Radboud University Nijmegen

Department of Social Sciences

Artificial Intelligence

Towards a Simplified Audio Setup of a BCI-based Language Training Paradigm: a Behavioral Study

Bachelor's Thesis

July 4, 2022

Crispijn Aalberts, s1037764

crispijn.aalberts@ru.nl

First supervisor: M. W. Tangermann

Second supervisor: Dr. P. M. W. Desain

Abstract

Brain-computer interface (BCI) systems use brain signals to create a communication link between the brain and an external device. Auditory BCIs detect focused attention on acoustic stimuli based on event-related potentials (ERPs) in signals from electroencephalography. A recent study by Musso et al. (2022) investigated the use of an auditory BCI for aphasia rehabilitation in stroke patients. Musso et al. propose a new BCI-based language training paradigm that reinforces effective language processing strategies through brain-state-dependent feedback. The paradigm uses a ring of six loudspeakers for the spatial presentation of stimuli which has been shown to improve classification accuracy in auditory BCI paradigms. However, this setup is not feasible for everyday BCI usage. Thus, the present study aims to investigate the possibility of a simplified audio setup using stereo headphones. The proposed setup will be evaluated by conducting a behavioral study and assessing the workload and ergonomic ratings. Eight healthy participants were tested in a within-subject design using four audio conditions: six loudspeakers, stereo headphones, stereo headphones incorporating pitch and mono headphones. Results show the two stereo headphone conditions are comparable to the six loudspeaker condition in terms of counting accuracy, workload ratings and ergonomic ratings. This indicates that a spatial headphone paradigm might be a promising compromise between the goal of high classification accuracy and easy applicability for everyday BCI use.

Note: This bachelor's thesis was part of a group project with the goal of investigating the feasibility of a simplified audio setup of the BCI-based language training paradigm by Musso et al. (2022). The group project was divided into three parts: optimizing the stereo audio transformations (Kortenbach, 2022), EEG analysis of the experimental paradigm (Milosevska, 2022) and a behavioral study which will be covered in the present study. The group worked together to design and test the simplified audio setup. Each student wrote an individual thesis covering their part of the analysis.

Content

1. Introduction

Brain-computer interface (BCI) systems provide an artificial communication link between an individual's brain and an external device via the processing and classification of detectable brain signals. When users perform specific mental tasks, the system identifies the activity patterns of their brain and this information can be used to control an external application device without depending on motoric output pathways via the limbs (Allison et al., 2020). Brain-computer interfaces create new possibilities for individuals suffering from severe neuromuscular disorders such as amyotrophic lateral sclerosis, brainstem stroke, and spinal cord damage. These patients, who may be completely paralyzed, can use the basic communication abilities the BCI system provides to communicate their requests to caregivers or even operate word processing programs or neuroprostheses (Wolpaw et al., 2002).

Moreover, BCI systems can also be used for cognitive rehabilitation. Patients with impaired cognitive functions can use a neurofeedback BCI system to restore cognitive functions. Neurofeedback therapy measures brain signals to improve neural functions: patients are shown a suitable graphical representation of their brain activity and are taught to self-regulate this activity to bring it to a goal state (Carelli et al., 2017). This method has been used to treat a variety of neurological and psychiatric disorders, including attention deficit hyperactivity disorder, anxiety, epilepsy, and addictive disorders (Angelakis et al., 2007). A recent study by Musso et al. (2022) has investigated the use of BCI for aphasia rehabilitation in stroke patients. Musso and colleagues propose a new language training approach for the rehabilitation of aphasia patients that uses an auditory BCI paradigm. In this training approach, patients have to perform an auditory target word detection task whilst their EEG is recorded. During training, the patient is provided with immediate brain-state-dependent feedback which reflects how well they perform the task during a training session. As the BCI system is able to provide feedback time-locked to a brain state, it can reward effective language processing strategies which increases brain plasticity. The results are promising: patients substantially improved in functional communication, expressed by higher self-reported quality and quantity of language production. Moreover, pre-post fMRI recordings showed significant changes in functional connectivity between several language domains (Musso et al., 2022).

The BCI language training paradigm by Musso et al. (2022) is intended to be used as assistive technology in end users' homes or in post-stroke rehabilitation centers. Therefore, efforts are necessary to transfer the BCI setup from the laboratory to these environments. The goal of the present study is to investigate whether a simplified audio setup that utilizes stereo headphones is possible for the BCI-based language training paradigm by Musso et al. This study will analyze the performance of the simplified audio setup by conducting a behavioral study. In addition, this study will evaluate the subjective ergonomic experience and workload of the proposed setup.

First, the core concept of BCI is discussed (1.1). Second, an overview of current research on stimulus properties in auditory BCI systems is presented (1.2). Third, the use of BCI in aphasia rehabilitation is introduced (1.3). Lastly, the aims and research questions of this study are presented (1.4).

1.1 Brain-Computer Interfaces

BCI systems acquire brain signals, classify them, and translate them into commands that are transferred to output devices. Researchers have used a variety of different brain signals for device control including electroencephalographic, intracortical, electrocorticographic, and even single-neuron-based brain signals (Shih et al. 2012). Most modern BCI systems rely on an electroencephalogram (EEG) (Allison et al., 2020). An electroencephalogram is a measure of electrical activity produced by neurons in the cerebral cortex, recorded non-invasively from the scalp (Kübler et al., 2001). Hence, EEG activity results from the summation of excitatory and inhibitory postsynaptic potentials from underlying areas of the cerebral cortex, with some contribution of granular and glial cell activity (Speckmann & Elger, 1993). For BCIs, EEG offers several advantages over other input signals. Because of its relatively low costs, high temporal resolution, and portability EEG is the preferred choice for many BCI applications (Käthner et al., 2013).

In an EEG-based BCI system stimulus-dependent brain signals, also known as event-related potentials (ERPs), are recorded. ERPs are electrical brain responses time-locked to sensory, cognitive, or motor stimuli and are characterized by their voltage amplitude and their latency in relation to stimulus onset. ERPs are labeled by their latency and electrical polarity (positive or negative). For example, a prominent negative peak 200 milliseconds post-stimulus is referred to as N200; a positive peak 300 milliseconds after stimulus onset is referred to as P300 (Kübler et al., 2001). An ERP component can also be defined by its scalp distribution or its relation to experimental variables. The ERP is made up of voltage deflections that reflect the processing of sensory information as well as higher-level processing such as selective attention, memory updating, semantic comprehension, and other kinds of cognitive activity (Duncan et al., 2009). The sequence and latencies of ERP components capture the time course of processing activity in milliseconds, whilst their amplitudes show the extent of neural resource allocation to distinct cognitive processes. ERPs are sensitive to information processing characteristics (e.g., auditory discrimination accuracy, expectation, semantic processing) and can be used to complement traditional performance measures, like the accuracy and speed of behavioral responses (Duncan et al., 2009).

In the case of an attention-based ERP approach, the subject is asked to concentrate on one target stimulus in a sequence of non-target and target stimuli. During this task, the signal is processed by extracting relevant features from it. The relevant features are transferred to a machine learning model, such as a classifier. The classifier then separates ERPs elicited by target stimuli from ERPs elicited by non-target stimuli. One classification method often used for this is linear discriminant analysis (LDA). LDA is a classification method based on finding a decision boundary to separate two classes. In this case the true class labels are known (target vs non-target), which means this is a supervised classification problem. The separating hyperplane between the two classes is calculated by minimizing the variance, while maximizing the means of the classes. LDA assumes that the covariance matrices of either class are equal, that both classes are normally distributed, and that the classes are linearly separable. In the case of EEG data we are dealing with high-dimensional data, so shrinkage regularization needs to be applied to the LDA model in order to increase classification accuracy (Höhne et al., 2015).

After classification, a control signal can then be transmitted to the application device based on the classifier output (Kleih et al., 2015). Thus, for a BCI system to be successful, it needs to have a high classification accuracy. This study will use an auditory ERP-based BCI paradigm, the principle of which is described in Section 1.1.3.

1.1.1 P300 ERP component

A paradigm that is often used in ERP-based BCIs is the oddball paradigm. In the oddball paradigm, a random sequence of stimuli is presented. The stimuli consist of infrequently occurring target stimuli and frequently occurring non-target stimuli (Sutton et al., 1965, 1967). The task of the participant is to classify these stimuli correctly. The target stimulus that is repeated infrequently (the “oddball”) will elicit a P300 (Duncan-Johnson & Donchin, 1977). The P300 is a positive deflection in the EEG which typically peaks 300 milliseconds or more after the onset of a rare, task-relevant stimulus. However, the latency of the P300 ERP component can vary between 200 to 700 milliseconds after stimulus onset (Furdea et al., 2009; Käthner et al., 2013). The amplitude of the P300 is related to task-processing demand (Gopher & Donchin, 1986; Polich, 2007), and its latency is linked to the difficulty in differentiating target stimuli from non-targets (Duncan et al., 2009).

The P300 wave has a centro-parietal scalp distribution with a maximum amplitude over the midline scalp sites (Picton & Hillyard, 1974; Ritter et al., 1972; Squires et al., 1975). It has been hypothesized that the P300 consists of two components (Dien et al., 2004; Polich, 2007; Squires et al., 1975). The first component is the P300a recorded at fronto-central portions of the scalp, related to attentional processing (Polich, 2007), possibly including “novelty-P300” (Dien et al., 2004; Polich, 2007; Simons et al., 2001). And the second component is the parietally recorded P300b which is elicited when target stimuli are processed (Wascher et al., 2020). It is well established that the P300 amplitude is largest when evoked by events that the subject considers important (Johnson & Donchin, 1978; Sellers & Donchin, 2006). Therefore, subjects can produce more distinct P300 signals by attending to specific target stimuli while ignoring others.

It has been shown that increasing workload and task-processing demand reduces the amplitude of the P300 (Gopher & Donchin, 1986; Käthner et al., 2013; Kramer et al., 1986; Wintink et al., 2001). Recent studies have used the NASA-TLX questionnaire (NASA Human Performance Research Group, 1987) as a measure of subjective workload for BCI applications (Käthner et al., 2013; Pasqualotto et al., 2011; Riccio et al., 2011, Simon et al., 2015). The present study will also make use of the NASA-TLX questionnaire to evaluate subjective workload. For further details see Section 2.3.2.
Furthermore, several studies have found that other factors such as mood and motivation (Nijboer et al., 2010) influence either P300 amplitude (Baykara et al., 2016; Kleih et al., 2010) or BCI performance (Kleih et al., 2011). These studies used the Questionnaire for Current Motivation (QCM) for BCI (Nijboer et al., 2008) which is a slightly altered version of the original QCM (Rheinberg et al., 2001). The QCM for BCI consists of 18 items divided into four subscales: incompetence fear, mastery confidence, interest and challenge which have to be rated on a seven point Likert scale. The present study will not use the QCM for BCI but rather a simplified questionnaire partially adapted from Höhne et al. (2012) asking for the subjective ergonomic experience of the participant. For further details see Section 2.3.3.

1.1.2 N200 ERP component

The N200 is a negative deflection in the EEG which peaks 200 to 350 milliseconds post-stimulus (Folstein & Van Petten, 2008) and reflects the detection of novelty or mismatch during auditory stimulation. It can be separated into three components: a subcomponent following detection of novelty or mismatch over anterior regions, an attention-related stimulus categorization subcomponent across central areas, and a subcomponent related to focused attention over contralateral central areas (Luck, 2014). It has been shown that the N200 amplitude is higher, when the target stimulus is attended (Oknina et al., 2011). Furthermore, the later subcomponent of the N200 can be used as an indicator of attention allocation and for estimating the detection duration (Luck, 2014)

1.1.3 Auditory ERP-based BCI paradigms

Oddball paradigms where the participant is instructed to focus on specific audio stimuli are applied in many auditory ERP-based BCI paradigms (e.g. Höhne et al., 2011; Käthner et al., 2013; Kleih et al., 2015; Musso et al., 2022). Stimuli can be classified as target or non-target based on the differences in ERP responses for the attended target stimulus and the unattended non-target stimulus. The classification result can then be used as a control signal for an application device. Auditory ERP-based BCI paradigms have distinct advantages over their visual counterparts, especially for patients with locked-in syndrome or amyotrophic lateral sclerosis (ALS). While most visual ERP-based BCI paradigms require adequate control of the user's eye movement or muscle-dependent gaze direction, auditory BCI paradigms overcome this restriction (Höhne et al., 2011). This is critical in medical applications, as these movements may be exhausting or impossible to perform for some patients (Hill et al., 2014). Even patients in late stage ALS are usually able to hear, which means they can make use of an auditory BCI. Although listening to auditory stimuli is more demanding than attending visual stimuli (Klobassa et al., 2009), it is still possible to set up an ERP-based BCI paradigm that exclusively uses auditory stimuli.

1.2 Stimulus properties in auditory BCI paradigms

As stated earlier, the amplitude of the P300 ERP component is linked to task-processing demand (Gopher & Donchin, 1986; Polich, 2007), and its latency is linked to the difficulty in differentiating target stimuli from non-targets (Duncan et al., 2009). This implies that different stimulus properties can affect the amplitude and latency of the P300, and would therefore also affect classification accuracy. The section below will highlight studies of different stimulus properties in auditory BCI paradigms.

1.2.1 Spatial stimulus presentation: AMUSE paradigm

Human listeners are able to distinguish sounds in space, according to several behavioral studies (Brungart et al., 1999; Mondor & Zatorre, 1995; Teder-Sälejärvi & Hillyard, 1998). Several of these studies also found that when subjects focus their attention on a specific direction, their attentional resources appear to be distributed in a gradient, with decreased alertness as they move away from the attended direction (Mondor & Zatorre, 1995; Teder-Sälejärvi & Hillyard, 1998).

An offline study by Schreuder et al. (2010) investigated the influence of spatial stimulus presentation in a BCI paradigm using artificial tone stimuli. In this “Auditory Multi-class Spatial ERP” paradigm, also referred to as the AMUSE paradigm, a participant is surrounded by six equally-spaced loudspeakers. These loudspeakers are used to enhance stimulus discriminability by cueing focused attention towards a certain direction. Schreuder et al. found that the mean classification accuracy between target and non-target stimuli was higher with spatial versus non-spatial presentation. All but one subject reached selection scores higher than 90% in the spatial condition. Subject performance in the non-spatial control condition was below the 70% threshold for all but one subject, showing that spatial location adds vital information to the stimulus (Schreuder et al., 2010).

An altered version of the AMUSE paradigm was also used in a master's thesis by Denzer (2016) on the effect of spatial word presentation in an auditory ERP paradigm for BCIs (Denzer, 2016). In this study, participants would listen to a cueing sentence, after which a sequence of target and non-target words was played. Similar to the setup used by Musso et al. described in Section 1.3.2 participants were instructed to focus attention on the target word, evoking a P300 response that could then be classified using an LDA model. The study found that the experimental condition with six equally-spaced loudspeakers resulted in a higher classification accuracy compared to the mono audio headphone condition and the one-directional speaker condition (Denzer, 2016).

Another study by Gao et al. (2011) examined the effect of spatial stimulus presentation via five simulated headphone directions compared to one headphone direction in an auditory BCI paradigm. The auditory stimulus was a sequence of five spoken digits with the same duration and intensity. In the spatial paradigm with five headphone directions, each digit was presented from a fixed direction, and each digit could be a male or female voice. The participant was instructed to pay attention to the direction of the target digit, while ignoring digits from the other directions, then discriminate the target voice’s gender. In the non-spatial paradigm, all stimulus voices were presented from directly ahead of the participant using only one spatial location. The spatial paradigm significantly outperformed the non-spatial paradigm with more averaged trials (Gao et al., 2011).

1.2.2 Effect of pitch on stimulus discriminability: PASS2D paradigm

Höhne et al. (2011) proposed a novel approach that uses auditory ERPs to control a BCI spelling application. This paradigm - called “Predictive Auditory Spatial Speller with two-dimensional stimuli,” or PASS2D - was investigated in an online study with healthy participants. To control the auditory ERP speller, BCI users had to focus their attention on two-dimensional auditory stimuli presented via headphones that varied in both, pitch (high/medium/low) and direction (left/middle/right). Similarly to the approach of Schreuder et al. (2010), however, the information transmitted by the two dimensions (pitch and direction) is independent and not redundant as in Schreuder’s AMUSE paradigm.The resulting 3 × 3 design offers an arrangement of nine stimuli that are easy to discriminate from each other, see Figure 1. These nine different stimuli can be used as control signals to drive a predictive text entry system. Which enables the user to spell a letter by a single nine-class decision plus two additional decisions to confirm a spelled word (Höhne et al., 2011).

In the PASS2D paradigm, the nine auditory stimuli are not completely independent: for each target there are four non-targets being equal in one dimension, that is, two stimuli with the same pitch (same row) and two stimuli with the same direction (same column). Analysis of the binary classifier outputs and multiclass decisions revealed that the classifier could resolve the dimension “pitch” better than the dimension “direction” (Höhne et al., 2011). Effectively showing that altering the pitch of auditory stimuli adds another dimension to the discriminability of the stimulus which could potentially improve classifier performance.

Schreuder et al., (2010) showed, that both dimensions contain valuable information for a discrimination task, and that a redundant combination can enhance the discriminability compared to the single stimulus types. Hence, the choice to use the two stimulus dimensions independently rather than redundantly encoding the same information represents a trade-off between large steps in pitch/direction and a relatively high number of nine classes (Höhne et al., 2011). The PASS2D paradigm was limited to only three directions. Because of this, the hardware complexity and space requirements for setting up the system at a patient's house can be lowered, since three-direction audio can be achieved using basic stereo headphones.

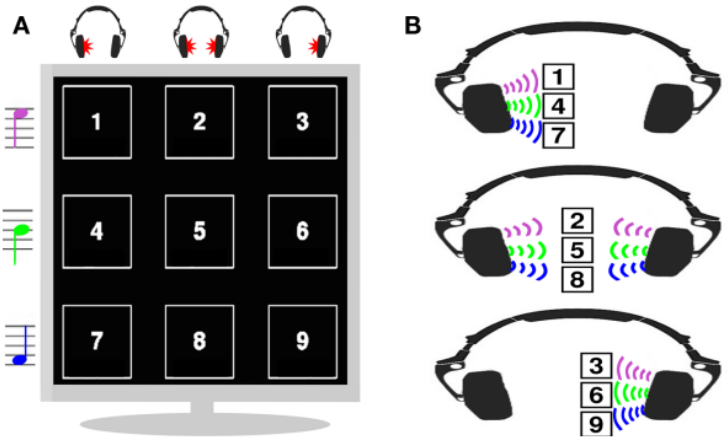


Figure 1: Visualization of the nine auditory stimuli, varying in pitch and direction. The 3 × 3 design (**A.**) was shown on the screen. (**B.**) Distribution of the nine stimulus tones to the stereo channel of the headphone. Adapted from: *Höhne, J., Schreuder, M., Blankertz, B., Tangermann, M. (2011). A Novel 9-Class Auditory ERP Paradigm Driving a Predictive Text Entry System. Frontiers in neuroscience. 5. 99. 10.3389/fnins.2011.00099.*

1.2.3 Natural stimuli

Both the AMUSE paradigm (Schreuder et al., 2010) and the PASS2D paradigm (Höhne et al., 2011) use short artificially generated tones to elicit auditory ERP responses. Two practical limitations were observed that were related to this choice of stimuli. First, these highly regulated and uniform tone settings were perceived as difficult to use and even described as unpleasant by some users (Höhne et al., 2011). Indicating a limited overall acceptance of a final BCI spelling system and possibly even affecting performance, given that user motivation is correlated with BCI performance (Kleih et al., 2010; Tangermann et al., 2011). Second, posterior analysis of the spelling performance in both paradigms indicated a number of systematic multi-class confusions in the classification of target versus non-target stimuli (Höhne et al., 2012). Schreuder et al. (2010) hypothesized that a BCI with

spoken word stimuli might be a better alternative to the somewhat unnatural tone stimuli. Since the increased semantic and acoustic information in natural stimuli may improve stimulus discrimination, which is critical for classification accuracy. However, it also introduces problems such as higher latency jitter in the P300 onset, which could make classification more difficult (Schreuder et al. 2010). Spoken words containing spatial information may elicit a stronger P300 response since it is easier to focus on the direction of the stimulus.

A study by Höhne et al. (2012) investigated the effect of using natural stimuli and found a higher mean classification accuracy using natural stimuli versus artificial stimuli in a spatial auditory BCI paradigm. In the experiment, artificially generated tones, spoken syllables, and sung syllables were presented via headphones. The nine artificially generated stimuli consisted of three tones with different pitch (high/medium/low). Each of the three tones was presented from three different directions (left/middle/right). Thereby the 3 × 3 design of the PASS2D paradigm (Höhne et al., 2011) was maintained, see Figure 1. The spoken and sung syllables were varied in speaker voice (bass/tenor/soprano) and vowels (i/æ/o). Every speaker voice was presented only from one fixed direction, following the same 3 × 3 design as the artificially generated tones. Participants were instructed to concentrate on the target stimulus and to ignore all non-target stimuli. In addition, they were asked to count the target stimuli and to indicate the number of occurrences at the end of each trial. The study found that the mean classification accuracy was higher for spoken syllables and sung syllables compared to artificial tones, showing that natural stimuli improve classification accuracy (Höhne et al., 2012). Furthermore, the subjective ergonomic ratings showed that it was easier for the subjects to concentrate on natural stimuli than on artificial stimuli. Participants also counted the number of targets more accurately for natural stimuli than for artificial stimuli. Additionally, a positive correlation between subjective ergonomic ratings and classification accuracies was found (Höhne et al., 2012). These findings suggest that the higher amount of stimulus information contained in natural stimuli could improve classification accuracy in auditory BCI paradigms.

1.3 Aphasia

Aphasia is a condition that affects the capacity to comprehend or produce language. It is caused by damage to specific brain regions in the left hemisphere and is common in stroke patients (Damasio, 1992). People with aphasia face a variety of challenges, ranging from occasional difficulty finding words to complete loss of the ability to speak, read, or write. Based on the location of the brain lesion, different forms of aphasia have been recognized, each with its unique set of symptoms. One of the most prevalent methods of describing aphasia is by the fluency of language output. A general distinction can be made between fluent and non-fluent aphasia. Fluent aphasia, associated with posterior lesions, is characterized by continuous runs of speech with a variety of syntactic structures, in which phrase length is normal, but output is often incorrect (Edwards, 2005; Feyereisen et al., 1991; Goodglass & Kaplan, 1983). In contrast, non-fluent aphasia, which is associated with anterior lesions, is characterized by increased effort, impaired prosody, articulatory mistakes, and limited grammaticality (Feyereisen et al., 1991; H. Goodglass & Kaplan, 1983). Fluent aphasia syndromes include Wernicke’s aphasia, transcortical sensory aphasia, conduction aphasia, and anomic aphasia, whereas non-fluent aphasias include global aphasia, Broca’s aphasia, and transcortical motor aphasia (Clough & Gordon, 2020).

