Proceedings of the Master's Programme Cognitive Neuroscience of the Radboud University

Programme Director: **Rob van Lier**
Journal Logo: **Claudia Lüttke**
Cover Image: **Layout Team**

Contact Information:
Journal CNS
Radboud University
Postbus 9104
6500 HE Nijmegen
The Netherlands
nijmegencns@gmail.com

Table of Contents

From the Editors-in-Chief

Dear reader,

We are very happy to present to you the first volume of the 14th issue of our journal. As always, the papers published in this issue consist entirely of students' master theses written within the Cognitive Neuroscience program. It was a pleasure to read all the amazing submissions we had this time. We believe that the theses printed in this issue are a good representation of the different high-quality research done by the students within this program.

This present issue is quite language-oriented. In total it comprises six theses out of which four come from the Language & Communication track. We are excited that with this distribution a formerly small track of the CNS master program is represented well. The remaining two theses come from the Perception, Action and Control track. We want to congratulate the authors for being published.

Last but not least we want to thank our amazing team. Without you the current issue would not have been possible. You did a fantastic job! Special thanks go out to Ilaria and Lenno for taking up more work than was expected from them. We are very lucky to have team members like you.

We are excited to get started on the next issue 14.2. But for now, enjoy reading.

Nijmegen, February 2019

Antonia & Katharina

Editors-in-Chief

Hello, I’m a cognitive neuroscientist...

Imagine you’re about to board a seven hour flight to Hawaii. Before you take off, you quickly tweet a joke about how your profession is always portrayed differently in movies: “Hello, I’m a professor in a movie, I only reach the main point of my lecture right as class is ending. Then I yell at students about the reading/ homework as they leave.” When you land, 5331 people responded to your tweet, 25903 people retweeted it, and 184479 people liked it (@_roryturnbull, January 8). This is what happened to Rory Turnbull, an assistant professor in Linguistics from the University of Hawaii.

Rory Turnbull’s tweet prompted many researchers to point out how their field of science constantly gets mispresented in the media. This is equally true for the field of cognitive neuroscience: who hasn’t seen an actor playing with their phone near an MRI, or watched a movie where a ‘doctor’ read someone’s mind by looking at a colourful brain scan?

Innovative techniques to research the brain, as well as recent developments in artificial intelligence, have clearly led to an increase of public interest in our field of research, and not just in the movies. Social media, such as Twitter, have also dramatically increased the speed and ease of dissemination of information to the media and the public. As social media allows for information to spread quickly, this often can come at the cost of the reliability of information to facilitate the ease of processing for the reader. This can easily lead to a misrepresentation of often complicated, and highly uncertain research findings.

Our current media landscape perfectly captures our desire to read and share ‘stories’. One of the many skills I had the pleasure to acquire during my time as a student of the Cognitive Neuroscience master’s program was how to translate those complicated findings into such stories.

We are clearly effective in communicating our science to other scientists. Now it is time to use those skills to make our science accessible to all. It should be our pleasure, privilege, and responsibility to communicate our science to the people who surround us in our daily lives. This will not only result in better movies, but will also open up meaningful debates with all layers of society. Tell your story. The exciting findings of the papers in this edition of the Nijmegen CNS journal deserve nothing less.

Linda Drijvers

Postdoctoral researcher at the CoSI lab, Donders Institute
CNS Alumna (2014)

About the cover

What we are looking at on the cover of the current issue of the Cognitive Neuroscience Journal is one of the most popular icons in the world, a drawing created by Leonardo da Vinci around 1490 AD. This design is called the 'Vitruvian Man', known also as 'The proportions of the human body according to Vitruvius' (Italian: Le proporzioni del corpo umano secondo Vitruvio). The original drawing is accompanied by notes that are based on the building guide 'De Architectura' which was written by Vitruvius, a famed Roman architect. Vitruvius, in his book, described the correlations of the ideal human body proportions with geometry as being the principal source of proportion among the classical orders of architecture. Some of his ideal determinations were that four fingers equal one palm, four palms equal one foot, six palms make one cubit, four cubits equal a man's height and that twenty-four palms equal one man. Leonardo's drawing depicts a nude male in two superimpossed positions, displaying his four arms and legs apart, inscribed in a cricle and a square, allowing him to strike 16 poses stimultaneously. The proportional relationship of the parts reflects universal design and a 'medical' equilibrium of elements that ensure a stable body structure. He, famously, wrote: "Man has been called by the ancients a lesser world, and indeed the name is well applied; because, as man is composed of earth, water, air, and fire…this body of the earth is similar." He compared the human skeleton with rocks ("supports of the earth") and the expansion of the lungs to the flow of the oceans. This picture represents a cornerstone of Leonardo da Vinci's attempts to relate man to nature. Leonardo's drawing was traditionally named in honor of the architect.

Leonardo da Vinci's 'Vitruvian Man' as a cover picture for this issue of the Cognitive Neuroscience Journal represent his beliefs that the workings of the human body are an analogy to the workings of the universe. Delving deeper, the relationship between the universe and the human brain has already been suggested and as neuroscientists and astrophysisists globally support, the complexities and structures of the brain and the cosmic web are actually similar. Crortical gray matter contains roughly 6 billion neurons and nearly 9 billion non-neuronal cells and the cerebellum has about 69 billion neurons and about 16 billion non-neuronal cells. What is interesting, though, is that the total number of neurons in the human brain falls in the ballpark of the number of galaxies in the observable universe. Even when looking at pictures with the naked eye, we can immediately grasp some smiliarities between images of the cosmic web and the brain. "It is truly a remarkable fact that the cosmic web is more similar to the human brain than it is to the interior of a galaxy; or that the neuronal network is more similar to the cosmic web than it is to the interior of a neuronal body. Despite extraordinary differences in substrate, physical mechanisms, and size, the human neuronal network and the cosmic web of galaxies, when considered with the tools of information theory, are strikingly similar" (Vazza & Feletti, 2017).

Vasilis Kougioumzoglou

P300 Versus MMN: The Clinical Potential of ERP Components to Assess Auditory Discrimination Abilities in Cochlear Implant Users

Rosanne Abrahamse[1]
Supervisors: Vitória Piai[1,4], Margreet Langereis[2,3], Anneke Vermeulen[2,3]

[1]Radboud University Nijmegen, Donders Centre for Cognition, Nijmegen, the Netherlands, [2]Radboud University Medical Centre, Donders Institute, dept. Hearing and Implants, Nijmegen, the Netherlands, [3]Radboud University Medical Centre, dept. of Otorhinolaryngology, Nijmegen, the Netherlands, [4]Radboud University Medical Centre, dept. of Medical Psychology, Nijmegen, the Netherlands

The current study was designed as a starting point in developing an electrophysiological marker of speech perception abilities in cochlear implant users. Two event-related potentials (ERPs), the Mismatch Negativity response (MMN) and the P300 response were compared in their ability to assess auditory discrimination abilities in prelingually deaf adolescent cochlear implant users (n = 8) and normal-hearing controls (n = 14). The ERPs were compared in terms of their robustness on an individual level, with an equally limited amount of stimuli in each condition. A frequency contrast (500 Hz vs. 1000 Hz tone) and a consonant contrast (/ba/ vs. /da/ syllable) were used as stimuli. The P300 response, as opposed to the MMN response, was elicited in all individuals in both contrast conditions and, therefore, was deemed the most robust ERP of the two. Further analyses on differences in amplitude and latency of the P300 response as a function of group and/or contrast condition yielded significantly longer latencies for the consonant as opposed to the frequency contrast condition. It is suggested that the absence of group differences can be ascribed to a ceiling effect in the auditory discrimination abilities of the cochlear implant group. The relations between amplitude of the P300 in response to the consonant contrast, behavioural speech perception scores and duration of deafness indicate that the amplitude of the P300 has the potential to objectively inform us about speech perception abilities of cochlear implant users, as well as about the development of the auditory cortex after implantation. Future research can focus on measuring the P300 response in younger cochlear implant users, as well as measuring the P300 response with more complex input.

Keywords: P300, MMN, cochlear implantation, discrimination, speech perception

Corresponding author: Rosanne Abrahamse, **E-mail:** abrahamseros@gmail.com

Cochlear implants (CIs) are devices used to provide profoundly deaf children and adults with better hearing function. These devices surpass the impaired cochlea and restore the hearing pathway towards the auditory nerve. Although the use of these devices considerably improves speech perception, the speech perception abilities of CI users remain limited compared to the speech perception abilities of normal-hearing children and adults (American Speech-Language-Hearing Association, ASHA, 2004). This is a challenge in itself, but the implant outcomes also vary greatly per individual. While some implant users seem to perform almost equivalently to normal-hearing peers on speech perception measures, others perform considerably below average (2004; Beynon, Snik & van den Broek, 2002; Pisoni & Geers, 2000).

Several factors have been shown to play a role in the variable speech perception outcomes of CI users, such as age at implantation (Pisoni & Geers, 2000; Ruffin, Kronenberger, Colson, Henning & Pisoni, 2013), communication mode (Geers, 2002; Pisoni & Geers, 2000; Ruffin et al., 2013) and IQ (Geers, 2002). Despite these examples, a large proportion of variability remains unexplained. Clinically, the need to detect the variability early in development is high. The earlier clinicians know that children are t risk for non-optimal speech perception outcomes, the sooner appropriate interventions can be applied.

In addition to speech perception problems, research shows non-normal and variable performance in linguistic domains among CI users (de Hoog, Langereis, Weerdenburg, Knoors & Verhoeven, 2016a; 2016b; Pisoni & Geers, 2000; Schorr, Roth & Fox, 2008; Svirsky, Robbins, Kirk, Pisoni & Miyamoto, 2000). Interestingly, these linguistic problems are not always directly related to speech perception problems (de Hoog et al., 2016a). If the origin of speech perception problems in CI users can be more properly located, it may be possible to either relate these to, or differentiate these from, the existing variability in linguistic performance. This way, appropriate interventions can be designed for each type of problem.

Differences in the central auditory processing function of the CI users provide a possible explanation for the observed variability in speech perception outcomes (Groenen, Beynon, Snik, Broek, 2001; de Hoog et al., 2016a; Kraus et al., 1993; Näätänen, Paavilainen, Rinne & Alho, 2007; Pisoni & Geers., 2000). These differences can be due to having experienced a period of auditory deprivation, or, when deafness occurs at a younger age or congenitally, to an overall immature auditory system. The central auditory processing function of the brain can be measured using electro-encephalography (EEG). This may offer a means to clinically detect the individual variation in speech perception outcomes after implantation. It provides insight into the development of the auditory cortex and it has been linked to subjective speech perception outcomes (Groenen et al., 2001; Kelly, Purdy & Thorne, 2005). Furthermore, it provides the objectiveness that is needed to assess the perception abilities of very young children.

The current study will firstly assess the clinical potential of two electrophysiological approaches to measure the central auditory processing function of CI users. To be suitable for the clinic, the measures should be robust on an individual level, their acquisition should have an appropriately short duration, and the measure should have task requirements that fit the attention span of young children. The two EEG-approaches will be tested in an adolescent population (a normal-hearing group, and a group of prelingually deaf CI users). Starting out with longer EEG-recordings (as a result of testing two methods instead of one) in an older, more flexible population, provides the opportunity to choose the best method for future research. This way, young children can in the future be exposed to less demanding tasks. As a secondary aim, the relation between the central processing function of the CI users, measured using EEG, and their speech perception scores as measured in the clinic, will be explored. This will be the first step in establishing a link between the objective measurement and the subjective scores.

Marking speech perception abilities in cochlear implant users: two approaches

The electrophysiological approaches we decided to focus on are the P300, or P3b component (Polich, 1987), and the mismatch negativity (MMN) component (Näätänen et al., 2007). These are both late event-related potentials (ERPs) that can be measured using EEG. Late ERPs are proposed to reflect the central auditory processing function of the brain. The MMN and P300 components, for example, appear when the brain performs auditory discrimination of two stimuli (Johnson, 2009). Auditory processing can be elicited by means of an auditory oddball paradigm, in which a participant hears a stream of frequent standard stimuli (for example 80% of the time) that are randomly alternated with infrequent deviant stimuli (for example 20% of the time). If auditory discrimination is performed, it is reflected in the difference between the averaged brain responses to the standard stimuli and the averaged brain responses to the deviant stimuli.

The MMN response requires no attention from the participant and is said to reflect how accurately the auditory sensory memory substrate of the brain can perform lower-level discrimination, based on perceptual stimuli characteristics (Näätänen, 2001). Auditory MMN response is a negative deflection in amplitude around 150-250 ms, which is observed over fronto-central regions of the brain. The P300 response is said to reflect a more conscious, higher-level cognitive process. Each new stimulus is evaluated against a model of the earlier one held in working memory. If a change in stimulus is detected, the model is updated. Besides perceptual discrimination, attention to the stimuli is required for the updating of the model (Polich, 2012). The auditory P300 response is characterized by a positive deflection in amplitude around 300 ms. It is often observed over centro-parietal areas of the brain (Johnson, 2009).

Speech perception problems may be specific to certain speech contrasts, and abilities may differ for speech as opposed to non-speech stimuli. Both the MMN and the P300 components vary in amplitude and latency with respect to the input they are given. This effect has been found for intensity contrasts (e.g., 80 dB stimulus vs. 90 dB stimulus), frequency contrasts (e.g., 500 Hz pure tone vs 1000 Hz pure tone) and speech-sound contrasts (e.g., consonant contrasts /ba/ vs. /da/ or vowel contrasts /i/ vs. /a/). More difficult contrasts lead to longer latencies and altered amplitudes (MMN: see Näätänen et al., 2007 for a review; P300: Polich, 1987; see Polich, 2004 for a review). Furthermore, differences are also evident across conditions, with more complex stimuli (speech-sound contrasts) yielding longer latencies and altered amplitudes as opposed to simple stimuli (frequency contrasts; MMN: Näätänen et al., 2007; P300: Polich, 2004). The MMN and the P300 components thus give the opportunity to investigate not only auditory processing in general, but to also distinguish between responses to complex as opposed to simple stimuli. With respect to future clinical marker abilities, varying input stimuli may lead to finding out which ones are more (or less) sensitive predictors for speech perception.

Earlier findings on the P300 and MMN components in cochlear implant users

In terms of task-requirements, the MMN is optimal to measure auditory processing in critical populations like infants and severely disabled people, because it does not require attention. The P300 requires attention and is therefore less attractive for this purpose. However, the task requirements for measuring the P300 are sufficiently low (participants are instructed to count the deviant in their heads or press a button when hearing the deviant stimulus), to be suitable for children from ages 3-4 onwards (Johnson, 2009).

Both components have been shown to be measurable on a group level and on an individual level in both (pre- and postlingually deaf) children and adults with cochlear implants (MMN: Kileny, Boerst & Zwolan, 1997; Kraus et al., 1993; Ponton et al., 2000; Singh, Liasis, Rajput, To & Luxon, 2004; Turgeon, Lazzouni, Lepore & Ellemberg, 2014; Watson, Titterington, Henry & Toner, 2007; P300: Beynon et al., 2002; Beynon, Snik, Stegeman & van den Broek, 2005; Groenen et al., 2001; Jordan et al., 1997; Kileny, 1991; Micco et al., 1995). In some studies, differences in the CI user ERPs compared to the normal-hearing control ERPs appeared, such as a prolonged latency (MMN: Turgeon et al., 2014; P300: Beynon et al., 2005) or a different amplitude (MMN: Ponton et al., 2000; Watson et al., 2007; P300: Beynon et al., 2005).

Furthermore, both components are relatively sensitive in distinguishing between well-performing and poor-performing users, on a group level and on an individual level. The ERPs of well-performing CI users are similar to that of normal-hearing controls, while the ERPs of poor-performing CI users (categorised as such due to low behavioural speech-perception scores or a below average subjective discrimination of the stimuli) are often found to be absent or different (MMN: Kraus et al., 1993; Singh et al., 2004; Turgeon et al., 2014; P300: Beynon et al., 2002; Groenen et al., 2001; Jordan et al., 1997; Kileny, 1991; Micco et al., 1995).

In terms of contrast conditions, there is a trend for more different responses as conditions become more complex. For the MMN component, longer latencies were found for increasing complexity of conditions. That is, the speech-sound contrast condition yielded longer latencies than the frequency or loudness conditions (Kileny et al., 1997). For the P300 component, well-performing CI users showed a P300 only when hearing a frequency and a vowel contrast (Beynon et al., 2005; Groenen et al., 2001). When hearing a consonant contrast, the P300 was absent in a significant number of participants (Beynon et al., 2005; Groenen et al., 2001). Furthermore, a poor-performing group of CI users in another study again showed only a P300 for the frequency contrast, albeit with a prolonged latency. The well-performing group in this study showed a P300 for the consonant contrast and performed therefore similar to the normal-hearing control group. Vowel contrasts were not addressed in this study (Beynon et al., 2002). On the basis of this research it is expected that the more complex consonant contrast condition may show more differences in robustness among CI users.

Interestingly, the simpler conditions seem to be best for predicting behavioural speech perception, although research is scarce and results are inconsistent. For the MMN, in studies with frequency and vowel contrasts, duration of the component is found to correlate with perception scores (Kelly et al., 2005; Kileny et al., 1997), while in studies with consonant contrasts, amplitude is found to correlate with perception scores (Turgeon et al., 2014). For the P300, a relation between amplitude and perception scores was found, again only in the frequency and the vowel conditions, not in the consonant condition (Groenen et al., 2001).

Direct comparisons of the two approaches are scarce. One study combined both approaches in three CI-participants and three participants without hearing problems. The ERPs were identified in both the inattentive and the attentive paradigm. Although the study does not address the clinical potential of the components in particular, it stresses that both play an important role in the comprehension of the central auditory processing function (Obuchi, Harashima & Shiroma., 2012).

The next step in finding a suitable clinical marker

All of these results are quite promising in terms of the clinical potential of the MMN and P300 response to indicate speech perception difficulties. There are, however, a few shortcomings of these studies. Firstly, most studies mentioned above are outdated. Several aspects of cochlear implantation have improved during the past 15 years. Children now are more often, as well as earlier in life, eligible for an implant. The adolescents that were chosen to participate in this study were some of the first from this 'new generation' of less strict implantation eligibility. There is a need for replicating older findings under these contemporary circumstances. Secondly, a number of studies had small sample sizes (Beynon et al., 2002; Jordan et al., 1997; Kileny, 1991; Obuchi et al., 2012), or did not compare their results to a group of normal-hearing controls (Kileny et al., 1997). The latter can hamper interpretation of the results. In the current study, sample size remains a problem. Still, eight CI users were tested, whereas the studies mentioned above drew their conclusions on only half the amount of participants.

Thirdly, analysis techniques that were used to determine the presence of the ERP in individual waveforms were not consistent over studies. The early studies all chose to manually determine the amplitude and latency of the response, making use of the difference between the response to the standard stimuli and the response to the deviant stimuli (the difference waveform). Statistical analysis was consequently performed over the manually determined amplitude and latency values in some, but not all studies (e.g., Kileny et al., 1997 did not perform statistical analysis). This manual method is subjective and prone to bias (e.g., Kilner, 2013). Only a few attempts have been made to make the analysis more objective (Ponton et al., 2000, and for healthy subject data see Bishop & Hardiman, 2010). We applied a statistical procedure (non-parametric cluster-based permutation tests, see Maris & Oosterveld, 2007) to our EEG-data to identify MMN or P300 responses. By doing this, we aimed to make ERP analyses more objective and reliable. In addition, not many studies give notice of the existence of CI artefacts. Some do and approach the problem by rejecting artefact above 100 mV (Singh et al., 2004) and 50 mV (Kelly et al., 2005). Another used a semi-automatic procedure to attenuate them (Turgeon et al., 2014). In our study, we paid specific attention to developing an efficient procedure for attenuating artefacts.

Lastly, almost no studies have compared the MMN and the P300 responses directly in one design and thus, in terms of clinical utility not much has been concluded. Comparisons are needed in order to evaluate which approach has potential to develop into a clinical marker for speech perception abilities in CI users. An advantage of the MMN compared to the P300 is that it can be measured in younger populations. However, an advantage of the P300 compared to the MMN is that the P300 response has been detected with only 12 minutes of recording (Beynon et al., 2002; 2005; Groenen et al., 2001; Kileny, 1991; Micco et al., 1995). The shorter a measurement takes to yield robust results, the more advantageous this is for the clinic. The MMN response has been detected using 25 minute EEG-recordings (Kraus et al., 1993), but there are also experiments which lasted 35 to 40 minutes (Singh et al., 2004; Turgeon et al., 2014). Interestingly, in one study (Obuchi et al., 2012) the P300 as well as the MMN response were elicited with only four minute recordings. Here, the MMN paradigm still yielded valuable results when restrictions were imposed on duration of the measurements. If this result can be replicated, the MMN might be perfectly discernible using only a limited amount of stimuli, and perhaps a better candidate for becoming a clinical marker than the P300.

Aims and objectives

This study focused on comparing the robustness of the MMN and the P300 response to measure auditory discrimination, with an equal limited amount of stimuli data (10 min EEG-recordings), on an individual level. Robustness was defined on the basis of how many individuals showed a statistically significant amplitude difference in their neurophysiological responses to the standard as opposed to the deviant stimuli. Results were obtained for two contrast conditions: a frequency contrast (500 vs. 1000 Hz tones) and a consonant contrast (/ba/ vs. /da/ syllables). EEG was measured in 14 normal-hearing participants, and in 8 prelingually deaf young-adult CI users. In the conditions (ERP component x contrast conditions) where the individual responses were most robust, the ERPs of a matched sample (n = 8) of normal-hearing participants were compared to the CI user ERPs. For those ERPs, group differences and within-group variation was assessed using the mean amplitude and latency measure. Furthermore, the effect of contrast conditions (frequency vs. consonant contrast) was addressed for both groups to be able to evaluate differences in simple (e.g., frequency contrast) vs. complex processing (e.g., consonant contrast). Lastly, behavioural speech perception as measured in the clinic and duration of deafness were related to P300 amplitude to explore whether the P300 would be a suitable marker for speech perception abilities, and whether this suitability differs for different input contrasts.

Methods

Participants

Eight Dutch adolescent prelingually deaf CI users (M_{age} = 19.9, ranging from 16-25; 6 males) were recruited for the ERP measurements through the otolaryngology department of the Radboudumc in Nijmegen, the Netherlands. All of the adolescents had profound bilateral hearing loss. Exclusion criteria consisted of having an IQ < 85, having a developmental or neurological disorder, or having had any serious head-trauma in the past. Table 1 describes the characteristics of the CI users. All participants used the same implant processor (Cochlear™ Nucleus®), and none of them used any additional hearing aids. For the participants with bilateral implants, EEG recording was done using only one implant. These users were allowed to choose on which implant (left or right) they wanted to be tested. Participation was on a voluntary basis and the participants received a monetary reward for their participation of 20 euros in vouchers.

A control group of 14 Dutch normal-hearing participants was also tested (M_{age} = 21.4, ranging from 18-25; 6 males). All of them had no history of hearing problems or speech/language problems. Furthermore, they had no psychiatric or neurological disorders. We restricted the educational levels of the normal-hearing participants to level 6 (out of 7), according to the Dutch neuropsychological educational level coding (Hendriks, Kessels, Gorissen, Schmand & Duits, 2014). This was done to achieve a more proper matching of the two participant groups. We recruited 11 participants with an educational level of 6, three participants with an education level of 5 and one participant with an educational level of 4. The control participants were recruited via flyers and experiment databases. They received a monetary reward for their participation of 20 euro in vouchers. This research was approved by the ethical review board of the Radboud University Medical Centre.

There was no significant difference between the mean age of the normal-hearing participants and the mean age of the CI user group (W = 73.5, p = .24). This was tested using a Wilcoxon rank-sum test.

Materials

Two conditions, an inattentive and an attentive condition, were designed to elicit the MMN component and the P300 component separately. In both conditions, the auditory oddball paradigm was used. Two stimuli types were used: a frequency contrast (a 500 vs. 1000 Hz tone) and a consonant contrast (syllables /ba/ vs. /da/).

Table 1.
Demographic and clinical information of the 8 CI users.
Note. F: female, M: male, Educ. Level: education level, DD: duration of deafness before implantation, Bi: bilateral, Uni: unilateral, CI: cochlear implant, MOC: mode of communication.

ID	Sex	Age (yrs)	Educ. Level	Etiology	Age at implantation (yrs)	DD (yrs)	Bi/ Uni	CI use per day (hrs)	Main MOC
1	M	24	5	Meningitis	3	2.08	Uni	14	Speech
2	M	25	6	Meningitis	3.6	2.17	Uni	16	Speech
3	M	23	5	Congenital	2.7	2.7	Uni	14	Speech
4	M	18	4	Meningitis	1.6 and 1.6	0.08	Bi	15	Speech
5	M	16	4	Prematurity	2.1 and 5	2.08	Bi	12	Speech
6	F	16	4	Unknown	3	3	Uni	16	Half/ half
7	F	18	6	Meningitis	1	0.5	Uni	16	Speech
8	M	19	6	Congenital (LADD syndrome)	5.8 and 19	5.8	Bi	14	Speech

was used as the standard stimulus, the 1000 Hz tone was used as the deviant stimulus. For the consonant contrast we used the syllable /ba/ as the standard stimulus and /da/ as the deviant stimulus. The duration of the stimuli was 170 ms. These synthesized stimuli were the same stimuli that Beynon et al., (2005) used for their ERP experiment. They, in turn, adapted these stimuli from the ones used in Groenen et al., (2001). For a detailed description of these stimuli please consult the articles mentioned above. The order of the four conditions was randomized with breaks in between. Half of the participants started with the frequency contrast (first: inattentive-frequency, second: attentive-frequency), and did the consonant contrast after that (third: inattentive-consonant, fourth: attentive-consonant). The other half of the participants did this the other way around (first: inattentive-consonant, second: attentive-consonant, third: inattentive-frequency, fourth: attentive-frequency).

Auditory stimuli.

For the frequency contrast, a 500 Hz pure tone burst and a 1000 Hz pure tone burst of 120 ms each were generated with Praat (Boersma & Weenink, 2018) (settings: stereo channels, 20 ms linear rise and fall time, 80 ms plateau time, sampling frequency of 44100 Hz and an amplitude of 0.2). The 500 Hz tone was used as the standard stimulus, the 1000 Hz tone was used as the deviant stimulus. For the consonant contrast we used the syllable /ba/ as the with breaks

Behavioural assessment.

Before each set of ERP conditions, two short reaction-time tasks were performed by the participants. There was a consonant contrast version and a frequency contrast version. This was to see whether the participants could subjectively distinguish between the contrasts. The same stimuli as the stimuli used in the ERP conditions were randomly presented 20 times (50% standard, 50% deviant). The participants were asked to press the left button when they heard the standard stimulus and the right button when they heard the deviant stimulus. When a button was pressed, the next stimulus was automatically presented. If no button was pressed within 1500 ms, the next stimulus appeared. For both versions there was a familiarization phase of five trials in which the stimuli were presented and it was shown on the screen which button to press for which stimulus. The reaction times of the participants were analysed.

When the first reaction time task was done, all participants were asked to judge the loudness of the sound on a 5-point scale, with 1 = too soft, 2 = a bit soft, 3 = good, 4 = a bit loud, 5 = too loud. Subsequently, the CI participants were given the opportunity to adjust their speech processor if they wished, to avoid any discomfort while listening to the stimuli. From a total of twenty-two participants (8 CI, 14 NH), fifteen participants rated the sound as 'good' (7 CI, 8 NH), four participants rated the sound as 'a bit soft' (1 CI, 1 NH), and three participants rated the sound as 'a bit loud' (0 CI, 3 NH). None of the CI users felt the need to adjust their processors.

Procedure

The ERP measurements were performed in a sound-proof EEG-lab. Subjects were seated in a comfortable chair. Sound was presented via speakers that were approximately 2.5m away from the participant. The sound presentation at ear-level was kept at 65 dB at all times, as measured by a measuring amplifier (Bruel & Kjaer Type 2610) and a microphone (Bruel & Kjaer Type 4192).

Stimuli were presented with Presentation® software (Version 18.0, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com). For both the inattentive and attentive measurements, the standard stimuli occurred at a probability rate of 85% and the deviant stimuli occurred at a probability rate of 15%. In each of the four conditions, there were two blocks of 220 stimuli, resulting in a total of 440 stimuli per condition. In each block, first 20 standard stimuli were presented, followed by 30 deviant stimuli that were randomly embedded in 170 standard stimuli. It was made sure that between two deviant stimuli at least three standard stimuli were presented. We controlled for a list-specific effect by generating multiple stimulus lists and assigning these at random to the participants. For each of the four conditions three random stimulus lists of 440 stimuli were generated, resulting in 12 lists in total. In the control group, one-third of the participants got four A-lists (n = 5), one-third got four B-lists (n = 5) and one-third got four C-lists (n = 4). Because the CI user group turned out to be smaller than expected, five participants in the CI user group got four A-lists, and three participants in the CI user group got four B-lists. The lists were made with the program Mix (van Casteren & Davis, 2006), and adjusted by hand to remove any presentation patterns that arose even after randomization.

For the inattentive measurements, the stimuli were presented with an inter-stimulus interval (ISI) of 1000 ms with 10% jitter. In both contrast conditions (frequency and consonant), the participants were asked to watch two different 10-minute silent snippets of a film. The snippets were selected for having emotionally neutral contents. They were instructed thereafter not to pay attention to the sounds that would be presented during the movie. The two inattentive measurements (frequency and consonant) lasted eight minutes each. The video-snippet was automatically quit when the inattentive measurement had fully run. There was no break between the two blocks of stimuli (note: there were breaks between all four experiments, but not within the inattentive experiment(s), so not between blocks).

For the attentive measurements, the stimuli were presented with an ISI of 1500 ms with 10% jitter. This differed from the inattentive measurements because the P300 is a component that spreads out over a longer time-window. We did not want to risk any overlap in neurophysiological responses to the stimuli. In both contrast conditions (frequency and consonant), the participants were instructed to, for each block of 220 stimuli, count in their heads the number of deviant stimuli that occurred. At the end of each block they were asked to type in how many deviant stimuli they had heard (30 deviants in each block). Between the two blocks there was a break to enhance the participants' attention and dismiss fatigue. The participants were told that they could close their eyes during the measurement if they had difficulty not-blinking, but that they had to be careful not to fall asleep during the measurement. As a result of this, some people closed their eyes during the measurements, but most kept their eyes open. Each of the two measurements (frequency and consonant) lasted 11 minutes.

EEG data acquisition.

The EEG was continuously recorded from 24 active electrodes embedded in a 10-20 international system electrode cap (Acticap 32Ch standard-2). The reference electrode was placed at Cz for online referencing and the EEG signal was re-referenced offline using the common average method. The ground electrode was placed at AFz. Due to too much space between the EEG-cap and the scalp, where the electrodes around the processor were supposed to be placed, and due to possible CI artefacts, we did not fill electrode places around the cochlear implant(s) and its contralateral side (CP6; T8; P8; TP10; TP9; T7; CP5; P7). This configuration was kept for both control participants and CI users to enhance consistency. Electrooculography (EOG) was recorded from two horizontal electrodes, placed at the left and right temples, and two vertical electrodes, placed above and below the left eye. Electrode impedance was kept below 20 KOhm.

The EEG signal was online filtered with the low cutoff of .016 Hz and the high cutoff of 125 Hz. The EEG was recorded at a sampling rate of 500 Hz.

Analysis

We analysed the EEG-data using the MATLAB-toolbox Fieldtrip (Oostenveld, Fries, Maris & Schoffelen, 2011). Data of the inattentive-frequency condition of one control participant (pp9) were missing due to an experimenter error. For the inattentive conditions, the data were cut into segments with a time-frame of -0.3 to 0.7 seconds before and after onset of the stimulus, respectively. For the attentive conditions this time-frame was -0.3 to 1 seconds. Vertical and horizontal EOG were re-referenced following a bipolar montage. The data were de-trended.

Data cleaning and CI artefact removal. Data were filtered with a low-pass filter of 80 Hz. For both the removal of eye-artefacts and the removal of CI-artefacts, we did an independent component analysis (ICA; Jung et al., 2000). We performed the ICA over all four conditions together. We visually inspected the component topographies, the component time-courses and the corresponding EEG segments. Eye-blink components were rejected.

We developed a procedure for the removal of possible cochlear implant EEG artefacts. The implant artefact is independent of brain processes or task design. It is a reaction of the implant electrode array to the presentation of a sound, that is detected by the EEG. The artefacts are described in the literature as a systematically occurring increased or decreased amplitude peak. (Gilley et al., 2006; Turgeon et al., 2014; Viola et al., 2012). The artefacts do not occur in each CI user, but only in some of them. To attenuate these CI artefacts, we performed a time-locked analysis over the ICA components. For any process to be time-locked to a stimulus, this means that in each trial during the EEG (regardless of paradigm, input stimulus or standard/deviant classification) there should be a deflection in amplitude as a consequence of a cognitive, lower-level, or external process. This deflection should occur at the exact same time in all 1600 trials. If an ICA analysis is performed and the mean of these trials per component is plotted, it shows which components from the ICA are time-locked to the stimulus. These components can then be removed. Removing activity related to the biological processes of the P300 or MMN is unlikely. These neural processes are known to occur later than at stimulus presentation. It is also unlikely that what comes out of the time-locked analysis is non-CI artefact noise. Non-artefact noise, such as eyeblinks, will cancel out because they occur at different time-points throughout the trials. Auditory presentation is the only factor that occurs roughly around the same time in all conditions. It is possible that biological processes related to this auditory presentation (such as the N1 or P2) are filtered out by the CI-artefact analysis. However, the input stimuli that were used were different. The biological processes that occur in reaction to the syllables may be different in time than the biological processes that occur in reaction to the tones. Even if it should be the case that we eliminate activity occurring from these processes, they are not the processes we focus on in this article. We did a time-locked analysis over ICA components for all participants with a CI. If, per participant, there were components that had time-locked amplitude deflections occurring all at the exact same time after stimulus presentation, they were removed from the data. Furthermore, we checked whether the spatial morphologies of these identified components matched morphologies from earlier papers (Gilley et al. 2006; Viola et al. 2012). The components could not occur later than 150 ms after stimulus presentation, otherwise they were not removed. In the end, we deleted artefact component(s) for CI users 1, 5, 6 and 7. Although this procedure was semi-automated, it still remains a subjective task to determine which components should be eliminated.

After the ICA, the data (per participant) were split into the four individual conditions. For each condition we used a semi-automatic artefact rejection approach (ft_rejectvisual in Fieldtrip) to identify and throw out any trials that were outliers. This approach shows the preprocessed data in all channels or trials and allows the user to select noisy data (trial and/or channel) and delete it. Furthermore, it is possible to compute the variance in each channel and/or trial and delete outliers based on this. For each participant on average 17 (for the CI users) or 18 (for the controls) out of 440 trials were deleted per condition. Channels that were noisy were noted down (not deleted) for each experiment. Later on, this information was taken in consideration when selecting the channels to perform statistics on.

ERP calculations and statistics

Group-level analysis.

The artefact-free data were used to compute group ERPs. The ERPs were computed by averaging waveforms across trials per stimulus condition (standard vs. deviant), per group (normal-hearing and CI user), per task by contrast condition (attentive-frequency, attentive-consonant, inattentive-frequency, inattentive-consonant). The data were filtered with a low-pass filter of 50 Hz and down-sampled to 512 Hz. We used cluster-based permutation tests (Maris & Oostenveld, 2007) to statistically evaluate the presence of the ERPs in all four conditions per group. This was done using a within-subjects design in which the grand average response to the standard trials was compared to the grand average response to the deviant trials. Statistics were performed as follows: first a dependent samples t-test was calculated for every electrode/ time-point. The comparison was based on all time-points from 150 ms to 800 ms post-stimulus onset for the attentive task-condition and 50 ms to 350 ms for the inattentive task condition. Statistical tests were based on channels 'CP1', 'CP2', 'P3', 'P4', 'Pz', 'C3', 'C4' for the attentive task-conditions and 'Fz', FCz', 'F3', 'F4', 'FC1' and 'FC2' for the inattentive task conditions. Decisions for these time-points and channels were based on previous literature describing the location of effects (Johnson, 2009) and on the exclusion of channels that were deemed excessively noisy during data acquisition. The electrodes/time points were clustered based on spatial and temporal adjacency at an alpha level of 0.05. Channels had on average 3.3 neighbours. Cluster-level statistics were calculated by taking the sum of the t-values within every cluster. The largest cluster-level statistic was taken for evaluation under a permutation distribution. This distribution under the null hypothesis of exchangeability between trial conditions was constructed by randomly re-assigning the standard trial and the deviant trial labels to the original individual ERP waveforms, followed by the construction of spatiotemporal clusters, in the same way as for the observed data. 1000 permutations were used to make the permutation distribution. The p-value was determined as the proportion of random permutations that yielded a more extreme cluster statistic than the cluster in the original data. The alpha-level was set to 0.05 (two-sided test). If the p-value was smaller than alpha, the difference between the standard and the deviant trials was deemed significant.

Individual-level analysis.

We also performed individual ERP analyses per stimulus-condition (standard vs. deviant) per task by contrast condition (attentive-frequency, attentive-consonant, inattentive-frequency and inattentive-consonant). To test for ERP presence, we used the same cluster-based permutation procedure as described for the group analysis. That is, we tested the difference between the standard and the deviant waveforms per individual. However, a between-trials design was used. This means that an independent samples t-test was performed.

Amplitude and latency analyses.

The attentive task-condition yielded robust results for all participants for both contrast conditions, while the inattentive task condition did not (see Results below). Therefore, we performed the amplitude and latency analyses only on the attentive task-condition data. To avoid the different sample sizes of the two groups, we took a sub-sample of normal-hearing participants (n = 8) to match the sample of CI user participants (n = 8). This sub-sample firstly was matched to the CI user group on the order in which the contrast conditions appeared during data collection. Then, each CI user was matched to a control on at least one of the following three criteria: education level, age, or sex. This was done because it was not possible to match the two groups on one criterion only and because all criteria were deemed equally important to match on. This semi-objective way of matching shielded the matching from a selection bias. In the end, the sub-sample consisted of normal-hearing controls 1, 3, 4, 5, 6, 8, 11 and 12. Their significant clusters and time-windows from the individual ERP statistics are reported in Appendix A. We performed the amplitude and latency analysis over one electrode: 'Pz'. This decision was based on earlier studies that also performed their analyses over one or two electrodes. (Beynon et al. 2002, 2005; Groenen et al., 2001; Kileny, 1991; Micco et al., 1995; Obuchi et al., 2012). This way, we would be able to more accurately relate our findings back to earlier ones. We did not pursue this decision for the ERP presence analysis (see above), because the cluster-based permutation approach is more conservative when using more channels.

Amplitude analysis.

We assessed amplitude differences in the attentive task condition difference waveforms between groups, per contrast condition. Because analysing amplitude differences in ERP designs has been shown to be prone to bias and may lead to incorrect conclusions (Luck, 2012; Woodman, 2010), we explored the outcomes of two different analyses. We calculated the mean amplitude (MA) and the peak amplitude (PA). As the MA, we calculated the mean voltage over a pre-specified time-window of the difference wave (standard deviant condition). We chose 220 ms to 705 ms for the frequency contrast condition and 315 ms to 760 ms for the consonant contrast condition as time-windows. These time-points were the minimum and maximum time-points of the time-windows in which the individual ERPs were significant, as present in the cluster-based statistics. As the PA, we took the peak amplitude of the difference waveform (standard-deviant condition). For the consonant contrast condition, we chose the same time-windows as in the MA approach. For the frequency contrast condition, however, we chose a time-window of 300 ms to 705 ms. This was done because CI user 3 showed two peaks: one around 250 ms and one around 400 ms. As the first peak was very early in comparison to all other CI users, it was suspected to be a CI artefact. We therefore wanted to make sure we got the PA of the second peak. The peaks of all other participants did not start until 300 ms so there was no danger of missing other peaks.

We performed statistics over the means and standard deviations of both the MA and PA outcome values. Because of a relatively small *n*, we used a non-parametric Wilcoxon rank-sum test to test for significant mean differences between groups and a non-parametric Fligner-Killeen test to test for homogeneity of variances between-groups. more, we used the Wilcoxon signed-rank test to test for significant mean differences between contrast conditions. Lastly, we used the Wilcoxon rank-sum test to explore an interaction effect between contrast condition and group.

By comparing the results of these two analyses, this paper may contribute to the discussion on the reliability of calculating the peak amplitude versus the mean amplitude of the difference wave. On the one hand, it is debated whether the peak of an ERP component has any meaningful value in itself. Also, the peak of an ERP is very sensitive to high-frequency noise (Luck, 2012; Woodman, 2010). On the other hand, the mean amplitude is not free of disadvantage either. A lower-to-noise level may cause the ERP waveform of an individual to be fluctuating at several places in the waveform (see for example the alpha noise in Figure 5, CI user 2), causing the mean amplitude (calculated over the entire window) to decrease.

Latency analysis.

We also assessed latency differences between groups and contrast conditions in the attentive task condition. Measuring ERP latency differences on a single-subject level is deemed a cautious undertaking. Firstly, the relationship between the underlying component and the local shape of a component is not obvious (Luck, 2005). Secondly, the signal-to-noise level is low due to averaging over a small amount of trials. Therefore, we measured latency differences with the jackknife-based approach (Kiesel et al., 2008). In this approach, latencies are scored for each of *n* grand average waveforms in a group, with each grand average waveform computed from a subsample of *n-1* individual waveforms. Each participant in a group is omitted from the analysis once, and each latency score is calculated not from a single-subject waveform, but from a grand average.

Using the peak latency of the component as a scoring method has been deemed misleading and arbitrary in the ERP literature (Luck, 2005; Woodman, 2010). Therefore, we again relied on Kiesel et al. (2008) for our scoring method. The scoring was done as follows: First, we determined a latency onset criterion of 300 ms to 700 ms, based on our set time-windows for determining the peak amplitude of the difference wave. The P300 ERP was determined as the first positive going peak from the set onset criterion (300 ms in our case). Then, for each subsample, ERP latency was calculated using a relative criterion technique: “the time-point at which the amplitude reaches a constant, pre-specified percentage of the peak value” (Kiesel et al., 2008). We chose to take 50% as the pre-specified value, one of the percentage values that came out to be most reliable to use when measuring latencies of the P300 in an oddball paradigm (Kiesel et al., 2008). We submitted the latency outcome values of this jackknife-based approach to a 2x2 repeated measures analysis of variance (ANOVA; group x condition). The F-value of this ANOVA was adjusted according to the following formula: Ulrich & Miller (2001). It was not possible to assess individual differences using the Jack-knife based approach. We therefore also calculated the latency values per single-subject waveform (with the same scoring method and settings) and used the Fligner-Killeen test to test for homogeneity of variance between-groups. Although we are aware of the pitfalls of the single-subject waveform method, we wanted to be as thorough as possible in exploring differences in variation per group.

Correlation with behavioural speech perception scores and duration of deafness

Lastly, we correlated MA with behavioural speech perception scores of the CI users as measured in the clinic and with duration of deafness. We used a non-parametric correlation method: the Spearman's rho test. Peak amplitude was not included. This decision was based on the fact that both amplitude measures showed a similar pattern of results and mean amplitude has been argued to be more reliable (Luck, 2012; Woodman, 2010). For consistency, we correlated single-subject waveform ERP latency with behavioural speech perception scores and duration of deafness.

The behavioural speech perception scores were obtained in the clinic using a word intelligibility task (NVA lijsten, Bosman Wouters & Dumman, 1995). In this task, isolated one-syllable words are presented at 70 dB in an audio-booth and the CI user is asked to repeat the word they are presented with. The percentage correctly repeated words is used as an outcome measure. Almost all scores were obtained within the same half year as the EEG was conducted. For CI user 1 and user 5, there were no up-to-date perception scores so their lastly available data were used. For CI user 1 this was obtained five years ago, for CI user 5 this was obtained one year ago. The decision was made to not exclude these values in the analysis because the CI users were all implanted at a young age and their perception scores are assumed to have relatively stabilized over the years. The bilaterally implanted CI users were all tested with only their right CI on, the same CI as they chose to have on during the EEG-measurements.

Duration of deafness was obtained from the demographic information of our participants. It is displayed in years in table 1.

Results

Behavioural assessments

Reaction time experiment.

Results for the behavioural reaction-time task - per group per contrast condition - are displayed in Figure 1. A two-way ANOVA revealed a main effect of group (*F* (1,40) = 15.56, *p* < .001). The normal-hearing group pressed significantly faster (*M* = 457.61, *SD* = 108.77) than the CI user group (*M* = 588.35, *SD* = 162.29) in both contrast conditions. There was also a main effect of contrast condition (*F* (1,40) = 21.17, *p* < .001). Both groups pressed significantly faster in the frequency contrast condition (*M* = 431.80, *SD* = 124.83) than in the consonant contrast condition (*M* = 431.81, *SD* = 124.31). Descriptively, the plot shows that the CI user group tends to be slower than the normal-hearing group in the consonant contrast condition. This interaction effect was, however, not significant (*F* (1,40) = 2.72, *p* = .100).

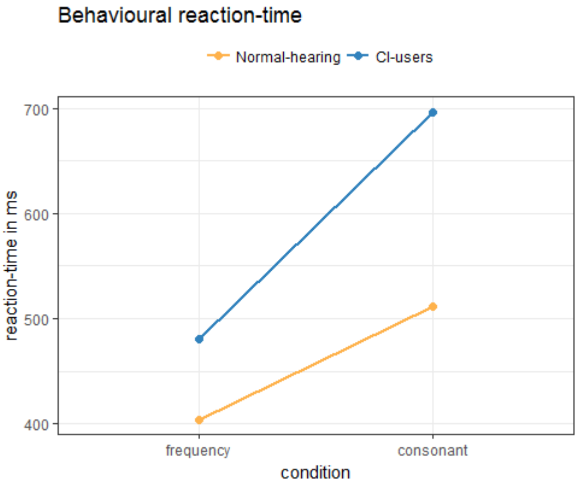


Fig. 1. Behavioural reaction time task results. Mean reaction time and standard deviations in milliseconds as a function of group (Normal-hearing (n = 14), CI users (n = 8)) and contrast condition (frequency, consonant)

Counting deviants during the attentive task-condition

Scatterplots with standard scores (score per participant in both groups minus the normal-hearing group's mean divided by the normal-hearing group's SD) of the amount of deviants counted per block during the attentive task-condition are displayed in Figure 2, for each contrast condition separately. In total, 30 deviants could be counted in each block. The normal-hearing group had a mean of 29.60 (*SD* = 2.47) over both blocks in the frequency contrast condition, and a mean of 29.82 (*SD* = 1.70) over both blocks in the consonant contrast condition. The CI user group had a mean of 30.50 (*SD* = 1.32) over both blocks in the frequency contrast condition and a mean of 29.38 (*SD* = 2.00) over both blocks in the consonant contrast condition.

Group ERP results

Group averaged ERP results for both groups collapsed over the electrodes 'CP1', 'CP2', 'P3', 'P4', 'Pz', 'C3', 'C4' for the attentive task-condition (P300) and 'Fz', FCz', 'F3', 'F4', 'FC1' and 'FC2' for the inattentive task-condition (MMN) are displayed respectively in Figure 3.1 A and B and 3.2 A and B. In the normal-hearing group, time-locked EEG-activity for the deviant trials was found to be significantly more positive in amplitude than activity for the standard trials in the attentive task condition. This was found for both the frequency (270-690 ms, *p* = .002) and the consonant contrast (390-690 ms, *p* = .002). The CI user group yeilded similar results. A significant positive deflection was found for the frequency (270-660 ms, *p* = .002) and the consonant contrast (350-660 ms, *p* = .002). The time-windows all correspond roughly to the P300 component as described in the literature (usually present from 350-500 ms). For the inattentive task-condition, time-locked EEG activity for the deviant trials was found to be significantly more negative in amplitude than activity in the standard trials, but only for the frequency contrast. This was found for the normal-hearing group (90-160 ms, *p* = .020) as well as for the CI user group (120-200 ms, *p* = .020). The time-windows correspond roughly to the MMN component as described in the literature (usually present from 150-250 ms). For the consonant contrast, significant negative deflections were found for neither the normal-hearing group (*p* = .090), nor the CI user group (*p* = 1.000). Scalp topographies of the difference wave (standard deviant) of the group-averaged ERPs are displayed respectively in Figure 3.1 C and D and 3.2 C and D. For the attentive task-conditions the clusters were detected over centro-parietal regions. For the frequency contrast in the inattentive task-condition, the clusters were detected over fronto-central regions.

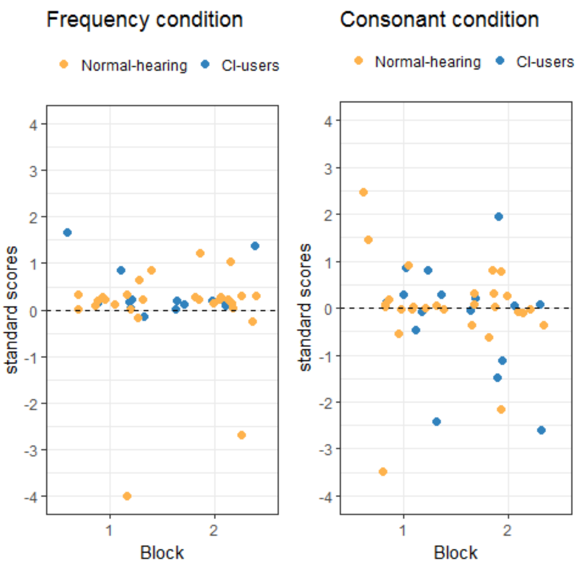


Fig. 2. Results of deviant counting. Scatterplots showing the standard scores in both groups of the amount of deviants counted during the attentive task-condition. On the left the results for the frequency contrast condition are displayed, on the right the results for the consonant contrast condition are displayed.

