.0716
11	0.1227	-0.1688	0.7084
12	***-0.2824***	-0.3851	-0.2859
13	0.4995	0.5571	-0.0993
Mean Correlation	-0.1131	-0.0406	0.2012

Table 16: *IBI correlation values for each participant combination in the 13 experiments (time window: 10 sec). The bold/italic values indicate the experiments for which the mean correlation is higher for participant combination 1+2 (non-excluded participants) than both other combinations.*

Then, the correlation values for the 10-second time window were examined. Table 16 shows the correlation values for each participant combination in the 13 experiments as well as the overall mean, but now for this larger time window of 10 seconds. As can be seen, the correlation from participant 1 and 2 (non-excluded participants) is only higher than both other combinations for experiment 5, 7, 10, and 12. In fact, the mean correlation between participant 1 and 2 is even the most negative of them all, and more negative

than the results from the 5-second window as well. Overall, there is once again quite a lot of variance between the experiments. The p-values which have been calculated for the comparison between these mean correlation values are even worse than for the 5-second window, and thus not significant at all; $p = 0.6339$ and $p = 0.9485$ for the comparison between participant 1 and 2 with 1 and 3, and 1 and 2 with 2 and 3, respectively.

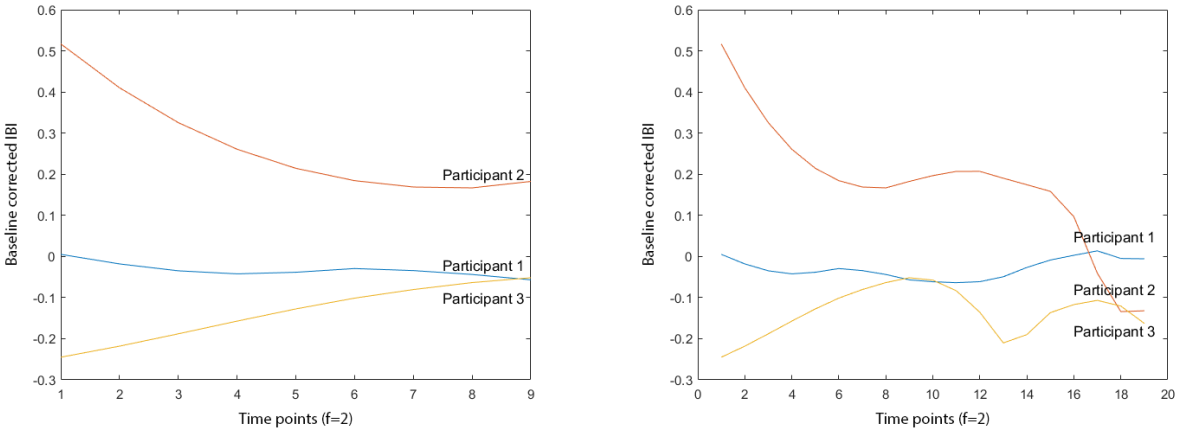
Similarly to the 5-second window, it was investigated whether significance could be found at an experiment level for this 10-second window. Table 17 shows the p-values for each experiment and participant combination. The experiments for which the correlation from participant 1 and 2 (non excluded participants) is higher than the other two combinations (experiment 5, 7, 10, 12), is a significant difference ($p < 0.05$). These experiments are almost completely different from the results with the 5-second time window. Only experiment 12 is statistically significant for both time windows.

Experiment	**Combination 1+2/1+3**	**Combination 1+2/2+3**
1	0.4982	0.4998
2	0.5031	0.4983
3	0.0001	0.4995
4	0.5059	0.4954
5	***0.0001***	***0.0001***
6	0.0001	0.4966
7	***0.0001***	***0.0001***
8	0.4928	0.5029
9	0.4997	0.5041
10	***0.0001***	***0.0001***
11	0.0001	0.5013
12	***0.0001***	***0.0001***
13	0.5057	0.001

Table 17: *IBI p-values for each participant combination in the 12 experiments (time window: 10 sec). The bold/italic values indicate the experiments for which the mean correlation is significantly higher for participant combination 1+2 (non-excluded participants) compared to both other combinations.*

5.3.3 Trend of IBI Curve

Additionally, the IBI value curve from the three participants was compared for both time windows. For this comparison, experiment 12 was chosen to depict a scenario in which participant 1 and 2 (non-excluded participants) had a higher correlation than both other combinations. See Figure 19. This was the only experiment for which this was the case for both time windows (as shown in Tables 14 and 16). Conversely, experiment 8 was chosen to depict a scenario in which participant 1 and 2 had a low correlation. See Figure 20.



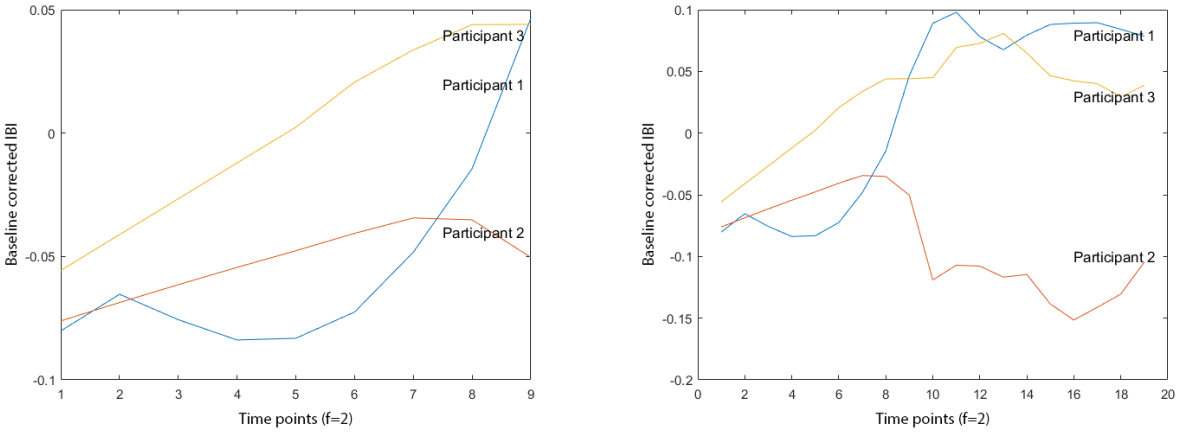
(a) 5-second window

(b) 10-second window

Figure 19: *IBI data from the three participants for the (a) 5-second time window and the (b) 10-second time window from experiment 12 (high correlation). The blue line indicates participant 1, the yellow line is for participant 2, and the red line indicates participant 3 (excluded participant).*

Figure 19 (a) shows that the IBI from participant 1 starts at the second-highest position and stays relatively constant throughout the 5-second time window, with only a slightly decreasing arch (which also increases a fraction just before the halfway point of the time window). The IBI from participant 2 starts the highest, then goes down, and finally results in still being the highest IBI. Lastly, the IBI from participant 3 (socially excluded participant) starts at the lowest position and steadily goes up, ending up with the same IBI value as participant 1. As can be seen, the trend from the curve of participant 1 and 2 is relatively similar (with IBI values which are quite close together as well), resulting in a high correlation between these two participants. Conversely, the curve from participant 3 goes in a completely different direction, resulting in a low correlation with both other participants.

In Figure 19 (b), the trend from participant 1 mostly continues in the larger time window (with some slight fluctuations), ending up with the highest IBI. The IBI from participant 2, on the other hand, goes up again near the middle of the time window and has a steep decline near the end, making it end up with the second-highest IBI. Lastly, the IBI from participant 3 decreases right in the middle of the larger time window and then goes up a bit again near the end, but it still ends up with the lowest IBI (although not by much). This means that throughout both time windows, the third (excluded) participant has a faster heart rate than the other two participants since the IBI is lower (shorter).



(a) 5 second window

(b) 10 second window

Figure 20: *IBI data from the three participants for the (a) 5 second time window and the (b) 10 second time window from experiment 8 (low correlation). The blue line indicates participant 1, the yellow line is for participant 2, and the red line indicates participant 3 (excluded participant).*

Figure 20 (a) shows that the IBI from participant 1 starts the lowest (although almost similar to participant 2), goes up, goes down a bit, and then goes up in a steep arch, ending up with the highest IBI (together with participant 3). The IBI from participant 2 steadily increases and then has a small dip near the end, making it end up with the lowest IBI. Lastly, the IBI from participant 3 (socially excluded participant) starts the highest and keeps steadily increasing until it ends up with the same highest IBI as participant 1. As can be seen, the curves from all three participants are quite different, causing low correlation overall.

In Figure 20 (b) it can be seen that the trend from participant 1 continues and then has a small dip before evening out to result in the highest IBI. The IBI from participant 3 also continues to increase more and then decreases, making it end up with the second-highest IBI. The IBI from participant 2, on the other hand, decreases even more, then evens out a little, only to have a small dip and ending in an increasing arch. However, its IBI is still the lowest out of the three participants. This means that in this case one of the participants who is excluding someone actually has a higher heart rate than the excluded participant themselves.

When comparing Figures 19 and 20 it becomes clear that there is quite a lot of variance between the experiments. Furthermore, both figures show that there is a fairly large change after the end of the 5-second time window, which might indicate that there is a delayed reaction and that the 10-second window is, thus, better suited for capturing the IBI from the three participants.

5.4 Galvanic Skin Response (GSR) Analysis

For the GSR analysis, the correlation and the trend of the curve were also examined, but now only with a time window of 5 seconds.

5.4.1 Correlation

Figure 18 shows the mean correlation values for each participant combination for the 12 remaining experiments (experiment 6 was discarded because of data loss) as well as the overall mean correlation value.

Experiment	**Correlation 1+2**	**Correlation 1+3**	**Correlation 2+3**
1	0.0144	0.9141	0.0018
2	0.9864	0.9969	0.9718
3	-0.1662	-0.2618	0.9551
4	***0.4040***	-0.1828	0.0286
5	***0.9620***	-0.0248	-0.1386
7	-0.8398	-0.7228	0.8644
8	***0.1826***	-0.7065	-0.2603
9	***0.7845***	-0.7727	-0.9769
10	-0.8861	0.9396	-0.8205
11	0.7250	0.9742	0.7929
12	-0.4448	-0.9873	0.5434
13	0.0177	0.1858	0.7793
Mean Correlation	0.1450	0.0293	0.2284

Table 18: *GSR correlation values for each participant combination in the 12 experiments (experiment 6 was discarded because of data loss). The bold/italic values indicate the experiments for which the mean correlation is higher for participant combination 1+2 than both other combinations.*

As can be seen in Table 18, the correlation from participant 1 and 2 (non-excluded participants) is only higher than both other combinations for experiment 4, 5, 8 and 9. Similar to the IBI data, there is quite a lot of variance between the experiments for the GSR data. The p-values which have been calculated for the comparison between these mean correlation values are not significant with $p = 0.3372$ and $p = 0.6178$ for the comparison between participant 1 and 2 with 1 and 3, and participant 1 and 2 with 2 and 3, respectively.

Thus, it was investigated whether significance could be found at an experiment level. Table 19 shows the p-values for each experiment and participant combination. The experiments for which the correlation from participant 1 and 2 (non-excluded participants) is higher than the other two combinations (experiment 4, 5, 8, 9), actually show a significant difference as well ($p < 0.05$).

Experiment	Combination 1+2/1+3	Combination 1+2/2+3
1	0.5028	0.0001
2	0.4986	0.0001
3	0.0001	0.4982
4	***0.0001***	***0.0001***
5	***0.0001***	***0.0001***
7	0.5003	0.5001
8	***0.0001***	***0.0001***
9	***0.0001***	***0.0001***
10	0.4978	0.5121
11	0.4936	0.5045
12	0.0001	0.5001
13	0.5033	0.5052

Table 19: *GSR p-values for each participant combination in the 12 experiments. The bold/italic values indicate the experiments for which the mean correlation is significantly higher for participant combination 1+2 compared to both other combinations.*

5.4.2 Trend of GSR Curve

The trend of the GSR signal after the language switch was also investigated (see Figure 21). As can be seen, participant 3 (socially excluded participant) shows a much more negative trend than the other two participants. Their GSR signal is much more flat in comparison (although this is slightly less so for the second participant). However, the scale of the y-axis of participant 1 is much larger, showing much greater variance. Therefore, Figure 22 also shows the GSR signals for the three participants on the same y-axis scale. Here, it can be seen that participant 1 actually has the greatest (negative) slope and the GSR signals for participant 2 and 3 are flatter. To show this, the directionality of the slope from the three mean signals was computed by fitting a linear polynomial on the data of all experiments and then averaging this direction. The direction of the slope of participant 1 is -0.0025, for participant 2 it is -0.0016 and for participant 3 it is -0.0013.

Then, the curves of one experiment for which participant 1 and 2 had a high correlation and one experiment with low correlation were examined (see Figure 23). In 23 (a), it can be seen that both the GSR from participant 1 and 2 (non-excluded participants) show a similar curve and decrease during the 5-second time window, while the GSR from participant 3 (socially excluded participant) slightly increases. This increased skin conductance response shows that the excluded participant has a higher sweat gland activity which indicates a higher level of arousal. Conversely, in 23 (b), participant 1 and 3 show the most similar curve (only slightly decreasing) while the curve from participant 2 decreases much more. This means that participant 1 also has an increased level of arousal.

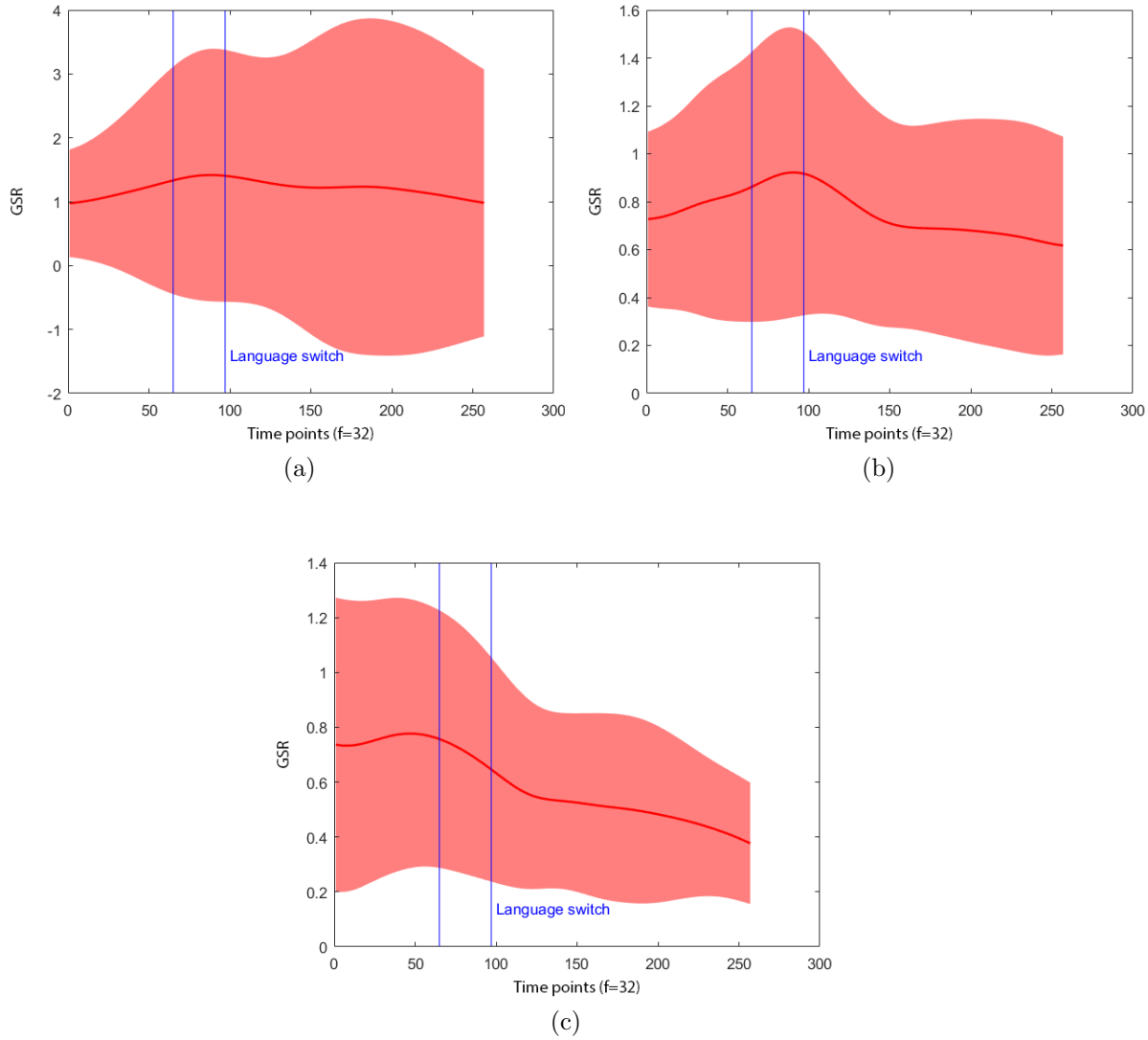
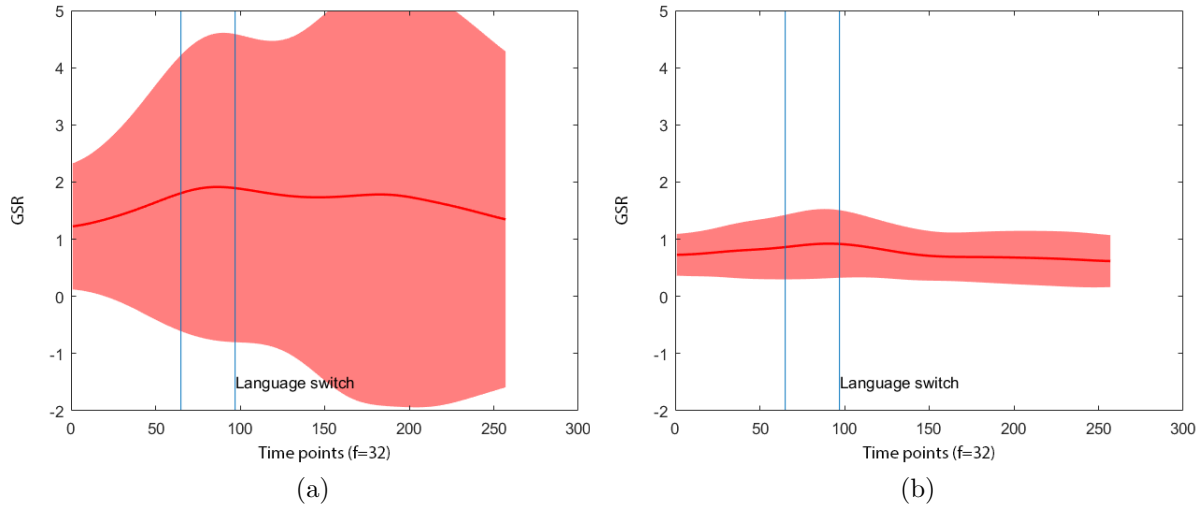
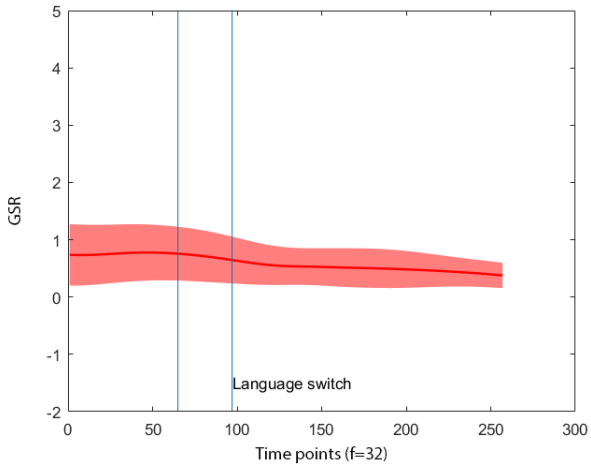


Figure 21: *Mean and standard deviation of the GSR data of the three participants. (a) shows the data from participant 1, (b) from participant 2, and (c) from participant 3 (socially excluded participant). The blue lines indicate the language switch. These are 2 lines because the language switch can vary somewhere between that 1-second window.*



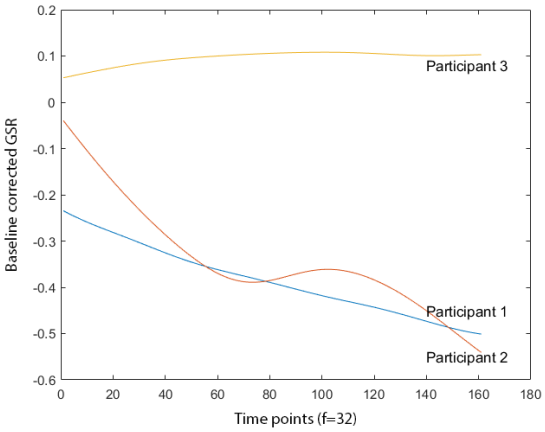
(a)

(b)

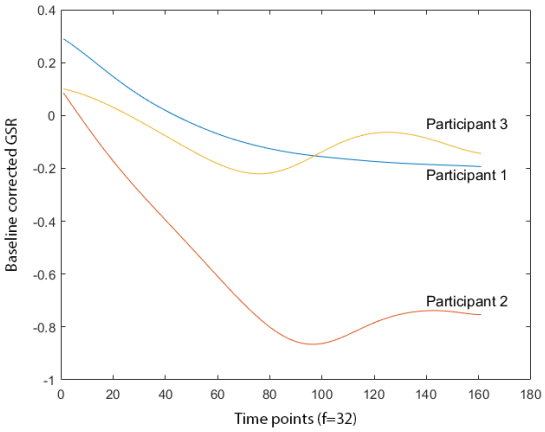


(c)

Figure 22: *Mean and standard deviation of the GSR data of the three participants plotted on the same y-axis. (a) shows the data from participant 1, (b) from participant 2, and (c) from participant 3 (socially excluded participant). The blue lines indicate the language switch. These are 2 lines because the language switch can vary somewhere between that 1-second window.*



(a) High correlation (experiment 9)



(b) Low correlation (experiment 10)

Figure 23: *GSR data from the three participants for (a) an experiment with high correlation and (b) an experiment with low correlation. The blue line indicates participant 1, the yellow line is for participant 2, and the red line indicates participant 3 (excluded participant).*

6 Discussion

The international world we live in makes it crucial to understand each other and not exclude anyone by linguistically ostracising them from participating in teamwork. In this thesis, one of the first hyperscanning studies with three participants has been conducted. In this hyperscanning study, the first steps have been taken to measure this language induced feeling of social exclusion. This chapter provides a discussion of the results that were generated by this research. The results of the statistical analyses of the EEG, HR, and GSR data are discussed and compared with the existing literature. Furthermore, the contributions and limitations of this research, along with recommendations for future work on this topic are provided.

6.1 Statistical Analyses

Before the statistical analyses took place, the effect of the imposed language barrier on inducing a feeling of social exclusion for one participant in each experiment was verified. The results from the post-experiment questionnaire showed that out of the 13 socially excluded participants, 4 felt very excluded, 7 felt excluded and 2 felt slightly excluded. All socially excluded participants, thus, actually felt excluded. Therefore, the language barrier from the other two participants talking in a language the excluded participant could not understand actually had the "desired" effect for this experiment.

This thesis proposed three modalities to measure this feeling of social exclusion: Electroencephalogram (EEG) data, Heart Rate (HR) data, and Galvanic Skin Response (GSR) data. Social exclusion was hypothesised to be reflected in the lack of synchronisation in the data, and that the two participants who are excluding the other participant will thus have a higher level of synchronisation with each other.

6.1.1 Electroencephalogram (EEG)

The statistical analysis of the EEG data was done on both a regional level and channel level by comparing the Phase-Locking Value (PLV) values for the level of synchronisation between the participants. These analyses were performed both with and without baseline correction. A summary of the analyses results will first be given before the findings will be related to the existing literature.

6.1.1.1 Regional Level

The results of the analysis at the regional level using Method M1 (first calculating PLV per channel and then averaging for each of the five brain regions) indicate that the central brain region shows the most promising results since the PLV value from participant combination 1 and 2 (non-excluded participants) is higher than the PLV value from both other combinations in this region. This is the case across all five frequency bands (see Table 3). Conversely, when using Method M2 (first averaging the EEG data for each of the brain regions and then calculating the PLV value) the temporal region has the best results for all frequency bands (see Table 4). The gamma frequency band showed the best results for both methods. The overall mean PLV value from all frequency bands and brain regions lies around +/- 0.32 out of 1.

Furthermore, the results show that Method M1 has a much lower standard deviation than the second method (see Tables 3 and 4). This suggests that Method M1 is a more robust method for the PLV calculation across regions. All further regional analyses results have therefore used this method of PLV calculation.

Statistical analyses of the results without baseline correction show that even though several frequency band and brain region combinations have a higher PLV value for participant combination 1 and 2 (non-excluded participants), none of these are significantly different (see Table 5). Using a subset of the experiments in which the excluded participant indicated to have felt either "excluded" or "very excluded" slightly improved the results in the central region. However, the results from comparing the PLV value participant combination 1 and 2 with the PLV values from the socially excluded participant with the other two participants were still not statistically significant (see Table 7).

Conversely, the statistical analysis of the EEG results when computing the PLV value after correcting for the baseline from before the language switch, actually did show a significant difference in the central region for the beta frequency band and the gamma frequency band. While both these comparisons were statistically significant, the results from the gamma frequency band had greater significance ($p = 0.006$, $p = 0.010$ compared to $p = 0.036$, $p = 0.047$)). However, after applying the False Discovery Rate (FDR) neither of these results remain significant.

Interestingly, the results show that the PLV values after baseline correction are slightly negative for all participant combinations (see Table 11). This shows that all participants were affected by the language switch. While it was expected for participant combinations 1 and 3 as well as 2 and 3, because the assumption was that the socially excluded participant (participant 3) would have the greatest change in PLV value, it is clear that this also happened for the other two participants. One possible explanation for this is that most of these participants mentioned that they were a little uncomfortable and nervous that they had to exclude a participant. This might have led them to focus on other things than team coordination with each other, thus decreasing the PLV level and, thereby, the synchronisation, between them.