Neuroimaging studies of language disorders, such as aphasia, have refuted Broca and Wernicke's theory of full functional modularity for the two main language centers (Eling & Whitaker, 2009). Broca's area and Wernicke's area are still vital for language functions, however, recent studies in the field of neurolinguistics investigate areas and connections that go far beyond these traditional areas (Tremblay & Dick, 2016). Brain lesions in any of the areas involved in linguistic functioning can result in different forms of aphasia depending on the lesion site. Because each of these areas contributes to the overall functioning of language processing and production. Furthermore, it has been shown by neurocognitive studies (Friederici & Gierhan, 2013; Indefrey & Levelt, 2000) and neuroimaging data (Okada & Hickok, 2006) that these areas of language processing and production interact with one another. As a result, language competencies could theoretically be trained as a network, i.e. training the language processing network could have an impact on the language production network.

Aside from the effects on language comprehension and production, post-stroke aphasia has severe consequences for the health-related quality of life of patients. Aphasia is associated with a reduction of independence, social isolation (Dalemans et al., 2010) and failure in returning to work (Doucet et al., 2012). In the long term, chronic aphasia may not only cause psychological distress (Gainotti, 1997), but it may also affect the economic situation of the patient (Hinckley, 1998), as many professions require elaborate communication.
Aphasia is more common than Parkinson's disease, nearly 180,000 Americans acquire the disorder each year (Aphasia.org, 2021). Furthermore, post-stroke aphasia is also associated with increased risk of mortality (Pedersen et al., 2004).

Even in the most severe cases of aphasia, recovery is still possible. Within six months following a stroke, spontaneous aphasia recovery is observed, but only minimal improvements are seen past that point. Intensive therapy is required to aid patients in this chronic stage. Conventional speech and language therapy (cSLT) is the most common treatment for aphasia. But there are also other therapy approaches, such as computer-based language therapy. Computer aided therapy programs for aphasia patients are either specific, designed to target a particular deficit area in the brain, or general, addressing many deficit areas (Archibald et al., 2009). Therapeutic benefits have been reported for specific programs addressing deficits such as, anomia (Raymer et al., 2006), sentence comprehension (Crerar et al. 1996), sentence construction (Linebarger et al. 2001), and spelling (Mortley et al. 2001).
While cSLT is still the widely used treatment for aphasia, its effectiveness has not been decisively proved (Berthier 2005). A meta-analysis by Brady et al. which included 27 randomized studies that compare cSLT versus no therapy in stroke patients showed clear evidence in favor of cSLT (Brady et al., 2016). Nevertheless, the effects were only short-term, with moderate effect sizes for functional communication, reading comprehension, and writing, low effect sizes for expressive abilities, and no effect for naming and auditory comprehension (Brady et al., 2016).

The combination of low therapy success, substantially reduced quality of life, and a high incidence rate calls for more effective therapeutic interventions. As a result, efforts have been made to integrate knowledge from the field of BCI in order to create new therapeutic approaches.

1.3.1 Visual BCI in aphasia rehabilitation

In a study by Shih et al. (Shih et al. 2013) the use of a BCI for communication was tested and proven successful in eight aphasic stroke patients. All stroke patients had Broca's aphasia and an NIH stroke scale language subscore of 2 or greater, indicative of severe aphasia. The task of the patients was to focus attention on a specific character of the word matrix and silently count the number of times the target character flashed. This was done using the copy speller mode of the P300 speller: words were presented on the top left of the monitor, and the target character was listed in parentheses at the end of the letter string as shown in Figure 2. Patients were able to achieve speller accuracies of between 60% and 65% on this task.

Figure 2: The 6 × 6 matrix used in the study by Shih et al. (2013). A row or column intensifies for 100 ms every 175 ms. The letter in parentheses at the top of the window is the current target letter "D." Adapted from: *Krusienski, D. J., Sellers, E. W., McFarland, D. J., Vaughan, T. M., & Wolpaw, J. R. (2008). Toward enhanced P300 speller performance. Journal of Neuroscience Methods, 167(1), 15–21. doi:10.1016/j.jneumeth.2007.07.017*

Later research by Kleih and colleagues (Kleih et al., 2016) also validated the feasibility of a visual P300-BCI speller communication system for aphasia patients. The study included five participants diagnosed with post-stroke aphasia (predominantly motor aphasia) according to the Bielefelder Aphasie Screening (Richter et al., 2006). In each training session, participants were instructed to copy-spell three short five-letter words. After copy-spelling, the participant could use free-spelling mode such that they spelled without the experimenter knowing the target word. In order for the participants to use the P300 speller successfully, modifications to the system were necessary. The participants reported major problems in ignoring non-target stimuli so the researchers supported focusing on the target letter by covering the word matrix with a piece of cardboard on which a square for the target letter was cut out. The use of cardboard altered the traditional BCI paradigm, allowing a target stimulus to be classified against background EEG rather than target vs. non-target stimuli. Nonetheless, the participant still had to concentrate on the target stimulus, or else a random letter would be selected. After participants had achieved 100% accuracy using the cardboard presentation, the cardboard was removed, and the participant could use the standard BCI presentation. In their last session, all participants were able to use the free-spelling mode without any extra alterations or assistance.

According to the researchers, implementing a visual P300 BCI could contribute to aphasia rehabilitation in two major ways: first, the P300 amplitude is dependent on attention allocation and therefore provides a measure of task attention (Johnson, 1986; Polich, 2007). If BCI training can be used as a form of attention training, it should result in an enhanced P300 amplitude on the psychophysiological level since letters are selected by focusing attention. Kleih et al. found an increase in P300 amplitude across presentation modes in two out of five patients. Another study by Baykara et al. (2016), which used healthy participants, also showed an increased P300 amplitude as a result of training to focus on the target stimulus in an auditory BCI paradigm. It should be noted, however, that this assumption cannot be directly applied to people diagnosed with post-stroke aphasia without further research since their brains might react differently compared to brains unaffected by stroke (Kleih et al., 2016).
Second, language that is thought by a patient, and therefore exists in the brain, can be communicated to the environment through the use of a BCI. Although the patient might not be able to speak, they may be able to communicate by using a BCI system to decode intended speech, e.g. from electrocorticogram signals (Miller et al., 2020; Panachakel & Ramakrishnan, 2021). As a result, neuronal networks that produce covert speech might be supported in cortical plasticity, which facilitates rehabilitation. This could increase neural plasticity more effectively compared to making the patient write, type, or express themselves in a different way (Kleih et al., 2016). Unfortunately, the study can not provide conclusive evidence for this hypothesis since it did not assess changes in patients' language abilities.

1.3.2 Auditory BCI in aphasia rehabilitation

Musso et al et al. (2022) developed an auditory ERP-based BCI paradigm for the cognitive rehabilitation of aphasia patients. Ten patients (age 58 ± 11 years, 9 male, 1 female) with different types of chronic aphasia participated in the study. Patients wearing an EEG cap were seated in a ring of six speakers (extending the AMUSE paradigm, Schreuder et al., 2010). During a trial, a cueing sentence was played from one of the loudspeakers. A cueing sentence is a short sentence of which the last word is missing, e.g. *Am Ende der großen Pause läutet die ... Glocke.* After the cueing sentence was presented to the patient, a rapid stimulus sequence consisting of one target word alongside five non-target words was played. Patients were instructed to recognize and focus attention to the target word while ignoring the non-target words, evoking the P300 ERP component. During a trial, the target word would be played from the same loudspeaker as the cueing sentence and the six words were tied one-to-one to the loudspeakers. This allowed patients to exploit the spatial information of the audio stimulus by focussing their attention solely on a single loudspeaker (Musso et al., 2022). A mono-presentation of stimuli via headphones could also be used alternatively to the spatial presentation using six loudspeakers.

To classify the target vs. non-target ERP differences and verify focused attention to the target word, a regularized LDA model was maintained for each patient. The model's capacity to discriminate between the attended and unattended word, including whether this attended word was the target, was communicated to the patient through auditory and visual feedback. The patient could use this feedback as an indicator of their task success throughout training sessions, even if they are unable to speak. Most importantly, this feedback aids the patient's improvement which is associated with a functional reorganization of the language-related brain areas (Lucchese et al., 2017; Musso et al., 1999). The advantage of a BCI approach is

that it does not require overt language production to provide feedback. It offers direct feedback based on language-related brain activity. Musso et al. cite other studies (Biasiucci et al., 2018; Cervera et al., 2018) that successfully apply a similar technique by using brain-state-dependent feedback for post-stroke motor rehabilitation. Because the BCI system is able to provide feedback time-locked to a brain state it can reward effective language processing strategies which might increase brain plasticity (Musso et al., 2022).

Musso et al. reported significant improvements of each Aachen aphasia test (Huber et al., 1983) subtest in a pre-post comparison. Patients with mild/moderate aphasia showed improvements in naming, repetition and writing while patients with moderate/severe aphasia mostly improved in the token test. In addition, patients substantially improved in functional communication, expressed by higher self-reported quality and quantity of language production. Lastly, patients also showed increased P300 peak amplitudes in channel Cz, earlier P300 onsets in channel Cz, and increased target/non-target classification accuracy.

Research has shown that a stroke in the left hemisphere of the brain can cause aphasia by directly damaging the dual language system (Ueno et al., 2011). However, it can also cause an imbalance between the default mode network (DMN) and the language network (Geranmayeh et al., 2016) located in left-dominant fronto-temporo-parietal brain regions (Catani et al., 2005; Musso et al., 2015; Saur et al., 2008). The DMN is a network of interacting brain regions that works as a domain-general system for attention and cognitive control. This imbalance between the DMN and the language network is partially reversed when speech production improves (Geranmayeh et al., 2016). This can also be observed in the pre-post fMRI recordings of the study by Musso et al. (2022) which show that their BCI-based language training rebalanced the language network and DMN. These fMRI recordings showed a decreased functional connectivity between the DMN and the main hubs of the language network (Musso et al., 2022).

The researchers attribute these significant results to their design decision to implement an elementary language task. Rather than providing feedback based on N400 or P600 ERP components as markers of language comprehension (Hagoort et al., 1996) or syntactic processing (Patel et al., 1998), respectively. They use a simpler setup that rewards effective language processing strategies based on the P300 ERP component, which even patients with severe aphasia can still perform. Using this elementary language task Musso et al. were able to train up basic language abilities which in turn facilitated higher language skills, such as language production. Since basic and higher language abilities partially share the same brain areas which have been proven to interact with each other (Friederici & Gierhan, 2013; Indefrey & Levelt, 2000; Okada & Hickok, 2006).
The BCI language training setup by Musso et al. is intended to be used as assistive technology in end users' homes or in post-stroke rehabilitation centers. Hence, efforts are necessary to transfer the BCI setup from the laboratory to these environments.

1.4 Aims and hypotheses

This study aims to investigate whether a simplified audio setup is possible for the BCI-based language training paradigm by Musso et al. (2022) by conducting a behavioral study and evaluating the subjective ergonomic experience and workload of the proposed setup.

For auditory BCIs, spatial stimulus presentation provides significant advantages (Denzer 2016; Gao et al. 2011; Schreuder et al. 2010). In the original paradigm by Musso et al. spatial information can only be exploited in the six loudspeaker condition. However, this complex setup requires a ring of six equally-spaced loudspeakers which is not feasible for everyday BCI use. Therefore, for the purpose of developing an applicable BCI paradigm for everyday usage, the question is whether a simplified audio setup using stereo headphones for the spatial presentation of words can achieve similar results. Hence, the present study aims to investigate the difference between spatial presentation in real space via speakers and virtually simulated spatial presentation in an auditory BCI paradigm. Since virtually presented sounds might be perceived differently than sounds presented in real space via loudspeakers, and the advantage of virtually simulated spatial presentation might not be as high as spatial presentation in real space. Furthermore, this study will investigate the effect of pitch on stimulus discriminability. As it has been shown to improve stimulus discriminability in an auditory BCI paradigm (Höhne et al., 2011).

This study will introduce two stereo conditions that will be used alongside the six loudspeaker and mono-presentation conditions described in the paper by Musso et al. (2022):

1. Stereo AMUSE: stimulus presentation via stereo headphones incorporating spatial information of six directions.
2. Stereo AMUSE + pitch: stimulus presentation via stereo headphones incorporating spatial information while shifting the pitch of the stimulus based on the spatial direction.

Multiple stereo conditions were tested before deciding on these two stereo conditions. For further details see Kortenbach (2022).

The remainder of this paper will use abbreviations for the different audio conditions. 6D = six loudspeaker, ST = stereo AMUSE, SP = stereo AMUSE + pitch, MO = mono headphone.

This study will analyze the performance of the newly introduced stereo conditions compared to the 6D and MO conditions by conducting a behavioral study. For the behavioral study, participants will be asked to count the number of target words. After every trial the participant will be asked to name the target word and the amount of times they counted it.
In addition, this study will evaluate the subjective ergonomic experience and workload for the four different audio conditions. As stated above, better subjective ergonomic ratings were related to better classification accuracy (Höhne et al., 2012) and other psychological factors such as mood and motivation (Nijboer et al., 2010) affect P300 amplitude (Baykara et al., 2016; Kleih et al, 2010) and BCI performance (Kleih et al., 2011; Tangermann et al., 2011). Moreover, increasing workload decreases the amplitude of the P300 ERP response (Gopher & Donchin, 1986; Käthner et al., 2013; Kramer et al., 1986; Wintink et al., 2001). The question is whether these subjective ratings differ between paradigms using six loudspeakers versus stereo headphones for the spatial presentation of stimuli.

The results of the behavioral study, subjective ratings, and workload ratings will be compared to the peak amplitudes and latencies of the P300. The amplitude and latency of the P300 ERP components affect classification accuracy. Factors influencing the P300 response like gender, age, and handedness (Polich, 2007), are controlled for via exclusion criteria and counterbalancing.
Furthermore, this study will investigate if the results of the behavioral study, subjective ratings, and workload ratings are correlated with the classification accuracy of the system for each of the four audio conditions.

The following research questions were defined:

- RQ1a) Behavioral Study: Is there a significant difference in counting accuracy when transitioning from the 6D condition to ST versus SP versus MO?
- RQ1b) Behavioral Study and Component Analysis: Is there a correlation between the results of the behavioral study and the amplitude and latency of the P300 ERP component when transitioning from the 6D condition to ST versus SP versus MO?
- RQ1c) Behavioral Study and Classification Accuracy: Is there a correlation between the results of the behavioral study and the classification accuracy when transitioning from the 6D condition to ST versus SP versus MO?

- RQ2a) Workload Rating: Is there a significant difference between the workload ratings when transitioning from the 6D condition to ST versus SP versus MO?
- RQ2b) Workload Rating and Component Analysis: Is there a correlation between the workload ratings and the amplitude and latency of the P300 ERP component when transitioning from the 6D condition to ST versus SP versus MO?
- RQ2c) Workload Rating and Classification Accuracy: Is there a correlation between the workload ratings and classification accuracy when transitioning from the 6D condition to ST versus SP versus MO?

- RQ3a) Subjective Rating: Is there a significant difference between the subjective ergonomic ratings when transitioning from the 6D condition to ST versus SP versus MO?
- RQ3b) Subjective Rating and Component Analysis: Is there a correlation between the subjective ratings and the amplitude and latency of the P300 ERP component when transitioning from the 6D condition to ST versus SP versus MO?
- RQ3c) Subjective Rating and Classification Accuracy: Is there a correlation between the subjective ratings and classification accuracy when transitioning from the 6D condition to ST versus SP versus MO?

2. Methods

2.1 Design

The influence of stimulus presentation was investigated in a one-way within-subject study, using an auditory ERP-based BCI paradigm. To account for interpersonal variability of BCI performance and EEG signals, a within-subject research design was selected (Blankertz et al., 2010). Dependent variables were the counting results of the behavioral study, offline classification accuracy, the peak amplitude and peak latency of class discriminative ERP components, and the workload and subjective ergonomic ratings of the four audio conditions. The independent variable is represented by the four stimulus presentation conditions:

1. Six loudspeaker (6D): stimulus presentation via six equally-spaced loudspeakers.
2. Stereo AMUSE (ST): stimulus presentation via stereo headphones incorporating spatial information of six directions.
3. Stereo AMUSE + pitch (SP): stimulus presentation via stereo headphones incorporating spatial information while shifting the pitch of the stimulus based on the spatial direction.
4. Mono headphones (MO): mono-presentation of stimuli via headphones.

2.2 Participants

Data of 8 healthy participants were analyzed in this study (4 male, 4 female, $Mage$ = 20.75 years, *SD* = 0.43, range = 20-21 years). All participants reported being native Dutch speakers, having been raised monolingual, being right-handed and non-musicians, having no hearing problems, no traumatic brain injury, and no neurological, psychological or psychiatric conditions. None of the participants had prior experience with BCI paradigms. Participants were recruited at Radboud University Nijmegen. Participation was voluntary and participants did not receive monetary reimbursement. All participants signed informed consent forms prior to participation in this study (see Appendix: A). Data recordings were conducted using pseudonyms and the study was approved by the Ethics committee of the Faculty of Social Science.

2.3 Instruments

2.3.1 EEG

An EEG signal was recorded with the BrainVision Recorder (Brain products) and amplified by a multichannel EEG amplifier (BrainAmp DC, Brain Products) with 32 passive Ag/AgCl electrodes (EasyCap), grounded at channel AFz and referenced behind the left ear. Channel list: ['Fp1', 'Fp2', 'F3', 'F4', 'C3', 'C4', 'P3', 'P4', 'O1', 'O2', 'F7', 'F8', 'T7', 'T8', 'P7', 'P8', 'Fz', 'Cz', 'Pz', 'FC1', 'FC2', 'CP1', 'CP2', 'FC5', 'FC6', 'CP5', 'CP6', 'Oz', 'TP10', 'PO9', 'PO10', 'FCz', 'vEOG']. The sampling rate was 1 kHz with impedances kept below 20 kΩ (ground and reference electrodes below 5 kΩ). The signal was processed using an analog bandpass filter between 0.016 and 250 Hz before digitizing and storing it for offline analysis. Electrooculographic (EOG) artifacts were detected using an additional electrode placed below the right eye. To avoid artifacts, participants were shown their EEG signal and taught how to avoid critical movements before starting the experiment .

2.3.2 Workload Rating

The workload of the four audio conditions was evaluated using the NASA-TLX questionnaire (NASA Human Performance Research Group, 1987). The NASA-TLX is a well validated instrument for the evaluation of subjective workload (Hart, 2006) and it has recently been used as a measure of workload for BCI applications (Käthner et al., 2013; Pasqualotto et al., 2011; Riccio et al., 2011; Simon et al. 2015). The questionnaire consists of six subscales: Mental, Physical and Temporal Demands, Performance, Effort and Frustration. Each subscale has to be rated on a 21 point scale with a score ranging from 0 (low) to 100 (high). For this study, the weighting procedure of the NASA-TLX was eliminated; the ratings were averaged to provide an estimate of overall workload. A meta-analysis in which unweighted NASA-TLX scores were compared to weighted NASA-TLX scores found no significant differences (Grier, 2015). See Appendix: B for the questionnaire.

2.3.3 Subjective Ergonomic Rating

The subjective ergonomic experience questionnaire partially adapted from Höhne et al. (2012) comprised five items asking for the subjective ergonomic experience, i.e. motivation, stimulus discriminability, concentration, confidence, and overall rating, for each of the four audio conditions of the BCI paradigm. Each item had to be answered on a visual analogue scale ranging from 0 (low) to 100 (high). Furthermore, the questionnaire contained an additional question asking the participant to indicate if a specific sound stimulus was too loud/quiet when using a certain audio condition. See Appendix: B for the questionnaire.

2.4 Stimuli

2.4.1 Auditory

The Dutch sentence and word stimuli were processed and edited with Audacity 3.1.3. (for further details see De Wit, 2022). Auditory word stimuli were selected according to the following constraints: Firstly, all word stimuli consisted of bisyllabic nouns with two consecutive starting consonants. Secondly, the initial consonant and vowel of the first syllable had to be different. Words stimuli were of neutral valence according to Moors et al. (2013), depictable, and similar in duration and frequency. Furthermore, word stimuli were not contained in clinical assessment tests used in the study by Musso et al. (2022), such as the Aachen Aphasia Test. Six sentence stimuli were formulated in a way that the beginning of each sentence semantically implied only one of the six word stimuli and the corresponding word was not present in the cueing sentence. The Dutch sentence stimuli (English translation in *italics*) and the corresponding target words (in red) are shown below:

1. Ik zie mijn reflectie in de . . . spiegel
   I see my reflection in the . . . mirror

2. Ik mix een milkshake in de . . . blender
   I mix a milkshake in the blender

3. We eten aardappelen, vlees en . . . groenten
   We're eating potatoes, meat, and vegetables

4. Om zijn nek draagt de zakenman een . . . stropdas
   The businessman is wearing a tie around his neck

5. Op de kermis win ik een grote . . . knuffel
   I win a big stuffed animal at the carnival

6. Het veld wordt omgeploegd met een . . . tractor
   The field gets plowed with a tractor

2.4.2 Visual

In the six loudspeaker, stereo AMUSE, and stereo AMUSE + pitch conditions visual stimulus was displayed during sentence presentation to indicate the target speaker direction. The visual cue was not displayed during the word sequence presentation. The visual cue consisted of a circle of six dots on a black background, five white and one red dot, with the red dot corresponding to the target speaker direction (see Figure 3).

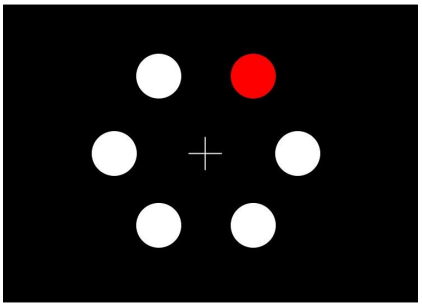


Figure 3: Visual cue to indicate target speaker direction

2.5 Auditory BCI Paradigm

2.5.1 Setup

The experimental paradigm was implemented in MATLAB (version 2020b, The Mathworks, Inc.) and Python (version 3.8.13). Sound files were transmitted through an external soundcard (Behringer X Air XR18) to loudspeakers (Audioengine A2+) or headphones Shure SE112). Due to soundcard characteristics, sound files were presented with a mean delay of 47.8 ms (SD = 1.2 ms). Sound delay was tested by playing a 440 hz sinusoid a 1000 times with a stimulus onset asynchrony (SOA) of 150 ms. Visual stimuli were displayed using a 24 inch LED flat-screen monitor (BenQ XL2420T).

For the BCI paradigm adapted from Musso et al. (2022), six loudspeakers were arranged in a ring (diameter 120 cm, angle between speakers 60°, see Figure 4). Participants were sat in a chair in the center of the ring, such that their eyes were on the diagonal line between left and right speaker (head-speaker distance 60 cm). In front of the participant was a computer screen with a fixation cross. For details see Appendix: C.