Individual ERP results

Individual ERP results per task condition per contrast condition collapsed over the electrodes 'CP1', 'CP2', 'P3', 'P4', 'Pz', 'C3', 'C4' for the attentive task condition (P300) and 'Fz', FCz', 'F3', 'F4', 'FC1' and 'FC2' for the inattentive task condition (MMN) are displayed respectively in Figures 4 (attentive-frequency), 5 (attentive-consonant), 6 (inattentive-frequency) and 7 (inattentive-tone). In each of these figures the standard trials and the deviant trials are plotted per individual, as well as the time-window(s) in which the difference between these trials was significant (if there were significant differences). This significant time-window is indicated by the dashed lines in the figures. In each of the figures two randomly picked individual ERP plots from the normal-hearing control group are displayed. In the other eight plots of each of the figures, all individual ERPs of the CI users are displayed. Figure 8 shows the number of individuals for whom the ERP effects were statistically present per group and per contrast condition. An ERP waveform was deemed present if there was at least one positive (only for the attentive task condition) or one negative (only

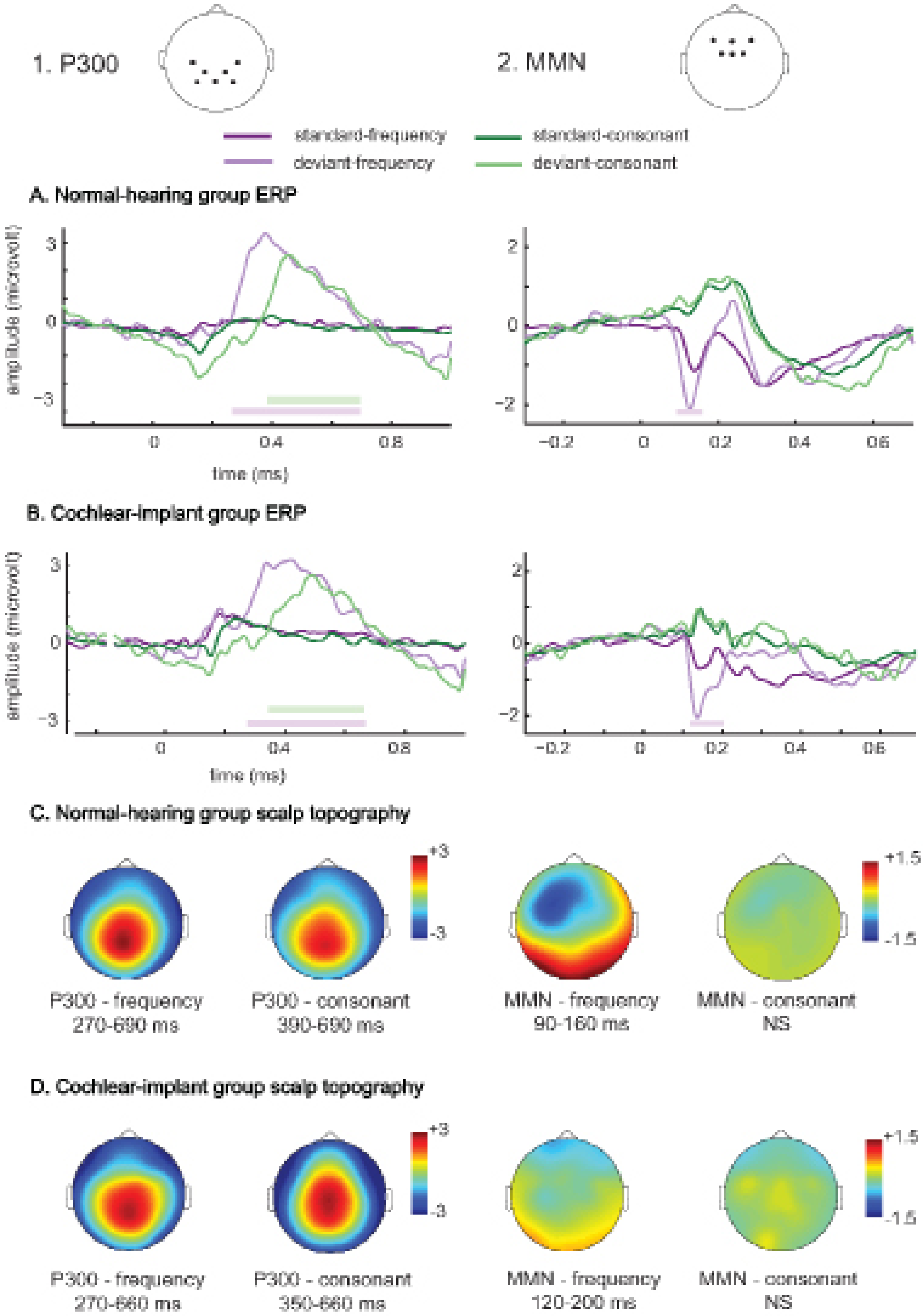


Fig. 3. Group ERP results. Group averaged ERP waveforms (**A, B**) for the standard-frequency, deviant frequency, standard-consonant and deviant-consonant contrast trials and corresponding difference-wave (standard-deviant) scalp topographies (μV) (**C, D**) per group (normal-hearing and CI user group) per task-condition (1. P300 [attentive], 2. MMN [inattentive]). EEG-cap configurations are also shown per task-condition. The time-windows in which the standard trials were significantly different from the deviant trials are indicated by the light purple and light green bars in the ERP plots

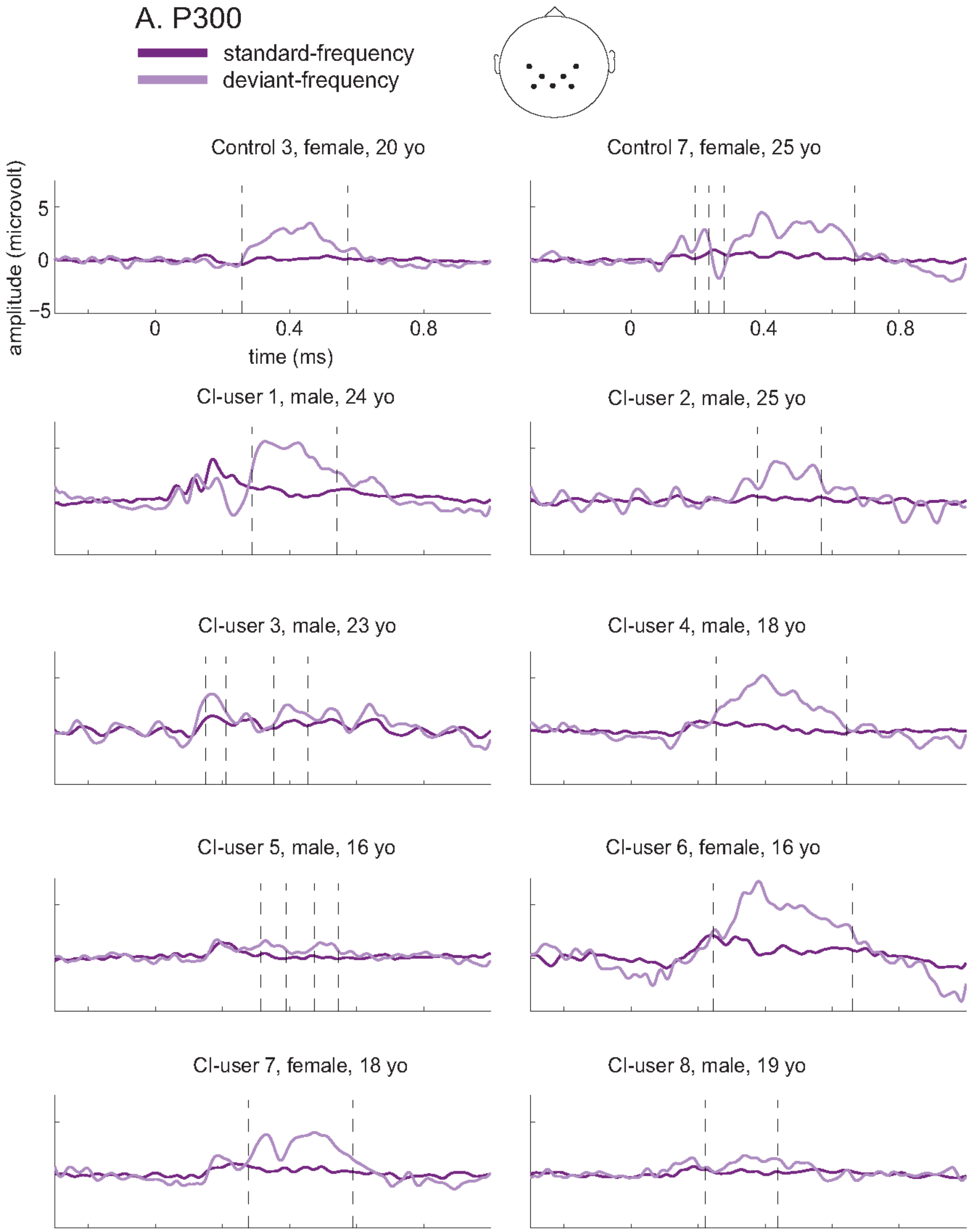


Fig. 4. Individual ERP results attentive-frequency. Individual ERP waveforms for the standard-frequency and deviant frequency trials. At the top of the graph, two randomly picked ERPs of normal-hearing controls are displayed as reference. The next eight ERPs correspond to each individual CI user (n = 8). The EEG-cap configuration is shown. Dashed lines indicate significance

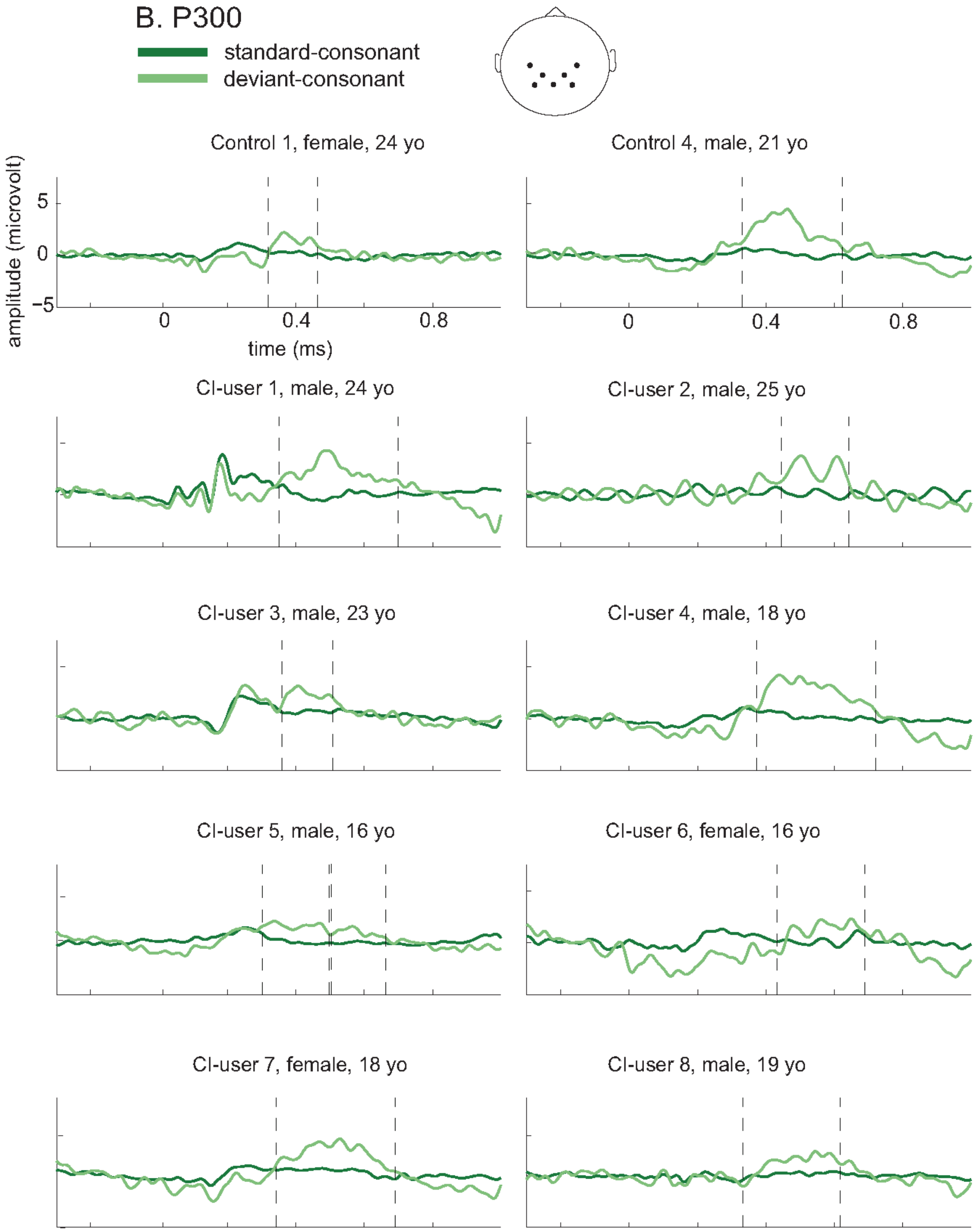


Fig. 5. Individual ERP results attentive-consonant. Individual ERP waveforms for the standard-consonant and deviant consonant trials. At the top of the graph, two randomly picked ERPs of normal-hearing controls are displayed as reference. The next eight ERPs correspond to each individual CI user (n = 8). The EEG-cap configuration is shown. Dashed lines indicate significance

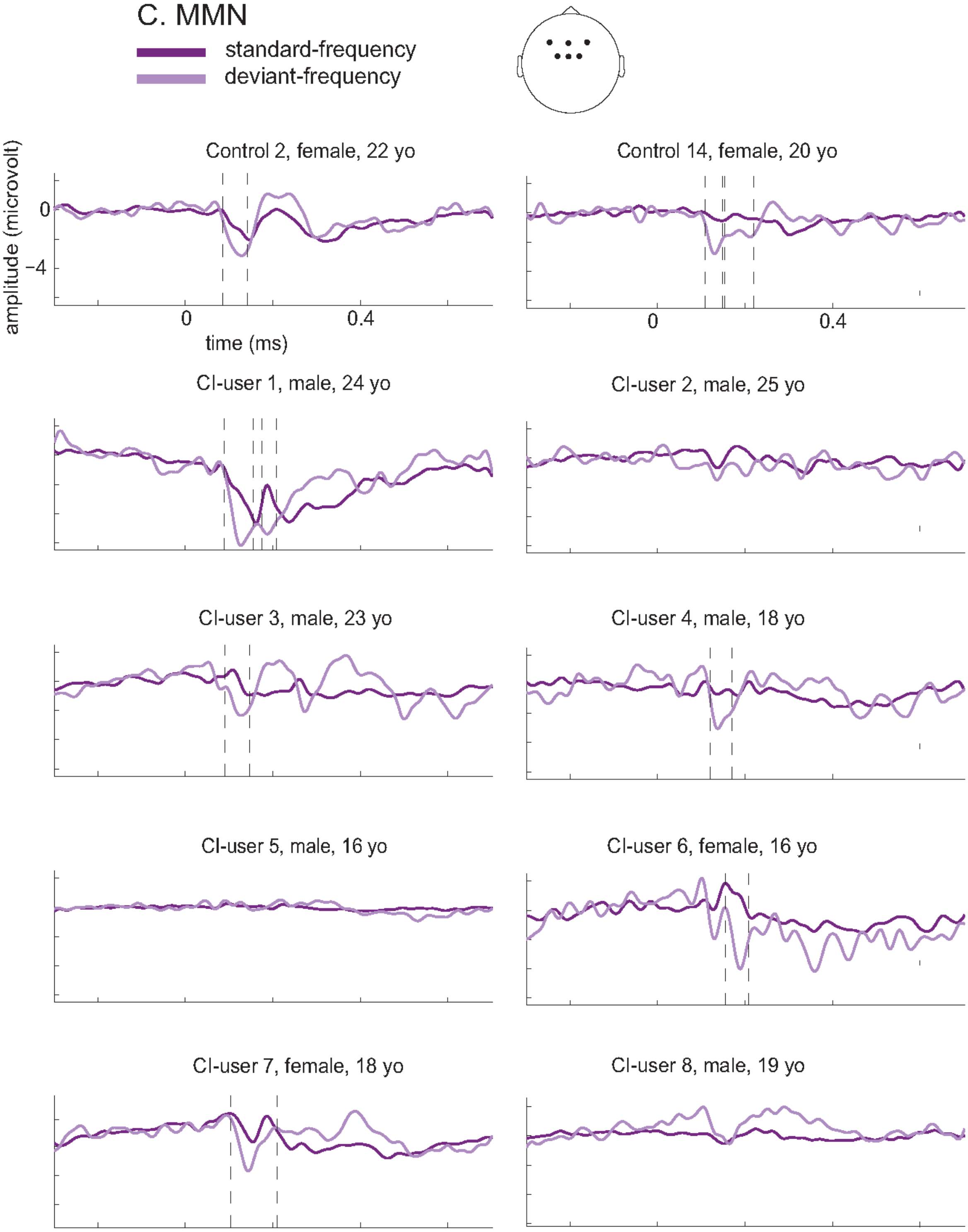


Fig. 6. Individual ERP results inattentive-frequency. Individual ERP waveforms for the standard-frequency and deviant frequency trials. At the top of the graph, two randomly picked ERPs of normal-hearing controls are displayed as reference. The next eight ERPs correspond to each individual CI user (n = 8). The EEG-cap configuration is shown. Dashed lines indicate significance

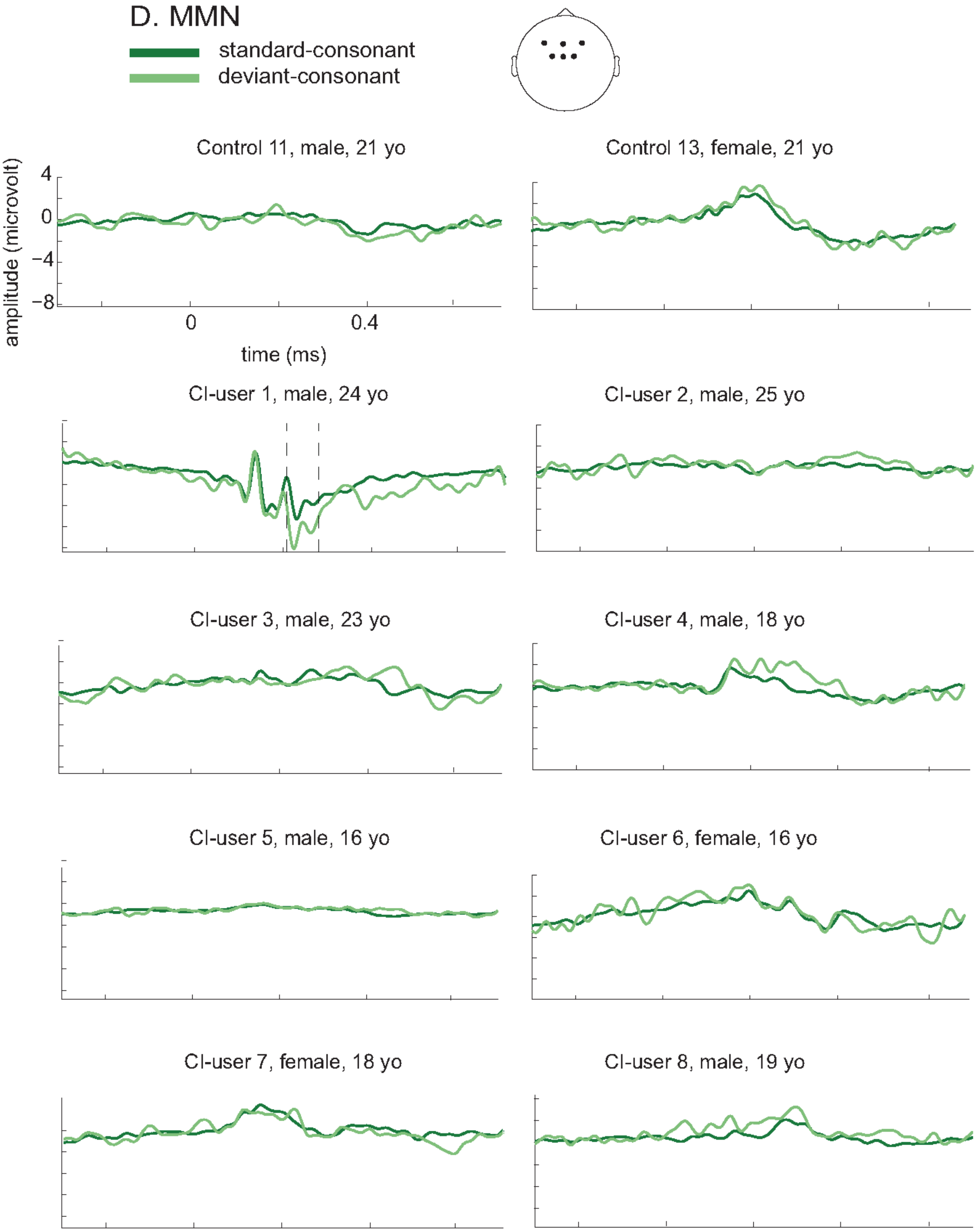


Fig. 7. Individual ERP results inattentive-consonant. Individual ERP waveforms for the standard consonant and deviant consonant trials. At the top of the graph, two randomly picked ERPs of normal-hearing controls are displayed as reference. The next eight ERPs correspond to each individual CI user (n = 8). The EEG-cap configuration is shown. Dashed lines indicate significance

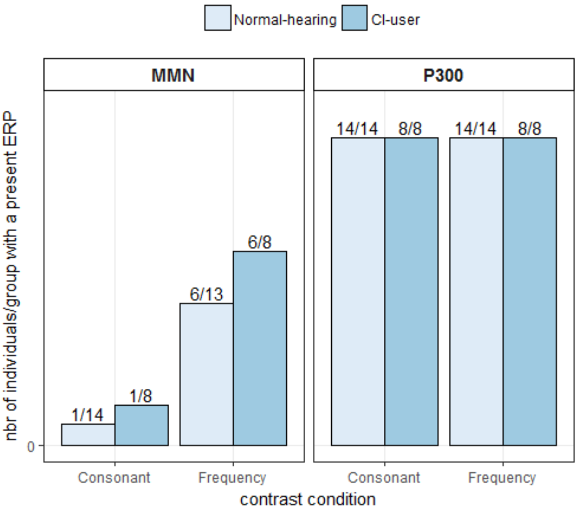


Fig. 8. Number of individuals with a significant individual ERP waveform. The number of individuals per group that had a significant ERP waveform, split by contrast condition

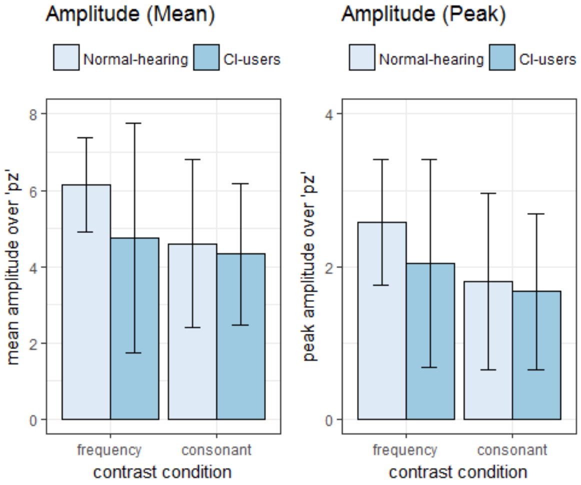


Fig. 9. Amplitude results. Plots showing the mean and standard deviations of the mean amplitude (left) and peak amplitude (right) in microvolt of the difference waveforms per group per contrast condition, as measured in the attentive condition. Amplitude was calculated over electrode 'Pz'

for the inattentive task condition) significant cluster in the pre-specified time-windows. A table with all p-values of the difference between standard and deviant trials and corresponding time-windows per participant (n = 22) is displayed in Appendix A.

Amplitude results

Means and standard deviations of our two amplitude measures, the mean amplitude and the peak amplitude - per group per contrast condition - are displayed in Figure 9. The Wilcoxon rank-sum test (for group) and signed-rank test (for condition) did not show significant differences between groups and contrast conditions, neither for the mean amplitude (W = 107, p = .45 for the group comparison, V = 38, p = .12 for the contrast condition comparison), nor the peak amplitude (W = 101, p = .32 for the group comparison, V = 40, p = .16 for the contrast condition comparison). To test for the interaction effect of group x condition we calculated the difference of the frequency minus the contrast condition for each individual. Consequently, we tested the difference as a function of group, again using the Wilcoxon rank-sum test. We did not find a significant interaction effect for either measure (Mean amplitude: W = 20, p = .23, Peak amplitude: W = 20, p = .23). Furthermore, the Fligner-Killeen test showed no main effect of group or condition on the variance within groups on either the mean amplitude measure (χ^2 (1) = 0.42, p = .51 for group, χ^2 (1) = 0.02, p = .90 for condition), or the peak amplitude measure (χ^2 (1) = 1.26, p = .26 for group, χ^2 (1) = 0.57, p = .44 for condition). There was a significant interaction between group and condition in variance for the peak amplitude (χ^2 (3) = 9.22, p = .03), but not for the mean amplitude measure (χ^2 (3) = 4.55, p = .20). This means that when using peak amplitude as a measure, the variance within the CI user group was significantly greater than the variance within the normal-hearing group, and that this is true only for the frequency condition, not for the consonant condition.

Latency results

Results for the latency analysis per group per contrast condition are displayed in Figure 10. A two-way ANOVA revealed a main effect of condition (*F* (1,28) = 23, *p* < .001). The latency was significantly later in the consonant (*M* = 0.402, *SD* = 0.01) than in the frequency contrast condition (*M* = 0.308, *SD* = 0.01) for both groups. There was no main effect of group (*F* (1,28) = 0.18, *p* = .67), nor an interaction effect of group x condition (*F* (1,28) = 0.004, *p* = .95). The Fligner-Killeen test we performed on the latencies calculated from the single-subject ERP waveforms was not significant. There was homogeneity of variance between groups (χ^2 (1) = 0.55, *p* = .46) and between contrast

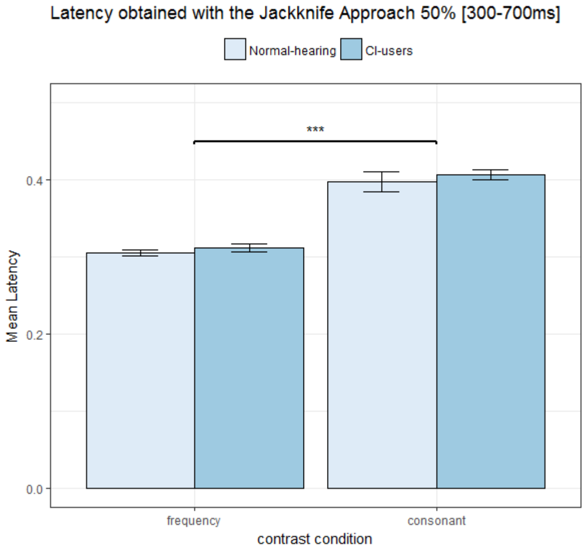


Fig. 10. Latency results. Plots showing the mean and standard deviations of the latency outcome values in seconds, as measured in the attentive task-condition. Results are shown per group per contrast condition. Results were obtained using a Jack-knife based approach and a 50% relative criterion scoring technique with a time-window of 300-700 ms. Latency was calculated over electrode 'Pz'. The asterisks indicate significant differences

conditions (χ^2 (1) = 0.65, p = .41), and there was no significant interaction effect between group x contrast condition (χ^2 (3) = 2.8, p = .41).

Relation between amplitude and latency, behavioural speech perception scores, and duration of deafness

Results of the non-parametric correlation between the behavioural speech perception scores and mean amplitude and duration of deafness and mean amplitude are displayed in Figure 11. As for speech perception, there is no significant relation between mean amplitude of the P300 components and the behavioural speech perception scores in the frequency condition (r_s = -.26, p = .53). For the consonant condition, however, there was a strong correlation between the mean amplitude of the P300 components and the behavioural speech perception scores. This showed a trend towards significance (r_s = .70, p = .05). The assumed relation is positive: the higher the behavioural speech perception score of the individual, the greater the amplitude of the P300 component. As for duration of deafness, there is no significant relation between mean amplitude of the

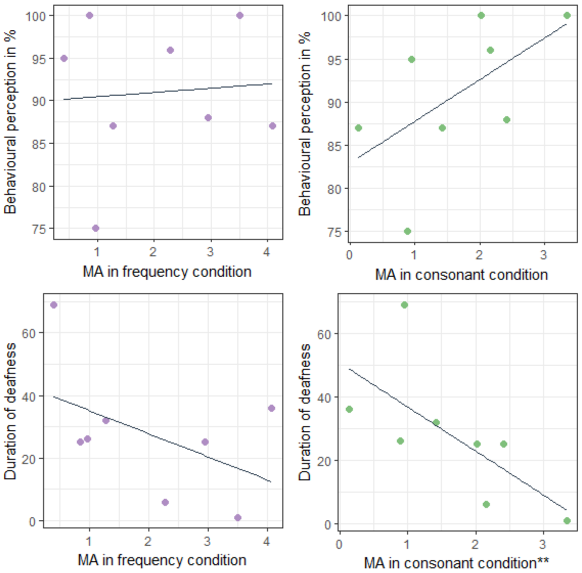


Fig. 11. Correlations between behavioural speech perception and duration of deafness. Scatterplots showing the correlations between the mean amplitude (MA) in microvolt and the behavioural speech perception scores in percentage correct (top) and duration of deafness in months (bottom), for each contrast condition (left: frequency, right: consonant)

P300 components and duration of deafness in the frequency condition (r_s = -.33, p = .41). There is, however, a significant and strong correlation between the mean amplitude of the P300 components and duration of deafness in the consonant condition (r_s = -.83, p = .009). The relation is negative, which means that the shorter a CI user has been deaf, the greater the amplitude of the P300 component.

Latency as measured using the single-subject waveform approach was not correlated with behavioural speech perception scores or duration of deafness, in either condition (duration of deafness and latency for the frequency contrast: r_s = .19, p = .64 and the consonant contrast: r_s = .06, p = .88; behavioural speech perception & latency for the frequency contrast: r_s = -.15, p = .75 and the consonant contrast: r_s = -.41, p = .30).

Discussion

Robustness of the MMN and the P300 and their suitability for the clinic

The primary aim of this study was to compare two ERP components, the P300 component and the MMN component, in terms of their ability to robustly measure auditory discrimination in normal-hearing adolescents and adolescents with a CI on an individual level. The group-averaged ERP results show that, despite a difference in sample size, the CI user group performs similar to the normal-hearing control group. The similarity of results between the two groups is in line with earlier research (MMN: Kraus et al., 1993; Ponton et al., 2000; P300: Groenen et al., 2001; Micco et al., 1995 [although not similar for the consonant contrast]), with the exception of poor-performing CI users (Beynon et al., 2002; Turgeon et al., 2014).

Furthermore, the individual results are clear-cut in indicating which ERP paradigm is most robust in measuring auditory discrimination under the set circumstances. For the attentive task condition, eliciting the P300 response, all participants showed a significant difference between the standard and the deviant waveforms, regardless of group or contrast condition. For the inattentive task condition, eliciting the MMN response, results were less robust. While in the frequency contrast condition half of the CI users and normal-hearing participants showed ERP presence, in the consonant contrast only one participant in each group showed ERP presence.

Although all individual waveforms showed significance in the attentive task-condition, the components seem to vary considerably in robustness. Although this variation is evidently present in CI users, it is also present in some controls (e.g., see control 1 in Figure 5). Therefore, on the basis of the figures, we should be careful in interpreting differences in the robustness of the waveforms as non-normal differences in auditory discrimination abilities. They may be differences that also appear in the normal-hearing population. Furthermore, it is equally likely that differences in signal-to-noise ratio may be underlying the variation in robustness.

The absence of robustness in the inattentive task condition corroborates the hypothesis that the set duration of the experiment is too short to robustly elicit the MMN. The difference in robustness between the frequency and the consonant contrast condition may be explained by a combination of a lack of power due to duration and a more complex contrast condition. This is emphasized by the findings from two earlier studies. These used almost four times as long EEG-recordings as were used in the current study to elicit the MMN with a consonant contrast (Singh et al., 2004; Turgeon et al., 2014). This finding has important implications for setting up ERP experiments in the future. When the aim is to elicit a robust MMN response on an individual level, a longer measurement than 10 minutes is needed. Interestingly, this finding contradicts the finding of an earlier study. This study found a significant MMN and P300 response in three CI users and three normal-hearing controls with only three-minute EEG-recordings, elicited using frequency contrasts (Obuchi et al., 2012). There is no straightforward explanation for this discrepancy. It may, however, be due to differences in the amount of electrodes that were used to perform the analysis over.

P300 only: differentiating between groups, within groups, and between contrasts

Amplitude and latency between- and within groups.

A second aim of this study was to evaluate whether the amplitude and latency of the individual waveforms in the attentive task condition could distinguish the CI users as a group from the normal-hearing participants as a group. Also, it was assessed whether variance on these measures was greater in the CI user group than in the normal-hearing group. This would indicate that differences in the P300 response of CI users cannot be ascribed to regular P300 variation as present in the normal-hearing population. Greater variation in the P300 response of CI users was expected based on studies that found individual differences in the behavioural speech perception abilities of prelingually deaf CI users (Pisoni et al., 2000; ASHA, 2004), and in the P300 results of prelingually deaf CI users (Beynon et al., 2002; Jordan et al., 1997; Kileny, 1991).

Results of both mean amplitude and peak amplitude showed that it was not possible to distinguish the CI user group from the normal-hearing group in either contrast condition. Latency results showed a similar pattern. This finding was not unexpected based on the results of Beynon et al. (2002), although in Beynon et al. (2005), postlinguals could be distinguished from normal-hearing controls.

A first explanation for this null result may be that differences in auditory discrimination abilities are present, but that our measurements are not able to reflect these. Firstly, because we used individual waveforms for our analysis, the signal-to-noise ratio may have been low. For the latency analysis however, this is less of an issue, because we used the Jackknife approach. Secondly, even if not only noise was measured, conclusions on the nature of an effect should be drawn with caution. It may well be that other (task-independent) cognitive or biological mechanisms such as general intelligence, attention, or the arousal state of subjects were underlying P300 components (Polich & Kok, 1994). These processes are not expected to differ between groups, which might explain why no differences were found.

A second explanation may be that the differences between the auditory discrimination abilities of CI users and normal-hearing participants are too subtle to be elicited by the chosen measurements and design in this study. Our contrast conditions are fairly simple compared to the level of difficulty in auditory perception CI users encounter on a daily basis. This assumption would contradict the findings of older studies that this contrast does not elicit a P300 for some CI users. However, in older studies, more poor-performers were included in the analysis (Beynon et al., 2002; Jordan et al., 1997; Kileny, 1991). Poor performance in those studies was defined as no behavioural discrimination ability (Kileny, 1991; Jordan et al., 1997) or a low behavioural speech perception score (e.g. <65% on an open-set speech recognition task in Beynon et al., 2002). All CI users in the current study could discriminate between the stimuli (although as a group they took somewhat longer). Their behavioural speech perception scores were high (mean 90%, ranging from 75-100%).

Another explanation for such a ceiling performance may be that differences in perception abilities (at least for speech in isolation) between (prelingually deaf) CI users and normal-hearing participants have diminished over the years. It is possible nowadays to be implanted from a very early age. Compare for example the age of implantation in older studies on prelingually deaf CI users (range 5-33 yo in Beynon et al., 2002 and Jordan et al., 1997 combined), to the age of implantation in our study (range 1-5 yo). For prelingually deaf users in general, this early implantation means that their period of auditory deprivation (duration of deafness) diminishes considerably and, furthermore, that their auditory cortex can start developing while it is still flexible.

Considering the variance within groups, greater variance in peak amplitude was found within the CI user group as opposed to the normal-hearing group. This was found only for the frequency contrast. While this implies to corroborate the expectation of non-normal individual differences in the auditory perception abilities of CI users, it does not rhyme with our behavioural results. For the deviant counting, the spread in the consonant condition was greater for both groups. For the reaction time task, although only descriptively, the normal-hearing group showed greater variance in the frequency condition, while the CI users showed greater variance in the consonant contrast condition. Moreover, because this result was not found for the mean amplitude measure, even though this measure has been argued to be more reliable (Luck, 2012; Woodman, 2010), this result should be interpreted with caution.

Latency variance as obtained with the Jackknife approach was not used in our analysis of the variance within groups. It was not possible to correct for the reduced variance as a result of this approach (Miller & Ullrich, 2001). When using the latency of the single-subject ERP waveforms, there were no differences in latency variation between the two groups. The absence of greater variance in auditory processing in the CI user group as opposed to the normal-hearing group again confirms the hypothesized ceiling effect for our CI users.

Amplitude and latency between contrast conditions.

We replicate earlier studies in finding that, using the P300 as a measure for auditory discrimination, it is likely that the discrimination of stimuli by the brain lies on a continuum of complexity, with more complex stimuli being more difficult to discriminate (see Polich, 2004 for a review on healthy subjects).

Although we found a longer latency for the consonant contrast as opposed to the frequency contrast, we did not find the same result for the amplitude. Descriptively, however, there was a trend towards the more complex condition yielding a lower amplitude. The greater variance in the frequency condition for the CI users may explain why the difference is not significant. Furthermore, the difference in robustness of the effects for amplitude as opposed to latency may be again due to the fact that we obtained the latency from eight times an n-1 group sample, while we used the individual waveforms to obtain the amplitude.

These results are fruitful for the long-term goal of this study to develop a neurophysiological predictor for speech perception abilities. If input conditions are made more complex in the future, latency and amplitude differences may increase. That way, it may be possible to highlight the more subtle differences in speech perception that were not picked up by the current design.

P300 only: relation between neurophysiological results, behavioural results, and duration of deafness

In the light of our long-term aim to develop a marker for speech perception abilities it is important to link neurophysiological results to behavioural results. Despite the small sample size of this study, it was found that lower behavioural speech perception resulted in a lower amplitude of the P300 response. This was found only for the consonant contrast. This result implies that the lower the amplitude of the P300 component, the harder it is to perceptually discriminate between phonemes. Phoneme discrimination is a very important aspect of the speech perception process. Important to keep in mind is that perceptual discrimination is necessary, but not sufficient for the P300 to appear. The P300 amplitude is also influenced by task-dependent cognitive processes such as immediate working memory of the stimulus and attention allocation (Polich, 2004, 2010). Differences in the development of the auditory processing function may have resulted in differences in these memory and attention processes used for speech perception, resulting in more difficulty for some users as opposed to others. Conclusions on the exact contribution of processes underlying the link between the P300 amplitude and speech perception should be drawn with caution. Differences in working memory and attention allocation processes may appear independently of CI or deafness. This also has implications for the specificity of the P3 amplitude as a marker for speech perception ability. If a CI user presents with an absent P3, we cannot be sure whether this is the result of a CI and/or deafness related auditory discrimination deficit, the result of a (general) deficit in working memory updating, or simply lack of attention from the CI user during the task. However, in terms of sensitivity of the marker, it is unlikely that a P3 is present while perceptual discrimination is not.

The relation between behavioural speech perception and P300 amplitude in the consonant condition furthermore shows that tone discrimination (frequency contrast) says less about speech perception abilities than speech discrimination (consonant contrast). As was already laid out in the introduction, the relation between the consonant contrast /ba/ vs. /da/ and behavioural speech perception has not been investigated much in CI populations. Earlier articles on the P300 (Groenen et al., 2001) and the MMN (Kelly et al., 2005) found a relation between speech perception scores and frequency and vowel contrasts as opposed to consonant contrasts. However, it is hard to compare our findings to theirs for several reasons. Firstly, they tested postlingually deaf adults. Secondly, ERPs for consonant contrasts were not measured (Kelly et al.,

While the former relation was on the verge of significance, the relation between duration of deafness and amplitude of the P300 was robust. This was again found for the consonant contrast condition only. This implies that the longer a CI user has been deaf before implantation, the lower their P300 amplitude in response to speech stimuli. This finding is evidence for the fact that the auditory cortex is flexible enough to adapt to speech input after implantation, when performed early in development as is nowadays more and more the case with prelingually deaf CI users. Adaptation success seems to decrease as a function of the duration of speech deprivation (other factors that may play a role in this process put aside). This finding has been robustly established in the literature on obligatory cortical auditory evoked potentials (CAEP; Sharma, Dorman & Spahr, 2002a; Sharma, Spahr & Dorman, 2002b) but has not often been confirmed using the discriminative CAEPs MMN and P300. Our relatively short P300 experiment was able to capture this, and the findings correspond well to the correlation between behavioural speech perception and P300 amplitude in response to speech. This is evidence that the P300 response, as we measured it, is meaningful, despite individual waveform noise.

The results of the correlation measures are in line with our other behavioural results. That is, for the deviant counting during the P300 experiments, CI users (and this is also true for some normal-hearing participants) showed more deviations from the normal-hearing mean in the consonant condition than in the frequency condition. This implies that the stimuli in the consonant condition were somewhat more difficult to discriminate. However, results should be interpreted with caution. Someone with a low score on deviant counting does not have to have a lower discrimination ability. Deviant counting also measures processes other than perceptual discrimination, such as context-updating and attention allocation. The main function of the deviant-counting task was to make sure our participants paid attention to the stimulus. The context updating required for the deviant counting task is not expected to present a confound to our results, because it is expected that perceptual difficulties occur prior to context updating and also influence it linearly. The results of the behavioural reaction time task also showed a trend towards the CI users being slower than the normal-hearing controls, but only for the consonant condition. It may be due to a small sample size that this trend was not significant. This greater spread and difficulty for CI users in the contrast condition may have led to this condition being a more sensitive measure for differences in auditory processing of speech.

A relation between latency and duration of deafness or behavioural speech perception was not found. An explanation for this may be that the single-subject waveform method is low in power. Rho-values were found in the right direction for both conditions, but they were very small. Latency may not be the most suitable measure to elicit relations between individual ERPs and behavioural outcomes, if any present.

Recommendations and limitations

The current study shows that as a clinical predictor, the P300 is more robust on an individual level with a limited amount of duration. For future research it is important to focus on this P300 response and to extend the current findings to a population of young children. For this population, it is even more important than for adolescents to obtain objective and all-round information on their auditory abilities.

Furthermore, the current study replicates, under contemporary implantation circumstances, the earlier found quality of the P300 response as a possible marker of perceptional challenges for the cochlear implant population. It paves the way for further research into predicting the auditory processing of speech in more difficult conditions. For example, recent research has focused on measuring the P300 in noise in postlingually deaf elderly CI users. It was found that for the CI users as a group, the P300 response was absent in the measurements with white noise as opposed to the measurements in quiet (Soshi et al., 2014). Measuring the P300 in noise may be the first step towards highlighting more subtle differences in processing. Furthermore, on a behavioural level, testing speech perception abilities in different conditions of noise also yields promising opportunities to investigate in more detail the challenges for CI users. Differences have already been found for speech perception ability masked by two-talker babble as opposed to steady-state noise (Hillcock-Dunn, Taylor, Buss & Leibold, 2015 on hearing-aid users) or white noise (Soshi et al., 2014 on CI users)

The assumed ceiling effect in our study showed that speech in isolation for early implanted, prelingually deaf adolescent CI users may be peer-like. This implies that the results of our study are in favour of the hypothesis of de Hoog et al. (2016a), who proposed a discrepancy between linguistic problems and speech perception problems for some, but not all, CI users. The linguistic problems of our sample were not taken as a variable in this study, but it was observed in the clinic that linguistic ability varied considerably and that it did not show a one-to-one mapping with behavioural speech perception scores. This indicates that this discrepancy should gain attention in future research. It should, furthermore, be investigated whether the discrepancy persists when measuring speech perception in more ecologically valid conditions.

This study is limited by its small sample size. It should also be taken into account that the signal-to-noise ratio of individual waveforms remains low (both found for CI users and for controls) and this may have distorted interpretations. We, however, have tried our best to diminish this risk by using new, well-argued-for analysis methods. The trade-off between more robust results for the group level (latency analysis) or more information on individual performance (amplitude analysis) remains a problem.

Conclusions

Our study shows that when taking into account suitability for children and clinical utility in general, an attentive task paradigm (eliciting the P300 component), as opposed to an inattentive task paradigm (eliciting the MMN component), is the most robust in highlighting auditory discrimination abilities, at least when using frequency and consonant contrasts. The P300 component is robust on an individual level. When latency of the component is taken as a measure the component can distinguish between complex vs. simple conditions, even with a small sample size. The P300 component, however, cannot (yet) distinguish between CI users and normal-hearing controls, or between individual CI users. This may be due to a ceiling effect in the performance of the CI users: their speech perception is relatively good, and the input conditions may have made the auditory discrimination too easy. The suitability of the P300 as a marker for speech perception ability is backed up by the findings that the amplitude of the P300 is related to behavioural speech perception as measured in the clinic and duration of deafness. Future research should focus on investigating the P300 in relation to behavioural speech perception in younger CI users, as well as using more challenging, ecologically valid experimental conditions.

Acknowledgements

A special thanks to all the cochlear implant users that were willing to participate in this study. I furthermore would like to thank all the normal-hearing students that participated, Milou van Helvert for helping me with the data collection, and Randi Goertz and Marlijn ter Bekke for proof-reading this thesis. I thank Dr. Margreet Langereis, Dr. Anneke Vermeulen and Dr. Andy Beynon for providing me with the creative foundations for this thesis, thorough theoretical support and their clinical expertise. Lastly, I thank my supervisor Vitória Piai for her excellent supervision during the past year, for sharing with me her ideas and her in-depth expertise, and for educating and supporting me where and whenever possible.

References

American Speech-Language-Hearing Association. (2004). *Cochlear implants*. [Technical Report]. Retrieved April 14, 2018, https://www.asha.org/policy/TR2004-00041/

Boersma, P., & Weenink, D. (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0.39, retrieved 12 April 2018 from http://www.praat.org/

Bosman, A. J., Wouters, J., & Damman, W. (1995). Realisatie van een cd voor spraakaudiometrie in Vlaanderen. *Logopedie en foniatrie: maandblad van de Nederlandse vereniging voor logopedie en foniatrie*, *67*(9), 218-225.

Beynon, A. J., Snik, A. F., & van den Broek, P. (2002). Evaluation of cochlear implant benefit with auditory cortical evoked potentials. *International journal of audiology*, *41*(7), 429-435.

Beynon, A. J., Snik, A. F. M., Stegeman, D. F., & Van den Broek, P. (2005). Discrimination of speech sound contrasts determined with behavioral tests and event-related potentials in cochlear implant recipients. *Journal of the American Academy of Audiology*, *16*(1), 42-53.

Bishop, D. V. M., & Hardiman, M. J. (2010). Measurement of mismatch negativity in individuals: A study using single-trial analysis. *Psychophysiology*, *47*(4), 697-705.

van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behavior research methods*, *38*(4), 584-589.

Geers, A. E. (2002). Factors affecting the development of speech, language, and literacy in children with early cochlear implantation. *Language, Speech, and Hearing Services in Schools*, *33*(3), 172-183.

Gilley, P. M., Sharma, A., Dorman, M., Finley, C. C., Panch, A. S., & Martin, K. (2006). Minimization of cochlear implant stimulus artifact in cortical auditory evoked potentials. *Clinical Neurophysiology*, *117*(8), 1772-1782.

Groenen, P. A., Beynon, A. J., Snik, A. F., & Broek, P. V. D. (2001). Speech-evoked cortical potentials recognition in cochlear implant users and speech. *Scandinavian audiology*, *30*(1), 31-40

Hendriks, M. P., Kessels, R. P. C., Gorissen, M. E. E., Schmand, B. A., & Duits, A. A. (2014). *Neuropsychologische diagnostiek. De klinische praktijk*. Amsterdam: Boom.

Hillock-Dunn, A., Taylor, C., Buss, E., & Leibold, L. J. (2015). Assessing speech perception in children with hearing loss: What conventional clinical tools may miss. *Ear and hearing*, *36*(2), e57-60.

de Hoog, B. E., Langereis, M. C., Weerdenburg, M., Knoors, H. E., & Verhoeven, L. (2016a). Linguistic profiles of children with CI as compared with children with hearing or specific language impairment. *International journal of language & communication disorders*, *51*(5), 518-530.

de Hoog, B. E., Langereis, M. C., van Weerdenburg, M., Keuning, J., Knoors, H., & Verhoeven, L. (2016b). Auditory and verbal memory predictors of spoken language skills in children with cochlear implants. *Research in developmental disabilities*, *57*, 112-124.

Johnson, J. M. (2009). Late auditory event-related potentials in children with cochlear implants: a review. *Developmental neuropsychology*, *34*(6), 701-720.

Jordan, K., Schmidt, A., Plotz, K., Von Specht, H., Begall, K., Roth, N., & Scheich, H. (1997). Auditory event-related potentials in post-and prelingually deaf cochlear implant recipients. *The American journal of otology*, *18*(6 Suppl), S116-117.

Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., Mckeown, M. J., Iragui, V., & Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, *37*(2), 163-178.

Kelly, A. S., Purdy, S. C., & Thorne, P. R. (2005). Electrophysiological and speech perception measures of auditory processing in experienced adult cochlear implant users. *Clinical Neurophysiology*, *116*(6), 1235-1246.

Kiesel, A., Miller, J., Jolicœur, P., & Brisson, B. (2008). Measurement of ERP latency differences: A comparison of single-participant and jackknife-based scoring methods. *Psychophysiology*, *45*(2), 250-274.

Kileny, P. R. (1991). Use of electrophysiologic measures in the management of children with cochlear implants: brainstem, middle latency, and cognitive (P300) responses. *The American journal of otology*, *12*(Suppl), 37-42.