6.1.1.2 Channel Level

Subsequently, the PLV values at channel level were also investigated. Results show that only channel C4 is statistically significant when comparing the PLV value of participant 1 and 2 (non-excluded participants) with the other two combinations. This is the case for all frequency bands, except the gamma frequency band. Instead, the delta and theta frequency bands performed best with p-values of $p = 0.017$, $p = 0.022$ and $p = 0.019$, $p = 0.020$ for the delta and theta frequency bands, respectively (see Table 6). However, the statistical significance of this channel does not hold after FDR correction.

On a channel level, the results were also greatly improved after baseline correction, with the results showing greater statistical significance. In total, there are also more channels for which one of the combinations shows statistical significance. However, for the delta, theta, alpha, and beta frequency bands, it is still only channel C4 which shows statistical significance for both comparisons. Interestingly, it is channel FC2 (and not channel C4) which shows statistical significance in the gamma frequency band (see Table 13). Nevertheless, both of these channels belong to the central brain region. The theta and alpha band both show the greatest significance ($p = 0.008$, $p = 0.007$ and $p = 0.009$, $p = 0.006$,

respectively). However, these results do not remain significant after FDR correction as well. This is even the case after using just a subset of the least noisy channels to reduce the number of multiple comparisons.

6.1.1.3 Overall Conclusions and Comparison to Literature

From this EEG data analysis it is clear that the central brain region has the most promising results. This is shown in both the statistical analysis at regional level as well as at channel level since channel C4 (the most significant channel) is part of the central region. These results are somewhat similar to the results from Lindenberger et al. (2009) since their research showed that brain-to-brain synchrony primarily involved the fronto-central regions of the brain. Furthermore, Lindenberger et al. (2009) used the PLI method to compute brain-to-brain synchrony, which is a very similar measure to the PLV measure used in this thesis research and should give the same results when the sources of the data are separated between brains (Aydore et al., 2013). However, the performed task was completely different, since the participants were asked to play guitar together in the research by Lindenberger et al. (2009). Social exclusion was, thus, not involved in this experiment at all. Moreover, these findings are also somewhat in line with the findings from Wang et al. (2018). After all, channel C4 is part of the Mirror Neuron System (MNS) since it is located on the primary motor cortex (Wang et al., 2006). The research by Wang et al. (2018) showed that the brain areas in the MNS (and MS) neural systems are most likely to show brain-to-brain synchrony, which matches the findings in this thesis research. Furthermore, since the experimental task and subsequent feelings of social exclusion are all language-based, it is also interesting to note that the primary motor cortex is related to language (Wang et al., 2018).

Interestingly, the frequency band with the strongest result is not the same for the regional level and channel level. For the regional level, the best results are achieved with the gamma frequency band, while the channel level achieved the best results with either the theta or alpha frequency bands. Although the results at channel level are, thus, in stark contrast with the results on a regional level, they are more in line with previous research. After all, previous research showed that the frequency bands below 12 Hz (delta, theta, alpha) show the strongest brain-to-brain synchrony (Babiloni et al., 2006; Czeszumski et al., 2020; Mueller et al., 2013; Davidesco et al., 2019a,b; Kawasaki et al., 2013). These results are to be expected since these studies also investigated brain-to-brain synchrony at channel level instead of at a broader regional level. It seems that averaging over multiple channels to investigate the brain region more broadly caused the frequency band for which the synchronisation was sensitive to become much higher. Nevertheless, these studies did not investigate social exclusion or another psychological construct. Instead, they studied other social interactions, such as cooperation during a card game, simultaneous guitar playing, and speech tasks (Babiloni et al., 2006; Czeszumski et al., 2020; Mueller et al., 2013; Davidesco et al., 2019a,b; Kawasaki et al., 2013).

Even though the EEG results from this research are not significant there is at least an indication that the central brain region in combination with the gamma frequency band at the regional level and theta/alpha frequency band at the channel level is the most important when comparing brain-to-brain synchrony after social exclusion. Nevertheless, this is merely an indication and is, unfortunately, not proven yet with this research. This exact combination of results cannot be compared to the literature since, to the best of the researchers' knowledge, no previous work has ever investigated the effect of social

exclusion on the brain-to-brain synchrony in an EEG hyperscanning study such as this. However, the brain-to-brain synchrony for social inclusion (team coordination) situations, have been investigated using EEG hyperscanning studies before. An example of such a study is the study by Szymanski et al. (2017), although their experimental setting was slightly different since they had participants perform a visual search task. Their findings show similar results to the ones indicated by the current research since they have shown that the brain-to-brain synchrony was also highest in the central region with the theta (and delta) frequency band. These results give further indication that the central brain area is important for showing high brain-to-brain synchrony when people are socially included as well as the lack thereof when people are being socially excluded.

One possible reason for the lack of significance in the research results could be found in the low average PLV value (+/- 0.32) that was calculated in this research. Other literature has shown much higher PLV values of at least 0.6 on average (Jian et al., 2017; Wang et al., 2006). While Wang et al. (2006) had some results with similar PLV values, their research also had much higher PLV values of around 0.85. The current research was far from reaching PLV values as high as that. This suggests that a psychological construct such as social exclusion (instead of other social interactions) might simply be much harder to quantify using brain-to-brain synchrony measures such as this.

A possible cause for this low PLV might be because of the pseudo-trials that needed to be created to compensate for the fact that PLV computation does not work with only one single trial (as was the case because of the short time window) (Lachaux et al., 1999). If multiple "real" trials could have been used, then every trial would presumably show a similar synchronisation trend across each trial. After all, the idea of using trials is that a certain "action" is repeated for each trial and that a similar trend should be shown across all these trials. With these pseudo-trials on the other hand, this similar synchronisation trend would be much less pronounced since that "action" is not repeated for each trial. Instead, only a small segment of the "action" would be used for each of the pseudo-trials because a sliding time window is used to create these pseudo-trials. If there is a lot of variance during the time window, the PLV computation across these pseudo-trials would result in a low PLV value. Gysels and Celka (2004) also used a sliding time window to create pseudo-trials for their PLV calculation, although their purpose was to determine the usefulness of synchronisation in classifying mental tasks. However, in contrast to the results acquired by this thesis research, Gysels and Celka (2004) found PLV values of around 0.6 which is similar to the findings from Jian et al. (2017) and Wang et al. (2006). This would imply that the use of pseudo-trials does not influence the resulting PLV values (or at least not by much). This makes it less credible that the PLV values are this low because of the use of pseudo-trials. The most credible explanation for these low PLV values is, thus, still the suggested theory that social exclusion is simply much harder to quantify using such brain-to-brain synchrony measures.

6.1.2 Heart Rate (HR)

For the heart rate analysis, a 5-second and 10-second window after the first language switch were compared to each other. The results show that the correlation from participant 1 and 2 (non-excluded participants) is only higher than the correlation from the other two participant combinations for experiment 2, 3, and 12 in the 5-second window and experiment 5, 7, 10, and 12 in the 10-second window. For these individual experiments, the differences are statistically significant ($p = 0.0001$ for both comparisons).

Thus, only experiment 12 is significant for both time windows. Conversely, the mean correlation values from all experiments are far from having a statistically significant difference since the correlation from participant 1 and 2 (non-excluded participants) is only the second-highest for the 5-second window, and the lowest for the 10-second window (as opposed to the expected highest correlation). Moreover, the overall mean correlation is even negative for these non-excluded participants (see Tables 14 and 16). This suggests that the (mean) IBI correlation values for each participant combination are not a very good representative for the synchronisation between participants. One possible explanation for this lack of consistent correlation for the participant combinations could be the large variation between all experiments; only a small number of experiments show a positive correlation for the non-excluded participants while the rest of the experiments even show varying degrees of negative correlations. This large variation could indicate that IBI correlation might not be sufficient for comparing HR synchronisation from an experiment in which social exclusion is measured. These results are in contrast to the research by Reinerman-Jones et al. (2011) which suggested that correlation was a good measurement for detecting synchrony between HR data.

Nevertheless, the results suggest that there is at least an indication that social exclusion might be able to be measured using HR. After all, the analysis results from experiment 2, 3, and 12 for the 5-second window, and experiment 5, 7, 10, and 12 for the 10-second window showed the synchronisation as expected. Even though there are clear differences between the two time windows, both show that the socially excluded participant had the lowest IBI, indicating a faster heart rate (see Figure 19). This faster heart rate indicates higher arousal for this excluded participant which, in turn, could indicate that they were feeling more negative emotions such as anger or anxiety because of the social exclusion which they were experiencing, as is shown in Figure 2 (Molden et al., 2009). This is in line with previous work from Williams (2002) which researched the emotions social exclusion could evoke.

The fact that there is quite a change in the trend of the IBI curve after the 5-second time window (and is still changing quite a bit in the 10-second time window), suggests a delay of several seconds (or more) before the effect of the exclusion is shown in the heart rate. Perhaps this also means that the correlation measure actually could indicate synchrony, but that it needs to be computed for IBI data over a later and/or longer time window. If that were the case, the findings would be in line with the work by Reinerman-Jones et al. (2011).

However, most previous literature analysed the HR data based on the HRV instead of the IBI (Acharya et al., 2006). The reason that this was not done in the current study is because it was not possible since the used time windows were too short. If it is indeed the case that a longer time window is needed to reliably use the PCC method to measure synchrony, then using HRV might be better suited.

6.1.3 Galvanic Skin Response (GSR)

The results of the GSR analysis showed that the mean correlation of participant 1 and 2 (non-excluded participants) is higher than the other two combinations for experiment 4, 5, 8, and 9. The differences within these experiments are statistically significant as well (with $p = 0.0001$ for both comparisons). See Tables 18 and 19. However, these experiments are completely different from the significant experiments after HR analysis (2, 3, 12 and 5, 7,

10, 12). Only experiment 5 is also significant for the 10-second window in the HR analysis. Furthermore, the mean correlation from participant 1 and 2 across all experiments is only the second-highest, and not the highest as was expected. Similar to the HR analysis results, these results suggest that the (mean) GSR correlation values for each participant combination are not a very good representative for synchrony between participants. This is once more in contrast with the findings by Reinerman-Jones et al. (2011) since their overview of the comparison between several different statistical analysis methods showed that correlation (and ANOVA) provide the best direct methods for analysing physiological data such as HR and GSR.

Additionally, the trend of the GSR signal after the first language switch was also investigated. At first glance, it appeared that the socially excluded participant had a much more negative trend than the other two participants (see Figure 21). However, after calculating the directionality of the curve and plotting the GSR signals of the three participants on the same scale, it became clear that participant 1 actually had the most negative directionality and that the direction of the curve of the socially excluded participant was the most positive (or least negative). See Figure 22. This suggests that the directionality of the curve might have influenced the correlation between the participants, causing the measure not to work as expected. Nevertheless, when participant 1 and 2 (non-excluded participants) had a higher correlation to each other (experiment 4, 5, 8, 9), the GSR curve from the socially excluded participant increases during the analysed time window. This increased skin conductance response shows that the socially excluded participant has a higher sweat gland activity which indicates higher arousal and could thus indicate more negative emotions (Molden et al., 2009).

However, even though the literature suggests that skin conductance measures are generally regarded as the most sensitive of emotional arousal such as this (Frazier et al., 2004), the results from the HR analysis were actually more in line with expectations. After all, the findings from the GSR analysis show that experiment 4, 5, 8, and 9 are significantly different, while two of socially excluded participants in these experiments indicated that they only felt slightly excluded. Conversely, for the HR analysis, the socially excluded participant indicated to have felt very excluded for two of these experiments and excluded for the rest of these experiments (2, 3, 12 and 5, 7, 10, 12).

6.2 Contributions

There are several contributions made by this thesis research. First of all, this study has pioneered in using three neurophysiological signals (EEG, HR, and GSR) to measure feelings of social exclusion induced by a language barrier. Until now, such psychological constructs had not been investigated by actually measuring neurophysiological signals. Instead, earlier research has only ever used surveys to measure social exclusion. Furthermore, previous research has mostly only conducted hyperscanning studies using two participants while this thesis research has been one of the first to simultaneously measure the data from three participants during an experiment (Xie et al., 2019; Burgess, 2013). Lastly, while social exclusion has previously only been researched in workplace settings, this study has investigated social exclusion among students at a university (Oshri et al., 2008; Fiset and Bhave, 2019; Thau et al., 2007; Williams, 2007).

6.3 Limitations

From the results that have been discussed, it can be concluded that there are still some limitations and possibilities for future work when it comes to measuring social exclusion. First of all, perhaps the most obvious limitation is the number of experiments that have been conducted. This is especially the case because of the data loss of the first three experiments, leaving only 13 experiments in total. If there are more experiments the results could be different and the conclusions will be more meaningful.

Furthermore, another limitation is the high level of variance between the experiments. One explanation for this might be that not all the participants that spoke in their mother tongue excluded the third participant quite as much. Sometimes, one of the two participants tried to give a brief summary to the socially excluded participant after their discussion was over, while others only reacted to the socially excluded participant when they directly asked them to speak in English again. Furthermore, the reactions from the socially excluded participant also differed from person to person. Some went completely quiet after the first two riddles and did not engage in helping behaviours anymore, while others kept trying to insert themselves into the conversation. This last group indicated that something similar to this has happened to them before and that they were simply used to it. Inserting themselves is the only way in which they can take part in the discussion in such cases. However, that does not lessen the fact that they felt socially excluded. Nevertheless, this might not be a problem which can easily be overcome. People are all different and will thus also always react differently in situations such as this.

Additionally, the experimental situation in which the feeling of social exclusion was induced is, of course, different from the natural scenario. This is especially the case for the two participants who are excluding the other participant. After all, linguistically ostracising someone usually happens involuntarily (Robinson et al., 2013), while it is done on purpose in this experiment. While it might have the same effect on the socially excluded participant, some of other two participants have indicated to have felt uncomfortable and guilty while doing it (conversely, some indicated to have even enjoyed it). This might have influenced the level of team coordination between them as well since they might have been a little too worried about their task of excluding the other participant. Nevertheless, they both needed to do that so it could also be argued that they at least felt similarly and should still have a greater level of synchrony with each other than with the excluded participant. Additionally, they have a shared goal of excluding the other participant for which it might be argued that their level of synchronisation should increase as well. Furthermore, the socially excluded participant sometimes guessed that there was at least something going on because the researchers did not step in to ask the other two participants to switch back to English. Perhaps it might have been better to at least say that they should switch back to English a couple of times during the experiment (and telling them that they should basically just ignore the researcher when that happens) to make the socially excluded participant less suspicious. This could be investigated in a future study. Of course, investigating a natural scenario would be most desirable, but it is not actually possible to have a completely natural scenario. After all, it has to happen in a research environment since the participants need to be equipped with the sensors. Furthermore, when people are specifically asked to work together to solve a task, they probably will not switch over to their mother tongue during that time. Unless, of course, the task is so difficult or emotional that they cannot express themselves in English (Tenzer et al., 2014).

Moreover, participant 1 and 2 (non-excluded participants) were now determined based on the seat arrangement. This way, participant 1, 2, and 3 were measured with the same measuring equipment across all experiments. However, there might have been a better way to determine who participant 1 and 2 were. Perhaps, it could have been decided based on who took the lead in solving the riddles the most. The one who took the lead the most could, for example, be participant 1. Nevertheless, the goal is to see that the synchrony between participant 1 and 2 is higher than the synchrony of both other combinations. This means that deciding who from the two participants speaking in their mother tongue is participant 1 and who is participant 2 should not matter. Even so, another division of participant 1 and 2 could have potentially improved the results, since it is now often the case that participant 2 and 3 have a high level of synchrony. However, these results would not necessarily have been meaningful, because seat arrangement should not matter.

Lastly, it was assumed that the PLV value becomes lower for each trial (each riddle). After all, the first time people switch language, the surprise effect would be the strongest. This feeling of exclusion might change over time, and it is, thus, expected that the contrast in PLV becomes smaller in the later trials, since people would have basically gotten used to being socially excluded. However, this assumption was not confirmed. Therefore, it would be interesting to compare the data (from the HR and GSR data as well) from the first language switch to the last language switch in another study, to see if this assumption is true.

6.4 Recommendations for Future Work

Based on the limitations described above, several recommendations for future work can be made. A first recommendation would be to verify and improve the results of this thesis by conducting the experiments with more participants. Data from more participants will make the statistical analyses, and thus the results, more reliable and meaningful than they are now.

Furthermore, there were quite a lot of individual differences between the socially excluded participants in the way how they reacted to the social exclusion. Some people fell quiet relatively quickly and engaged in fewer helping behaviours while others kept inserting themselves into the conversation. It would be interesting to see if dividing these participants into two groups (people who become quiet and people who keep inserting themselves) could resolve the issue of variation. It is therefore recommended to study the differences in the level of social exclusion between the participants who insert themselves into the conversation as opposed to the participants who stay quiet.

Another recommendation would be to conduct the experiment in a more natural setting. While a completely natural scenario is not possible, the occurrence of the language barrier could arise more naturally. Since the switch to the mother tongue usually happens in a non-purposeful way, simply because it is easier to express yourself in your own language when something difficult needs solving (Robinson et al., 2013), it would be interesting to conduct an experiment with a very difficult language-oriented task. This would hopefully cause the language barrier to arise involuntarily. There are two main advantages to this: (1) the two participants who are excluding the other participant would not constantly think and worry about it (which might have influenced the results), and the socially excluded participant would become less suspicious because of this more natural language barrier. However, beware that this kind of experiment could be a hit or miss scenario

since it is possible that a language barrier does not occur in every conducted experiment. In such a scenario, movement artefacts could also play a larger role, since people will likely move around more if they are to behave as natural as possible. Perhaps the use of fNIRS brain-imaging equipment instead of EEG would be preferable in such a case because of its greater resistance to motion artefacts (Czeszumski et al., 2020).

Additionally, the PLV value that was computed per region was based on all the channels in that region (no matter if the first (M1) or second method (M2) was used). Even though all channels have been extensively pre-processed, some channels are still noisier than others. Therefore, it might be interesting to pick only the two cleanest (least noisy) channels per region for each participant, and only use those representative channels to compute the PLV values. This might improve the results since only the best channels are used. This will at least make the regional results more reliable, since there will be less of a chance of achieving synchrony by luck, simply because of noise.

Moreover, since the PLV values were quite low in comparison to other researches (Jian et al., 2017; Wang et al., 2006; Gysels and Celka, 2004) it is advised to investigate whether another EEG synchrony measure would improve these results. Partial Directed Coherence (PDC) for example, is a frequently used synchrony measure (Burgess, 2013) which could potentially lead to improvements in the significance level of the comparisons of the non-excluded participants with the socially excluded participant. This method is more focused on the causal links and information flow between brains (Czeszumski et al., 2020). This PDC algorithm has been successfully used by Babiloni et al. (2006) to measure brain-to-brain synchrony between participants.

Furthermore, the results from the HR data suggested that there might be a delay in the expected IBI response. Another recommendation is therefore to increase the window length of the analysis and to start this time window several seconds after the language switch. However, if an even larger time window of several minutes would be used, the use of Heart Rate Variability (HRV) instead of IBI could potentially improve the results even more since most previous research also uses this method (Acharya et al., 2006). It would thus be recommended to use HRV instead of IBI. Increasing the window length should ensure that there is enough information available to calculate the HRV. Much more research into the emotional arousal indicated by the HR have been conducted using HRV (Acharya et al., 2006) so using the HRV will likely improve the reliability of the results. As for the GSR data, correlation showed to not be a very good measure for synchrony. Since there are many other different measures to analyse GSR data, it might be better to use such an other measure. For example, it might be interesting to look into detecting the peaks of the SCR and compare those peaks across participants.

Lastly, while this research looked at the synchronisation from each of the three modalities (EEG, HR, GSR) separately, future work could also focus on actually combining these three modalities into a multi-modal system to improve the measurement of social exclusion. An interesting way to do this would be to make a classification system to classify social exclusion. The synchrony measures that have been used in this thesis research, in addition to other signal features, could potentially be used as input features for such a classification system. How well this classifier performs will indicate how well social exclusion can be measured using these neurophysiological signals.

7 Conclusion

In this thesis, one of the first hyperscanning studies with three participants has been conducted to answer the following main research question:

How does social inclusion/exclusion induced by a language barrier reflect in neurophysiological signals?

This chapter provides the conclusions that can be drawn from the research described in this thesis. As has been specified in the Introduction, the main research question has been divided into three sub-questions, which will be answered one by one.

To what extent can a language barrier cause feelings of social exclusion?

Through the results of a post-experiment questionnaire, it was found that, out of the 13 socially excluded participants, 4 felt very excluded, 7 felt excluded, and 2 participants felt slightly excluded. None of them selected one of the other two possibilities (neutral, not excluded at all). All socially excluded participants, thus, indicated that they actually felt excluded. From this, it can be concluded that social exclusion can be reliably induced by a language barrier.

To what extent is social inclusion/exclusion reflected in the synchrony between individuals?

Even though the feeling of social exclusion was successfully induced, it is not reflected very well in the synchrony between the participants. The EEG synchrony computed with the PLV measure was relatively low overall, with a mean value around +/- 0.32 out of 1. The greatest synchrony difference between the socially excluded participant and the other two participants was found in the central region of the brain with the gamma frequency band for the regional comparison and the theta or alpha frequency bands for comparison at channel level. While these results were significant before FDR correction, they do not remain significant after correction. Nevertheless, there is at least an indication that the central brain region is the most important when comparing brain-to-brain synchrony after socially excluding someone.

Furthermore, the overall HR and GSR synchrony between the participants, computed with the PCC, was not a good indication of the social exclusion felt by the socially excluded participants. The mean correlation between the two non-excluded participants across all experiments was not even the highest (and sometimes even negative) despite the expectation that it would be, and was, thus, not significantly higher as well. The HR and GSR results only show a significant difference in synchrony for a few individual experiments. Thus, there is merely an indication that social inclusion/exclusion is reflected in the synchrony between individuals and it is not proven yet with this research.

To what extent can social inclusion/exclusion be measured using EEG, HR, and GSR?

Since social inclusion/exclusion was not reflected well in the synchrony between individuals, and this synchrony was based on the measured EEG, HR, and GSR data, it can be concluded that social inclusion/exclusion was not measured well using these three modalities in this thesis research. Out of the three modalities used in this research, the best results for measuring social inclusion/exclusion are achieved with the EEG modality, although, as has been discussed, these results are not significant. However, this does not necessarily mean that EEG, HR and GSR cannot be used to measure social inclusion/ex-

clusion at all. Several recommendations for improvements have already been discussed in Section 6.4. For example, the usage of a different EEG synchrony measure (such as PDC), using HRV instead of IBI, and detecting and comparing the SCR instead of just looking at the GSR correlation. These changes could potentially improve the extent to which social inclusion/exclusion can be measured using EEG, HR, and GSR.

In conclusion, the fact that there is a significant difference in the PLV values of the EEG data before the calculation of the FDR could at least be an indication that it might be possible to use the PLV values to measure synchrony and, thereby, social exclusion. Furthermore, the HR and GSR data also indicate that there is at least some difference between the data from the socially excluded participants compared to the other two participants. However, further investigation is needed to make any meaningful and definitive conclusions about the extent that social exclusion induced by a language barrier can be measured with these three neurophysiological signals.

A Measures for Reducing Discomfort

As has been stated, it is possible that participants might feel uncomfortable because of the social exclusion taking place. However, in order to mitigate this effect, a time limit for the experiment will be set at 15 minutes. This should give ample time to achieve the desired data as well as minimise the time participants could be feeling uncomfortable. Knowing the full extent of the experiment, as well as the fact that the experiment only takes 15 minutes, there is no reason for the two participants who will ignore the third participant during communication with each other to feel so uncomfortable that they wish to withdraw. Furthermore in order to make the discomfort for the third participant as low as possible, every time a new riddle is addressed (there are seven in total) this is spoken aloud in English, so everyone can understand it. This should reduce the level of discomfort before the third person is ignored again during the largest bulk of communication for solving the riddle. Nevertheless, if it should happen that the third participant still feels so uncomfortable that he/she wishes to withdraw, they are, of course, free to do so. This just means that the experiment will have to be repeated with new participants.

The researchers realise that there is still a possibility that the participants (and the third participant in particular) could feel discomfort and stress. While this experiment setting is necessary to get the desired response from the participants, extra measures have been taken to ensure that this discomfort and stress will not lead to problematic situations. In addition to the time limit of 10 minutes and the switch back to English for every new riddle the level of discomfort has been minimised for the participants in the following three ways:

- The Informed Consent Form and Information Brochure both already indicate that there could be some discomfort during the execution of the task. This way, the participants are at least partly prepared for any feelings of discomfort they may have
- During the experiment, there is a stop protocol in place. Both the researcher and supervisor will be nearby at all times to observe the participants and determine whether the participants are feeling too uncomfortable (even though this is not expected). Signs that show discomfort that will be watched out for are: leaning away, a frown, tense or crossed arms, covering the neck dimple (or grabbing a necklace), rubbing of the forehead, and neck touching . Taking these, and other, signs into account, the researchers will continuously estimate whether it is necessary to ask the participants if he/she still wants to continue. After a continuation of the experiment, if one of these signs has occurred multiple times (or multiple signs have occurred), the experiment will be stopped immediately.
- After the experiment, the participants (with the third participant in particular) will be debriefed about the true purpose of the experiment and the researchers will apologise for the, unfortunately, necessary experiment setting. In addition to that, the participants will also have the opportunity to give their opinion and other thoughts about the experiment. The third participant will also be asked if they still consent their data being used for the study (see Post Experiment Approval document)

Thankfully, executing the stop protocol has been necessary during any of the experiments.