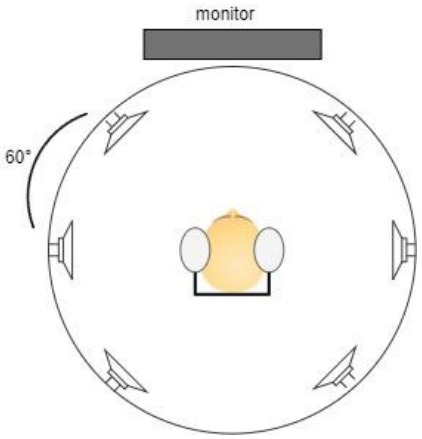


Figure 4: Experimental setup of the paradigm for the four conditions. The angle between speakers is 60° and eyes-monitor-distance is approximately 80 cm. During the experiment, stimuli were presented either via six loudspeakers, stereo headphones, stereo headphones incorporating pitch, or mono audio via headphones.

2.5.1.1 Six loudspeaker condition (6D)

In the six loudspeaker condition, stimuli were presented from all six speakers. During a trial, each of the six words appeared from the same six speaker directions. Likewise, the cueing sentence and corresponding target word were also presented from the same speaker direction. For every trial the relation between word stimuli and loudspeaker direction was pseudorandomized.

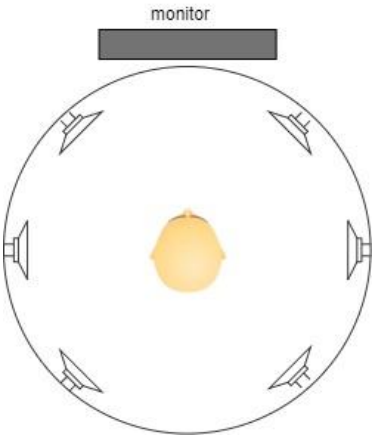


Figure 5: Six loudspeaker condition

2.5.1.2 Stereo AMUSE condition (ST)

The stereo AMUSE condition reduces the six loudspeaker setup to stereo headphones. Stimuli are presented via six virtually simulated loudspeaker directions. These loudspeaker directions are the same as the ones used in the six loudspeaker condition. This allows the participant to exploit spatial information with a less complicated setup.

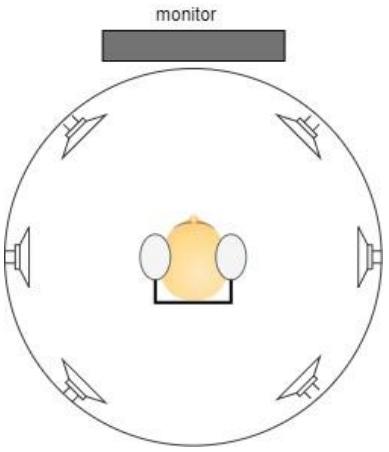


Figure 6: Stereo AMUSE condition

2.5.1.3 Stereo AMUSE + pitch condition (SP)

The stereo AMUSE + pitch condition uses the same virtually simulated loudspeaker directions as the aforementioned condition. Additionally, stimuli are varied in pitch to increase discriminability. Stimuli from certain directions are always presented in the same pitch. In this condition, stimuli presented via the two frontal speakers have a higher pitch (+1 tone), while stimuli presented via the two speakers in the back have a lower pitch (- 1.5 tone).

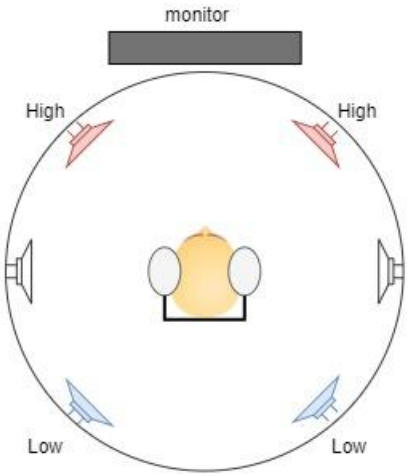


Figure 7: Stereo AMUSE + pitch

2.5.1.4 Mono headphone condition (MO)

In the mono headphone condition stimuli were presented via headphones with one mono-channel. As a consequence, spatial information of the stimulus could not be exploited. For the first two subjects audio stimuli were presented only in the left ear (see Section 4.5.1 for further details). For the remaining six subjects audio stimuli were presented in both ears simultaneously.

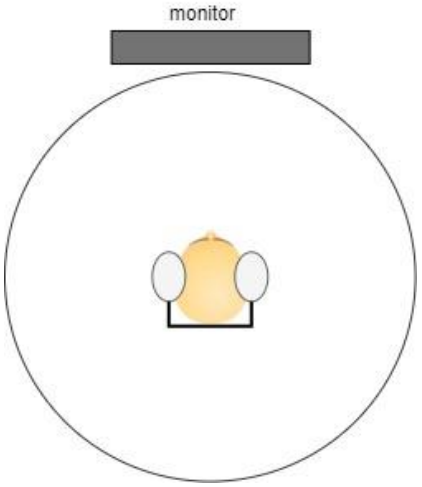


Figure 8: Mono headphone condition

2.5.2. Trial Structure

During a trial, one word was the target and the other five words were non-targets. A single run consisted of six trials. During a run each of the six words was the target exactly once. The target word was indicated by a cueing sentence. In all four conditions a trial started with a ‘get ready’ cue (“Herstart”) while a fixation cross was displayed on the screen (see Figure 9). Followed by an indication of the target word via the cueing sentence with the last word missing, i.e. the target word. The cueing sentence was played from a specific speaker or virtually simulated direction to indicate the direction of the target word. Between trials of the same run, the target word changed in pseudorandom order. In the 6D, ST, and SP conditions target direction also changed in each trial. The minimum distance between target directions of two successive stimuli was one speaker. Target direction was indicated on the screen using a visual cue (see Figure 3). After the visual cue, a word sequence was presented consisting of 15 repetitions of the six words (90 total stimuli, 15 targets, 75 non-targets, SOA=250 ms, duration: 22s) in pseudorandom order such that between two targets minimum one and maximum ten non-targets were presented. Because the SOA was set to 250 ms word stimuli slightly overlapped.
Participants were instructed to fully concentrate on the target word and count how many times it was presented. During familiarization participants were told the target word would be presented exactly 15 times during each trial. The end of a trial was indicated by another audio cue (“Ontspan”). Participants were asked to report the target word and their total count after the ‘end of trial’ cue was played. Trial length was approximately 35 s due to different sentence lengths.

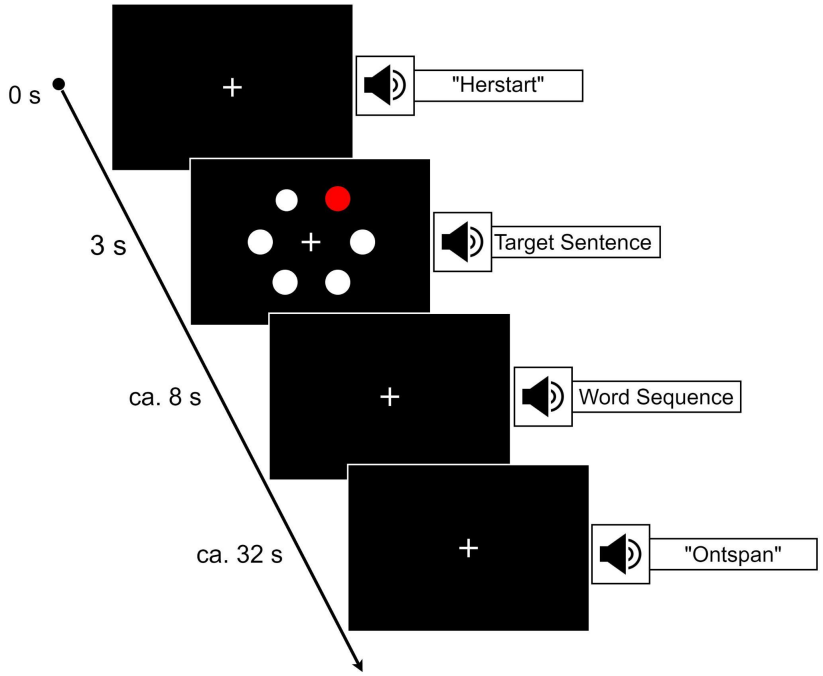


Figure 9: Example of one trial including screen display during condition 6D, ST, and SP.

2.6 BCI Session

BCI sessions took place at the Donders Institute in Nijmegen. A single session lasted about 3 hours and included the following phases: (a) preparation of EEG cap, ca. 45 min, (b) familiarization phase, ca. 15 min, (c) resting state recording, 1 min eyes open, 1 min eyes closed, (d) two recordings of the standard oddball paradigm, 2 x 5 min, (e) the auditory BCI paradigm, 1.5 - 2 hours, and (f) the workload and subjective rating questionnaires, ca. 10 min, as listed in Table 1.

2.6.1 Familiarization Phase

Familiarization took place immediately before the BCI paradigm. The familiarization phase is used to introduce participants to the audio stimuli, to learn the combinations of the target word and cueing sentences, and to prepare participants for the relatively high SOA. Familiarization was adjusted based on the performance of the participant. It generally consisted of seven consecutive steps. In the first step, each of the six words was played three times from the same speaker direction, after which the participants had to repeat the word. In step two, each cueing sentence was presented from the same speaker direction with a short pause before the last word, during which participants had to say the missing target word. Thirdly, participants completed a practice run of the six loudspeaker condition with slow SOA and less stimuli (3 targets, SOA = 1 s). Participants were instructed to count the target words in silence without moving their lips or using their fingers to count. Participants had to wait until the end of trial cue (“Ontspan”) to repeat the target word and the total count to not disrupt the EEG recording. When participants felt confident about the task, they moved on to step four in which they practiced a run of the six loudspeaker condition with the actual number of target words and SOA used in the experimental paradigm (15 targets, SOA = 0.25 s). In the last three steps, the participant practiced with the other three audio conditions (ST, SP, MO) while also using the actual number of target

words and SOA of the experimental paradigm. After every step, participants were asked if they wanted to practice more or if they felt confident enough to move on to the next step. No EEG signal was recorded during familiarization.

Phase	Element	Duration (min)
Preparation	EEG Cap	ca. 45
Resting State Recording	eyes open	1
	eyes closed	1
Oddball	OB1	5
	OB2	5
Familiarization*	F1: words	1.5
	F2: sentences	2
	F3: 6D slow	2
	F4: 6D fast	2
	F5: ST fast	2
	F6: SP fast	2
	F7: MO fast	2
BCI Paradigm	Block 1: (6D, ST, SP, MO)	24
	Block 2: (ST, SP, MO, 6D)	24
	Block 3: (SP, 6D, MO, ST)	24
	Block 4: (MO, SP, ST, 6D)	24
Questionnaires	NASA-TLX	ca. 5
	Subjective Ergonomic Rating	ca. 5

Table 1: OB = Oddball, F = Familiarization, 6D: six speakers condition, ST: stereo AMUSE condition, SP: stereo AMUSE + pitch condition, MO: mono headphone condition. *repeated if necessary.

2.6.2 Auditory Oddball

After familiarization and two resting state recordings, two runs of the oddball paradigm (OB) were presented to the participant. The OB consisted of infrequently presented target tones (high pitch: 1000 Hz) and frequently presented non-target tones (low pitch: 500 Hz). A single OB run consisted of 50 targets and 250 non-targets presented in a pseudorandom order with an SOA of 1 s, amounting to 5 min total task duration. Before the OB, participants were instructed to fully focus on the target tone and ignore the non-target tone. In addition, they needed to count the total number of target tones. The OB was presented using the six loudspeaker condition, with the target tone always presented from the same speaker direction.

2.6.3 BCI Paradigm

The auditory ERP-based BCI paradigm consisted of auditory word sequences presented in four conditions 6D, ST, SP, and MO. The trials were presented as follows. A single run consisted of six trials, such that each word was the target exactly once (see Figure 10b). Six runs were presented in one block, and the paradigm consisted of four blocks coinciding with

the four audio conditions (see Figure 10a). The order of the runs was pseudorandomized to prevent succession of auditory conditions with the goal of minimizing learning effects. The experimental paradigm consisted of 4 blocks × 4 runs × 6 trials × 90 stimuli = 8640 stimuli in total. One run took approximately 4 min, one block lasted approximately 24 min, and the complete paradigm had a duration of 96 min excluding breaks. Participants could take longer breaks between blocks (5-15 min) and short breaks between runs (0-5 min). During the longer breaks participants were provided food and drinks and they conversed with the researchers about their opinion of the different audio conditions.

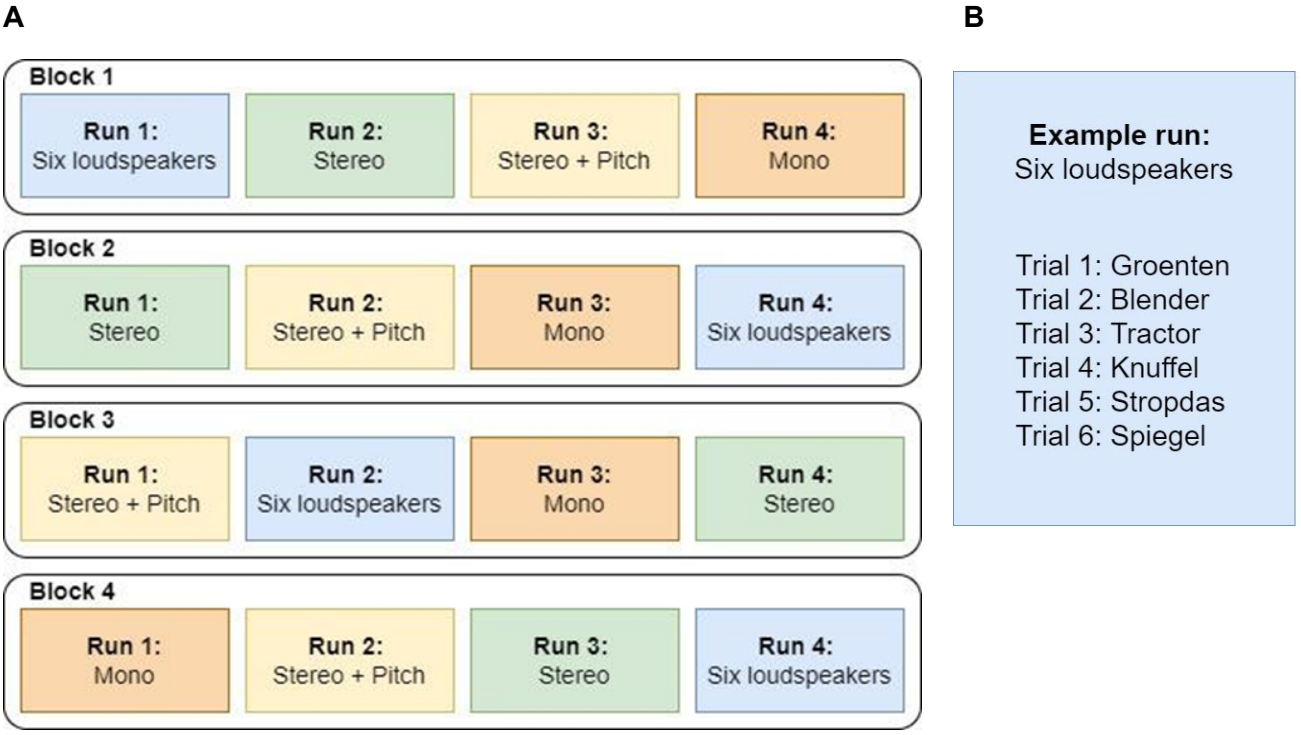


Figure 10: Paradigm procedure. (**A**) The BCI paradigm consisted of four blocks. For each audio condition, a single run was presented in pseudorandomized order in one block. A run consisted of six trials. (**B**) Example of a single run in the 6D condition. Between trials there is a break of 3 s.

2.7 Offline Signal Processing

EEG signals were analyzed offline using Python (version 3.8) and the MNE toolbox (version 1.0.3). EEG analysis was part of the bachelor's thesis by Milosevska (2022), the sections below provide a summary of the methods used in the analys.

2.7.1 Pre-processing

Sampled EEG data was pre-processed for each participant. The signal was bandpass-filtered to [0.5, 12] Hz for visualization of the averaged epochs and to [0.5, 8] Hz for classification after which it was sub-sampled to 100 Hz. The signal was split into epochs containing data from -260 ms before stimulus onset (0 ms) to +1200 ms after stimulus onset.

Epochs were rejected based on a maximum amplitude of 200μV for the EEG channels and 300μV for the EOGv, for further details see Milosevska (2022). For the remaining epochs, data from -260 ms to -10 ms (the length of a single word stimulus) were used for baseline correction. Then, the ERP signal per participant is obtained by averaging over all target epochs and all non-targets separately for each electrode.

2.7.2 Feature Extraction

Feature extraction finds a smaller subset of class-discriminative features from the high-dimensional pre-processed EEG data. For ERP analysis, feature dimensionality can be reduced by extracting suitable features like the average amplitude in certain time intervals (Blankertz et al., 2011). Based on the study by Musso et al. (2022), ten time intervals of interest [80, 150; 151, 210; 211, 280; 271, 350; 351, 440; 450, 560; 561, 700; 701, 850; 851, 1000; 1001, 1200] were selected, leading to 32 channels × 10 time intervals = 320 features.

2.8 Offline Classification

It is important to note that this was an offline study. Unlike the study by Musso et al. (2022), participants did not receive any feedback about the model's ability to discriminate between the attended and unattended word. Since the goal of this experiment was not to train the language abilities of the participants but rather to investigate the viability and user experience of the simplified audio setup.
For each participant and for each of the four audio conditions of the experimental paradigm a binary-choice classifier was trained on the pre-processed ERP signals to classify between target and non-target stimuli. Class sizes were imbalanced: the classifier was trained on 1440 targets and 7200 non-targets per participant. Moreover, to investigate whether certain word stimuli were perceived as easier or more difficult, additional binary classifiers were trained on all target and non-target presentations of each of the six word stimuli.

2.8.1 Linear Discriminant Analysis

To classify between the target and non-target responses, a shrinkage-regularized linear discriminant analysis was used (Blankertz et al., 2011). As stated previously, linear discriminant analysis (LDA) is a classification method based on finding a decision boundary to separate two classes. LDA assumes that the covariance matrices of either class are equal, that both classes are normally distributed, and that the classes are linearly separable . LDA uses a weight vector w and a bias term b to project the data x to a new space. Within this new space, a separating hyperplane can be calculated which divides the two classes. This hyperplane is defined by $w^t x + b = 0$. The output of the classifier ranges between [-1, 1], depending on the input ERPs. With -1 being the target class and 1 being the non-target class. Weights w are calculated to minimize the variance within the projected distributions, while maximizing the distance between the estimated means of the two classes. In the case of EEG data, covariance matrices show high dimensionality and there is relatively little training data (Blankertz et al., 2011). Because of this, LDA tends to overestimate large eigenvalues and underestimate small eigenvalues contained in the original covariance matrix, which in turn leads to a lower classification accuracy (Blankertz et al., 2010). Therefore, shrinkage regularization needs to be applied to the LDA model in order

to shrink the extreme eigenvalues towards the average values in the estimated covariance matrix with the goal of obtaining a lower estimation error (Höhne et al., 2015). For further details on the EEG analysis pipeline used for this study see Milosevska (2022).

2.8.2 Classification Accuracy

To evaluate classifier output, the classification accuracy was estimated using chronological 5-fold cross-validation. Data were split into five groups keeping their chronological order. The classifier was trained five times on different combinations of the other four data groups and the remaining fifth part was used as a holdout set to test classification accuracy. The accuracy of each individual participant was reported as the mean of the five test accuracies. Classification accuracy was expressed as the Area under the Receiver Operating Characteristics (ROC) Curve (AUC). The ROC curve expresses the true positive rate against the false positive rate taking into account all possible class-separation thresholds. The AUC is the area under the ROC curve and it describes the degree of separability between two classes. The AUC is independent of class sizes and robust against different numbers of epochs (Schreuder, 2014). Classification accuracies were tested for significance above chance level with permutation testing, see Milosevska (2022).

2.9 ERP Analysis

2.9.1 Grand Average ERP

The grand average ERP response was obtained by applying the signal processing procedure as described in Section 2.7 to the EEG data of each participant before averaging over all stimulus locked epochs for targets and non-targets.

2.9.2 ERP Component Analysis

For each participant, the peak amplitude and latency of the P300 ERP component were analyzed. Peak amplitudes and latencies were determined using bootstrapping in which 80 percent of the data was randomly sampled 10 times before averaging peak readouts. The peak amplitude was defined as the maximum voltage in a selected time window. The selected time windows were set to [200-400] for the negativity and [300-700] for the positivity based on visual inspection of the data. Peak latency was defined as the time between stimulus onset and maximum/minimum amplitude of the ERP component. Visual inspection of the data revealed that channels Cz and F3 had the highest target ERP amplitudes for the P300 and N200 components, respectively (Milosveska, 2022).

2.10 Statistical Analysis

After conducting the Shapiro-Wilk for all dependent variables it was concluded that the data was not normally distributed due to the small sample size (n=8) of the study. For that reason, only non-parametric tests were used for statistical analysis.
To test for differences in counting accuracy, workload ratings, and subjective ratings, a Friedman test with the within-subject factor ‘stimulus presentation’ (6D vs. ST vs. SP vs. MO) was conducted. As a measure of effect size for the Friedman test, eta squared η^2 was used with $\eta^2 < .06$ being a small effect, $.06 \leq \eta^2 \leq .13$ being a medium effect, and $\eta^2 > .13$

being a large effect (Cohen, 1988). The Friedman test is a non-parametric alternative to the repeated measure one-way ANOVA test for testing whether two or more independent samples originate from the same distribution (Friedman, 1937). A significant Friedman test indicates that there is a difference between groups. However, the test does not identify in which groups this difference occurs. Therefore, to further analyze differences between conditions, pairwise two-tailed Wilcoxon signed-rank tests were conducted (reported with Bonferroni-adjusted confidence intervals and p-values for multiple comparisons). The Wilcoxon signed-rank test is the nonparametric test equivalent to the paired-samples t-test. It can be used to compare two matched samples to assess whether their population mean ranks differ (Wilcoxon, 1945). To test if the results of the behavioral study and the workload/subjective questionnaires are correlated with classification accuracy and the peak amplitudes and latencies of the P300 ERP component, Kendall's tau-b correlations were conducted for each condition (reported with Bonferroni-adjusted p-values for multiple comparisons). Kendall's tau-b is the non-parametric equivalent to the Pearson's correlation coefficient. In this study it is used as an alternative to the non-parametric Spearman's rank-order correlation coefficient. Since Kendall's tau-b is specifically adapted to handle ties in the data and will produce a more accurate p-value in the case of a small sample size (Best & Gipps, 1974; Kendall, 1976). This makes it well suited for this study since the subjective ratings dataset contains excessive ties.