Kileny, P. R., Boerst, A., & Zwolan, T. (1997). Cognitive evoked potentials to speech and tonal stimuli in children with implants. *Otolaryngology-head and neck surgery*, *117*(3), 161-169.

Kilner, J. M. (2013). Bias in a common EEG and MEG statistical analysis and how to avoid it. *Clinical Neurophysiology*, *124*(10), 2062-2063.

Kraus, N., Micco, A. G., Koch, D. B., McGee, T., Carrell, T., Sharma, A., Wiet, R. J. & Weingarten, C. Z. (1993). The mismatch negativity cortical evoked potential elicited by speech in cochlear-implant users. *Hearing research*, *65*(1-2), 118-124.

Luck, S. J. (2005). Ten simple rules for designing ERP experiments. In T. C. Hardy (Ed.), *Event-related potentials: A methods handbook* (pp. 17-32). Cambridge, Massachussetts: The MIT Press

Luck, S. J. (2012). Event-related potentials. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology, Vol. 1. Foundations, planning, measures, and psychometrics* (pp. 523-546). Washington, DC, US: American Psychological Association.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of neuroscience methods*, *164*(1), 177-190.

Micco, A. G., Kraus, N., Koch, D. B., McGee, T. J., Carrell, T. D., Sharma, A., Nicol, T. & Wiet, R. J. (1995). Speech-evoked cognitive P300 potentials in cochlear implant recipients. *American Journal of Otology*, *16*(4), 514-520.

Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology, 38*(1), 1-21.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clinical neurophysiology*, *118*(12), 2544-2590.

Obuchi, C., Harashima, T., & Shiroma, M. (2012). Auditory evoked potentials under active and passive hearing conditions in adult cochlear implant users. *Clinical and experimental otorhinolaryngology*, *5*(Suppl 1), 6-9.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational intelligence and neuroscience*, *2011*, 1.

Pisoni, D. B., Cleary, M., Geers, A. E., & Tobey, E. A. (1999). Individual differences in effectiveness of cochlear implants in children who are prelingually deaf: New process measures of performance. *The Volta Review*, *101*(3), 111-164.

Pisoni, D. D., & Geers, A. E. (2000). Working memory in deaf children with cochlear implants: Correlations between digit span and measures of spoken language processing. *The Annals of otology, rhinology & laryngology. 109*(Suppl 12) , *92-93*.

Polich, J. (1987). Task difficulty, probability, and inter-stimulus interval as determinants of P300 from auditory stimuli. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *68*(4), 311-320.

Polich, J., & Kok, A. (1995). Cognitive and biological determinants of P300: an integrative review. *Biological psychology*, *41*(2), 103-146.

Polich, J. (2004). Neuropsychology of P3a and P3b: A theoretical overview. *Brainwaves and mind: Recent developments*, 15-29.

Polich, J. (2012). Neuropsychology of P300. In S.J. Luck, E.S. Kappenman (Eds.), *Oxford handbook of event-related potential components*,159-188. New York: Oxford University Press Inc.

Ponton, C. W., Eggermont, J. J., Don, M., Waring, M. D., Kwong, B., Cunningham, J., & Trautwein, P. (2000). Maturation of the mismatch negativity: effects of profound deafness and cochlear implant use. *Audiology and Neurotology*, *5*(3-4), 167-185.

Ruffin, C. V., Kronenberger, W. G., Colson, B. G., Henning, S. C., & Pisoni, D. B. (2013). Long-term speech and language outcomes in prelingually deaf children, adolescents and young adults who received cochlear implants in childhood. *Audiology and Neurotology*, *18*(5), 289-296.

Schorr, E. A., Roth, F. P., & Fox, N. A. (2008). A comparison of the speech and language skills of children with cochlear implants and children with normal hearing. *Communication Disorders Quarterly*, *29*(4), 195.

Sharma, A., Dorman, M. F., & Spahr, A. J. (2002a). A sensitive period for the development of the central auditory system in children with cochlear implants: implications for age of implantation. *Ear and hearing*, *23*(6), 532-539.

Sharma, A., Spahr, A., Dorman, M., & Todd, N. W. (2002b). Early cochlear implantation in children allows normal development of central auditory pathways. *Annals of Otology, Rhinology & Laryngology*, *111*(Suppl 5), 38-41.

Singh, S., Liasis, A., Rajput, K., Towell, A., & Luxon, L. (2004). Event-related potentials in pediatric cochlear implant patients. *Ear and hearing*, *25*(6), 598-610.

Soshi, T., Hisanaga, S., Kodama, N., Kanekama, Y., Samejima, Y., Yumoto, E., & Sekiyama, K. (2014). Event-related potentials for better speech perception in noise by cochlear implant users. *Hearing research*, *316*(C), 110-121.

Svirsky, M. A., Robbins, A. M., Kirk, K. I., Pisoni, D. B., & Miyamoto, R. T. (2000). Language development in profoundly deaf children with cochlear implants. *Psychological science*, *11*(2), 153-158.

Turgeon, C., Lazzouni, L., Lepore, F., & Ellemberg, D. (2014). An objective auditory measure to assess speech recognition in adult cochlear implant users. *Clinical Neurophysiology*, *125*(4), 827-835.

Ulrich, R., & Miller, J. (2001). Using the jackknife-based scoring method for measuring LRP onset effects in factorial designs. *Psychophysiology*, *38*(5), 816-827.

Viola, F. C., De Vos, M., Hine, J., Sandmann, P., Bleeck, S., Eyles, J., & Debener, S. (2012). Semi-automatic attenuation of cochlear implant artifacts for the evaluation of late auditory evoked potentials. *Hearing research*, *284*(1-2), 6-15.

Watson, D. R., Titterington, J., Henry, A., & Toner, J. G. (2007). Auditory sensory memory and working memory processes in children with normal hearing and cochlear implants. *Audiology and Neurotology*, *12*(2), 65-76.

Woodman, G. F. (2010). A brief introduction to the use of event-related potentials in studies of perception and attention. *Attention, Perception, & Psychophysics*, *72*(8), 2031-2046.

Group	Attentive (P300)		Inattentive (MMN)	
	Frequency	Consonant	Frequency	Consonant
Patients (n=8)				

1	0.288-0.542(**)	0.350-0.700(**)	0.089-0.155(*)
			0.175-0.209(*)
				0.208-0.280(*)

2	0.378-0.567(**)	0.444-0.641(**)	NS	NS

3	0.149-0.212(*)
	0.352-0.454(*)
		0.358-0.509(**)	0.091-0.147(**)	NS

4	0.255-0.643(**)	0.374-0.720(**)	0.120-0.170(**)	NS

5	0.475-0.546(*)
	0.315-0.389(*)
		0.302-0.497(**)
		0.501-0.663(**)
			NS	NS

6	0.244-0.660(**)	0.432-0.688(**)	0.155-0.208(**)	NS
7	0.278-0.589(**)	0.343-0.690(**)	0.104-0.210(**)	NS
8	0.221-0.438(**)	0.334-0.616(**)	NS	NS
Controls (n=14)				

1	0.223-0.401(**)
	0.470-0.538(**)
	0.413-0.454(*)
		0.319-0.464(**)	NS	NS

2	0.221-0.577(**)	0.333-0.755(**)	0.085-0.141(*)	NS
3	0.259-0.575(**)	0.411-0.645(**)	NS	NS
4	0.274-0.653(**)	0.331-0.624(**)	0.302-0.350(**)	NS

5	0.300-0.575(**)	0.427-0.523(**)
		0.545-0.630(**)
			NS	0.270-0.311(*)

6	0.309-0.704(**)	0.244-0.315(**)
		0.421-0.741(**)
			0.085-0.136(**)	NS

7	0.278-0.667(**)
	0.192-0.233(*)
		0.401-0.741(**)	0.079-0.202(**)	NS

8	0.286-0.602(**)	0.346-0.712(**)	0.096-0.138(*)	NS
9	0.298-0.501(**)	0.372-0.565(**)	Missing	NS
10	0.288-0.554(**)	0.386-0.651(**)	NS	NS
11	0.270-0.532(**)	0.427-0.561(**)	NS	NS

12	0.263-0.471(**)	0.339-0.579(**)
		0.587-0.671(*)
			NS	NS

13	0.296-0.440(**)
	0.546-0.632(*)
		0.401-0.528(**)
		0.532-0.598(**)
		0.610-0.749(**)
			NS	NS

14	0.366-0.452(**)
	0.497-0.563(**)
		0.487-0.575(**)
		0.594-0.659(**)
			0.108-0.149(*)
			0.153-0.220(*)
				NS

Note. * p < .05, ** p < .01, NS = not significant

Relationship Between Language and Music Processing: Evidence from Cross Linguistic Influence on Rhythmic and Melodic Perception

Mrudula Arunkumar[1]
Supervisor: Dr. Makiko Sadakata[1,2]

[1]Radboud University Nijmegen, Donders Institute for Brain, Cognition and Behaviour, The Netherlands
[2]Institute for Logic, Language and Computation, University of Amsterdam, The Netherlands

Previous studies on the relationship between language and music have looked at the influence of music on language and vice versa. The current study consisted of two experiments that observed the transfer of learning effect from language to music, specifically the influence on rhythmic and melodic perception. Both experiments used the Musical Ear Test (MET) to assess the rhythmic and melodic aptitude. Working memory and phonological memory tasks were administered to control for individual differences. The first experiment investigated the differences in rhythmic perception among English monolinguals and Finnish multilinguals and revealed that there was no significant difference in their rhythmic aptitude. This could be attributed to the monolingual nature of the English participants and the rhythmic properties of English and Finnish since they do not differ in metric preference. In the second experiment, Dutch speakers learning Chinese were recruited to compare their performances on the melodic aptitude test with Chinese-English bilinguals and Dutch-English bilinguals. Only Chinese-English bilinguals showed a significantly higher score on melodic aptitude task than the Chinese learners and Dutch-English bilinguals. This finding suggests that learning a tonal language does not provide sufficient sensitivity in pitch processing as seen among native tonal language speakers to yield a significant transfer effect from language to music. Results also showed that Chinese learners had no correlation between the language task and musical task similar to the Chinese-English bilinguals, indicating that they tend to split the processing of lexical tones from the musical pitch variations unlike the Dutch-English bilinguals who show correlation as they perceive the pitch input from both language and music tasks as general psychoacoustic information. This study adds supporting evidence to the existing literature on the transfer of learning effect and cross domain relationship between language and music.

Keywords: Language, music, transfer effect, rhythm, pitch.

Corresponding author: Mrudula Arunkumar; **E-mail:** mrudula.arunkumar@mpi.nl

"Two sides of the same coin" is a phrase that aptly fits the description of the relationship between language and music. Language and music are two defining features present in our daily life that are universal and present across all cultures and societies (Nettl, 2000; Williamson, 2009). On putting more thought into what constitutes language and music, we can see how both may seem different, especially in terms of their functionality, and yet they have various commonalities. This invites an interesting topic of research involving the study of two systems – language and music - that are also specific and relatable to human beings. It also brings views from an interdisciplinary standpoint ranging from linguistics, musicology, cognitive neuroscience, philosophy, and even evolution. Exploring the relationship between language and music opens doors to understanding how they are represented in the brain, whether they function in parallel or share overlapping brain regions, and whether expertise can be transferable between the two sound systems. This thesis aims to explore the transfer of learning effect between language and music, and especially focuses on the transfer of enhanced sensitivity of rhythm and pitch differences from language to music.

Both language and music are similar auditory inputs constituting of features like pitch, timbre, and rhythm that span across both (Williamson, 2009). The two sound systems contain patterns of sound which are put together to form meaningful phrases (Arbib, 2011). Auditory processing enables us to make sense of these different sounds that we hear and to associate them with meaningful relevant information (Kraus & Banai, 2007). Irrespective of whether the input is language or music, the first step in the processing of both language and music is the same, that involves combining the smaller units (musical notes or syllables) into meaningful larger units (like melodies or sentences) based on certain rules or syntax relevant for each system (Patel, 2008).

Auditory processing is dynamic and malleable and can change based on the exposure or experience gained, in particular sound systems (Kraus & Banai, 2007). With respect to language, one can see the impact of language experience on processing of sounds as early as infancy. Newborns show the ability to distinguish different phonemes across all languages but as months go by, their processing is tuned towards sounds from their native language (Kuhl et al., 2006). Studies performed among adults, for instance Mandarin speakers, show changes in neural circuitry in the cortical and subcortical areas indicating strong encoding of pitch content that arose due to their experience with tones (Krishnan, Xu, Gandaour & Cariani, 2005).

Similar to language, musical experience can also shape auditory processing as evidended by studies comparing musicians with non-musicians. Musicians show better sensitivity to incoming pitch information (Schön, Magne & Besson, 2004) and better responses towards artificial tones than non-musicians (Peretz & Zatorre, 2005). On a neural level, musicians show more robust encoding of pitch related information in the subcortical areas of the auditory pathway as shown by Wong, Skoe, Russo, Dees & Kraus (2007). Mussachia, Strait & Kraus (2008) also found that musical training enhanced the auditory memory in addition to shaping the pitch-specific encoding. There is also evidence showing that the N400 effect is seen both in music and language. Similar to language, tones and chords can elicit N400 effects similar to that seen in words (Koelsch, 2011).

All in all, it is evident that the language or musical experience can modify auditory processing which also shapes the brain and neural circuitry. Influence of experience is not only specific to auditory processing but can also be seen in cognitive domains like intelligence and executive control. Studies have shown that being a bilingual improves overall executive functioning (e.g., Bialystok, 2006) and similarly, being a musician also enhances executive control (Schellenberg, 2006). Looking at how language and music influence auditory processing and lead to cognitive benefits in their respective ways, brings the spotlight back to realizing the commonalities that exist between the two domains. Rhythm and pitch are two acoustic properties that exist in both language and music and by focusing on common features such as these, it is possible to study the relationship between language and music. To delve into this deeper, it is important to have a clear understanding about what constitutes rhythm and pitch in language and music.

Rhythm in language and music

Linguistic rhythm

In order to understand speech, segmentation of the speech input is necessary, which is done using the basis of linguistic rhythm (Cutler, 1994). Linguistic rhythm refers to the way language units are arranged periodically in time (Patel & Daniele, 2003) and can also rely on the position of lexical stress on the syllables in a word (Liberman, 1975). The two forms of linguistic rhythm are explained as follows.

Unit level classification

Pike (1945) proposed a classification of languages based on rhythm that depended on their stress and syllable patterns. This led to categorising languages as stress-timed and syllable-timed languages. Stress-timed languages are those that have roughly equal time intervals between stresses. English, German, and Dutch are few examples of stress-timed languages. In case of syllable-timed languages, syllables occur in a periodic fashion which are seen in languages like Turkish, Finnish, and French. There is also a third category that represents the periodic interval of morae in a language. Mora is a unit which is smaller than a syllable consisting of a consonant and a vowel or just a consonant or a vowel (Patel, 2008). Japanese is an example of languages that are mora-timed. Overall, the rhythmic classification of languages is based on the isochrony in speech, in this case, recurrence of a particular type of speech unit: stress, syllable, or mora (Low, 2006). Although studies have opposed the idea of isochrony (e.g., Roach 1982), the core principle behind this classification lies in how the rhythm is structured in the language which tends to be similar among certain languages leading to this categorisation (Dauer, 1983).

Metric foot preference

Another rhythmic aspect in a language is the metric foot that represents the rhythmic structure of a word based on stress patterns. The rhythmic pattern of stressed and unstressed syllables enables better comprehension (Jusczyk, Cutler & Redanz, 1993). The stress pattern of a word indicates the strength and dominance of the word. The stress could be placed either in the beginning of the word for the first syllable or at the end of the word. This leads to the classification of metric preference based on the position of stressed syllables: word-initial stress (Trochaic) and word-final stress (Iambic; Hayes, 1985). English, German and Finnish follow the trochaic metric preference whereas Turkish is an example of a language with iambic stress patterns.

Musical rhythm

Rhythm is also an important trait in music. Upon listening to certain songs, our instinct is to tap our feet or move in line with the rhythm of the song. Rhythm provides a temporal reference for the musical piece and characterizes the periodicity that is seen in music (Patel, 2008). Rythm is not only present by mean of beats, but also by means of grouping of various tones in particular patterns of phrases (Patel, 2003).

Although rhythm in language and music have distinct purposes, features like statistical patterning and rhythming grouping present in both language and music emphasize their commonality (Patel, 2003). This common property of rhythm opens doors to explore the possibilities of transfer of learning effect that can occur from language to music since rhythm is featured in both.

Pitch in language and music

Pitch is an important feature that is prevalent in language and music. In language, pitch plays a key role in pragmatics through prosody. Differences in pitch, in the form of intonation, convey different emotions or intentions that the speaker wishes to express. Apart from being a common feature in intonations, pitch also takes a lexical role in tonal languages. Pitch variations indicate difference in word meaning which are the traits of a tonal language. Mandarin Chinese is an example of one such language that has four distinct lexical tones (Duanmu, 2000). If a tonal speaker uses a different tone in place of a right tone, it can lead to a different semantic context and change the course of the conversation. Such tonal variations that indicate different word meanings are not present in non-tonal languages. Therefore, it is crucial for tonal speakers to understand and enunciate the pitch differences (in tones) correctly, as it has a major impact on communication.

As commonly seen in music, pitch takes a major role in defining a musical piece. The universality of how every music across cultures relies on pitch differences in notes from an octave makes pitch a defining feature of music (McDermott & Hauser, 2005). The scales, which are a set of pitches, encompass a musical melody (Bidelman, Gandour & Krishnan, 2011).

Similar to how rhythm is represented in language and music, pitch also tends to serve different functions for each domain but remains a shared feature between language and music. This makes it interesting to look at the possible effects of pitch exposure in one domain onto the other domain that may lead to a transfer of learning effect across the two domains of language and music.

Language & music – two sides of the same coin?

Going back to the claim that the relationship between language and music is like two sides of the same coin, the above information regarding rhythm and pitch in language and music provides support for this claim by showing that language and music share features like rhythm and pitch even though they serve different purposes. Neuroimaging studies have also shown support for shared processing between language and music. Certain areas in the brain like the Heschl's gyrus show activation for both words and musical tones (Binder et al.,1996). Primary auditory regions have also shown to respond in similar manner for both speech and music (Zatorre, Evans & Meyer, 1992). While reading musical notes and understanding symbols in a language, the supramarginal gyrus is involved (Falk, 2000). It is also interesting to note that similar to how Broca's region is responsible language production, it is also active during a music performance (Falk, 2000). In addition to this, there are numerous studies (e.g. Brown, Martinez & Parsons, 2006) showing overlapping brain regions for language and music processing. Koelsch (2000) stated music and language are two poles in the language-music continuum and experiences in music or language can have an impact on both.

Transfer effect

As it is evident that language and music share features both acoustically and cortically, there is a chance that expertise or experience in one domain can transfer to the other domain. This can happen due to the fact that the two domains share overlapping physical properties and neural features that can lead to enhancement in one domain due to the effect of the other. This can either occur from music to language, wherein musical abilities can improve language processing, or from language to music, in which language skills may help in performing musical tasks better. For this project, the transfer effect was studied based on the physical properties shared between language and music behaviourally without focusing on the neural and cortical properties.

Music to language transfer

Many studies showing the transfer of training effects from music to language have been reported. These include evidence that show an enhancement in second language learning especially of the phonetic structure, as well as an enhancement in phonological processing due to musical abilities (Anvari, Trainor, Woodside & Levy, 2002; Slevc & Miyake, 2006). Musicians also tend to have higher sensitivity to prosodic cues in language (Thompson et al. 2004) and better perception of metric structure of words (Marie, Magne & Besson, 2011).

These results show that there are benefits in language-related skills due to the expertise in music. While comparing the influence of musicality on lexical tone identification, musicians showed superior performance in identifying lexical tones accurately (Chua & Brunt, 2015). In fact, musicians with no experience of a tonal language performed as good as those who were experts in a tonal language, like Mandarin Chinese speakers (Delogu, Lampi & Belardinelli, 2010). These transfer effects could possibly stem from musical training providing an overall enhancement in sensory perception and cognitive mechanism that operate on different levels enriching the auditory processing that is seen in language-related skills (Bidelman, Hutka & Moreno, 2013).

Language to music transfer

As both language and music operate similarly, even with shared networks on the cortical level (Patel, 2011), the transfer effect could also occur in the other direction: from language to music. Tonal language speakers provide an interesting group to study with respect to transfer of learning effect as they are exposed to pitch differences in the form of lexical tones. Bidelman, Hutka & Moreno (2013) studied this bidirectionality by comparing musicians with Cantonese speakers and English speakers on their performance on auditory acuity, music perception, and general cognitive abilities. Apart from musicians, who performed better in all tasks, Cantonese speakers also showed better performance in music perception than English speakers indicating the effect of their tonal background on musical perception. Bilinguals have also be shown to have benefits in the musical domain as bilinguals, specifically second language learners of English outperformed monolinguals in the melodic and rhythmic aptitude task (Roncaglia-Denissen, Roor, Chen & Sadakata, 2016). By comparing bilingual speakers of languages that have different rhythmic classification, such as syllable-timed and stress-timed languages, Roncaglia-Denissen, Schmidt-Kassow, Heine, Vuust & Kotz (2013) found that speaking two languages with distinct rhythmic features helped in perceiving musical rhythms more than speaking languages that shared similar rhythmic features. This was argued to be due to the auditory enhancement that the rhythmic variation provides which is transferred to the musical domain where rhythm plays a role.

Present Study

Although there is evidence showing language to music transfer, the literature is still scarce compared to the numerous studies done on music to language transfer. This study focusses on the transfer of learning effect from language to music, specifically in the ability to distinguish rhythmic and pitch differences. As mentioned earlier, speaking languages that are rhythmically diverse or speaking a tonal language can have an influence on the musical pitch and musical rhythmic perception (Chen, Lui, & Kager, 2016; Roncaglia-Denissen et al., 2013; 2016). However, groups with various types of language background differing in rhythmic or tonal exposure need to be studied and compared to explore the transfer effect further. For example, studying languages that are different in unit level rhythmic classification but share rhythmic feature of metric preference might show us whether enhanced musical rhythmic perception is present similar to the effects seen by studying languages that are rhythmically different in metric preference and unit level classification in Roncaglia Denissen et al. (2013). With respect to enhanced pitch perception, studies so far have shown the effects of being a native tonal speaker on increased melodic skills (Chen, Lui, & Kager, 2016). But it is yet to be investigated whether adult learners of a tonal language can also contribute to enhanced melodic pitch perception like native tonal speakers. In this study, two independent experiments were conducted to analyse the influence of linguistic background on rhythmic and melodic perception. Experiment 1 aimed to determine whether there is a difference in rhythmic perception between English monolinguals and Finnish multilinguals that have different rhythmic properties at the unit level but share metric preference, whereas Experiment 2 aimed to find if learners of a tonal language, namely Mandarin Chinese, show enhanced melodic perception like Chinese-English bilinguals. Further details on the background, motivation, along with the methods and results will be discussed individually for each experiment in the following sections.

EXPERIMENT 1 – RHYTHMIC PERCEPTION

Background

The aim of this experiment is to observe whether there is an enhanced rhythmic perception among multilinguals speaking languages with varied characteristics compared to monolinguals. This is done by comparing the rhythmic aptitude of the two language groups: English monolinguals and Finnish multilinguals by controlling for factors such as musical experience, working memory, and phonological ability. Previous studies have found that mastering languages with different rhythmic features leads to enhancement in musical rhythmic perception which supports the transfer of learning effect (Roncaglia-Denissen et al., 2013; 2016). The diversity in rhythmic features of the languages that were studied corresponds to rhythmic classification based on unit level and metric preference. In the past, the language groups that were compared were either both stress-timed languages with same metric preference (Dutch-English, German-English), or different in unit level, one being syllable-timed language and one being a stress-timed language with different metric preference (Turkish-English, Turkish-German). On comparing the rhythmic perception of Turkish–German learners and German-English learners, the former had better rhythmic aptitude than the latter. The authors attributed this to the factor that Turkish-German learners were more varied in their rhythmic background than German-English learners as Turkish and German have different rhythmic properties in both unit level and metric preference compared to German and English which are both stress-timed languages having the same metric preference. Hence being sensitive to varied rhythmic cues from the language learnt can help perceiving musical rhythms better. The same trend was also seen while comparing Turkish-English participants with Dutch-English participants owing to the expansive rhythmic exposure of Turkish-English that was lacking in Dutch-English group. In contrast to previous studies, this experiment studies the influence of rhythmic variation by comparing languages that are syllable-timed (Finnish) and stress-timed (English) but share the metric preference of trochee which is word-initial stress. As previous studies focused on a more general view on rhythmic variability in terms of differences in both unit level classification and metric preference, studying English and Finnish enables us to disentangle the factor of metric preference in order to check whether that is crucial for rhythmic transfer to take place. English was chosen as the monolingual group since previous studies only looked at Dutch monolinguals and Turkish monolinguals. Finnish was chosen due to the feasibility of approaching participants and it also fit with the target language that needed to be studied. In addition to studying rhythmically diverse languages, Roncaglia-Denissen et al. (2016) also found that Dutch, Turkish, and Chinese learners of English performed better than Turkish monolinguals in both melodic and rhythmic perception indicating that bilingual exposure might improve overall musical perception. As Turkish monolinguals constituted the only monolingual group that was investigated in previous studies, adding the English monolingual dataset from this experiment will prove useful to further substantiate the bilingual advantage. Hence comparing English monolinguals with Finnish native speakers who were multilinguals provided a platform to explore whether speaking more than one language enhances overall musical perception as shown in Roncaglia-Denissen et al. (2016). Since previous studies have shown enhancement caused by bilingualism, we expected that Finnish multilinguals will have higher rhythmic aptitude than the English monolinguals. Also, due to the fact that English and Finnish share the same trochaic metric preference (Jusczyk et al., 1993; Livonen & Harnud, 2005) but differ in terms of being stress-timed and syllable-timed languages, the Finnish multilingual group having more exposure is expected to show better performances in musical rhythmic aptitude than English monolinguals. In case they show no differences then it would provide a newer insight about the influence of having the same metric preference on rhythmic perception as that was not covered in previous studies.

Method

Participants

The study comprised of 15 English monolinguals (N = 15, female = 11, Mean Age = 22.81 years) and 15 Finnish multilinguals (N = 15, female = 12, Mean Age =24.6 years). They were recruited from Nijmegen, Amsterdam, Ghent, and Brussels. Most of them were university students. English monolingual participants were from English-speaking countries, namely the United Kingdom (N = 8), the United States of America (N = 4), the Caribbean (N = 2) and Indonesia (N = 1) who were studying at the Radboud University in Nijmegen or the University of Amsterdam. Finnish multilingual participants were native Finnish speakers whose second language was English and additionally learnt mostly Swedish, German, or Dutch. This sample size was chosen due to the challenging nature of finding English monolinguals in a non-English-speaking country.

Participants from the English and Finnish groups had an average of 2.7 years and 4.1 years of musical experience respectively. All the participants had normal or corrected-to-normal vision and did not have any neurological impairment, epilepsy, hearing or visual impairments. Upon giving detailed instructions about the experiment, written consent was obtained from the participants for the purpose of data collection and publication use. They were provided with a monetary compensation of 10 Euros.

This study was approved by the Ethics Committee of the Faculty of Humanities of the University of Amsterdam and Faculty of Social Sciences of the Radboud University in Nijmegen.

Materials

The following tests were administered to the participants.

Musical Ear Test

The musical ear test (MET) by Wallentin Nielson, Fris-Olivarius, Vuust & Vuust (2010) was designed to measure the musical abilities of musicians and non-musicians in a relatively shorter duration. This test was used in our study in order to assess the participants' musical aptitude. The test consists of two parts: Melody and Rhythm, and the participants had to judge whether the two melodic or rhythmic phrases were similar or different. Each part comprised of 52 trials, making it a total of 104 trials with 2 practice questions for each part. The melodic part was represented by short piano phrases that ranged from 3 to 8 tones. They were presented in pairs, where the melodies had a duration of one measure played at the speed of 100 bpm. Half of the 52 trials were "same" and the other half comprised of "different" trials, but the order in which they were presented, was randomized. The 26 trials that were different were characterized by a pitch violation among which, 13 of them had a pitch violation along with a contour violation.

The rhythmic part was characterised by rhythmical phrases or beats that were played with a wood block. This subtest also consisted of 52 trials where 26 of them were "same" and 26 were "different" with respect to one rhythmic change. Thirty-seven of the 52 trials began on the downbeat and rest of the trials started later. They also varied in rhythmic complexity, however the order of the trials was randomized.

The entire test took about 18 minutes to complete and didF not provide feedback at the end of the test. Participants were advised to wear headphones during the test and asked to answer as quick as possible and intuitively in case of any difficult trials. The participants listened to the trials (melody and rhythm) and were asked to judge whether the pair comprised of same melodies/rhythms or different melodies/rhythms by clicking the "same" or "different" option.

Chinese Tone discrimination task

The Chinese Tone discrimination task was administered to all the groups of participants irrespective of their language background. This task was used by Chen, Lui, & Kager (2016) to study the differences between the processing of lexical tone by asking the participants to discriminate between monosyllabic and bisyllabic pairs of Chinese tones. This test consisted of two parts: Monosyllabic (MT) and Bisyllabic (BT) discrimination. The stimuli consisted of monosyllables such as: /ba/, /bwɔ/, /bi/, /da/, /dwɔ/, /di/, /la/, /lwɔ/, /li/, /ma/, /mwɔ/, /mi/, /na/, /nwɔ/, /ni/, which were recorded separately by a female native speaker. Every syllable was recorded with all the possible tones: high level (T1), rising (T2), low-dipping (T3), and high-falling (T4). For MT, the participants had to distinguish between the tonal pairs of T1 and T4 and between T2 and T3. For each tonal pair all possible combinations were presented, for example, T1-T1, T1-T4, T4-T1, T4-T4. This part consisted of 120 trials. For BT, the stimuli consisted of two syllables followed by another set of two syllables which would either be the same as the first set or different. The possible combinations included T3T3–T2T3, T3T3–T3T2, T3T3–T2T2, T4T4–T1T4, T4T4–T4T1, and T4T4–T1T1. This part had 180 trials and the participants had to click "same" or "different" based on what they heard. The experiment was designed in such a way that the trials progress quickly, leaving only a second for the participant to answer, after which the next trial is presented automatically.

Phonological and Working memory measures

Phonological memory has been shown to be important for the processing of novel sounds and word learning (Baddeley, Gathercole, & Papagno, 1998). This is relevant to our study as it involves the ability to retain words and working memory efficiency which tends to vary between monolinguals and multilinguals (e.g., Bialystok et.al., 2004). In order to study the phonological memory capacity of an individual, the Mottier test (Mottier, 1951) was conducted. The Mottier test consists of pseudowords (words that have no meaning) which were recorded by a native English and a native Finnish speaker and was administered to the participant groups accordingly. The pseudowords began with two syllables and after each set, which contained 6 words, another syllable was added, thus increasing the length of the pseudowords. Maximum of six sets were used which meant that the maximum length of the pseudowords presented was 7 syllables. After listening to each word, the participants were asked to repeat the word.

To assess the participants' working memory, the Backward Digit Span (BDS) task was used. In this experiment, the test consisted of 14 sets of two trials that begin with a series of two numbers. There is an increase of one number after every set. Similar to the Mottier test, the stimuli for BDS were recorded by a native speaker of each group (English and Finnish) which were then administered to the respective group.

Self-Reported Language Questionnaire

Before the experiment, a language background questionnaire was given to the participants. This extensive language questionnaire contained information that tapped into their proficiency level and experience for the languages they know (Marian, Blumenfeld, & Kaushanskaya, 2007). This was done in order to assess their language skills, giving us a better picture of their language background. It was especially important to administer this test to verify the monolingual nature of English participants. The questionnaire included questions about their age of acquisition, how long they have been learning, along with rating scales about their reading, speaking, writing, and listening skills in each language. The questionnaires for the participants were administered through an online link that was created using Qualtrics.

Musical Background Questionnaire

It was important to assess the participants' musical background as it could be a potential factor that can contribute to their performance in the MET (Wallentin, et al., 2010). This was done with the help of Goldsmith Musical Sophistication Index (Müllensiefen, Gingras, Stewart & Musil, 2014), specifically using the subsets Perceptual Ability & Musical training. It included questions such as their ability to perceive an out of tune/beat of a song, how long they have been learning, etc. This test was also administered online through Qualtrics.

Procedure

The Chinese Tone Discrimination task, MET, Mottier test, and BDS were conducted using a laptop, inside a quiet room on the day of experiment. The Self-reported language questionnaires and musical background questionnaires were sent to the participants through an online link prior to the day of experiment and they were advised to complete it before participating. The entire experiment lasted for an hour. For the listening tasks, i.e., MET and Chinese tone discrimination, the participants used over-the-ear headphones while for the repetition tasks, i.e., Mottier test and BDS, they listened to the stimuli from the laptop speaker and repeated aloud while the experimenter scored their responses.

Mottier Test

The audio files with the pseudo-words were played one by one from the laptop following which the participant repeated them. It was scored simultaneously on the respective Answer Sheet. The test was terminated if the participant failed to recall more than four pseudo words from a set. The scores were calculated based on the number of correctly repeated pseudowords with a maximum score of 36.

Backward Digit Span

This follows a similar set up as the Mottier test, in which the audio is played from the laptop and the participant repeats the series of numbers that were heard, in the reverse order. The test was terminated when two consecutive errors were made while repeating backwards, irrespective of the length of the series. The total number of correct trials was counted as the score obtained by the participant, with the maximum of 14 as the total score.

Musical Ear Test

The melodies and rhythms for the MET were presented on the laptop through an online link using Qualtrics. The participants played each of the trials and clicked "same" or "different" based on their response. The accuracy percentage of each part was calculated by dividing the correct number of trials by 52 for both melody and rhythm.

Chinese tone Discrimination task

This task was programmed using ZEP (Veenker, 2017) and was opened on the laptop, separately for monosyllabic and bisyllabic discrimination. BT always preceded the MT. The reason to follow this particular order is due to the fact that non-tonal speakers were more accurate in the MT than the BT. Hence, in situations where MT preceded BT, a possible learning effect could occur that might influence the performance in discrimination disyllables (Chen et. al, 2016). However, it is unlikely that BT could lead to this effect as accuracy in MT was already quite high and hence BT is administered first following which MT is administered. During both these tasks the participants were asked to click the "same" or "different" option that appeared on the screen after the presentation of the stimuli. The experiment proceeded quickly and the participant had only a second to answer the question after which it automatically proceeded to the next question. For every correct answer, incorrect answer and a skipped question the scores were coded as 1, 0, and -1 respectively. The accuracy percentage was calculated based on the number of correct trials divided by the total number of trials.

Statistical Analysis

The musical background of the participants was compared between the groups to identify group differences using Mann-Whitney U tests. This was done to avoid any confounding factor of musical experience which could interfere with their performance in the MET (Wallentin et al., 2010). The scores from the language background questionnaires were used to check whether monolingual participants learnt any other languages and the scores revealed that none of the participants showed no formal learning of any other language. However, the limitation here is the fact that an exposure to other languages cannot be controlled since they were living in the Netherlands.

Additionally, we also checked for group differences in the working memory and phonological memory measures by comparing the groups' scores on BDS and Mottier test using a Mann-Whitney U test. Analysis of Covariances (ANCOVA) was used to compute the group differences in rhythmic perception. Their mean scores percentage in the MET-Rhythm subtest was entered as the dependent variable and the language groups were the between-subjects factor. The covariates that were considered for this analysis were the participants' scores of BDS and Mottier test as well as their performance in the MET-Melody subtest. These were the covariates that were used in previous studies as well (Roncaglia-Denissen et al., 2013; 2016) as it is known that working memory and the other subtest of MET have an influence in the MET performance (Wallentin et al., 2010).

The participants' performances in the Chinese Tone discrimination tasks, specifically the number of missed trials, were also assessed using independent *t*-tests in order to see differences in terms of how quickly the participants responded that might relate to the executive control advantage seen in multilinguals. The language groups were entered as between-subjects factor and the number of missed trials of MT and BT tasks and their accuracy percentage in both the tasks as dependent variables.

Experiment 1 Results

Musical Background

Mann-Whitney U test was performed to compare differences in the musical background of the English and Finnish groups. Their perceptual abilities, musical training, and number of years of formal training were entered as dependent variables keeping the language groups as the between-subject factor. With respect to the musical background, no significant differences were present between the groups' musical perceptual abilities (*U* = 90.5, *p* = .367), training (*U* = 107.5, *p* = .838), and years of formal training (*U* = 143, *p* = .217). Their mean scores in the perceptual ability and musical training questionnaires from the MSI along with the mean of number of years of training are shown in Table 1.

Phonological and working memory measures

The results exhibited a significant difference between the groups in their Mottier Test performances, *U* = 179, *p* <.05, indicating that the Finnish Multilingual group (*M* = 28.8, *SD* = 3.1) outperformed the English Monolingual group (*M* = 24.8, *SD* = 3.6). However, there was no significant difference in the BDS scores, *U* = 87.5, *p* = .305 between the Finnish multilinguals (*M* = 7.4, *SD* = 2.2) and English monolinguals (*M* = 8.4, *SD* = 2.5).

Rhythmic Aptitude

To compare the rhythmic aptitude between the two groups, ANCOVA was used, and the scores on BDS, Mottier Test and other subset of MET (MET-Melody) were entered as covariates. No significant group difference in their MET-Rhythm scores, *F*(1,28) = 3.415, *p* = .076 (r^2 = .229) was found. This shows that the Finnish multilinguals did not significantly differ in their rhythmic perception from English monolinguals. However, as seen in Figure 1, Finnish multilinguals showed a higher percentage in their accuracy of MET-Rhythm test than the English monolinguals but this was not significant.

Chinese Tone Discrimination Task

On comparing the participants' discrimination accuracy in the Chinese Tone discrimination task, no group differences were found, for both MT and BT. Since the task was designed in such a way that the participants had only one second to answer, their ability to respond on time was assessed by comparing the number of missed trials between the groups. For MT, the missed trial count between groups did not differ significantly, *t*(28) = 0.440, *n.s.* and for BT, similar results were seen, showing no significant difference, *t*(28) = 0.090, *n.s.* The mean values of the number of missed trials per group are shown in Table 2.

Table 1.
Mean of the number of years of training and Musical Sophistication Index (MSI) subscales: Perceptual Ability, Musical training.

	English Monolinguals		Finnish Multilinguals	
Variables	Mean	SD	Mean	SD
Number of years of musical training	2.2	2.5	4.1	3.8
MSI-Perceptual Abilities	46.8	6.5	44.06	8.11
MSI-Musical Training	21.7	7.7	21.8	9.1

Table 2.
Mean and Standard Deviation of Number of missed trials in Chinese Tone Discrimination tasks.

English Monolinguals			Finnish Multilinguals	
	Mean	SD	Mean	SD
Missed Trials in MT[1]	16.16	19.45	10.33	8.3
Missed Trials in BT[2]	16.28	12.63	15.90	11.09

[1] Monosyllabic Discrimination
[2] Bisyllabic Discrimination

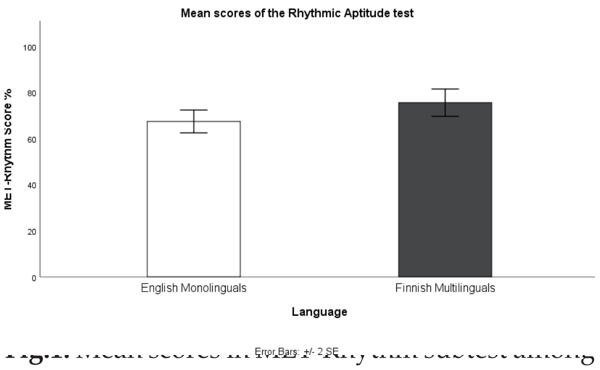


Fig.1. Mean scores in MET-Rhythm subtest among English Monolinguals and Finnish Multilinguals showing the error bars that indicate standard error.

Discussion

This experiment aimed to explore the enhancing effects of being a multilingual on musical rhythmic perception. Additionally, we also looked at whether speaking languages that are different on unit level but share the same metric preference can contribute to this effect. English monolinguals and Finnish multilinguals represented a language grouping that enabled us to observe the effects of multilingualism. With respect to rhythmic diversity, English is a stress-timed language and Finnish is a syllable-timed language and both share the metric preference for trochee, which was in contrast to the languages grouped in previous studies that differed in both unit level classification and metric preference (e.g., Roncaglia-Denissen et al., 2013; 2016). The groups were administered with a musical aptitude test using MET, cognitive tests on working memory and phonological memory and a Chinese tone discrimination task. To answer our research question whether Finnish multilinguals are better in musical rhythmic perception than English monolinguals, we compared MET-Rhythm scores of the two groups, controlling for the cognitive measures and MET-Melody score.

The results showed that Finnish multilinguals did not show a significantly higher performance in the rhythmic perception task compared to the English monolinguals. Although the average of the rhythmic aptitude score was higher in Finnish multilinguals the difference was not significant. This is contrary to the work done earlier (Roncaglia-Denissen et al., 2016) which showed that second language learners performed better than monolinguals. In terms of rhythmic variability, the differences in Finnish, and English, being syllable-timed and stress-timed languages respectively, did not account for an enhancement in musical rhythmic perception. This result also does not conform with the findings from Roncaglia-Denissen et al. (2013) which showed that Turkish-German learners performed better than German-English learners in the rhythmic aptitude test. However, it is to be noted that the Turkish and German have differences in rhythmic properties both at the unit level and metric preference as they are syllable-timed and stress-timed languages respectively, with preference for the iambic foot for Turkish and trochaic foot for German. This is not the case for Finnish, and English that were studied in this experiment as they are syllable-timed and stress-timed languages respectively but shared the metric preference of trochee. This could mean that, it is important for the languages to differ not only in unit level classification (syllable-timed or stress-timed), but also in word level (metric preference of trochaic or iambic) to be able to see a significant enhancement in musical rhythmic perception. As the Finnish multilinguals were not exposed to rhythmic diversity in terms of the metric preference, their sensitivity to rhythmic differences might not have been as high as individuals who are exposed to languages having more variations in their rhythmic properties of unit level classification and metric preference, as seen in Turkish-German learners. This could possibly explain why the transfer of learning effect from language to music was not evident. However, Arvaniti & Ross (2011) have mentioned that this type of rhythmic classification cannot be reliable or translated based on listeners' perception so it is questionable to use this mode of classification. Being the first study looking at rhythmic transfer from the perspective of sharing a metric preference, no strong claim regarding the effects can be made without a replication or a follow-up study. It would be useful to analyse the language properties further to see what features could be teased apart in order to see a significant transfer effect. For example, as this experiment looked at languages that differed in unit level rhythmic classification having shared metrical preference, a next step could probably move in the direction of comparing languages that have same unit level rhythmic classification but differ in metric preference. For instance, participants who speak both Turkish and Spanish could be recruited and compared with Turkish-German speakers. Since both Spanish and Turkish are syllable-timed languages they differ in metrical preference as Turkish is iambic, preferring word final stress (Inkelas & Orgun, 2003) while Spanish is trochaic (Schmidt-Kassow et al., 2011). By comparing these two language pairs, more insight on the influence of rhythmic variation could be established.

In addition to rhythmic perception, the groups were also compared with their performance on the Chinese tone discrimination task. We expected to see similar trends of bilingual advantage in discrimination accuracy as shown in Chen et al., (in prep) but found that the results of English monolinguals were on par with Finnish multilinguals in discrimination accuracy and number of missed trials. This failed to show a multilingual advantage which made us rethink about the true monolingual nature of the English participants. The English monolinguals that participated in this experiment were exchange or graduate students studying in the Netherlands who also differed in origin of English-speaking countries. Recruiting English monolinguals from the same country living in an English speaking environment like England might prove to be a better group for comparison with multilinguals. However, in studies by Roncaglia-Denissen et al. (2016) and Chen et al. (in prep), only Turkish monolinguals were used to compare the rhythmic perception and Chinese tone discrimination accuracy with second language learners, and the monolingual group had lower scores in all the tasks. Future studies can broaden the dataset by studying other monolinguals to further validate the presence of a bilingual advantage in musical rhythmic perception.

EXPERIMENT 2 – PITCH PERCEPTION

Background

While looking at effects of music on language perception, it has been found that musical training enhances mandarin tone perception among non-tonal speakers (e.g., Hung & Lee, 2008; Mok & Zuo, 2012). The auditory brainstem response is said to be domain general since the brainstem shows activation for pitch processing in both music and language (Bidelman, Hutka & Moreno, 2013). This could mean that possible effects can be seen while observing effects of language skills on musical perception. Studies based on the transfer of learning from language to music have found that tonal language speakers show better perceptual discrimination of musical pitch (e.g., Pfordresher & Brown, 2009). Roncaglia-Denissen et al. (2016) also studied the transfer effect focusing on pitch perception, and found that Chinese-English bilinguals show better performance in melodic aptitude compared to Dutch-English bilinguals and Turkish Monolinguals. Chen et al. (2016) also found that Chinese native listeners outperformed the Dutch native listeners in musical tasks. They also looked at the correlation of the performances of both the musical aptitude test and the lexical tone discrimination task for the Dutch learners and Chinese listeners in order to explore the relationship in pitch processing in language and music. No correlation between language and music tasks existed among Chinese-English bilinguals, but Dutch-English bilinguals showed correlation between the musical task and the lexical tone discrimination task. This difference led the authors to propose the "split hypothesis": Native tone language listeners tend to split the input of lexical tones from other types of pitch variation, in this case, musical pitch. Although cross domain benefits, like better melodic pitch perception among tonal language speakers, are evident between the domains of language and music, it is important to note the acoustic pitch input that tonal language speakers receive, is contextually different from nontonal language speakers as the former has tonal exposure with pitch playing a lexical role. This could explain why a correlation between musical and language tasks existed only for nontonal speakers and not Mandarin Chinese speakers as nontonal speakers perceive the input from a general psychoacoustic perspective without any contextual differences.

Experiment 2 aimed to look at the effect of learning a tonal language on the melodic perception by comparing the performance of adult tonal language learners in melodic aptitude tasks with that of tonal and nontonal speakers. For this purpose, native Dutch speakers who are learning Mandarin Chinese were recruited and their performance in melodic aptitude tests were compared with those of Chinese-English bilinguals and Dutch-English bilinguals, representing a native tonal group and nontonal group respectively. The learners of Chinese were further divided into two categories: Beginners and Advanced, based on the number of weeks of learning. The performance of both groups of Chinese learners in melodic aptitude test were compared with Chinese-English bilinguals and Dutch-English bilinguals to check whether the enhanced musical pitch perception is prominent among learners of Mandarin Chinese.

Furthermore, this experiment also aimed to explore the relationship between the melodic tasks and the lexical tone discrimination tasks to see whether the split hypothesis holds for learners of a tonal language. Since the split hypothesis proposes that individuals with tonal exposure tend to split the perception of lexical tones from musical pitch variations, it would be interesting to see whether Chinese learners show similar trend like the Chinese-English participants from Chen et al. (2016). Therefore, we hypothesized that advanced learners especially, will not show any correlation between the performance of lexical task and musical task due to their exposure to tone language that enables them to split the perception of lexical tones from musical pitch variation.

Methods

Participants

The participants in Experiment 2 consisted of 27 Dutch learners of Mandarin Chinese, who were categorised into Beginner Learners (N=14, Mean Age = 22.07 years) and Advanced learners (N = 13, Mean Age = 23.31 years) based on the number of weeks spent in learning Mandarin Chinese. They were students who were enrolled in Chinese language learning courses from Radboud in'to Languages, Lischerijn college in Utrecht and the programme "China studies" from Universiteit Leiden. Only those who had less than 3 years or no musical experience were recruited.

This experiment also consisted of data of 15 Dutch-English Bilinguals (N = 15, 8 females, Mean Age = 25.53 years) from the study by Roncaglia-Denissen et al. (2016) and 15 Chinese-English bilinguals (N = 15, Mean age = 25.13 years) from Chen et al. (2016) that were studied along with second language learners of Mandarin Chinese.

All the participants had normal or corrected-to-normal vision and did not have any neurological impairment, epilepsy, hearing and visual impairments. Upon giving detailed instructions about the experiment, written consent was obtained from the participants for the purpose of data collection and publication use. They were provided with a monetary compensation of 10 Euros.

This experiment was approved by the Ethics Committee of the Faculty of Humanities of the University of Amsterdam, Faculty of Social Sciences of the Radboud University in Nijmegen and by the Ethics Committee Psychology (CEP) of Leiden University.

Materials

The same tasks: MET, Chinese Tone discrimination task, Mottier test and BDS that were used in Experiment 1 (see Materials under Experiment 1) were used for this Experiment. The differences between the materials used in Experiment 1 and Experiment 2 were in the recorded Stimuli for Mottier test and BDS as they were recorded by a Native Dutch speaker since the participants in this Experiment were native Dutch speakers.

For this experiment, the language and musical background questionnaires were combined into a single questionnaire that focused more on the participants' Chinese learning history unlike the extensive language and musical background questionnaires that were used for Experiment 1. It included questions related to their proficiency in reading, speaking, writing and listening skills in Chinese as well as their first and second language and number of years of musical training.

Procedure

The procedure of administration and scoring of BDS, Mottier Test and Chinese Tone discrimination was identical to Experiment 1. With respect to MET, the test was administered through PsychoPy in place of Qualtrics. In this case, the stimuli automatically played following a fixation cross after which "Same" or "Different" appeared on the screen. Depending on their response the participants pressed the left-arrow key for "Same" and right-arrow key for

"Different".

Statistical Analysis

The Mandarin Chinese proficiency among Beginners and Advanced Learners was analysed by observing the participants' competency in Mandarin Chinese skills such as understanding, reading, writing and speaking. This is useful to show a clear distinction between the beginners and advanced learners. To compare their skills, Mann-Whitney U tests were performed by using each of the skills as dependent variables and the Chinese Learner groups (Beginners, Advanced Learners) as between-subjects factor.