B Extra Figures and Tables

B.1 Regional Level Without Baseline Correction

		Frontal	**Central**	**Parietal**	**Temporal**	**Occipital**
D	1+2	0.3196 ± 0.0185	***0.3230*** ± 0.0108	0.3184 ± 0.0106	0.3138 ± 0.0120	0.3228 ± 0.0253
	1+3	0.3222 ± 0.0193	0.3210 ± 0.0106	0.3249 ± 0.0133	0.3231 ± 0.0221	0.3369 ± 0.0367
	2+3	0.3196 ± 0.0158	0.3178 ± 0.0127	0.3174 ± 0.0079	0.3099 ± 0.0170	0.3141 ± 0.0182
T	1+2	0.3196 ± 0.0185	***0.3230*** ± 0.0108	0.3184 ± 0.0106	0.3138 ± 0.0120	0.3228 ± 0.0253
	1+3	0.3222 ± 0.0193	0.3210 ± 0.0106	0.3249 ± 0.0133	0.3231 ± 0.0221	0.3369 ± 0.0367
	2+3	0.3196 ± 0.0158	0.3178 ± 0.0127	0.3174 ± 0.0079	0.3099 ± 0.0170	0.3141 ± 0.0182
A	1+2	0.3196 ± 0.0185	***0.3230*** ± 0.0108	0.3184 ± 0.0106	0.3138 ± 0.0120	0.3228 ± 0.0253
	1+3	0.3222 ± 0.0193	0.3210 ± 0.0106	0.3249 ± 0.0133	0.3231 ± 0.0221	0.3369 ± 0.0367
	2+3	0.3196 ± 0.0158	0.3178 ± 0.0127	0.3174 ± 0.0079	0.3099 ± 0.0170	0.3141 ± 0.0182
B	1+2	0.3199 ± 0.0176	***0.3222*** ± 0.0103	0.3189 ± 0.0104	0.3128 ± 0.0128	0.3229 ± 0.0235
	1+3	0.3216 ± 0.0186	0.3200 ± 0.0099	0.3244 ± 0.0123	0.3213 ± 0.0212	0.3356 ± 0.0346
	2+3	0.3193 ± 0.0145	0.3170 ± 0.0114	0.3172 ± 0.0071	0.3104 ± 0.0154	0.3136 ± 0.0169
G	1+2	***0.3187*** ± 0.0077	***0.3187*** ± 0.0083	0.3208 ± 0.0078	0.3122 ± 0.0196	0.3186 ± 0.0134
	1+3	0.3167 ± 0.0113	0.3146 ± 0.0078	0.3210 ± 0.0093	0.3173 ± 0.0194	0.3237 ± 0.0208
	2+3	0.3173 ± 0.0074	0.3142 ± 0.0071	0.3182 ± 0.0045	0.3079 ± 0.0140	0.3160 ± 0.0123

Table 20: *Mean and Standard Deviation of PLV values for each frequency band (D=delta, T=theta, A=alpha, B=beta, G=gamma) and participant combination (3=excluded participant). The bold/italic values indicate the frequency band and brain region combinations for which the mean PLV value is higher for participant combination 1+2 (non-excluded participants) than both other combinations.* → ***Method M1***

		Frontal	**Central**	**Parietal**	**Temporal**	**Occipital**
D	1+2	0.3178 ± 0.0301	0.3251 ± 0.0348	0.3208 ± 0.0256	***0.3162*** ± 0.0185	0.3248 ± 0.0401
	1+3	0.3218 ± 0.0456	0.3325 ± 0.0340	0.3213 ± 0.0338	0.3087 ± 0.0241	0.3363 ± 0.0575
	2+3	0.3198 ± 0.0442	0.3133 ± 0.0249	0.3267 ± 0.0350	0.3052 ± 0.0255	0.3071 ± 0.0246
T	1+2	0.3178 ± 0.0300	0.3250 ± 0.0346	0.3210 ± 0.0254	***0.3162*** ± 0.0184	0.3248 ± 0.0398
	1+3	0.3218 ± 0.0454	0.3323 ± 0.0337	0.3213 ± 0.0336	0.3087 ± 0.0239	0.3361 ± 0.0573
	2+3	0.3197 ± 0.0438	0.3135 ± 0.0250	0.3265 ± 0.0348	0.3052 ± 0.0256	0.3071 ± 0.0245
A	1+2	0.3178 ± 0.0298	0.3248 ± 0.0341	0.3212 ± 0.0251	***0.3162*** ± 0.0181	0.3248 ± 0.0392
	1+3	0.3218 ± 0.0450	0.3317 ± 0.0330	0.3212 ± 0.0331	0.3087 ± 0.0233	0.3359 ± 0.0568
	2+3	0.3194 ± 0.0430	0.3141 ± 0.0250	0.3261 ± 0.0343	0.3050 ± 0.0256	0.3071 ± 0.0242
B	1+2	0.3176 ± 0.0286	0.3238 ± 0.0316	0.3221 ± 0.0237	***0.3159*** ± 0.0171	0.3248 ± 0.0360
	1+3	0.3221 ± 0.0426	0.3293 ± 0.0304	0.3198 ± 0.0302	0.3094 ± 0.0199	0.3347 ± 0.0536
	2+3	0.3181 ± 0.0389	0.3168 ± 0.0250	0.3246 ± 0.0311	0.3045 ± 0.0261	0.3075 ± 0.0222
G	1+2	***0.3244*** ± 0.0160	***0.3223*** ± 0.0226	0.3196 ± 0.0166	***0.3184*** ± 0.0207	0.3198 ± 0.0179
	1+3	0.3139 ± 0.0210	0.3215 ± 0.0205	0.3060 ± 0.0191	0.3161 ± 0.0234	0.3208 ± 0.0337
	2+3	0.3161 ± 0.0224	0.3204 ± 0.0252	0.3219 ± 0.0231	0.3012 ± 0.0163	0.3134 ± 0.0159

Table 21: *Mean and standard Deviation of PLV values for each frequency band (D=delta,T=theta,A=alpha,B=beta,G=gamma) and participant combination (3=excluded participant). The bold/italic values indicate the frequency band and brain region combinations for which the mean PLV value is higher for participant combination 1+2 (non-excluded participants) than both other combinations.* → ***Method M2***

Method	Freq band	Combination	Region	p-value
1	D	1+2/1+3	Central	0.32217
		1+2/2+3		0.13459
	T	1+2/1+3	Central	0.31187
		1+2/2+3		0.13229
	A	1+2/1+3	Central	0.32067
		1+2/2+3		0.13079
	B	1+2/1+3	Central	0.29977
		1+2/2+3		0.12099
	G	1+2/1+3	***Central***	***0.10129***
		1+2/2+3		***0.080092***
		1+2/1+3	Frontal	0.29857
		1+2/2+3		0.31037
2	D	1+2/1+3	Temporal	0.19888
		1+2/2+3		0.11289
	T	1+2/1+3	Temporal	0.19348
		1+2/2+3		0.11339
	A	1+2/1+3	Temporal	0.18378
		1+2/2+3		0.11169
	B	1+2/1+3	Temporal	0.19598
		1+2/2+3		0.10699
	G	1+2/1+3	Temporal	0.39106
		1+2/2+3		0.01229
		1+2/1+3	Frontal	0.077792
		1+2/2+3		0.14419
		1+2/1+3	Central	0.46815
		1+2/2+3		0.42146

Table 22: *p-values for the frequency bands and regions for which the mean PLV value from participant combination 1 and 2 is higher than both other combinations. The bold/italic value indicates the frequency band and brain region combination for which the p-value is closest to being significant for both participant combination comparisons.*