All statistical analysis was performed with Python (version 3.8) and R (version 4.1.3). The statistical significance level was set to $\alpha = .05$.

3. Results

3.1 Behavioral Data

3.1.1 Counting Task

Counting task accuracies are reported in percentage terms. Wilcoxon signed-rank test reported with Bonferroni-adjusted confidence intervals and *p*-values for multiple comparisons. In the results section Bonferroni-adjusted *p*-values will be reported as p^b. Six pairwise comparisons so α = 0.05/6 = 0.0083. In the 6D condition, subjects reached a mean counting accuracy of 98.61% (SD = 1.09, range = 96.1-99.7%), in the ST condition a mean counting accuracy of 96.33% (SD = 2.72, range = 90.0-98.6%), in the SP condition a mean counting accuracy of 97.53% (SD = 2.38, range = 92.5-99.4%), and in the MO condition a mean counting accuracy of 89.85% (SD = 6.32, range = 75.8-96.4%), see Figure 11a. It should be noted that subject four is a clear outlier in the data, see Figure 11b. During the experiment this subject indicated that they felt fatigued and they had difficulty concentrating. The counting accuracy of subject four was subpar, mainly in the MO condition, which impacted the mean counting accuracies. Counting accuracy differed significantly between conditions, as indicated by the Friedman test, $\chi 2$ = 22.77, p = <.001, η^2 = .706. Accuracies were higher in the SP condition than in the ST condition (Z = 2.66, p^b = .047) and MO condition (Z = 2.66, p^b = .047). Counting accuracy in the ST condition was higher than in the MO condition (Z = 2.66, p^b = .047). No significant difference was found between the 6D condition and the other three audio conditions, see Table 3 for all pairwise comparisons.

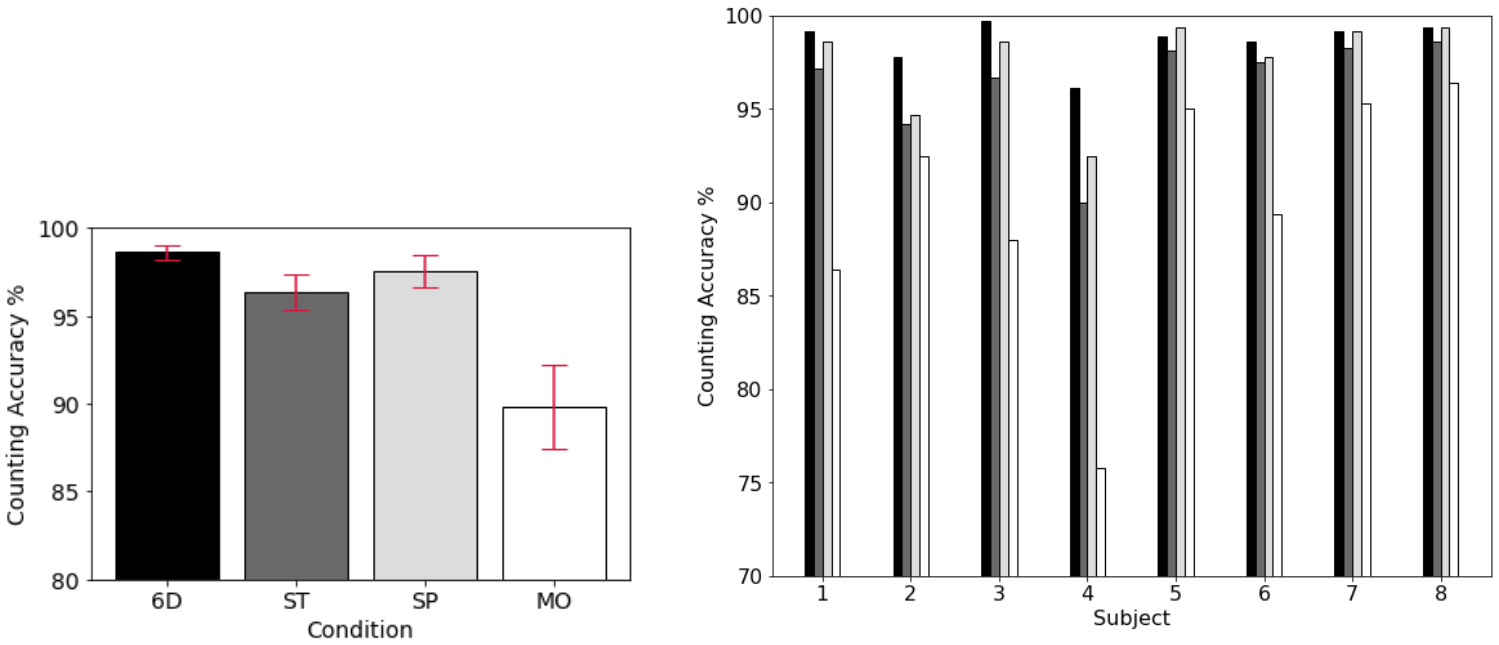


Figure 11: Counting accuracies for the four audio conditions. Bars on the left represent mean counting accuracy values, red error bars represent standard error of mean (SEM). Bars on the right depict individual counting accuracies. 6D = six loudspeaker condition, ST = stereo AMUSE condition, SP = stereo AMUSE + pitch condition, MO = mono headphone condition.

3.1.2. Correlation between counting accuracy and P300 amplitude and latency

Kendall's tau-b correlation was Bonferroni adjusted for multiple comparisons. Four comparisons in total (6D, ST, SP and MO), so $\alpha = 0.05/4 = 0.0125$.
For all four audio conditions, correlation between counting accuracy and P300 amplitude was not significant (6D: τb =.182, *p* = .618, ST: τb = .214, *p* = .548, SP: τb =0, *p* = 1, MO: τb = .357, *p* = .275). Moreover, correlation between counting accuracy and P300 latency was not significant (6D: τb = -.255, *p* = .454, ST: τb = .357, *p* = .275, SP: τb = .222, *p* = .530, MO: τb = -.071, *p* = .905). See Table 2 for the full correlation matrix.

3.1.3. Correlation between counting accuracy and classification accuracy

For all four audio conditions, correlation was not significant after Bonferroni correction (6D: τb = .618, *p* = .046, ST: τb = -.071, *p* = .905, SP: τb = .296, *p* =.379, MO: τb = .286, *p* =.399). Without Bonferroni correction there is a significant positive correlation between counting accuracy and classification accuracy in 6D (τb = .618, *p* = .046), such that the higher the counting accuracy the higher the classification accuracy.

Item	P300 Amplitude		P300 Latency		Class. Accuracy	
	τb	p	τb	p	τb	p
6D						
Counting Accuracy	.182	.618	-.255	.454	.618	.046
Motivation	-.036	1	.182	.618	.182	.618
Discriminability	.109	.803	.109	.803	.036	1
Concentration	-.286	.399	.071	.905	.071	.905
Confidence	-.071	.905	0	1	.286	.399
Overall Rating	-.182	.618	-.036	1	-.109	.803
Mean Workload	.500	.109	0	1	.143	.720
ST						
Counting Accuracy	.214	.548	.357	.275	-.071	.905
Motivation	.741	.017	-.222	.530	.667	.033
Discriminability	.109	.803	.473	.135	-.182	.618
Concentration	.327	.319	.109	.803	.182	.618
Confidence	.182	.618	.109	.803	.036	1
Overall Rating	.255	.454	.036	1	.109	.803
Mean Workload	-.071	.905	-.071	.905	.214	.548
SP						
Counting Accuracy	0	1	.222	.530	.296	.379
Motivation	.445	.167	-.148	.706	.741	.017
Discriminability	-.071	.905	.143	.720	.143	.720
Concentration	.255	.454	-.036	1	.327	.319
Confidence	.182	.618	.036	1	.036	1
Overall Rating	.296	.379	-.148	.706	.296	.379
Mean Workload	.036	1	-.109	.803	-.036	1
MO						
Counting Accuracy	.357	.275	-.071	.905	.286	.399
Motivation	.567	.075	.189	.611	.038	1
Discriminability	.473	.135	-.036	1	-.036	1
Concentration	-.109	.803	-.036	1	.182	.618
Confidence	.255	.454	.400	.213	.327	.319
Overall Rating	.340	.308	.113	.799	-.113	.799
Mean Workload	-.500	.109	-.071	.905	0	1

Table 2: Kendall's tau-b correlation matrix. 6D = six loudspeaker condition, ST = stereo AMUSE condition, SP = stereo AMUSE + pitch condition, MO = mono headphone condition. Ergonomic rating values range from 0 (low) to 100 (high).

Measurement	Item		p^b	Z	95% CI Diff.^{b} LL	UL
Behavioral Study	Counting Task	6D - ST	.085	2.45	0.85	3.60
		6D - SP	.355	1.89	0.05	3.35
		6D - MO	.085	2.45	3.45	16.00
		ST - SP*	.047	2.66	-2.50	-0.30
		ST - MO*	.047	2.66	1.70	14.20
		SP - MO*	.047	2.66	2.20	16.70
Subjective Rating	Motivation	6D - ST	.124	2.31	5.00	35.00
		6D - SP	.345	1.90	<0.01	24.50
		6D - MO	.085	2.45	22.50	70.00
		ST - SP	.511	1.72	-12.50	<0.01
		ST - MO	.085	2.45	10.00	40.00
		SP - MO	.134	2.29	27.50	48.00
	Discriminability	6D - ST	.084	2.46	20.00	35.00
		6D - SP	.132	2.29	10.00	27.50
		6D - MO	.084	2.46	30.00	73.00
		ST - SP	.213	2.10	-22.5	-7.00
		ST - MO	.123	2.32	2.50	46.50
		SP - MO	.094	2.42	-10.00	70.00
	Concentration	6D - ST	.443	1.79	<-0.01	31.50
		6D - SP	.827	1.48	-2.50	30.00
		6D - MO	.213	2.10	<-0.01	45.00
		ST - SP	1.000	1.29	-9.50	-2.00
		ST - MO	.213	2.10	7.50	20.00
		SP - MO	.347	1.90	10.00	28.00
	Confidence	6D - ST	.134	2.29	15.00	27.00
		6D - SP	.213	2.10	7.50	25.50
		6D - MO	.085	2.45	20.00	59.00
		ST - SP	.840	1.48	-20.00	1.00
		ST - MO	.129	2.30	15.00	32.00
		SP - MO	.084	2.46	15.00	42.00
	Overall	6D - ST	.134	2.29	10.00	35.00
		6D - SP	.132	2.29	5.00	23.00
		6D - MO	.082	2.47	25.00	55.00
		ST - SP	.186	2.16	-12.50	-7.50
		ST - MO	.081	2.47	15.00	32.00
		SP - MO	.082	2.47	20.00	44.50
Workload Rating	Mean Workload	6D - ST*	.047	2.66	-20.00	-2.50
		6D - SP	.200	2.13	-12.50	-2.50
		6D - MO	.085	2.45	-35.00	-13.80
		ST - SP	.469	1.76	-3.40	20.00
		ST - MO*	.047	2.66	-20.80	-5.00
		SP - MO*	.047	2.66	-40.00	-6.60

Table 3: Paired Wilcoxon signed-rank test for differences in counting task accuracy, subjective ratings, and workload ratings between conditions.

Note: 6D = six loudspeaker condition, ST = stereo AMUSE condition, SP = stereo AMUSE + pitch condition, MO = mono headphone condition. CI = confidence interval, LL = lower limit, UL = upper limit. Counting Task accuracies reported in percentage terms, Subjective Rating and Workload Rating values range from 0 (low) to 100 (high). [b]Value Bonferroni-adjusted for multiple comparisons, *p < .05

3.2 Workload Data

3.2.1 Workload Rating

The 6D condition had a mean workload of 40.73 (SD = 14.78, range = 18.3-65.0), the ST condition had a mean workload of 51.88 (SD = 10.69, range = 35.8-67.5), the SP condition had a mean workload of 46.86 (SD = 14.35, range = 18.3-70.8), and the MO condition had a mean workload of 66.56 (SD = 11.53, range = 45.8-84.2). Mean workload differed significantly between conditions, $\chi 2$ = 20.54, p = <.001, η^2 = .627. The workload rating was lower in the 6D condition than in the ST condition (Z = 2.66, p^b = .047). For both the ST and SP condition the workload rating was lower than for the MO condition (Z = 2.66, p^b = .047 for ST-MO and SP-MO). No significant differences were found between the other pairwise comparisons, see Table 3. These workload ratings show a clear trend: perceived workload was lower in the spatial presentation conditions 6D, ST, and SP than in the non-spatial presentation condition MO, see Figure 12. All subjects, except for subject 2, indicated that the counting task was the most difficult in the MO condition, see Appendix: D/E for counting scores and conversation notes.

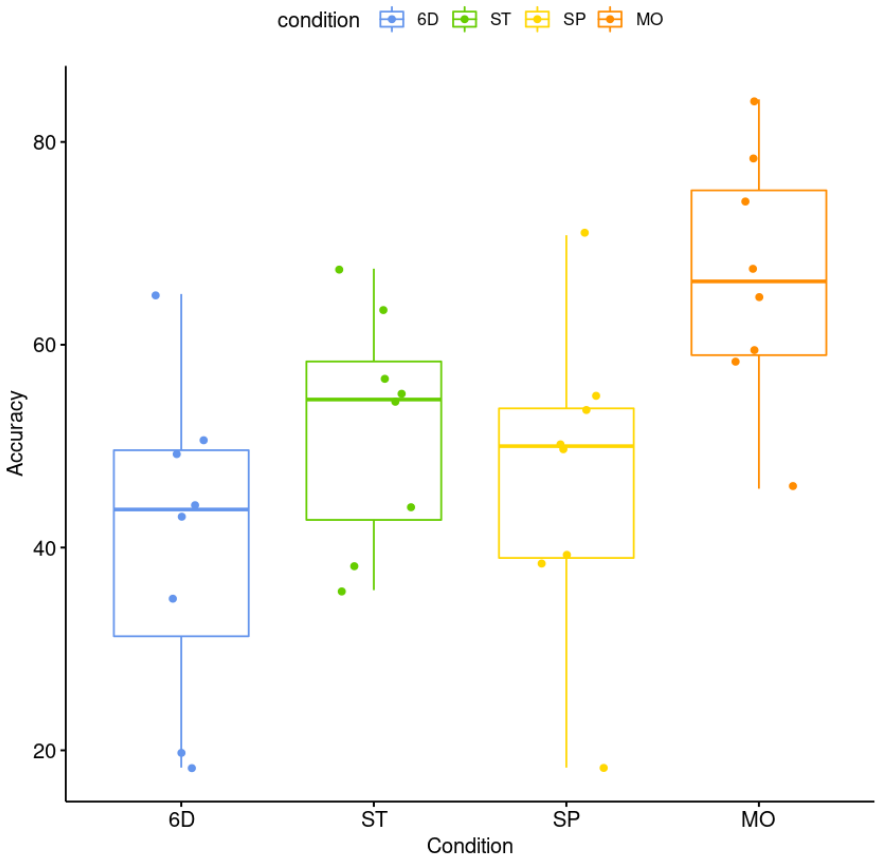


Figure 12: Boxplot showing the spread of the workload data and the outliers.

	6D		ST		SP		MO	
Item	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Mental	50.00	26.10	65.00	20.62	58.13	24.87	82.5	15.21
Physical	35.63	23.77	48.75	19.80	40.63	21.71	58.13	19.52
Temporal	50.00	18.87	56.25	19.65	55.00	20.62	73.75	22.19
Performance	20.00	18.20	27.5	12.25	28.75	21.76	43.13	20.30
Effort	49.38	20.98	60.00	17.50	53.75	21.03	75.00	14.14
Frustration	39.38	17.40	53.75	14.52	45.00	18.03	66.88	22.07
Mean Workload	40.73	14.78	51.88	10.69	46.86	14.35	66.56	11.53

Table 4: Descriptive Values for Workload Data. 6D = six loudspeaker condition, ST = stereo AMUSE condition, SP = stereo AMUSE + pitch condition, MO = mono headphone condition. Workload rating values range from 0 (low) to 100 (high).

3.2.2. Correlation between workload rating and P300 amplitude and latency

Kendall's tau-b correlation was Bonferroni adjusted, the α value was set to 0.0125. For all four audio conditions, correlation between workload ratings and P300 amplitude was not significant (6D: τb = .500, *p* = .109, ST: τb = -.071, *p* = .905, SP: τb = .036, *p* = 1, MO: τb = -.500 , *p* = .109). Moreover, correlation between workload ratings and P300 latency was not significant (6D: τb = 0, *p* = 1, ST: τb = -.071, p = .905, SP: τb = -.109, *p* = .803, MO: τb = -.071, *p* = .905).

3.2.3. Correlation between workload rating and classification accuracy

For all four audio conditions, correlation was not significant (6D: τb = .143, *p* = .720, ST: τb = .214, *p* = .548, SP: τb = -.036, *p* = 1, MO: τb = 0, *p* = 1).

3.3 Subjective Data

3.3.1 Ergonomic Rating

Ergonomic aspects concerning the four audio conditions were rated significantly different for all items, as indicated by main effects Condition for 'Motivation', $\chi 2$ = 17.81, *p* = <.001, η^2 = .529; 'Discriminability', $\chi 2$ = 19.83, *p* = <.001, η^2 = .601; 'Concentration', $\chi 2$ = 12.09 , *p* = .007, η^2 = .325, 'Confidence', $\chi 2$ = 21.00, *p* = <.001, η^2 = .643; 'Overall Rating'; $\chi 2$ = 22.68, *p* = <.001, η^2 = .703. Post-hoc Wilcoxon signed-rank tests did not find any significant differences between audio conditions, see Table 3. This is due to the small sample size of the study and excessive ties in the subjective data which reduces the statistical power of the Wilcoxon signed-rank test (discussed in Section 4.5.7). Descriptively, ergonomic ratings were higher in the spatial presentation conditions 6D, ST, and SP than in the non-spatial presentation condition MO, see Table 5. Subject 2 was the only participant to rate the MO condition higher in 'Discriminability' compared to the stereo conditions ST and SP, see Figure 13.

Item	6D		ST		SP		MO	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Motivation	76.75	12.76	57.25	9.61	63.88	12.66	29.63	12.49
Discriminability	78.63	15.92	50.75	15.71	61.88	20.23	23.50	15.46
Concentration	73.25	11.27	61.00	16.83	63.88	15.02	51.38	19.33
Confidence	84.50	10.36	66.25	13.94	73.50	17.97	46.38	21.18
Overall Rating	82.50	8.28	63.00	12.07	70.50	10.52	40.00	17.31

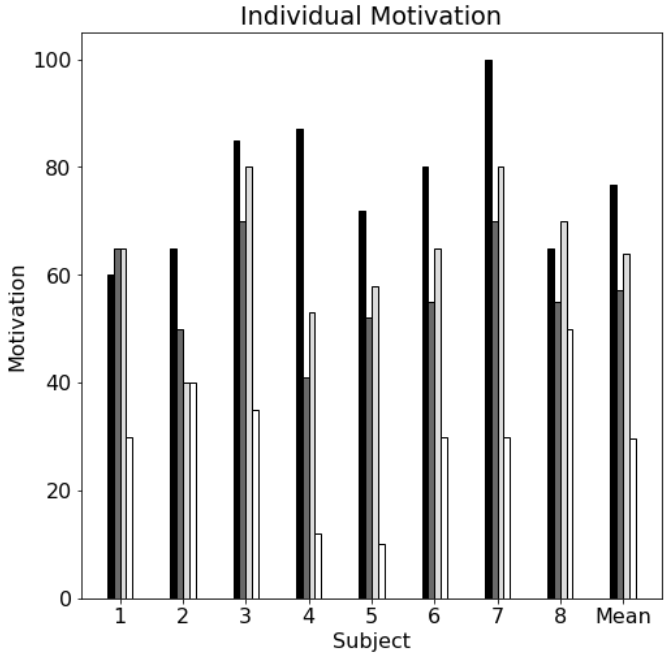
Table 5: Descriptive Values for Subjective Data. 6D = six loudspeaker condition, ST = stereo AMUSE condition, SP = stereo AMUSE + pitch condition, MO = mono headphone condition. Ergonomic rating values range from 0 (low) to 100 (high).

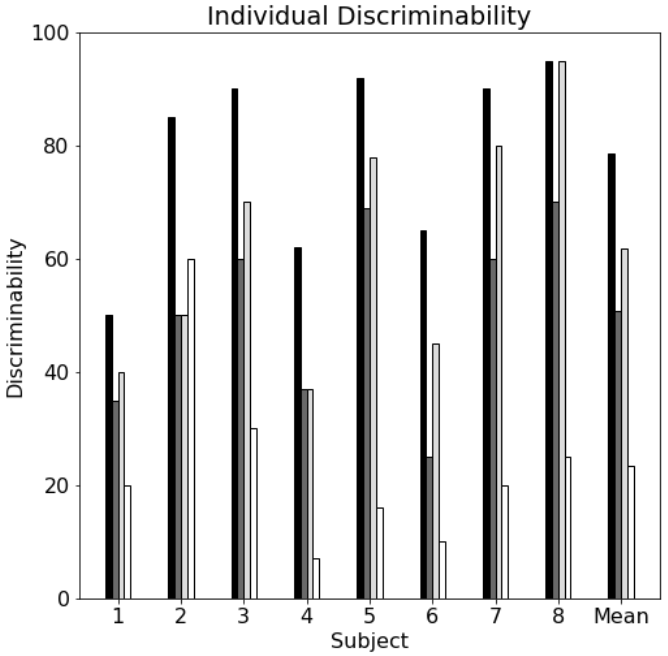
3.3.2. Correlation between ergonomic rating and P300 amplitude and latency

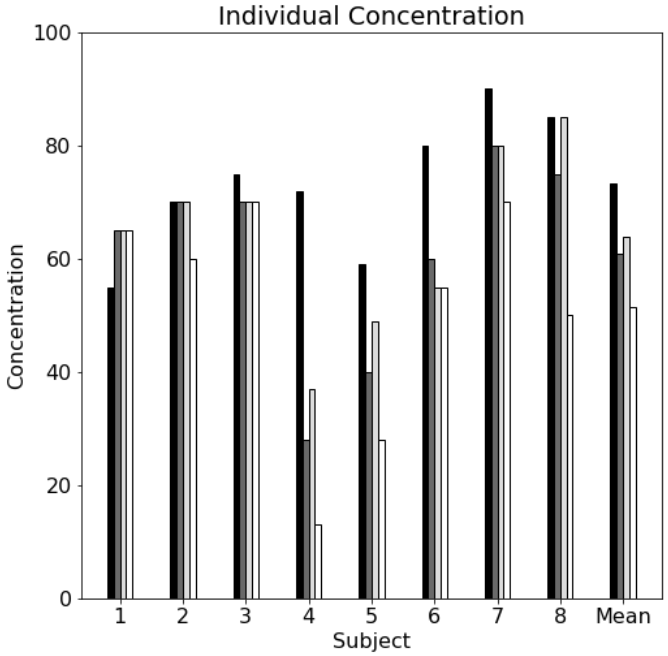
Kendall's tau-b correlation was Bonferroni adjusted, the α value was set to 0.0125.
For 'Motivation', correlation with P300 amplitude was not significant after Bonferroni correction (6D: τb = -.036, *p* = 1, ST: τb = .741, *p* = .017, SP: τb = .445, *p* = .167, MO: τb = .567, *p* = .075). Without Bonferroni correction there is a significant positive correlation between motivation and P300 amplitude in ST (τb = .741, *p* = .017), such that the higher the motivation of the subject the higher the P300 amplitude. For 'Motivation', correlation with P300 latency was not significant (6D: τb = .182, *p* = .618, ST: τb = -.222, *p* = .530, SP: τb = -.148, *p* = .706, MO: τb = .189, *p* = .611)
For 'Discriminability', 'Concentration', 'Confidence' and 'Overall Rating' correlation with P300 amplitude and latency was not significant, see Table 2.