The scores of phonological and working memory measures were also compared between all the four groups: Dutch-English, Chinese-English, Beginners and Advanced learners of Mandarin Chinese. This was done with the help of two Analyses of Variance (ANOVAs), one having Mottier scores as a dependent variable and the language groups as a between-subjects factor, and the other used BDS score as a dependent variable with the language groups as between-subjects factor. It is important to note that for this part of the analysis the Dutch-English bilingual data were obtained from an existing data set (from Roncaglia-Denissen et al., 2016).

Moving towards the aim of study we focused on the group differences in the performances in the melodic aptitude test. For this, ANCOVA was used, having the MET-Melody score percentage as the dependent variable while the scores on Mottier test, BDS and their accuracy percentage in the MET-Rhythm subtest were entered as covariates. The language groups were the between-subjects factor. This was also followed from previous studies since working memory measures and MET subtests influence the MET performance (Wallentin et al., 2010)

To analyse the cross-domain correlation between melodic perception and Chinese tone discrimination, Pearson's product moment correlation was used. This was done across groups for both MT and BT tasks. In this section of analysis, the Dutch-English data were from a different study by Chen, Liu & Kager (2016) as the data of Chinese tone discrimination task of Dutch-English group (from Roncaglia-Denissen et al., 2016) were not available[1].

Experiment 2 Results

Mandarin Chinese proficiency

The Mandarin Language skills – understanding, speaking, reading and writing were compared among the beginners and advanced learners of Chinese. For "understanding" skills the two groups differed significantly, $U = 166, p < .001$.

Significant differences were also found in speaking skills, $U = 143$, $p < .05$. Reading ($U = 157.5$, $p < .001$) and Writing skills ($U = 132.5$, $p < .05$) also showed significant differences between the beginners and advanced learners of Chinese.

The mean value of the total number of weeks for which the beginners and advanced learners spent time in learning Mandarin Chinese were also computed. Beginners spent an average of 33 weeks in learning Chinese whereas the advanced learners spent an average of 220.77 weeks in learning Chinese. The categorisation of the Chinese learners were based on the number of weeks spent in learning and seeing significant differences in Mandarin Chinese skills assessing the proficiency validates the usage of criteria of weeks as method of splitting the group.

Phonological & working memory measures

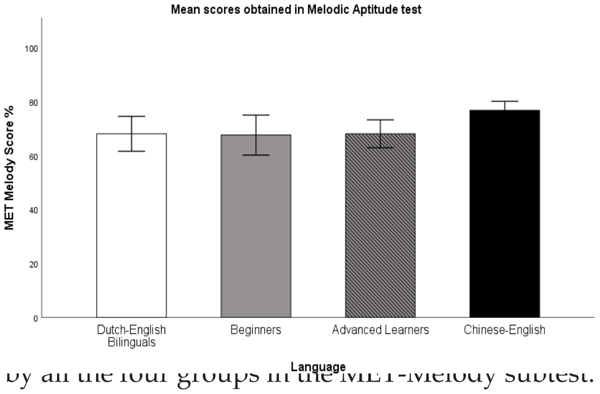
The ANOVA that was carried out to assess differences in working memory performance using BDS scores, revealed significant group differences, $F(3,44) = 3.222$, $p < .05$. On performing post-hoc analysis using Bonferroni corrections, significant differences were present between Dutch-English Bilinguals ($M = 8$, $SD = 2.2$) and Chinese-English Bilinguals ($M = 11.11$, $SD = 3.3$). Other between-group differences including Beginners and Advanced learners of Chinese were not significant. While comparing the Mottier Test scores, no significant differences were found between groups.

Melodic Aptitude

On performing ANCOVA with MET-melody score as dependent variable, the results showed that there was no significant difference between the groups regarding their melodic perception. However, while performing a planned comparisons of the

[1] The data of MET scores for both the sets of Dutch-English data were compared in order to check for group differences. No significant differences were found for MET Melody ($U = 104.5, p =.744$) and MET-Rhythm ($U = 126, p =.595$). As the groups do not differ in their performances in MET, using the different data set would not have influenced the outcome of analysis.

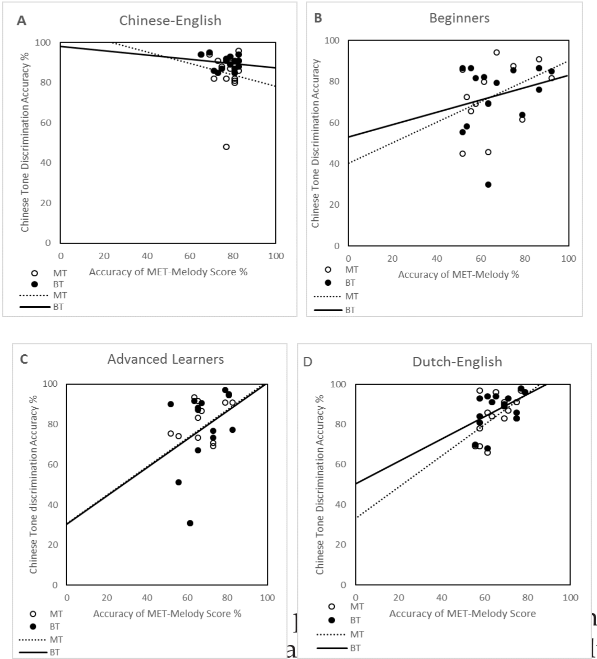
means of the MET Melody scores, the Chinese-English participants had significantly higher mean score (*M* = 76.66, *SD* = 6.24) compared to Dutch-English bilinguals (*M* = 68.06, *SD* =12.5), Beginners (*M* = 67.5, *SD*=13.8) and Advanced learners of Chinese (*M* = 68.04, *SD* = 9.30) as seen in Figure 2.



by all the four groups in the MET-Melody subtest. Chinese English group had the highest mean which differs significantly while comparing it with other groups.

Chinese Tone vs Melodic Pitch Discrimination

Pearson's product moment correlation was used to observe the relationship between the MT and BT performance and the MET-Melody scores for each group. For the Dutch-English bilingual group,



Scores % for the four groups. A. Denotes the correlation analysis in Dutch-English Bilinguals in which MT and MET-Melody scores are significantly correlated (r = .581) and there is also a positive correlation for BT and MET-Melody but they are not significantly correlated (r = .488). B. Correlation plot for Beginner learners of Mandarin Chinese which show a less positively correlated trend which is not significant. C. Advanced Learners also show no significant correlation for both BT and MT (Table 3. Illustrates more details on the relationship between the variables). D. Chinese English participants also show no significant correlation between Chinese tone discrimination task and MET-Melody Scores.

Table 3.
Correlation between MET-Melody scores and discrimination accuracy for Monosyllabic Discrimination (MT) and Bisyllabic Discrimination (BT)

		Dutch-English		Beginner Learners		Advanced Learners		Chinese-English	
		MT	BT	MT	BT	MT	BT	MT	BT
MET-Melody	Correlation Coefficient	.581*	.488(*)	.441	.252	.384[a]	.340[a]	-.133	-.166
	p-value	.023	.065	.115	.385	.195	.256	.637	.554

*Significant at p < .05.

[a] Among the advanced learners there was one participant as an outlier with the mean Accuracy of approximately 30% for both MT and BT, by removing this outlier the correlation coefficient for advanced learners is r = .182 (for MT) and r = .090 (for BT).

there was a significant correlation between the MET-Melody score and the MT discrimination

accuracy, $r = .581$, $p < .05$. Although there was a similar type of correlation for BT, it was not statistically significant, $r = .488$, $p = .065$ (Fig. 3A). There was no significant correlation between MET-Melody and BT and MT tasks for the other three groups – Chinese-English Bilinguals, Beginners and Advanced Learners of Chinese. There was also a trend in the degree of correlation between MET and MT, BT discrimination accuracy between the four groups as the correlation coefficients decreased from Dutch-English bilinguals to beginners and advanced learners and this showing no correlation among Chinese-English bilinguals (see Table 3. For correlational plot, see Fig. 3).

Discussion

The aim of this experiment was to look at effects of learning a tonal language on melodic perception. To observe possible transfer of pitch perception skills from language to music, we compared the melodic aptitude of learners of a tonal language, in this case Mandarin Chinese, with that of Dutch-English and Chinese-English Bilinguals. As previous studies have shown that native Chinese speakers show better melodic perception than non- tonal speakers (e.g., Roncaglia-Denissen et al., 2016; Wong et al., 2012) we expected tonal learners, especially advanced learners, to show better melodic perception than nontonal speakers. The results indicated that Chinese-English bilinguals performed significantly higher than the other three groups: Dutch-English bilinguals, beginners and advanced learners of Mandarin Chinese. This result showing enhanced melodic perception of Chinese-English bilinguals replicated the findings of previous studies (e.g., Chen et al., 2016). This enhancement in melodic perception could be attributed to the exposure and proficiency in discriminating lexical tones that enable the Chinese-English bilinguals to perform better in perception of musical melodies. Although beginners and advanced learners of Chinese were exposed to the Chinese tones for roughly 33 weeks and 220 weeks respectively, no significant enhancement in melodic perception was seen while being compared with the nontonal group of Dutch-English bilinguals. Tonal language learners and native speakers differ in terms of amount of exposure which could possibly be a reason as to why tonal language learners did not perform as well as the native speakers. This is supported by the findings of Bidelman, Gandour & Krishnan (2010) who pointed out that length of exposure is positively associated with the tonal speakers' melodic discrimination abilities and that the neural representation of cross-domain transfer relied on the experience and amount of training an individual has in one of the domains. This could mean that learning a tonal language may not be sufficient to reap the benefits of enhanced melodic perception that is evident among native tonal speakers.

We also aimed to explore the relationship between language and music by observing the cross-domain correlation following up on the "split hypothesis" proposed by Chen et al. (2016). The correlation analysis between the melodic aptitude test (Musical domain) and Chinese tone discrimination task (Language domain) was done for each of the four groups: Chinese-English bilinguals, Dutch-English bilinguals, beginners and advanced learners of Chinese. The Chinese tone discrimination task included both monosyllabic discrimination (comparison between two monosyllabic tones) and bisyllabic discrimination (comparison between two pairs of bisyllabic tones). The results showed that Chinese-English bilinguals had no correlation between the lexical tone discrimination task (both monosyllabic and bisyllabic discriminations) and their melodic aptitude performance, which was similar to the results seen in Chen et al. (2016). The beginners and advanced learners of Chinese also did not show any correlation between the melodic aptitude task and monosyllabic and bisyllabic tone discrimination present in the Chinese tone discrimination task. As portrayed in Chen et al. (2016), the reason behind why there is no correlation among native tonal speakers could be due to the fact that the pitch information that they receive from both the tasks (MET and Chinese tone discrimination) contextually differ. Being tonal speakers, the Chinese-English bilinguals have an intact representation of Chinese tones that carry a lexical role due to which they may split the processing of lexical tones from the processing of other pitch variation from other domains, in this case, music (Chen et al., 2016). Likewise, the same reasoning is valid to explain the absence of correlation between the melodic task and Chinese the tone discrimination task among beginners and advanced learners of Chinese. Both groups have been exposed to lexical tones through learning Chinese, which enabled them to perceive the incoming pitch variations of Chinese tone discrimination task differently from the melodic aptitude task leading to split processing. There has been evidence from Nan, Sun & Peretz, (2010) supporting the independent nature of pitch processing as seen in native tonal speakers who have congenital amusia (inability to discriminate or reproduce different melodic tones). This shows that there may not unified processing of pitch input from both language and music for speakers of tonal language who learn them in different contexts. However, for non-tonal speakers, there is no difference in context in terms of the pitch information they receive from both the tasks. From our results, we noticed that the Dutch-English bilinguals showed significant correlation between the monosyllabic discrimination and melodic task similar to the results of Chen et al. (2016). There was also a similar trend seen between bisyllabic discrimination and melodic task performances among Dutch-English bilinguals. The significant correlation between language task and music task among Dutch-English speakers can be attributed to the fact that they did not undergo any training in learning Chinese tones or music and therefore processed the incoming pitch input as a general psychoacoustic level that led to their correlation. Studies in the past have shown how lexical tones are processed differently among tonal language speakers and nontonal language speakers. For example, Halle, Change & Best (2004) found that French listeners processed lexical tone differently from Cantonese listeners but were still sensitive to the variations. Another study by Francis, Ciocca, Ma & Fenn (2008) found differences in perceptual spaces for tonal speakers and nontonal speakers while processing lexical tone. These findings are in line with the reasoning indicating that nontonal speakers do not split the incoming pitch input as lexical tone from melodic pitch variations.

There is also visible trend of decrease in amount of correlation between melodic task and Chinese tone task: Higher correlation between language and music was evident among Dutch-English bilinguals and this degree of correlation decreased from beginners to advanced learners of Chinese, further leading to no correlation seen among Chinese-English bilinguals. This trend gives us an idea on how the amount of exposure to tone variations in a language can possibly have an influence on nature of relationship between the performances in the language and music tasks.

Exploring the relationship between language and music by observing the “split processing” can give us more details regarding the factors responsible for transfer from language to music to take place. Although the Chinese-English group shows no correlation between the musical task and Chinese tone discrimination task, they still show higher accuracy in melodic aptitude than Dutch-English bilinguals, beginners and advanced learners. This indicates that even though performance in language and music tasks may not be directly related as seen in Chinese-English bilinguals and learners of Chinese, the exposure in tonal variations in Chinese leads to perceptual enhancement of factors like pitch acuity that is shared both by music and language. This enhancement of the common factor like pitch acuity, in turn, helps in perceiving musical tones better. It has been found that expertise in language and music show neural enhancement in auditory brain stem that may lead to better pitch acuity (Bidelman, Gandour & Krishnan, 2011). Hence, looking beyond the direct influence of lexical tone performance on musical task, and knowing that they are not directly correlated shows us the possibility of underlying factors like pitch acuity that could lead to the transfer of learning effect. Future studies can explore this further, by implementing neuroimaging methods in addition to the behavioural evidence.

GENERAL DISCUSSION & CONCLUSION

Both experiments carried out in this study aimed at investigating the influence of experience from speaking certain languages on musical rhythmic and melodic perception. Experiment 1 focused on the transfer of rhythmic perception skills from language to music, while Experiment 2 looked at the transfer of pitch perception skills obtained from learning a tonal language to perceiving melodic differences. Carrying forward from the past studies, the first experiment compared the rhythmic aptitude among two language groups that differed in unit level classification but shared metric preference of trochee. The results did not show a significant enhancement in rhythmic aptitude, possibly because the languages that were studied shared the metric preference. Another important factor to take into account is that the monolingual group may not have been representative of a general monolingual population, because they performed better than another monolingual group tested in the previous study (Roncaglia-Denissen et al., 2016). Therefore, further research needs to be conducted to include other monolingual samples in order to validate the bilingual advantage observed in rhythmic perception skills.

In the second experiment, the native tonal speakers who were Chinese-English bilinguals outperformed all the groups including the Chinese learners, in the melodic aptitude test. This indicates that although the learners of Chinese were exposed to the lexical tone variations, learning a tonal language was not sufficient to exhibit cross-domain transfer that was seen among native tonal language speakers. Instead, interesting patterns of correlations between their lexical tone discrimination accuracy and melodic perception were found. Only the Dutch-English bilinguals group showed significant correlation between their performances in the lexical task and melodic task. The rest of the groups that had been exposed to lexical tones, although varying in amount of exposure, showed no correlation between the two tasks. This may suggest that the participants with no exposure to lexical tones (Dutch-English) and participants with exposure to lexical tones (Chinese-English bilinguals, Beginners and Advanced learners of Chinese), process the incoming tonal information in different ways: As the pitch input from the Chinese tone discrimination task and MET are contextually different for the Chinese learners and Chinese Native speakers, their processing of the lexical tone input is split from the musical pitch variations, implying differences in the way they perceive the pitch input. This ability to split the pitch information is absent among non-tonal speakers as they perceive the input as general acoustic information in the psychophysical form.

Knowing that tonal experience influences the way tonal input is perceived, we understand that the cross-domain transfer between language and music might rely on a deeper underlying shared factor since the two domains do not show any direct correlation in terms of their performance in the language and musical tasks. However, this needs to be further supported by neuroimaging evidence that might give a better picture on the split processing of lexical and musical tones that differ contextually for tonal and non-tonal speakers.

Hence, findings from both the experiments expand the literature on transfer of learning effect from language with respect to the influence of linguistic experience, in terms of speaking languages with rhythmic variability and learning a tonal language on musical rhythmic and melodic perception. Future research can take the findings of both experiments forward, by expanding the study of transfer effect through investigating the mechanisms and cortical areas responsible for music and language and how they rely on each other. This will in turn provide more insight on the relationship between language and music and establish the fact that they are indeed two sides of the same coin.

References

Arbib, M. A. (2013). *Language, music, and the brain: a mysterious relationship*. Cambridge, MA: The MIT Press.

Anvari, S.H., Trainor, L.J., Woodside, J., & Levy, B.A (2002). Relations among musical skills, phonological processing, and early reading ability in preschool children. *Journal of Experimental Psychology 83(2)* 111-130.

Arvaniti, A. & Tristie, R. (2010), Rhythm Class & Speech perception. *Speech Prosody 2010 100887:1-4*

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychol. Rev., 105*, 158–173.

Bialystok, E. (2006) Effect of bilingualism on computer video games experience on the Simon task. *Canadian Journal of Experimental psychology, 60,* 68-79.

Bidelman, G.M., Gandour, J.T., Krishnan, A., (2011). Musicians and tone-language speakers share enhanced brainstem encoding but not perceptual benefits for musical pitch. *Brain and Cognition 77,* 1-10.

Bidelman, G. M., Hutka, S., & Moreno, S. (2013). Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: evidence for bidirectionality between the domains of language and music. *PLoS One, 8*(4):e60676.

Besson, M., & Schön, D. (2001). Comparison between language and music. *Biological Foundations of Music, 930*, 232-258.

Binder, J.R., Forst, A. Jr. Hammeke, T.A., …… Prierto, T. (1997). *The Journal of Neuroscience 17(1),* 353-362.

Brown, S., & Martinez, M.J., & Parsons, L.M. (2006). Music and language side by side in the brain: a PET study of the generation of melodies and sentences. *European Journal of Neuroscience, 23(10),* 2791 – 2803.

Chen, A., Liu, L., & Kager, R. (2016). Cross-domain correlation in pitch perception, the influence of native language. *Language, Cognition and Neuroscience, 3798(April),* 1– 10.

Chen, A., Çetinçelik, M., Ronacaglia-Denissen, P., & Sadakata, M. (in prep.) Native language and bilingualism on pitch processing in different domains

Chua, A.J., & Brunt, J., (2015). Effect of musical experience on tonal language perception.

Proceedings of Meetings on Acoustics 060009 21(1)

Cutler, A. (1994). The perception of rhythm in language, *Cognition* 50. 79-81.

Dauer, R.M. (1983). Stress-timing and syllable-timing reanalyzed. *J. Phon.,* 11, 51-62.

Delogu, F., Lampis, G., & Belardinelli, M.O. (2010). From melody to lexical tone: Musical ability enhances specific aspects of foreign language perception. *European journal of cognitive psychology, 22(1),* 46-61.

Duanmu, S. (2000). *The phonology of standard chinese* . Oxford: Oxford University Press.

Falk, D. (2000). Hominid brain evolution and the origin of music. In The *Origins of*

Music. N.L. Wallin, B. Merker & S. Brown, Eds.: 197–216. MIT Press. Cambridge, MA.

Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics, 36(2),* 268–294.

Hallé, P. A., Chang, Y., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics, 32(3),* 395–421.

Hayes, B. (1985). Iambic and trochaic rhythm in stress rules. *Proceedings of the Eleventh Annual Meeting of the Berkeley Linguistics Society (California),* 429–446.

Inkelas, S., & Orgun, O. (2003). Turkish Stress: a review. *Phonology* 20, 139–161.

Jusczyk, P.W., Cutler, A., & Redanz, N.J. (1993). Infants' Preference for the Predominant Stress Patterns of English Words. *Child Development. 64(3),* 675-687.

Kraus, N., & Banai, K. (2007) Auditory-Processing Malleability: Focus on language and music, *Current Directions in psychological Science, 16(2)* 105-110.

Krishnan, A., Xu, Y., Gandour, J., Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Brain Res Cogn Brain Res*. 25(1):161-168.

Koelsch S. (2000). Brain and Music – A Contribution to the Investigation of Central Auditory Processing with a New Electrophysiological Approach. Leipzig: Risse

Koelsch, S. (2011). Towards a neural basis of neural basis of music perception – a review and updated model. *Frontiers of Psychology 2,* 110.

Kuhl, P.K., Williams, K.A., Lacerda, F., Stevens, K.N., Lindblom, B., (1992). Linguistic Experience Alters Phonetic Perception in Infants by 6 Months of Age, *Science,* 255(5044), 606-08.

Lee, C. Y., & Hung, T. H. (2008). Identification of Mandarin tones by English-speaking musicians and nonmusicians. *Journal of the Acoustical Society of America, 124(5),* 3235–3248.

Livonen, A., & Harnud, H. (2005). Acoustical comparison of the monophthong systems in Finnish, Mongolian and Udmurt. *Journal of the International Phonetic Association, 35(1)* 59–71

Low E.L. (2006). A Review of Recent Research on Speech Rhythm: Some Insights for

Language Acquisition, Language disorders and Language Teaching. In: *Hughes R.*

(eds) Spoken English, Tesol and Applied Linguistics. Palgrave Macmillan, London

Marie, C., Magne, C., & Besson,M. (2011). Musicians and the Metric Structure of Words.

Journal of Cognitive Neuroscience 2011, 23(2), 294-305

McDermott, J., & Hauser, M. (2005). The origins of music: Innateness, uniqueness, and evolution. *Music Perception, 23*, 29–59.

Mok, P., & Zuo, D., (2012). The separation between music and speech: Evidence from the perception of Cantonese tones. *Journal of the Acoustical Society of America, 132(4),* 2711-20.

Mottier, G. (1951). Über Untersuchungen der Sprache lesegestörter Kinder. *Folia Phoniatr.*

Logop. 3, 170–177.

Müllensiefen, D., Gingras, B., Stewart, L. & Musil, J. (2014). The Musicality of Non- Musicians: An Index for Measuring Musical Sophistication in the General Population. *PLoS ONE 9(*2): e89642.

Mussachia, C., Strait, D., Kraus, N., (2008). Relationships between behavior, brainstem and cortical encoding of seen and heard speech in musicians and non-musicians. *Hearing research, 241(1-2)* 34-42.

Nan, Y., Sun, Y., & Peretz, I. (2010). Congenital amusia in speakers of a tone language: Association with lexical tone agnosia. *Brain, 133(9),* 2635–2642.

Nettl, B. (2000). An ethnomusicologist contemplates universals in musical sound and musical culture. *In N.L. Wallin, B. Merker & S. Brown (Eds.) The origins of music (pp.463– 472). Cambridge, MA: MIT Press*

Oberauer, K., Süß, H.-M., Schulze, R., Wilhelm, O., & Wittmann, W. W. (2000). Working memory capacity–facets of a cognitive ability construct. *Pers. Individ. Dif., 29*, 1017– 1045.

Patel, A.D. (2008). *Music, language, and the brain.* New York: Oxford University Press.

Patel, A.D., (2003). Rhythm in Language and Music: Parallels and Differences. Annals of the New York Academy of Sciences, 999(1), 140-143.

Patel, A.D., & Daniele, J.R. (2003). The empirical comparison of rhythm in language and music. *Cognition 87 (1),* 35-45

Patel, A.D., (2011). Why would Musical Training Benefit the Neural Encoding of Speech?

The OPERA Hypothesis. *Frontiers in psychology, 2,* 142.

Peretz, I., & Zatorre, R. J. (2005). Brain organization for music processing. *Annual Review of Psychology*, 56:89–114.

Pfordresher, P.Q., & Brown, S. (2009) Enhanced production and perception of musical pitch in tone language speakers. Attention, Perception & Psychophysics, 71(6), 1385- 1398.

Pike, K. L. (1945). *The Intonation of American English*. Ann Arbor, MI: University of Michigan Press.

Roach, P. (1983). On the distinction between "stress-timed" and "syllable-timed" languages. *In Crystal, D. (ed.), Linguistic Controversies*, Arnold, London

Roncaglia-Denissen, M. P., Schmidt-Kassow, M., Heine, A., Vuust, P., & Kotz, S. A. (2013). Enhanced musical rhythmic perception in Turkish early and late learners of German. *Frontiers in Psychology*, *4*(SEP), 1–8.

Roncaglia-Denissen, M. P., Roor, D. A., Chen, A., & Sadakata, M. (2016). The Enhanced Musical Rhythmic Perception in Second Language Learners. *Frontiers in Human Neuroscience*, *10*(June), 1–10.

Schellenberg, E. G. (2006). Long-term positive associations between music lessons and IQ. *Journal of Educational Psychology, 98*(2), 457-468.

Schmidt-Kassow, M., Roncaglia-Denissen, M.P., & Kotz, S. A. (2011).Why pitch sensitivity matters: event-related potential evidence of metric and syntactic violation detection among Spanish late learners of German. *Frontiers of Psychology:Language Sciences*. 2:131.

Schon, D., Magne, C., & Besson,M., (2004). The music of speech: Music training facilitates pitch processing in both music and language. *Psychophysiology, 41,* 341–349

Slevc, L.R., & Miyake, A.L. (2006). Individual Differences in Second-Language Proficiency *Does Musical Ability Matter?* Psychological Science 17(8).

Thompson, W.F., Ilie, G., & Schellenberg, E. (2004) Decoding Speech Prosody: Do Music Lessons Help? *Emotion, 4(1)*, 46-64.

Veenker, T.J.G. (2017). The Zep Experiment Control Application (Version *1.8*) [Computer software]. Beexy Behavioural Experiment Software. Available from http://www.beexy.org/zep/

Wallentin, M., Nielsen, A. H., Friis-Olivarius, M., Vuust, C., & Vuust, P. (2010). The musical ear test, a new reliable test for measuring musical competence. *Learning and Individual Differences., 20,* 188–196.

Wong, P.C., Skoe, E., Russo, N.M., Dees, T., Kraus, N., (2007) Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience 10(4),* 420-422.

Wong, P.C.M., Ciocca, V., Chan, A.H.D., Ha, L.Y.Y., Tan, L., & Peretz, I. (2012). Effects of Culture on Musical Pitch Perception. *PLoS One 7(4):* e33424.

Williamson, V. (2009). In search of language and music, *The Psychologist 22(12),* 1022-1025

Zatore, R.J., Ecans, A.C., Meyer, E., Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science 256 (5058),* 846-849.

An Electrophysiological Perspective on the Resolution of Anaphoric Dependencies: Evidence from Event-Related Potentials and Neural Oscillations

Cas Coopmans[1,2]
Supervisor: Mante Nieuwland[1,2]

[1]Max-Planck-Institute for Psycholinguistics, Nijmegen, The Netherlands
[2]Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands

In this study, we used electroencephalography to investigate the effects of givenness and discourse coherence of proper names on the electrophysiological correlates of anaphoric reference resolution. Participants read two-sentences mini-discourses in which repeated and new proper names were coherent or incoherent with the preceding discourse. Preregistered analyses revealed effects of givenness and coherence in both event-related potentials (ERPs) and neural oscillatory dynamics. In comparison with new names, repeated names elicited a reduced N400 and reduced Late Positive Component, and an increase in theta-band (4-7 Hz) synchronization. Discourse-coherent names elicited an increase in gamma-band (60-80 Hz) synchronization in comparison to discourse-incoherent ones, while no effects of discourse were observed in the ERP signal. We additionally observed an Nref ERP effect for new proper names that could not be properly linked to a discourse referent. We argue that the observed theta old/new effect reflects the reactivation of a representation held in working memory, and interpret the gamma-band effect in terms of successful semantic integration of coherent proper names. These results demonstrate that discourse-level anaphoric reference can be studied through neural oscillations and further establish the role of memory mechanisms in discourse-level language comprehension.

Keywords: Proper names, discourse comprehension, reference resolution, oscillations, ERPs, theta, gamma

Corresponding author: Cas Coopmans ; **E-mail**:cas.coopmans@mpi.nl

A cognitive perspective on anaphoric reference

Anaphoric dependency refers to the referential relationship between two linguistic representations that are separated from each other in time. This property of human language poses a real challenge to the comprehension system, because successful reference resolution requires multiple cognitive systems to work in concert. Consider the following discourse context: 'Lionel Messi and Cristiano Ronaldo are two of the world's greatest football players. Unsurprisingly, last year's Ballon d'Or was won by Ronaldo'. In order to understand that Ronaldo in the second sentence refers to the Portuguese Cristiano Ronaldo and not to the resigned Brazilian football player who happens to bear the same name, one has to establish a link between the anaphor 'Ronaldo' in the second sentence and its antecedent 'Cristiano Ronaldo' in the preceding discourse. In addition, in order to evaluate the correctness of this statement, the anaphor has to be integrated into the semantic representation of this sentence. In other words, in order to understand anaphoric dependencies, the language processor has to be able to distinguish between what is new and what has been referred to before, a distinction termed givenness, and to build a coherent and meaningful discourse representation on the basis of the meaning of the individual words. A recent proposal by Nieuwland and Martin (2017) links these two processes (i.e., memory retrieval and semantic integration), assumed to be core components underlying anaphor resolution, to increased oscillatory synchronization in two ranges of the gamma band. Nieuwland and Martin argue that anaphor comprehension draws on an interaction between the brain's recognition memory network and the language system. This specific proposal forms part of a bigger enterprise that aims to answer the question whether the neural mechanisms classically thought to be involved in memory functioning might also underlie online language processing (Covington & Duff, 2016; Duff & Brown-Schmidt, 2012; Piai et al., 2016).

In an attempt to test Nieuwland and Martin's proposal, we performed an electroencephalography (EEG) study to further investigate the effects of givenness and coherence on neural oscillations. In order to study the effects of givenness, we utilized a key attribute of proper names, which is that they can both establish reference by introducing a new referent into the discourse and maintain reference by referring to a given discourse referent. These proper names were embedded in two-sentence mini-discourses, in which we additionally manipulated the coherence of the proper names with respect to the preceding discourse. By looking at how these effects play out in electrophysiological brain recordings, we aim to contribute to the emerging view on the role of memory processes in language comprehension.

An electrophysiological perspective on anaphoric reference

One prominent memory-based approach to linguistic dependency resolution is known as the cue-based retrieval framework (Martin, 2016; McElree, 2000; McElree, Foraker & Dyer, 2003). The cue-based retrieval framework argues that the memory system subserving sentence comprehension has a content-addressable architecture, in which access and retrieval of content-addressable memory representations is cue-induced and direct, without a search through irrelevant representations. Retrieval success is dependent on the match between the cue that triggered retrieval and the target representation (McElree, 2006). Specific to anaphor comprehension, this equates to identifying the referent of an anaphor by evaluating the overlap between the representation of the anaphor that triggered retrieval and the representation of the antecedent held in working memory.

Recent studies have demonstrated the potential of the cue-based retrieval framework to account for the electrophysiological correlates of anaphoric reference by using event-related potentials (ERPs) to study cue-based retrieval during online language comprehension (e.g., Martin, Nieuwland & Carreiras, 2012, 2014). ERPs are voltage fluctuations in the electroencephalogram that are evoked by an event, such as an external stimulus. Comparing the average ERP signal between conditions reveals ERP effects, which, by virtue of their multidimensional nature (i.e., polarity, amplitude, time course, scalp topography), can provide both quantitative and qualitative information about the cognitive event at hand (Van Berkum, 2004). The downside of ERPs is that, due to averaging, any event-related response that is nonstationary (i.e., has a jitter in phase and/or latency) will be canceled (Tallon-Baudry & Bertrand, 2000). ERPs therefore only capture neural activity that is strictly time- and phase-locked to the stimulus event.

Non-phase-locked neural activity, known as event-related neural oscillations, are event-induced modulations of ongoing rhythmic patterns of neural activity. Because oscillations are always present, the phase of the oscillatory signal at the onset of the event is variable, meaning that event-induced changes in the oscillatory EEG activity are not strictly phase-locked (Bastiaansen, Mazaheri & Jensen, 2012). This non-phase-locked activity can be visualized by a time-frequency approach that computes the power of the activity at each individual frequency prior to averaging. As power (amplitude squared) is a positive measure, averaging the time-frequency signal over multiple trials leaves in both phase-locked and non-phase-locked activity. The benefit of studying both ERPs and neural oscillations is that they, because they reflect dissociable aspects of the same neuronal activity, might provide a complementary view on the neural activity related to cognitive processes (Davidson & Indefrey, 2007; Kielar, Panamsky, Links & Melzer, 2015). In the following sections, we will review the electrophysiological literature on memory reactivation and semantic integration during language comprehension.

Event-related potentials and anaphoric reference

Givenness

Behavioral studies have established that word repetition leads to repetition priming: repeated words are easier to process than new words, as demonstrated via lexical decision tasks (e.g., Scarborough, Cortese & Scarborough, 1977). In EEG studies of recognition memory, this repetition priming effect is known as the ERP old/new effect, which often takes the form of a reduction of the amplitude of the N400 component and a subsequent Late Positive Component (LPC) for repeated compared to new words (Rugg, 1985, 1990; Van Strien, Hagenbeek, Stam, Rombouts & Barkhof, 2005; Rugg & Curran, 2007). The N400 component is a negative deflection peaking between 300-500 ms after the onset of each content word and is largest over centro-parietal electrodes (the difference between two components constitutes the N400 effect; Kutas & Hillyard, 1980; Kutas & Federmeier, 2011), while the LPC has a parietal distribution, begins approximately around 500 ms after word onset and can last until up to 1000 ms (for a review, see Van Petten & Luka, 2012). In sentence and discourse context, where repetition is a natural means to establish co-reference, repeated words also elicit a reduced N400 component (Van Petten, Kutas, Kluender, Mitchiner & McIsaac, 1991; Streb, Henninghause & Rösler, 2004; Camblin, Ledoux, Boudewyn, Gordon & Swaab, 2007a; Ledoux, Gordon, Camblin & Swaab, 2007; Almor, Nair, Boiteau & Vendemia, 2017), but new words elicit a larger LPC than repeated words (Burkhardt, 2006, 2007; Kaan, Dallas & Barkley, 2007; Schumacher, 2009; Wang & Schumacher, 2013). The attenuated N400 in response to a repeated word (i.e., anaphor) seems to reflect the ease with which its referent (i.e., antecedent) is identified within the preceding discourse, while the LPC for new words indexes augmentation of the current discourse model with a new entity (Burkhardt, 2006; Schumacher, 2009; Schumacher & Hung, 2012; Wang & Schumacher, 2013). Evidence for this two-process interpretation was provided by Burkhardt (2006), who demonstrated that inferentially-bridged noun phrases, such as 'the conductor' in (1), first pattern with repeated words (2) in terms of a reduced N400 and then with new words (3) in terms of an enhanced LPC.

1) Tobias visited a *concert* in Berlin. He said that **the conductor** was very impressive.

2) Tobias visited a *conductor* in Berlin. He said that **the conductor** was very impressive.

3) Tobias talked to *Nina*. He said that **the conductor** was very impressive.

Although 'the conductor' in (1) might rather effortlessly be anchored to the discourse based on inferential knowledge (i.e., *concert* implies *conductor*), explaining the reduced N400, it will still have to be introduced into the discourse model as an independent representation, explaining the enhanced LPC (Burkhardt, 2006).

Discourse coherence

In line with these findings, the amplitude of the N400 has been argued to reflect the ease or difficulty with which conceptual knowledge associated with words or names is retrieved (Kutas & Federmeier, 2011; Van Berkum, 2012). This process can be modulated by context, but only to the comprehender's benefit. That is, while a coherent context can reduce the N400 amplitude by facilitating ease of retrieval, no additional retrieval cost (i.e., enhanced N400) seems to be incurred from an incoherent context (Hagoort & Van Berkum, 2007; Kutas & Federmeier, 2011; Van Petten & Luka, 2012). The effects of discourse on semantic retrieval have been addressed by numerous studies, which all showed that the N400 is reduced for words that are coherent with respect to the preceding discourse compared to discourse-incoherent words (Camblin, Gordon & Swaab, 2007b; Filik & Leutholdt, 2008; Nieuwland & Van

Berkum, 2006a; Salmon & Pratt, 2002; St. George, Mannes & Hoffman, 1994; Van Berkum, Hagoort & Brown, 1999a; Van Berkum, Zwitserlood, Hagoort & Brown, 2003). An important finding of these studies was that the discourse-level and sentence-level N400 were identical in terms of time course, morphology and scalp distribution (Van Berkum et al., 1999a, 2003; Salmon & Pratt, 2002; Nieuwland & Van Berkum, 2006a), indicating that the processes responsible for the N400 do not seem to be sensitive to where the semantic constraints come from (Van Berkum, 2004, 2012).

With particular relevance to the current study are the findings from a study by Wang and Yang (2013), who demonstrated the beneficial effects of context on the retrieval of proper names from working memory. They set up a two-sentence discourse context in which two proper names were introduced and described with contrastive characteristics (e.g., 'John is a *singer*, Peter is an *actor*'). In the third sentence, the interpretation of the critical proper name was either coherent or incoherent with respect to the preceding discourse (e.g., 'a film producer/music producer came to *Peter*'). Compared to discourse-coherent names, incoherent proper names elicited a widely distributed N400 effect and an additional P600 effect, showing that the learned meaning of previously unknown proper names can easily and rapidly be reactivated from working memory and integrated into the context.

Givenness in discourse

Although repetition in language context seems to show a similar N400 profile as repetition in word lists, the discourse context has been shown to modulate this effect. That is, when repeated proper names refer to antecedents that are in discourse focus, as in (4), they elicit a larger N400 than repeated proper names that refer to an antecedent that is not prominent in the discourse model, as in (5) (Swaab, Camblin & Gordon, 2004).

4) At the office *Daniel* moved the cabinet because **Daniel** needed room for a desk.

5) At the office *Daniel and Amanda* moved the cabinet because **Daniel** needed room for a desk.

This effect is an electrophysiological manifestation of the so-called repeated name penalty (Gordon, Hendrick, Ledoux & Yang, 1999), and has been observed in both reading (Swaab et al., 2004; Ledoux et al., 2007) and listening to fully connected natural speech (Camblin et al, 2007a). Notably, the N400 elicited by repeated name anaphors that co-refer with a discourse-prominent antecedent is comparable to the N400 elicited by mentioning a new name, suggesting that the repeated-name penalty reflects the effect of discourse prominence overriding otherwise facilitatory effects of repetition (Camblin et al., 2007a).

A referentially induced negativity

The ERP effect most prominently associated with referential processing is a sustained negativity with an anterior distribution (for a review, see Van Berkum, Koornneef, Otten & Nieuwland, 2007). Van Berkum, Brown and Hagoort (1999b) observed that referentially ambiguous noun phrases, such as 'the girl' in a discourse that involved two girls, elicited a rapidly (~300 ms after noun onset) emerging and relatively long-lasting negativity with a predominantly frontal distribution. This sustained frontal negativity in response to written referential ambiguity was subsequently replicated in the auditory modality, using the same mini-discourses presented as naturally produced connected speech (Van Berkum, Brown, Hagoort & Zwitserlood, 2003). Later studies by Van Berkum, Nieuwland and colleagues showed that this referentially induced frontal negativity, termed Nref, is reliably elicited in the comparison between unambiguous and ambiguous anaphors, such as 'he' in 'David told Peter that he …' where both 'David' and 'Peter' are equally plausible antecedents (Nieuwland, 2014; Nieuwland, Otten & Van Berkum, 2007a; Nieuwland & Van Berkum, 2006b; Van Berkum, Zwitserlood, Bastiaansen, Brown & Hagoort, 2004). Note that the results of these studies do not imply that "the processes directly responsible for the negativity here must be uniquely tied to resolving referential ambiguity" (Van Berkum, 2009, p. 301), but merely show that "the processing consequences of referential ambiguity quite consistently show up as sustained frontal negativities" (Van Berkum, 2012, p. 600).

In line with these latter statements, a recent ERP study found an Nref-like effect in response to the resolution of anaphors that were referentially unambiguous (Barkley, Kluender & Kutas, 2015). Barkley and colleagues presented participants pronouns and proper names that either did (6) or did not (7) have an antecedent earlier in the sentence.

6) After a covert mission that deployed *Will*$_i$/*him*$_j$ for nine terrible months, **he**$_i$**/Will**$_j$ longed for home.

7) After a covert mission that required deployment for nine terrible months, **he/Will** longed for home.

In comparison with pronouns without antecedents, pronouns with antecedents elicited a large negativity with an anterior distribution. No difference was observed between proper names with and without antecedents. These findings were interpreted in terms of the cue-based retrieval properties of certain anaphoric forms (McElree et al., 2003). According to the authors, pronouns contain a [+ antecedent] feature that cues a process they call 'back association', which involves the reactivation of an already encoded antecedent in order to allow the establishment of a referential dependency. Proper names, in contrast, are known to be primarily used to introduce entities into the discourse rather than to maintain reference (Gordon & Hendrick, 1998; Gordon et al., 1999). Accordingly, proper names were argued not have the [+ antecedent] feature and therefore do not trigger back association (Barkley et al., 2015).

The fact that several ERP effects are seen in studies aimed at identifying the electrophysiological correlates of anaphoric reference indicates that it does not rely on a single process. Rather, it relies on the cooperation of multiple cognitive operations, among others retrieval from working memory and semantic integration. ERPs are inherently limited in what they can tell us about the interaction between these operations. The study of neural oscillations, however, might allow us to gain more insight into not only which cognitive components are involved in anaphor comprehension (e.g., via investigation of local oscillatory power) but also how these components work together in order to establish a coherent discourse representation (e.g., via studying global phase coherence).

Neural oscillations and anaphoric reference

Different aspects of language comprehension have been related to different oscillatory frequency bands. Memory reactivation and semantic integration have most prominently been associated with oscillations in the theta and gamma frequency bands (for a recent review, see Meyer, 2017).

Theta oscillations

With particular relevance for the current study are two lines of research that have observed a relationship between retrieval from both long-term memory and working memory and activity in the theta band (4-7 Hz). First, theta-band oscillations have been related to the ease with which lexico-semantic information is retrieved from long-term memory, as evidenced by larger theta power for semantically rich compared to semantically lean words (Bastiaansen, Van der Linden, Ter Keurs, Dijkstra & Hagoort, 2005), larger theta power for semantically anomalous compared to semantically coherent words (Davidson & Indefrey, 2007; Hald, Bastiaansen & Hagoort, 2006; Wang, Zhu & Bastiaansen, 2012) and differences in the scalp distribution of theta power as a function of the modality-specific properties of words (Bastiaansen, Oostenveld, Jensen & Hagoort, 2008). Second, in the recognition memory literature, theta is related to the recognition and retrieval of successfully remembered probes (see Nyhuis & Curran, 2010 for a review). These studies either employ a study-test or a continuous recognition paradigm. In the former, participants are asked to study a list of items, often words, and after some time (sometimes on a different day) are given another list of items and are asked to indicate whether these items had been studied before ('old') or not ('new'). In continuous recognition paradigms, participants walk through a list of items and have to judge continuously for each individual item whether it has been presented in the list before or not. Correctly remembered old items reliably elicited larger theta-band synchronization than correctly recognized new items, both in the study-test paradigm (Chen & Caplan, 2016; Jacobs, Hwang, Curran & Kahana, 2006; Klimesch, Doppelmayr, Schimke & Ripper, 1997; Klimesch, Doppelmayr, Schwaiger, Winkler & Gruber, 2000; Osipova et al., 2006) and in the continuous recognition task (Burgess & Gruzelier, 1997, 2000; Van Strien et al., 2005; Van Strien, Verkoeijen, Van der Meer & Franken, 2007). These theta oscillations have therefore been linked to the process of matching the probe to representations in working memory, where old and new items are differentiated (Jacobs et al., 2006; Chen & Caplan, 2016).

One could argue that, despite potential differences in task-induced strategies, the retrieval demands of continuous recognition and anaphor comprehension are conceptually similar. In both cases, an item (probe, anaphor) triggers the retrieval of a recently encoded item (target, antecedent) that is held in working memory. In addition, in both cases the items are separated in time and intervened with other items that might interfere with retrieval success. The core computational operations involved in recognition memory and anaphor comprehension (i.e., distinguishing old from new information; matching input to a representation in working memory) are thus conceptually related and might even be shared (Covington & Duff, 2016; Martin, 2016; Nieuwland & Martin, 2017).

Gamma oscillations

Oscillations in the gamma-band (>30 Hz), as found in language experiments, have been related to semantic integration processes (called semantic unification; Bastiaansen & Hagoort, 2006; Hagoort, Willems & Baggio, 2009). These experiments show that whenever semantic unification is successful, gamma-band power is increased. For instance, gamma-band synchronization is increased for referentially coherent pronouns relative to pronouns that are referentially ambiguous or do not have a proper referent (Van Berkum, Zwitserlood, Bastiaansen, Brown & Hagoort, 2004), for semantically coherent sentences compared to syntactic prose (i.e., syntactically correct but semantically anomalous sentences; Bastiaansen & Hagoort, 2015), and for semantically coherent compared to anomalous words (Hald et al., 2006; Penolazzi, Angrili & Job, 2009), the latter effect being modulated by the semantic relatedness between the anomalous word and the sentence context (Rommers, Bastiaansen & Dijkstra, 2013).

In a recent paper, Nieuwland and Martin (2017) reported time-frequency analyses of four previous ERP studies on referential processing. All four studies compared referentially coherent anaphors to referentially incoherent ones (ambiguous or problematic; e.g., 'Peter thought that *he/she* would win the race'). Two bursts of gamma-band power were observed in response to referentially coherent anaphors: an increase in 35-45 Hz ('low') gamma-band power between 400 and 600 ms and another gamma-band increase of 60-85 Hz ('high gamma') between 500 and 1000 ms. Source localization revealed a strong source for the low gamma effect in the left posterior parietal cortex (LPPC), while the high gamma effect was localized to the left inferior frontal-temporal cortex, including the left inferior frontal gyrus (LIFG). Largely because the LPPC has been shown to differentiate old from new information (Gonzalez et al., 2015) and has been related to successful retrieval from episodic memory (for a review, see Wagner, Shannon, Kahn & Buckner, 2005), the low gamma effect was taken to reflect the reactivation of the antecedent by the brain's recognition memory network. As the LIFG is known to be involved in sentence-level semantic unification (Hagoort, 2005; Hagoort & Indefrey, 2014), the high gamma effect was argued to reflect the integration of the reactivated antecedent into the semantic representation of the sentence.

Notably, Nieuwland and Martin (2017) acknowledge that these interpretations face several challenges. First of all, although the two gamma effects were assigned different functional interpretations, it is questionable whether they are truly distinct. Secondly, the low gamma effect was not found in all of their experiments, and it shows up in a frequency range that is not commonly associated with successful language comprehension, nor with successful memory retrieval. It is therefore not yet clear whether these gamma effects reflect the workings of a recognition memory network or whether they are more specifically related to processes involved in resolving referential ambiguity. Third, memory processes have been associated most prominently with the oscillations in the theta band, but it is unclear how these relate to the memory mechanisms that are relevant for anaphor comprehension. We aim to address these questions in a dedicated EEG experiment that focuses on the effects of givenness and coherence of proper names in discourse.

Present study

The present EEG study uses ERPs and neural oscillations to investigate the involvement of memory retrieval and semantic integration in anaphor comprehension. We use two-sentence mini-discourses in which the variables givenness (old/anaphor vs. new/non-anaphor) and coherence (discourse-coherent vs. discourse-incoherent) are orthogonally manipulated in a two-by-two design. An example of a stimulus item is given in Table 1. The first sentence of each mini-discourse introduces two entities by name (e.g., John and Peter). In the second sentence, a critical proper name was used either anaphorically, referring back to one of the two previously mentioned entities ('old'; e.g., John when John and Peter are introduced), or non-anaphorically, introducing a new entity into the discourse model ('new'; e.g., John when David and Peter are introduced). In addition, the interpretation of the critical proper name in the second sentence is either coherent or incoherent with respect to the preceding discourse. In order to examine the brain's response to a new, non-anaphoric proper name when there are no specific referents in the preceding discourse, we added a neutral condition in which a new proper name could be anaphorically linked to a non-specific, generic antecedent in the preceding

Table 1.
Example of one stimulus item, containing all five conditions of an original Dutch two-sentence mini-discourse. Approximate English translations of each sentence are provided below. The critical proper names (CW) are in bold. Characteristic information that was manipulated is underlined.

	Sentence 1	Sentence 2
Old-coherent	**Jan** en Peter zijn de beste spelers uit het voetbalelftal. *(**John** and Peter are the best players in the football team.)*	De topscorer van het team was **Jan** met dertig doelpunten in totaal. *(The top scorer of the team was **John** with thirty goals in total.)*
Old-incoherent	**Jan** en Peter zijn de slechtste spelers uit het voetbalelftal. *(**John** and Peter are the worst players in the football team.)*	
New-coherent	David en Peter zijn de slechtste spelers uit het voetbalelftal. *(David and Peter are the worst players in the football team.)*	
New-incoherent	David en Peter zijn de beste spelers uit het voetbalelftal. *(David and Peter are the best players in the football team.)*	
New-neutral	De spelers in het voetbalelftal zijn erg goed. *(The players in the football team are very good.)*	

sentence (e.g., the players in the football team), and is neutral with respect to the coherence manipulation. This additionally allowed us to investigate Barkley et al.'s (2015) claim that proper names do not trigger back association. To be specific, if proper names do not trigger back association, no differences should be observed between new names for which a (generic) discourse referent is available (new-neutral) and new names that do not have an antecedent and therefore must introduce an entity into the discourse (new-coherent).