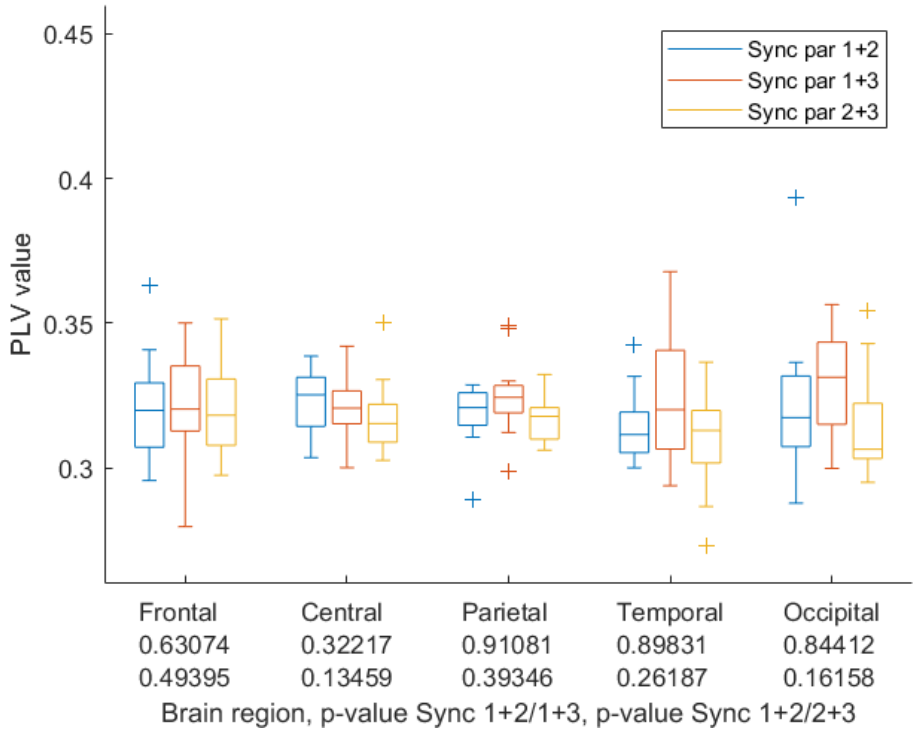


Figure 24: *Method M1: Delta frequency band*

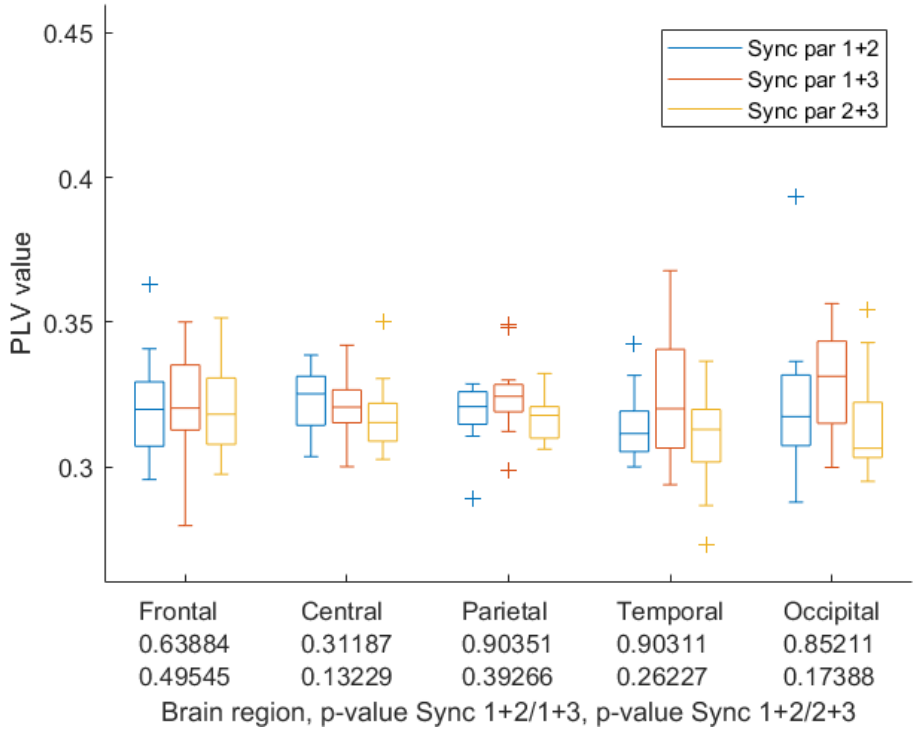


Figure 25: *Method M1: Theta frequency band*

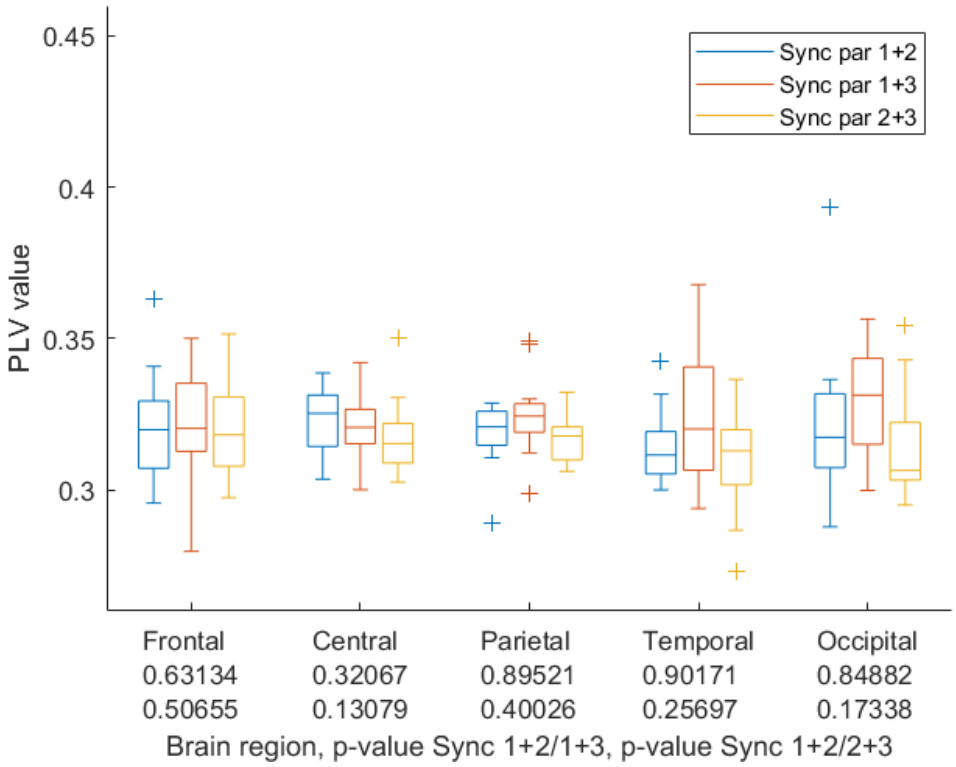


Figure 26: *Method M1: Alpha frequency band*

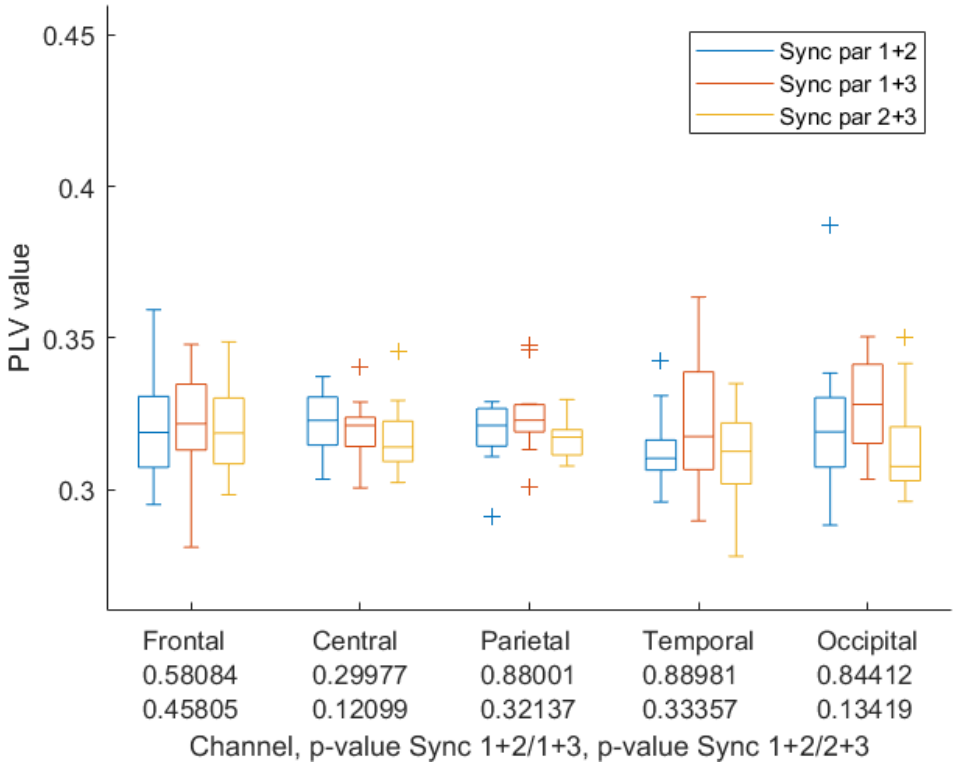


Figure 27: *Method M1: Beta frequency band*

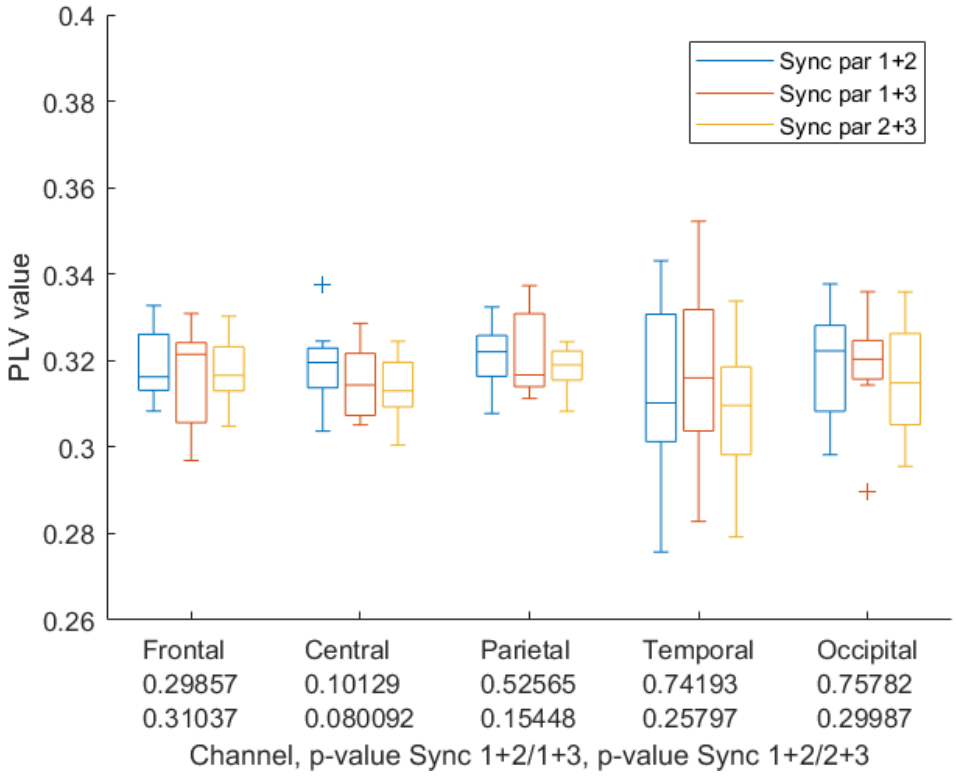


Figure 28: *Method M1: Gamma frequency band*

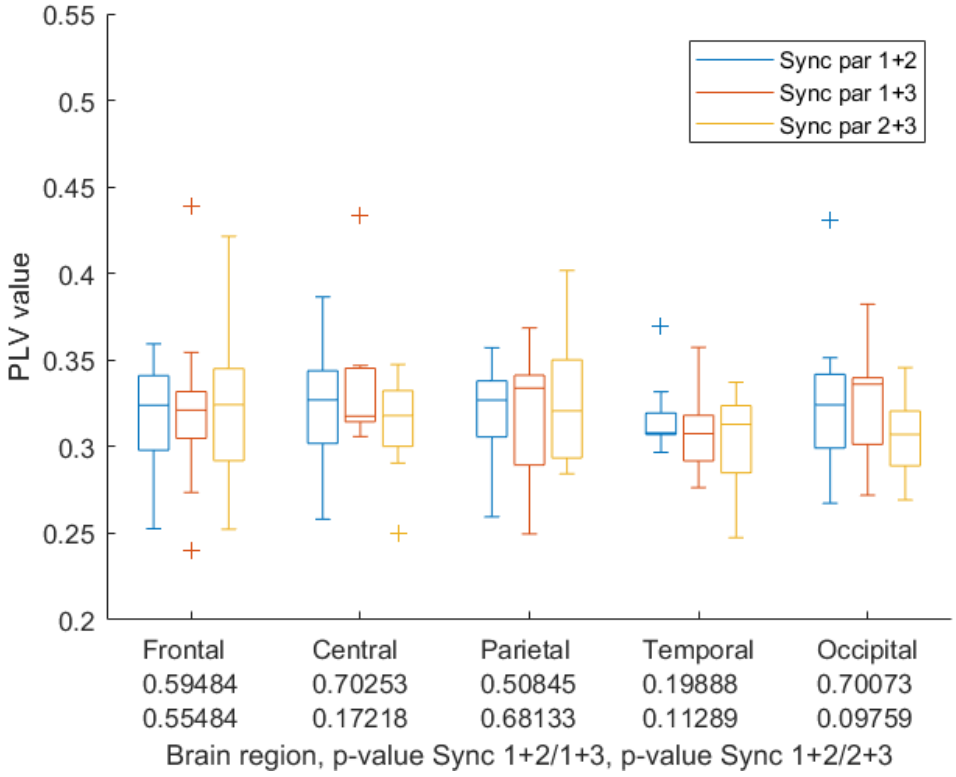


Figure 29: *Method M2: Delta frequency band*

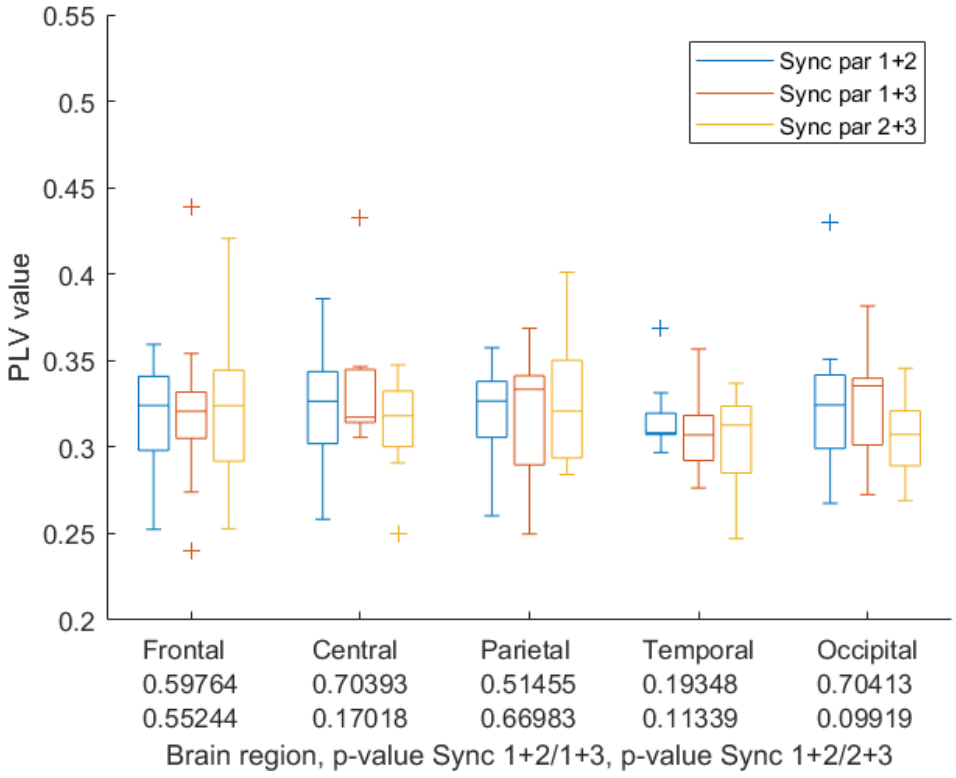


Figure 30: *Method M2: Theta frequency band*

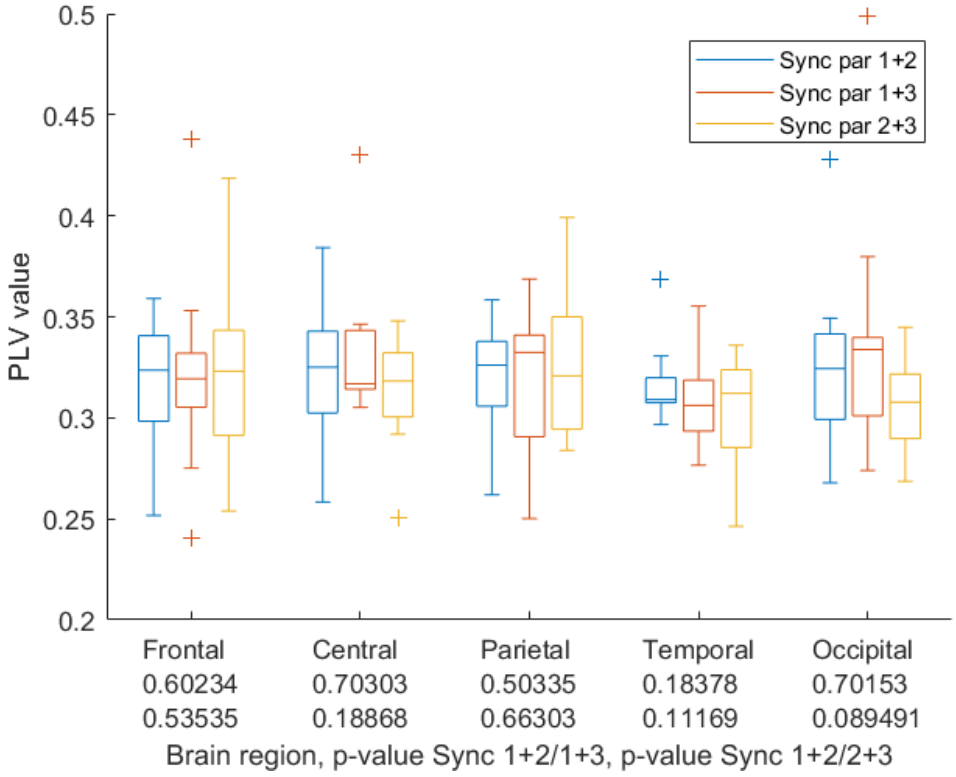


Figure 31: *Method M2: Alpha frequency band*

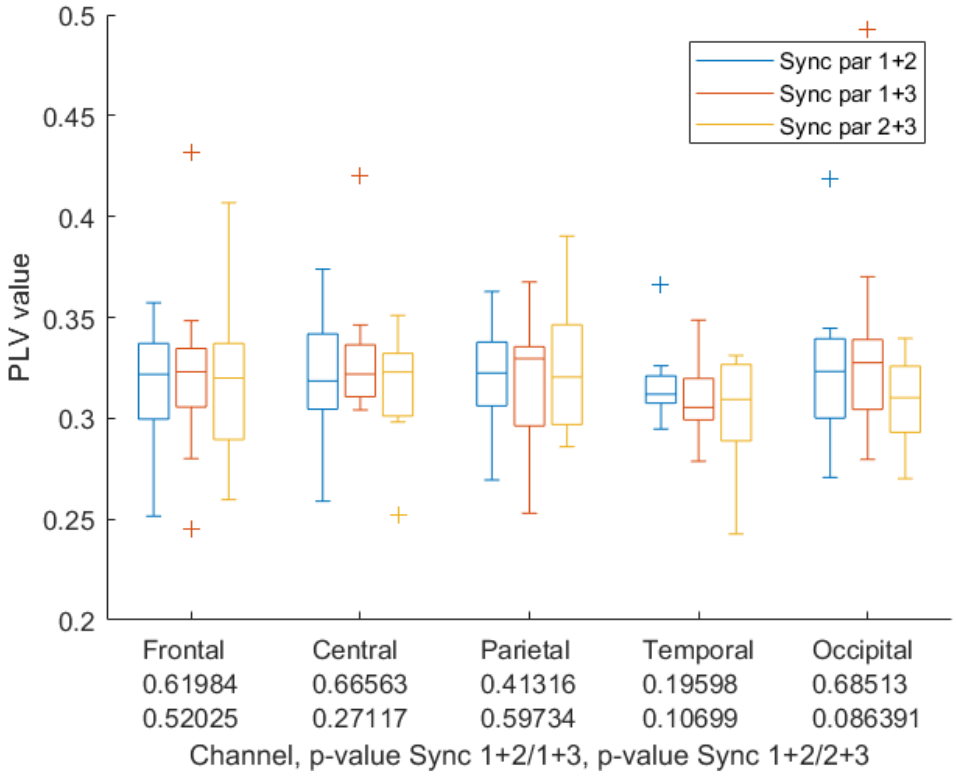


Figure 32: *Method M2: Beta frequency band*

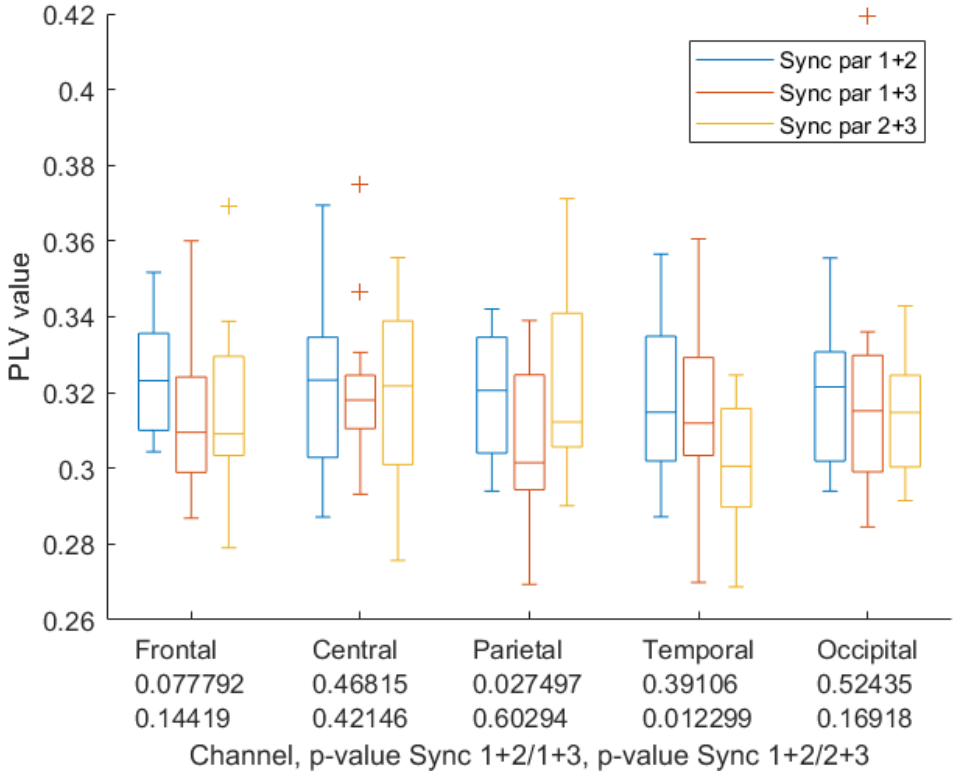


Figure 33: *Method M2: Gamma frequency band*

B.2 Channel Level Without Baseline Correction

Freq band	Combination	Channel	p-value
D	1+2/1+3	5 (FC1)	0.68123
	1+2/2+3		0.023898
	1+2/1+3	21 (CP6)	0.037796
	1+2/2+3		0.21478
	1+2/1+3	***23 (C4)***	***0.016898***
	1+2/2+3		***0.021698***
	1+2/1+3	29 (AF4)	0.66843
	1+2/2+3		0.028897
T	1+2/1+3	5 (FC1)	0.69543
	1+2/2+3		0.021798
	1+2/1+3	21 (CP6)	0.035396
	1+2/2+3		0.21728
	1+2/1+3	***23 (C4)***	***0.019198***
	1+2/2+3		***0.019598***
	1+2/1+3	29 (AF4)	0.62614
	1+2/2+3		0.024698
A	1+2/1+3	5 (FC1)	0.69023
	1+2/2+3		0.023598
	1+2/1+3	21 (CP6)	0.039096
	1+2/2+3		0.21188
	1+2/1+3	***23 (C4)***	***0.020298***
	1+2/2+3		***0.021498***
	1+2/1+3	29 (AF4)	0.61864
	1+2/2+3		0.024798
B	1+2/1+3	5 (FC1)	0.62514
	1+2/2+3		0.021098
	1+2/1+3	16 (Oz)	0.47145
	1+2/2+3		0.047195
	1+2/1+3	21 (CP6)	0.045795
	1+2/2+3		0.19288
	1+2/1+3	***23 (C4)***	***0.020298***
	1+2/2+3		***0.021598***
	1+2/1+3	29 (AF4)	0.61114
	1+2/2+3		0.018598
G	1+2/1+3	19 (P4)	0.12719
	1+2/2+3		0.017098
	1+2/1+3	26 (FC2)	0.037696
	1+2/2+3		0.058394
	1+2/1+3	29 (AF4)	0.43766
	1+2/2+3		0.026897

Table 23: *p-values for the frequency bands and channels for which the mean PLV value from participant combination 1 and 2 is higher than both other combinations. The bold/italic values indicate the chan- nels for which both comparisons are statistically signi cant.*

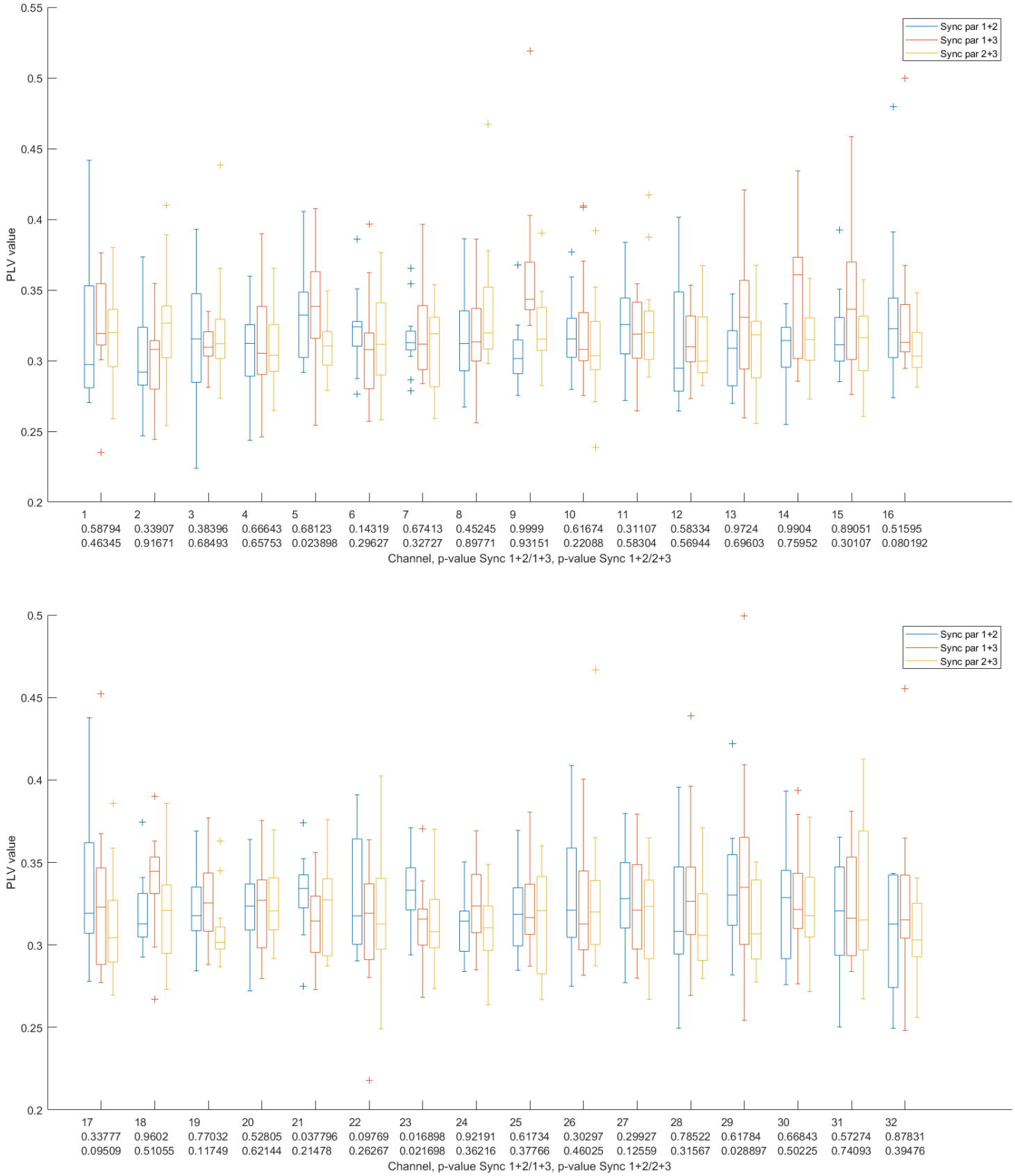


Figure 34: *All channels: Delta frequency band*

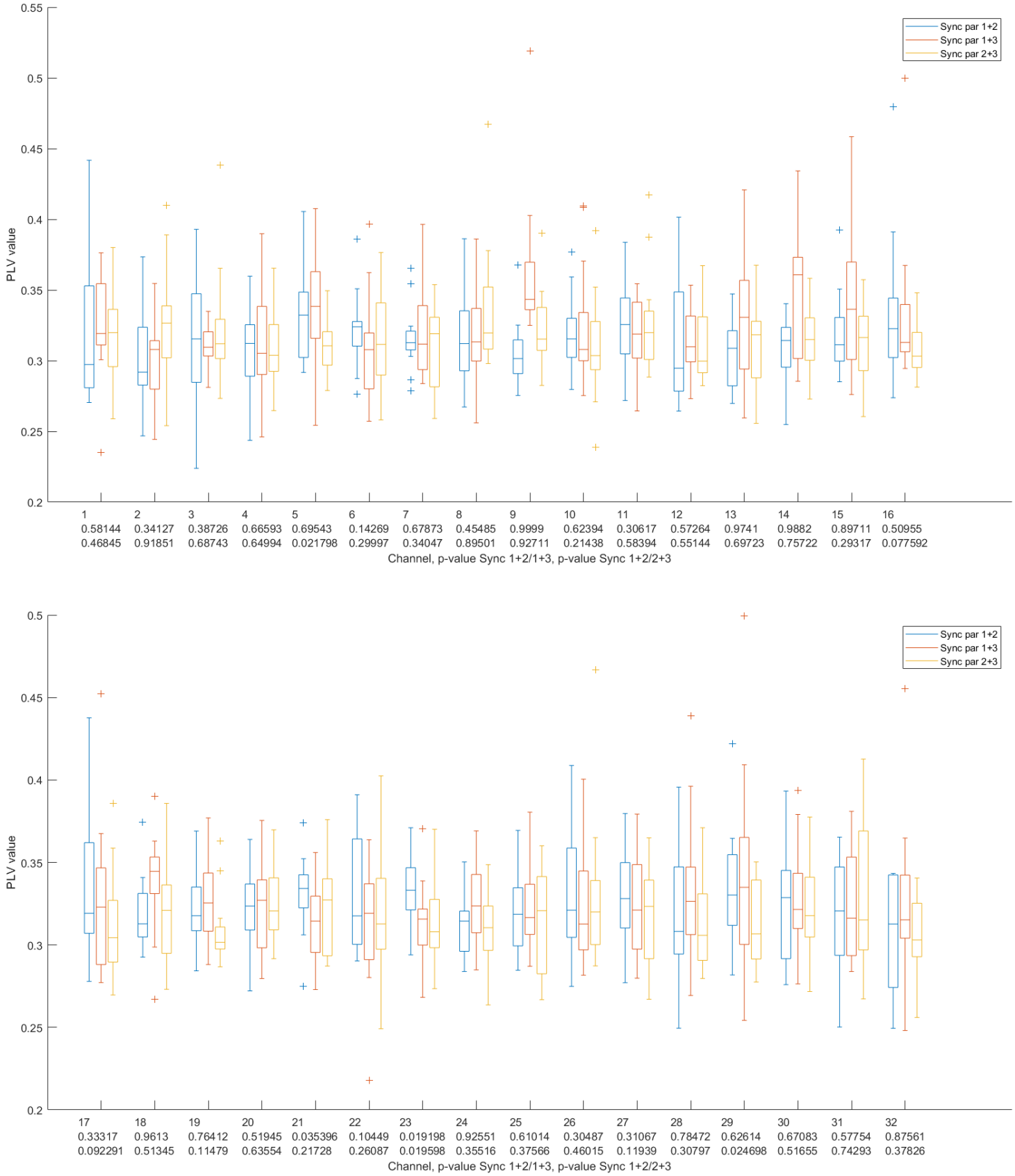
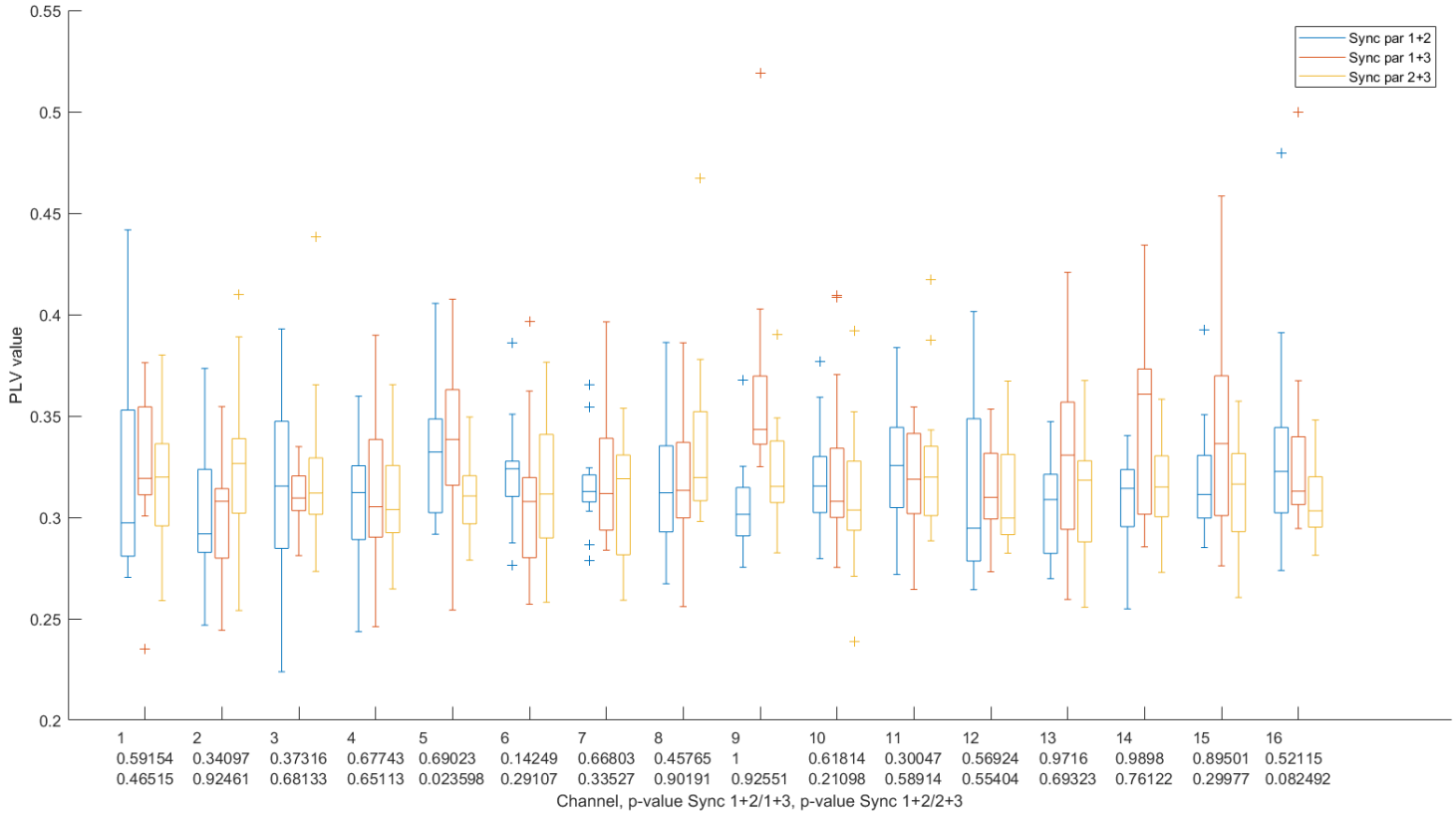