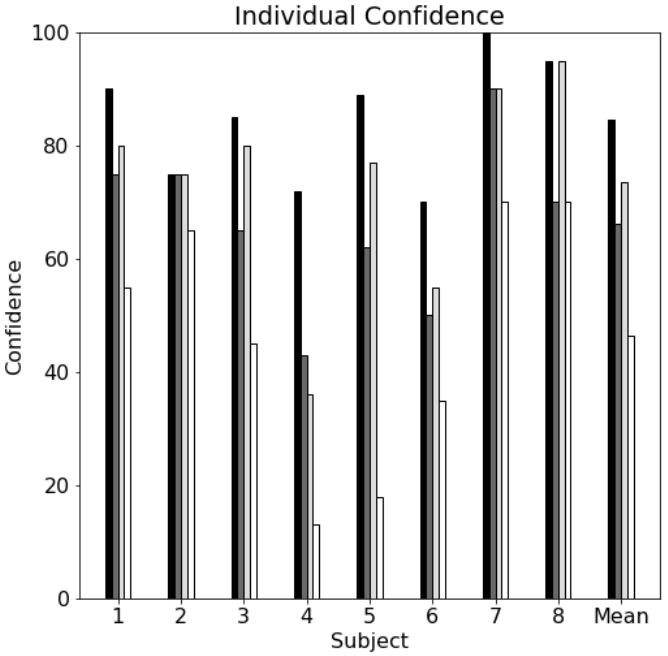
3.3.3. Correlation between ergonomic rating and classification accuracy

For 'Motivation', correlation was not significant after Bonferroni correction (6D: τb = .182, *p* = .618, ST: τb = .667, *p* = .033, SP: τb = .741, *p* = .017, MO: τb = .038, *p* = 1). Without Bonferroni correction there is a significant positive correlation between motivation and classification accuracy in ST (τb = .667, *p* = .033) and SP (τb = .741 *p* = .017), such that the higher the motivation of the subject the higher the classification accuracy.
For 'Discriminability', 'Concentration', 'Confidence' and 'Overall Rating' correlation with classification accuracy was not significant, see Table 2.









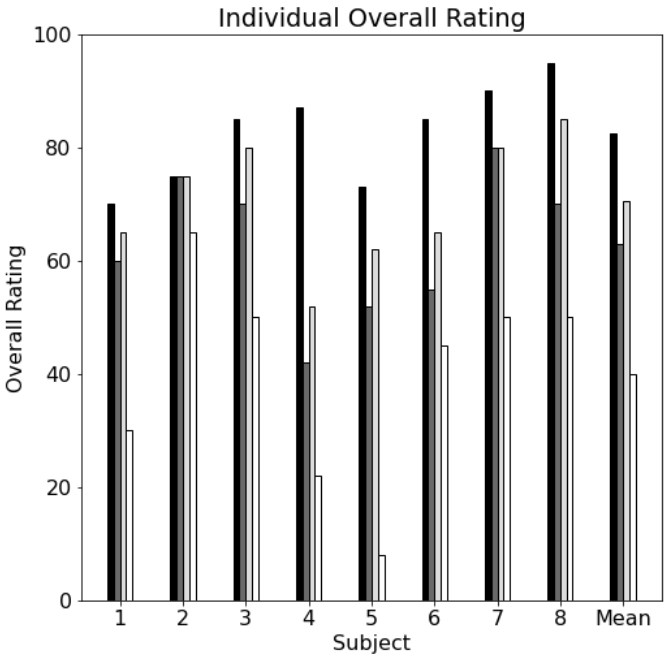


Figure 13: Individual and Mean Ergonomic Ratings. 6D = six loudspeaker condition, ST = stereo AMUSE condition, SP = stereo AMUSE + pitch condition, MO = mono headphone condition. Ergonomic rating values range from 0 (low) to 100 (high).

4. Discussion

The aim of this study was to investigate whether a simplified audio setup is possible for the BCI-based language training paradigm by Musso et al. (2022). Two new stereo conditions, stereo AMUSE (ST) and stereo AMUSE + pitch (SP), were tested versus the six loudspeaker (6D) and mono headphone (MO) conditions of the original paradigm by Musso et al. The results show that the stereo conditions are comparable to the 6D in terms of counting task performance, ergonomic ratings and workload. The 6D condition outperforms the other conditions in all aspects. However, the SP is a close second, slightly outperforming the ST condition.

4.1 Summary of Results

First, there was a significant difference in counting accuracy between the SP condition and the ST and MO condition, as well as a significant difference between the ST and MO condition, showing that the current offline BCI paradigm greatly benefits from spatial presentation of stimuli. Descriptively, counting accuracy was highest in the 6D condition. Furthermore, a significant positive correlation between counting accuracy and classification accuracy was observed in the 6D condition (without Bonferroni correction).
Spatial presentation conditions 6D, ST and SP were also superior in workload data. The 6D condition was rated significantly lower in workload than the ST condition, and the workload ratings for both the ST and SP condition were lower than for the MO condition. No significant correlation was found between workload rating and classification accuracy or the peak amplitude and latency of the P300. Lastly, ergonomic ratings concerning the four audio conditions were rated significantly different for all items according to the Friedman test. However, post-hoc Wilcoxon signed-rank tests did not reveal any significant differences between audio conditions. Descriptively, ergonomic ratings were highest in the 6D condition, followed by the SP and ST condition in second and third place respectively. Moreover, a significant positive correlation was observed between motivation and P300 amplitude in the ST condition, as well as a positive correlation between motivation and classification accuracy (both without Bonferroni correction). In the next sections the individual research questions will be addressed.

4.2 Behavioral Study

4.2.1. RQ1a: Difference in counting accuracy

Counting accuracy differed significantly between conditions, as indicated by the Friedman test. Descriptively, counting accuracy was higher for the spatial conditions 6D, ST and SP than for the non-spatial condition MO. This may indicate that, in the current attention task, with a relatively fast SOA and short stimulus length, spatial presentation helped participants discriminate the target word by cueing them to focus on a single direction. As a result, it added additional information to the stimulus information, facilitating the discrimination between stimuli and the focusing of attention. Without extra spatial information, the high semantic and phonological information of spoken word stimuli might not have been enough to identify and discriminate stimuli. Thus, this study was able to replicate the findings by Schreuder et al. (2010) and Denzer (2016) showing that spatial presentation of words adds vital information to the stimuli. Regarding research using headphones to virtually simulate

spatial presentation, the current findings are similar to the study that used spoken digits (Gao et al., 2011), which found higher classification accuracy with spatial presentation. However, counting accuracy cannot be compared directly to classification accuracy (Höhne et al., 2011). A subject can have low counting performance but a high classification accuracy, the two measurements are not always correlated. Rather, counting accuracy should be used as an additional measurement to verify target detection, i.e. a low counting accuracy means a low number of recognized targets.

Pairwise Wilcoxon signed-rank tests found a significant difference in counting accuracy between the SP condition and the ST and MO condition, as well as a significant difference between the ST and MO condition. However, the test did not find any significant differences between the 6D condition and the other three audio conditions. This is because of the way the Wilcoxon signed-rank test handles ties in the data. There are two types of ties that occur when using the Wilcoxon signed-rank test: observations in the sample may be exactly equal, or the difference between two observations may be equal. Due to the nature of the counting accuracy data set both of these types of ties occur in the pairwise comparisons. The Wilcoxon signed-rank test assigns ranks to observations based on their absolute values relative to each other. In the case of equal observations or equal differences, R takes the average of their ranks which means the Wilcoxon rank sum distribution cannot be used to calculate exact p-values. Instead, the wilcox.test function returns a normal approximated p-value along with a warning message that says “cannot compute exact p-value with ties” (R Core Team, 2016). This decreases the statistical power of the test, meaning the test is more likely to make a type II error by wrongly failing to reject the null hypothesis. In the present study the effect of ties is even larger because of the small sample size of the data. Hence, since the counting accuracy data set for the 6D condition contained too many ties, the Wilcoxon signed-rank test did not find a significant result when comparing 6D to ST, SP and MO.

In the present study, it was hypothesized that the advantage of virtually simulated spatial presentation might not be as high as spatial presentation in real space. But, as it turns out, the counting accuracy of the two stereo conditions is comparable to the 6D condition. Descriptively, the 6D condition had the highest counting accuracy (98.61%), however, the ST (96.33%) and SP (97.53%) conditions are not far behind. The counting accuracies can be compared to the study by Höhne et al. (2012) which also implemented a counting task to verify target detection. This study used the same number of target stimuli as the present study (15 targets), and also used stereo headphones to simulate spatial direction and altered the pitch of stimuli. The counting accuracy of the present study is comparable to the counting accuracy for the natural stimuli conditions used in the study by Höhne et al. (~92% for spoken syllables, and ~94% for sung syllables). The difference in counting accuracy between these studies could be explained by the difference in audio stimuli. Since spoken word stimuli contain higher semantic and acoustic information than syllables this may lead to facilitated stimulus discrimination, which is important for higher counting accuracy.

The present study also found counting accuracy to be higher in SP condition compared to the ST condition. This is in line with the findings of the PASS2D study by Höhne et al. (2011) which also used headphones for spatial presentation and altered the pitch of stimuli. This experimental paradigm, however, used artificial tone stimuli instead of spoken words. By analyzing classifier output and multiclass decisions Höhne et al. found that the classifier

could resolve the dimension “pitch” better than the dimension “direction”. Effectively showing that altering the pitch of auditory stimuli adds another dimension to the discriminability of the stimulus. Again, counting accuracy cannot be compared directly to classifier performance. Nonetheless, the use of pitch alterations may lead to improved stimulus discriminability, which in turn is important for higher counting accuracy and classification accuracy.

In sum, the present study provides evidence that spatial presentation (both virtually simulated and in real space) facilitates focusing of attention in an auditory BCI paradigm. Moreover, pitch alterations are used successfully in this paradigm to improve stimulus discriminability.

4.2.2. RQ1b: Correlation between counting accuracy and P300 amplitude and latency

Based on previous findings that P300 amplitude is linked to task-processing demand (Gopher & Donchin, 1986; Polich, 2007) and P300 latency is linked to the difficulty in differentiating target stimuli from non-targets (Duncan et al., 2009), a significant positive correlation was expected. Since a higher counting accuracy would mean the task is relatively simple and the target stimuli are easy to distinguish, this should theoretically lead to a higher P300 amplitude and shorter latency. However, no significant correlation was found between counting accuracy and P300 amplitude and latency for all four audio conditions. This might be caused by a higher latency jitter in the P300 onset due to the use of natural stimuli which contain increased semantic and acoustic information (Schreuder et al., 2010). Latency jitter is the variability between peak latencies among stimuli as well as among subjects. In general, this makes peak analysis more difficult. As a result, there may be some inaccuracy in the latency data, such that greater variability in latencies might have reduced P300 amplitude in the subjects' ERP (Luck, 2014), for further details see Milosevska (2022). Another factor affecting peak analysis is that word stimuli were presented with a mean hardware delay of 47.8 ms. This would reduce mean latency for peaks by 47.8 ms.

4.2.3. RQ1c: Correlation between counting accuracy and classification accuracy

The present study found no correlation between counting accuracy and classification accuracy. As previously stated, counting accuracy cannot be used as a direct predictor for classification accuracy (Höhne et al., 2011). Rather, it should be used as an additional measurement to verify target detection.

4.3 Workload Data

4.3.1. RQ2a: Difference in workload ratings

Mean workload differed significantly between conditions, as indicated by the Friedman test. The mean workload was lower in the 6D condition than in the ST condition. For both the ST and SP condition the workload rating was lower than for the MO condition. Lower workload ratings of the two stereo conditions ST and SP indicate that a spatial headphone paradigm might be a promising compromise between the goal of high classification accuracy and easy applicability for everyday BCI use. Interestingly enough, the Wilcoxon signed-rank test did not find a significant difference in mean workload between the 6D and the MO condition. Despite the absolute difference in mean workload being the largest between these two

conditions. Again, this might be caused by ties in the data set and the small sample size of the study. Furthermore, the standard deviation of the different subscale ratings (Mental, Physical, etc.) is substantial. This is due to the small sample size of the study and varying opinions across subjects. For example, subjects 1, 4, and 6 often indicated they felt fatigued and took multiple longer breaks (5-15 min) while the other subjects only required short breaks (0-5 min) in between blocks. See Appendix: D for breaks and an overview of comments by participants.

Even though the NASA-TLX has been the most widely used workload measure in recent years (Grier, 2015), global workload analyses have been limited to comparisons within the same test. This is due to the lack of published guidelines on the interpretation of NASA-TLX results (Hart, 2006). Specifically, there is no reference that a researcher could cite to state if an observed workload score was high or low. Therefore, to indicate if an observed workload is high or low, a comparison with similar systems or conditions is required.

When comparing the four audio conditions of the experimental paradigm, it is clear that the spatial presentation conditions 6D, ST, and SP outperform the non-spatial condition MO in terms of perceived workload. Still, the difference in mean workload between 6D and ST is significant (6D: 40.73 - ST: 51.88), as indicated by the Wilcoxon signed-rank test. The difference between 6D and SP, however, is small (6D: 40.73 - SP: 46.86). This is due to lower ratings on the ‘Mental’, ‘Physical', ‘Effort’, and ‘Frustration’ subscales compared to the ST condition. Effectively showing that the addition of pitch decreases the difficulty of the counting task and improves user experience. The MO condition was rated significantly higher in workload compared to the two stereo conditions. The largest values are observed in the ‘Mental’ (82.5), ‘Temporal’ (73,75), and ‘Effort’ (75.0) subscales. This indicates that the fast presentation of short word stimuli is not eligible for non-spatial presentation via headphones.

Unfortunately, NASA-TLX workload evaluations of auditory BCI paradigms are scarce. This makes it difficult to compare the workload evaluations of the present study to similar studies. Käthner et al. (2013) and Simon et al. (2015) both assessed workload ratings for auditory BCI paradigms. These studies will be used as a reference to compare the workload ratings of the present study.
Käthner et al. (2013) evaluated the workload of an auditory BCI spelling system which uses a combination of pitch and directional cues presented via stereo headphones. The audio stimuli consisted of a modified version of the tones introduced by Schreuder et al. (2010). Käthner et al. reported a mean workload of 57.15. The two stereo conditions used in the present study had a lower mean workload (ST: 51.88, SP: 46.86). The difference in workload can be attributed to the higher task complexity of the auditory speller. Participants had to consecutively focus on two different tones to select a target letter. This relation of task difficulty and workload was demonstrated by Käthner et al., the study found a significant negative correlation between workload and spelling accuracy (Käthner et al., 2013). The difference in mean workload could also be explained by the different audio stimuli used in the study. Käthner et al. used artificial tone stimuli while the present study used natural spoken word stimuli. Natural stimuli have been shown to improve concentration (ergonomic rating) and counting accuracy in auditory BCI paradigms (Höhne et al., 2012). Moreover, it has been found that natural stimuli reduce workload and increase BCI performance (Höhne & Tangermann, 2014; Lopez-Gordo et al., 2012).

Simon et al. (2015) measured the workload of an auditory BCI speller that uses animal sounds combined with directional cues to code rows and columns of a letter matrix. Participants faced a screen displaying a 5 × 5 letter matrix with the letters A to Y. Since the study used an auditory BCI paradigm the matrix only served as a static visual aid. Each row and each column of the matrix was coded by one of five different animal tones. To better differentiate the animal tones via headphones, each tone was coded by a single direction. Participants reported a mean workload of 62.24 in the first session and a mean workload of 69.42 in the second session. These ratings are a lot higher than the mean workload ratings of the stereo conditions used in the present study (ST: 51.88, SP: 46.86). Again, this is likely due to the higher task complexity of the auditory speller. Even though the study by Simon et al. used a longer SOA (400 ms) compared to the present study (250 ms), which has been shown to improve BCI performance (Höhne & Tangermann, 2012), the spelling task is still much more challenging compared to the simple counting task used in the current study. Furthermore, in the study by Simon et al. audio stimuli were only presented from the front. Although this does eliminate potential front-back confusions, it also lowers the interaural time difference between audio stimuli since the virtual speaker locations have to be placed closer together. This might have made it harder for participants to discriminate between spatial directions.

4.3.2. RQ2b: Correlation between workload ratings and P300 amplitude and latency

No significant correlation was found between workload ratings and amplitude for all four audio conditions. This is in contrast to the findings of other studies which showed that increasing workload and task-processing demand reduces the amplitude of the P300 (Gopher & Donchin, 1986; Käthner et al., 2013; Kramer et al., 1986; Wintink et al., 2001). Based on these findings a significant negative correlation between workload rating and amplitude was expected. Moreover, no significant correlation was found between workload ratings and latency. Again, this might be caused by a higher latency jitter in the P300 onset due to the use of natural stimuli (Schreuder et al., 2010).

4.3.3. RQ2c: Correlation between workload ratings and classification accuracy

Based on the findings by Käthner et al. (2013) a significant negative correlation between workload ratings and classification accuracy was expected for all four audio conditions of the experimental paradigm. Denzer (2016) also found a negative correlation between the ergonomic rating ‘Fatigue’, which is related to workload, and classification accuracy. However, the present study did not find a significant correlation between workload ratings and classification accuracy.

4.4 Subjective Data

4.4.1 RQ3a: Difference in subjective ergonomic ratings

According to the Friedman test, there is a significant difference in subjective ergonomic ratings between the four audio conditions. However, post-hoc pairwise Wilcoxon signed-rank tests found no significant differences between conditions. This is due to the fact that the ergonomic ratings data contains excessive ties. To illustrate: the data set for the ergonomic rating ‘Concentration’ contained the value ‘70’ seven times. Seeing as the data set only

contains 32 values in total, this is a big problem. With such a small sample size and many ties in the data the statistical power of the Wilcoxon signed-rank test is severely reduced (as discussed in section 4.5.7). Moreover, the standard deviation of the ergonomic ratings is substantial. Again, this is a result of varying opinions across subjects and a small sample size.

The same trend can be seen for all ergonomic ratings as for the counting accuracy and workload ratings: participants gave better ergonomic and workload ratings and reported the number of targets more accurately for the spatial audio conditions. Descriptively, the non-spatial MO condition was the only condition to be evaluated negative on the ergonomic scale for ‘Motivation’, ‘Discriminability’, ‘Confidence’ and ‘Overall Rating’. Based on these findings, one might suggest that the non-spatial paradigm can not be recommended as a BCI-based paradigm for aphasia rehabilitation. However, the MO condition can still prove useful in this paradigm. By removing spatial presentation as a helpful cue, the condition can increase training pressure for subjects. This means that the subject will have to focus solely on the word stimuli since they can no longer rely on spatial presentation.

Descriptively, ‘Motivation’ is higher for the 6D condition (76.75) than for the stereo conditions ST (57.25) and SP (63.88). The goal is for the simplified audio setup to be as motivating as possible. In this regard, there is a distinction to be made between aphasia patients and the healthy participants used in the present study. Aphasia patients are intrinsically motivated since they want to improve their language abilities, as opposed to healthy participants who will not use the setup for rehabilitation purposes.
Furthermore, the present study found an increase in the ergonomic aspect ‘Discriminability’ when comparing the SP (61.88) condition to the ST (50.75) condition. This is in line with the findings presented by Höhne et al. (2011).
Descriptively, the subjective ratings for ‘Concentration’ differed less between conditions compared to other ergonomic aspects. The spatial presentation conditions still outperformed the non-spatial MO condition. The ST (61.0) and SP (63.88) were rated similarly in terms of concentration.
The ratings for ‘Confidence’ show the same pattern as counting accuracy and workload ratings. Subjects who achieved high counting accuracies gave low workload ratings and high confidence ratings.
The ‘Overall Rating’ for the stereo conditions ST (63.0) and SP (70.5) was positive. This is important for the realization of an online paradigm using the simplified audio setup. Based on the findings in subjective data, again the stereo audio setup can be recommended as a simplified setup of the BCI-based language training paradigm by Musso et al. (2022). Both stereo conditions achieved adequate scores in terms of counting accuracy, workload and ergonomic ratings, further research into the effect of pitch is required to decide if the SP condition is appropriate for aphasia rehabilitation.

4.4.2 RQ3b: Correlation between ergonomic ratings and P300 amplitude and latency

Correlation between ‘Motivation’ and P300 amplitude was not significant after Bonferroni correction. Without Bonferroni correction there is a significant positive correlation between motivation and P300 amplitude in the ST condition. This result is in line with the finding that higher motivation increases the P300 amplitude (Baykara et al., 2016; Kleih et al., 2010;

Nijboer et al., 2010). These studies which investigated the effects of motivation on P300 amplitude (Nijboer et al., 2010; Baykara et al., 2016; Kleih et al., 2010) used the Questionnaire for Current Motivation (QCM) for BCI (Nijboer et al., 2008) to assess motivation in subjects. The QCM for BCI consists of 18 items which have to be rated on a seven point Likert scale. The present study, however, used a simplified questionnaire partially adapted from Höhne et al. (2012). In this questionnaire the ergonomic aspect 'Motivation' is measured by a single question: "How motivating does the 6D/ST/SP/MO condition appear to you?" In this regard, the present study is similar to Denzer (2016). The study by Denzer also evaluated 'Motivation', 'Concentration', and 'Discriminability' (although it was referred to as 'Pop-Out'). Denzer, however, did not investigate the correlation between ergonomic ratings and P300 amplitude and latency. Other ergonomic aspects that were assessed by Denzer, such as 'Fatigue', 'Exhaustion' and 'Easiness', were not included in the subjective questionnaire of the present study since these aspects were already contained in the workload questionnaire.
The ergonomic ratings 'Discriminability', 'Concentration', and 'Confidence' were not correlated with P300 amplitude and latency. This is in contrast to previous findings that the P300 amplitude is linked to task-processing demand (Gopher & Donchin, 1986; Polich, 2007), and P300 latency is linked to the discriminability of target stimuli (Duncan et al., 2009). Based on these findings, better discriminability of stimuli and a higher concentration and confidence level of the subject should have resulted in a higher P300 amplitude and better latency.