Hypotheses

For the ERP analysis, we expect that name repetition would elicit a biphasic ERP pattern: compared to repeated names, new names are expected to elicit a larger N400 and a subsequent LPC (e.g., Burkhardt, 2006). Similar to Wang and Yang (2013), we expect to observe a smaller N400 for discourse-coherent than discourse-incoherent proper names. A possible interaction effect might reveal itself in two ways. First, an incoherent discourse context could override the facilitatory effects of repetition (e.g., Camblin et al., 2007a; Ledoux et al., 2007). As a result, there will be no attenuation of the N400 for old-incoherent proper names. Alternatively, under the assumption that a possible N400 in our experiment reflects reactivation of an item in working memory, we expect the N400 in response to new names to be insensitive to discourse coherence. That is, a coherent discourse cannot not facilitate the reactivation of a new name, simply because a new name does not yet have a representation in working memory. Concerning the new-neutral condition, if proper names trigger back association, we expect an Nref effect for new-neutral compared to old-coherent proper names. No difference is expected between new-neutral and new-coherent, as both conditions contain a reference group to which the new proper name can be linked. However, if proper names do not trigger back association, as assumed by Barkley et al. (2015), no Nref effect is expected between new-neutral proper names and the coherent conditions.

We investigated oscillatory dynamics in three frequency ranges: theta (4-7 Hz), low gamma (34-45 Hz) and high gamma (60-80 Hz). Based on the observations of an increase in theta-band synchronization for old compared to new words (Burgess & Gruzelier, 1997, 2000; Klimesch et al., 1997, 2000), we tentatively hypothesize that increased theta-band synchronization reflects retrieval from working memory. These reactivation processes might also be linked to the gamma frequency band, as suggested by Nieuwland and Martin (2017). We therefore predict an increase in both theta- and low gamma-band power in response to old compared to new proper names. In line with the semantic unification literature (Bastiaansen & Hagoort, 2006, 2015), we expect to see an increase in high gamma-band power for discourse-coherent compared to discourse-incoherent proper names.

Methods

Preregistration

The design and settings of our analysis procedures (i.e., preprocessing, time-frequency analysis, and statistical analysis) have all been preregistered at Open Science Framework[1]. All non-preregistered analyses are exploratory and will explicitly be referred to as such.

Participants

A total sample size of 40 participants, after exclusions, was preregistered. In total, 45 native speakers of Dutch participated in the experiment. They were paid 18 euros for participation. All of them had normal or corrected-to-normal vision and none of them reported being dyslexic. Three participants had to be excluded after we found out that they were left-handed. Two additional participants had to be excluded from ERP analysis because too many trials had been rejected during ERP preprocessing. Similarly, after preprocessing the oscillatory data one participant ended up with too few trials and had to be excluded from the time-frequency analysis. Therefore, ERP analysis was done on 40 right-handed speakers of Dutch (30 females, average age: 23 years, age range: 19-33 years), while 41 right-handed speakers of Dutch (29 females, average age: 23 years, age range: 19-33 years) were included in the time-frequency analysis.

Materials

Experimental items

A total of 225 two-sentence mini-discourses were created. The second sentence of a given mini-discourse was identical for all conditions, while the first sentence varied between the conditions. The first sentence introduced two people by proper names within a conjoined noun phrase (e.g., John and Peter). This type of embedding has been shown to reduce the prominence of both referents, making subsequent use of a repeated name felicitous. The second half of the first sentence added characteristic information about these two people. This information could regard a social or personality characteristic (e.g., being friendly/unfriendly), a physical characteristic (e.g., being strong/weak) or a behavioral characteristic (e.g., getting good/bad grades). In addition, the first sentence always contained 'a reference group' to which both players belong (e.g., the football team). This was added to make sure the second sentence is not interpreted as referring to a situation in which only two entities are present. In the second sentence, a proper name (the critical word; CW) was used either anaphorically or non-anaphorically. Anaphoric use of the critical proper name represented the 'old' condition, because the exact same name had been introduced in the first sentence, whereas non-anaphoric use of the critical proper name represented the 'new' condition. The factor coherence ('coherent' vs. 'incoherent') was manipulated by using antonyms to denote the characteristic information across conditions within each item (e.g., friendly-unfriendly, getting good grades-getting bad grades). This made the interpretation of the critical proper name in the second sentence either coherent or incoherent with the characteristic information that was assigned to this proper name in the first sentence. In the fifth, new-neutral condition, the first sentence was kept semantically similar to the other conditions, with the exception that the proper names were removed and the reference group (e.g., the players in the football team) had become the subject of the sentence. This fifth condition was called new-neutral, because the proper name in the second sentence introduces a new entity (albeit interpretable as belonging to the reference group) and its interpretation in the second sentence is neutral with respect to the characteristic information provided in the first sentence. To avoid

[1] The project is called 'Proper names in discourse' and can be reached via https://osf.io/nbjfm/

sentence-final wrap-up effects contaminating the brain's response in the time window of interest, the CW was always followed by five words that ended the sentence without (anaphorically) referring to any of the entities.

The factors givenness (old, new) and coherence (coherent, incoherent) were orthogonally manipulated in a within-subjects two-by-two design, rendering the critical proper name in the second sentence old and coherent, new and coherent, new and incoherent or old and incoherent.

Approximately half (113) of the items only contained male names, the other half (112) contained only female names. All names were unambiguously either a male or female name. Each name was used only once in the entire experiment, meaning that we used 675 proper names (225 items x 3 names) in total.

To control for potential effects of order of mention, half of the anaphoric proper names in the old conditions referred to the first-mentioned name in the first sentence (John in the examples in Table 1), while the rest of the time it referred to the second-mentioned name in the first sentence (Peter in the examples in Table 1).

Filler items

We added 25 filler items to prevent participants from becoming too familiar with incoherent items. The first sentence of each filler item was very similar to the first sentence in the new-neutral condition. It contained a reference group (e.g., the candidates in the elections) with characteristic information (e.g., being popular). The second sentence was very similar to the second sentence in all experimental items, except for the fact that the proper name was replaced by a definite noun phrase referring to a specific person that belongs to the reference group (e.g., the politician). A translated example of a filler item: *The candidates in the elections are very popular. The majority of the people voted for the politician with the extraordinarily creative ideas.*

Comprehension questions

Eighty comprehension questions (50% with correct answer 'yes', 50% with correct answer 'no') were included to ensure that participants paid attention. These had to be answered by means of a button press. Questions were either about the general gist of the discourse story (50/80) or about specific entities in the stories (30/80). The average percentage of correctly answered questions was 92,4%. None of the participants scored below the preregistered cutoff of 70%.

Experimental lists

Five lists were made, each containing one condition of an experimental item. This was done to ensure that participants only saw one condition of each item. All lists had the same number of items of each condition. The filler items were the same for all lists. For each of the five lists, two versions were created by pseudorandomizing the order of the trials, with the only restriction that the same condition was never presented more than three times in a row. The participants were equally divided over all ten lists.

Procedure

Participants were comfortably seated in front of a computer screen in a fully shielded soundproof booth. After they had been informed about the procedure of the experiment, they were instructed to attentively and silently read sentences for comprehension and answer the comprehension questions. The sentences were presented in black letters (font Times New Roman, size 34) at the center of the screen, which had a light grey background.

Each trial started with a fixation cross (+) presented at the center of the screen. When participants pressed a button, the first sentence of each mini-discourse would be presented as a whole. After participants had carefully read the sentence, they pressed a button to start the presentation of the second sentence. This second sentence was presented word-by-word, which allowed us to control the time point at which the critical word was presented. Each word was presented for a duration of 400 ms, with the exception of the sentence-final word, which had a duration of 800 ms. The inter-word-interval was 200 ms in length. The sentence-final word was either followed by a fixation cross, indicating the start of the next trial, or a comprehension question. Participants were asked to minimize eye blinks and body movements during the word-by-word presentation of the second sentence.

The 250 (225 experimental + 25 filler) items were presented in five blocks of 50 items. Each block contained nine items of each condition and five filler items. Sixteen items per block were followed by a comprehension question. Participants were allowed to take short breaks between blocks. In total, the experiment lasted approximately 70 minutes (excluding EEG set-up time).

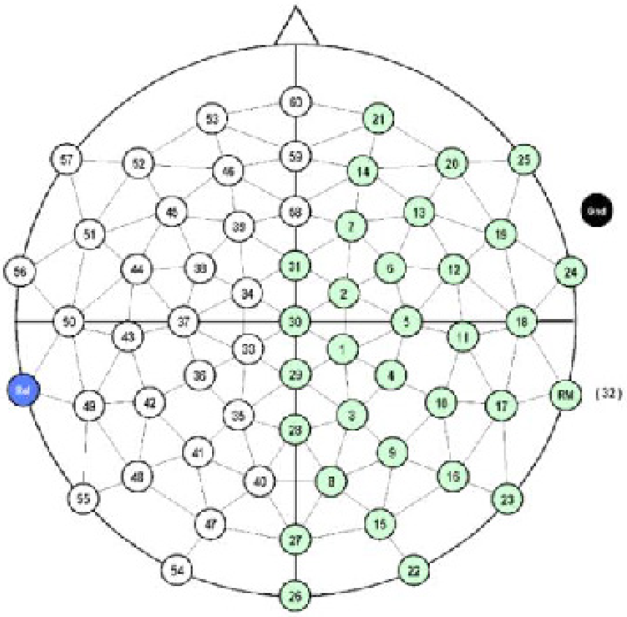


Fig. 1. Schematic representation of the 59-electrode array layout.

EEG recording and preprocessing

The electroencephalogram (EEG) was recorded using an MPI custom actiCAP 64-electrode montage (Brain Products, Munich, Germany), of which 59 electrodes were mounted in the electrode cap (see Figure 1). Horizontal eye movements (horizontal EOG) were recorded by one electrode placed on the outer canthus of the right eye, and eye blinks (vertical EOG) were recorded by two electrodes placed below both eyes. One electrode was placed on the right mastoid, the reference electrode was placed on the left mastoid and the ground was placed on the forehead. The EEG signal was amplified through BrainAmp DC amplifiers, referenced online to the left mastoid, sampled at 500 Hz and filtered with a passband of 0.016-249 Hz.

For ERPs, the data was band-pass filtered at 0.03-40 Hz (24 db/oct). Then, the data was re-referenced to the average of the left and right mastoid. Epochs were created ranging from -500 to 1500 ms relative to CW onset, and these were normalized to a 250 ms pre-CW baseline. Bad trials, containing low-frequency drifts, spikes or line noise were rejected through visual inspection. Independent Component Analysis (using ICA weights from a 1 Hz high-pass filtered version of the data) was used to filter artefacts resulting from eye movements and steady muscle activity. Epochs containing voltage values exceeding ± 90 μV were automatically rejected. On average, 13.5 ERP segments (average per condition ranged from 1.9 to 2.3) per subject were rejected. After preprocessing, two participants ended with less than 160 trials and were replaced.

For oscillations, the data was band-pass filtered at 0.1-100 Hz (24 db/oct) and re-referenced to the average of the left and right mastoid. Then, the data was segmented into epochs ranging from -1000 to 2500 ms relative to CW onset. Bad trials were again rejected through visual inspection. Independent Component Analysis (using ICA weights from a 1 Hz high-pass filtered version of the data) was used to filter artifacts resulting from eye movements and steady muscle activity. Because we did not apply baseline correction on the oscillatory data, using the preregistered automatic artifact rejection procedure based on an amplitude criterion (of ± 100 μV) would have excluded too many (good) trials. Therefore, we used a difference criterion that excluded segments in which the difference between the maximum and minimum voltage exceeded 200 μV. On average 12.1 segments (average per condition ranged from 2.1 to 2.5) per subject were rejected. One participant ended with less than 160 trials and was replaced.

Time-frequency (TF) analysis of oscillatory power was performed using the Fieldtrip toolbox (Oostenveld, Fries, Maris & Schoffelen, 2011). In order to find a right balance between time and frequency resolution, we performed time-frequency analysis in two different, but partially overlapping frequency ranges. For the low (2-30 Hz) frequency range, power was extracted from each individual frequency by moving a 400-ms Hanning window with ± 5 Hz spectral smoothing along the time axis in time steps of 10 ms. For the high (30-90 Hz) frequency range, we computed power changes with a multitaper approach (Mitra & Pesaran, 1999) based on discrete prolate spheroidal (Slepian) sequences as tapers, with a 400 ms time-smoothing and a ± 5-Hz spectral-smoothing window, in frequency steps of 2.5 Hz and time steps of 10 ms. On each individual trial, power changes in the post-CW interval were computed as a relative change from a baseline period ranging from -500 to -250 ms relative to CW onset. Per subject, we computed average power changes for each condition separately.

Statistical analysis

Statistical analysis of ERPs

We performed a linear mixed effects analysis (Baayen, Davidson & Bates, 2008) in R (R Core Team, 2018), using the lme4-package (Bates, Maechler & Bolker, 2012). The analyses were done separately for the N400, LPC and Nref regions-of-interest (ROIs).

At the N400 ROI, the dependent variable was, the average voltage across the centroparietal electrodes 35, 28, 3, 41, 40, 8, 9, 47, 27, 15 in a 300-500 ms window after CW onset (based on the spatial and temporal characteristics of the N400; e.g., Kutas & Federmeier, 2011). At the LPC ROI, the dependent variable was average across the same centroparietal electrodes 35, 28, 3, 41, 40, 8, 9, 47, 27, 15 in a 500-1000 ms window after CW onset (e.g., Van Petten & Luka, 2012). Dependent variables of the N400- and LPC ROIs were computed separately for each trial and each participant. The predictors coherence and givenness were deviation coded. We started with a full model that included the main effects of givenness (new, old) and coherence (coherence, incoherent), as well as their two-way interaction as fixed effects terms. Subject and item were included as random effects. Following Barr, Levy, Scheepers and Tily (2013), we attempted to use a maximal random effects structure by including the interaction between givenness and coherence as by-subject and by-item random slope. Because these models did not converge, we tested the same models without random correlations. As this still led to non-convergence, we removed the interaction term from the random slope and tested models with the same predictors and a by-participant and by-item random intercept only. Models are specified in the footnote[2]. In order to locate the model with the best fit, we started from a full model that included the interaction term and both main effects, and reduced its complexity stepwise, by first removing the interaction and then the main effects. Models were compared using R's `anova()` function, and *p*-values below α = .05 were treated as significant.

At the Nref ROI, the dependent variable was the average voltage across the frontal electrodes 53, 60, 21, 46, 59, 14, 39, 58, 7 in a 300-1500 ms window after CW onset (e.g., Van Berkum et al., 2007). In two separate analyses, we tested the effect of condition, where condition either had the levels 'new-neutral' and 'new-coherent' or 'new-neutral' and 'old-coherent'[3]. Subject and item were entered as random effects, and we included condition as by-subject and by-item random slope. We compared the models with and without condition using R's `anova()` function. *P*-values below α = .05 were treated as significant.

Statistical analysis of oscillatory power

We used cluster-based random permutation tests (Maris & Oostenveld, 2007) to compare differences in oscillatory power across conditions. This non-parametric statistical test deals with the multiple comparisons problem by statistically evaluating cluster-level activity rather than activity at individual data points, and is based on the fact that effects in electrophysiological data tends to be clustered in time, space and frequency (Maris, 2012). By evaluating the test statistic of the multidimensional cluster it retains statistical sensitivity while controlling the false alarm rate.

In brief, the cluster-based random permutation test works as follows: first, by means of a two-sided dependent samples t-test we performed the comparisons described below, yielding uncorrected p-values. Neighboring data triplets of electrode, time and frequency-band that exceeded a critical α-level of .05 were clustered. Clusters of activity were evaluated by comparing their test cluster-level statistic (sum of individual *t*-values) to a Monte-Carlo permutation distribution that was created by computing the largest cluster-level *t*-value on 1000 permutations of the same dataset. Clusters falling in the highest or lowest 2.5th percentile were considered significant. We used the correct-tail option that corrects *p*-values for doing a two-sided test, which allowed us to evaluate *p*-values at α = .05.

The following comparisons had been preregistered: a contrast between old (average of old-coherent and old-incoherent) and new (average of new-coherent and new-incoherent) in the 4-7 Hz

[2] N400/LPC:
Model1: N400/LPC ~ givenness*coherence + (1 | subject) + (1 | item)
Model2: N400/LPC ~ givenness + coherence + (1 | subject) + (1 | item)
Model3: N400/LPC ~ givenness + (1 | subject) + (1 | item)
Model4: N400/LPC ~ coherence + (1 | subject) + (1 | item)
Model5: N400/LPC ~ (1 | subject) + (1 | item)

[3] Nref:
Model1: Nref ~ condition + (condition|subject) + (condition|item)
Model2: Nref ~ (condition|subject) + (condition|item)

(theta) frequency range in a 0-1000 ms time window and in the 35-45 Hz (low gamma) frequency range in a 400-600 ms time window. Coherent (average of old-coherent and new-coherent) was compared to incoherent (average of old-incoherent and new-incoherent) in the 60-80 Hz (high gamma) frequency range in a 500-1000 ms time window. As the cluster-based permutation test is designed to compare two conditions at a time, we tested for an interaction effect by comparing the difference between coherent and incoherent in the old condition to the same difference in the new condition.

Non-preregistered analyses

TF analysis of new-neutral proper names

We performed exploratory time-frequency analysis on the conditions new-neutral, new-coherent and old-coherent in the same way as described in the section *EEG recording and preprocessing.* A cluster-based permutation test was used to compare new-neutral to both new-coherent and old-coherent in the 4-7 Hz (0-1000 ms), 35-45 Hz (400-600 ms) and 60-80 Hz (500-1000 ms) frequency ranges. Settings for the permutation test were identical to those described in the section *Statistical analysis of oscillatory power.*

TF analysis of ERP signal

To rule out the possibility that event-related potential activity contaminated our time-frequency analyses, we performed time-frequency analysis on the ERP signal that was obtained after within-subject averaging (see Wang et al., 2016 for a similar approach). As discussed in the introduction, ERPs contain phase-locked activity only, whereas time-frequency data contains both phase-locked and non-phase-locked activity. If the time-frequency results are driven by phase-locked activity, TF analysis of the ERP signal is expected to show a similar pattern as the TF analysis that was based on single-trial data. TF analysis of the subject-averaged ERP data was done in a similar way as described in the section *EEG recording and preprocessing.* A cluster-based permutation test was then used to compare the differences between old and new within the 4-7 Hz frequency range and a 300-500 ms time window.

Beamformer source localization

In an attempt to identify the sources underlying the observed differences in the 4-7 Hz theta and 60-80 Hz gamma oscillatory power, we applied a beamformer technique called Dynamical Imaging of Coherent Sources (Gross et al., 2001). This method uses a frequency-domain implementation of a spatial filter to estimate the source strength at a large number of previously computed grid locations in the brain. These grid locations are points in a three-dimensional grid that forms a discretized representation of a brain volume. Because the increase in 4-7 Hz theta activity for old compared to new names was most pronounced between 300-500 ms post-CW onset, this time period was subjected to source reconstruction. Following Nieuwland and Martin (2017), the increase in 60-80 Hz gamma activity for coherent compared to incoherent names was analyzed within a 500-1000 ms interval post CW-onset. The procedure and settings of the beamformer approach are adopted from Nieuwland and Martin (2017).

In addition to these condition-specific time windows, we extracted the data of all conditions in a 500-300 ms pre-CW baseline window. All data was re-referenced to the average of all electrodes. We identified the theta-activity to be most prominent between 4-7 Hz and therefore performed time-frequency analysis on 5 Hz, using a Hanning taper with ± 2 Hz spectral smoothing. In the gamma time window, we estimated power at 70 Hz, using a Slepian sequence taper with ± 10 Hz spectral smoothing.

We aligned the electrode positions of the montage to a standard Boundary Element Method (BEM) head model, which is a volume conduction model of the head based on an anatomical MRI (magnetic resonance imaging) template (Oostenveld, Praamsta, Stegeman & van Oosterom, 2001). This head model was subsequently discretized into a three-dimensional grid with a 5 mm resolution, and for each grid point an estimation of source power was calculated. For the 5 Hz and the 70 Hz frequencies-of-interest separately, a common inverse filter was computed on the basis of the combined dataset containing the pre-CW and post-CW intervals of both conditions (i.e., old-new for 5 Hz, coherent-incoherent for 70 Hz), which was then separately applied to all trials of each condition in order to estimate source power. After averaging over trials, we computed the difference between post-CW and pre-CW activity for each condition separately in the following way: (post-CW - pre-CW)/pre-CW. In order to visualize the estimated activity, we computed grand averages over subjects and subsequently interpolated the grid of the estimated power values to the anatomical MRI.

These estimates of source power were subjected to statistical analysis by means of a cluster-based permutation test (see section *Statistical analysis of oscillatory power*). On each source location in the three-dimensional grid we performed a one-sided dependent samples t-test (at α = .05, yielding uncorrected *p*-values) on trial-averaged data of respectively old and new (5 Hz) and coherent and incoherent (70 Hz). Neighboring grid points with significant *t*-values were clustered. A cluster-level test statistic was calculated by summing the individual *t*-values within each cluster and evaluated relative to a permutation distribution that was based on 1000 randomizations of the same dataset. In order to localize the spatial coordinates of the areas exhibiting significant differences, we interpolated only the *t*-values of the significant, clustered source points to the anatomical MRI. We identified brain areas using a template atlas (Tzourio-Mazoyer et al., 2002).

Results

ERP analysis

At the N400 region-of-interest, the effect of coherence was similar in the old and new conditions, $\chi^2 = 0.97$, $p = .33$. In addition, coherent and incoherent names elicited a similar N400 (mean difference = 0.20, *SE* = 0.19), $\chi^2 = 1.16$, $p = .28$. New names elicited a more negative N400 than old names (mean difference = 2.14, *SE* = 0.19), $\chi^2 = 127.45$, $p < .001$.

At the LPC region-of-interest, the effect of coherence was similar in the old and new conditions, $\chi^2 = 3.01$, $p = .08$. In addition, coherent and incoherent names elicited a similar LPC (mean difference = 0.12, *SE* = 0.19), $\chi^2 = 0.41$, $p = .52$. New names elicited a more positive LPC than old names (mean difference = 0.73, *SE* = 0.19), $\chi^2 = 14.03$, $p < .001$. Figure 2 contains the ERPs and corresponding scalp distributions of the difference between old and new (2A) and coherent and incoherent (2B).

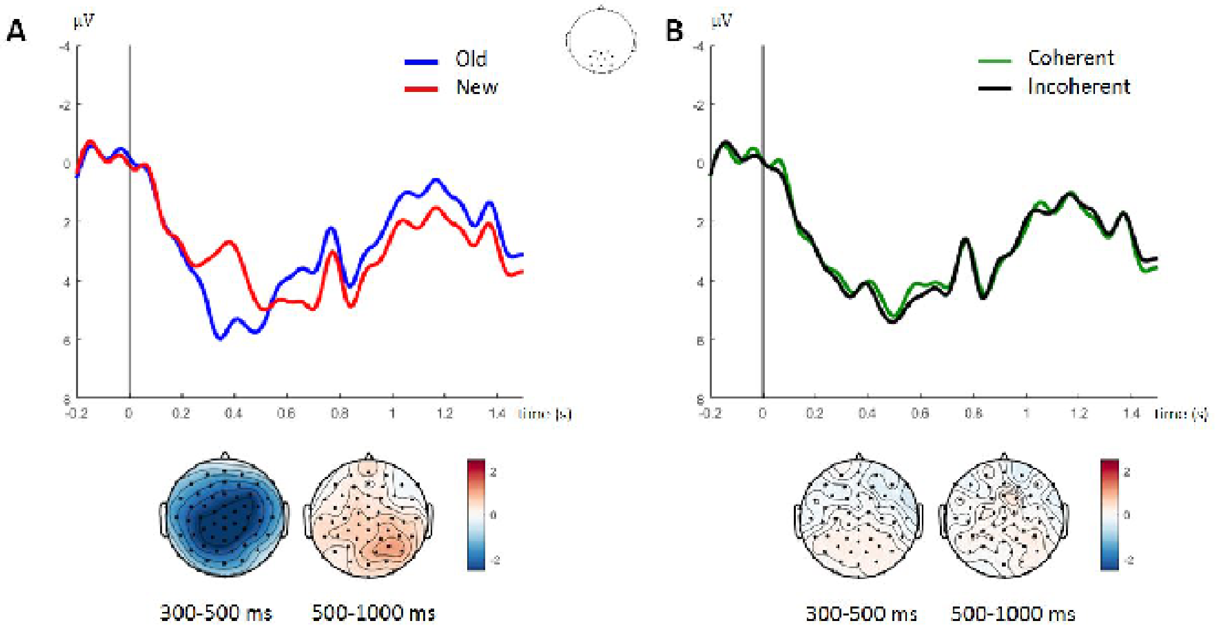


Fig. 2. N400 (300-500 ms) and LPC (500-1000 ms) responses, averaged over the centroparietal cluster, as a function of (A) givenness and (B) coherence. Scalp topographies represent the difference between (A) old and new and (B) coherent and incoherent. ERPs are low-pass filtered at 10 Hz for presentation purposes only.

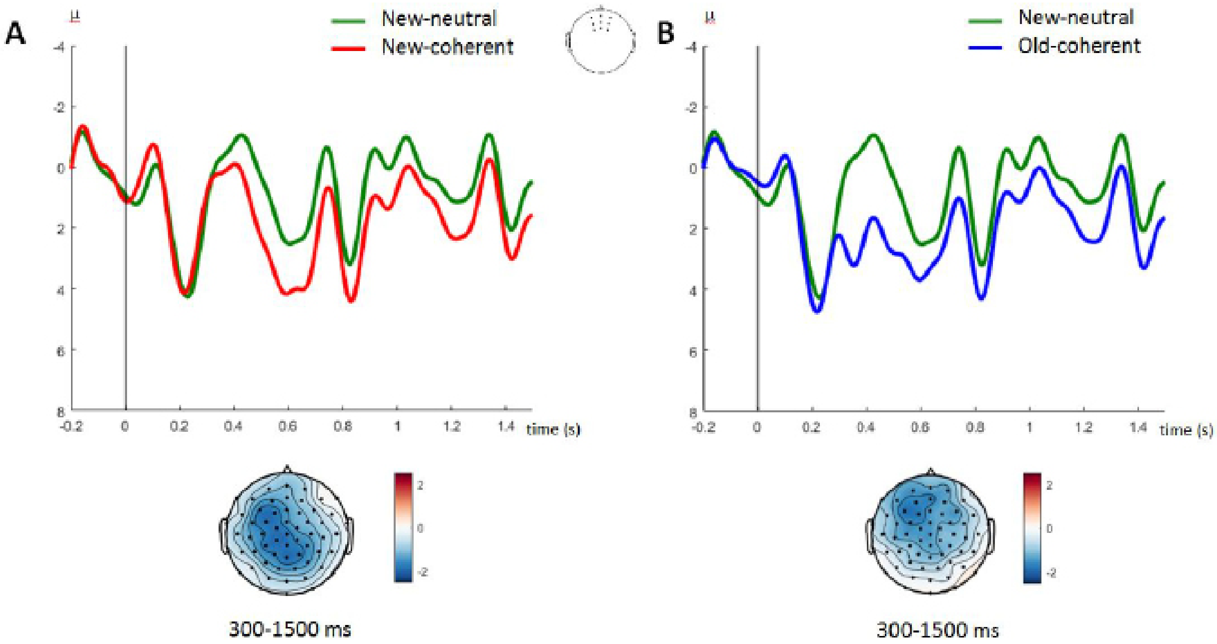


Fig. 3. Nref (300-1500 ms) responses, averaged over the frontal cluster, for the comparisons between new-neutral and new-coherent (A) and old-coherent (B). Scalp topographies reflect the difference between (A) new-neutral and new-coherent and (B) new-neutral and old-coherent. ERPs are low-pass filtered at 10 Hz for presentation purposes only.

At the Nref region-of-interest, the average ERP to new-neutral was significantly more negative than the average ERP to old-coherent (mean difference = 1.45, *SE* = 0.36), χ^2 = 14.61, p < .001. Similarly, the average ERP to new-neutral was significantly more negative than the average ERP to new-coherent (mean difference = 1.17, *SE* = 0.34), χ^2 = 10.46, *p* = .001. The ERPs and corresponding scalp distributions are shown in Figure 3.

TF analysis

In the 4-7 Hz (theta) frequency range, we found significantly larger theta-band power in the old compared to the new condition (*p* = .034). TF representations and scalp topography are presented

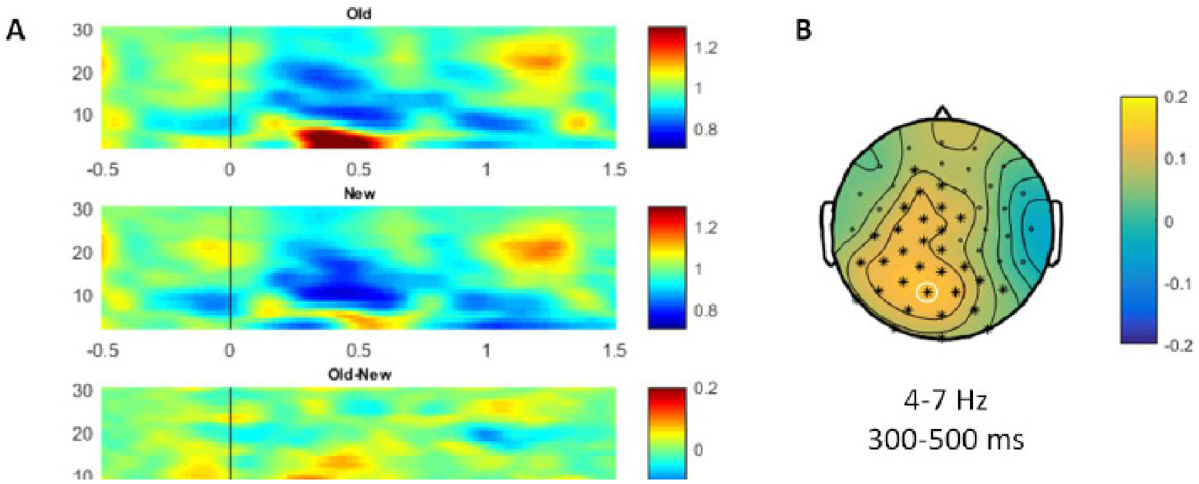


Fig. 4. Results of the TF analysis in the frequency range of 2-30 Hz. (A) TF representations (parietal-midline electrode 40) for old, new and the difference between old and new. (B) Topographical distribution of the 4-7 Hz theta effect in the 300-500 ms time window. Electrodes participating in the significant cluster are marked by an asterisk (*).

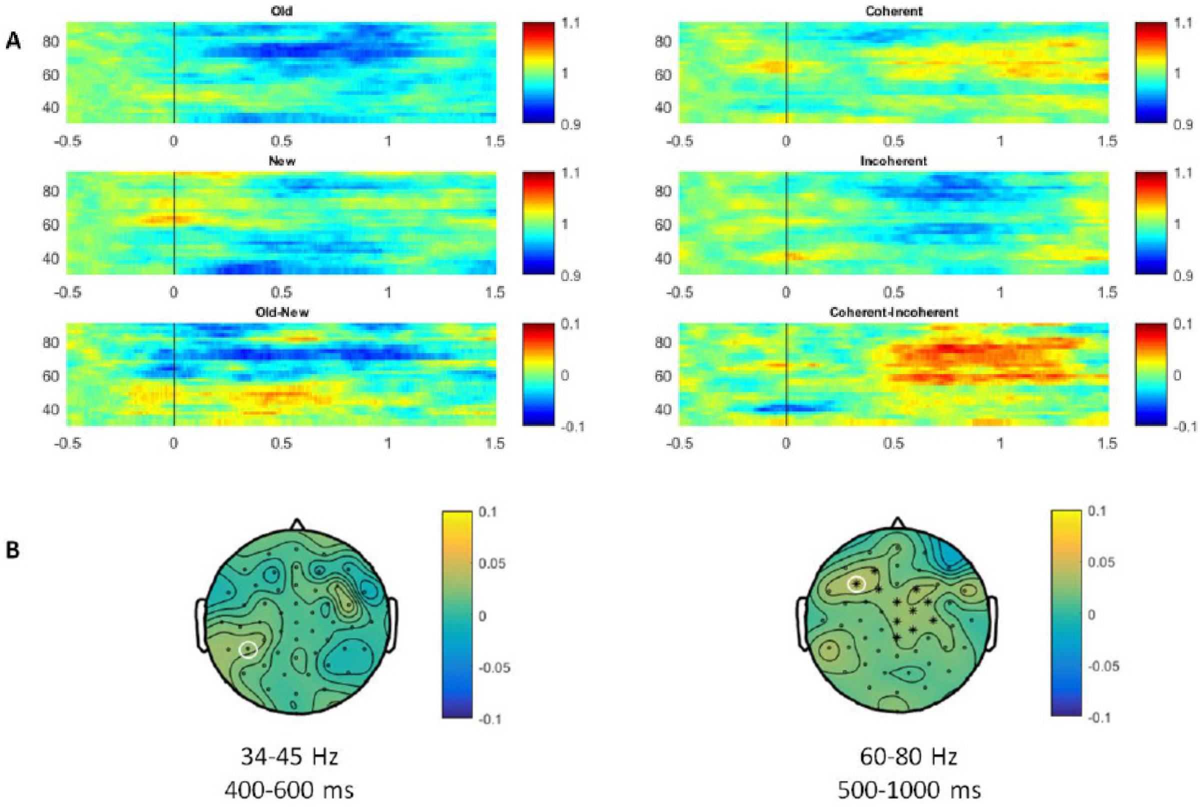


Fig. 5. Results of the TF analysis in the frequency range of 30-90 Hz. (A) The left panel shows the TF representations (left parietal electrode 42) of old, new and the difference between old and new. The right panel shows the TF representations (left frontal electrode 45) of coherent, incoherent and the difference between coherent and incoherent. (B) Topographical distributions for the 34-45 Hz old-new difference in the 400-600 ms time window (left) and for the 60-80 Hz coherent-incoherent difference in the 500-1000 ms time window (right). Electrodes participating in the significant cluster are marked by an asterisk (*).

in Figure 4.

In the 35-45 Hz (low gamma) frequency range, no significant clusters were observed for the contrast old-new (Figure 5a). In the 60-80 Hz (high gamma) frequency range, we observed significantly larger power in the coherent condition than in the incoherent condition (p = .04) within the preregistered time window of 500-1000 ms (Figure 5b). No significant clusters were observed for the interaction between givenness and coherence.

Exploratory analyses

TF comparisons of neutral and coherent conditions

In the 400-600 ms time window, old-coherent proper names elicited significantly stronger 35-45 Hz gamma-band synchronization than proper names in the new-neutral condition (p = .004). None of the other contrasts yielded significant differences. These results are visualized in Figure 6.

Beamformer source localization

A beamformer procedure was applied to localize the sources of the 4-7 Hz theta and 60-80 Hz gamma effects. We first applied a spatially unrestricted cluster-based permutation test to the power differences in the entire source space. This did not yield significant sources for the theta effect or the high gamma effect.

We therefore performed both literature-driven and data-driven exploratory region-of-interest analysis of the gamma effect (Figure 6A). Nieuwland and Martin (2017) localized the source of their high gamma activity to left frontal-temporal regions, encompassing inferior frontal lobe, inferior temporal lobe and anterior temporal lobe. We performed exploratory region-of-interest (ROI) analysis by restricting a cluster-based permutation test to these regions. It revealed a significant difference between coherent and incoherent, p = .047. Within this ROI, the difference was only significant in the left inferior frontal lobe (LIFG). However, as the effect seems to extend into dorsal regions of the frontal cortex

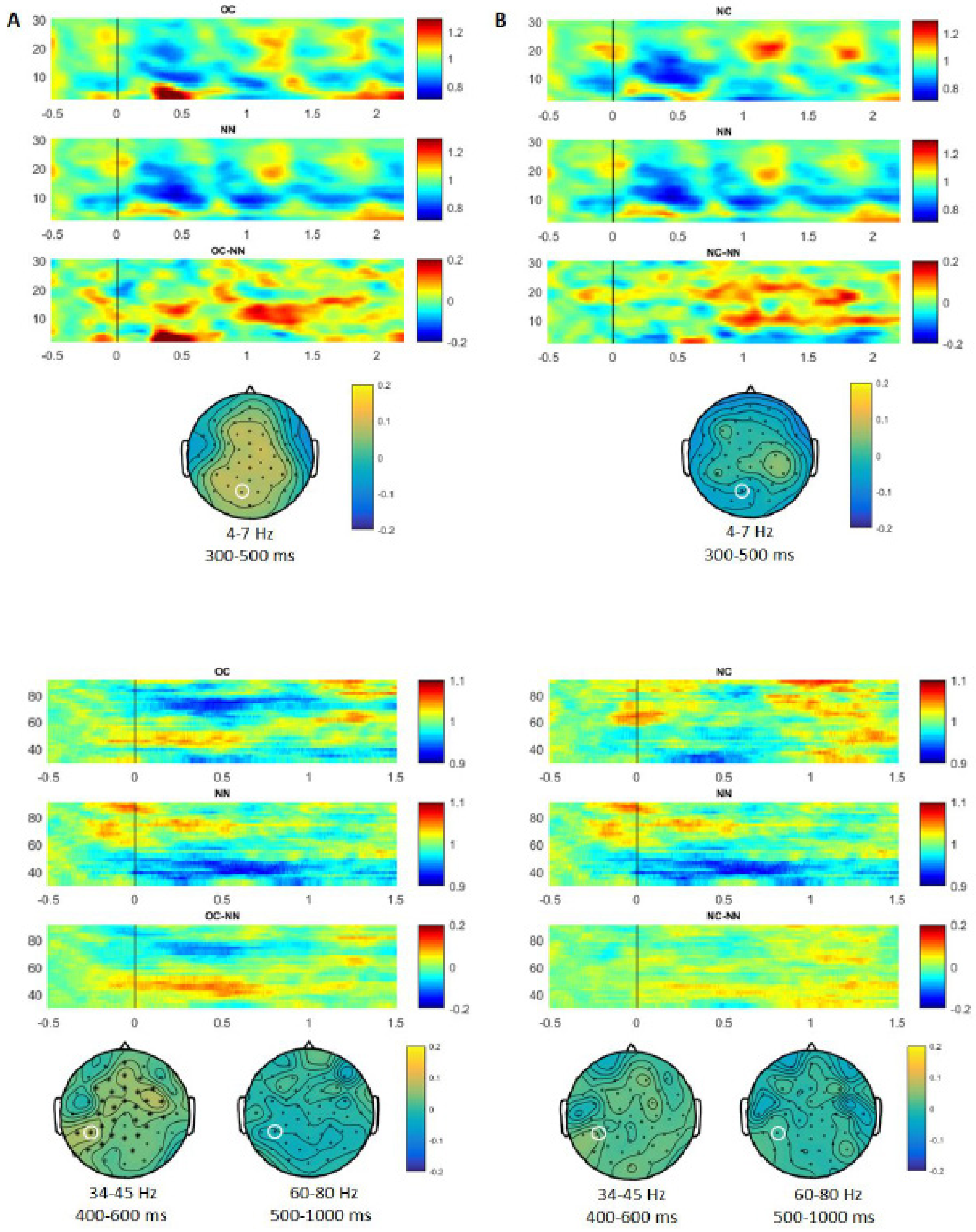


Fig. 6. Results of the TF analysis in the new-neutral (NN), old-coherent (OC) and new-coherent (NC) conditions. (A) TF representations of old-coherent, new-neutral and the difference between old-coherent and new-neutral, in both the low and high frequency ranges (parietal-midline electrode 40). (B) TF representations of new-coherent, new-neutral and the difference between new-coherent and new-neutral, in both the low and high frequency ranges (left parietal electrode 42). Below each TF representation are the topographical distributions of the respective differences. Electrodes participating in the significant cluster are marked by an asterisk (*).

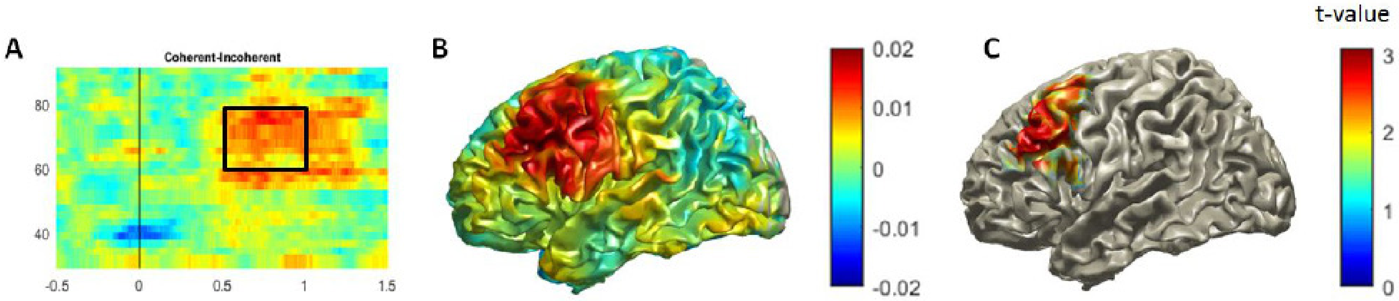


Fig. 7. Source localization results for the 60-80 Hz gamma effect. (A) TF representation of the difference between coherent and incoherent (black outline indicates the TF window of interest). (B) Surface plot of the power differences. (C) Surface plot of the ROI-based statistical results. Colors represent t-values, masked for significance.

(Figure 7B) and does not encompass any areas in the temporal lobe, we performed additional data-driven ROI analysis on the entire left frontal lobe in order to explore where the effect was strongest. Again, a significant difference between conditions was found (*p* = .027), in an area encompassing the LIFG and the left medial frontal gyrus. The source localization results are shown in Figure 7.

TF of ERP data

TF analysis of the ERP data revealed no significant differences between old and new (*p* > .1). Figure 8 presents the results of the TF analysis based on single-trial data and based on averaged ERP data. The TF representation of the ERP data seems to show a difference in power only up to 3 Hz, while we analyzed only frequencies within the frequency range of 4-7 Hz. Therefore, we believe that the theta old-new difference was not driven by phase-locked activity but rather represents 'true' oscillatory activity.

Discussion

In the present EEG study, we examined event-related potentials (ERPs) and neural oscillations in order to investigate the involvement of language and memory processes in anaphor comprehension. More specifically, we aimed to test the idea that gamma-band synchronization in response to coherent referential dependencies reflects the workings of the recognition memory and language networks. Subjects were presented two-sentence mini-discourses in which the interpretation of anaphoric (old/ repeated) and non-anaphoric (new) proper names was either coherent or incoherent with respect to the preceding discourse. As expected, we observed that in comparison to new names, repeated names elicited an attenuated N400 followed by a reduced Late Positive Component (LPC). Surprisingly, the ERPs in response to discourse-coherent and discourse-incoherent names did not show any differences. Proper names in the new-neutral condition elicited an Nref effect in comparison to both coherent conditions. In the time-frequency signal, we

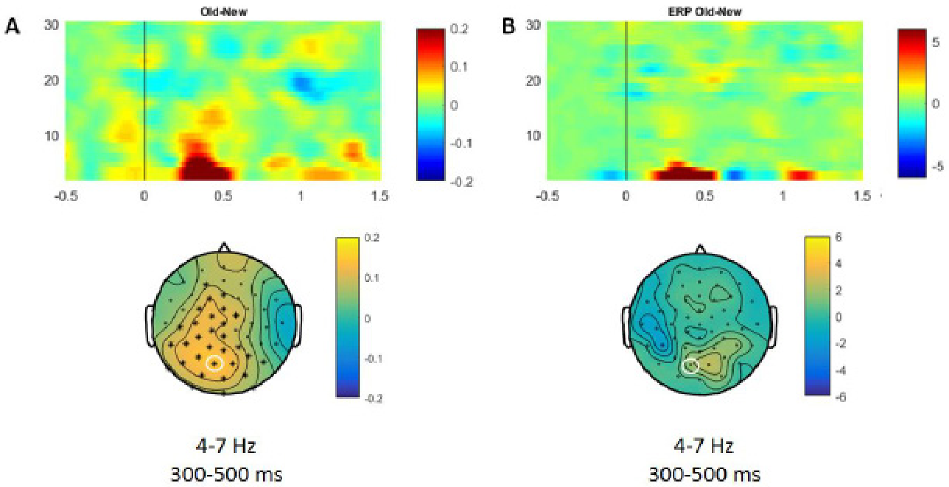


Fig. 8. Comparisons of TF representations (parietal-midline electrode 40) (A) on single trials and (B) on averaged ERP data. Topographical distributions of the 4-7 Hz average difference in the 300-500 ms time window are provided in the lower panel.

observed larger theta (4-7 Hz) power for repeated compared to new names. In the 60-80 Hz gamma range, we observed an increase in synchronization for coherent compared to incoherent proper names, which was most strongly associated with areas in the left frontal lobe. Exploratory analyses revealed an increase in 35-45 Hz gamma synchronization for old-coherent compared to new-neutral proper names.

Givenness and the N400-LPC complex

The expected biphasic ERP pattern observed for givenness is consistent with frameworks that view reference resolution as a two-stage process. These distinguish between a first stage at which the referent is lexically identified (i.e., reactivated from working memory), and a second stage at which the anaphor is integrated into the discourse model (Almor & Nair, 2007). In the current experiment, lexical identification of the referent of repeated proper names was facilitated because the antecedent had just been processed and was still available in working memory by the time the critical name was presented (see Wang & Yang, 2013 for a similar conclusion). The N400 effect in our experiment thus seems to be related to identifying the referent and reactivating its representation from working memory, a process that is facilitated for repeated names. New names do not reside in working memory and therefore elicit a large N400 component. The LPC might reflect updating of the discourse model by establishing an independent referential representation for the new name (Burkhardt, 2006, 2007; Schumacher, 2009; Schumacher & Hung, 2012; Wang & Schumacher, 2013; see also Kaan et al., 2007).

The Nref effect and dependency formation

Proper names in the new-neutral condition elicited a larger Nref than both the old-coherent and new-coherent proper names. Although the old-coherent names additionally elicited an attenuated N400 compared to new-neutral names, the fact that there was also an Nref for the comparison between new-neutral and new-coherent indicates that the difference between new-neutral and old-coherent is not solely due to the downstream consequences of name repetition. The Nref has been argued to reflect processes involved in resolving referential ambiguity (e.g., Nieuwland et al., 2007b; Van Berkum, 2009), but as the new-neutral names in our experiment are not genuinely ambiguous, we do not find this interpretation particularly compelling. Instead, we will argue that the Nref effects in response to our stimuli reflect difficulty in forming referential dependencies. Our interpretation derives from the cue-based retrieval framework (McElree, 2000, 2006; McElree et al., 2003), in which it is argued that the second element in a referential dependency (i.e., the anaphor) triggers reactivation of already encoded information that is held in working memory (i.e., the antecedent), which is addressable by virtue of content overlap between anaphor and antecedent (i.e., retrieval cues; Lewis & Vasishth, 2005; Lewis, Vasishth & Van Dyke, 2006). With respect to our stimuli in the new-neutral condition, we suggest that retrieval cues on the proper name trigger the reactivation of the reference group, which contains a sufficient amount of content overlap with the proper name (i.e., in terms of gender, animacy, etc.). However, because the match between anaphor and antecedent is not perfect, dependency formation does not run smoothly, as reflected in the Nref effect.

Evidence in favor of our interpretation of the Nref in terms of retrieval difficulty and its consequences for dependency formation comes from three recent ERP studies. First, Martin, Nieuwland and Carreiras (2012, 2014) investigated ERPs in response to elliptic determiners in Spanish, which either did or did not agree with their antecedent in terms of grammatical gender (e.g., … t-shirt$_{\text{FEM}}$ ... another$_{\text{MASC}}$). They additionally examined whether this process was modulated by the gender of a structurally unavailable local attractor noun. Both studies found that for fully grammatical sentences with unambiguous elliptic determiners the gender of the attractor modulated the amplitude of the Nref (albeit in opposite direction in both studies). These findings indicate that the gender of the attractor can interfere with ellipsis-based retrieval of the correct antecedent, even in the absence of referential ambiguity, and that the Nref might be an electrophysiological correlate of attempted retrieval and subsequent dependency formation. Similarly, Karimi, Swaab and Ferreira (2018) showed that retrieval difficulty, modulated as a function of the representational richness of antecedents (e.g., 'the actor' vs. 'the actor who was visibly upset') modulates the amplitude of the Nref. In all, this suggests that the Nref in our experiment represents difficulty establishing a referential dependency. The observation that such difficulty is also perceived in response to proper names provides evidence against Barkley et al.'s (2015) proposal that proper names do not trigger back association.

One caveat to our interpretation is that the ERPs in response to the new-coherent proper names did not show signs of referential dependency formation, although these conditions also contained a reference group (e.g., David and Peter are the worst players *in the football team*) to which the new proper name could have been linked. Two differences between the new-neutral and new-coherent condition might explain this result. That is, in the context sentence of the neutral condition, the reference group was the grammatical subject and it did not contain proper names. These differences might be relevant, because both factors have been shown to increase the discourse prominence of the denoted referent (Gordon & Hendrick, 1998; Gordon et al., 1999; Sanford, Moar & Garrod, 1988). As a result, the reference group in the new conditions might not have been accessible enough to be considered available for co-reference. We are currently designing a follow-up experiment in which the neutral condition is adjusted such that the presence of proper names is manipulated, and the reference group is mentioned more explicitly and thereby possibly made more available for co-reference. Note that this does not affect our interpretation of the Nref in the new-neutral condition in terms of dependency formation, but merely aims to answer the question why this process was not triggered in the new-coherent condition.