Figure 35: *All channels: Theta frequency band*



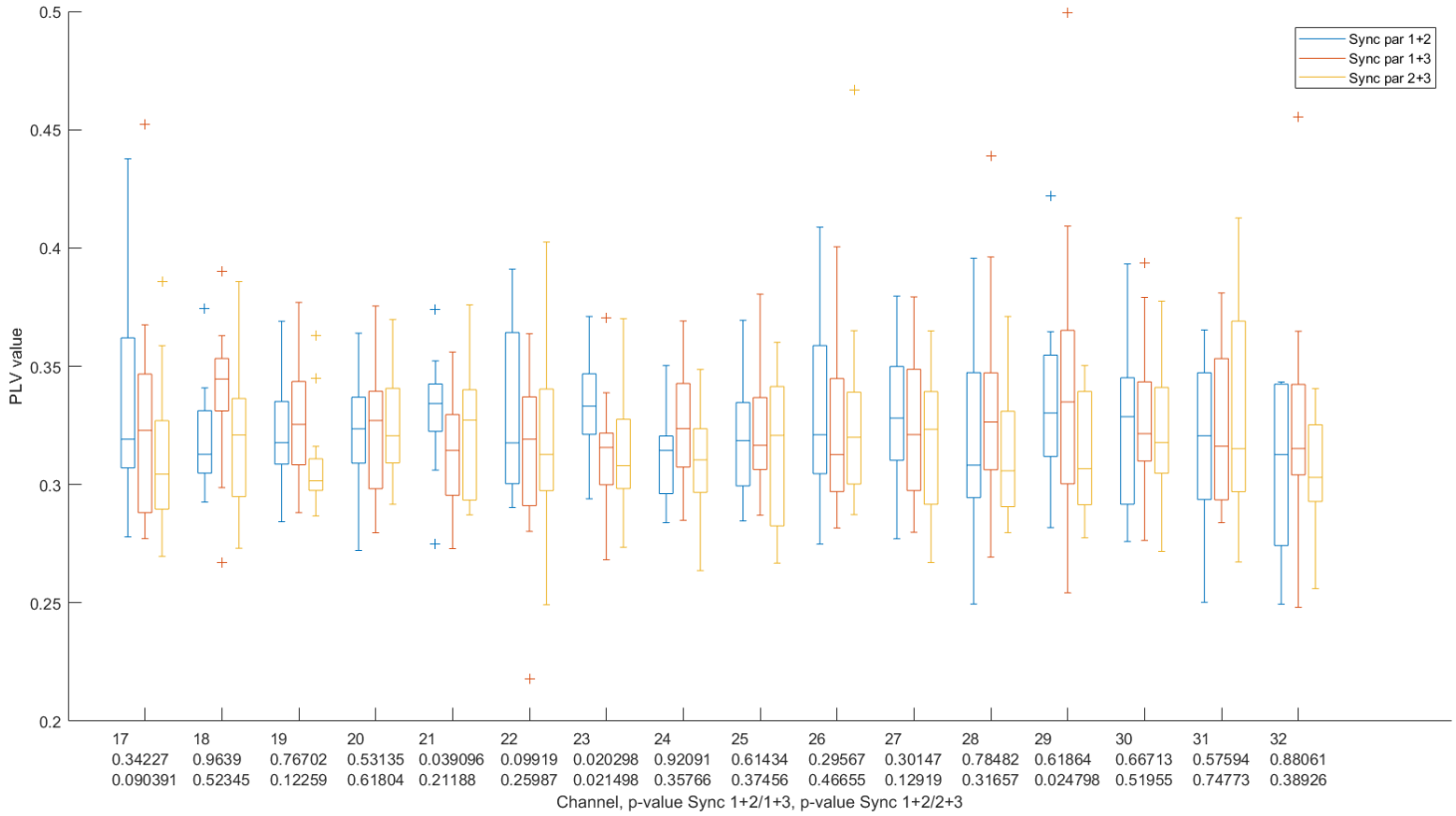


Figure 36: *All channels: Alpha frequency band*

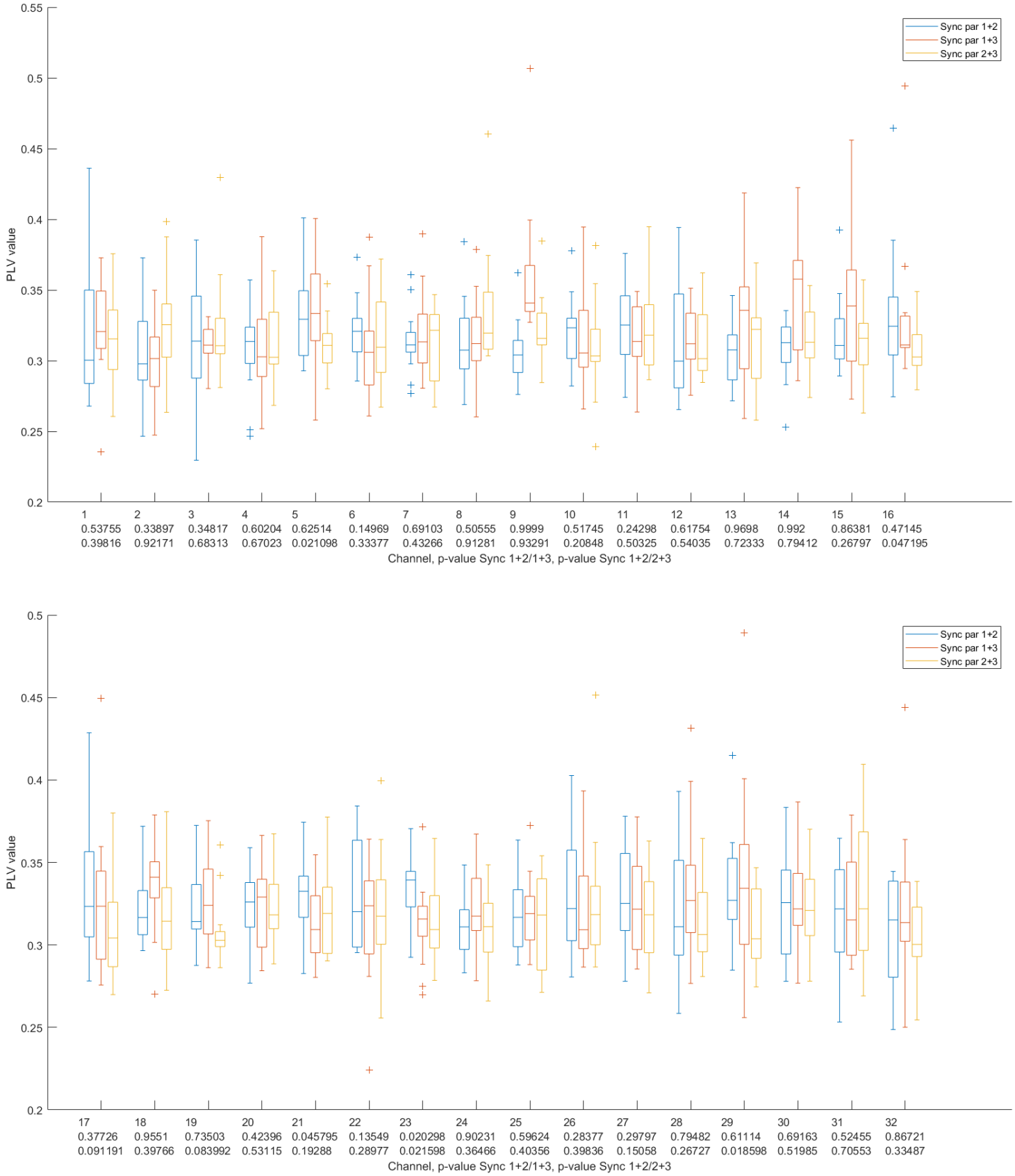


Figure 37: *All channels: Beta frequency band*

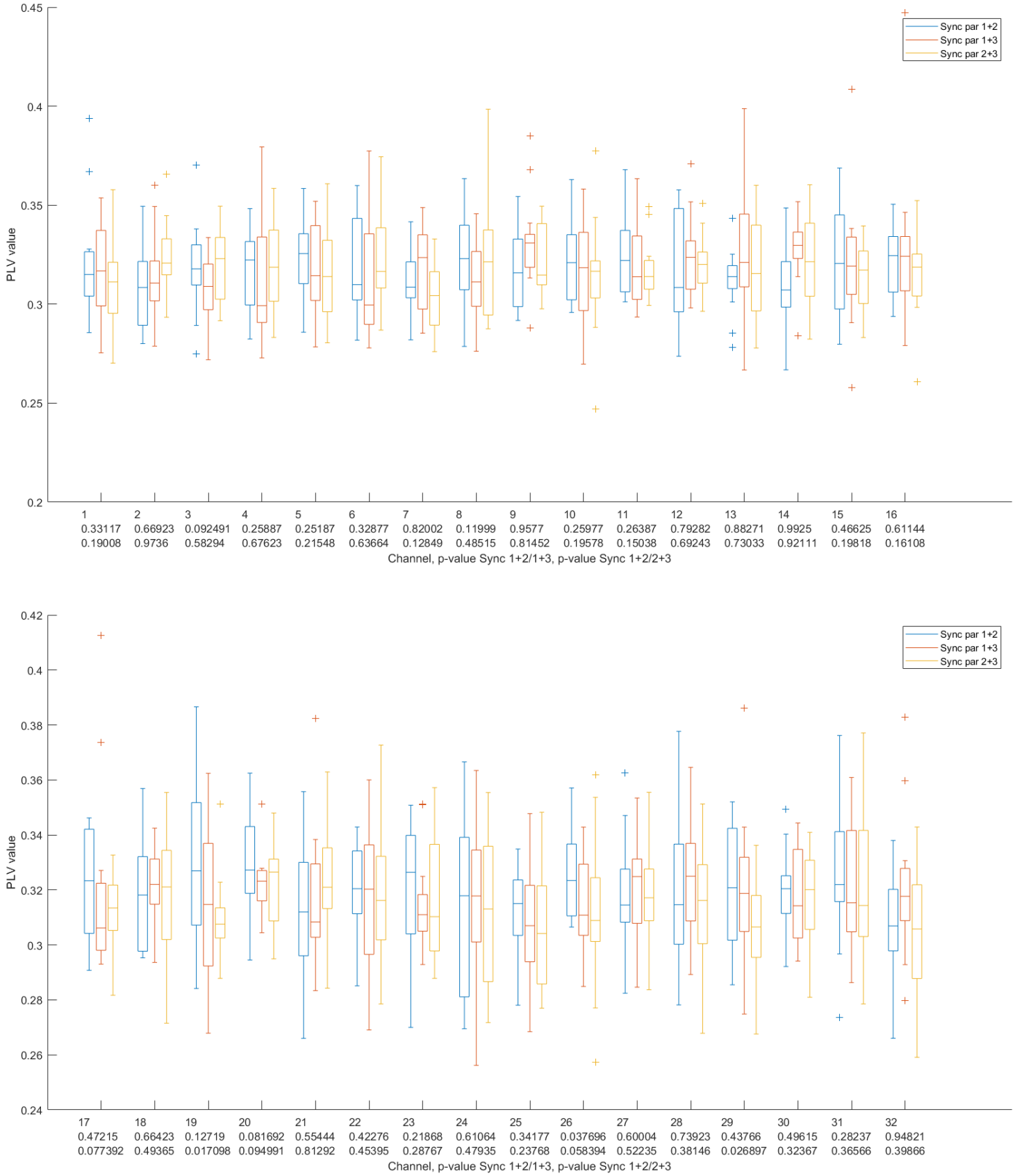


Figure 38: *All channels: Gamma frequency band*

B.3 Regional Level Without Baseline Correction: Subsets

Subset		Frontal	Central	Parietal	Temporal	Occipital
all	1+2	***0.3187*** ± 0.0077	***0.3187*** ± 0.0083	0.3208 ± 0.0078	0.3122 ± 0.0196	0.3186 ± 0.0134
	1+3	0.3167 ± 0.0113	0.3146 ± 0.0078	0.3210 ± 0.0093	0.3173 ± 0.0194	0.3237 ± 0.0208
	2+3	0.3173 ± 0.0074	0.3142 ± 0.0071	0.3182 ± 0.0045	0.3079 ± 0.0140	0.3160 ± 0.0123
VE	1+2	***0.3239*** ± 0.0095	***0.3187*** ± 0.0026	***0.3191*** ± 0.0086	0.3103 ± 0.0247	0.3197 ± 0.0168
	1+3	0.3178 ± 0.0142	0.3179 ± 0.0081	0.3161 ± 0.0059	0.3248 ± 0.0224	0.3253 ± 0.0075
	2+3	0.3187 ± 0.0060	0.3113 ± 0.0081	0.3167 ± 0.0037	0.3006 ± 0.0146	0.3119 ± 0.0111
E	1+2	***0.3198*** ± 0.0079	***0.3197*** ± 0.0076	***0.3219*** ± 0.0078	0.3110 ± 0.0200	0.3201 ± 0.0130
	1+3	0.3192 ± 0.0104	0.3154 ± 0.0080	0.3219 ± 0.0100	0.3206 ± 0.0179	0.3246 ± 0.0226
	2+3	0.3162 ± 0.0069	0.3143 ± 0.0073	0.3174 ± 0.0045	0.3064 ± 0.0126	0.3166 ± 0.0133
EC	1+2	***0.3233*** ± 0.0083	0.3161 ± 0.0046	0.3177 ± 0.0076	0.3043 ± 0.0214	0.3202 ± 0.0130
	1+3	0.3216 ± 0.0126	0.3179 ± 0.0092	0.3183 ± 0.0075	0.3229 ± 0.0232	0.3339 ± 0.0252
	2+3	0.3189 ± 0.0052	0.3146 ± 0.0082	0.3180 ± 0.0035	0.3037 ± 0.0136	0.3173 ± 0.0130

Table 24: *Mean and standard deviation of PLV values per brain regions for each subset (VE=Very Excluded, E=Excluded, EC=Excluded+Connected) and each participant combination (3=excluded participant). Results are only shown from the gamma frequency band with Method M1. The bold/italic values indicate the frequency band and brain region combinations for which the mean PLV value is higher for participant combination 1+2 (non-excluded participants) than both other combinations.*

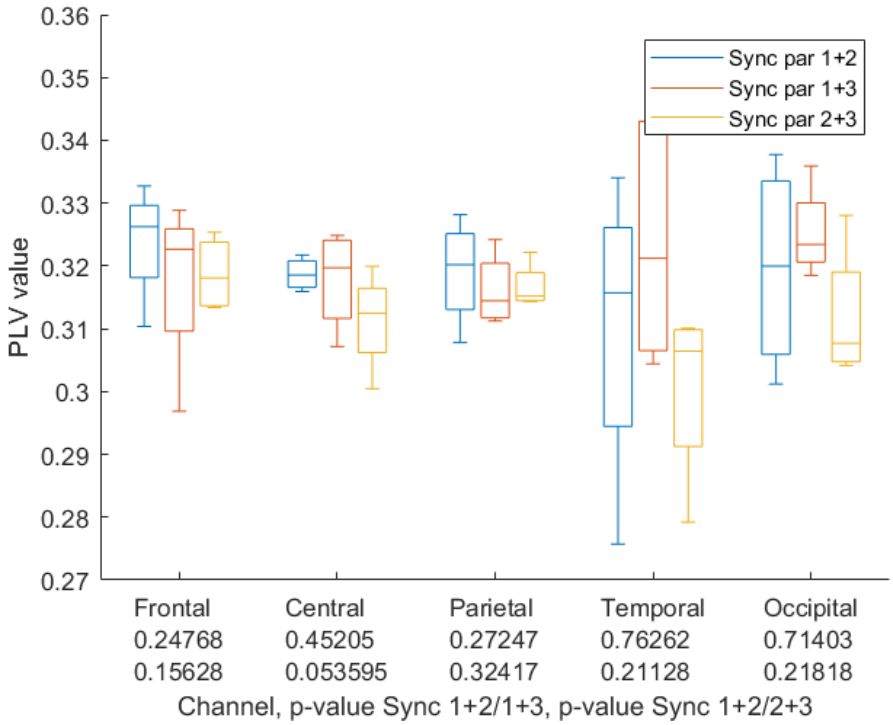


Figure 39: *Subset VE with gamma frequency band: all experiments in which the excluded participant actually felt very excluded.*

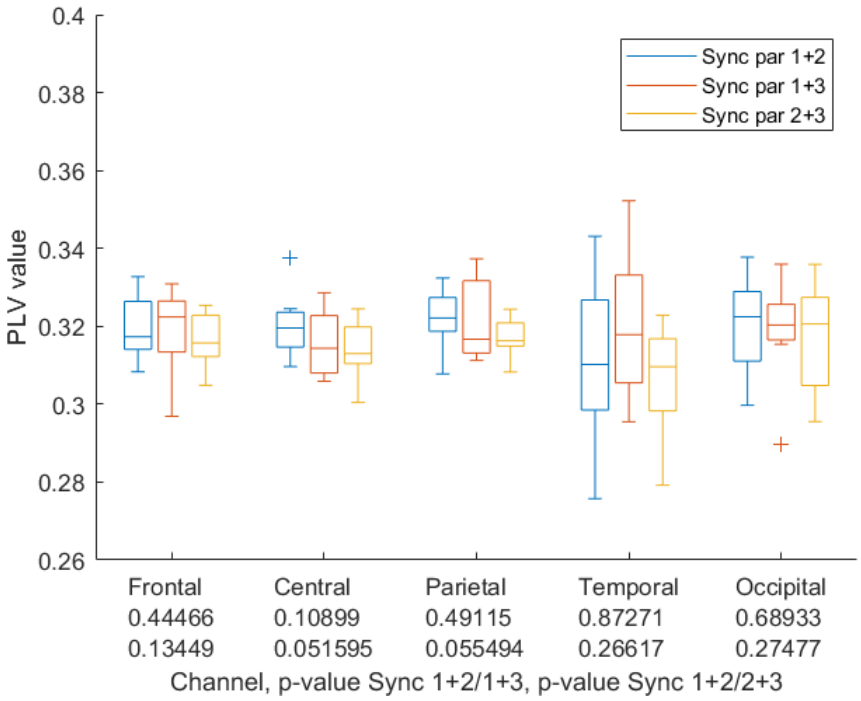


Figure 40: *Subset E with gamma frequency band: all experiments in which the excluded participant actually felt (very) excluded. Best performing combination for central brain region.*

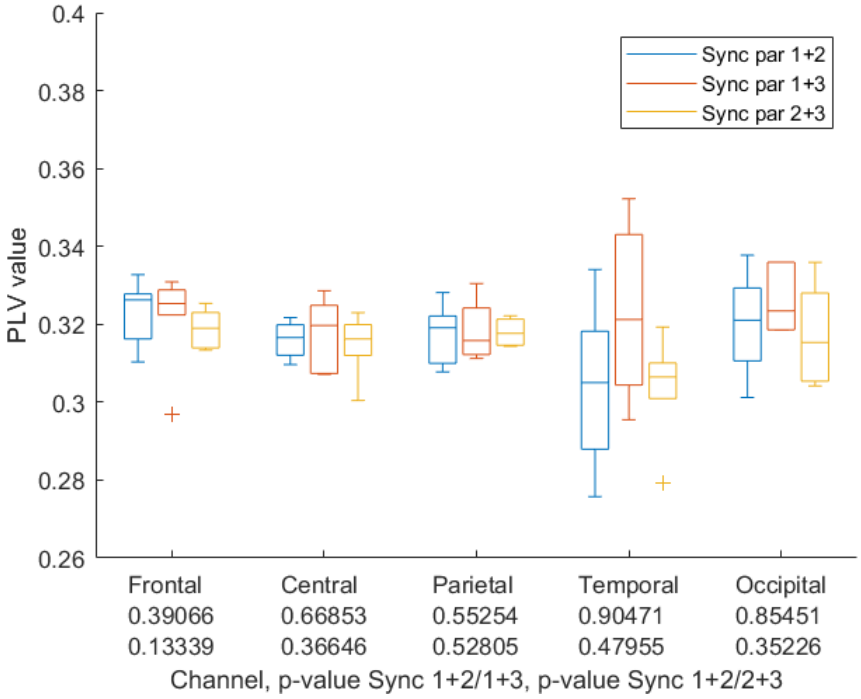


Figure 41: *Subset EC with gamma frequency band: all experiments in which the excluded participant actually felt (very) excluded and the other two participants felt most connected to each other.*

B.4 Channel Level Without Baseline Correction: Subsets

Subset	Combination	Channel	p-value
all	1+2/1+3	5 (FC1)	0.68123
	1+2/2+3		0.023898
	1+2/1+3	21 (CP6)	0.037796
	1+2/2+3		0.21478
	1+2/1+3	***23 (C4)***	***0.016898***
	1+2/2+3		***0.021698***
	1+2/1+3	29 (AF4)	0.66843
	1+2/2+3		0.028897
VE	1+2/1+3	5 (FC1)	0.51665
	1+2/2+3		0.0050995
	1+2/1+3	16 (Oz)	0.20428
	1+2/2+3		0.035396
	1+2/1+3	17 (O2)	0.015098
	1+2/2+3		0.073093
	1+2/1+3	18 (PO4)	0.53225
	1+2/2+3		0.016398
	1+2/1+3	23 (C4)	0.21618
	1+2/2+3		0.043096
	1+2/1+3	29 (AF4)	0.42206
	1+2/2+3		0.0082992
E	1+2/1+3	5 (FC1)	0.75212
	1+2/2+3		0.026597
	1+2/1+3	6 (FC5)	0.052395
	1+2/2+3		0.28217
	1+2/1+3	16 (Oz)	0.46585
	1+2/2+3		0.046695
	1+2/1+3	19 (P4)	0.66233
	1+2/2+3		0.016998
	1+2/1+3	23 (C4)	0.068093
	1+2/2+3		0.035296
	1+2/1+3	29 (AF4)	0.74453
	1+2/2+3		0.010699
EC	1+2/1+3	5 (FC1)	0.71673
	1+2/2+3		0.00072993
	1+2/1+3	23 (C4)	0.076592
	1+2/2+3		0.0060994
	1+2/1+3	29 (AF4)	0.80492
	1+2/2+3		0.024098

Table 25: *p-values for the delta frequency band and channels for which the mean PLV value from participant combination 1 and 2 (non-excluded participants) is higher than both other combinations (3=excluded participant). Significance values are shown for each sub-set (VE=Very Excluded, E=Excluded, EC=Excluded+Connected). The bold/italic value indicate best performing subset and brain region combination.*

Subset	Combination	Channel	p-value
all	1+2/1+3	5 (FC1)	0.69543
	1+2/2+3		0.021798
	1+2/1+3	21 (CP6)	0.035396
	1+2/2+3		0.21728
	1+2/1+3	***23 (C4)***	***0.019198***
	1+2/2+3		***0.019598***
	1+2/1+3	29 (AF4)	0.62614
	1+2/2+3		0.024698
1	1+2/1+3	5 (FC1)	0.51515
	1+2/2+3		0.0038996
	1+2/1+3	16 (Oz)	0.19478
	1+2/2+3		0.033797
	1+2/1+3	17 (O2)	0.016398
	1+2/2+3		0.070093
	1+2/1+3	18 (PO4)	0.52815
	1+2/2+3		0.013999
	1+2/1+3	23 (C4)	0.21868
	1+2/2+3		0.043596
	1+2/1+3	29 (AF4)	0.42256
	1+2/2+3		0.010099
2	1+2/1+3	5 (FC1)	0.74363
	1+2/2+3		0.027897
	1+2/1+3	16 (Oz)	0.46145
	1+2/2+3		0.049995
	1+2/1+3	19 (P4)	0.67923
	1+2/2+3		0.015498
	1+2/1+3	23 (C4)	0.070193
	1+2/2+3		0.041196
	1+2/1+3	29 (AF4)	0.74843
	1+2/2+3		0.010399
3	1+2/1+3	5 (FC1)	0.71283
	1+2/2+3		0.010499
	1+2/1+3	23 (C4)	0.076492
	1+2/2+3		0.0058994
	1+2/1+3	29 (AF4)	0.80342
	1+2/2+3		0.023498

Table 26: *p-values for the theta frequency band and channels for which the mean PLV value from participant combination 1 and 2 (non-excluded participants) is higher than both other combinations (3=excluded participant). Significance values are shown for each sub-set (VE=Very Excluded, E=Excluded, EC=Excluded+Connected). The bold/italic value indicate best performing subset and brain region combination.*