4.4.3 RQ3c: Correlation between ergonomic ratings and classification accuracy

In line with the findings by Kleih et al. (2011) and Tangermann et al. (2011), the present study found a significant positive correlation (without Bonferroni correction) between the ergonomic aspect 'Motivation' and classification accuracy in the ST condition.
The present study found no significant correlation between the other ergonomic rating items and classification accuracy. This is in contrast to the findings presented in the study by Höhne et al. (2012) which showed a positive correlation between subjective ergonomic ratings and classification accuracies. This suggests that there are notable individual differences in the participants' ergonomic experiences across conditions, such that some individuals with higher classification accuracies may have exerted more effort and thus rated their ergonomic experience worse. These differences could be taken into account in future studies by modifying stimulus properties and task demand based on individual ratings as well as BCI performance in a baseline session.

4.5 Limitations and Future Research

4.5.1 Implementation of mono headphone condition

The goal was for the MO condition to be the same as the mono headphone condition used in the study by Musso et al. (2022). Unfortunately, due to a mistake in the Python code of the experimental paradigm, audio stimuli in the MO condition were presented only in the left ear for the first two subjects. After the second subject the mistake was corrected and audio stimuli were presented in both ears simultaneously for the remaining six subjects. The counting accuracy data seems to have been impacted by this mistake, subject 1 had an accuracy of only 86.4% in the MO condition which is relatively low compared to other

subjects (excluding outlier subject 4). Subject 2 was unaffected in terms of counting accuracy, but was the only subject to rate the MO condition higher in ‘Discriminability’ compared to the stereo conditions ST and SP.

4.5.2 Stimulus length

The first three subjects listened to a word sequence with a different ‘rhythm’. This was caused by some audio files being longer in duration than others. The word stimuli were adjusted in length so that the word sequence had a better rhythm. Furthermore, some of the word stimuli were changed: the ‘P’ sound in ‘Spiegel’ and the ‘T’ sound in ‘Stropdas’ were made more noticeable and the word ‘Knuffel’ was slowed down by 15% to make it easier to distinguish. The adjusted set of word stimuli were used for the five remaining subjects. The different ‘rhythm’ of word stimuli in combination with left ear mono presentation might have caused the low counting accuracy of subject 1. Subject 3 was unaffected by the change of word stimuli in terms of counting accuracy, workload and ergonomic ratings.

4.5.3 SOA

In the BCI-based language training paradigm by Musso et al. (2022), fast feedback is necessary to close the loop between top-down and bottom-up processing in language training for aphasia patients. Short word stimuli and fast SOA facilitate fast feedback. Although a fast SOA allows for a short trial length of around 35 s, it also increases task complexity. This could pose a problem for aphasia patients who are typically much older than the healthy participants used in the present study. Previous studies reported that optimal stimulation speed varied between participants in auditory (Käthner et al., 2013) and visual (Tangermann et al., 2011) oddball paradigms. Therefore, an individually adjusted SOA might be a good option for BCI-based aphasia rehabilitation. Höhne and Tangermann (2012) also suggested adjusting the SOA per subject, as they found a high variability in classification accuracy depending on the SOA.

4.5.4 Order of conditions and learning effects

In the experimental paradigm of this study, conditions switch every run. Hence, the learning phase of one condition ends when the participant switches to a new condition. Therefore, the learning curve might have been affected by the relatively quick switching between the four audio conditions. A possible fix to the current paradigm's interruption of learning would be to present at least two consecutive runs of the same audio condition. Another option is to train the paradigm task in multiple sessions. In an online paradigm, however, participants receive feedback on their performance after each trial. This feedback has been proven to be necessary for initial learning of P300 ERP-based BCI abilities (Arvaneh et al., 2015). Feedback also improves motivation (Shute, 2008), which is correlated with increased P300 amplitude (Baykara et al., 2016; Kleih et al., 2010) and BCI performance (Kleih et al., 2011). As a result, learning might have been slower in the present offline study with no feedback. Still, the complete paradigm contained four runs for each audio condition, for a total of 24 trials. This might have been enough repetition to enable learning. Since learning was not investigated in the present study, further analysis could assess the difference between classification accuracies over time.

4.5.5 Counting task

In the present paradigm procedure, participants had to silently count the number of target words. This was a relatively simple task for the healthy subjects that participated in the study, as shown by the high counting accuracies (6D: 98.61%, ST: 96.33%, SP: 97.53%, MO: 89.85%). Therefore, other methods could have provided more informative measures. Future studies might add other measures, such as the reaction time of the subject, to better evaluate the difference in target recognition between conditions.

4.5.6 Adverse effects of pitch

While pitch improves stimulus discriminability it might also distract subjects from the true goal of language training. For example, subjects 3, 6 and 8 reported that they only focused on the pitch of audio stimuli in the SP condition, instead of the spatial direction or the words. Moreover, subjects 1 and 3 indicated that they heard the voices of multiple people in the SP condition, with the high-pitched stimuli being a childish voice and the low-pitched stimuli being a manly voice. If subjects match the word stimuli to the cueing sentence based solely on pitch rather than semantic meaning it makes language training less effective. The present study did not assess changes in subjects' language abilities since the study used healthy participants instead of aphasia patients. Hence, further investigation into the effect of altered pitch stimuli on language training is needed.

4.5.7 Ties in data

As stated previously, the data sets for counting accuracy, workload ratings and ergonomic ratings all contained ties. In the case of ties, the Wilcoxon rank sum distribution cannot be used to calculate exact p-values. Instead, the Wilcoxon signed-rank test returns a normal approximated p-value. This decreases the statistical power of the test, meaning the test is more likely to make a type II error by wrongly failing to reject the null hypothesis. In the present study the effect of ties is even larger because of the small sample size of the data. There are many ways of dealing with ties in the data when using the Wilcoxon signed-rank test. Wilcoxon (1945) suggested dropping the ties from the data altogether, and performing the test on the reduced data set but because of the small sample size of the present study this is not an option. Dropping values from the already small data set would reduce the statistical power of the test even further. Another method for handling ties involves dropping the tied ranks and assigning the remaining ranks to the other observations (Pratt, 1959). However, since the consecutive integers used in the test no longer start at one, this procedure uses normal approximation to find the critical values. A third method of handling ties randomly assigns signs to the tied ranks, but this results in a loss of efficiency (Putter, 1955). Unfortunately, the sample size of the present study is not enough to obtain sufficient statistical power to determine the significance of the results. To solve this issue, further analysis could use permutation tests using the mean differences as a metric. Permutation tests solve the issue of a small sample size by randomly permuting the signs of the observations thousands of times, finding the mean each time, and comparing the observed mean difference with the simulated permutation distribution of means.

4.5.8 Processing speed and age

The age of the participants used in this study may have had a big influence on classification accuracy and ERP components. It has been shown that age has an effect on both amplitude and latency of the P300 ERP component (Dujardin et al., 1993; Mueller et al., 2008; Tsolaki et al., 2015). Furthermore, with increasing age, processing speed slows down in terms of reaction time and ERP components (Pinal et al., 2015) and inhibition of irrelevant information is impaired (Gazzaley et al., 2005). Therefore, age definitely had an impact on the processing of stimuli and on ERP responses in this study, which in turn affected classification accuracy. Thus, further studies should include participants of the same age as the average aphasia patient to test if the current experimental paradigm is suitable for older users. The short stimuli and fast SOA used in this study could potentially complicate information processing for elderly patients. Moreover, the high number of different non-target words might be a distraction since inhibition of irrelevant stimuli is impaired with increasing age.

4.6 Conclusion

The present study was able to show that the stereo headphone paradigm might be a promising compromise between the goal of high classification accuracy and easy applicability for everyday BCI use. Moreover, the study highlights the advantage of spatial stimulus presentation and pitch in a stereo headphone paradigm while using a relatively high word sequence speed. The counting accuracy of the two stereo headphone conditions is comparable to the six loudspeaker setup. Furthermore, the stereo headphone paradigm received positive workload ratings, which is important for everyday BCI use. Subjective ergonomic ratings for the stereo conditions were positive. Additionally, a positive correlation was observed between motivation and P300 amplitude as well as classification accuracy in the ST condition. Testing young healthy participants, the present results might not be transferable to older aphasia patients, the potential users of the BCI-based language training paradigm. Thus, age is a factor to be considered in future studies. Still, the limitations of the present study provide a basis for future research. Thereby, the results of this study are valuable for designing a simplified audio setup for a future online BCI paradigm used in aphasia rehabilitation.

References

Allison, B. Z., Kübler, A., & Jin, J. (2020). 30+ years of P300 brain–computer interfaces. Psychophysiology, 57(7). https://doi.org/10.1111/psyp.13569

Angelakis E, Stathopoulou S, Frymiare JL, Green DL, Lubar JF, Kounios J. (2007) EEG neurofeedback: a brief overview and an example of peak alpha frequency training for cognitive enhancement in the elderly. Clin Neuropsychol; 21(1):110-29. doi: 10.1080/13854040600744839. PMID: 17366280.

Aphasia FAQs. (2021, 28/05). National Aphasia Association. Retrieved from https://www.aphasia.org/aphasia-faqs/ on 24/04/2022

Archibald LM, Orange JB, Jamieson DJ. Implementation of computer-based language therapy in aphasia. Ther Adv Neurol Disord. 2009 Sep;2(5):299-311. doi: 10.1177/1756285609336548. PMID: 21180620; PMCID: PMC3002597.

Arvaneh, M., Ward, T. E., & Robertson, I. H. (2015). Effects of feedback latency on P300-based brain-computer interface. Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, 2015-November, 2315–2318. http://doi.org/10.1109/EMBC.2015.7318856

Baykara, E., Ruf, C. A., Fioravanti, C., Käthner, I., Simon, N., Kleih, S. C., et al. (2016). Effects of training and motivation on auditory P300 brain-computer interface performance. Clin. Neurophysiol. 127, 379–387. doi: 10.1016/j.clinph.2015.04.054

Berthier, M.L. Poststroke Aphasia. Drugs Aging 22, 163–182 (2005). https://doi.org/10.2165/00002512-200522020-00006

Best, D.J., Gipps, P.G. (1974), Algorithm AS 71: The Upper Tail Probabilities of Kendall's Tau Applied Statistics, Vol. 23, No. 1. (1974), pp. 98-100.

Biasiucci A, Leeb R, Iturrate I, Perdikis S, Al-Khodairy A, Corbet T, Schnider A, Schmidlin T, Zhang H, Bassolino M, Viceic D, Vuadens P, Guggisberg AG, Millán JDR. (2018) Brain-actuated functional electrical stimulation elicits lasting arm motor recovery after stroke. Nat Commun. 20;9(1):2421. doi: 10.1038/s41467-018-04673-z. PMID: 29925890; PMCID: PMC6010454.

Blankertz B, Lemm S, Treder M, Haufe S, Müller KR. (2011) Single-trial analysis and classification of ERP components--a tutorial. Neuroimage. 15;56(2):814-25. doi: 10.1016/j.neuroimage.2010.06.048. Epub 2010 Jun 28. PMID: 20600976.

Blankertz, B. & Tangermann, M. & Vidaurre, C. & Fazli, S. & Sannelli, C. & Haufe, S. & Friedrichs-Maeder, C. & Ramsey, L. & Sturm, I. & Curio, G. & Müller, K. (2010). The Berlin Brain–Computer Interface: Non-Medical Uses of BCI Technology. Frontiers in neuroscience. 4. 198. 10.3389/fnins.2010.00198.

Brady, M. C., Kelly, H., Godwin, J., Enderby, P., & Campbell, P. (2016). Speech and language therapy for aphasia following stroke. Cochrane Database of Systematic Reviews. doi:10.1002/14651858.cd000425.pub4

Brungart DS, Durlach NI, Rabinowitz WM. (1999) Auditory localization of nearby sources. II. Localization of a broadband source. J Acoust Soc Am;106(4 Pt 1):1956-68. doi: 10.1121/1.427943. PMID: 10530020.

Byers, J. C., Bittner, A. C., & Hill, S. G. (1989) Traditional and raw task load index (TLX) correlations: Are paired comparisons necessary? Advances in Industrial Ergonomics and Safety. A. Mital (Ed.) Taylor & Francis., 481-485

Carelli L, Solca F, Faini A, Meriggi P, Sangalli D, Cipresso P, Riva G, Ticozzi N, Ciammola A, Silani V, Poletti B. (2017) Brain-Computer Interface for Clinical Purposes: Cognitive Assessment and Rehabilitation. Biomed Res Int.;2017:1695290. doi: 10.1155/2017/1695290. Epub 2017 Aug 23. PMID: 28913349; PMCID: PMC5587953.

Catani M, Jones DK, ffytche DH. Perisylvian language networks of the human brain. (2005) Ann Neurol. ;57(1):8-16. doi: 10.1002/ana.20319. PMID: 15597383.

Cervera MA, Soekadar SR, Ushiba J, Millán JDR, Liu M, Birbaumer N, Garipelli G. Brain-computer interfaces for post-stroke motor rehabilitation: a meta-analysis. (2018) Ann Clin Transl Neurol. 25;5(5):651-663. doi: 10.1002/acn3.544. PMID: 29761128; PMCID: PMC5945970.

Clough, S., & Gordon, J. K. (2020). Fluent or nonfluent? Part A. Underlying contributors to categorical classifications of fluency in aphasia. Aphasiology, 1–25. doi:10.1080/02687038.2020.1727709

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Hillsdale, NJ: Erlbaum.

Crerar M.A., Ellis A.W., Dean E.C. (1996) Remediation of sentence processing deficits in aphasia using a computer-based microworld. Brain Lang 52/1: 229–275

Dalemans RJ, De Witte LP, Beurskens AJ, Van Den Heuvel WJ, Wade DT. An investigation into the social participation of stroke survivors with aphasia. (2010) Disabil Rehabil.;32(20):1678-85. doi: 10.3109/09638281003649938. PMID: 20367500.

Damasio, A. R. (1992). Aphasia. New England Journal of Medicine, 326(8), 531–539. https://doi.org/10.1056/nejm199202203260806

Denzer, S. (2016). Influence of Spatial Word Presentation in an Auditory ERP Paradigm for Brain-Computer Interfaces. Unpublished.

De Wit, K. (2022). Extending a German Aphasia Rehabilitation BCI into Dutch and English Domains. Unpublished.

Dien J, Spencer KM, Donchin E. Parsing the late positive complex: mental chronometry and the ERP components that inhabit the neighborhood of the P300. (2004) Psychophysiology.41(5):665-78. doi: 10.1111/j.1469-8986.2004.00193.x. PMID: 15318873.

Doucet T, Muller F, Verdun-Esquer C, Debelleix X, Brochard P. Returning to work after a stroke: a retrospective study at the Physical and Rehabilitation Medicine Center La Tour de

Gassies. (2012) Ann Phys Rehabil Med. 55(2):112-27. English, French. doi: 10.1016/j.rehab.2012.01.007. Epub 2012 Feb 18. PMID: 22386687.

Dujardin, K., Derambure, P., Bourriez, J., Jacquesson, J., and Guieu, J. (1993). P300 component of the event-related potentials (erp) during an attention task: effects of age, stimulus modality and event probability. International Journal of Psychophysiology, 14(3):255– 267.

Duncan, C. C., Barry, R. J., Connolly, J. F., Fischer, C., Michie, P. T., Näätänen, R., … Van Petten, C. (2009). Event-related potentials in clinical research: Guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. Clinical Neurophysiology, 120(11), 1883–1908. doi:10.1016/j.clinph.2009.07.045

Duncan-Johnson CC, Donchin E. On quantifying surprise: the variation of event-related potentials with subjective probability. Psychophysiology. 1977 Sep;14(5):456-67. doi: 10.1111/j.1469-8986.1977.tb01312.x. PMID: 905483.

Edwards, S. (2005). Fluent Aphasia (Cambridge Studies in Linguistics). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511486548

Eling P, Whitaker H. Chapter 36: history of aphasia: from brain to language. Handb Clin Neurol. 2010;95:571-82. doi: 10.1016/S0072-9752(08)02136-2. PMID: 19892139.

Feyereisen, P., Pillon, A., & De Partz, M.-P. (1991). On the measures of fluency in the assessment of spontaneous speech production by aphasic subjects. Aphasiology, 5, 1–21. doi:10.1080/02687039108248516

Folstein JR, Van Petten C. Influence of cognitive control and mismatch on the N2 component of the ERP: a review. Psychophysiology. 2008 Jan;45(1):152-70. doi: 10.1111/j.1469-8986.2007.00602.x. Epub 2007 Sep 10. PMID: 17850238; PMCID: PMC2365910.

Friederici AD, Gierhan SM. The language network. Curr Opin Neurobiol. 2013 Apr;23(2):250-4. doi: 10.1016/j.conb.2012.10.002. Epub 2012 Nov 9. PMID: 23146876.

Friedman, M. (1937) The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance. American Statistical Association, 32, 675-701. http://dx.doi.org/10.1080/01621459.1937.10503522

Furdea, A., Halder, S., Krusienski, D. J., Bross, D., Nijboer, F., Birbaumer, N., & Kübler, A. (2009). An auditory oddball (P300) spelling system for brain-computer interfaces. Psychophysiology, 46(3), 617–625. doi:10.1111/j.1469-8986.2008.00783.x

Gainotti, G. (1997). Emotional, psychological and psychosocial problems of aphasic patients: an introduction. Aphasiology 11, 635–650. doi: 10.1080/02687039708249412

Gazzaley, A., Cooney, J.W., Rissman, J., & D'Esposito, M. (2005). Top-down suppression deficit underlies working memory impairment in normal aging. Nature Neuroscience, 8, 1298–1300. doi:10.1038/nn1543

Geranmayeh F, Leech R, Wise RJS. Network dysfunction predicts speech production after left hemisphere stroke. Neurology. 2016 Apr 5;86(14):1296-1305. doi: 10.1212/WNL.0000000000002537. Epub 2016 Mar 9. PMID: 26962070; PMCID: PMC4826341.

Goodglass, H., & Kaplan, E. (1983). The Assessment of Aphasia and Related Disorders. Philadelphia, PA: Lea & Febiger https://doi.org/10.1002/ana.410160524

Gopher, D., & Donchin, E. (1986). Workload: An examination of the concept. In L. Kaufman & K. Boff (Eds.), Handbook of Perception and Human Performance (pp. 41.1–41.49). New York: Wiley.

Grier, R. A. (2015). How High is High? A Meta-Analysis of NASA-TLX Global Workload Scores. Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 59(1), 1727–1731. doi:10.1177/1541931215591373

Hagoort P, Brown CM, Swaab TY. Lexical-semantic event-related potential effects in patients with left hemisphere lesions and aphasia, and patients with right hemisphere lesions without aphasia. Brain. 1996 Apr;119 (Pt 2):627-49. doi: 10.1093/brain/119.2.627. PMID: 8800953.

Haiyang Gao, Minhui Ouyang, Dan Zhang, & Bo Hong. (2011). An auditory brain-computer interface using virtual sound field. 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. doi:10.1109/iembs.2011.6091131

Halder S, Hammer EM, Kleih SC, Bogdan M, Rosenstiel W, Birbaumer N, Kübler A. Prediction of auditory and visual p300 brain-computer interface aptitude. PLoS One. 2013;8(2):e53513. doi: 10.1371/journal.pone.0053513. Epub 2013 Feb 14. PMID: 23457444; PMCID: PMC3573031.

Hart SG. Nasa-Task Load Index (NASA-TLX); 20 Years Later. Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 2006;50(9):904-908. doi:10.1177/154193120605000909

Hill, N. J., Ricci, E., Haider, S., McCane, L. M., Heckman, S., Wolpaw, J. R., & Vaughan, T. M. (2014). A practical, intuitive brain–computer interface for communicating "yes" or "no" by listening. Journal of Neural Engineering, 11(3), 035003. doi:10.1088/1741-2560/11/3/035003

Hinckley, J. J. (1998). Investigating the predictors of lifestyle satisfaction among younger adults with chronic aphasia. Aphasiology 12, 509–518. doi: 10.1080/02687039808249554

Höhne J, Bartz D, Hebart MN, Müller KR, Blankertz B. (2015) Analyzing neuroimaging data with subclasses: A shrinkage approach. Neuroimage. 2016 Jan 1;124(Pt A):740-751. doi: 10.1016/j.neuroimage.2015.09.031. Epub 2015 Sep 25. PMID: 26407815.

Höhne, J.& Krenzlin, K. & Dähne, S. & Tangermann, M. (2012). Natural stimuli improve auditory BCIs with respect to ergonomics and performance. Journal of neural engineering. 9. 045003. 10.1088/1741-2560/9/4/045003.

Höhne, J., Schreuder, M., Blankertz, B., Tangermann, M. (2011). A Novel 9-Class Auditory ERP Paradigm Driving a Predictive Text Entry System. Frontiers in neuroscience. 5. 99. 10.3389/fnins.2011.00099.

Höhne, J., Tangermann, M. (2012). How stimulation speed affects event-related potentials and BCI performance. Conf. Proc. IEEE Eng. Med. Biol. Soc. 2012, 1802–1805. doi: 10.1109/EMBC.2012.6346300

Höhne J, Tangermann M. (2014) Towards user-friendly spelling with an auditory brain-computer interface: the CharStreamer paradigm. PLoS One. 2;9(6):e98322. doi: 10.1371/journal.pone.0098322. Erratum in: PLoS One. 2014;9(7):e102630. PMID: 24886978; PMCID: PMC4041754.

Huber, W., Poeck, K., Weniger, D., & Willmes, K. (1983). AAT-Aachener aphasie test. Hogrefe, Göttingen.

Indefrey P, Levelt W J M. In: The New Cognitive Neurosciences. 2nd ed. Gazzaniga M, editor. Cambridge, MA: MIT Press; 2000. pp. 845–865.

Johnson, R. (1986). For distinguished early career contribution to psychophysiology: award address, 1985: a triarchic model of P300 amplitude. Psychophysiology 23, 367–384. doi: 10.1111/j.1469-8986.1986.tb00649.x

Johnson, R. Jr., & Donchin, E. (1978). On how P300 amplitude varies with the utility of the eliciting stimuli. Electroencephalography and Clinical Neurophysiology Volume 44, Issue 4, April 1978, Pages 424-437 https://doi.org/10.1016/0013-4694(78)90027-5

Käthner, I., Ruf, C. A., Pasqualotto, E., Braun, C., Birbaumer, N., & Halder, S. (2013). A portable auditory P300 brain–computer interface with directional cues. Clinical Neurophysiology, 124(2), 327–338. doi:10.1016/j.clinph.2012.08.006

Kendall, M.G. (1976). Rank Correlation Methods. 4th Ed. Griffin.