Absence of ERP effects for discourse coherence

To our surprise, we did not observe an effect of discourse coherence on the N400. A first possible interpretation of the absence of a coherence effect could be that participants did not notice the incoherence because they were not engaged enough. During the post-experiment debriefing all participants reported to have noticed the discourse incoherence, they all scored very high on the comprehension questions (average percentage correct of 92%), and the Nref observed in the comparisons between the neutral and coherent conditions indicates that participants were engaged in co-reference processes. This suggests that participants were adequately paying attention and did notice the incoherence, at least for some items. Yet, in order to check whether the coherence manipulation was effective for each individual item, we are currently setting up a norming test in which an additional group of participants will be asked to rate the coherence of each discourse item (e.g., on a 5-point Likert scale, as done by Wang, Verdonschot & Yang, 2016).

Two alternative, but related explanations are offered for the absence of a coherence effect. First, the very strong repetition effects (i.e., having a peak-to-peak difference of approximately 3 μV) might have washed out any effects of coherence. In terms of the absence of any difference between old-coherent and old-incoherent, it is conceivable that name repetition produced a ceiling effect, such that the addition of a coherent discourse did not further reduce the N400. This does not have any bearing on the absence of an N400 difference between new-coherent and new-incoherent, which, assuming that the N400 reflects semantic retrieval, had been hypothesized to elicit a similar N400 anyway (Kutas & Federmeier, 2000, 2011; Van Berkum, 2009). Also note that this inseparability in the time domain is perfectly consistent with the fact that we observed coherence-induced changes in the high gamma range, which are separable from the oscillatory effects of givenness in the time-frequency domain by virtue of different frequency characteristics. An alternative possibility is that the link between the initially meaningless proper names and the associated information was not strong enough to immediately affect processing difficulty. The relative coherence of each item hinged on the strength of the association between proper name and characteristic information described in the first sentence, which in turn is highly dependent on how semantically constraining the sentence is. Word-learning studies have shown that the meaning of novel words can be acquired very fast, but only when these novel words are learned in a strongly constraining context from which their meaning can easily be derived (Borovsky, Kutas & Elman, 2010; Mestress-Missé, Rodriguez-Fornells & Münte, 2007). On a similar note, Wang and Yang (2013) used a two-sentence discourse context to set up an explicit contrast between two newly introduced discourse entities. Linguistic contrast is known to facilitate word learning (Au & Markman, 1987; Kupferborg & Ohlstain, 1996) and is likely to have affected the fast mapping between proper name and characteristic information in the Wang and Yang (2013) study too. On top of that, participants in their study had to judge the congruence of the whole discourse after each trial, making it essentially inevitable that the incoherence was noticed. It is certainly possible that the proper name's meaning in our experiment was not yet established enough to yield immediate processing difficulty in the case of a discourse-incoherent interpretation. This, too, is compatible with the fact that we observed coherence effects in the time-frequency signal. Event-induced oscillations do not necessarily have to be time-locked in order to be picked up by time-frequency analysis. As long as the latency variability is not too large and does not exceed the length of the taper used for convolution with the EEG time course (Tallon-Baudrey & Bertrand, 2000), time-frequency analysis can pick up induced oscillations, albeit in relatively time-smoothed appearance (Cohen, 2014). The fact that the high gamma effects seem to be smoothed in time, extending beyond the expected time window of 1000 ms (denoted by the black contour in Figure 7A), is compatible with this possibility.

Increased theta-band synchronization for repeated names

Compared to new names, given names elicited a widespread increase in theta-band synchronization that was most prominent over left parietal electrodes. Our time-frequency analysis of the ERP signal in the N400 time window showed that the theta effect was not driven by phase-locked activity, instead reflecting 'true' oscillatory activity (Wang et al., 2012). This is in agreement with the findings that the given/new ERP and theta effects are independent phenomena (Jacobs et al., 2006; Klimesch et al., 2000).

As briefly discussed in the introduction, one line of research has related theta oscillations in language processing to retrieval from semantic long-term memory (Bastiaansen et al., 2002, 2005, 2008), where theta synchronization is increased when retrieval is difficult (Hagoort et al., 2004; Hald et al., 2006). As repeated and new proper names place similar demands on retrieval from long-term memory, this is not likely to have caused the theta effects. If anything, retrieval of new names would be more difficult than retrieval of repeated names, suggesting an increase in theta for new compared to repeated names, which is the opposite of what we found.

Rather, we interpret the theta effect as reflecting successful retrieval from working memory. To reiterate, studies of recognition memory have found an increase in theta-band synchronization for correctly remembered targets compared to correctly rejected distractors (Burgess & Gruzelier, 1997, 2000; Chen & Caplan, 2016; Jacobs et al., 2006; Klimesch et al., 1997, 2000, 2006; Osipova et al., 2006; Van Strien, 2005, 2007), suggesting that theta oscillations might reflect a relational process that matches the probe to a representation held in working memory (Chen & Caplan, 2016; Jacobs et al., 2006). This is in line with the idea that theta oscillations also underlie retrieval of information in during online language processing (Covington & Duff, 2016; Duff & Brown-Schmidt, 2012; Meyer et al., 2015).

No influence of givenness on low gamma oscillations

Against our prediction, we did not observe a difference between repeated and new names in the 34-45 Hz (low) gamma range. Interestingly, however, exploratory analyses revealed that old-coherent proper names elicited larger low gamma synchronization than proper names in the new-neutral condition. It should be noted that our prediction was solely based on Nieuwland and Martin's interpretation of their gamma effects for coherent anaphors. They localized the origin of these effects to the left posterior parietal cortex (LPPC), an area that has been related to successful memory retrieval (e.g., Öztekin et al., 2008; Wagner et al., 2005) and the ability to make old/new judgments (Gonzalez et al., 2015). However, activity in this area is often related to aspects of memory that are arguably different from the memory mechanisms that underlie language comprehension. For instance, fMRI (functional magnetic resonance imaging) studies have indicated that activity in the LPPC relates to a subjective perception of memory strength (Hutchinson, Uncapher & Wagner, 2015) and the phenomenological experience of remembering (Wagner et al., 2005). In addition, the LPPC has been implicated in recovery of information about the temporal ordering of to-be-remembered items, which is thought to require a slow serial search operations (Öztekin et al., 2008). As argued before, there is good reason to believe that the memory mechanisms underlying language comprehension work via direct access rather than serial search (McElree, 2000, 2006; McElree et al., 2003). Combined with the absence of an old/new effect in the low gamma range, this suggests that language comprehension does not rely on the memory-preserving functions of the LPPC. This begs the question what the gamma effects do reflect. One commonality between our findings and those by Nieuwland and Martin is that both were elicited by a comparison that also elicited an Nref effect. Although the specific processes underlying both Nref effects are probably different (i.e., dependency formation vs. resolving referential ambiguity), this might indicate that the gamma effects are related to general processes involved in resolving a (linguistically) complicated situation. Further research is needed to find out whether the low gamma effects are specific to referential processes or whether they reflect domain-general cognitive mechanisms involved in complex tasks, as has been proposed for the LPPC (e.g., Chein, Ravizza & Fiez, 2003).

Increased high gamma synchronization for coherent names

In line with our expectations, we observed an increase in 60-80 Hz gamma-band synchronization in response to discourse-coherent compared to discourse-incoherent proper names. Literature-based exploratory ROI analysis revealed that the effect was generated by left frontal regions, encompassing the left inferior frontal gyrus (LIFG). Previous studies of semantic unification on discourse-level have shown that the language processor immediately uses information from both sentence-level and discourse-level sources (Van Berkum et al., 1999a, 2003) and integrates these in a 'single unification space', predominantly active in the LIFG (Hagoort & Van Berkum, 2007; Hagoort & Indefrey, 2014). Semantic unification on the sentence-level has been shown to modulate oscillations in the gamma band, whereby it is generally observed that synchronization is increased whenever semantic unification is successful (Bastiaansen & Hagoort, 2015; Hald et al., 2006; Peña & Melloni, 2012; Penolazzi, Angrili & Job, 2009). We provide converging evidence that combines these patterns: discourse-level manipulations of semantic unification modulate gamma oscillations in the LIFG.

It has recently been noted that these gamma-band modulations might be more easily explained in terms of prediction rather than semantic unification (Lewis & Bastiaansen, 2015; Lewis, Wang & Bastiaansen, 2015; Wang, Zhu & Bastiaansen, 2012). Bastiaansen and colleagues propose that a match between pre-activated representations and incoming language material translates into gamma-band synchronization. Specific to our stimuli, it could be argued that the discourse-coherent proper names were predictable and therefore elicited a gamma-band increase. However, new proper names were never predictable, whether coherent with the preceding discourse or not, suggesting that some of our conditions do not lend themselves for predictive processes. In addition, we localized the source of the gamma effect to the LIFG, which has been strongly linked to unification operations (Hagoort, 2005; Hagoort & Indefrey, 2015), and differentiates between discourse-coherent and incoherent anaphors (Hammer, Jansma, Tempelmann & Münte, 2011). Context-based predictive processes, in contrast, typically show up as activation in medial temporal regions (e.g., Lau & Nguyen, 2015). Last, no effect of coherence was found on the N400, which has been strongly linked to predictability (Kutas & Hillyard, 1984). Although one could argue that the absence of any coherence-related differences on the N400 also constitutes evidence against the view that our gamma findings reflect semantic unification, in the preceding paragraphs we suggested that the absence of coherence-related modulations of the N400 might have been related to time variability in the moment at which the incoherence was noticed. If preparatory processes had been able to preactivate the (old-)coherent proper names, one would expect the incoherence to be noticed as soon as the word had been recognized (i.e., predicted representations might be given a 'head-start' in activation; Lau & Nguyen, 2015), and not leading to a possible incoherence response that has a variable latency.

Thus, while previous studies have shown the contextually constraining effects of discourse on semantic unification in the time domain (i.e., modulations of the N400 amplitude; e.g., Van Berkum et al., 1999a, 2003), the present study is the first to demonstrate these effects of discourse coherence in the time-frequency domain.

Conclusion

The present EEG study used ERPs and neural oscillations to study the involvement of recognition memory and semantic unification in anaphor comprehension via respectively givenness and coherence. We manipulated givenness by utilizing the ability of proper names to introduce new reference and maintain old reference, and found that it modulates oscillatory activity in the theta band. Coherence of proper names was reflected in gamma-band synchronization. In all, our study is the first to show that givenness and coherence of discourse-level anaphors modulates oscillatory synchronization in separated frequency bands, thereby encouraging future time-frequency research into the role of memory mechanisms in discourse-level language comprehension.

References

Almor, A., & Nair, V. A. (2007). The form of referential expressions in discourse. *Language and Linguistics Compass, 1*(1-2), 84-99.

Almor, A., Nair, V. A., Boiteau, T. W., & Vendemia, J. M. C. (2017). The N400 in processing repeated name and pronoun anaphors in sentences and discourse. *Brain and Language, 173*, 52-66.

Au, T. K., & Markman, E. M. (1987). Acquiring word meanings via linguistic contrast. *Cognitive Development, 2*, 217-236.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390-412.

Barkley, C., Kluender, R., Kutas, M. (2015). Referential processing in the human brain: an event-related potential (ERP) study. *Brain Research, 1629*, 143-159.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278.

Bastiaansen, M., & Hagoort, P. (2006). Oscillatory neuronal dynamics during language comprehension. *Progress in brain research*, *159*, 179-196.

Bastiaansen, M., & Hagoort, P. (2015). Frequency-based segregation of syntactic and semantic unification during online sentence level language comprehension. *Journal of cognitive neuroscience*, *27*(11), 2095-2107.

Bastiaansen, M. C., Mazaheri, A., & Jensen, O. (2012). Beyond ERPs: Oscillatory neuronal dynamics. In S. J. Luck, & E. S. Kappenman (Eds.), *The Oxford handbook of event-related potential components* (pp. 31-50). New York, NY: Oxford University Press.

Bastiaansen, M., Oostenveld, R., Jensen, O., & Hagoort, P. (2008). I see what you mean: Theta power increases are involved in the retrieval of lexical semantic information. *Brain and Language, 106*, 15–28.

Bastiaansen, M., Van der Linden, M., ter Keurs, M., Dijkstra, T., & Hagoort, P. (2005). Theta responses are involved in lexicosemantic retrieval during language processing. *Journal of Cognitive Neuroscience, 17*, 530–541.

Bastiaansen, M., Berkum, J. J. A., & Hagoort, P. (2002). Event-related theta power increases in the human EEG during online sentence processing. *Neuroscience Letters, 323*, 13-16.

Bates, D. M., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes.

Bertrand, O., & Tallon-Baudry, C. (2000). Oscillatory gamma activity in humans: a possible role for object representation. *International Journal of Psychophysiology*, *38*(3), 211-223.

Borovsky, A., Kutas, M., & Elman (2010). Learning to use words: event-related potentials index single-shot contextual word learning. *Cognition, 116*, 289-296.

Burgess, A.P., & Gruzelier, J.H., (1997). Short duration synchronization of human theta rhythm during recognition memory. *NeuroReport, 8*, 1039–1042.

Burgess, A.P., Gruzelier, J.H. (2000). Short duration power changes in the EEG during recognition memory for words and faces. *Psychophysiology, 37*, 596–606.

Burkhardt, P. (2006). Inferential bridging relations reveal distinct neural mechanisms: evidence from event-related brain potentials. *Brain and Language, 98*, 159–168.

Burkhardt, P. (2007). The P600 reflects cost of new information in discourse memory. *NeuroReport, 18*, 1851–1854.

Camblin, C. C., Ledoux, K., Boudewyn, M., Gordon, P. C., & Swaab, T. Y. (2007a). Processing new and repeated names: effects of coreference on repetition priming with speech and fast rsvp. *Brain Research, 1146*, 172-184.

Camblin, C. C., Gordon, P. C., & Swaab, T. Y. (2007b). The interplay of discourse congruence and lexical association during sentence processing: evidence from ERPs and eye tracking. *Journal of Memory and Language, 56*, 103-128.

Chen, Y. Y., & Caplan, J. B. (2016). Rhythmic activity and individual variability in recognition memory: theta oscillations correlate with performance whereas alpha oscillations correlate with ERPs. *Journal of Cognitive Neuroscience, 29*(1), 183–202.

Chein, J. M., Ravizza, J. A., & Fiez, J. A. (2003). Using neuroimaging to evaluate models of working memory and their implications for language processing. *Journal of Neurolinguistics, 16*, 315-339.

Cohen, M. X. (2014). *Analyzing neural time series data*. Cambridge, MA: MIT Press.

Covington, N. V., & Duff, M. C. (2016). Expanding the language network: direct contributions from the hippocampus. *Trends in Cognitive Sciences, 20*(12), 869-870.

Davidson, D. J., & Indefrey, P. (2007). An inverse relation between event-related and time-frequency violation responses in sentence processing. *Brain Research, 1158*, 81-92.

Duff, M. C., & Brown-Schmidt, S. (2012). The hippocampus and the flexible use and processing of language. *Frontiers in human neuroscience*, *6*, 69.

Filik, R., & Leuthold, H. (2008). Processing local pragmatic anomalies in fictional contexts: evidence from the N400. *Psychophysiology, 45*, 554-558.

Gonzalez, A., Hutchinson, J. B., Uncapher, M. R., Chen, J., LaRocque, K. F., Foster, B. L. ..., & Wagner, A. D. (2015). Electrocorticography reveals the temporal dynamics of posterior parietal cortical activity during recognition memory decisions. *Proceedings of the National Academy of Sciences, 112*(34), 11067-11071.

Gordon, P. C., & Hendrick, R. (1998). The representation and processing of coreference in discourse. *Cognitive Science, 22*(4), 389-424.

Gordon, P. C., Hendrick, R., Ledoux, K., & Yang, C. L. (1999). Processing of reference and the structure of language: an analysis of complex noun phrases.

Language and cognitive processes, 14, 353-379.

Gross, J., Kujala, J., Hämäläinen, M., Timmermann, L., Schnitzler, A., & Salmelin, R. (2001). Dynamic imaging of coherent sources: Studying neural interactions in the human brain. *Proceedings of the National Academy of Sciences, 98*(2), 694–699.

Hagoort, P. (2005). On Broca, brain, and binding. *Trends in Cognitive Science, 9*, 416-423.

Hagoort, P., & Indefrey, P. (2014). The neurobiology of language beyond single words. *Annual Review of Neuroscience*, *37*, 347-362.

Hagoort, P., Baggio, G., & Willems, R. M. (2009). Semantic unification. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences, 4th ed.* (pp. 819-836). Cambridge, MA: MIT Press.

Hagoort, P., & Van Berkum, J. J. A. (2007). Beyond the sentence given. *Philosophical Transactions of the Royal Society B, 362,* 801-811.

Hagoort, P., Hald, L., Bastiaansen, M., & Petersson, K. M. (2004). Integration of word meaning and world knowledge in language comprehension. *Science*, *304*(5669), 438-441.

Hald, L. A., Bastiaansen, M. C., & Hagoort, P. (2006). EEG theta and gamma responses to semantic violations in online sentence processing. *Brain and Language*, 96, 90–105.

Hammer, A., Jansma, B. M., Tempelman, C., & Münte, T. F. (2011). Neural mechanisms of anaphoric reference as revealed by fMRI. *Frontiers in Psychology, 2*(32).

Hutchinson, J. B., Uncapher, M. R., & Wagner, A. D. (2015). Increased functional connectivity between dorsal posterior parietal and ventral occipitotemporal cortex during uncertain memory decisions. *Neurobiology of learning and memory*, *117*, 71-83.

Jacobs, J., Hwang, G., Curran, T., & Kahana, M. J. (2006). EEG oscillations and recognition memory: Theta correlates of memory retrieval and decision making. *NeuroImage, 32*, 978-987.

Kaan, E., Dallas, A. C., & Barkley, C. M. (2007). Processing bare quantifiers in discourse. *Brain Research, 1146*, 199-209.

Karimi, H., Swaab, T., & Ferreira, F. (2018). Electrophysiological evidence for an independent effect of memory retrieval on referential processing. *Journal of Memory and Language, 102*, 68-82.

Kielar, A., Panamsky, L., Links, K. A., & Meltzer, J. A. (2015). Localization of electrophysiological responses to semantic and syntactic anomalies in language comprehension with MEG. *NeuroImage, 105*, 507-524.

Klimesch, W., Doppelmayr, M., Schimke, H., & Ripper, B. (1997). Theta synchronization and alpha desynchronization in a memory task. *Psychophysiology, 34*, 169–176.

Klimesch, W., Doppelmayr, M., Schwaiger, J., Winkler, T., & Gruber, W. (2000). Theta oscillations and the ERP old/new effect: independent phenomena? *Clinical Neurophysiology, 111*, 781–793.

Kupferborg, I., & Olshtain, E. (1996). Explicit contrastive instruction facilitates the acquisition of difficult L2 forms. *Language Awareness, 5*(3/4), 149-165.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: brain potentials reflect semantic incongruity. *Science, 207*, 203-205.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947), 161.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4*, 463-470.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62*, 621-647.

Lau, E. F., & Nguyen, E. (2015). The role of temporal predictability in semantic expectation: An MEG investigation. *Cortex*, *68*, 8-19.

Ledoux, K., Gordon, P. C., Camblin, C. C., & Swaab, T. Y. (2007). Coreference and lexical repetition: neural mechanisms of discourse integration. *Memory & Cognition, 35*, 801-815.

Lewis, R. L., & Vasishth, S. (2005). An activation‐based model of sentence processing as skilled memory retrieval. *Cognitive science*, *29*(3), 375-419.

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in cognitive sciences*, *10*(10), 447-454.

Lewis, A. G., & Bastiaansen, M. (2015). A predictive coding framework for rapid neural dynamics during sentence-level language comprehension. *Cortex*, *68*, 155-168.

Lewis, A. G., Wang, L., & Bastiaansen, M. (2015). Fast oscillatory dynamics during language comprehension: Unification versus maintenance and prediction? *Brain and Language*, *148*, 51-63.

Maris, E. (2012). Statistical testing in electrophysiological studies. *Psychophysiology, 49*(4), 549-65.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of neuroscience methods*, *164*(1), 177-190.

Martin, A. E. (2016). Language processing as cue integration: grounding the psychology of language in perception and neurophysiology. *Frontiers in Psychology, 7*(120).

Martin, A. E., Nieuwland, M. S., & Carreiras, M. (2012). Event-related brain potentials index cue-based retrieval interference during sentence comprehension. *Neuroimage, 59*, 1859-1869.

Martin, A. E., Nieuwland, M. S., & Carreiras, M. (2014). Agreement attraction during comprehension of grammatical sentences: ERP evidence from ellipsis. *Brain and Language, 135*, 42-51.

McElree, B. (2000). Sentence comprehension is mediated by content-addressable memory structures. *Journal of Psycholinguistic Research, 29*, 111-123.

McElree, B. (2006). Accessing recent events. In B. H. Ross (Eds.), *The psychology of learning and motivation: Vol.*

Beta-Band Desynchronization Reflects Competition Between Movement Plans of the Left and Right Hand

Milou van Helvert[1]
Supervisors: Leonie Oostwoud Wijdenes[1], Linda Geerligs[1], Pieter Medendorp[1]

[1]Radboud University Nijmegen, Donders Institute for Brain, Cognition and Behaviour, The Netherlands

Parallel processing of multiple movement plans enables a smooth interaction with the environment. It allows for rapid switching between response alternatives, which is crucial in threatening situations. The internal representations of these response alternatives are thought to compete during movement preparation. When preparing a reaching movement towards multiple possible target locations, beta-band desynchronization is modulated by this movement plan competition. We tested whether the same applies when multiple effectors are available and we sought for neural evidence of movement plan competition related to hand use. Behavioral evidence suggests that movement plans of the left and right hand compete during hand selection, as greater ambiguity with respect to the hand to use comes with a cost, reflected in slower reaction times and movement variability. We recorded brain activity with electroencephalography (EEG) while participants (n = 17) performed a speeded hand-selection reaching task. To estimate the effect of competition between movement plans for the left and right hand, trials were included during which the hand to use was predetermined and movement plan competition was thus thought to be minimal. We focused on event-related desynchronization (ERD) in the beta band in response to a cue marking the onset of movement preparation. Results indicate that beta-band ERD is indeed modulated by competition between movement plans of the left and right hand: for reaches to the point of subjective equality, a point in space where left and right hand use is equiprobable and movement plan competition is thus thought to be maximal, beta-band ERD was smaller when participants were free to select the hand to use than when the hand to use was predetermined. These results indicate that hand selection is based on a competitive process between movement plans for the left and the right hand and underline the idea of parallel processing of multiple movement plans simultaneously. These findings provide us with valuable insight into the way the brain processes information necessary to plan goal-directed movements.

Keywords: movement preparation, effector selection, hand choice, electroencephalography, beta-band desynchronization

Corresponding author: Milou van Helvert; **E-mail:** mvanhelvert@live.nl

Imagine having a picnic with friends and trying to reach for the orange juice. Which hand do you use to pick up the glass, your left or right hand? How does your brain make this decision? Complex environments, such as a picnic, provide us with extensive sensory information and give rise to many potential actions. The way this sensory information is processed to plan the execution of goal-directed movements has been a topic of debate. It was long believed that the brain processes information about the environment in a serial manner (Marr, 1982; Poggio, 1981). This serial processing view proposes that you first determine the goal of the movement based on sensory information, i.e., reach for and presumably drink the orange juice. After this, the brain computes a more detailed plan on how to achieve this goal by specifying, for example, which hand to use for the reaching movement and how fast and precise the movement should be.

More recently, Cisek (2007) put forward the affordance competition hypothesis. This hypothesis contradicts the serial processing view by proposing that the brain prepares multiple potential actions in parallel. The internal representations of these potential actions have been described as affordances (Gibson, 1979). The parallel processing of affordances enables continuous interaction in a complex environment and quick action in response to hazardous events (Cisek & Kalaska, 2010). Building on the affordance competition hypothesis, the brain might construct movement plans for multiple movements simultaneously. These movement plans can be based on the same computational goal, i.e., reaching for the orange juice. This goal can be achieved with different potential actions, i.e., reaches with the left hand or the right hand. The affordance competition hypothesis is underlined by the finding that having multiple action possibilities comes with a cost (Oostwoud Wijdenes, Ivry, & Bays, 2016). This cost is reflected in greater movement variability, suggesting that movement plans are indeed processed simultaneously but that processing capacity is limited. Due to this limited capacity, the parallel processing of information gives rise to a constant competition between the internal representations of potential actions, hence the name affordance competition hypothesis (Cisek, 2007; Cisek & Kalaska, 2010). This competition is eventually resolved when a certain movement plan prevails, resulting in movement.

Hand choice experiments have been used to test the affordance competition hypothesis (Oliveira, Diedrichsen, Verstynen, Duque, & Ivry, 2010). This decision process is often encountered in daily life, as most actions require moving one of the hands, and is mostly resolved unconsciously. In general, people use the hand ipsilateral to the reach goal (Bryden, Pryde, & Roy, 2000; Gabbard & Rabb, 2000). However, for reach goals close to the body midline people usually show a preference to use their dominant hand.

Oliveira et al. (2010) investigated whether hand choice evokes a competitive process between simultaneously prepared left and right hand movement plans. They hypothesized that competition between these movement plans is greatest when the decision uncertainty is greatest, i.e., when the evidence to use one hand over the other is most ambiguous. They found that reaction times of a reach were shorter if the hand to use was predetermined by the experimenter than if the hand to use was undetermined and the participants were thus free to choose the hand to use. Also, for the undetermined condition, reaction times were longer for reach directions for which the choice of left or right hand use was equiprobable. This point of equal choice is referred to as the point of subjective equality (PSE).

To examine the neural computations underlying hand choice, Oliveira et al. (2010) investigated the effect of transcranial magnetic stimulation (TMS) on the posterior parietal cortex (PPC) on hand choice. The PPC comprises the parietal reach region, a brain region associated with the planning of reaching movements. Single-pulse TMS to the left PPC increased the amount of left hand reaches and thus induced a bias in hand choice. However, TMS to the right PPC did not induce a similar bias in hand choice in the opposite direction. While the reason for this hemispheric asymmetry remains unclear, these results suggest that the PPC is part of the network involved in hand selection.

Here, we test the idea of parallel processing of movement plans for the decisions of hand choice by investigating the neural synchronization in sensorimotor regions. We will concentrate on beta-band activity (13 to 30 Hz), which has often been associated with movement preparation (Jasper & Penfield, 1949). Typically, event-related desynchronization (ERD) in the beta band is thought to reflect cortical activation and, more specifically, preparation of the execution of a movement (Pfurtscheller, 1992). However, beta-band ERD is not an undifferentiated reflection of neural activity. The level of desynchronization appears to be modulated by the level of uncertainty about the direction of the upcoming movement. Studies investigating this effect based this directional uncertainty either on the number of possible reach directions (Tzagarakis, Ince, Leuthold, & Pellizzer, 2010) or the separation of two possible reach directions in space (Grent-'t-Jong, Oostenveld, Jensen, Medendorp, & Praamstra, 2014; Grent-'t-Jong, Oostenveld, Medendorp, & Praamstra, 2015). In both cases, greater directional uncertainty corresponded to less beta-band ERD prior to the reaching movement.

Based on this information, we hypothesized that beta-band ERD is modulated in a similar way by decision uncertainty. This decision uncertainty would be based on the amount of competition between movement plans for the left and the right hand. To test this, we compared reaction times and beta-band ERD between predetermined and undetermined (freely selected) hand choice trials. Additionally, we compared reaction times and beta-band ERD for reaching movements towards target locations that evoke low competition between the left and the right hand to those associated with high competition between the hands. If movement plans are indeed prepared in parallel, we should expect more competition to result in longer reaction times (Oliveira et al., 2010) and less beta-band ERD (Grent-'t-Jong et al., 2014; Grent-'t-Jong et al., 2015; Tzagarakis et al., 2010).

Methods

Participants

Twenty participants took part in the study (5 males and 15 females, M = 21 years, age range 19-26 years). All participants reported to be right-handed, and this was confirmed by their responses on the Edinburgh Handedness Inventory (Oldfield, 1976). The participants had normal or corrected-to-normal vision and reported no history of neurological or psychiatric diseases or use of psychoactive medication or substances in the recent past. The ethics committee of the Faculty of Social Sciences of Radboud University Nijmegen, the Netherlands, approved the study. All participants gave written informed consent prior to the start of the study and were reimbursed for their participation in form of course credits, if applicable.

Experimental set-up

Participants performed a speeded hand-selection reaching task. Figure 1 illustrates the experimental paradigm, which is based on the task introduced by Oliveira et al. (2010). Visual stimuli were presented on a 42-inch touch monitor (Iiyama, Tokyo, Japan) with full HD resolution (1080p) and a refresh rate of 80 Hz. The room in which the task was completed was dark, except for the light emitted by the touch monitor and the monitors of the experimental computers, positioned approximately three meters away from the participant with the rear-side of the monitors facing the participant. At the start of the experiment, two start positions were presented on the touch monitor as grey disks with a diameter of 3.5 cm. These start positions were visible throughout the experiment and were positioned approximately 20 cm away from the participant's diaphragm and 9 cm on either side of the body midline. The color of these disks changed to white when touched to indicate correct placement of the participant's index fingers (Fig. 1A). A gaze fixation cross with a width of 2.5 cm was presented close to the center of the screen, 12 cm in front of the two start positions. There were five different possible cue and target positions on the screen. Colored disks with a diameter of 3.5 cm could be positioned in one of the following five directions on a semi-circular array with a 30 cm radius with its center at the fictitious point in the middle of the two starting positions: -40°, -10°, 0°, 10°, or 40°. Negative and positive directions on the semi-circle indicate positions to the left and the right of the body midline, respectively. Cue stimuli were orange and target stimuli were light blue.

The presentation of the task on the touch monitor was controlled by software that was custom-written in the Python programming language (Python Software Foundation, Beaverton, United States of America). To measure the onset of visual stimuli on the touch monitor, a photodiode was connected to the touch monitor and registered the presentation of cue and target stimuli. The output of the photodiode was recorded at 500 Hz with an electroencephalography (EEG) system (described later).

Experimental paradigm and procedure

The task of the participants was to reach with one of their index fingers towards the target as fast and accurately as possible. Participants were free to decide which hand to use for the movement in the majority of the trials. A trial was initiated by placing the two index fingers on the start positions presented on the touch monitor. After a fixed period of 1 s the cue appeared at one of the five cue positions (Fig. 1B and 1C). The duration of the presentation of the

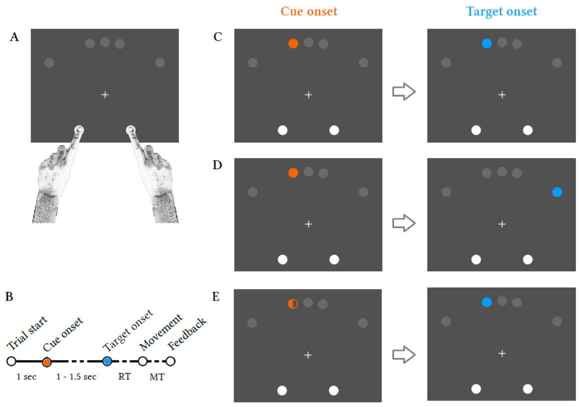


Figure 1. Illustration of the experimental set-up, paradigm, and procedure. (A) Top view of the experimental set-up (hands are not to scale relative to the set-up). Start positions (white disks), gaze fixation cross, and five potential cue and target positions (light grey disks) are shown. (B) Summary of the order of events over time for one trial. Variable periods of time are visualized with dashed lines. Reaction time and movement time are abbreviated as RT and MT, respectively. (C) Example of a correctly cued trial; the cue (orange) appeared at the same position as the target (light blue). Note that the other potential cue and target positions were not shown during the experiment. (D) Example of an incorrectly cued trial; the cue appeared at a different position than the target. (E) Example of a predetermined trial during which a modified cue stimulus instructed which hand to use. In this example, the participant was instructed to reach with the left hand.

and the participants were explicitly instructed to use the cue to prepare the movement. After the cue period the target was presented and the participants initiated the movement. The onset of the target was accompanied by a short beep sound. The cue and target could either be presented at the same position (correctly cued trials, 450 repetitions, Fig. 1C) or at different positions (incorrectly cued trials, 450 repetitions, Fig. 1D). Incorrectly cued trials were introduced to verify that the participants used the cue to prepare the reaching movement. Note that the participants were unaware of the type of trial during the cue period, as they were not provided with any other information apart from the target presented later on. When the participant touched the target it disappeared and a feedback message about the response time was presented close to the gaze fixation cross on the touch monitor. The participants were awarded with virtual points if the sum of the reaction time and the movement time was shorter than 0.7 s. This reward was implemented to motivate the participants to reach towards the target as fast as possible and therefore use the position of the cue to prepare the movement. If the sum of the reaction time and the movement time was indeed shorter than 0.7 s, the message “Well done! +1 point” was presented. Next to the feedback message, the total score of the participant was presented. If the sum of the reaction time and the movement time exceeded 0.7 s, the feedback message was “Too slow”. The feedback message disappeared when the participants placed the index finger of the hand used for the reaching movement back on the start position, thereby initiating the next trial. If the movement was initiated prior to the onset of the target, “Please wait for the target” was presented and the trial was restarted.

In one out of nine trials, the left or the right half of the cue stimulus was colored black instead of orange (predetermined trials, 100 repetitions, Fig. 1E). The orange colored half of the cue stimulus predetermined the hand to use for the reaching movement following the onset of the target. The participants were informed about these modified cue stimuli prior to the experiment and were able to dissociate the different cue stimuli during the practice trials.

All participants completed 900 trials in total. These comprised of 450 correctly cued trials (90 repetitions of each cue x target combination) and 450 incorrectly cued trials (22 or 23 repetitions of each cue x target combination). Participants were free to use the hand of their choice in 800 of these trials. In the remaining 100 trials the hand to use was predetermined (50 left hand and 50 right hand trials). The amount number of correctly cued trials was equal to the amount number of incorrectly cued trials in both the choice and the predetermined hand condition. In the predetermined hand condition, these trials were also balanced across hands. All trials were presented in a random order that differed for each participant and were subdivided in six blocks of 150 trials, separated by short breaks.

To familiarize with the experimental paradigm, participants completed 30 practice trials prior to the main experiment. Practice trials included all trial types. To make sure that participants were able to distinguish predetermined trials, the proportion of these trials was higher for the practice trials than for the main experiment (8/30 versus 100/900). In total, completion of both the practice trials and the main experiment took about one hour.

EEG acquisition and preprocessing

A 64-channel active electrode EEG system was used to record brain activity throughout the experiment (Brain Products, Gilching, Germany). Horizontal and vertical electro-oculograms (EOGs) were recorded by placing electrodes at the supraorbital and infraorbital ridges of the left eye and the outer canthi of the left and right eye. Impedance values were kept below 20 kΩ and the signal of all electrodes was referenced to the signal on left mastoid electrode TP9. The data was filtered online with a low cutoff value of 0.016 Hz and a high cutoff value of 200 Hz and digitized with a sampling frequency of 500 Hz and a resolution of 0.1 μV. To avoid excessive eye movements during the experiment, the participants were instructed to look at the gaze fixation cross throughout the experiment. However, to enable the participants to accurately touch the target, the participants were free to move their eyes during the presentation of the target.

The FieldTrip toolbox was used to process the EEG data off-line in MATLAB (Oostenveld, Fries, Maris, & Schoffelen, 2011). The data was re-referenced to the average signal of all EEG electrodes. Slow drifts in the signal and noise originating from the power lines were eliminated by applying a high-pass filter of 1 Hz and a band-stop filter at frequencies of 50 Hz, 100 Hz, and 150 Hz, respectively. Trials were time-locked to the onset of the cue as recorded by the photodiode. Trials with blinks around the onset of the cue were removed from the dataset because the blinks could have altered the timing of movement preparation. Blinks were automatically identified with Fieldtrip toolbox. To do so, first, the difference between the signal of the vertical EOG electrodes was computed, after which a fourth order Butterworth band-pass filter with a frequency range of 1 to 15 Hz was applied to the signal. The band-pass filtered data was transformed by applying a Hilbert transformation, after which the data was z-transformed. If z-transformed values exceeded the cutoff value of 3, this indicated the detection of a blink. Trials in which participants blinked around the onset of the cue, in the time window from 75 ms prior to cue onset to 25 ms after cue onset, were removed from further analyses. On average, this resulted in removal of 21 trials per participant (*SE* = 4.85).

Ocular artifacts not centered around the onset of the cue were removed from the data by running an independent component analysis (ICA). Rejection of components with an evident ocular origin was done according to the criteria described by McMenamin et al. (2010). After removal of these components, trials with excessive muscle activity during and preceding the cue period were automatically identified and removed by looking into high-frequency components of the data. To do so, the data was band-pass filtered by means of a ninth-order Butterworth band-pass filter with a frequency range of 110 to 140 Hz. The band-pass filtered data was transformed by applying a Hilbert transformation, after which the data was z-transformed. Trials in which z-transformed values exceeded the cutoff value of 13 in the time window from the start of the baseline period (0.3 s prior to cue onset) to target onset were considered trials with excessive muscle activity. These trials were removed from further analyses. On average, this resulted in removal of 103 trials per participant (*SE* = 12.30).

After identification and removal of trials containing artifacts, as well as removal of ocular components in the data, the data was low-pass filtered with a frequency threshold of 90 Hz and down-sampled from 500 Hz to 200 Hz to reduce the size of the data set and lower the processing capacity needed for further analyses. Excessively noisy or *dead* electrodes were identified by visually inspecting the preprocessed data for continuous high frequency noise or a lack of signal, respectively, and were replaced through interpolation of the EEG signal of electrodes adjacent to the respective electrode. Adjacent electrodes were identified based on a two-dimensional projection of the position of the electrodes and data of four participants contained such a noisy or irresponsive electrode (FC1, TP7, PO3, or Oz).

Analysis of natural choice behavior

Hand choice was determined for each trial as the hand that released the touch screen after the onset of the target. Trials during which the participant released both hands were not taken into account in further analyses. First, each participant's natural choice behavior was described by focusing on correctly cued choice trials. The preference to use the hand ipsilateral to the target position was tested with a non-parametric Wilcoxon signed-rank test comparing the proportion of right hand reaches for targets presented in the left-hemifield and targets presented in the right-hemifield. The general preference to use the dominant hand was tested with a non-parametric Wilcoxon rank-sum test comparing the overall proportion of right hand reaches with 0.5.

Next, the proportion of right hand choices for each target position during correctly cued choice trials was described by fitting a cumulative Gaussian distribution. The cumulative Gaussian distribution is described as follows:

$$\mathrm{P}\,(\mathrm{x}) = \lambda + (1 - 2\lambda)\,\frac{1}{\sigma\sqrt{(2\pi)}}\int_{-\infty}^{\mathrm{x}} \mathrm{e}^{-\frac{(\mathrm{t}-\mu)^2}{2\sigma^2}}\,\mathrm{dt} \quad (1)$$

P (x) represents the proportion of right hand reaches for target position x. The mean of the fitted curve, μ, represents the participant's PSE. The standard deviation of the curve is represented by σ and is related to the steepness of the curve, whereas λ represents the lapse or error rate. This error rate was controlled to improve the fit of the curve and was limited to values smaller than 0.1. The value of t ranged from -∞ to x. Fitting the cumulative Gaussian distribution to the experimental data was done according to a maximum Likelihood approach and was carried out with MATLAB functions 'normcdf' and 'fmincon'.

The preference to use the dominant hand to reach towards targets presented close to the body midline was tested with a two-tailed independent t-test comparing PSE values with 0˚. Based on the cumulative Gaussian distribution fit for each individual participant, the target was determined for which the competition between movement plans for each of the two hands was greatest. This high competition target was defined as the target closest to the participant's PSE and will be referred to as the PSE target. Two low competition targets were defined as the leftmost target (-40°) and the rightmost target (40°). Participants were expected to show a clear preference to reach for these targets with the left hand or the right hand, respectively, and these targets will be referred to as the extreme targets. Analyses of reaching movements to these extreme targets will focus only on the extreme target ipsilateral to the hand used (left extreme target for left hand reaches and right extreme target for right hand reaches).

Three participants showed such a strong preference to reach with their dominant right hand for the correctly cued choice trials that it was not possible to determine a low competition extreme target at which reaches with the left hand were evidently preferred. Their overall proportion of right hand reaches for correctly cued choice trials was 0.868, 0.833, and 0.993, and the proportion of right hand reaches towards the leftmost target was 0.663, 0.413, and 0.988, respectively. These three participants were excluded from further analyses.

Analysis of cue and target-based choice behavior

As mentioned earlier, incorrectly cued choice trials were introduced to verify that the participants used the cue to prepare the reaching movement. If the participants did indeed prepare a reaching movement towards the cue, hand choice should be biased based on cue position. Such a bias in hand choice would also indicate that the competition between movement plans for reaching movements with one of the two hands is, at least partially, resolved prior to target onset. Main effects of cue and target position on hand choice, as well as an interaction effect, were tested with a factorial repeated-measures ANOVA. If the results of Mauchly's test indicated that the assumption of sphericity was violated, the degrees of freedom were corrected according to the Greenhouse-Geisser estimates of sphericity. The outcome of particular cue x target combinations was assessed post-hoc.

Analysis of reaction times

Reaction times were defined as the first moment after target onset at which one of the hands released the screen as registered by the touch monitor. Trials with reaction times exceeding 1 s were not taken into account in any of the following analyses (similar to Tzagarakis et al., 2010). The effect of competition on reaction time was assessed with a factorial repeated-measures ANOVA. To additionally assess the effect of the length of the cue period on reaction times, cue period was added as an independent variable. The design was thus trial type (hand predetermined or choice) x target (extreme or PSE) x cue period (1.00, 1.25, or 1.50 s). If the results of Mauchly's test indicated that the assumption of sphericity was violated, the degrees of freedom were corrected according to the Greenhouse-Geisser estimates of sphericity.

Analysis of beta-band ERD

Beta-band ERD was determined by performing a time-frequency analysis of the EEG data. The time-frequency analysis was based on multiplication in the frequency domain and made use of a single Hanning taper with variable window length. The window length was dependent on the frequency of interest and was set to 5 divided by the frequency of interest, resulting in five cycles per time window. Frequencies of interest initially ranged from 13 to 30 Hz ($M_{\text{frequency resolution}} = 4.30$ Hz, $M_{\text{window length}} = 0.25$ s). Power values were computed every 10 ms starting from 0.3 s prior to cue onset up to 1.0 s after cue onset in steps of 0.5 Hz. Computed power values were corrected relative to baseline power in the period of 300 to 1000 ms prior to cue onset.

The beta-band frequency range appropriate in this study was determined by calculating the mean power relative to baseline over the entire frequency range (13 to 30 Hz) and for all electrodes in the time period of 0.8 to 1.0 s after cue onset. Power values for each frequency were averaged across participants, resulting in a two-dimensional dataset (electrode x frequency). This analysis focused on predetermined trials only, and separate averages were computed for left and right hand reaches. By subtracting the mean power preceding right hand reaches from the mean power preceding left hand reaches for each electrode, the frequency range that showed the clearest lateralization in activity across the two hands could be determined. In a similar way, the electrodes that showed where this lateralization was greatest were identified. Mean power was computed only across the frequency range that showed the clearest lateralization in activity across the two hands. The topographic distribution of the contrast was plotted and electrodes that showed the greatest lateralization in activity were selected for further analyses.

Beta-band power over time was calculated by averaging power values across the appropriate frequency range. This analysis resulted in a three-dimensional dataset with power values for each participant (trial x electrode x time). For the choice trials, this analysis focused on correctly cued trials only. We assumed that, for these trials, participants did not deviate from the movement plan that was dominant during the cue period after the onset of the target and thus eventually reached with the hand that *won the competition*. After the time-frequency analysis, trials were grouped based on the following criteria: trial type (hand predetermined or choice), target (extreme or PSE), and the hand used for the reaching movement (left or right).

The effect of movement plan competition on beta-band ERD was assessed with a nonparametric cluster-based permutation test (Maris & Oostenveld, 2007). This statistical analysis is based on the calculation of cluster-level statistics, connecting samples based on temporal adjacency, and circumvents the multiple comparisons problem often encountered during the analysis of large multidimensional neuroimaging datasets. To compute the cluster-based permutation test statistic all samples were first compared across conditions using multiple dependent t-tests. After this, samples with a t-value greater than a certain threshold were connected based on temporal adjacency and cluster-level statistics were computed. Differences between conditions were then evaluated by using the cluster-level statistic with the largest absolute value as a test statistic and determining the p-value under a permutation distribution. Here, the cluster-based permutation test was performed with the function 'ft_freqstatistics' of the FieldTrip toolbox and the permutation distribution was constructed with the maximum number of random partitions. Samples for which cluster-level statistics were computed ranged from 0.1 to 1 s after cue onset. Both the initial threshold for forming the clusters and the ultimate critical alpha-level to assess differences between conditions were set to 0.05.

First, to assess the effect of competition on beta-band ERD based on the ability to choose the hand to use, beta-band ERD was compared for predetermined and choice trials, both for reaching movements to the PSE target and the extreme targets. Second, to assess the effect of competition on beta-band ERD based on reaching movements towards target locations that evoke high competition between the left and the right hand to those associated with low competition between the hands, beta-band ERD was compared for reaching movements to the PSE target and the extreme targets, both for predetermined and choice trials.

Results

Natural choice behavior

When participants were free to choose which hand to use, they showed an overall preference to reach with the hand ipsilateral to the target position for the correctly cued choice trials. The proportion of right hand reaches was significantly smaller for the two targets presented at -40° and -10° (median = 0.28) than for the two targets presented at 10° and 40° (median = 0.98), $z = -3.62$, $p < 0.001$. Overall, however, the participants preferred to use their dominant right hand. The proportion of right hand reaches across all targets was significantly greater than 0.5 (median = 0.67), $z = -3.48$, $p < 0.001$. These results are in line with previous observations (Bryden, Pryde, & Roy, 2000; Gabbard & Rabb, 2000).

Figure 2 shows the Cumulative Gaussian

distribution fits to the natural choice behavior of three participants who demonstrated different preferences in hand choice. Fit parameters for all participants are described in Supplementary Table 1. Overall, the PSE was significantly smaller than 0° (*M* = -9.97, *SE* = 2.16), *t*(16) = -4.61, $p < 0.001$ (Fig. 2D), indicating that the PSE was usually left of the body midline. On average, participants used their left and right hand with approximately equal amounts to reach for the -10° target. The target closest to, or with the minimum absolute distance from, each individual participant's PSE was selected as the high competition target for that particular participant. Thirteen out of seventeen participants had the -10° target as PSE target, and three participants the 0° target. One participant showed a preference to reach with the left hand: the 10° target was the PSE target.

Cue and target-based choice behavior

To verify that participants prepared the reaching movement during the cue period and did not delay movement preparation until target presentation, we examined whether the position of the cue-biased hand choice using a factorial repeated-measures ANOVA which showed that both cue *F*(1.70, 27.20) = 27.66, $p < 0.001$, and target *F*(1.70, 27.12) = 104.28, $p < 0.001$, affected the proportion of right hand reaches. Additionally, there was an interaction effect between cue and target, *F*(7.01, 112.13) = 7.02, $p < 0.001$. The effects are shown in Figure 3. Modulations of target position on hand choice are reflected by a shift in the proportion of right hand reaches along the x-axis. Modulations of cue position on hand choice are reflected by a shift in the proportion of right hand reaches between the cue positions (different colored lines). Main effects showed that the proportion of right hand reaches increased for both cues and targets presented more to the right on the semicircle. Bonferroni corrected post-hoc tests confirmed significant interactions and indicated that the effect of the cue was largest for targets presented in line with or left of the body midline. In general, these results confirm that participants used the cue to prepare the reaching movement prior to target onset.

Reaction times

Effects of decision uncertainty and the length of the cue period on reaction times were tested with a factorial repeated-measures ANOVA. No significant interactions or main effects of trial type or target position were found (Fig. 4). The length of the cue period, however, significantly affected reaction times, *F*(2, 32) = 9.00, $p < 0.001$. Post-hoc tests revealed that reaction times were significantly longer following the shortest cue period of 1.00 s (*M*= 349 ms, *SE* = 6.74) compared to the intermediate cue period of 1.25 s (*M* = 336 ms, *SE* = 7.02), *t*(16)= 3.38, *p* = 0.003, and the longest cue period of 1.50 s (*M* = 336 ms, *SE* = 8.09), *t*(16) = 3.20, *p* = 0.006. Reaction times did not significantly differ across the intermediate and the longest cue periods of 1.25 and 1.50 s, respectively. These results suggest that longer cue periods resulted in greater response preparation, but that this response preparation did not differ when the cue period was lengthened from 1.25 to 1.50 s.

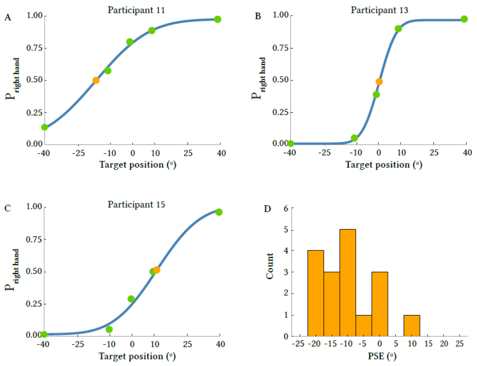


Figure 2. Cumulative Gaussian distribution fit for three participants showing different hand choice preferences and the distribution of PSE values. (A, B, C) Proportion of right hand reaches for each target position for three individual participants (green dots). Note that these data only comprise the correctly cued choice trials. The Cumulative Gaussian distribution fit is shown in blue and the mean of this distribution, i.e., the PSE, is indicated with an orange dot. (D) Histogram shows the distribution of PSE values for all participants.