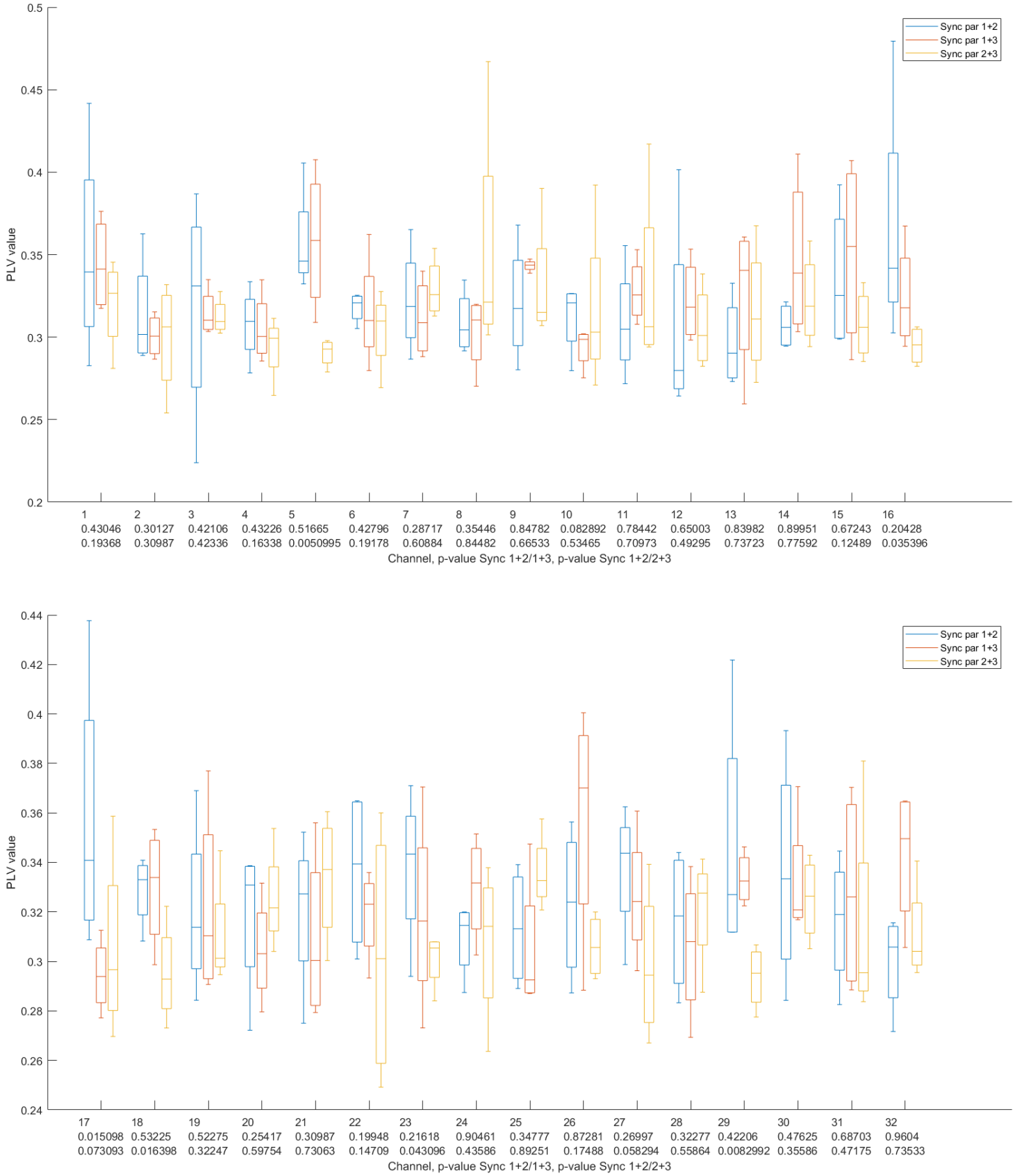


Figure 42: *Delta frequency band: All channels, subset VE*

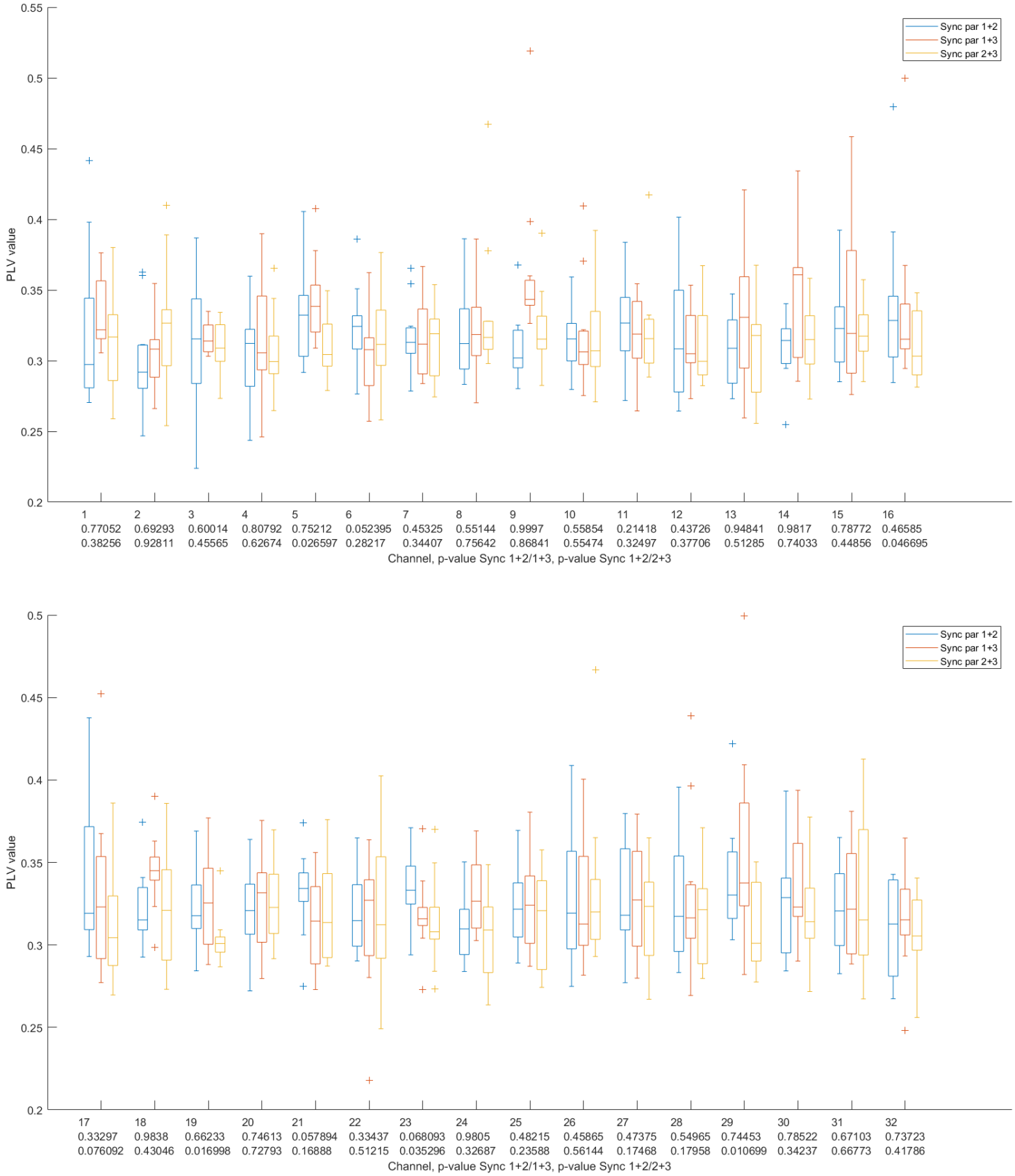


Figure 43: *Delta frequency band: All channels, subset E*

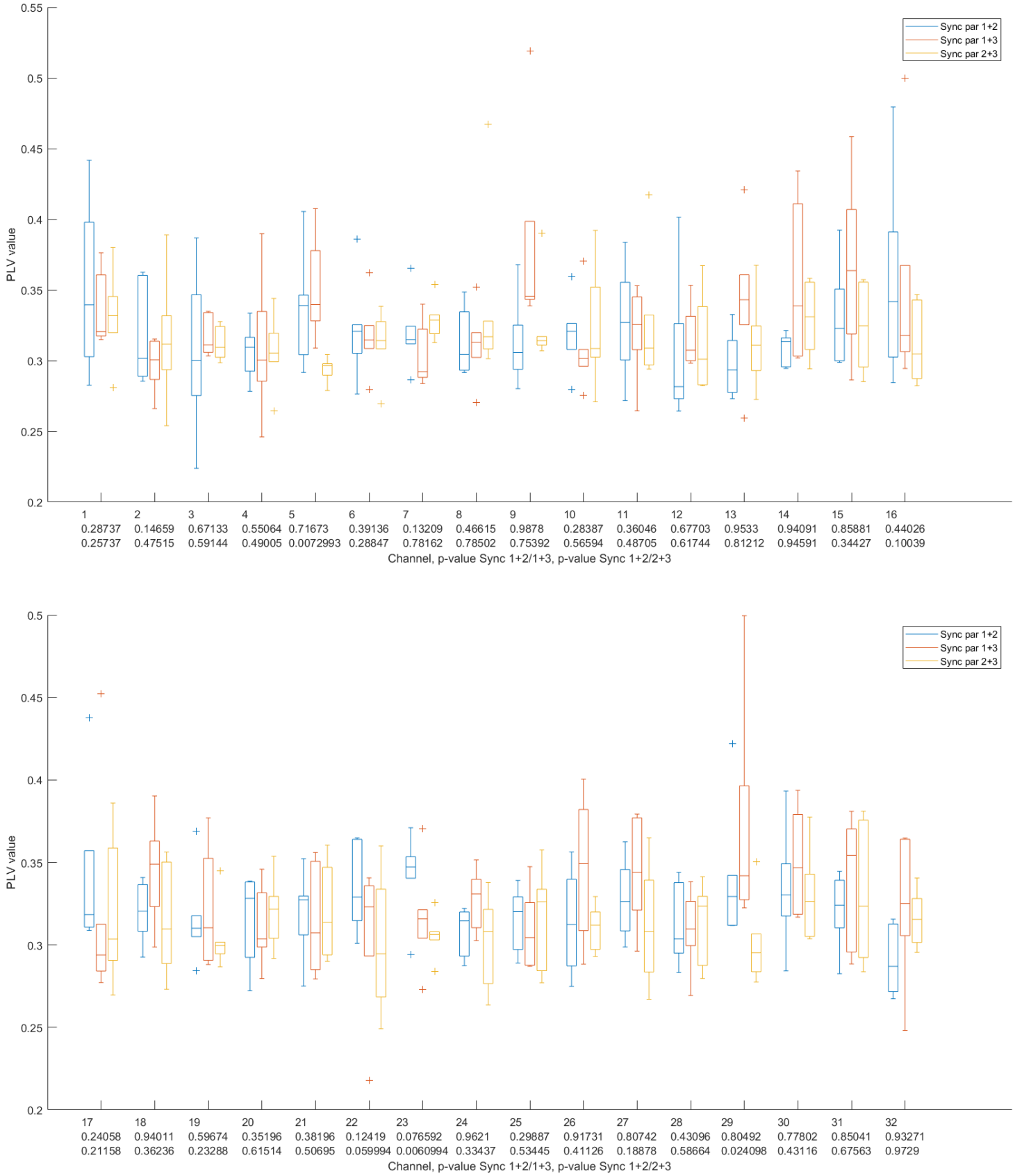
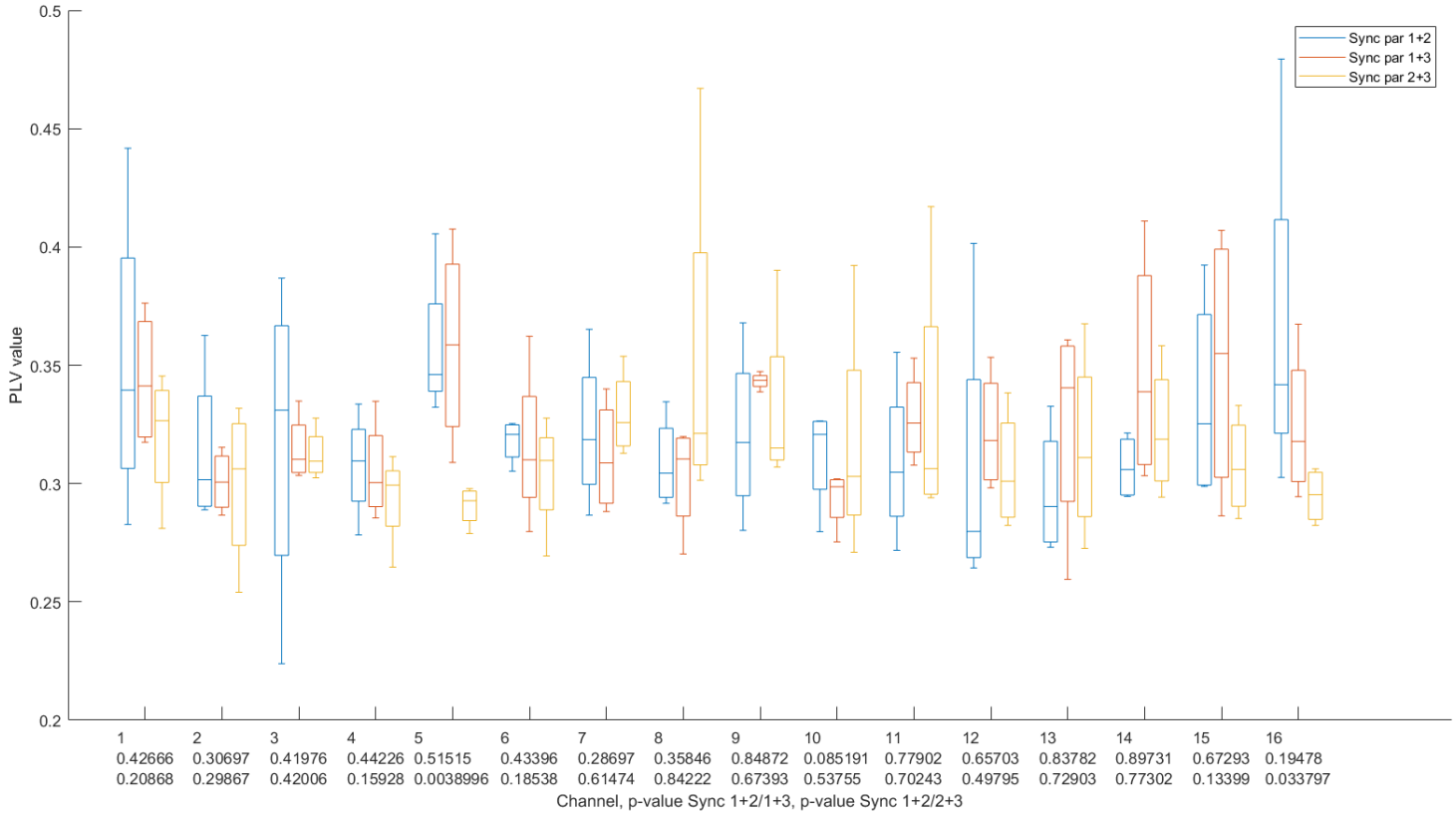


Figure 44: *Delta frequency band: All channels, subset EC*



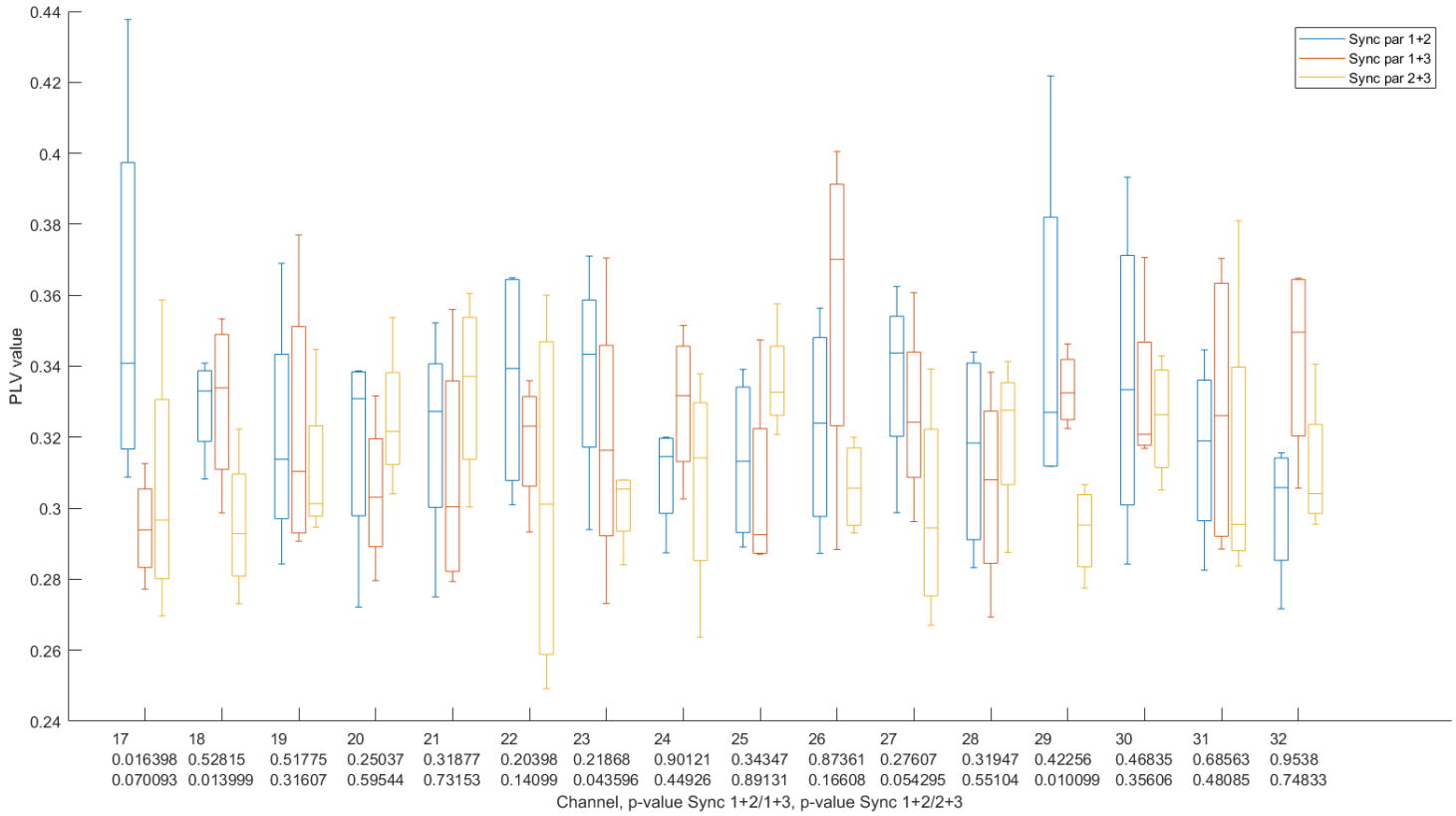


Figure 45: *Theta frequency band: All channels, subset VE*

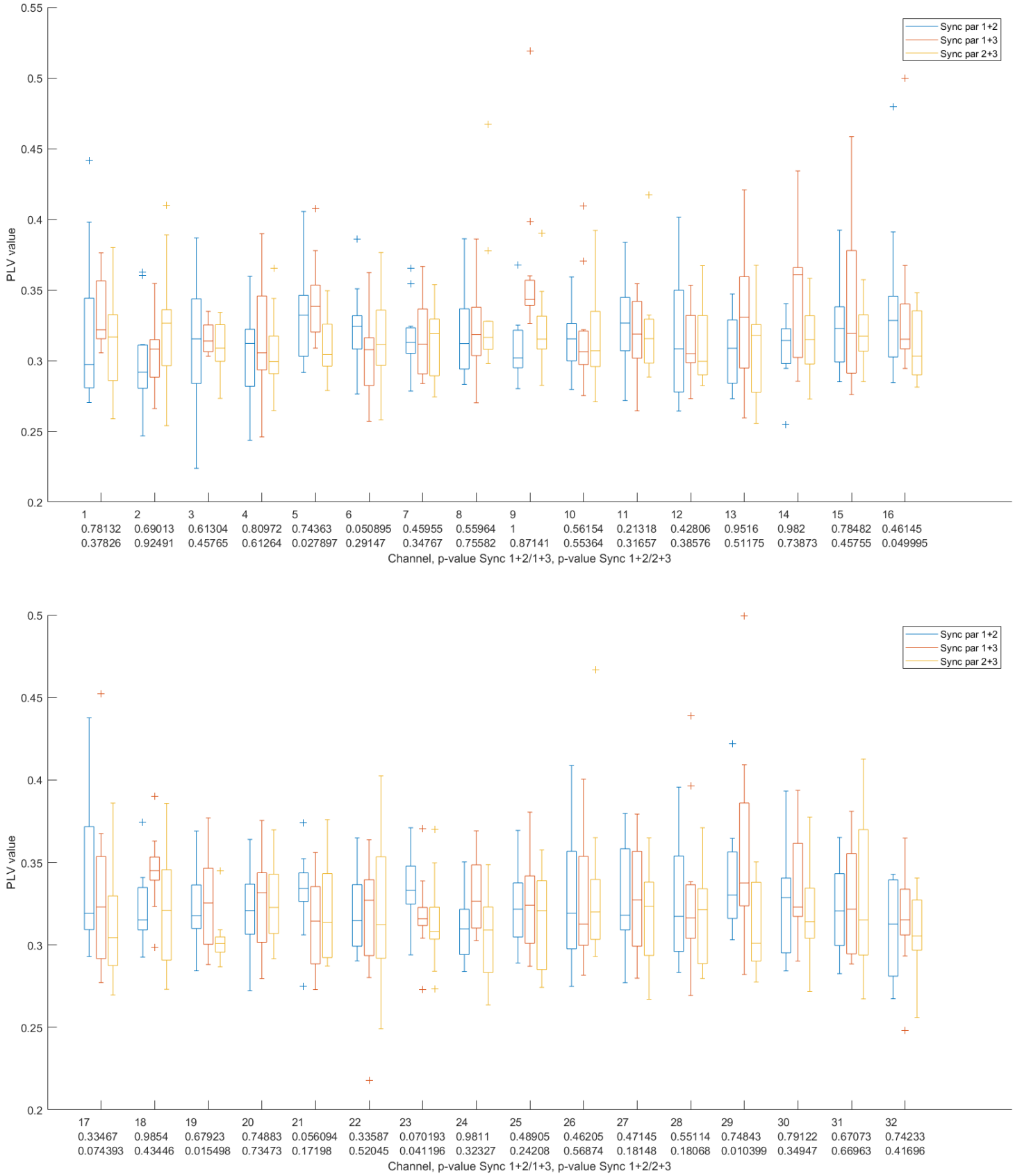


Figure 46: *Theta frequency band: All channels, subset E*

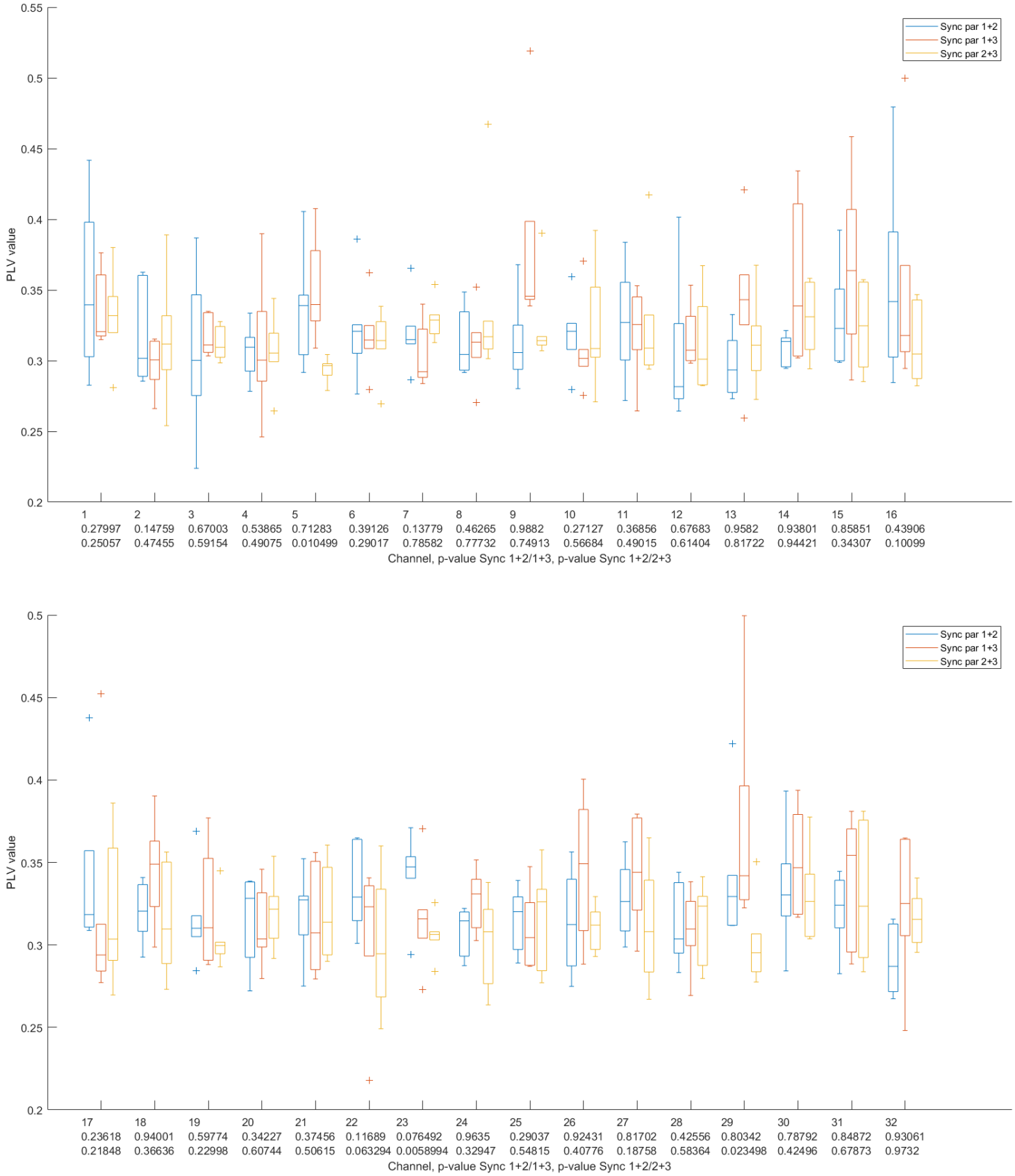


Figure 47: *Theta frequency band: All channels, subset EC*

B.5 Regional Level With Baseline Correction

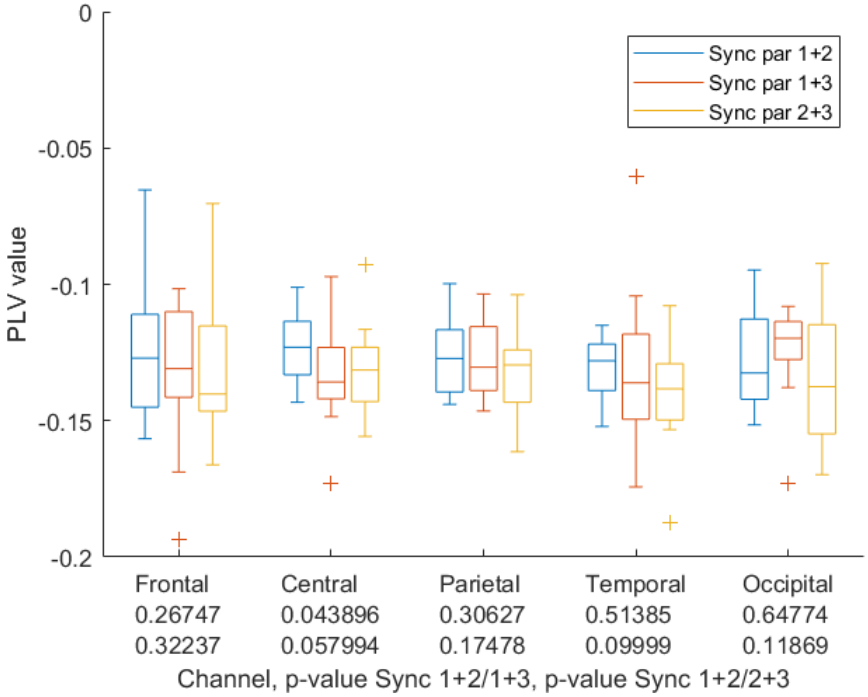


Figure 48: *Method M1: Delta frequency band. Best performing combination for central brain region after baseline correction.*

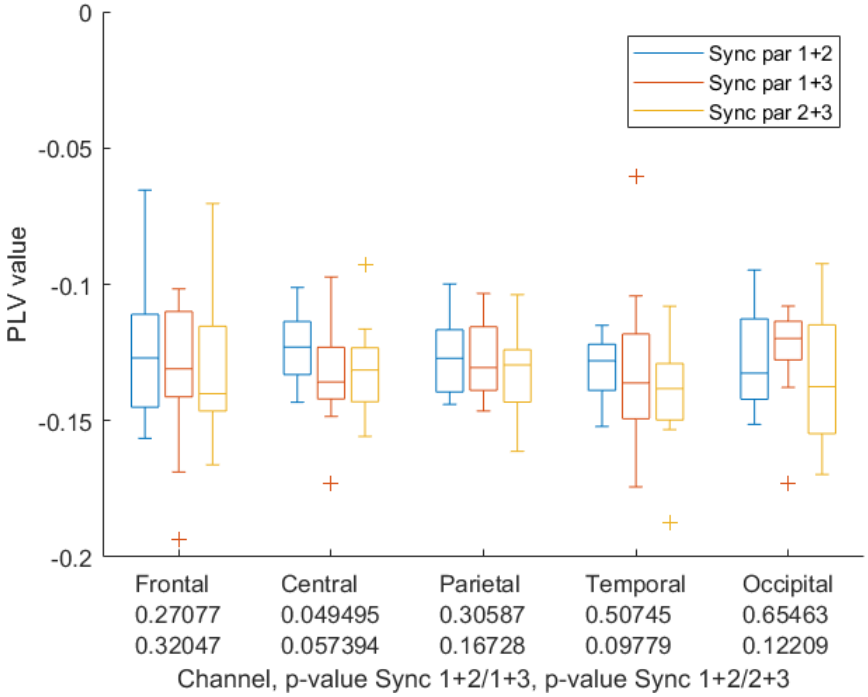


Figure 49: *Method M1: Theta frequency band. Best performing combination for central brain region after baseline correction.*

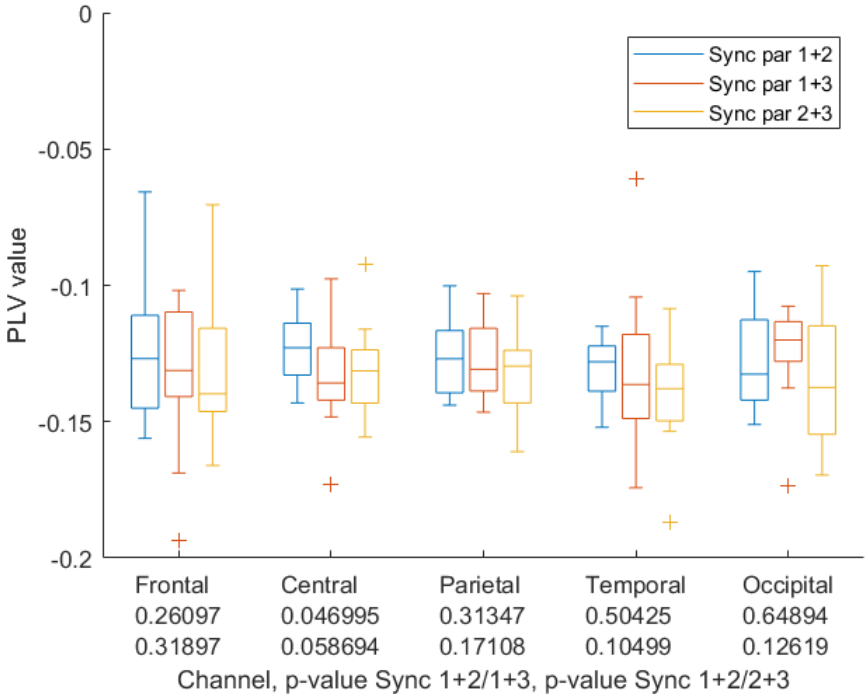


Figure 50: *Method M1: Alpha frequency band. Best performing combination for central brain region after baseline correction.*

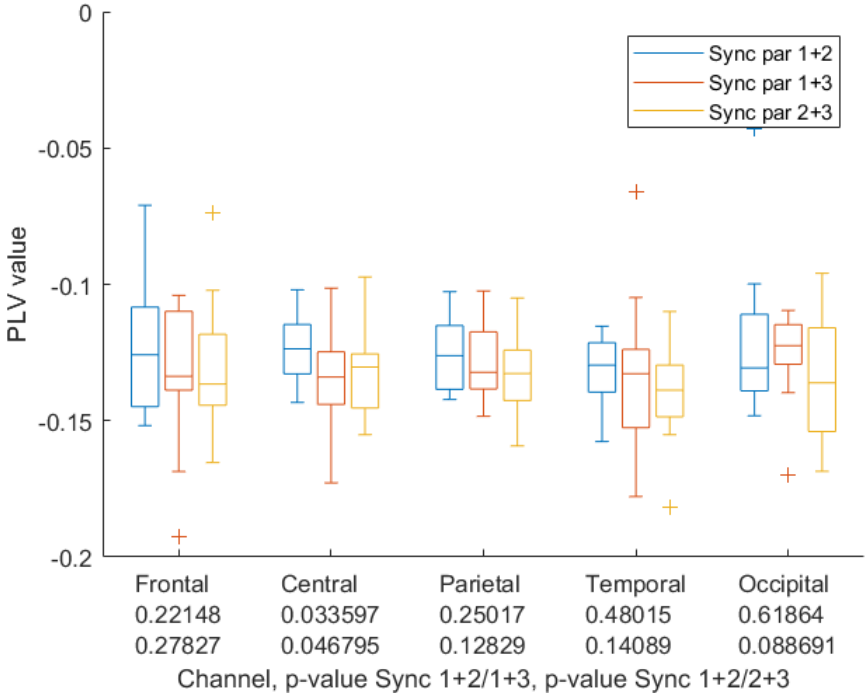


Figure 51: *Method M1: Beta frequency band. Best performing combination for central brain region after baseline correction.*