Kleih, S. C., Gottschalt, L., Teichlein, E., & Weilbach, F. X. (2016). Toward a P300 Based Brain-Computer Interface for Aphasia Rehabilitation after Stroke: Presentation of Theoretical

Considerations and a Pilot Feasibility Study. Frontiers in human neuroscience, 10, 547. https://doi.org/10.3389/fnhum.2016.00547

Kleih, S. C., Herweg, A., Kaufmann, T., Staiger-Sälzer, P., Gerstner, N., & Kübler, A. (2015). The WIN-speller: a new intuitive auditory brain-computer interface spelling application. Frontiers in Neuroscience, 9. https://doi.org/10.3389/fnins.2015.00346

Kleih SC, Nijboer F, Halder S, Kübler A. Motivation modulates the P300 amplitude during brain-computer interface use. Clin Neurophysiol. 2010 Jul;121(7):1023-31. doi: 10.1016/j.clinph.2010.01.034. Epub 2010 Feb 25. PMID: 20188627.

Kleih, Sonja & Riccio, Angela & Mattia, Donatella & Schreuder, Martijn & Tangermann, Michael & Zickler, Claudia & Kübler, Andrea. (2011). Motivation affects performance in a P300-Brain-Computer Interface. International Journal of Bioelectromagnetism Vol. 13, No. 1, pp. 46- 47, 2011

Kortenbach B.E. (2022) Six-channel to stereo audio transformations optimizing stimuli distinguishability in the AMUSE paradigm. Unpublished.

Kramer A, Schneider W, Fisk A, Donchin E. The effects of practice and task structure on components of the event-related brain potential. Psychophysiology. 1986 Jan;23(1):33-47. doi: 10.1111/j.1469-8986.1986.tb00590.x. PMID: 3945706.

Krusienski, D. J., Sellers, E. W., McFarland, D. J., Vaughan, T. M., & Wolpaw, J. R. (2008). Toward enhanced P300 speller performance. Journal of Neuroscience Methods, 167(1), 15–21. doi:10.1016/j.jneumeth.2007.07.017

Kübler, A., Furdea, A., Halder, S., Hammer, E. M., Nijboer, F., & Kotchoubey, B. (2009). A brain-computer interface controlled auditory event-related potential (p300) spelling system for locked-in patients. Annals of the New York Academy of Sciences, 1157, 90–100. http://doi.org/10.1111/j.1749-6632.2008.04122.x

Kübler, A., Kotchoubey, B., Kaiser, J., Wolpaw, J. R., & Birbaumer, N. (2001). Brain–computer communication: Unlocking the locked in. Psychological Bulletin, 127(3), 358–375. https://doi.org/10.1037/0033-2909.127.3.358

Linebarger M.C., Schwartz M.F., Kohn S.E. (2001) Computer-based training of language production: An exploratory study. Neuropsychol Rehab 11: 57–96

Lopez-Gordo, M. A., Fernandez, E., Romero, S., Pelayo, F., & Prieto, A. (2012). An auditory brain–computer interface evoked by natural speech. Journal of neural engineering, 9(3), 036013.

Lucchese G, Pulvermüller F, Stahl B, Dreyer FR, Mohr B. Therapy-Induced Neuroplasticity of Language in Chronic Post Stroke Aphasia: A Mismatch Negativity Study of (A)Grammatical and Meaningful/less Mini-Constructions. Front Hum Neurosci. 2017 Jan 6;10:669. doi: 10.3389/fnhum.2016.00669. PMID: 28111545; PMCID: PMC5216683.

Luck, S. J. (2014). An Introduction to the Event-Related Potential Technique. (2nd ed.,). Cambridge, MA: MIT Press.. http://doi.org/10.1118/1.4736938

Miller KJ, Hermes D, Staff NP. The current state of electrocorticography-based brain-computer interfaces. Neurosurg Focus. 2020 Jul;49(1):E2. doi: 10.3171/2020.4.FOCUS20185. PMID: 32610290.

Milosevska S. (2022) Influence of Word Presentation via Stereo Loudspeakers in an Auditory ERP Paradigm for Brain-Computer Interfaces

Moors, A., De Houwer, J., Hermans, D. et al. Norms of valence, arousal, dominance, and age of acquisition for 4,300 Dutch words. Behav Res 45, 169–177 (2013). https://doi.org/10.3758/s13428-012-0243-8

Mortley J., Wade J., Enderby P. (2004) Superhighway to promoting a client-therapist partnership? Using the internet to deliver word-retrieval computer therapy, monitored remotely with minimal speech and language therapy input. Aphasiology 18: 193–211

Mueller, V., Brehmer, Y., von Oertzen, T., Li, S.-C., & Lindenberger, U. (2008). Electrophysiological correlates of selective attention: A lifespan comparison. BMC Neuroscience, 9(1), 1–21. http://doi.org/10.1186/1471-2202-9-18

Musso M, Hübner D, Schwarzkopf S, Bernodusson M, LeVan P, Weiller C, Tangermann M. Aphasia recovery by language training using a brain-computer interface: a proof-of-concept study. Brain Commun. 2022 Feb 8;4(1):fcac008. doi: 10.1093/braincomms/fcac008. PMID: 35178518; PMCID: PMC8846581.

Musso M, Weiller C, Horn A, Glauche V, Umarova R, Hennig J, Schneider A, Rijntjes M. A single dual-stream framework for syntactic computations in music and language. Neuroimage. 2015 Aug 15;117:267-83. doi: 10.1016/j.neuroimage.2015.05.020. Epub 2015 May 19. PMID: 25998957.

Musso M, Weiller C, Kiebel S, Müller SP, Bülau P, Rijntjes M. Training-induced brain plasticity in aphasia. Brain. 1999 Sep;122 (Pt 9):1781-90. doi: 10.1093/brain/122.9.1781. PMID: 10468516.

Mondor TA, Zatorre RJ. Shifting and focusing auditory spatial attention. J Exp Psychol Hum Percept Perform. 1995 Apr;21(2):387-409. doi: 10.1037//0096-1523.21.2.387. PMID: 7714479.

NASA Human Performance Research Group. Task Load Index (NASA-TLX) NASA Ames Research Centre 1987. Available from: http://humansystems.arc.nasa.gov/groups/TLX

Nijboer. (2010). The influence of psychological state and motivation on brain-computer interface performance in patients with amyotrophic lateral sclerosis - a longitudinal study. Frontiers in Neuroscience. doi:10.3389/fnins.2010.00055

Nijboer F, Furdea A, Gunst I, Mellinger J, McFarland DJ, Birbaumer N, Kübler A. An auditory brain-computer interface (BCI). J Neurosci Methods. 2008 Jan 15;167(1):43-50. doi: 10.1016/j.jneumeth.2007.02.009. Epub 2007 Feb 20. PMID: 17399797; PMCID: PMC7955811.

Okada K, Hickok G. Left posterior auditory-related cortices participate both in speech perception and speech production: Neural overlap revealed by fMRI. Brain Lang. 2006 Jul;98(1):112-7. doi: 10.1016/j.bandl.2006.04.006. Epub 2006 May 23. PMID: 16716388.

Oknina, L. B., Kuznetsova, O. A., Belostotskyi, A. P., Nechaeva, N. L., Kutakova, E. V., Masherow, E. L., & Romanov, A. S. (2011). Amplitude-time characteristics of the long-latency components (N1, N2, and P300) of acoustic evoked potential in healthy subjects. Human Physiology, 37(1), 49–56. http://doi.org/10.1134/S0362119710061052

Panachakel, J. T., & Ramakrishnan, A. G. (2021). Decoding Covert Speech From EEG-A Comprehensive Review. Frontiers in Neuroscience, 15. https://doi.org/10.3389/fnins.2021.642251

Pasqualotto, Emanuele & Simonetta, Alessandro & Gnisci, V & Federici, Stefano & Belardinelli, Marta. (2011). Toward a Usability Evaluation of BCIs. International Journal of Bioelectromagnetism. 13. 121-122.

Patel AD, Gibson E, Ratner J, Besson M, Holcomb PJ. Processing syntactic relations in language and music: an event-related potential study. J Cogn Neurosci. 1998 Nov;10(6):717-33. doi: 10.1162/089892998563121. PMID: 9831740.

Pedersen PM, Vinter K, Olsen TS. Aphasia after stroke: type, severity and prognosis. The Copenhagen aphasia study. Cerebrovasc Dis. 2004;17(1):35-43. doi: 10.1159/000073896. Epub 2003 Oct 3. PMID: 14530636.

Picton TW, Hillyard SA. Human auditory evoked potentials. II. Effects of attention. Electroencephalogr Clin Neurophysiol. 1974 Feb;36(2):191-199. doi: 10.1016/0013-4694(74)90156-4. PMID: 4129631.

Pinal, D., Zurrón, M., & Díaz, F. (2015). Age-related changes in brain activity are specific for high order cognitive processes during successful encoding of information in working memory. Frontiers in Aging Neuroscience, 7, 75. doi:10.3389/fnagi.2015.00075http://doi.org/10.3389/fnagi.2015.00075

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. Clinical Neurophysiology, 118(10), 2128–2148. doi:10.1016/j.clinph.2007.04.019

Pratt, J. W., "Remarks on Zeros and Ties in the Wilcoxon Signed Rank Procedures," Journal of the American Statistical Association, 54 (1959), 655-67.

Putter, Joseph, "The Treatment of Ties in Some Nonparametric Tests," The Annals of Mathematical Statistics, 26 (September 1955), 368-86

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Raymer, A. M., Kohen, F. P., & Saffell, D. (2006). Computerized training for impairments of word comprehension and retrieval in aphasia. Aphasiology, 20(02-04), 257-268.

Rheinberg, Falko & Vollmeyer, Regina & Burns, Bruce. (2001). FAM: Ein Fragebogen zur Erfassung aktuller Motivation in Lern- und Leistungssituationen. Diagnostica. 47. 57-66. 10.1026//0012-1924.47.2.57.

Riccio A, Leotta F, Bianchi L, Aloise F, Zickler C, Hoogerwerf EJ, Kübler A, Mattia D, Cincotti F. Workload measurement in a communication application operated through a P300-based brain-computer interface. J Neural Eng. 2011 Apr;8(2):025028. doi: 10.1088/1741-2560/8/2/025028. Epub 2011 Mar 24. PMID: 21436511.

Richter, K., Wittler, M., and Hielscher-Fastabend, M. (2006). Bielefelder Aphasie Screening. Hofheim, Germany: NAT-Verlag

Ritter W, Simson R, Vaughan HG Jr. Association cortex potentials and reaction time in auditory discrimination. Electroencephalogr Clin Neurophysiol. 1972 Dec;33(6):547-555. doi: 10.1016/0013-4694(72)90245-3. PMID: 4117332.

Saur D, Kreher BW, Schnell S, Kümmerer D, Kellmeyer P, Vry MS, Umarova R, Musso M, Glauche V, Abel S, Huber W, Rijntjes M, Hennig J, Weiller C. Ventral and dorsal pathways for language. Proc Natl Acad Sci U S A. 2008 Nov 18;105(46):18035-40. doi: 10.1073/pnas.0805234105. Epub 2008 Nov 12. PMID: 19004769; PMCID: PMC2584675.

Schreuder M, Blankertz B, Tangermann M. A new auditory multi-class brain-computer interface paradigm: spatial hearing as an informative cue. PLoS One. 2010 Apr 1;5(4):e9813. doi: 10.1371/journal.pone.0009813. PMID: 20368976; PMCID: PMC2848564.

Sellers EW, Donchin E. A P300-based brain-computer interface: initial tests by ALS patients. Clin Neurophysiol. 2006 Mar;117(3):538-48. doi: 10.1016/j.clinph.2005.06.027. Epub 2006 Feb 7. PMID: 16461003.

Shih, J. J., Krusienski, D. J., & Wolpaw, J. R. (2012). Brain-computer interfaces in medicine. Mayo Clinic proceedings, 87(3), 268–279. https://doi.org/10.1016/j.mayocp.2011.12.008

Shih, J., Townsend, G., Krusienski, D., Shih, K., Shih, R., Heggeli, K., et al. (2013). Comparison of the checkerboard P300 speller vs. the row-column speller in normal elderly and an aphasic stroke population (S21. 006). Neurology 82:S21.006. doi:10.3217/978-3-85125-260-6-20

Shute, V. J. (2008). Focus on Formative Feedback. Review of Educational Research, 78(1), 153–189. http://doi.org/10.3102/0034654307313795

Simons, R. F., Graham, F. K., Miles, M. A., & Chen, X. (2001). On the relationship of P3a and the Novelty-P3. Biological Psychology, 56(3), 207–218. doi:10.1016/s0301-0511(01)00078-3

Simon N, Käthner I, Ruf CA, Pasqualotto E, Kübler A, Halder S. An auditory multiclass brain-computer interface with natural stimuli: Usability evaluation with healthy participants and a motor impaired end user. Front Hum Neurosci. 2015 Jan 9;8:1039. doi: 10.3389/fnhum.2014.01039. PMID: 25620924; PMCID: PMC4288388.

Speckmann, E.-J., & Elger, C. E. (1993). Introduction to the neurophysiological basis of the EEG and DC potentials. In E. Niedermeyer & F. Lopes da Silva (Eds.), Electroencephalography—Basic principles, clinical applications, and related fields (3rd ed., pp. 15-26). Baltimore: Williams & Wilkins

Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. Electroencephalography and Clinical Neurophysiology, 38(4), 387–401. doi:10.1016/0013-4694(75)90263-1

Sutton, S., Braren, M., Zubin, J., & John, E.R. (1965). Evoked-potential Correlates of Stimulus Uncertainty. Science 1965 Nov 26;150 (3700):1187-8. DOI: 10.1126/science.150.3700.1187

Sutton S, Tueting P, Zubin J, John ER. Information delivery and the sensory evoked potential. Science. 1967 Mar 17;155(3768):1436-9. doi: 10.1126/science.155.3768.1436. PMID: 6018511.

Tangermann, M., Schnorr, N., & Musso, M. (2014). Towards Aphasia Rehabilitation with BCI. In: G. Müller-Putz, G. Bauernfeind, C. Brunner, D. Steryl, S. Wriessnegger, & R. Scherer (Eds.). Proceedings of the 6th International Brain-Computer Interface Conference, Technical University of Graz, Graz, Austria, Sept. 16-19, 65–68. doi: 10.3217/978-3-85125-378-8-93.

Tangermann M, Schreuder M, Dahne S, H ¨ ohne J, Regler S, ¨ Ramsay A, Quek M, Williamson J and Murray-Smith R (2011) Optimized stimulation events for a visual ERP BCI Int. J. Bioelectromagn. 13 119–20

Teder-Sälejärvi, W.A., Hillyard, S.A. The gradient of spatial auditory attention in free field: An event-related potential study. Perception & Psychophysics 60, 1228–1242 (1998). https://doi.org/10.3758/BF03206172

Tremblay P, Dick AS. Broca and Wernicke are dead, or moving past the classic model of language neurobiology. Brain Lang. 2016 Nov;162:60-71. doi: 10.1016/j.bandl.2016.08.004. Epub 2016 Aug 30. PMID: 27584714.

Tsolaki, A., Kosmidou, V., Hadjileontiadis, L., Kompatsiaris, I. Y., and Tsolaki, M. (2015). Brain source localization of mmn, p300 and n400: aging and gender differences. Brain research, 1603:32–49.

Ueno T, Saito S, Rogers TT, Lambon Ralph MA. Lichtheim 2: synthesizing aphasia and the neural basis of language in a neurocomputational model of the dual dorsal-ventral language pathways. Neuron. 2011 Oct 20;72(2):385-96. doi: 10.1016/j.neuron.2011.09.013. PMID: 22017995.

Wascher, E., Arnau, S., Schneider, D., Hoppe, K., Getzmann, S., & Verleger, R. (2020). No effect of target probability on P3b amplitudes. International Journal of Psychophysiology. doi:10.1016/j.ijpsycho.2020.04.023

Wilcoxon, F. (1945). Individual comparisons by ranking methods. Biom. Bull., 1, 80–83.

Wintink AJ, Segalowitz SJ, Cudmore LJ. Task complexity and habituation effects on frontal P300 topography. Brain Cogn. 2001 Jun-Jul;46(1-2):307-11. doi: 10.1016/s0278-2626(01)80090-7. PMID: 11527356.

Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., & Vaughana, T. M. (2002). Brain–computer interfaces for communication and control. *Clinical Neurophysiology 113 767–791* https://doi.org/10.1016/S1388-2457(02)00057-3

Appendix A: Consent Form

Donders Centre for Cognition
STUDY-SPECIFIC INFORMED CONSENT FORM

For participation in:* ☐ Behavioural ☐ EEG ☐ Sled ☐ Robot ☐ EEG-FES
***tick the applicable box(es)**

To be filled out by the PARTICIPANT prior to the start of the experiment:
I confirm that:

- I was satisfactorily informed about the study both verbally and in writing, by means of the general information brochure and additional study specific information brochure(s) (version 2.1, December 2018).
- I have had the opportunity to put forward questions regarding the study and that these questions have been answered satisfactorily.
- I have carefully considered my participation in the experiment.
- I participate voluntarily.

I agree that:

- My research data will be acquired and stored for scientific purposes as mentioned in the general information brochure until 10 years after the research has been finalized.
- Personal data is acquired for administrative and scientific purposes.
- The connection between my personal and research data is stored until maximally one month after finalization of this study.
- Demographic data or data concerning my health, background or preferences is collected for scientific purposes.
- My not directly identifiable experimental data will be made public for verification, re-use and/or replication.
- Regulatory authorities can access my data for verification purposes.
- I will be informed by a designated expert, my general practitioner or a general practitioner of the Academic General Practitioner Center Heyendaal about any information which is of clinical relevance to me.

I understand that:

- I have the right to withdraw from the experiment at any time without having to give a reason.
- I have the right to request disposal of my experimental data up to 1 month after participation.
- My privacy is protected according to applicable European law (European General Data Protection Regulation (GDPR).
- My consent will be sought every time I participate in a new experiment.

I agree that I can be approached for a future study for comparable scientific research and to this end my contact details are stored until maximally one month after finalization of this study YES/NO
(make a choice)

I give my consent to take part in this experiment:

Name:…….. Date of birth:……..... (dd/mm/yy)

Signature:..... Date and place:……..

To be filled out by the RESEARCHER prior to the start of the experiment:
The undersigned declares that the person named above has been informed both in writing and in person about the experiment. He /she guarantees subjects' privacy protection according to Dutch law.

Name:…….. PI group:……

DCC PPF number:……

Signature:..... ……Date and place:……

December 2018, version 2.1

DONDERS INSTITUTE

DONDERS INSTITUTE

Donders Centre for Cognition

SCREENING FORM EEG research*
Version 2.1

To be filled out by the PARTICIPANT prior to the start of the experiment

Please answer the following questions first	Yes	No
- Are you younger than 18 years?		
- Have you had brain or head surgery?		
- Do you suffer from epilepsy?		

If you answered Yes to one of the above questions, you cannot participate in the experiment.

Signature:	Date:

Name of general practitioner:

Address:

** This form is only to be used for research with people of 18 years or older, who are of sound mind and judgment. The person involved has to give his or her consent personally.*

SCREENING FORM EEG research
Version 2.1

To be filled out completely by the RESEARCHER after the experiment

Adverse event		**YES/NO****
If YES:	dd/mm/yyyy	time
• Date and time of occurrence:	……..	……
• Description:		
• Severity:	mild/moderate/serious**	
• Relation to measurement procedure:	none/unlikely/possible/likely/definite**	
• Action taken:		
• Abated/ follow up:		
○ **Follow Standard Operating Procedure Adverse Event**		
Incidental Finding		**YES/NO****
If YES:		
Date:	dd/mm/yyyy ……	
○ **Follow Standard Operating Procedure Incidental Finding**		
**make a choice		

Appendix B: Workload and Ergonomic Rating Questionnaires

NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date

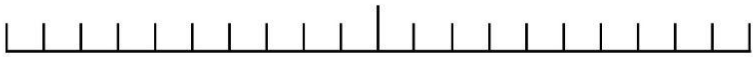
Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?



Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

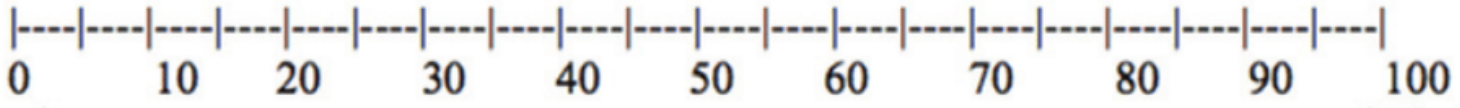
Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

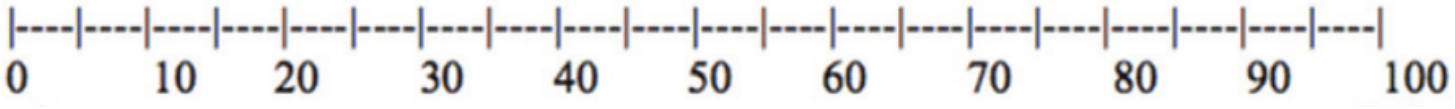
Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

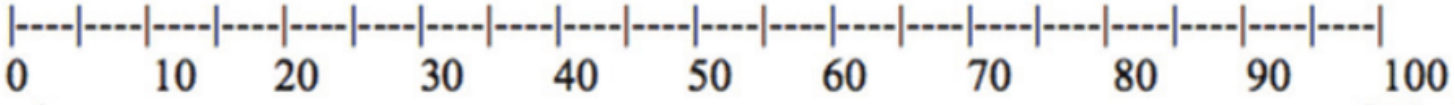
How motivating does condition X appear to you?



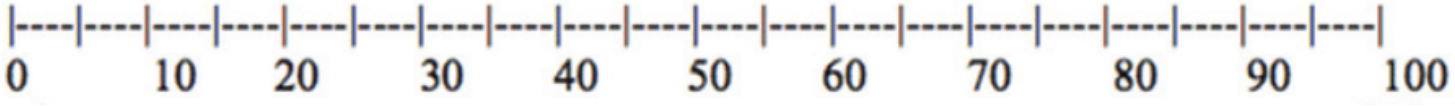
How easy was it to discriminate the stimuli in condition X?



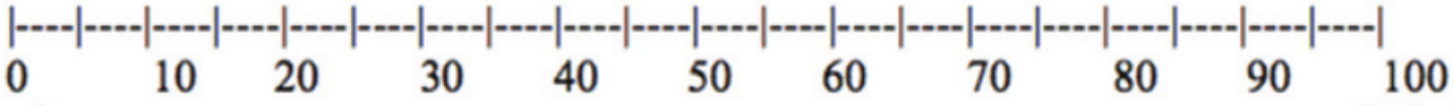
How do you judge your concentration while attending to stimuli in condition X?



How confident did you feel in your ability to count the target word in condition X?