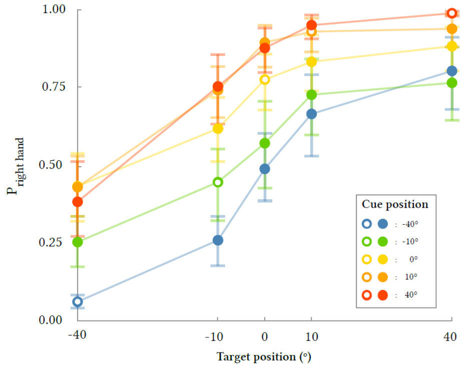


Figure 3. Effect of cue and target position on hand choice for each target position for incorrectly cued (filled dots) and correctly cued (open dots) choice trials. Mean proportion of right hand reaches over participants (± SE) is shown as a function of cue and target position.

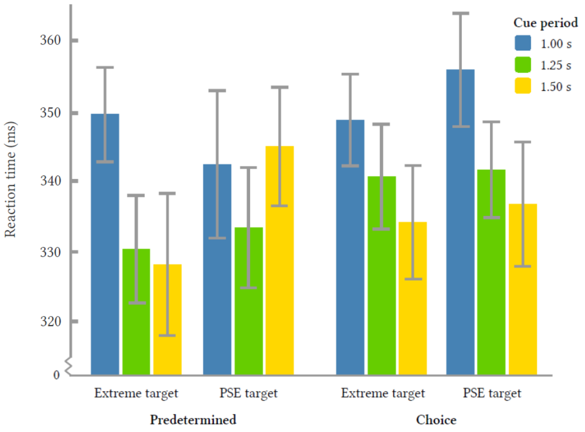


Figure 4. Mean reaction times over participants (± inter-participant SE).

Beta-band ERD

The frequency range that showed the greatest lateralization in activity preceding left and right hand reaches ranged from 16 to 23 Hz. For the sake of clarity, this frequency range will be referred to as beta-band. Figure 5A shows the topographic distribution of beta-band power preceding the reaching movement for left and right hand trials separately. Computing the contrast between these beta-band power distributions showed that the greatest lateralization could be found at central electrodes C3 and C4 (Fig. 5B). Further analyses therefore focus on these two electrodes; C3 for right hand reaches, and C4 for left hand reaches.

Modulations of beta-band ERD due to movement plan competition were tested by comparing predetermined and correctly cued choice trials for reaching movements to the PSE target (Fig. 6). Reaches to the PSE target were thought to involve high movement plan competition for choice trials because left and right hand choices were close to equiprobable. For predetermined trials, on the other hand, movement plan competition was expected to be low, or even absent, because the hand to use was already specified. If beta-band ERD preceding the reaching movement reflects movement plan competition, beta-band ERD is expected to be smaller with greater movement plan competition (Tzagarakis et al., 2010). We found that beta-band ERD was indeed significantly smaller for choice trials than for predetermined trials. This effect was found preceding both left ($p = 0.044$) and right hand reaches ($p < 0.001$) and was found approximately 0.6s after cue onset (mean onset of the effect across hands). These results are in line with our hypothesis and indicate that movement plan competition is reflected in the level of beta-band ERD, with greater movement plan competition resulting in less beta-band ERD.

Next, we investigated whether the modulation of beta-band ERD described above could be due to movement plan competition induced by having to decide which hand to use instead of making a predetermined reaching movement. Figure 7 illustrates the level of beta-band ERD preceding reaching movements for predetermined and correctly cued choice trials to the extreme targets. As participants showed a clear preference to reach with the left and right hand to the left and right extreme targets respectively, we expect movement plan competition to be low for both choice and predetermined trials.

Indeed, for left hand reaches, there was no significant difference in beta-band ERD between predetermined and choice trials. For right hand reaches, however, beta-band ERD was significantly smaller for choice trials than for predetermined trials, $p = 0.020$. This effect was found late in the cue period, approximately 0.8 s after cue onset.

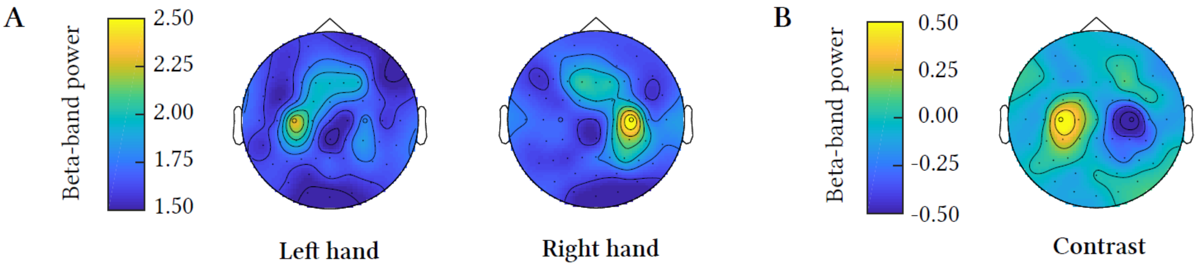


Figure 5. Localization of beta-band ERD. Topographic distribution of beta-band ERD preceding the reaching movement in the predetermined trials. Shown is the mean power as a ratio of baseline power in the frequency range from 16-23 Hz in the time period between 0.8 and 1.0 s after cue presentation. (A) Beta-band ERD preceding left and right hand reaches separately. (B) Difference in beta-band ERD preceding left and right hand reaches shown in panel A (left hand – right hand). Electrodes at which the contrast was greatest are C3 (left hemisphere) and C4 (right hemisphere).

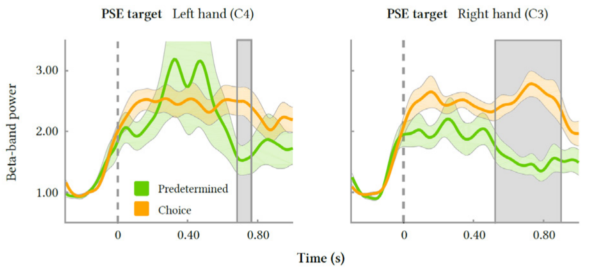


Figure 6. Beta-band ERD over time for the PSE target: predetermined versus choice. Shaded areas represent the SE over participants. Time point 0 indicates the onset of the cue. Grey areas indicate significant differences between the two conditions based on a nonparametric cluster-based permutation test ($p < 0.05$). Note that the line representing choice trials is based on more data than the line representing predetermined trials.

These results suggest that, at least for right hand reaches, having to choose the hand to use increases movement plan competition relative to making a reaching movement with a predetermined hand.

Based on the finding that movement plan competition modulates beta-band ERD when comparing predetermined and choice trials, we further examined the effect of movement plan competition within trials of the same condition. To do so, we compared beta-band ERD preceding correctly cued reaches to an extreme target versus the PSE target (Fig. 8). Reaches to the PSE target are thought to involve maximum movement plan competition, whereas movement plan competition is thought to be low for reaches to extreme targets. If movement plan competition based on target position is reflected in beta-band ERD, it is expected to be smaller preceding reaches to the PSE target. However, no significant differences in beta-band ERD were observed across the two target positions.

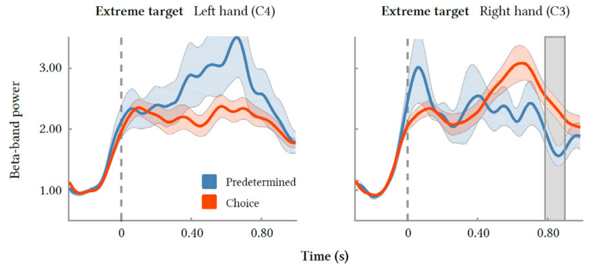


Figure 7. Beta-band ERD over time for the extreme targets: predetermined versus choice. Shaded areas represent the SE over participants. Time point 0 indicates the onset of the cue. Grey areas indicate significant differences between the two conditions based on a nonparametric cluster-based permutation test ($p < 0.05$).

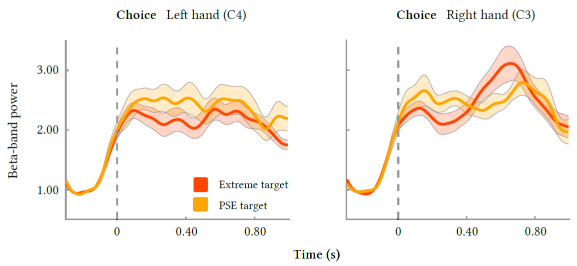


Figure 8. Beta-band ERD over time for choice trials: extreme target versus PSE target. Shaded areas represent the SE over participants. Time point 0 indicates the onset of the cue.

To investigate whether target position did not modulate beta-band ERD for predetermined reaches either, we compared beta-band ERD preceding predetermined reaches to an extreme target versus the PSE target (Fig. 9). Movement plan competition is expected to be low for all predetermined reaches, as the hand to use was already specified. We found, however, that beta-band ERD was significantly greater for reaches to the PSE target than for an extreme target. This effect was found for both left ($p = 0.024$) and right hand reaches ($p = 0.002$) and was found approximately 0.6 s after cue onset (mean onset of the effect across hands). This result suggests that beta-band ERD is modulated by target position. However, the modulation is opposite from the expected modulation, but not observed, for choice reaches. The difference in modulation patterns between the latter two comparisons suggests that the beta-band ERD modulation for predetermined reaches based on target position is not merely an effect of the position of the target in space, as this would have resulted in a similar modulation pattern of beta-band ERD for choice reaches. More likely, modulations based on target position involve processes other than movement plan competition.

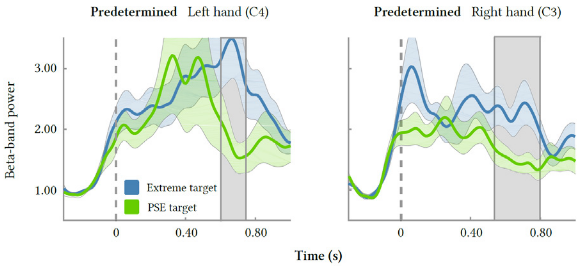


Figure 9. Beta-band ERD over time for predetermined trials: extreme target versus PSE target. Shaded areas represent the SE over participants. Time point 0 indicates the onset of the cue. Grey areas indicate significant differences between the two conditions based on a nonparametric cluster-based permutation test ($p<0.05$).

Discussion

In this study, we sought to find neural evidence for parallel processing of movement plans for hand choice. We hypothesized that beta-band ERD preceding a reaching movement is modulated by movement plan competition and tested this by recording brain activity during a speeded hand-selection reaching task. More specifically, greater movement plan competition was expected to be associated with greater decision uncertainty and less beta-band ERD preceding the reaching movement. The results indicate that beta-band ERD is indeed modulated by the level of movement plan competition: when reaching to a target close to the PSE, beta-band ERD was significantly smaller when participants freely selected the hand to use compared to when the hand to use was predetermined. Movement plan competition is thought to be maximal for reaches to the PSE when freely selecting the hand to use, as participants are equally likely to reach with their left hand or their right hand. For predetermined reaches, on the other hand, movement plan competition is expected to be low, independent of the target position. The observed modulation of beta-band ERD is therefore consistent with contemporary theories that suggest parallel processing of movement plans; ambiguity causes the internal representations of simultaneously prepared movement plans to compete (Cisek, 2007; Cisek & Kalaska, 2010).

It is noteworthy that shortly after cue onset, we observed an increase in power in the beta-band frequency range relative to baseline, instead of a decrease as the term beta-band ERD implies. This initial increase in power is thought to be due to the short inter-trial interval of the experimental paradigm, during which beta-band power has not fully recovered to true baseline values. While the short inter-trial interval is a shortcoming of the present study, the initial increase in power relative to baseline was observed in all comparisons and occurred at a time period outside the interval of interest in the cluster-based permutation test. We therefore do not think that this observation has an effect on the interpretation of the results discussed.

The interpretation of the modulation of beta-band ERD being based on ambiguity in hand choice is underlined by the finding that, for reaches to low competition targets, the ability to choose the hand to use was also associated with significantly smaller beta-band ERD compared to predetermined reaches. Both for choice and predetermined reaches, the decision uncertainty for reaches to these extreme targets is low. However, movement plan competition appears to be greater for choice reaches due to the fact that participants have to select the hand to use. The fact that no significant differences in beta-band ERD were observed within these choice reaches, when comparing reaches to the PSE target with reaches to an extreme target, is therefore surprising. This finding either suggests that movement plan competition does not differ for reaches to the PSE target and reaches to an extreme target, or that this difference in movement plan competition is not reflected in the level of beta-band ERD. However, based on the finding that beta-band ERD was modulated based on target position within predetermined trials, we would like to propose a third alternative explanation: movement plan competition does differ across target positions and does modulate the level of beta-band ERD, but this modulation is not observed for choice reaches due to an attentional benefit for reaches to the PSE target, additionally reflected in the level of beta-band ERD.

This explanation is based on the finding that, for predetermined reaches, beta-band ERD was significantly greater for reaches to the PSE target than for reaches to an extreme target. This unexpected finding is comparable to reaction time results reported by Oliveira et al. (2010), who studied behavioral correlates of movement plan competition for hand choice with a similar experimental paradigm, except for the fact that the target position was not cued. For predetermined trials, they found that reaction times for reaches to the PSE target were shorter than for reaches to extreme targets. The authors argued that this unanticipated result might be due to the possibility that participants focus their attention on the center of the experimental set-up, detecting central targets more readily than extreme targets. The idea of an attentional benefit for targets presented in the center of the screen could also explain the differences in beta-band ERD found for predetermined reaches in the present study. Beta-band activity in the frontal eye fields, located close to the premotor cortex, is known to be suppressed in a spatial selective fashion with attention (Siegel, Donner, Oostenveld, Fries, & Engel, 2008). Given the smeared spatial resolution of EEG signals, this attentional suppression might underlie the greater beta-band ERD for predetermined reaches to the PSE target compared to an extreme target. For choice trials, the attentional benefit for reaches to the PSE target and the accompanying enhancement of beta-band ERD might overpower the effect of movement plan competition on beta-band ERD, resulting in a net difference of zero between reaches to the PSE target and reaches to the extreme targets.

Next to the attentional benefit reflected in reaction times for predetermined reaches to the PSE target, Oliveira et al. (2010) found that reaction times for choice reaches to the PSE target were longer than for reaches to extreme targets. Though these patterns in reaction times are roughly in line with our findings on beta-band ERD modulation, we did not find similar differences in reaction times. This is perhaps due to experimental differences, as the experimental paradigm of Oliveira et al. (2010) did not include the presentation of a cue prior to target onset. Here, on the other hand, participants were instructed to prepare the reaching movement based on the position of the cue during the cue period (ranging from 1.00 to 1.50 s). Reaction time differences, similar to the ones that Oliveira et al. (2010), observed might not hold with this adapted experimental paradigm: based on the bias in hand choice due to incorrect cueing, movement plan competition is thought to be, at least partially, resolved prior to target onset. This suggests that the movement has been prepared during the cue period, underlined by the finding that reaction times appeared to be shorter with longer cue periods in this study. These ideas are corroborated by general differences in reaction times across the two studies: mean reaction times reported here were approximately 340 ms, whereas Oliveira et al. (2010) reported reaction times of approximately 410 ms.

Here, all observed differences in beta-band ERD were found rather late in the cue period, from approximately 0.6 or 0.8 s after cue onset. If participants started preparing the movement immediately after cue onset, differences in beta-band ERD due to movement plan competition were expected to be exhibited in the beginning of the cue period. However, it appears that participants delayed movement preparation until later in the cue period. This idea is supported by the finding that reaction times did not differ between cue periods with a duration of 1.25 and 1.50 s, but were significantly longer for cue periods with a duration of 1.00 s. Perhaps participants considered the average duration of the cue period, 1.25 s, as the standard time period within which the movement had to be prepared: with a longer cue period of 1.50 s the movement had already been prepared when the target appeared, but with a shorter cue period of 1.00 s movement preparation was still in progress at target onset. As movement preparation is generally thought to take less than 1.00 s, participants might have efficiently delayed the onset of movement preparation in time. This delay could explain the late onset of differences in the level of beta-band ERD. Even though it is known that the onset of beta-band ERD is related to the onset of movement preparation (Kaiser, Birbaumer, & Lutzenberger, 2001), to the best of our knowledge the human ability to intentionally delay movement preparation has not been studied.

A shortcoming of the present study is the large spread of PSE values. Even though for most participants the target at -10° was picked as being the PSE target, the variability in the distance between the PSE target and the extreme targets might have complicated the interpretation of the results. To avoid this asymmetry in the experimental set-up, future studies could assess the PSE value prior to the experiment and align the target in the middle of the experimental set-up with this PSE value.

In conclusion, this study focused on competition between movement plans for the left and right hand by investigating neural synchronization during a speeded hand-selection reaching task. Beta-band ERD was shown to decrease with greater competition between the two hands: for reaches to the PSE target, beta-band ERD was smaller for choice trials than for predetermined trials. These results support the idea that hand choice is based on a competitive process between movement plans for the left and right hand and therefore provide us with valuable information about the way the brain processes sensory information to prepare goal-directed movements and enables us to interact within complex environments.

References

Bryden, P.J., Pryde, K.M., & Roy, E.A. (2000). A performance measure of the degree of hand preference. *Brain and Cognition, 44*, 402-414. doi: 10.1006/brcg.1999.1201

Cisek, P. (2007). Cortical mechanisms of action selection: The affordance competition hypothesis. *Philosophical Transactions of the Royal Society B, 362*, 1585-1599. doi: 10.1098/rstb.2007.2054

Cisek, P., & Kalaska, J.F. (2010). Neural mechanisms for interacting with a world full of action choices. *Annual Review of Neuroscience, 33*, 269-298. doi: 10.1146/annurev.neuro.051508.135409

Gabbard, C., & Rabb, C. (2000). What determines choice of limb for unimanual reaching movements? *The Journal of General Psychology*, *127*(2), 178-184. doi: 10.1080/00221300009598577

Gabbard, C., Tapia, M., & Rabb Helbig, C. (2003). Task complexity and limb selection in reaching.

International Journal of Neuroscience, *113*, 143-152. doi: 10.1080/00207450390161994

Gibson, J.J. (1979). *The ecological approach to visual perception.* New York, United States of America: Taylor & Francis Group. doi: 10.1002/bs.3830260313

Grent-'t-Jong, T., Oostenveld, R., Jensen, O., Medendorp, P.W., & Praamstra, P. (2014). Competitive interactions in sensorimotor cortex: oscillations express separation between alternative movement targets. *Journal of Neurophysiology, 112*, 224-232. doi: 10.1152/jn.00127.2014

Grent-'t-Jong, T., Oostenveld, R., Medendorp, P.W., & Praamstra, P. (2015). Separating visual and motor components of motor cortex activation for multiple reach targets: A visuomotor adaptation study. *The Journal of Neuroscience, 35*(45), 15135-15144. doi: 10.1523/JNEUROSCI.1329-15.2015

Jasper, H., & Penfield, W. (1949). Electrocorticograms in man: Effect of voluntary movement upon the electrical activity in the precentral gyrus. *Archiv für Psychiatrie und Zeitschrift Neurologie*, *183*, 163-174. doi: 10.1007/BF01062488

Kaiser, J., Birbaumer, N., & Lutzenberger, W. (2001). Event-related beta desynchronization indicates timing of response selection in a delayed-response paradigm in humans. *Neuroscience Letters, 312*(2001), 149-152. doi: 10.1016/S0304-3940(01)02217-0

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods, 164*(2007), 177-190. doi: 10.1016/j.jneumeth.2007.03.024

Marr, D.C. (1982). *Vision.* San Francisco, United States of America: W.H. Freeman & Company. doi: 10.1002/neu.480130612

McMenamin, B.W., Shackman, A.J., Maxwell, J.S., Bachhuber, D.R.W., Koppenhaver, A.M., Greischar, L.L., & Davidson, R.J. (2010). Validation of ICA-based myogenic artifact correction for scalp and source-localized EEG. *NeuroImage, 49*(3), 2416-2432. doi: 10.1016/j.neuroimage.2009.10.010

Oldfield, R.C. (1971). The assessment and analysis of handedness: The Edinburgh Inventory. *Neuropsychologica, 9*, 97-113. doi: 10.1016/0028-3932(71)90067-4

Oliveira, F.T.P., Diedrichsen, J., Verstynen, T., Duque, J., & Ivry, R.B. (2010). Transcranial magnetic stimulation of posterior parietal cortex affects decisions of hand choice. *Proceedings of the National Academy of Sciences, 107*(41), 17751-17756. doi: 10.1073/pnas.1006223107

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience, 2011*, 1-9. doi: 10.1155/2011/156869

Oostwoud Wijdenes, L., Ivry, R.B., & Bays, P.M. (2016). Competition between movement plans increases motor variability: Evidence of a shared resource for movement planning. *Journal of Neurophysiology, 116*(3), 1295-1303. doi: 10.1152/jn.00113.2016

Pfurtscheller, G. (1992). Event-related synchronization (ERS): An electrophysiological correlate of cortical areas at rest. *Electroencephalography and Clinical Neurophysiology*, *83*, 62-69. doi: 0.1016/0013-4694(92)90133-3

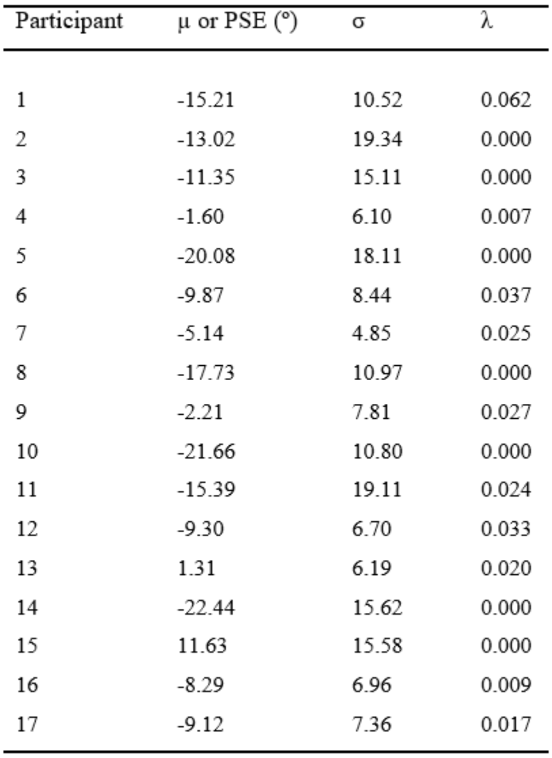
Poggio, T. (1981). Marr's computational approach to vision. *Trends in Neurosciences, 4,* 258-262. doi: 10.1016/0166-2236(81)90081-3

Siegel, M., Donner, T.H., Oostenveld, R., Fries, P., & Engel, A.K. (2008). Neuronal synchronization along the dorsal visual pathway reflects the focus of visual attention. *Neuron, 60*(4), 709-719. doi: 10.1016/j.neuron.2008.09.010

Tzagarakis, C., Ince, N.F., Leuthold, A.C., & Pellizzer, G. (2010). Beta-band activity during motor planning reflects response uncertainty. *The Journal of Neuroscience*, *30*(34), 11270-11277. doi: 10.1523/JNEUROSCI.6026-09.2010

Appendix A

Supplementary table 1.
Cumulative Gaussian parameter fits for each participant. Mean of the cumulative Gaussian fit, the standard deviation of the curve, and the lapse rate. The lapse rate was restricted to values smaller than 0.1.



Participant	μ or PSE (°)	σ	λ
1	-15.21	10.52	0.062
2	-13.02	19.34	0.000
3	-11.35	15.11	0.000
4	-1.60	6.10	0.007
5	-20.08	18.11	0.000
6	-9.87	8.44	0.037
7	-5.14	4.85	0.025
8	-17.73	10.97	0.000
9	-2.21	7.81	0.027
10	-21.66	10.80	0.000
11	-15.39	19.11	0.024
12	-9.30	6.70	0.033
13	1.31	6.19	0.020
14	-22.44	15.62	0.000
15	11.63	15.58	0.000
16	-8.29	6.96	0.009
17	-9.12	7.36	0.017

Enhanced Spatial Navigation Skills in Sequence-Space Synesthetes

Eline van Petersen[1]
Supervisor: Tessa M. van Leeuwen[1]
Co-authors: Rob van Lier[1], Mareike Altgassen[1]

[1]Radboud University Nijmegen, Donders Institute for Brain, Cognition and Behaviour, The Netherlands

Individuals with sequence-space synesthesia (SSS) perceive sequences like months, days and numbers in certain spatial arrangements. Several cognitive benefits have been associated with sequence-space synesthesia, such as enhanced mental rotation, more vivid visual imagery and an advantage in spatial processing. The current study aimed to further investigate these cognitive benefits, focusing on spatial navigation skills, to explore if the previously reported cognitive benefits are reflected in enhanced navigational performance. Synesthetes were distinguished from controls by means of a questionnaire, a consistency test and drawings. A virtual Morris Water Maze (MWM) task with two allocentric and two egocentric navigation conditions was used to assess spatial navigation abilities. For the allocentric tasks, participants had to use object cues to find a hidden platform and for the egocentric tasks, they had to use their own position as a reference. Results showed that synesthetes performed significantly better compared to controls on the allocentric and egocentric tasks that reflected real life situations more accurately. Further analyses revealed that specifically synesthetes with the ability to mentally rotate their spatial arrangements seemed to learn faster on the allocentric task. Results add to the existing literature concerning the cognitive benefits of SSS and are consistent with the previously found mental rotation advantage.

Keywords: sequence-space synesthesia, cognitive benefits, spatial navigation, allocentric, egocentric, virtual Morris Water Maze task

Corresponding author: Eline van Petersen; **E-mail:** elinevanpetersen@gmail.com

Synesthesia is a phenomenon in which sensory stimulation leads to automatic and involuntary additional experiences. Many forms of synesthesia have been reported some of which are common, like perceiving coloured letters (grapheme-colour synesthesia), and some very rare, like tasting words (lexical-gustatory synesthesia). The current study focused on one of the more common forms: sequence-space synesthesia (SSS; Jonas & Price, 2014). Individuals with SSS perceive months, days of the week, numbers or other sequences in certain spatial arrangements. For instance, months might be seen in a circular form, days in a U-shaped alignment or numbers on a spiralling line (Jonas & Price, 2014). These additional visuospatial associations in SSS have previously been associated with several cognitive benefits. Sequence-space synesthetes seem to show a memory advantage (Simner, Mayo, & Spiller, 2009), they perform better on mental rotation tasks (Brang, Miller, McQuire, Ramachandran, & Coulson, 2013; Havlik, Carmichael, & Simner, 2015), they report stronger visual imagery (Havlik et al., 2015; Price, 2009; Rizza & Price, 2012), they demonstrate higher visuospatial working memory accuracy and show an advantage in spatial processing (Hale, Thompson, Morgan, Cappelletti, & Cohen Kadosh, 2014). The current study aimed to further investigate these cognitive benefits of SSS, specifically focusing on spatial navigation skills, since the previously reported benefits might be reflected in enhanced navigational performance.

Characteristics of sequence-space synesthesia

Some sequence-space synesthetes have a synesthetic percept of only one sequence, whereas others see more than a dozen different sequences (Eagleman, 2009). Visualizing numbers and time units, like days, months and years are most common, but all kinds of sequences can elicit visuospatial impressions, even the alphabet, temperatures, or shoe sizes (Eagleman, 2009). The specific sequences that elicit visuospatial impressions comprise just one of the many aspects that can vary among individuals with SSS. For instance, the perceived sequences can take many different shapes of varying complexity (Jonas & Price, 2014), from simple lines, bent lines, and zigzag lines to circles, squares, and triangles (Eagleman, 2009) and even elaborate three-dimensional landscapes (Brang et al., 2013). Some synesthetes experience the shapes in mental space (associators), others experience the shapes outside of their body (projectors; Jonas & Price, 2014). For some, the forms are fixed, while others are able to apply spatial transformations, like mentally rotating them and zooming in or out in order to see them from multiple viewpoints (Jonas & Price, 2014). Some synesthetes additionally experience detailed visual content, such as colour or texture (Jonas & Price, 2014). Spatial forms might also be seen in two or three dimensions or in first person or third person perspective (Eagleman, 2009).

For each sequence-space synesthete, the perceived spatial arrangements and characteristics are likely to be unique and can even be different for the different sequences that he or she is able to visualize (Jonas & Price, 2014). For example, the spatial form for months might be seen from different viewpoints with the passage of time, whereas the form for the alphabet might not involve spatial transformations because it does not change over time (Jonas & Price, 2014). Moreover, the spatial arrangements often involve personal importance, like distortions of date lines to mark personally significant events (Price & Pearson, 2013) or important months occupying more space than others (Brang et al., 2013). Despite these various manifestations of SSS, it is common for all sequence-space synesthetes that seeing, hearing or thinking about particular sequences, as a whole or in parts, automatically elicits additional visuospatial experiences that are consistent over time (Cohen Kadosh, Gertner, & Terhune, 2012; Price & Pearson, 2013).

Cognitive benefits and costs of sequence-space synesthesia

Previously, it has been shown that visuospatial associations in SSS are beneficial for several cognitive tasks. Sequence-space synesthetes with time forms (e.g., months or years) outperform non-synesthetes in tests that assess recall of dates of public events and content of events in their own life. Synesthetes subsequently reported that events were retrieved from spatial locations within their visuospatial forms (Simner et al., 2009). It seems that the additional experiences in SSS lead to richer encoding and retrieval opportunities during memory tasks (Rothen, Meier, & Ward, 2012). Moreover, sequence-space synesthetes perform better than controls on mental rotation tasks (Brang et al., 2013; Havlik et al., 2015), they report stronger visual imagery than controls (Havlik et al., 2015; Price, 2009; Rizza & Price, 2012) and demonstrate higher visuospatial working memory accuracy (Hale et al., 2014). Furthermore, SSS has been linked to increased spatial processing. This was demonstrated with a task comprising spatial stimuli (i.e., several circles differing in size), where participants were required to make judgements about the overlapping order (i.e., leftmost circle on the bottom and rightmost on the top, or the other way around; Hale et al., 2014), as well as by means of a questionnaire about cognitive styles (Mealor, Simner, Rothen, Carmichael, & Ward, 2016).

However, the consistent and automatic synesthetic experiences have some costs as well. Sequence-space synesthetes respond more slowly on target detection tasks when the spatial relationship between items of presented sequences, like months (Smilek, Callejas, Dixon, & Merikle, 2007) and numbers (Gertner, Henik, & Cohen Kadosh, 2009), is incongruent with their own visuospatial percept of those sequences. This suggests that SSS "impairs the ability to represent items of sequences in a flexible manner according to task demands" (Gertner et al., 2009, p. 366). Furthermore, it has been shown that synesthetes with number forms are slower in doing simple calculations (Ward, Sagiv, & Butterworth, 2009). Perhaps because they are relying on their visuospatial forms when solving these arithmetic problems instead of using rote retrieval (i.e., a memorization strategy based on repetition), which is more optimal in this case (Hale et al., 2014).

Sequence-space synesthesia and spatial navigation

The association of SSS with the previously reported benefits leads to the expectation that sequence-space synesthetes might have enhanced spatial navigation skills. Mental rotation, spatial processing, memory, and imagery are involved in spatial navigation (Harris, Wiener, & Wolbers, 2012). Evidence from patient studies support the role of memory and imagery in spatial navigation. Patients suffering from representational neglect are affected in memory and spatial imagery performance (Chersi & Burgess, 2015) and it has been shown that they experience deficits in navigation when they have to re-orient themselves (Guariglia, Piccardi, Iaria, Nico, & Pizzamiglio, 2005). So, when memory and imagery performance are affected, spatial navigation abilities are likely to be affected. Thus it is quite plausible that when memory and imagery performance are enhanced, as is reported for sequence-space synesthetes, spatial navigation abilities are enhanced as well.

Guariglia et al. (2005) used a human version of the Morris Water Maze (MWM) task in real space to test spatial navigation skills in patients with mental representation disorders. The original MWM task was developed for rats and required them to find a platform by using various cues while moving around in a pool (Morris, 1981). In the human version of the task used by Guariglia et al. (2005), participants had to explore a room and find a target location. This task only required target place learning from different starting positions. The current study used a computerized version of the MWM task similar to the one used by Ring, Gaigg, Altgassen, Barr and Bowler (2018) that was adapted from Feigenbaum and Morris (2004). In this version of the task, a virtual pool was presented on a touchscreen and participants were required to find a hidden platform by moving over the screen. Over trials they had to work out and learn the shortest possible path from the starting point to the platform. For some conditions, they had to make use of object cues (allocentric) and for other conditions they had to use their own position as a reference (egocentric) in order to find the platform.

Differentiating between allocentric and egocentric conditions is relevant because normally when navigating in an environment both allocentric and egocentric strategies can be used. When using an allocentric strategy – also called place strategy or spatial memory strategy – one uses cognitive maps (i.e., mental representations of an environment) by thinking about landmarks and their positions relative to each other (e.g., Di Tore, Corona, & Sibilio, 2014; Iaria, Petrides, Dagher, Pike, & Bohobot, 2003; Konishi & Bohobot, 2013). When using an egocentric strategy – also called response or route strategy – one navigates by following a learned sequence of self-movements, such as a series of left and right turns at precise decision points from a given starting position (e.g., turn right after the park; e.g., Bohobot, Lerch, Thorndycraft, Iaria, & Zijdenbos, 2007; Chersi & Burgess, 2015; Di Tore et al., 2014; Konishi & Bohobot, 2013). The MWM task involved two allocentric and two egocentric conditions of which the first conditions (i.e., Allocentric 1 and Egocentric 1) were more similar to navigation in daily environments.

Previous studies have demonstrated that the virtual version of the MWM task is sensitive to detect differences in spatial navigation between certain groups. Ring et al. (2018) demonstrated that individuals with autism spectrum disorder (ASD) have difficulties in allocentric navigation, particularly when the task required them to change position while the platform and objects kept the same locations (i.e., the Allocentric 1 condition). Feigenbaum and Morris (2004) also found impairment on this allocentric navigation task in patients who had undergone right temporal lobectomy (RTL). Using this virtual MWM task, the current study set out to investigate whether such differences in navigational performance exist between sequence-space synesthetes and non-synesthetes.

Besides comparing spatial navigation skills between the two groups, individual differences among sequence-space synesthetes were examined when assessing spatial navigation abilities in order to see if some specific synesthetic features contributed to performance. Specifically, the ability to mentally rotate spatial forms was expected to enhance performance in at least the allocentric condition, in which the display had to be mentally rotated in order to find the platform. Further influences of synesthetic features on navigational performance were explored as well. Individual differences among sequence-space synesthetes have previously been shown to influence performance on visuospatial tasks. For example, synesthetes who are able to project forms into space (i.e., projector synesthetes) are shown to perform best on mental rotation tasks (Havlik et al., 2015).

The current study aimed to further investigate the cognitive benefits of SSS. More specifically, do sequence-space synesthetes have enhanced spatial navigation skills? Knowing whether sequence-space synesthetes outperform non-synesthetes at spatial navigation tasks and knowing whether individual differences among synesthetes are associated with enhanced performance may reveal information about the cognitive processes involved in SSS. This study therefore contributes to a better understanding of SSS at a cognitive level and may extend our knowledge about the cognitive processes involved in allocentric and egocentric navigation strategies.

Methods

Participants

Participants with SSS were recruited through poster advertisements and via the SONA Radboud research participation system. Age- and sex-matched control participants were recruited via the SONA system as well. Based on an online screening questionnaire about synesthetic experiences, 23 potential synesthetes and 22 controls were invited to take part in the study. After the tasks in the lab, one potential synesthete was not included in the synesthete group (due to insufficient responses at the consistency task and a drawing without typical synesthetic characteristics) and one potential control participant appeared to be a synesthete. The groups that were taken into account in analysis comprised 23 individuals with SSS (20 women, M_{age} = 23.22 years, age range 18-25 years with three exceptions (34, 38 and 44 years)) and 21 controls (19 women, M_{age} = 21.57 years, age range 18-25 years). An independent samples t-test indicated no significant age difference between groups ($t(26) = 1.15$, $p = .263$). Because of unequal variances between groups (Levene's test was significant), the degrees of freedom were adjusted accordingly. Of those included in the synesthete group, 13 reported having spatial forms for numbers, 21 for days and 23 for months. All participants were educated at university level. Informed consent was obtained before filling out the online screening questionnaire and again in the lab before taking part in the experiment. Participation was voluntary and compensated with 15 euros or 1.5 credit points. The study was approved by the Ethics Committee of the Faculty of Social Sciences (ECSS) at Radboud University Nijmegen.

Tasks and procedure

General procedure

Before participating in the study, all participants filled out a self-report questionnaire about synesthetic experiences. In the lab, by means of a consistency test, drawings, and additional questions, sequence-space synesthetes were distinguished from control participants. Participants were asked to select locations on a computer screen for numbers, days, and months, yielding a consistency score of the placement of items, and to draw their visuospatial experiences (synesthetes) or intuitive representations (controls) of those sequences on paper. Then participants performed a virtual MWM task to assess spatial navigation skills. During allocentric and egocentric navigation tasks, participants were asked to find a hidden platform by moving over a touchscreen. For the allocentric tasks, they had to use object cues to find the platform and for the egocentric tasks, they had to use their own position as a reference to find the platform. Performance on these different tasks was assessed between groups and individual differences in the manifestation of SSS were taken into account in further analyses. We now describe each task in detail.

Screening questionnaire

An online SSS self-report questionnaire was developed in LimeSurvey and used as a screening tool to find participants for our study. The questions were based on descriptions of SSS in the literature and comprised some general screening questions (e.g., "Is the synesthetic experience automatically elicited when thinking of this sequence?") and some detailed questions about the perceived spatial forms (e.g., "Do you see this arrangement from a fixed perspective or are you able to rotate the form and adopt multiple viewpoints?"). When participants reported to have SSS for numbers, days and/or months, detailed questions followed about the spatial forms of each of those sequences separately. These questions covered all the characteristics of SSS as mentioned in the theoretical background. The complete questionnaire can be found in the Supplementary Materials (available in the online version). Filling out the questionnaire took about 10 to 20 minutes. Participants who reported having SSS received an invitation to participate in the study at Radboud University. Control participants filled out the questionnaire as well. They could just simply answer the first question ("Do you think you have sequence-space synesthesia?") with "no" and the specific questions about SSS did not appear. Before filling out the questionnaire, all participants were provided with a description of SSS and some examples of visuospatial forms to familiarize all of them with SSS prior to completing the questionnaire.

Consistency test

Because the screening questionnaire was based on self-report, participants' subjective reports of SSS were verified in the lab. Participants were asked again about the details of their spatial forms and performed a consistency test. The consistency test (Rothen, Jünemann, Mealor, Burckhardt, & Ward, 2016), written in E-prime 2.0, was obtained from Rothen et al. (2016) and adapted to the current study. Numbers 0-9, 50 and 100 ($N = 12$), days ($N = 7$) and months ($N = 12$) were centrally presented on a white background with font style Courier New and font size 18 in bold black. These stimuli were presented on a 24" BenQ screen with display resolution set to 1920x1080, controlled by a Dell computer running Windows 7.

Participants were comfortably seated in front of the screen at normal viewing distance. Stimuli were presented one by one in random order and participants had to select a location for each stimulus on the screen by making a mouse click. SSS participants were instructed to imagine the screen as the space in which they experienced the spatial arrangements of the sequences and choose locations that best fit their synesthetic experience. When a presented stimulus did not induce a synesthetic experience, they could press the space bar and the next stimulus appeared. There were five practice trials to get familiar with the task. Control participants were asked to find an intuitive location for each stimulus. They were instructed to try to choose the same location every time the stimulus reappeared, but they were not allowed to choose the same location for every single stimulus. Control participants did not have the opportunity to press the space bar. Afterwards they were asked if they had used a certain strategy for placing the different stimuli.

Each stimulus was presented for 1 s in the centre of the screen, then a cross appeared and participants could choose a location for the stimulus. All stimuli were presented three times resulting in a total of 93 (= 31x3) trials. Completing this task took about 15 minutes. The three chosen locations for each item formed a triangular area and we used the mean surface of all these areas together as the consistency score of each participant, according to the procedure described by Rothen et al. (2016).

Drawings

After the consistency test, SSS participants were asked to draw their spatial forms on a piece of paper and control participants were asked to draw a representation of numbers, days and months. Sequence-space synesthetes could use coloured pencils if they experienced additional colours with their spatial forms. It was verified whether control participants really associated those locations with the sequences or whether they just chose the same locations as remembered from the consistency test. These drawings were used as a control measure for the consistency test because control participants could achieve high consistency scores as well when adopting a certain strategy, for example placing items of a particular sequence in a straight line from left to right. Besides giving more confidence in distinguishing real sequence-space synesthetes from controls, these drawings were for some SSS participants an easier method to express their visuospatial experiences. This was especially relevant for those who perceived sequences with high visual content like colours because the consistency test was only a purely spatially-based estimate (Jonas & Price,

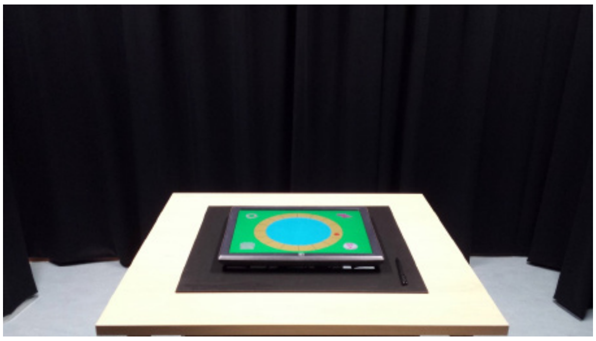


Fig. 1. Experimental set-up.

2014). This task took an additional 5 to 15 minutes, depending on the complexity of the drawings.

Morris Water Maze task

Finally, after the consistency test and drawings, participants performed a computerized version of the Morris Water Maze (MWM) task (Ring et al., 2018) to assess spatial navigation abilities. This task was written in Microsoft Visual Basic 6 and presented on a 19” ELO touchscreen with display resolution set to 1280x1024, controlled by a Dell computer running Windows XP. The screen was placed on a square table at comfortable height such that participants could easily reach the screen while standing. The table was surrounded by black curtains hanging from the ceiling to the ground and formed a 9 m2 separate area inside the room (Fig. 1). This prevented participants from using environmental cues, like doors and features on the wall, to guide navigation. There was enough space around the table such that participants could easily walk around the screen as instructed during the task. Lights were turned off to further reduce the influence of cues in the room.

On every trial, a virtual swimming pool environment was presented on the touchscreen. The display consisted of a blue circular area surrounded by an orange wall, representing water and the border of the pool, respectively. The green area outside the pool represented grass. There were four object cues (life ring, towel, chair, beach ball) around the pool, one in each corner of the screen (Fig. 2). Participants were instructed to find a hidden platform in the pool by moving over the touchscreen. They always had to start at the red dot that was presented in a fixed randomized order at the orange border. While searching for the platform, they used a touchscreen sensitive pen because this pen moved more easily over the screen than their finger and was therefore more accurate in registering the path that was taken. Participants were not allowed to lift the pen from the screen while searching for the platform and they were not allowed to cross the border of the pool. The platform, presented as a brown box, appeared once they passed the right location.

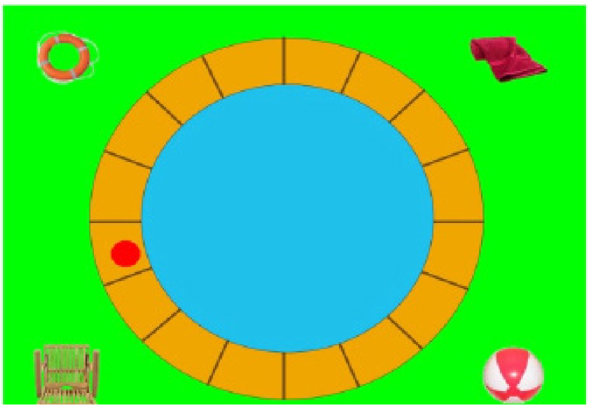


Fig. 2. Example of display.

Participants were asked to work out and learn the shortest possible path from the starting point to the hidden platform over several trials. There were three practice trials to get familiar with the task and to learn how to move properly over the screen. After these practice trials, participants had to perform five tasks, each consisting of 16 trials. The first task was always a place learning block and then two allocentric and two egocentric navigation blocks followed. These allocentric and egocentric blocks could be presented in any possible order, making a total of 16 different task orders. It was made sure that every order was performed by at least one synesthete and one control participant. For the allocentric tasks, participants had to use the object cues to find the platform and for the egocentric tasks, they had to use their own position as a reference. Importantly, participants had to figure out ‘the rule’ for finding the platform themselves.

Place learning

Place learning was used as a control condition to ensure that participants were able to perform the task properly and to check whether they showed learning over trials. There were no systematic manipulations during this condition. The objects, platform and participant kept the same positions (Fig. 3A).

Allocentric conditions

The two allocentric conditions were used to measure the strength of allocentric processing. In the first allocentric condition (Allocentric 1), the objects and platform stayed in the same location, but the participant changed position. The participant had to move to another side of the screen after every trial. This happened in a fixed randomized order. This condition was the original allocentric condition as developed by Feigenbaum and Morris (2004; Fig. 3B). In the second allocentric condition (Allocentric 2), the participant stayed in the same position, but the objects and platform changed in a fixed randomized order. The platform moved along with the objects, so they kept the same positions relative to each other. Participants did not see the platform and objects rotating, they only saw the new rotated order. This condition was added to the task by Ring et al. (2018; Fig. 3C).

Egocentric conditions

The two egocentric conditions were used to measure the strength of egocentric processing. In the first egocentric condition (Egocentric 1), the platform and participant stayed in the same location, but the objects rotated in a fixed randomized order. Participants did not see the objects rotating, they only saw the new rotated order. This condition was the original egocentric condition as developed by Feigenbaum and Morris (2004; Fig. 3D). In the second egocentric condition (Egocentric 2), the objects stayed in the same location, but now the platform and participant changed position in a fixed randomized order. The platform moved along with the participant, so platform and participant kept the same positions relative to each other. This condition was added to the task by Ring et al. (2018; Fig. 3E).

For each trial, participants had 60 seconds to find the platform. When a participant could not find the platform within these 60 seconds, a time out message appeared together with the platform. Every trial was followed by a distractor task. Participants had to 'pop' ten blue bubbles that appeared one by one at random locations on a black screen. After this distractor task, a black screen appeared with a yellow dot at one of the four sides of the screen, indicating on which side the participant had to stand for the upcoming trial.

After performing this virtual MWM task, participants were asked about their strategy for solving the task in order to control for the use of allocentric strategies in the allocentric conditions and egocentric strategies in the egocentric conditions. Additionally, they were asked about their navigational strategies in daily life. Performing the MWM task took about 30 minutes.

Data analysis

Consistency test

The three chosen locations for each item of the consistency test formed a triangular area and the mean surface in pixels across all these areas was our measure of consistency. The lower the score, the higher a participant's consistency of placing the items. Since sequence-space synesthetes always have the same spatial association for certain sequences, they were expected to consistently choose the same locations for the presented items. Controls do not have these spatial associations, therefore they were expected to choose these locations less consistently. However, because a substantial amount of control participants obtained high consistency scores by using certain strategies for placing the items, we eventually did not use a synesthesia cut-off score as in Rothen et al. (2016) to classify sequence-space

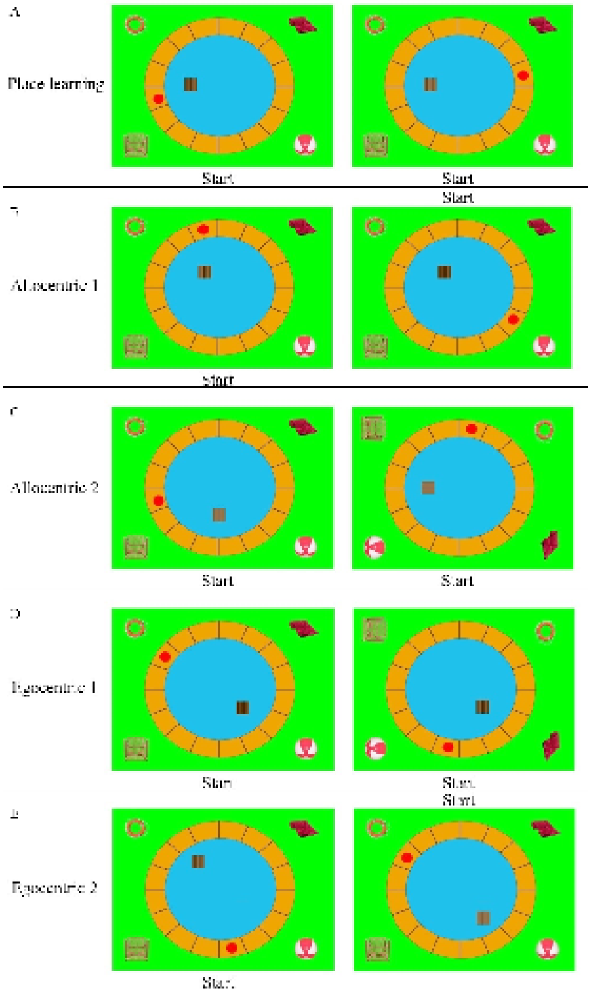


Fig. 3. Two example trials of each condition (A-E). 'Start' indicates the position of the participant relative to the screen. The platform was not visible for participants during the task, only after they passed the right location.

synesthetes and controls. Using only a cut-off score would have led to incorrect classification of participants.

Drawings

Instead of the consistency test, drawings and participants' descriptions of their synesthetic experiences – both from the questionnaire and questions in the lab – gave more confidence in correctly distinguishing sequence-space synesthetes from controls. Drawings were compared with the spatial forms generated by the consistency test and a comparison was made between drawings from synesthetes and controls. In addition to the drawings, synesthetes were asked how they perceived their visuospatial forms (e.g., projection vs. association, fixed perspective vs. multiple viewpoints). These subjective reports were compared to the answers given earlier (i.e., one up to five months prior to testing) in the self-report questionnaire. Controls were asked whether they always perceived the sequences like their drawings or whether the drawn locations were intuitive and not automatically elicited when thinking of those sequences.