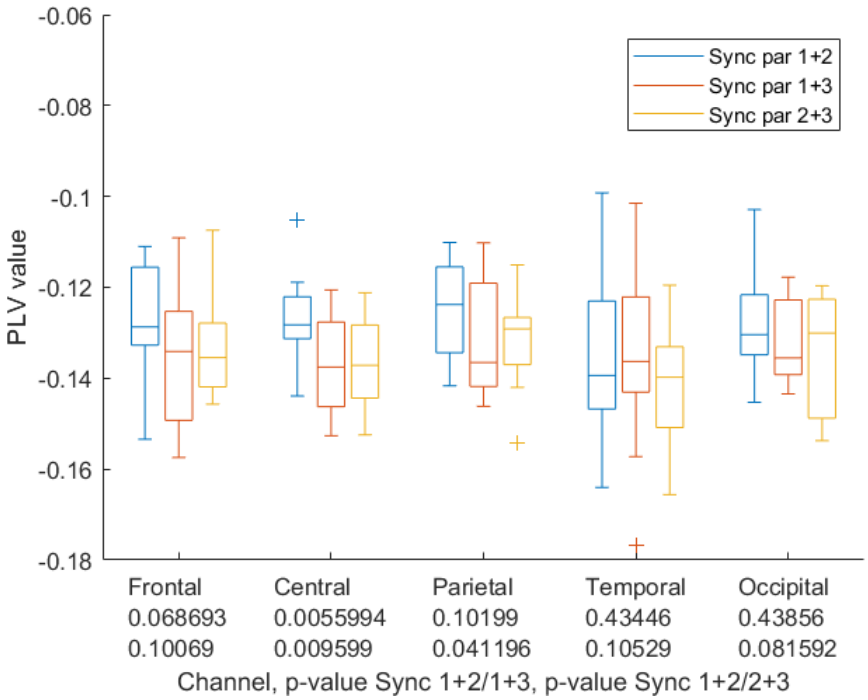


Figure 52: *Method M1: Gamma frequency band. Best performing combination for central brain region after baseline correction.*

B.6 Channel Level With Baseline Correction

Freq band	Combination	Channel	p-value
D	1+2/1+3	5 (FC1)	0.45455
	1+2/2+3		0.010699
	1+2/1+3	6 (FC5)	0.042696
	1+2/2+3		0.18208
	1+2/1+3	19 (P4)	0.41666
	1+2/2+3		0.044596
	1+2/1+3	21 (CP6)	0.0032997
	1+2/2+3		0.13999
	1+2/1+3	22 (CP2)	0.048895
	1+2/2+3		0.20118
	1+2/1+3	***23 (C4)***	***0.010399***
	1+2/2+3		***0.0065993***
	1+2/1+3	29 (AF4)	0.44946
	1+2/2+3		0.019198
T	1+2/1+3	5 (FC1)	0.45115
	1+2/2+3		0.011399
	1+2/1+3	6 (FC5)	0.043996
	1+2/2+3		0.18048
	1+2/1+3	19 (P4)	0.41686
	1+2/2+3		0.040796
	1+2/1+3	21 (CP6)	0.0035996
	1+2/2+3		0.14199
	1+2/1+3	22 (CP2)	0.042996
	1+2/2+3		0.20688
	1+2/1+3	***23 (C4)***	***0.0083992***
	1+2/2+3		***0.0068993***
	1+2/1+3	29 (AF4)	0.45895
	1+2/2+3		0.016598
A	1+2/1+3	5 (FC1)	0.45475
	1+2/2+3		0.012999
	1+2/1+3	6 (FC5)	0.043696
	1+2/2+3		0.17728
	1+2/1+3	19 (P4)	0.42176
	1+2/2+3		0.044096
	1+2/1+3	21 (CP6)	0.0035996
	1+2/2+3		0.13889
	1+2/1+3	22 (CP2)	0.046895
	1+2/2+3		0.20588
	1+2/1+3	***23 (C4)***	***0.0089991***
	1+2/2+3		***0.0064994***
	1+2/1+3	29 (AF4)	0.44836
	1+2/2+3		0.020698

Table 27: *p-values for the delta, theta and alpha frequency bands and channels for which the mean PLV value from participant combination 1 and 2 is higher than both other combinations. With baseline correction.*

Freq band	**Combination**	**Channel**	**p-value**
B	1+2/1+3	5 (FC1)	0.37456
	1+2/2+3		0.009599
	1+2/1+3	6 (FC5)	0.040296
	1+2/2+3		0.19518
	1+2/1+3	16 (Oz)	0.29377
	1+2/2+3		0.028297
	1+2/1+3	19 (P4)	0.36706
	1+2/2+3		0.028297
	1+2/1+3	21 (CP6)	0.0054995
	1+2/2+3		0.12529
	1+2/1+3	***23 (C4)***	***0.0089991***
	1+2/2+3		***0.0089991***
	1+2/1+3	29 (AF4)	0.44466
	1+2/2+3		0.015198
G	1+2/1+3	3 (F7)	0.028597
	1+2/2+3		0.37386
	1+2/1+3	7 (T7)	0.53035
	1+2/2+3		0.042196
	1+2/1+3	8 (C3)	0.044096
	1+2/2+3		0.34777
	1+2/1+3	17 (O2)	0.27057
	1+2/2+3		0.026697
	1+2/1+3	19 (P4)	0.069093
	1+2/2+3		0.0080992
	1+2/1+3	20 (P8)	0.036396
	1+2/2+3		0.062594
	1+2/1+3	***26 (FC2)***	***0.014999***
	1+2/2+3		***0.035096***
	1+2/1+3	29 (AF4)	0.23908
	1+2/2+3		0.016698

Table 28: *p-values for the beta and gamma frequency bands and channels for which the mean PLV value from participant combination 1 and 2 is higher than both other combinations. With baseline correction.*

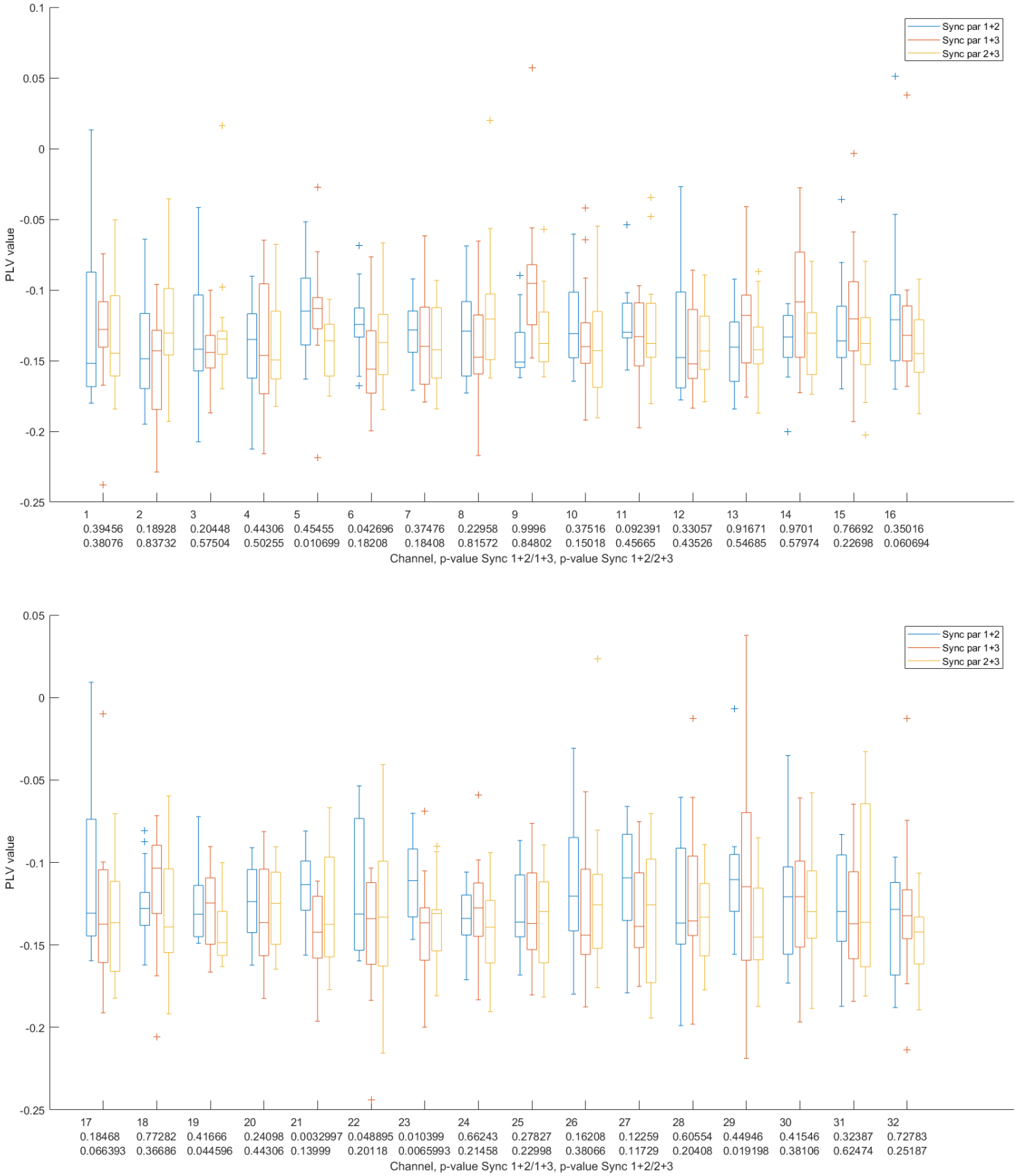


Figure 53: *All channels: Delta frequency band. With baseline correction*

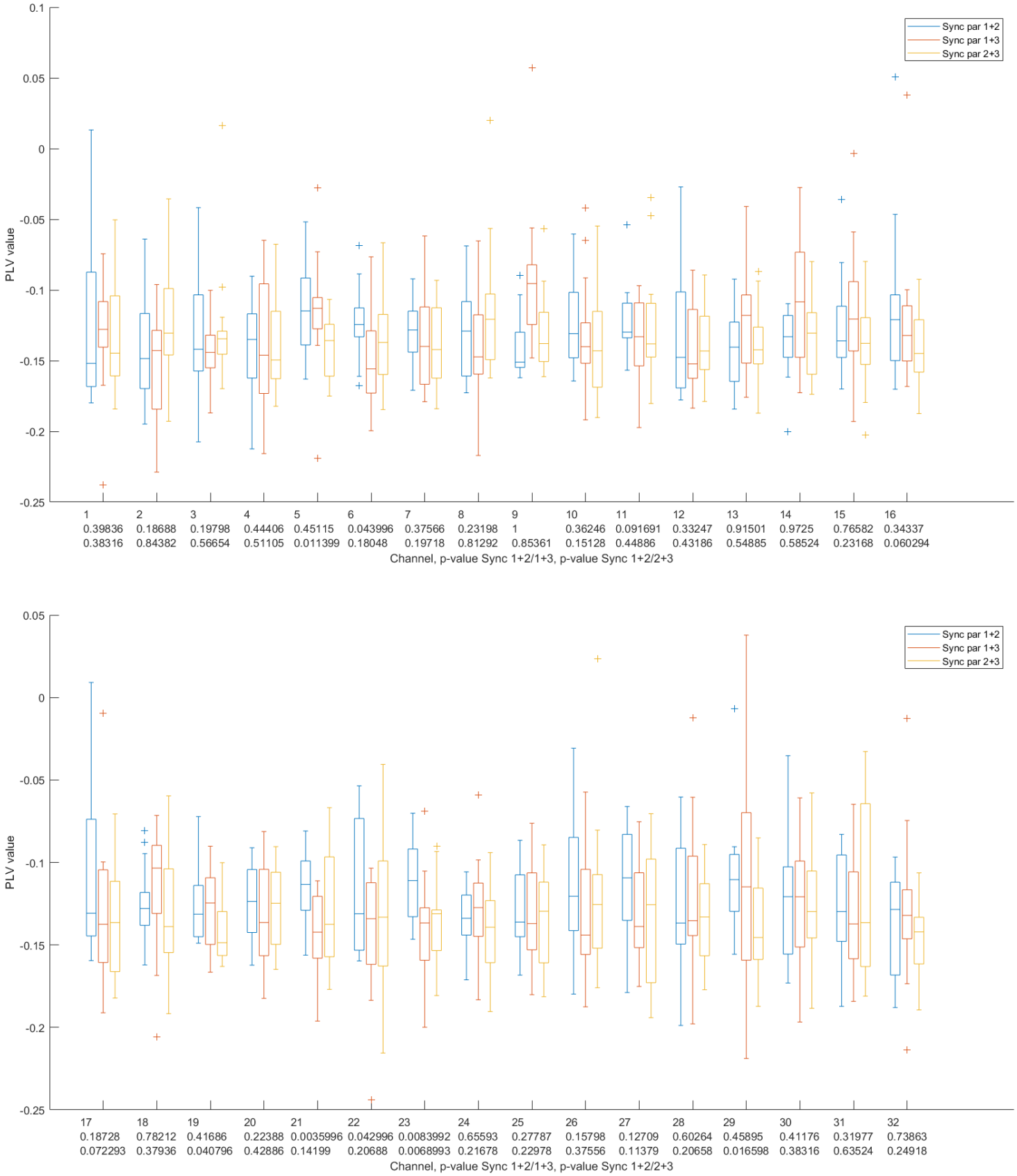


Figure 54: *All channels: Theta frequency band. With baseline correction*

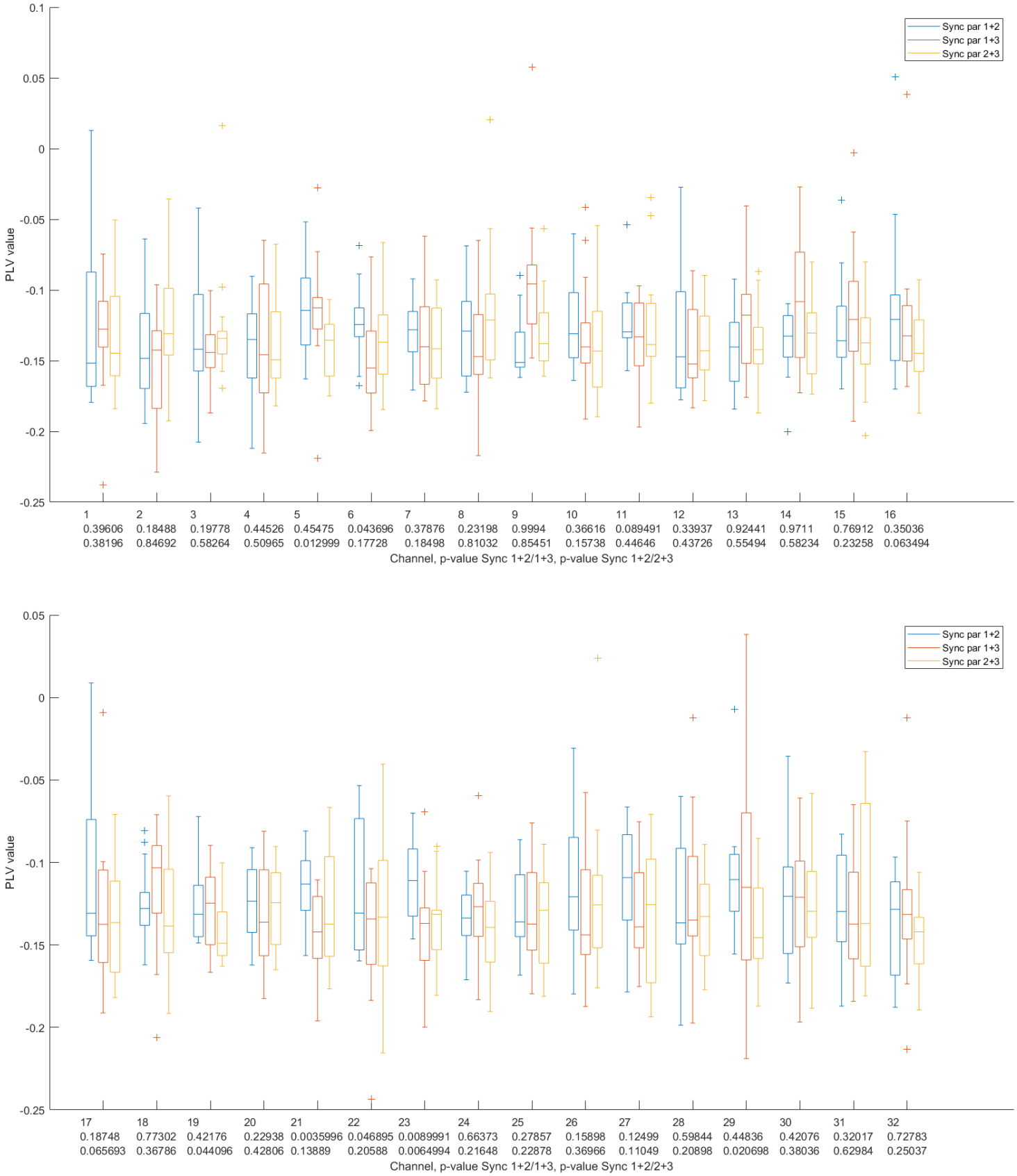


Figure 55: *All channels: Alpha frequency band. With baseline correction*

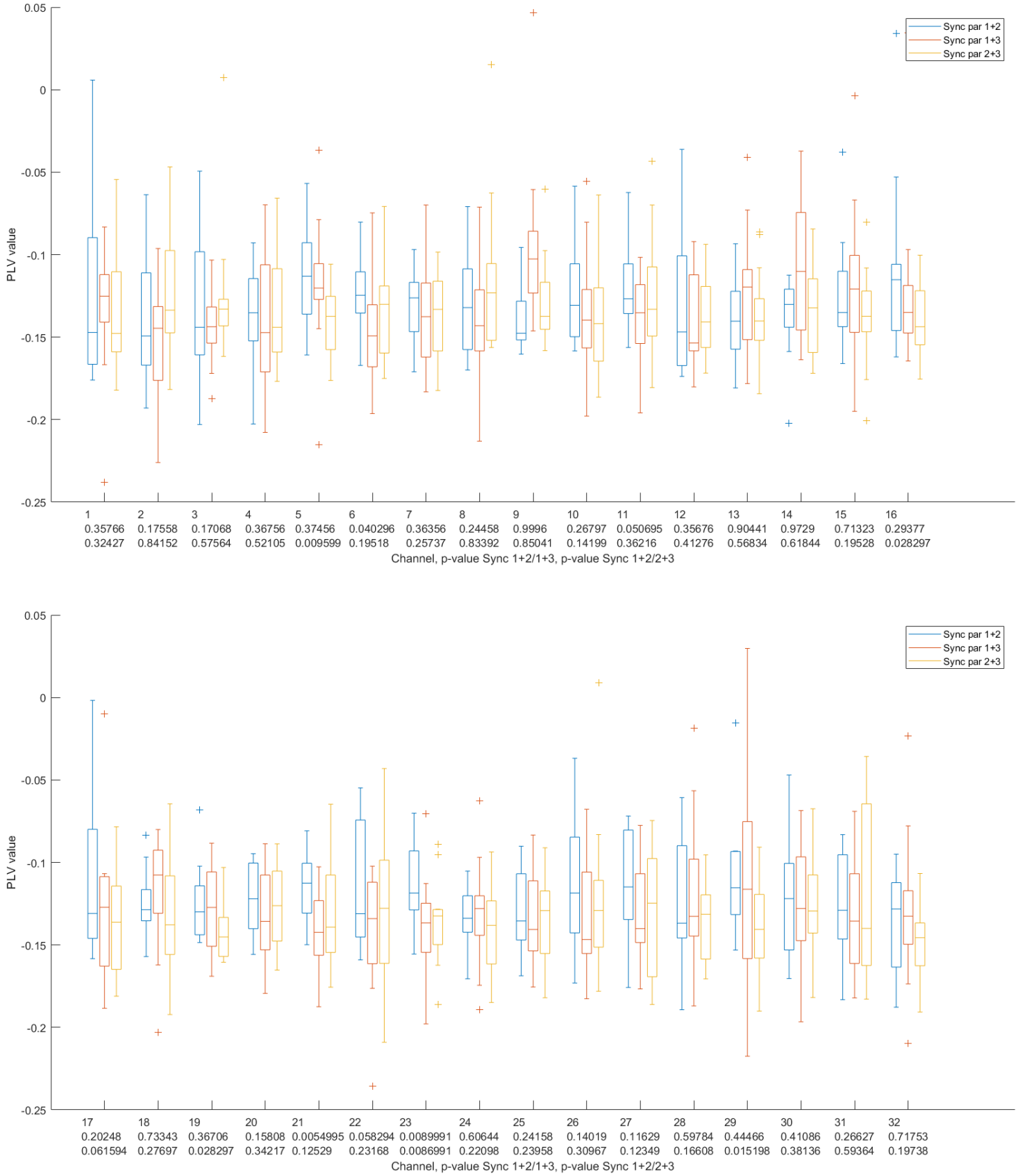
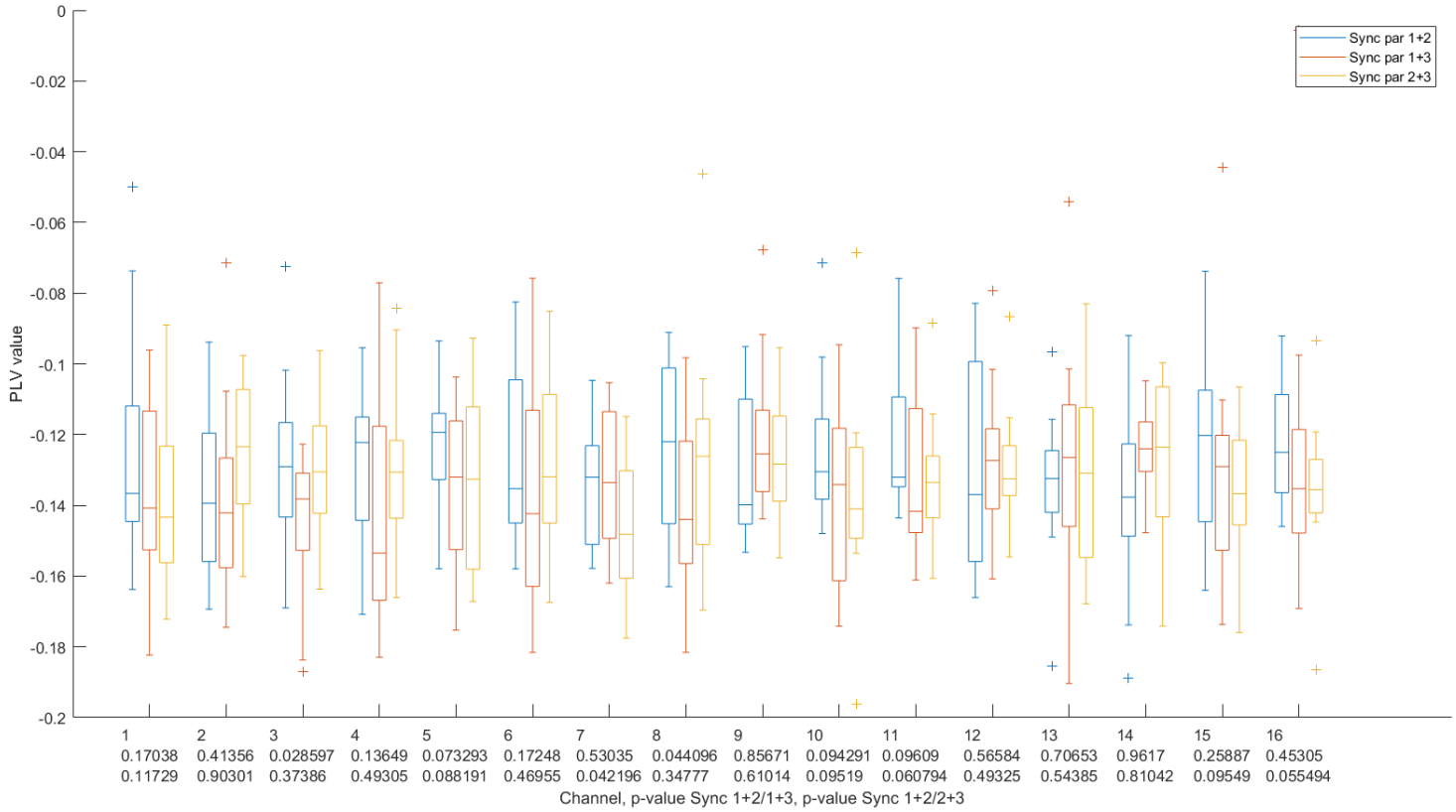


Figure 56: *All channels: Beta frequency band. With baseline correction*



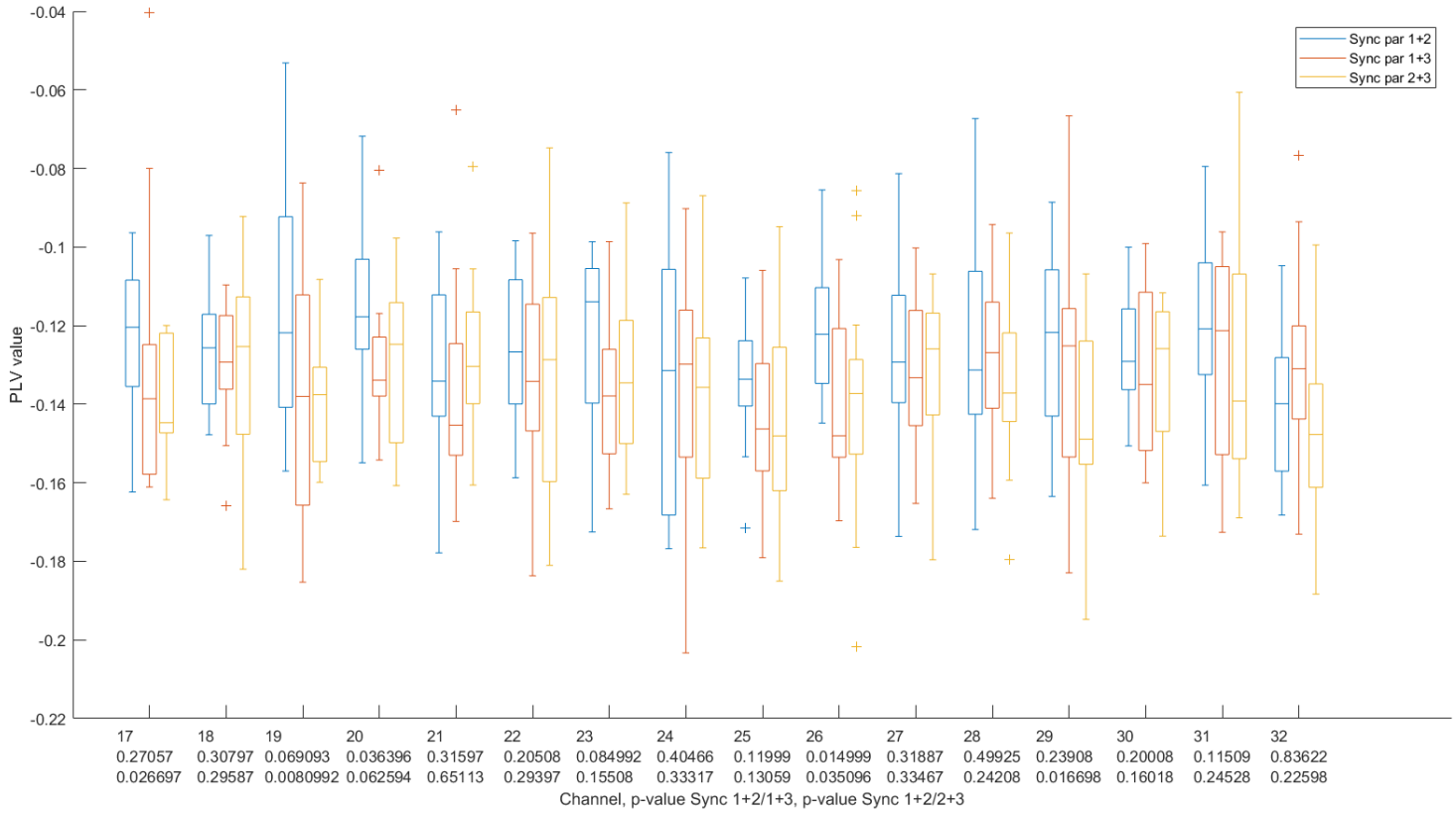


Figure 57: *All channels: Gamma frequency band. With baseline correction*

References

Acharya, U. R., Joseph, K. P., Kannathal, N., Lim, C. M., and Suri, J. S. (2006). Heart rate variability: a review. *Medical and biological engineering and computing*, 44(12):1031–1051.