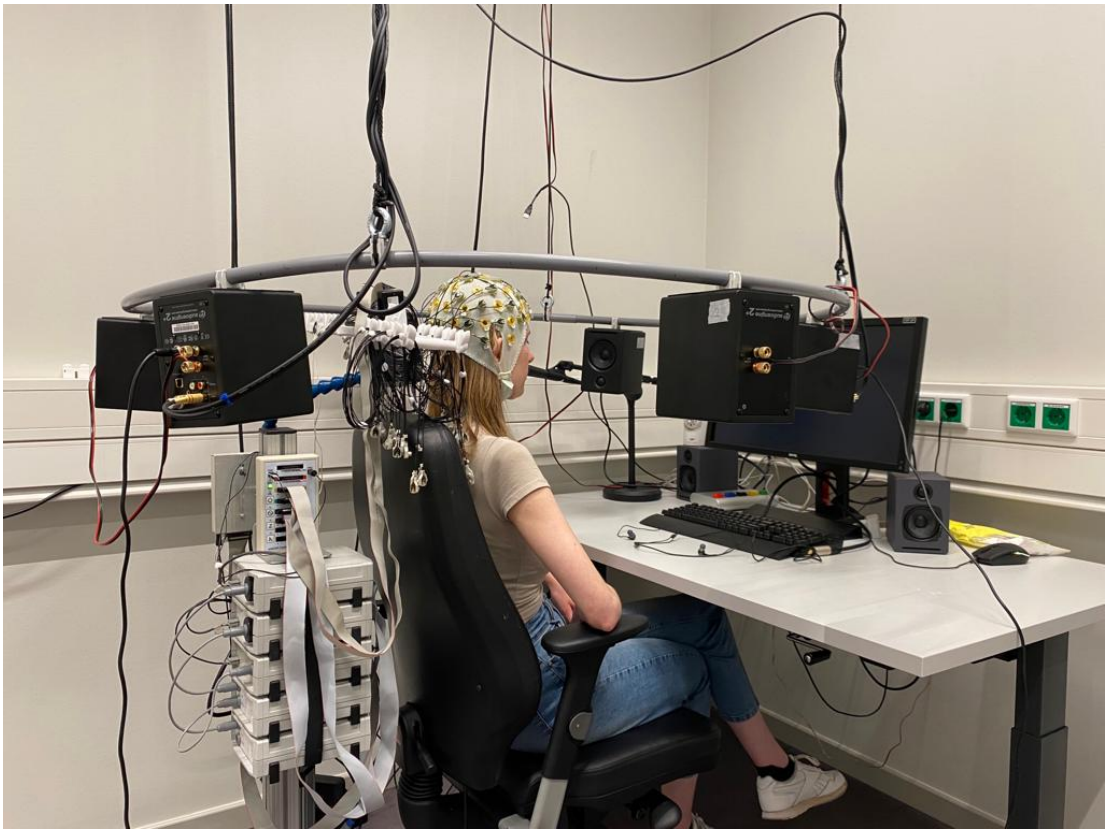


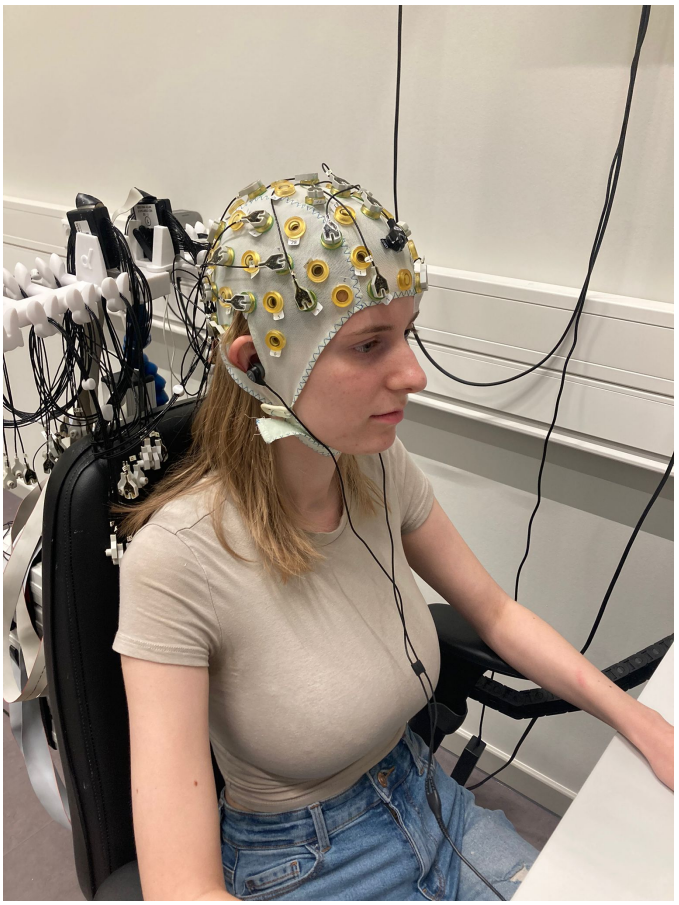
What is your overall impression of condition X?



Is there a sound stimulus that is too loud/quiet in condition X? If so, please indicate the word for which this is the case.

Appendix C: BCI Setup





Appendix D: Conversation Notes and Subject breaks

Subject 1

- The participant says SOA feels faster in mono condition, counting task is the most difficult in MO condition
- The participant indicates it is hard to discriminate between two words with the same pitch in the stereo pitch condition.
- According to the participant the audio stimuli in the SP condition feel like different voices, childish voice for high pitch and manly voice for low pitch
- Breaks: after block 2 (10 min), after block 3 (10 min)

Subject 2

- Participant indicates she cannot hear a difference between headphone conditions ST and SP.
- Breaks: after block 1 (5 min), after block 2 (5 min)

Subject 3

- Counting task is the most difficult in MO condition
- The participant indicates that they focus more on the pitch than location in the SP condition.
- According to the participant the audio stimuli in the SP condition feel like different voices, childish voice for high pitch and manly voice for low pitch
- The word ' Blender' is easier to distinguish according to the participant. The word 'Spiegel' sounds more like 'Piegel'
- Breaks: after block 2 (5 min)

Subject 4

- The participant could not recognize the word 'Blender' during familiarization
- Counting task is the most difficult in MO condition, the participants they confuse the 'A' sounds in the words 'Tractor' and 'Stropdas' in the MO condition
- Participants repeatedly indicates that they are tired, counting performance is far worse in block 4
- Breaks: after block 2 (20 min), after block 3 (10 min)

Subject 5

- The participant indicates that mono condition SOA feels faster, counting task is more difficult in MO
- Breaks: after block 3 (3 min)

Subject 6

- Counting task is more difficult in MO
- The words ‘Stropdas’ and ‘Groente’ are more apparent. ‘Knuffel’ and ‘Spiegel’ sound similar
- Participant indicates they focus solely on pitch in the SP condition
- Breaks: after block 2 (10 min), after block 3 (5 min)

Subject 7

- Counting task is more difficult in MO
- Breaks: after block 2 (5 min)

Subject 8

- Counting task is more difficult in MO
- The participant indicates SP is just as easy as 6D in terms of counting task.
- Participant mostly focused on pitch in the SP condition rather than spatial location
- Breaks: after block 2 (5 min), after block 3 (3 min)

Appendix E: Counting Scores per Subject

Note: 1 = 6D, 2 = ST, 3 = SP, 4 = MO, counting accuracies displayed below each cell

Subject 1

Block 1	1. Tractor 15 Stropdas 15 Blender 14 Groente 15 Spiegel 15 Knuffel 15 0.98888888888	2. Knuffel 14 Blender 15 Tractor 15 Stropdas 13 Spiegel 15 Groente 15 0.96666666611	3. Blender 14 Spiegel 15 Tractor 16 Knuffel 16 Groente 15 Stropdas 15 0.9666665	4. Groente 11 Tractor 15 Knuffel 10 Blender 12 Spiegel 13 Stropdas 9 0.7777778
Block 2	2. Spiegel 15 Tractor 14 Stropdas 15 Blender 15 Groente 15 Knuffel 15 0.98888888888	3. Tractor 15 Stropdas 15 Knuffel 15 Spiegel 14 Blender 15 Groente 15 0.98888888888	4. Knuffel 14 Stropdas 13 Spiegel 14 Blender 14 Tractor 13 Groente 12 0.8888889	1. Blender 15 Tractor 14 Spiegel 15 Groente 15 Knuffel 15 Stropdas 15 0.9888888888
Block 3	3. Spiegel 15 Stropdas 15 Tractor 15 Blender 15 Knuffel 15 Groente 14 0.98888888888	1. Groente 15 Tractor 15 Stropdas 14 Blender 15 Knuffel 15 Spiegel 15 0.98888888888	4. Stropdas 14 Spiegel 13 Blender 15 Knuffel 15 Groente 11 Tractor 12 0.8888889	2. Spiegel 15 Blender 15 Knuffel 15 Tractor 13 Groente 15 Stropdas 14 0.9666667
Block 4	4. Tractor 13 Spiegel 14 Knuffel 14 Blender 15 Stropdas 10 Groente 15 0.9	3. Stropdas 15 Groente 15 Knuffel 15 Spiegel 15 Blender 15 Tractor 15 1.0	2. Knuffel 15 Blender 15 Tractor 15 Spiegel 14 Groente 14 Stropdas 14 0.9666667	1. Knuffel 15 Spiegel 15 Tractor 15 Stropdas 15 Blender 15 Groente 15 1.0

Subject 2

Block 1	1. Tractor 15, Groente 15, Spiegel 15, Knuffel 14, Blender 15, Stropdas 15 0.988888	2. Blender 15, Groente 14, Tractor 13, Knuffel 16, Spiegel 14, Stropdas 14 0.933333	3. Groente 15, Stropdas 14, Knuffel 14, Spiegel 14, Blender 14, Tractor 15 0.955555	4. Groente 13, Stropdas 15, Blender 15, Tractor 14, Spiegel 11 Knuffel 13 0.9
Block 2	2. Knuffel 10, Groente 15, Spiegel 14, Blender 15, Stropdas 14, Tractor 15 0.922222	3. Groente 14, Tractor 13, Blender 14, Spiegel 15, Knuffel 14, Stropdas 15 0.944444	4. Knuffel 13, Blender 15, Tractor 14, Stropdas 14, Groente 15, Spiegel 14 0.94444	1. Groente 15, Knuffel 15, Spiegel 15, Blender 16, Stropdas 15, Tractor 15 0.988888
Block 3	3. Groente 15 Knuffel 13, Tractor 13, Spiegel 14, Blender 13, Stropdas 14 0.91111	1. Blender 15, Spiegel 14, Knuffel 15, Tractor 14, Stropdas 13, Groente 15 0.9555555	4. Groente 14 Blender 16 Stropdas 13 Spiegel 15 Knuffel 13 Tractor 14 0.92222	2. Tractor 14 Knuffel 15 Blender 14 Stropdas 14 Groente 15 Spiegel 15 0.9666666
Block 4	4. Blender 15 Groente 13 Stropdas 14 Tractor 15 Knuffel 13 Spiegel 14 0.933333	3. Spiegel 14 Blender 15 Stropdas 15 Knuffel 15 Tractor 15 Groente 14 0.977777	2. Knuffel 14 Tractor 14 Groente 14 Spiegel 15 Blender 15 Stropdas 13 0.9444444	1. Groente 14 Tractor 14 Stropdas 15 Knuffel 15 Blender 15 Spiegel 15 0.97777777

Subject 3

Block 1	1. Tractor 15 Knuffel 15 Blender 15 Stropdas 15 Groente 15 Spiegel 15 1.0	2. Knuffel 15 Blender 15 Spiegel 13 Stropdas 15 Groente 15 Tractor 14 0.96666	3. Tractor 15 Spiegel 15 Knuffel 15 Groente 15 Stropdas 15 Blender 15 1.0	4. Spiegel 14 Blender 16 Knuffel 14 Groente 12 Stropdas 10 Tractor 13 0.855555
Block 2	2. Tractor 15 Spiegel 14 Blender 15 Groente 14 Knuffel 15 Stropdas 13 0.955555	3. Knuffel 15 Groente 15 Tractor 15 Spiegel 14 Blender 15 Stropdas 15 0.988888	4. Groente 11 Knuffel 13 Tractor 14 Spiegel 12 Blender 15 Tractor 13 0.866667	1. Tractor 15 Groente 15 Knuffel 15 Spiegel 15 Blender 15 Stropdas 15 1.0
Block 3	3. Knuffel 14 Blender 14 Tractor 15 Stropdas 15 Groente 15 Spiegel 15 0.977777	1. Groente 15 Spiegel 15 Knuffel 15 Stropdas 15 Tractor 15 Blender 15 1.0	4. Stropdas 12 Blender 15 Tractor 15 Groente 12 Spiegel 9 Knuffel 11 0.8222222	2. Spiegel 15 Blender 15 Stropdas 15 Knuffel 14 Groente 13 Tractor 15 0.96666
Block 4	4. Groente 15 Blender 15 Knuffel 14 Stropdas 14 Tractor 15 Spiegel 15 0.977777	3. Stropdas 15 Knuffel 14 Tractor 15 Spiegel 15 Blender 15 Groente 14 0.977777	2. Blender 15 Knuffel 15 Groente 14 Spiegel 14 Stropdas 15 Tractor 15 0.97777	1. Spiegel 15 Knuffel 15 Groente 14 Blender 15 Stropdas 15 Tractor 15 0.98888

Subject 4

Block 1	1. Tractor 15, Stropdas 15, Knuffel 15, Groente 15, Spiegel 15, Blender 15 1.0	2. Groente 13, Stropdas 15, Tractor 15, Blender 13, Knuffel 15, Spiegel 15 0.95555	3. Spiegel 13, Stropdas 15, Blender 15, Tractor 15, Knuffel 15, Groente 15 0.977777	4. Knuffel 15, Blender 14, Spiegel 15, Tractor 13, Stropdas 11, Groente 11 0.877777
Block 2	2. Spiegel 15, Tractor 13, Knuffel 14, Stropdas 14, Blender 14, Groente 14 0.9333333	3. Blender 15, Knuffel 15, Groente 15, Tractor 15, Stropdas 15, Spiegel 14 0.988888	4. Knuffel 10, Tractor 12, Groente 11, Blender 12, Stropdas 9, Spiegel 11 0.7222222	1. Tractor 15, Groente 15, Knuffel 15, Stropdas 15, Blender 14, Spiegel 15 0.98888
Block 3	3. Tractor 16 Groente 14 Knuffel 14 Stropdas 12 Spiegel 14 Blender 15 0.922222	1. Groente 15 Spiegel 15 Stropdas 14 Knuffel 14 Blender 15 Tractor 13 0.95555	4. Spiegel 12 Blender 13 Groente 11 Stropdas 9 Knuffel 12 Tractor 10 0.74444444	2. Stropdas 14 Groente 13 Spiegel 13 Blender 13 Knuffel 8 Tractor 14 0.833333
Block 4	4. Knuffel 12, Blender 12, Stropdas 9, Tractor 9, Groente 9, Spiegel 11 0.688888	3. Tractor 15, Blender 11, Spiegel 14, Knuffel 10, Stropdas 11, Groente 12 0.8111111	2. Stropdas 10, – Tractor 12, Spiegel 15, Blender 15, Knuffel 12 0.877777	1. Knuffel 15, Groente 13, Tractor 11, Stropdas 14, Blender 13, Spiegel 15 0.9

Subject 5

Block 1	1. Blender 15 Groente 15 Spiegel 15 Knuffel 15 Tractor 15 Stropdas 15 1.0	2. Blender 15 Spiegel 15 Knuffel 14 Groente 15 Stropdas 15 Tractor 15 0.98888	3. Stropdas 15 Groente 15 Spiegel 14 Knuffel 14 Blender 15 Tractor 15 0.97777	4. Tractor 13 Blender 15 Spiegel 15 Knuffel 15 Stropdas 15 Groente 14 0.96666
Block 2	2. Spiegel 15 Tractor 15 Groente 15 Stropdas 15 Knuffel 13 Blender 16 0.96666	3. Knuffel 15 Groente 15 Stropdas 15 Spiegel 15 Blender 15 Tractor 15 1.0	4. Groente 15 Tractor 12 Stropdas 15 Blender 14 Knuffel 14 Spiegel 12 0.911111	1. Spiegel 15 Blender 15 Tractor 15 Stropdas 15 Groente 14 Knuffel 15 0.98888
Block 3	3. Knuffel 15 Spiegel 15 Tractor 15 Groente 15 Blender 15 Stropdas 15 1.0	1. Stropdas 15 Knuffel 15 Tractor 15 Blender 15 Groente 15 Spiegel 15 1.0	4. Blender 15 Spiegel 15 Tractor 15 Stropdas 14 Knuffel 17 Groente 14 0.95555	2. Blender 15 Spiegel 14 Stropdas 15 Groente 15 Knuffel 15 Tractor 15 0.98888
Block 4	4. Knuffel 15 Spiegel 15 Tractor 15 Groente 14 Blender 13 Stropdas 15 0.96666	3. Spiegel 15 Groente 15 Knuffel 15 Blender 15 Stropdas 15 Tractor 15 1.0	2. Knuffel 15 Groente 14 Tractor 15 Blender 15 Stropdas 15 Spiegel 14 0.977777	1. Knuffel 14 Stropdas 15 Tractor 16 Spiegel 15 Groente 15 Blender 14 0.96666

Subject 6

Block 1	1. Groente 15 Blender 15 Tractor 14 Knuffel 15 Stropdas 15 Spiegel 14 0.977777	2. Tractor 14 Blender 15 Knuffel 13 Groente 14 Spiegel 15 Stropdas 14 0.9444444	3. Knuffel 15 Spiegel 14 Tractor 15 Stropdas 15 Blender 14 Groente 15 0.977777	4. Stropdas 11 Spiegel 13 Blender 15 Knuffel 13 Groente 12 Tractor 14 0.8666667
Block 2	2. Groente 13 Tractor 15 Blender 15 Stropdas 15 Spiegel 15 Knuffel 15 0.977777	3. Tractor 15 Blender 15 Groente 15 Stropdas 15 Knuffel 15 Spiegel 16 0.9888888	4. Stropdas 15 Groente 11 Knuffel 15 Spiegel 12 Blender 14 Tractor 14 0.9	1. Groente 15 Tractor 15 Blender 15 Spiegel 15 Knuffel 15 Stropdas 16 0.988888
Block 3	3. Blender 15 Stropdas 16 Groente 15 Tractor 16 Knuffel 15 Spiegel 16 0.966666	1. Groente 15 Tractor 15 Blender 16 Knuffel 15 Spiegel 15 Stropdas 15 0.988888	4. Knuffel 13 Tractor 14 Spiegel 14 Stropdas 14 Groente 9 Blender 15 0.8777778	2. Tractor 15 Blender 15 Knuffel 15 Spiegel 14 Stropdas 16 Groente 15 0.977777
Block 4	4. Knuffel 14 Stropdas 16 Tractor 14 Blender 16 Groente 13 Spiegel 15 0.933333	3. Tractor 15 Spiegel 15 Stropdas 15 Blender 16 Groente 14 Knuffel 15 0.977777	2. Groente 15 Stropdas 15 Knuffel 15 Tractor 15 Spiegel 15 Blender 15 1.0	1. Tractor 14 Knuffel 15 Stropdas 15 Spiegel 15 Blender 15 Groente 15 0.98888

Subject 7

Block 1	1. Blender 15 Tractor 15 Spiegel 15 Knuffel 15 Stropdas 15 Groente 17 0.97777	2. Knuffel 15 Spiegel 15 Blender 15 Groente 15 Tractor 15 Stropdas 15 1.0	3. Tractor 17 Groente 15 Spiegel 15 Blender 15 Stropdas 15 Knuffel 15 0.97777	4. Spiegel 15 Blender 15 Tractor 16 Groente 13 Stropdas 14 Knuffel 16 0.944444
Block 2	2. Groente 15 Tractor 16 Spiegel 16 Stropdas 15 Blender 15 Knuffel 13 0.955555	3. Knuffel 15 Spiegel 15 Stropdas 15 Blender 15 Tractor 15 Groente 15 1.0	4. Tractor 12 Knuffel 15 Blender 15 Spiegel 15 Groente 14 Stropdas 13 0.93333	1. Knuffel 15 Blender 15 Tractor 16 Groente 15 Spiegel 15 Stropdas 15 0.988888
Block 3	3. Tractor 15 Stropdas 15 Blender 15 Groente 15 Spiegel 14 Knuffel 15 0.988888	1. Tractor 15 Knuffel 15 Groente 15 Blender 15 Stropdas 15 Spiegel 15 1.0	4. Blender 15 Stropdas 15 Tractor 15 Knuffel 15 Spiegel 15 Groente 16 0.988888	2. Tractor 15 Stropdas 15 Knuffel 15 Groente 15 Spiegel 15 Blender 15 1.0
Block 4	4. Groente 17 Knuffel 15 Stropdas 13 Blender 15 Spiegel 14 Tractor 15 0.944444	3. Blender 15 Knuffel 15 Stropdas 15 Tractor 15 Spiegel 15 Groente 15 1.0	2. Spiegel 15 Groente 16 Knuffel 14 Tractor 15 Blender 15 Stropdas 15 0.977777	1. Knuffel 15 Stropdas 15 Groente 15 Tractor 15 Spiegel 15 Blender 15 1.0

Subject 8

Block 1	1. Knuffel 15, Blender 15, Groente 15, Stropdas 15, Spiegel 15, Tractor 15 1.0	2. Groente 14, Spiegel 15, Stropdas 15, Tractor 15, Blender 14, Knuffel 16 0.966666	3. Stropdas 15, Spiegel 15, Knuffel 15, Blender 15, Tractor 15, Groente 15 1.0	4. Stropdas 13, Tractor 14, Spiegel 15, Knuffel 14, Groente 13, Blender 15 0.933333
Block 2	2. Spiegel 15, Groente 15, Stropdas 15, Blender 15, Knuffel 15, Tractor 15 1.0	3. Blender 15, Tractor 14, Knuffel 15, Spiegel 15, Stropdas 15, Groente 15 0.988888	4. Blender 15, Tractor 14, Stropdas 15, Groente 15, Knuffel 13, Spiegel 15 0.966666	1. Knuffel 15, Tractor 15, Groente 14, Stropdas 15, Spiegel 15, Blender 15 0.9888888
Block 3	3. Blender 15, Spiegel 15, Knuffel 15, Stropdas 15, Tractor 14, Groente 15 0.988888	1. Tractor 15, Spiegel 15, Blender 16, Groente 15, Knuffel 15, Stropdas 15 0.988888	4. Stropdas 15, Tractor 15, Knuffel 15, Spiegel 15, Groente 14, Blender 15 0.988888	2. Tractor 15, Stropdas 15, Knuffel 15, Blender 14, Groente 15, Spiegel 16 0.977777
Block 4	4. Stropdas 14 Tractor 14 Spiegel 15 Blender 15 Knuffel 14 Groente 15 0.966666	3. Groente 15 Tractor 15 Spiegel 15 Stropdas 15 Knuffel 15 Blender 15 1.0	2. Stropdas 15 Knuffel 15 Spiegel 15 Tractor 15 Groente 15 Blender 15 1.0	1. Stropdas 15 Blender 15 Tractor 15 Spiegel 15 Knuffel 15 Groente 15 1.0