Ahn, S., Cho, H., Kwon, M., Kim, K., Kwon, H., Kim, B. S., Chang, W. S., Chang, J. W., and Jun, S. C. (2018). Interbrain phase synchronization during turn-taking verbal interaction—a hyperscanning study using simultaneous eeg/meg. *Human brain mapping*, 39(1):171–188.

Anderson, J. R. (2009). *Cognitive psychology and its implications*. Worth publishers.

Anwar, A. R., Muthalib, M., Perrey, S., Galka, A., Granert, O., Wolff, S., Heute, U., Deuschl, G., Raethjen, J., and Muthuraman, M. (2016). Effective connectivity of cortical sensorimotor networks during finger movement tasks: a simultaneous fnirs, fmri, eeg study. *Brain topography*, 29(5):645–660.

Astolfi, L., Toppi, J., Borghini, G., Vecchiato, G., Isabella, R., Fallani, F. D. V., Cincotti, F., Salinari, S., Mattia, D., He, B., et al. (2011). Study of the functional hyperconnectivity between couples of pilots during flight simulation: An eeg hyperscanning study. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2338–2341. IEEE.

Asuero, A. G., Sayago, A., and Gonzalez, A. (2006). The correlation coefficient: An overview. *Critical reviews in analytical chemistry*, 36(1):41–59.

Aydore, S., Pantazis, D., and Leahy, R. M. (2013). A note on the phase locking value and its properties. *Neuroimage*, 74:231–244.

Babiloni, F., Cincotti, F., Mattia, D., Mattiocco, M., Fallani, F. D. V., Tocci, A., Bianchi, L., Marciani, M. G., and Astolfi, L. (2006). Hypermethods for eeg hyperscanning. In *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 3666–3669. IEEE.

Baccalá, L. A. and Sameshima, K. (2001). Partial directed coherence: a new concept in neural structure determination. *Biological cybernetics*, 84(6):463–474.

Baess, P., Zhdanov, A., Mandel, A., Parkkonen, L., Hirvenkari, L., Mäkelä, J. P., Jousmäki, V., and Hari, R. (2012). Meg dual scanning: a procedure to study real-time auditory interaction between two persons. *Frontiers in human neuroscience*, 6:83.

Bakker, J., Pechenizkiy, M., and Sidorova, N. (2011). What's your current stress level? detection of stress patterns from gsr sensor data. In *2011 IEEE 11th international conference on data mining workshops*, pages 573–580. IEEE.

Baumeister, J., Barthel, T., Geiss, K.-R., and Weiss, M. (2008). Influence of phosphatidylserine on cognitive performance and cortical activity after induced stress. *Nutritional neuroscience*, 11(3):103–110.

Benedek, M. and Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *Journal of neuroscience methods*, 190(1):80–91.

Benesty, J., Chen, J., Huang, Y., and Cohen, I. (2009). Pearson correlation coefficient. In *Noise reduction in speech processing*, pages 1–4. Springer.

Bevilacqua, D., Davidesco, I., Wan, L., Chaloner, K., Rowland, J., Ding, M., Poeppel, D., and Dikker, S. (2019). Brain-to-brain synchrony and learning outcomes vary by student–teacher dynamics: Evidence from a real-world classroom electroencephalography study. *Journal of cognitive neuroscience*, 31(3):401–411.

Bigdely-Shamlo, N., Mullen, T., Kothe, C., Su, K.-M., and Robbins, K. A. (2015). The prep pipeline: standardized preprocessing for large-scale eeg analysis. *Frontiers in neuroinformatics*, 9:16.

Bilek, E., Stößel, G., Schäfer, A., Clement, L., Ruf, M., Robnik, L., Neukel, C., Tost, H., Kirsch, P., and Meyer-Lindenberg, A. (2017). State-dependent cross-brain information flow in borderline personality disorder. *JAMA psychiatry*, 74(9):949–957.

Boto, E., Holmes, N., Leggett, J., Roberts, G., Shah, V., Meyer, S., and Brookes, M. (2018). Moving brain imaging towards real-world applications using a wearable meg system. *Nature*, pages 17–19.

Burchardt, T. and Le Grand, J. (2002). Degrees of exclusion: Developing a dynamic, multidimensional. *Understanding social exclusion*, page 30.

Burgess, A. P. (2013). On the interpretation of synchronization in eeg hyperscanning studies: a cautionary note. *Frontiers in human neuroscience*, 7:881.

Cao, P., Ye, B., Yang, L., Lu, F., Fang, L., Cai, G., Su, Q., Ning, G., and Pan, Q. (2020). Preprocessing unevenly sampled rr interval signals to enhance estimation of heart rate deceleration and acceleration capacities in discriminating chronic heart failure patients from healthy controls. *Computational and Mathematical Methods in Medicine*, 2020.

Carter, R. M. and Huettel, S. A. (2013). A nexus model of the temporal–parietal junction. *Trends in cognitive sciences*, 17(7):328–336.

Cha, K.-M. and Lee, H.-C. (2019). A novel qeeg measure of teamwork for human error analysis: An eeg hyperscanning study. *Nuclear Engineering and Technology*, 51(3):683–691.

Chang, C. and Glover, G. H. (2010). Time–frequency dynamics of resting-state brain connectivity measured with fmri. *Neuroimage*, 50(1):81–98.

Cohen, P. R., Jensen, D., et al. (1997). Overfitting explained. In *Preliminary Papers of the Sixth International Workshop on Artificial Intelligence and Statistics*, pages 115–122.

Cohen, S. S., Henin, S., and Parra, L. C. (2017). Engaging narratives evoke similar neural activity and lead to similar time perception. *Scientific reports*, 7(1):1–10.

Cohen, S. S., Madsen, J., Touchan, G., Robles, D., Lima, S. F., Henin, S., and Parra, L. C. (2018). Neural engagement with online educational videos predicts learning performance for individual students. *Neurobiology of learning and memory*, 155:60–64.

Cohen, S. S. and Parra, L. C. (2016). Memorable audiovisual narratives synchronize sensory and supramodal neural responses. *ENeuro*, 3(6).

Collingridge, D. S. (2013). A primer on quantitized data analysis and permutation testing. *Journal of Mixed Methods Research*, 7(1):81–97.

Cui, X., Bryant, D. M., and Reiss, A. L. (2012). Nirs-based hyperscanning reveals increased interpersonal coherence in superior frontal cortex during cooperation. *Neuroimage*, 59(3):2430–2437.

Czeszumski, A., Eustergerling, S., Lang, A., Menrath, D., Gerstenberger, M., Schuberth, S., Schreiber, F., Rendon, Z. Z., and König, P. (2020). Hyperscanning: A valid method to study neural inter-brain underpinnings of social interaction. *Frontiers in Human Neuroscience*, 14:39.

Das, D., Datta, S., Bhattacharjee, T., Choudhury, A. D., and Pal, A. (2018). Eliminating individual bias to improve stress detection from multimodal physiological data. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5753–5758. IEEE.

Davidesco, I., Laurent, E., Valk, H., West, T., Dikker, S., Milne, C., and Poeppel, D. (2019a). Brain-to-brain synchrony between students and teachers predicts learning outcomes. *bioRxiv*, page 644047.

Davidesco, I., Laurent, E., Valk, H., West, T., Dikker, S., Milne, C., and Poeppel, D. (2019b). Brain-to-brain synchrony predicts long-term memory retention more accurately than individual brain measures. *bioRxiv*, page 644047.

DeChurch, L. A. and Mesmer-Magnus, J. R. (2010). The cognitive underpinnings of effective teamwork: a meta-analysis. *Journal of applied psychology*, 95(1):32.

Dehzangi, O., Rajendra, V., and Taherisadr, M. (2018). Wearable driver distraction identification on-the-road via continuous decomposition of galvanic skin responses. *Sensors*, 18(2):503.

Delorme, A. (2020). Preprocessing data in eeglab. `https://training.incf.org/lesson/eeglab-preprocessing-part-4-filtering`. Accessed: 2020-12-10.

Delorme, A. and Makeig, S. (2004). Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis. *Journal of neuroscience methods*, 134(1):9–21.

Dikker, S., Silbert, L. J., Hasson, U., and Zevin, J. D. (2014). On the same wavelength: predictable language enhances speaker–listener brain-to-brain synchrony in posterior superior temporal gyrus. *Journal of Neuroscience*, 34(18):6267–6272.

Dikker, S., Wan, L., Davidesco, I., Kaggen, L., Oostrik, M., McClintock, J., Rowland, J., Michalareas, G., Van Bavel, J. J., Ding, M., et al. (2017). Brain-to-brain synchrony tracks real-world dynamic group interactions in the classroom. *Current Biology*, 27(9):1375–1380.

Dotan-Eliaz, O., Sommer, K. L., and Rubin, Y. S. (2009). Multilingual groups: Effects of linguistic ostracism on felt rejection and anger, coworker attraction, perceived team potency, and creative performance. *Basic and Applied Social Psychology*, 31(4):363–375.

Dumas, G., Nadel, J., Soussignan, R., Martinerie, J., and Garnero, L. (2010). Inter-brain synchronization during social interaction. *PloS one*, 5(8).

Ferdinando, H., Ye, L., Seppänen, T., and Alasaarela, E. (2014). *Emotion recognition by heart rate variability*. PhD thesis, Petra Christian University.

Ferris, D. L., Chen, M., and Lim, S. (2017). Comparing and contrasting workplace ostracism and incivility. *Annual Review of Organizational Psychology and Organizational Behavior*, 4:315–338.

Fiset, J. and Bhave, D. P. (2019). Mind your language: The effects of linguistic ostracism on interpersonal work behaviors. *Journal of Management*, page 0149206319833445.

Frazier, T. W., Strauss, M. E., and Steinhauer, S. R. (2004). Respiratory sinus arrhythmia as an index of emotional response in young adults. *Psychophysiology*, 41(1):75–83.

Funane, T., Kiguchi, M., Atsumori, H., Sato, H., Kubota, K., and 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.

Genovese, C. R., Lazar, N. A., and Nichols, T. (2002). Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage*, 15(4):870–878.

Glover, G. H. (2011). Overview of functional magnetic resonance imaging. *Neurosurgery Clinics*, 22(2):133–139.

Good, P. I. (2006). *Permutation, parametric, and bootstrap tests of hypotheses*. Springer Science & Business Media.

Gorman, J. C. (2014). Team coordination and dynamics: two central issues. *Current Directions in Psychological Science*, 23(5):355–360.

Gorman, J. C., Amazeen, P. G., and Cooke, N. J. (2010). Team coordination dynamics. *Nonlinear dynamics, psychology, and life sciences*, 14(3):265.

Gysels, E. and Celka, P. (2004). Phase synchronization for the recognition of mental tasks in a brain-computer interface. *IEEE Transactions on neural systems and rehabilitation engineering*, 12(4):406–415.

Hasan, M. K., Rusho, R. Z., Hossain, T. M., Ghosh, T. K., and Ahmad, M. (2014). Design and simulation of cost effective wireless eeg acquisition system for patient monitoring. In *2014 International Conference on Informatics, Electronics & Vision (ICIEV)*, pages 1–5. IEEE.

Hasson, U., Furman, O., Clark, D., Dudai, Y., and Davachi, L. (2008). Enhanced intersubject correlations during movie viewing correlate with successful episodic encoding. *Neuron*, 57(3):452–462.

Henderson, J. K. (2005). Language diversity in international management teams. *International Studies of Management & Organization*, 35(1):66–82.

Hirata, M., Ikeda, T., Kikuchi, M., Kimura, T., Hiraishi, H., Yoshimura, Y., and Asada, M. (2014). Hyperscanning meg for understanding mother–child cerebral interactions. *Frontiers in human neuroscience*, 8:118.

Hirsch, J., Zhang, X., Noah, J. A., and Ono, Y. (2017). Frontal temporal and parietal systems synchronize within and across brains during live eye-to-eye contact. *Neuroimage*, 157:314–330.

Iacoboni, M. and Dapretto, M. (2006). The mirror neuron system and the consequences of its dysfunction. *Nature Reviews Neuroscience*, 7(12):942.

Jian, W., Chen, M., and McFarland, D. J. (2017). Eeg based zero-phase phase-locking value (plv) and effects of spatial filtering during actual movement. *Brain research bulletin*, 130:156–164.

Jiang, J., Dai, B., Peng, D., Zhu, C., Liu, L., and Lu, C. (2012). Neural synchronization during face-to-face communication. *Journal of Neuroscience*, 32(45):16064–16069.

Kasper-Fuehrera, E. C. and Ashkanasy, N. M. (2001). Communicating trustworthiness and building trust in interorganizational virtual organizations. *Journal of Management*, 27(3):235–254.

Kawasaki, M., Yamada, Y., Ushiku, Y., Miyauchi, E., and Yamaguchi, Y. (2013). Inter-brain synchronization during coordination of speech rhythm in human-to-human social interaction. *Scientific reports*, 3(1):1–8.

King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R., and Montague, P. R. (2005). Getting to know you: reputation and trust in a two-person economic exchange. *Science*, 308(5718):78–83.

Kinreich, S., Djalovski, A., Kraus, L., Louzoun, Y., and Feldman, R. (2017). Brain-to-brain synchrony during naturalistic social interactions. *Scientific Reports*, 7(1):1–12.

Koike, T., Tanabe, H. C., Okazaki, S., Nakagawa, E., Sasaki, A. T., Shimada, K., Sugawara, S. K., Takahashi, H. K., Yoshihara, K., Bosch-Bayard, J., et al. (2016). Neural substrates of shared attention as social memory: A hyperscanning functional magnetic resonance imaging study. *Neuroimage*, 125:401–412.

Konvalinka, I. and Roepstorff, A. (2012). The two-brain approach: how can mutually interacting brains teach us something about social interaction? *Frontiers in human neuroscience*, 6:215.

Krueger, F., McCabe, K., Moll, J., Kriegeskorte, N., Zahn, R., Strenziok, M., Heinecke, A., and Grafman, J. (2007). Neural correlates of trust. *Proceedings of the National Academy of Sciences*, 104(50):20084–20089.

Kulkarni, M. (2015). Language-based diversity and faultlines in organizations. *Journal of Organizational Behavior*, 36(1):128–146.

Lachaux, J.-P., Rodriguez, E., Martinerie, J., and Varela, F. J. (1999). Measuring phase synchrony in brain signals. *Human brain mapping*, 8(4):194–208.

Lauring, J. (2008). Rethinking social identity theory in international encounters: Language use as a negotiated object for identity making. *International Journal of Cross Cultural Management*, 8(3):343–361.

Legendre, P. and Legendre, L. (1998). Statistical testing by permutation. *Numerical ecology*, pages 17–26.

Liang, Z., Li, J., Xia, X., Wang, Y., Li, X., He, J., and Bai, Y. (2018). Long-range temporal correlations of patients in minimally conscious state modulated by spinal cord stimulation. *Frontiers in physiology*, 9:1511.

Likens, A. D., Amazeen, P. G., Stevens, R., Galloway, T., and Gorman, J. C. (2014). Neural signatures of team coordination are revealed by multifractal analysis. *Social neuroscience*, 9(3):219–234.

Lindenberger, U., Li, S.-C., Gruber, W., and Müller, V. (2009). Brains swinging in concert: cortical phase synchronization while playing guitar. *BMC neuroscience*, 10(1):22.

Mandel, A., Bourguignon, M., Parkkonen, L., and Hari, R. (2016). Sensorimotor activation related to speaker vs. listener role during natural conversation. *Neuroscience letters*, 614:99–104.

Melnik, A., Legkov, P., Izdebski, K., Kärcher, S. M., Hairston, W. D., Ferris, D. P., and König, P. (2017). Systems, subjects, sessions: to what extent do these factors influence eeg data? *Frontiers in human neuroscience*, 11:150.

Michel, C. M. and Brunet, D. (2019). Eeg source imaging: a practical review of the analysis steps. *Frontiers in neurology*, 10.

Miyakoshi, M. (2020). Makoto's pre-processing pipeline. `https://sccn.ucsd.edu/wiki/Makoto's_preprocessing_pipeline`. Accessed: 2020-12-04.

Molden, D. C., Lucas, G. M., Gardner, W. L., Dean, K., and Knowles, M. L. (2009). Motivations for prevention or promotion following social exclusion: Being rejected versus being ignored. *Journal of personality and social psychology*, 96(2):415.

Montague, P. R., Berns, G. S., Cohen, J. D., McClure, S. M., Pagnoni, G., Dhamala, M., Wiest, M. C., Karpov, I., King, R. D., Apple, N., et al. (2002). Hyperscanning: simultaneous fmri during linked social interactions.

Mu, Y., Guo, C., and Han, S. (2016). Oxytocin enhances inter-brain synchrony during social coordination in male adults. *Social Cognitive and Affective Neuroscience*, 11(12):1882–1893.

Mueller, V., Saenger, J., and Lindenberger, U. (2013). Intra-and inter-brain synchronization during musical improvisation on the guitar. *PloS one*, 8(9).

Neeley, T., Hinds, P. J., and Cramton, C. D. (2009). Walking through jelly: Language proficiency, emotions, and disrupted collaboration in global work. *Harvard Business School Organizational Behavior Unit Working Paper*, (09-138).

Nyaupane, G. P., Paris, C. M., and Teye, V. (2010). Why do students study abroad? exploring motivations beyond earning academic credits. *Tourism Analysis*, 15(2):263–267.

Oshri, I., Kotlarsky, J., and Willcocks, L. (2008). Missing links: building critical social ties for global collaborative teamwork. *Communications of the ACM*, 51(4):76–81.

Pan, J. and Tompkins, W. J. (1985). A real-time qrs detection algorithm. *IEEE transactions on biomedical engineering*, (3):230–236.

Parkinson, C., Kleinbaum, A. M., and Wheatley, T. (2018). Similar neural responses predict friendship. *Nature communications*, 9(1):1–14.

Piller, I. et al. (2012). Multilingualism and social exclusion. *The Routledge handbook of multilingualism*, pages 281–296.

Pion-Tonachini, L., Kreutz-Delgado, K., and Makeig, S. (2019). Iclabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*, 198:181–197.

Ravaja, N., Saari, T., Salminen, M., Laarni, J., and Kallinen, K. (2006). Phasic emotional reactions to video game events: A psychophysiological investigation. *Media psychology*, 8(4):343–367.

Rawal, N. (2008). Social inclusion and exclusion: A review. *Dhaulagiri Journal of Sociology and Anthropology*, 2:161–180.

Reindl, V., Gerloff, C., Scharke, W., and Konrad, K. (2018). Brain-to-brain synchrony in parent-child dyads and the relationship with emotion regulation revealed by fnirs-based hyperscanning. *NeuroImage*, 178:493–502.

Reinerman-Jones, L., Taylor, G., Cosenzo, K., and Lackey, S. (2011). Analysis of multiple physiological sensor data. In *International Conference on Foundations of Augmented Cognition*, pages 112–119. Springer.

Robinson, S. L., O'Reilly, J., and Wang, W. (2013). Invisible at work: An integrated model of workplace ostracism. *Journal of Management*, 39(1):203–231.

Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161.

Savitzky, A. and Golay, M. J. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry*, 36(8):1627–1639.

Saxe, R. (2006). Uniquely human social cognition. *Current opinion in neurobiology*, 16(2):235–239.

Scholkmann, F., Holper, L., Wolf, U., and Wolf, M. (2013). A new methodical approach in neuroscience: assessing inter-personal brain coupling using functional near-infrared imaging (fniri) hyperscanning. *Frontiers in human neuroscience*, 7:813.

Stam, C. J., Nolte, G., and Daffertshofer, A. (2007). Phase lag index: assessment of functional connectivity from multi channel eeg and meg with diminished bias from common sources. *Human brain mapping*, 28(11):1178–1193.

Stephens, G. J., Silbert, L. J., and Hasson, U. (2010). Speaker–listener neural coupling underlies successful communication. *Proceedings of the National Academy of Sciences*, 107(32):14425–14430.

Stuldreher, I. V., de Winter, C. J., Thammasan, N., and Brouwer, A.-M. (2019). Analytic approaches for the combination of autonomic and neural activity in the assessment of physiological synchrony. In *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, pages 4143–4148. IEEE.

Szymanski, C., Pesquita, A., Brennan, A. A., Perdikis, D., Enns, J. T., Brick, T. R., Müller, V., and Lindenberger, U. (2017). Teams on the same wavelength perform

better: Inter-brain phase synchronization constitutes a neural substrate for social facilitation. *Neuroimage*, 152:425–436.

Tang, H., Mai, X., Wang, S., Zhu, C., Krueger, F., and Liu, C. (2016). Interpersonal brain synchronization in the right temporo-parietal junction during face-to-face economic exchange. *Social cognitive and affective neuroscience*, 11(1):23–32.

Tatum, W. O. (2014). Ellen r. grass lecture: Extraordinary eeg. *The Neurodiagnostic Journal*, 54(1):3–21.

Tenzer, H., Pudelko, M., and Harzing, A.-W. (2014). The impact of language barriers on trust formation in multinational teams. *Journal of International Business Studies*, 45(5):508–535.

Thammasan, N., Stuldreher, I. V., Schreuders, E., Giletta, M., and Brouwer, A.-M. (2020). A usability study of physiological measurement in school using wearable sensors. *Sensors*, 20(18):5380.

Thatcher, R. W. (2012). Coherence, phase differences, phase shift, and phase lock in eeg/erp analyses. *Developmental neuropsychology*, 37(6):476–496.

Thau, S., Aquino, K., and Poortvliet, P. M. (2007). Self-defeating behaviors in organizations: The relationship between thwarted belonging and interpersonal work behaviors. *Journal of applied psychology*, 92(3):840.

UTwente (2018). For international students. `https://www.utwente.nl/en/education/bachelor/practical-information/`. Accessed: 2020-05-05.

Vecchiato, G., Astolfi, L., Fallani, F. D. V., Cincotti, F., Mattia, D., Salinari, S., Soranzo, R., and Babiloni, F. (2010). Changes in brain activity during the observation of tv commercials by using eeg, gsr and hr measurements. *Brain topography*, 23(2):165–179.

Voss, J., Albert, I., and Ferring, D. (2014). Language use and value orientations in multinational work teams in luxembourg: Conflict or harmony? *International Journal of Intercultural Relations*, 41:192–196.

Wang, M.-Y., Luan, P., Zhang, J., Xiang, Y.-T., Niu, H., and Yuan, Z. (2018). Concurrent mapping of brain activation from multiple subjects during social interaction by hyperscanning: a mini-review. *Quantitative imaging in medicine and surgery*, 8(8):819.

Wang, Y., Hong, B., Gao, X., and Gao, S. (2006). Phase synchrony measurement in motor cortex for classifying single-trial eeg during motor imagery. In *2006 international conference of the IEEE engineering in medicine and biology society*, pages 75–78. IEEE.

Williams, K. D. (2002). *Ostracism: The power of silence*. Guilford Press.

Williams, K. D. (2007). Ostracism: The kiss of social death. *Social and Personality Psychology Compass*, 1(1):236–247.

Xie, H., Howell, A., Schreier, M., Sheau, K. E., Manchanda, M. K., Ayub, R., Glover, G., Jung, M., Reiss, A. L., and Saggar, M. (2019). Finding the neural correlates of collaboration using a three-person fmri hyperscanning paradigm. *bioRxiv*, page 782870.

Yun, K., Watanabe, K., and Shimojo, S. (2012). Interpersonal body and neural synchronization as a marker of implicit social interaction. *Scientific reports*, 2:959.

Zhang, L. and Hung, H. (2016). Beyond f-formations: Determining social involvement in free standing conversing groups from static images. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1086–1095.