Also, a complexity score was created for synesthetes based on the drawings and descriptions of their forms. For each sequence, ten complexity features were chosen, and one point was given for each feature that was present. The total score was divided by the number of sequences the synesthete perceived (i.e., one, two or three), so the score could range between zero and ten points. This score was used as a covariate in the analyses of the MWM task. An overview of the complexity features is presented in the Supplementary Materials (Table S9).

Morris Water Maze task

For the MWM task, four dependent measures were used to assess performance: (1) the length of the path taken to find the platform (Path Length), (2) the time needed to find the platform (Time to Target), (3) the percentage of time spent in the quadrant containing the platform, and (4) the angle of the path taken heading towards the platform after the first movement on the screen. We focused in our analyses on Time to Target and Path Length since these variables captured performance most directly and were straightforward to interpret. Moreover, these variables showed a clear learning effect over trials in contrast to the other two measures (Fig. S1 in the Supplementary Materials). For more details on why these latter measures were not included in further analyses, see General Discussion. Path Length was calculated as the difference between the shortest possible path and the actual path that was taken in order to enable comparison between trials. This measure was then transformed from pixel into mm. Time to Target was measured in milliseconds. Because of a very high variation in the data, both Path Length and Time to Target were square root transformed. Due to this transformation, the variation became less extreme for trials with long search times and path lengths (i.e., the first few trials). Data were analysed using repeated measures ANOVAs. As in Ring et al. (2018), these analyses were done for the Allocentric 1 and Egocentric 1 conditions and for the Allocentric 2 and Egocentric 2 conditions. These analyses were done separately for conditions 1 and 2, because the first conditions were the original conditions of the MWM task developed by Feigenbaum and Morris (2004) – and reflected real life situations more accurately (see General Discussion) – and the second conditions were the added conditions developed by Ring et al. (2018). Individual differences in the manifestation of SSS were used as between-subject factors in further exploratory analyses.

Additionally, there were several measures to control for correct performance. It was counted how often participants left the pool area, how often they lifted the pen from the screen, when they were timed out (i.e., when they could not find the platform within 60 seconds), and how much time they needed to complete the distractor task. Control measures were analysed using repeated measures ANOVAs.

Results

Data of the consistency test were analysed using t-tests and data of the MWM task using repeated measures ANOVAs. If the sphericity assumption was violated, Greenhouse-Geisser correction (GG) was applied. The significance level for all analyses was set to $\alpha = .05$.

Consistency test

When taking the consistency scores for numbers, months and days together, an independent-samples t-test indicated that sequence-space synesthetes performed significantly more consistently than controls ($t(22) = -2.18$, $p = .040$, $\Delta = .49$). Also for numbers and months separately – but not for days – sequence-space synesthetes performed significantly

Table 1. Descriptive statistics of the consistency scores of both groups for numbers, months and days together and separately.

	SSS		Con		
	M	***SD***	***M***	***SD***	***N* SSS/Con**
Total	2320	2125	6477	8495	23/21
Numbers	1335	1119	5125	8176	13/21
Months	2317	1985	7651	11473	23/21
Days	2456	4066	5969	12417	21/21

more consistently: for numbers (*t*(21) = -2.09, *p* = .049, Δ = .46) and for months (*t*(21) = -2.10, *p* = .048, Δ = .46). Because of unequal variances between groups (Levene's test was significant), the degrees of freedom were adjusted, and Glass's delta was used for determining the effect size. The descriptive statistics of both groups are summarized in Table 1 and the spreading of the individual consistency scores for numbers, months and days together are shown in Figure 4. Despite the significant difference in consistency between groups, this figure shows a clear overlap in the consistency scores between synesthetes and controls. A few examples of generated figures of performance of both sequence-space synesthetes and control participants are presented in the Supplementary Materials (Fig. S8 and S9).

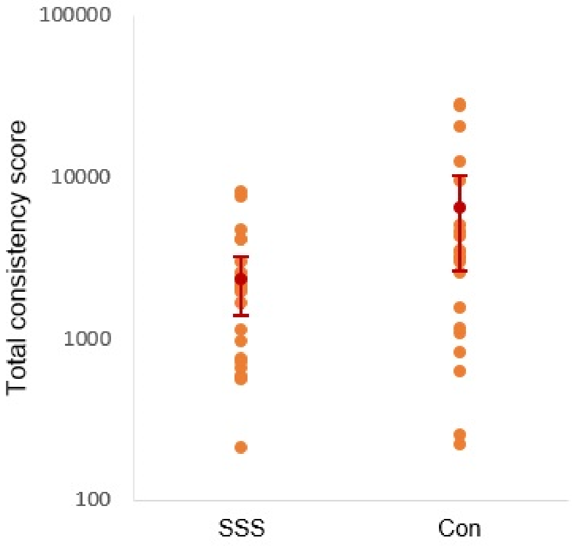


Fig. 4. Spreading of the individual consistency scores of both groups for numbers, months and days together. The red dot indicates the mean and error bars represent the 95% confidence interval. The scale of the y-axis is logarithmic.

Drawings

A comparison of drawings revealed some striking differences between sequence-space synesthetes and controls. Synesthetes showed a tendency to connect items of a sequence within a form (e.g., months connected as blocks in a circle), whereas controls just wrote the items down in isolation (e.g., months scattered on locations where they could remember them). For months, around 70% of the synesthetes connected items within a form, for days around 67% and for numbers around 54%. None of the control participants did this. In terms of complexity, drawings made by synesthetes were characterized by more elaborate and complex forms compared to those of controls. For months, around 65% of the synesthetes drew a circle, oval, square or a similar closed form while for days around 43%. For numbers, around 62% of the synesthetes drew a line with bends, corners or zigzags. None of the control participants did this. Most of them arranged the items of a sequence in rows or columns. For months, around 62% of the control participants did this, for days around 71% and for numbers around 67%. A detailed overview of the characteristics of each sequence for both groups can be found in the Supplementary Materials (Table S7 and S8) as well as a few examples of drawings made by sequence-space synesthetes and control participants (Fig. S10 and S11).

Morris Water Maze task

All analyses presented here focus on performance measured by the time needed to find the platform and the length of the path taken to find the platform. The results of the other two measures (i.e., Percentage of Time in Target Quadrant and Path Angle) are presented in the Supplementary Materials (Table S4 and S5 and Fig. S3 and S4). The complete tables with descriptive and inferential statistics for Time to Target and Path Length are presented in the

Supplementary Materials as well (Table S2 and S3). For Time to Target, not more than 2% of the trials per condition were outliers, and for Path Length this was not more than 3% (see Supplementary Materials for how outliers are dealt with).

Place learning

Data were analysed using a repeated measures ANOVA among the between-subjects factor group (SSS, controls) and the within-subjects factor trial (16 trials). For performance measured by Time to Target, a significant main effect was found for trial, $F(7.26, 304.98) = 20.21$, $p < .0001$, $\eta_p^2 = .33$, GG, meaning that the time needed to find the platform decreased over trials. There was no significant main effect for group or a significant group x trial interaction, which indicates similar learning over trials for both groups. Similar results were found for performance measured by Path Length (Table S1 and Fig. S1 in the Supplementary Materials).

Allocentric 1 and Egocentric 1

Data were analysed using a repeated measures ANOVA among the between-subjects factor group (SSS, controls) and the within-subjects factors trial (16 trials) and condition (allocentric, egocentric). For performance measured by Time to Target, significant main effects were found for trial, $F(6.40, 268.77) = 45.33$, $p < .0001$, $\eta_p^2 = .52$, GG, and for condition, $F(1, 42) = 6.44$, $p = .015$, $\eta_p^2 = .13$, showing that the time needed to find the platform decreased over trials and that it took more time to find the platform in the allocentric condition ($M = 47.47$, $SD = 15.67$) compared to the egocentric condition ($M = 41.71$, $SD = 14.02$) for most trials, suggesting that the allocentric condition was more

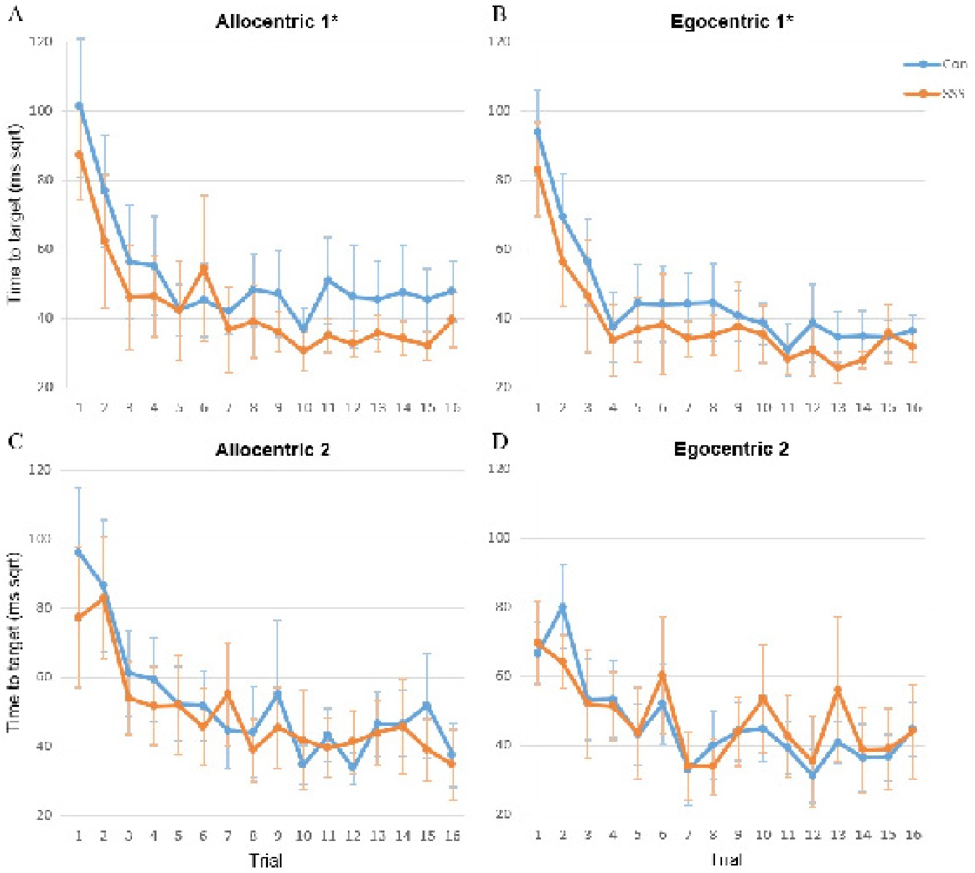


Fig. 5. Graphs showing the performance over trials measured by the time needed to find the platform in the Allocentric 1 and Egocentric 1 conditions (A and B) and the Allocentric 2 and Egocentric 2 conditions (C and D). The orange lines show the performance of sequence-space synesthetes, blue lines show the performance of control participants. Error bars reflect the 95% confidence interval. *Synesthetes performed significantly better in the Allocentric 1 and Egocentric 1 conditions compared to controls. The peaks in the learning curves can be explained by the fact that the starting point (red dot) appeared on the same location for every participant in a particular trial. For instance, the starting point in the 7th trial of Egocentric 2 was already close to the location of the platform, resulting in generally shorter times to find the platform. Likewise, the starting point in the 6th trial of Egocentric 2 was relatively far away from the platform, resulting in generally longer times to find the platform.

difficult. Importantly, a significant main effect of group, $F(1, 42) = 4.46$, $p = .041$, $\eta_p^2 = .01$, indicated that sequence-space synesthetes performed better at both conditions ($M = 40.83$, $SD = 13.17$) compared with controls ($M = 48.70$, $SD = 11.36$). There was no significant group x condition interaction. Figures 5A and B show performance over trials for both groups for the Allocentric 1 and the Egocentric 1 conditions, respectively. Similar results were found for performance measured by Path Length, except that the main effect of group was marginally significant (Table S2 and Fig. S2 in the Supplementary Materials).

Allocentric 2 and Egocentric 2

Data were analysed using a repeated measures ANOVA among the between-subjects factor group (SSS, controls) and the within-subjects factors trial (16 trials) and condition (allocentric, egocentric). For performance measured by Time to Target, significant main effects were found for trial, $F(6.07, 255.11) = 26.37$, $p < .0001$, $\eta_p^2 = .39$, GG, and for condition, $F(1, 42) = 4.17$, $p = .048$, $\eta_p^2 = .09$, showing that the time needed to find the platform decreased over trials and that it took more time to find the platform in the allocentric condition ($M = 51.00$, $SD = 16.12$) compared to the egocentric condition ($M = 46.94$, $SD = 18.35$) for most trials, suggesting that the allocentric condition was more difficult. There was a significant trial x condition interaction as well, $F(8.17, 343.08) = 3.54$, $p = .001$, $\eta_p^2 = .08$, GG, indicating that the time needed to find the platform decreased more over trials for the allocentric condition compared to the egocentric condition. There was no significant main effect of group or a significant group x condition interaction, meaning that sequence-space synesthetes and controls showed similar performance in both conditions. Figures 5C and D show performance over trials for both groups for the Allocentric 2 and the Egocentric 2 conditions, respectively. Similar results were found for performance measured by Path Length, except for a significant main effect of condition (Table S2 and Fig. S2 in the Supplementary Materials).

Control measures

There were no significant differences between groups for any of the control measures, meaning that any variations in the ability to correctly perform the task did not affect the results. The descriptive and inferential statistics are presented in the Supplementary Materials (Table S6).

Effects of synesthetic features on performance

Individual differences in the manifestation of SSS were taken into account in further analyses to see if some specific synesthetic features contributed to the enhanced performance of synesthetes in the Allocentric 1 and Egocentric 1 conditions. First, the specific hypothesis of enhanced performance in the Allocentric 1 condition of synesthetes among the ability to rotate their spatial forms was tested with a repeated measures ANOVA with the between-subjects factor group (rotation yes/no) and the within-subjects factor trial (16 trials). As shown in Figure 6, it seemed that synesthetes with the mental rotation ability learned faster in the Allocentric 1 condition compared to synesthetes who could not do this. This effect was however not significant (first eight trials ($F(1, 21) = 2.72$, $p = .11$, $\eta_p^2 = .12$)). The groups consisted of 15 synesthetes with the rotation ability and eight synesthetes who could not do this. A synesthete was classified as having the rotation ability if he or she had this for at least one of the sequences.

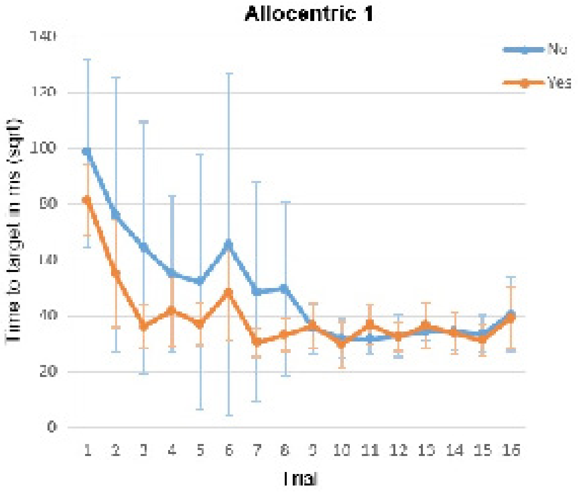


Fig. 6. Graph showing the performance over trials measured by the time needed to find the platform in the Allocentric 1 condition. The orange line shows the performance of synesthetes who were able to mentally rotate their spatial forms and the blue line the performance of synesthetes who could not do this. Error bars reflect the 95% confidence interval. Synesthetes with the mental rotation ability seemed to perform better during the learning phase.

Six other synesthetic features with different manifestations among sequence-space synesthetes were taken into further exploratory analyses to see if some of these features contributed to the better performance of synesthetes. The features taken into account were: association vs. projection, perceiving the spatial form in a two-dimensional vs. three-dimensional space, the ability to mentally move over the form (yes/no), the ability to zoom in and out (yes/no), whether the form itself moves with time (yes/no) and the presence of additional visual features like colour (yes/no). A synesthete was classified as having the ability if he or she demonstrated or reported this ability for at least one of the sequences. These features were taken as between-subject factors in separate repeated measures ANOVAs with only the synesthete group. No further effects of synesthetic features on performance were found for the Allocentric 1 and Egocentric 1 conditions. However, analyses amongst the Allocentric 2 and Egocentric 2 conditions revealed a few effects concerning synesthetic features that represent an allocentric (form itself moves) and egocentric (move over form) perspective on the spatial form. These results are presented in the Supplementary Materials (Fig. S5-S7). It is important to note that due to the many exploratory analyses that were performed (six synesthetic features x four conditions), some of these observed effects could have been significant by chance (i.e., at least one significant effect was expected for 24 analyses with a significance level set at $\alpha = .05$).

Next to these exploratory analyses with synesthetic features, it was examined whether the performance of synesthetes was modulated by the complexity of their synesthetic experience. A complexity score – based on synesthetic features and synesthetes' drawings – was taken as a covariate in a repeated measures ANCOVA with the within-subjects factor trial (16 trials). This complexity score did not significantly modulate the performance of synesthetes for any condition (all $F(1, 21) < 2.74$, n.s.).

Furthermore, participants were asked about their daily navigation strategies, but due to insufficient difference in used strategies among participants, this factor could not be taken into analysis. Most participants reported to use allocentric navigation strategies or a combination of both allocentric and egocentric strategies, whereas almost no one reported to favour egocentric navigation strategies.

Discussion

The aim of this study was to investigate whether SSS is beneficial for spatial navigation. To test this, sequence-space synesthetes and control participants performed a virtual Morris Water Maze task involving two allocentric and two egocentric navigation conditions in which they had to find a hidden platform. Known cognitive benefits of SSS, in particular, the ability of synesthetes to mentally rotate their spatial forms, were expected to be reflected in enhanced performance on this navigation task. Indeed, sequence-space synesthetes showed better performance in one of the allocentric conditions and one of the egocentric conditions (i.e., Allocentric 1 and Egocentric 1, the two original test conditions developed by Feigenbaum and Morris, 2004). Specifically, synesthetes with the ability to mentally rotate their spatial forms seemed to learn faster during the first trials of the allocentric task. As Ring et al. (2018) suggested, especially for this task, mental rotation is important for successful performance. Participants had to change position while the platform and the objects remained fixed. Participants thus saw the display from a different perspective on every trial and had to mentally rotate it back to its original perspective (Ring et al., 2018). For the egocentric condition, in which sequence-space synesthetes showed enhanced performance as well, none of the synesthetic features that we took into account were found to specifically contribute to their better performance. In this task, the platform and the participant kept the same positions, while the objects rotated. The current results add to the existing literature concerning the cognitive benefits of SSS and are consistent with the previously found mental rotation advantage.

It may seem surprising that sequence-space synesthetes did not perform significantly better than controls in the other allocentric and egocentric tasks (i.e., Allocentric 2 and Egocentric 2, the two conditions added by Ring et al., 2018). However, in this allocentric task, the platform moved along with the objects while the participant kept the same position and in this egocentric task, the platform moved along with the participant while the objects remained fixed. This movement of the platform with either the objects or the participant does never happen in everyday environments (Ring et al., 2018). Normally, when we navigate to certain destinations (e.g., buildings), these buildings do not change in space. They remain at fixed locations, like the platform did in the Allocentric 1 and Egocentric 1 tasks. Therefore, the first allocentric and egocentric tasks seemed to better reflect spatial navigation in daily life. An analysis with all four conditions in one repeated measures ANOVA confirmed that the Allocentric 2 and Egocentric 2 conditions were significantly more difficult than the Allocentric 1 and Egocentric 1 conditions (F(1, 42) = 4.35, p = .043, η_p^2 = .09). Averaged over trials, it took more time to find the platform in the second conditions (M = 48.97, SD = 15.88) compared to the first conditions (M = 44.59, SD = 12.83). Figure 5 clearly demonstrates this as well by less smooth learning curves for these conditions compared to the Allocentric 1 and Egocentric 1 conditions.

The enhanced performance of synesthetes compared to controls in the two original task conditions cannot be explained by any differences between synesthetes and control participants in the ability to follow the task instructions, since there were no differences between groups in any of the control measures (i.e., the number of times they left the pool area, the number of times they lifted the pen from the screen and the number of times they could not find the platform within 60 seconds). Also the experienced time interval between tasks, indicated by the time they needed to complete the distractor task, was not different between groups. An alternative explanation for the observed group difference in task performance is that sequence-space synesthetes might have been more interested in participating, causing them to be more motivated to perform well. However, this alternative explanation cannot fully account for the observed results, because there was no performance difference between synesthetes and controls in either the place learning condition (i.e., the control task), the control measures, the learning curve over trials or in the Allocentric 2 and Egocentric 2 conditions. The existence of a motivational difference between groups would have been evident in differences in performance here.

Ring et al. (2018) only found differences between groups on the original task conditions as well. Interestingly, they found that individuals with ASD performed significantly worse than control participants in the Allocentric 1 task, while the current study demonstrated that individuals with SSS performed significantly better than controls in this exact same task. This implies that the previously found link between ASD and synesthesia, i.e., synesthesia is more common among individuals with ASD, with a prevalence of 20% Baron-Cohen et al., 2013; Neufeld et al., 2013), is not reflected in performance on this virtual MWM navigation task. Since spatial navigation is a complex mental task involving many sub-processes, it is very well possible that the two groups do not converge to similar performance on the MWM task. Recent studies indicate that the shared cognitive characteristics between SSS and ASD seem to mainly involve elevated attention to detail (Mealor et al., 2016; Ward et al., 2017).

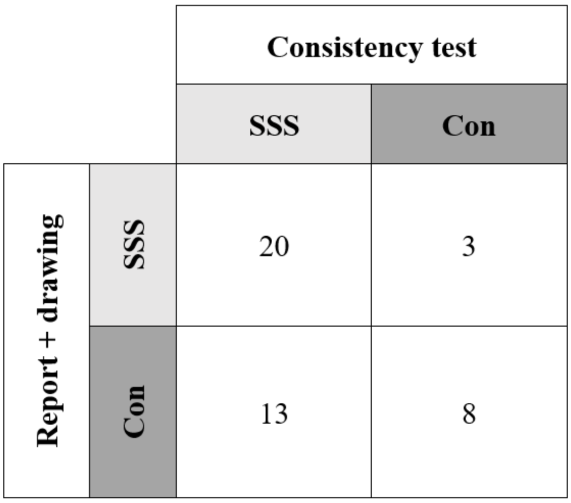
An interesting question is whether SSS is an adaptive, rather than an epiphenomenal, cognitive function. Perhaps the synesthetic visuospatial experiences remain to exist throughout generations because they are beneficial for a broad range of cognitive functions. The possibility of synesthetes to mentally manipulate the spatial forms (e.g., rotating the forms in order to see them from multiple perspectives) is beneficial for spatial thinking and, as the current study suggests, for spatial navigation. The ability to easily keep an overview of things that need to be done (without a planner) is clearly a memory related advantage and remembering important events and dates, like birthdays, is socially relevant as well. When sequence-space synesthetes are asked whether they experience any benefits of their visuospatial forms, most of them indeed report to experience a memory advantage and state that they cannot imagine living without the visuospatial forms. SSS thus clearly has personal importance. The current data adds to the debate whether SSS could indeed be an adaptive cognitive function.

Concerning the methodological aspects of the current study, there is one important difference compared to the study performed by Ring et al. (2018). We focused in our analyses on the time that was needed to find the target and the length of the path that was taken to find the target, while they focused on the percentage of time that was spent in the target quadrant. The latter measure, however, did not seem to be the most suitable, in contrast to what Ring et al. (2018) suggested. The first reason to doubt this measure is that during one of the conditions (Allocentric 2) the platform was always at the border of two quadrants (i.e., in between the ball and chair and moved along with these objects). Therefore, it was not possible to correctly define the target quadrant for trials in this condition. The second reason was the absence of a correct reflection of a learning effect over trials (Fig. S1 and S3 in the Supplementary Materials). We assumed that the time spent in the target quadrant should increase over trials, when participants started spending more time searching in the correct quadrant containing the platform. In practice, we observed that after learning, participants moved in one straight line from the starting point to the platform. In this way, they spent almost no time searching in the quadrant containing the platform. Ring et al. (2018) chose this measure because of its frequent use in the literature and its reduced vulnerability to variation among participants since it is expressed as a percentage of the total search time. Ring et al. (2018) suggested that a high variation among participants might have added noise to the data of Time to Target and Path Length and possibly obscured any differences between groups. However, based on our measurements and observations, the reasons they put forward do not seem to outweigh the two major problems that come along with this measure.

A caveat of the current study is that we did not include an additional mental rotation task in order to confirm whether mental rotation performance was indeed correlated with performance in the Allocentric 1 condition. The enhanced performance of synesthetes with the ability to mentally rotate their spatial forms, observed during the learning phase of this allocentric task, was not significant. A positive correlation between a mental rotation task and the Allocentric 1 condition therefore would have added to the evidence in favour of the contribution of synesthetes' mental rotation ability to the enhanced performance in this condition. Future studies should aim to replicate this observed mental rotation advantage of synesthetes for allocentric navigation with a larger sample size, because the current study only included 15 synesthetes with the mental rotation ability and eight synesthetes without this ability.

One might argue that the Allocentric 1 condition could still be egocentric due to continuous updating of participants' own spatial position relative to the platform during their movement in between trials to any of the four sides of the screen. So instead of finding the platform by using the object cues, participants might represent and update the relation between the platform and their own position (Simons & Wang, 1998). This so-called viewer-centered representation may trigger an egocentric strategy. In order to ensure that the Allocentric 1 condition can only be solved by using allocentric navigation strategies, in future studies, participants could be moved in a different way disrupting visual, vestibular and proprioceptive information. This prevents the updating mechanism to adjust for changes in participants' position (e.g., participants could be moved in a spinning wheelchair while covering their eyes, like in Simons and Wang, 1998).

Another suggestion for future studies with the aim to further investigate spatial navigation skills of sequence-space synesthetes is to conduct a navigation experiment in a virtual reality set-up. Such a set-up would more realistically reflect navigation in daily environments compared to the current approach. Therefore, it would convey different and perhaps converging information in order to answer the question whether sequence-space synesthetes have enhanced spatial navigation skills. A virtual reality experiment could thus expand on the current findings. These future studies could explore as well which specific synesthetic features contribute to the enhanced egocentric navigation of sequence-space synesthetes, since the features that we took into account could not explain this enhanced performance.



		Consistency test	
		SSS	Con
Report + drawing	SSS	20	3
	Con	13	8

Fig. 7. Scheme representing the suggested classification of participants based on the consistency test (Rothen et al., 2016) and the actual classification based on self-report and drawings. Three sequence-space synesthetes and thirteen control participants would have been misclassified when using a consistency cut-off score.

In addition to the results of the MWM task, this study gave more insights into reliable classification of sequence-space synesthetes and control participants. The consistency test (Rothen et al., 2016) was a valuable addition to check whether synesthetes consistently chose the same locations for items of sequences, but using only a consistency cut-off score would have led to an incorrect classification of participants. For the average area-based consistency score, Rothen et al. (2016) suggested to use a cut-off score of 0.2029% of the total monitor area, resulting in a cut-off score of 4207 for our study. Using this score would have led to the classification presented in Figure 7, which is highly deviant from the classification based on self-report and drawings. Rothen et al. (2016), however, increased the fit of their classification by excluding participants who had used certain strategies (e.g., placing items of a sequence on a horizontal straight line). We did not exclude these control participants which could explain this classification difference.

Instead of only using the results of the consistency test, drawings and participants' descriptions of their synesthetic experiences gave more confidence in correctly distinguishing sequence-space synesthetes from controls. Most studies investigating synesthesia have distinguished synesthetes from controls by using both subjective reports and consistency tests (Brang, Teuscher, Ramachandran, & Coulson, 2010). The current study demonstrated that drawings of spatial forms can serve as an important classification tool as well. Drawings of synesthetes were generally more complex and characterized by certain shapes (e.g., closed forms), while controls commonly arranged items of a sequence in rows or columns. This is consistent with previous reports of synesthetes experiencing months mostly in circular arrangements, while controls use rows or single straight lines as default (e.g., Brang et al., 2010; Eagleman, 2009). Importantly, synesthetes showed a tendency to connect items of a sequence within a form by blocks or lines. None of the control participants did this. This fits with synesthetes' reports that an item of a sequence – in particular, months and days – often "encompasses a region of space rather than a single location" (Brang et al., 2010, p. 316). Next to shape and complexity, this tendency of connection is therefore an important feature that characterizes synesthetes' drawings and may contribute to classification.

Conclusion

Sequence-space synesthetes have enhanced spatial navigation skills in a virtual navigation task. This study provides the first evidence for a spatial navigation benefit in SSS and the next question is whether these results translate to spatial navigation in daily environments. The findings of the current study add to the existing literature showing cognitive benefits of SSS and are consistent with the previously found mental rotation advantage. This study therefore contributes to a better understanding of SSS at a cognitive level and – the finding that mental rotation seems to be important for allocentric navigation – extends our knowledge about the cognitive processes involved in allocentric spatial navigation strategies.

Acknowledgements

We would like to thank Melanie Ring for providing the MWM task and Dr. Nicolas Rothen for providing the consistency task. We thank the Technical Support Group (Radboud University Nijmegen, Faculty of Social Sciences) for their invaluable help with making the MWM task run on the touchscreen and arranging the task set-up. We are grateful to all our participants for their interest and time investment and we would like to thank Dr. Surya Gayet for valuable comments and suggestions for improvement of the manuscript.

References

Baron-Cohen, S., Johnson, D., Asher, J., Wheelwright, S., Fisher, S. E., Gregersen, P. K., & Allison, C. (2013). Is synaesthesia more common in autism? *Molecular autism, 4*(1), 40.

Bohobot, V. D., Lerch, J., Thorndycraft, B., Iaria, G., & Zijdenbos, A. P. (2007). Gray matter differences correlate with spontaneous strategies in a human virtual navigation task. *Journal of Neuroscience, 27*(38), 10078-10083.

Brang, D., Miller, L. E., McQuire, M., Ramachandran, V. S., & Coulson, S. (2013). Enhanced mental rotation ability in time-space synaesthesia. *Cognitive Processing, 14*(4), 429-434.

Brang, D., Teuscher, U., Ramachandran, V. S., & Coulson, S. (2010). Temporal sequences, synesthetic mappings, and cultural biases: the geography of time. *Consciousness and Cognition, 19*(1), 311-320.

Chersi, F., & Burgess, N. (2015). The cognitive architecture of spatial navigation: hippocampal and striatal contributions. *Neuron, 88*(1), 64-77.

Cohen Kadosh, R., Gertner, L., & Terhune, D. B. (2012). Exceptional abilities in the spatial representation of numbers and time: insights from synesthesia. *The Neuroscientist, 18*(3), 208-215.

Di Tore, P. A., Corona, F., & Sibilio, M. (2014). Orienteering: spatial navigation strategies and cognitive processes. *Journal of Human Sport & Exercise, 10*(1).

Eagleman, D. M. (2009). The objectification of overlearned sequences: A new view of spatial sequence synaesthesia. *Cortex, 45*(10), 1266-1277. doi:10.1016/j.cortex.2009.06.012

Feigenbaum, J. D., & Morris, R. G. (2004). Allocentric versus egocentric spatial memory after unilateral temporal lobectomy in humans. *Neuropsychology, 18*(3), 462-472.

Gertner, L., Henik, A., & Cohen Kadosh, R. (2009). When 9 is not on the right: Implications from number-form synesthesia. *Consciousness and Cognition, 18*(2), 366-374.

Guariglia, C., Piccardi, L., Iaria, G., Nico, D., & Pizzamiglio, L. (2005). Representational neglect and navigation in real space. *Neuropsychologia, 43*(8), 1138-1143.

Hale, J., Thompson, J. M., Morgan, H. M., Cappelletti, M., & Cohen Kadosh, R. (2014). Better together? The cognitive advantages of synaesthesia for time, numbers, and space. *Cognitive Neuropsychology, 31*(7-8)*,* 545-564.

Harris, M. A., Wiener, J. M., & Wolbers, T. (2012). Aging specifically impairs switching to an allocentric navigational strategy. *Frontiers in Aging Neuroscience, 4,* 29.

Havlik, A. M, Carmichael, D. A, & Simner, J. (2015). Do sequence-space synaesthetes have better spatial imagery skills? Yes, but there are individual differences. *Cognitive Processing, 16*(3)*,* 245-253.

Iaria, G., Petrides, M., Dagher, A., Pike, B., & Bohobot, V. D. (2003). Cognitive strategies dependent on the hippocampus and caudate nucleus in human navigation: variability and change with practice. *Journal of Neuroscience, 23*(13), 5945-5952.

Jonas, C. N., & Price, M. C. (2014). Not all synesthetes are alike: spatial vs. visual dimensions of sequence-space synaesthesia. *Frontiers in Psychology, 5,* 1171.

Konishi, K., & Bohobot, V. D. (2013). Spatial navigational strategies correlate with gray matter in the hippocampus of healthy older adults tested in a virtual maze. *Frontiers in Aging Neuroscience, 5,* 1.

Mealor, A. D., Simner, J., Rothen, N., Carmichael, D. A., & Ward, J. (2016). Different dimensions of cognitive style in typical and atypical cognition: new evidence and a new measurement tool. *PLoS ONE, 11*(5), e0155483.

Morris, R. G. (1981). Spatial localization does not require the presence of local cues. *Learning and Motivation, 12*(2), 239-260.

Neufeld, J., Roy, M., Zapf, A., Sinke, C., Emrich, H. M., Prox-Vagedes, V., … & Zedler, M. (2013). Is synesthesia more common in patients with Asperger syndrome? *Frontiers in Human Neuroscience, 7,* 847.

Price, M. C. (2009). Spatial forms and mental imagery. *Cortex, 45*(10)*,* 1229-1245.

Price, M. C., & Pearson, D. G. (2013). Toward a visuospatial developmental account of sequence-space synaesthesia. *Frontiers in Human Neuroscience, 7,* 689.

Ring, M., Gaigg, S. B., Altgassen, M., Barr, P., & Bowler, D. M. (2018). Allocentric versus egocentric spatial memory in adults with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 48*(6), 2101-2111.

Rizza, A., & Price, M. C. (2012). Do sequence-space synaesthetes have better spatial imagery skills? Maybe not. *Cognitive Processing, 13,* 299-303.

Rothen, N., Jünemann, K., Mealor, A. D., Burckhardt, V., & Ward, J. (2016). The sensitivity and specificity of a diagnostic test of sequence-space synaesthesia. *Behavior Research Methods, 48*(4), 1476-1481.

Rothen, N., Meier, B., & Ward, J. (2012). Enhanced memory ability: Insights from synaesthesia. *Neuroscience and Biobehavioral Reviews, 36*(8), 1952-1963.

Simner, J., Mayo, N., & Spiller, M. J. (2009). A foundation for savantism? Visuo-spatial synaesthetes present with cognitive benefits. Cortex, 45(10), 1246-1260.

Simons, D. J., & Wang, R. F. (1998). Perceiving real-world viewpoint changes. Psychological Science, 9(4), 315-320.

Smilek, D., Callejas, A., Dixon, M. J., & Merikle, P. M. (2007). Ovals of time: Time-space associations in synaesthesia. Consciousness and Cognition, 16(2), 507-519.

Ward, J., Hoadley, C., Hughes, J. E., Smith, P., Allison, C., Baron-Cohen, S., & Simner, J. (2017). Atypical sensory sensitivity as a shared feature between synaesthesia and autism. Scientific Reports, 7, 41155.

Ward, J., Sagiv, N., & Butterworth, B. (2009). The impact of visuo-spatial number forms on simple arithmetic. Cortex, 45(10), 1261-1265.

Across-Session Consistency of Context-Dependent Language Processing: Towards a Clinical Tool

Natascha Roos[1]
Supervisors: Vitória Piai[1, 2]

[1]*Radboud University Nijmegen, Donders Institute for Brain, Cognition and Behaviour, The Netherlands*
[2]*Radboud University Medical Centre Nijmegen, Donders Institute for Brain, Cognition and Behaviour, The Netherlands*

Transient aphasias after tumor removal from the left hemisphere are commonly exhibited by patients after the surgery. In most cases, the deficits resolve within weeks. The present study serves to set the parameters for a clinical tool to track this process of language recovery over time. The main aims were to determine a suitable imaging method, the corresponding across-session consistency, and the effect size for a shorter testing duration. For this purpose, 30 native Dutch speakers were tested with magnetoencephalography (MEG) or functional magnetic resonance imaging (fMRI) while performing the same picture naming as sentence completion task (15 participants per method). Sentences were either constrained or unconstrained towards the picture, such that participants could retrieve the target word through sentence context (constrained sentences) or had to wait for the picture to appear (unconstrained sentences) to be able to name it. Behavioral results show a strong reaction time effect for picture naming in the MEG as well as the fMRI experiment, verifying that constrained sentence context primes the target word before the picture is shown. The MEG results reveal alpha-beta power decreases (10 - 20 Hz) in the left temporal and inferior parietal lobe that yield a Dice coefficient quantifying the across-session consistency of activation of 0.49. Analyzing only the first half of all MEG sessions reduces the area of the alpha-beta power decreases and lowers the Dice coefficient to 0.35 but increases the significance of the power decrease cluster. The fMRI results reveal BOLD signal increases for constrained over unconstrained sentences mostly in the left inferior temporal and parietal lobe but also bilaterally in motor areas, resulting in a Dice coefficient of 0.43. Analysis of the first half of the fMRI sessions diminishes the obtained BOLD increase clusters and lowers the Dice coefficient to 0.31. Based on spatiality, consistency, and significance of the obtained effect profiles with each method, the findings of the present study lead to conclude that MEG is a more suitable imaging method for the clinical tool than fMRI.

Keywords: MEG, fMRI, tumor, aphasia

Corresponding author: Natascha Roos; **E-mail:** natascha.roos1693@gmail.com

The present study is the first step of a project aiming to develop a clinical tool to track language-related changes of brain activity in patients. More precisely, the broader aim is to scan patients with tumors in language areas of the brain before as well as after they undergo tumor surgery to investigate the recovery process of language in the brain. Many of these patients show deficits in language use shortly after the surgery. However, these deficits are mostly temporary and can resolve within the following weeks. To capture and better understand this recovery process of the brain, the first step of the project is a study with healthy control participants to assess the optimal parameters for patient testing. This is what constitutes the present study and will be referred to as such in the following report. The longitudinal patient study will be a future step within the project, after the clinical tool is finalized and ready for patient testing.

Patients with tumors in language areas in the left hemisphere of the brain often suffer from aphasias after tumor removal. The term aphasia describes language deficits due to brain damage in one or more language domains such as speaking, comprehending, reading, or writing (see Dronkers & Baldo, 2010, for a review of different types of aphasia). However, these patients' language deficits due to resection surgery are mainly temporary and can resolve within the following weeks. For example, 110 patients performed language tests before undergoing surgery as well as two to three days and one month after (Wilson et al., 2015). Most patients showed normal language scores before surgery and a major decrease in performance two to three days after the surgery. But these decreases mostly resolved within one month such that the language scores of the second postsurgical test did not significantly differ from presurgical test scores. Further, this study also shows that the location where the tumor was resected determines the language domains that are affected, and thus the type of aphasia that is observed.

A meta-analysis of hemodynamic studies conducting language tasks with stroke patients suffering from aphasias concluded that the activated brain areas in patients are coherent across studies, including task-related activity in the right-hemisphere (Turkeltaub, Messing, Norise & Hamilton, 2011). Further, another study with aphasic stroke patients employing functional magnetic resonance imaging (fMRI) has shown that the amount of right-hemisphere activity for language tasks depends on the size of left-hemisphere lesions and influences language recovery (Skipper-Kallal, Lacey, Xing & Turkeltaub,2017). This suggests that the brain can compensate by employing the right hemisphere for language functions if the left hemisphere is lesioned.

To address this neuroplasticity of language with electrophysiological measures, Piai, Meyer, Dronkers and Knight (2017a) conducted a combined electroencephalography (EEG), behavioral, and structural connectivity study with patients that had suffered a stroke in the left hemisphere. In previous electrophysiological studies, employing a sentence completion task with neurotypical control subjects, the authors observed power decreases of brain activity in the alpha-beta frequency range (8-30 Hz) for context-driven word retrieval (Piai, Roelofs & Maris, 2014; Piai, Roelofs, Rommers & Maris, 2015). Interestingly, when conducting the same task with left-hemisphere stroke patients they discovered the same power decrease in the alpha-beta range as in neurotypical control subjects, but lateralized to the right instead of the left hemisphere in patients (Piai et al., 2017a; Piai, Rommers & Knight, 2017b).

These findings indicate that left and right hemispheres can perform similar neuronal computations and that patients with lesions in the left hemisphere can supportively draw on their intact right hemisphere for language use. As such, the brain can uniquely reorganize itself for language use in recovery. This stresses the individuality of the functional organization of the brain, which has been argued to be an important aspect for further investigation to enhance surgery outcomes and patient therapy (Duffau, 2005).

In accordance with this argumentation, a recent review shows that different language tasks have resulted in diverging findings in terms of the functional organization of language in the brain, suggesting that it might be diversely independent of the language domains and processes that are involved (Bradshaw, Thompson, Wilson, Bishop & Woodhead, 2017). This divergence can have either individual or task-dependent impacts, but might also merely derive from differences in methodology across studies. Thus, the authors of the review call for an increase in methodological consistency to increase the comparability of results from different studies. This would also help to gain further insights into the independent organization of language processes in the brain.

Following this line of reasoning, Wilson, Bautista, Yen, Lauderdale and Eriksson (2017) investigated the validity and reliability of different language production and comprehension paradigms to identify language areas with fMRI. Here, validity refers to the property of the paradigm to activate all and only those brain areas that have been shown to be essential for language processing, whereas reliability describes the consistency of measurements across more than one session. They conclude that sentence completion tasks provide the best-balanced combination of validity and reliability. However, they also point out general limitations of language mapping with fMRI in individuals for clinics and research, and prompt for equal assessments and comparisons of different paradigms.

To date, there have been several findings demonstrating transient aphasias after left-hemisphere damage (Skipper-Kallal et al., 2017; Wilson et al., 2015), with beneficial right hemisphere compensation in recovery (Piai et al., 2017a, 2017b; Turkeltaub et al., 2011). Additionally, many authors agree on the fact that more coherent methodology across studies would yield better comparable findings and help to obtain further insights to the functional organization of language in the brain (Bradshaw et al., 2017; Duffau, 2005; Wilson et al., 2017). If one paradigm was revealed to take the lead over others in assessing the processes of language recovery in patients, it could be widely employed as a standard clinical tool. Accumulating all findings obtained with the same tool would commonly contribute to gain further insights into neurorehabilitation. Finally, this would help to improve the predictions of language recovery as well as patient care, enriching it with more individualized therapy.

The Present Study

To develop such a clinical tool requires a spatially defined and reliable effect that can be captured within a short testing duration. As a first step in this direction, the present study was conducted with healthy control participants to determine the optimal parameters for this tool. More precisely, this study approaches three main questions: What is the most suitable imaging method for the clinical tool, how spatially reliable is the brain activity of our paradigm over time, and how does the effect size change for half the testing duration?

In order to reveal the appropriate imaging method for patient testing, participants were scanned with MEG or fMRI. This serves as a more direct comparison of the effect profiles obtained per method, while participants perform the same experimental task. As outlined above, the common standard method to localize brain functions in tumor patients so far has been fMRI, but not necessarily because it is methodologically more suitable.

In hemodynamic methods such as fMRI, the obtained signal is a blood oxygen level dependent (BOLD) response. This is based on the framework that neurons that are active consume oxygen from the blood, causing a subsequent increase in blood flow called a hemodynamic response function (HRF). This HRF leads to a higher blood oxygen level in the local vessels which increases the signal intensity for fMRI and yields the BOLD signal that is measured. As such, fMRI measures neuronal activity only indirectly with a rather slow temporal resolution depending on the HRF, which can have a delay of about two seconds and last between 6 to 12 seconds before the signal decays.

In contrast, electrophysiological methods, such as MEG, yield a direct measure of brain activation based on the magnetic fields of neuronal activity. Here, the requirement is that neurons are activated in synchrony which initiates an electric current in the brain and induces a magnetic field around it. Contrary to fMRI, MEG captures neuronal activity with a high temporal resolution at the level of milliseconds and thus allows to track the time course of neuronal sources. But as MEG is only measured close to the scalp, the spatial resolution of fMRI is more precise, especially for subcortical structures. Nevertheless, MEG is successfully being employed to determine the dominant hemisphere for language, as it is done in patients before undergoing brain surgery (Findlay et al., 2012).

This shows that both methods are certainly possible to employ for a clinical tool, even though they measure different aspects of neuronal activation and have been shown to provide divergent findings (Kujala et al., 2014; Liljestrom, Hulten, Parkkonen & Salmelin, 2009; Vartiainen, Liljeström, Koskinen, Renvall & Salmelin, 2011). Additionally, keeping in mind that the patient study involves surgery in between testing sessions, it is noteworthy that prior brain surgery can affect and impair the BOLD signal of the whole hemisphere for fMRI (Kim et al., 2005). Hence, the present study questions fMRI as the clinical standard to localize brain functions by comparing it to MEG for the same task. Employing a more suitable imaging method for these purposes could possibly improve patient care and rehabilitation therapy in the future.

To track the brain activity of patients before and after the surgery as well as the subsequent language recovery, the patient study will be conducted longitudinally. Accordingly, we would expect to see changes from pre- to post-surgery and subsequent sessions. But to be able to argue that the changes in patients derive from the surgery, we need to know how spatially reliable and consistent the captured effect really is. Since we do not expect any changes between sessions for healthy control subjects, the present study evaluates the across-session consistency of our paradigm. This is done by visualizing the areas of overlapping brain activity from session 1 and session 2. To quantify this overlap with a measure of overlapping brain activation across sessions, the Dice coefficient is calculated (Wilson et al., 2017).

Further, we evaluate the effect sizes obtained for half the testing duration by only analyzing the first half of each session. Especially for the patient study, the duration of testing sessions should be kept as short as possible. But since half a session equals shorter testing times, it also equals less acquired data. In other words, we want to shorten testing times without risking not capturing enough data for or sacrificing an effect. Therefore, we aim to delineate an amount of data that is necessary to obtain reliable and robust effects to serve the patient study design.

Regarding a spatially defined and reliable effect to employ for the clinical tool, we decided to use the same paradigm as in previous studies by Piai et al (2014, 2015, 2017a, 2017b). As outlined above, this context-dependent sentence completion paradigm has repeatedly elicited robust alpha-beta power decreases in control participants and patients. Herein, sentences are presented word-by-word on a screen in front of the participants. The last word of each sentence is the target word and presented as a picture. The task is to silently read the sentence and name the following picture. The sentences appear in two conditions such that the sentence context is either constrained or unconstrained towards the target word. This means that the sentence context either reveals information about the target word or not. To give an example, the picture for the target word *cow* was a photograph of a cow on white background. The corresponding constrained sentence was "*The farmer milked the [picture]* ", and the unconstrained sentence was "*The child drew a [picture]"*. Thus, the sentence context in constrained sentences enables participants to retrieve information about the target word and accordingly prepare to name the picture, already before it appears. Unconstrained sentences, however, do not give away information about the target word and participants must wait for the picture to appear until they can retrieve the information needed to name it.

This difference between the two conditions determines the time window of interest for the present study, which is the interval between the last word preceding the picture and the onset of the picture. All analyses focus on the differences in brain activity during this time window between the two conditions. More precisely, trials of different conditions only vary in sentence context and this variation also yields the effect of interest. Thereby, the present paradigm offers a precise contrast between conditions and prevents the capturing of condition-specific differences that are of no interest. Further, participants only perform one task which eliminates the risk of capturing possible task switching demands. As such, the present paradigm proofs to be highly suitable to be employed in a clinical tool.

In line with earlier studies by Piai et al (2014, 2015, 2017a, 2017b) from which the paradigm as well as the stimulus materials were adopted, we hypothesize for the present study faster naming times for the pictures in constrained compared to unconstrained sentences. Further, the MEG brain activity profiles are expected to correspond to the spectral power decrease in the alpha-beta frequency range observed for the same paradigm as outlined above. That is, a decrease of alpha-beta power in the constrained relative to the unconstrained condition. Regarding fMRI, there has been evidence for a correlation between alpha-beta power decreases and BOLD signal increases for covert picture naming (Conner, Ellmore, Pieters, DiSano & Tandon, 2011). Additionally, in another picture naming study comparing MEG to fMRI that correlation between both measures was highest for frequencies in the alpha-beta range (Liljeström, Stevenson, Kujala & Salmelin, 2015). Based on these findings, we expect the BOLD signal to increase for constrained over unconstrained sentences.

Method

The present study falls under the blanket approval for standard studies of the accredited ethical reviewing committee, CMO Arnhem-Nijmegen, following the declaration of Helsinki (2013). It was conducted at the Donders Institute for Cognitive Neuroimaging in Nijmegen in the Netherlands.

Participants

A total of 31 native Dutch speakers aged between 18 and 50 years (*Mdn* = 22) participated in the study for monetary compensation or course credits. All participants were healthy and right-handed, with normal or corrected-to-normal vision (no glasses), and compatible for MEG and MRI (MEG participants), or MRI only (fMRI participants). The dataset of one female fMRI participant was not considered for analysis due to a large amount of invalid trials and missing field maps for session 1. Thus, an additional participant was recruited so that both groups consisted of 15 participants. The MEG group included 7 females and ranged from 18 to 50 years (*Mdn* = 25), and the fMRI group included 11 females and ranged from 18 to 26 years (*Mdn* = 20).

Materials

The stimuli consisted of 224 target words with a corresponding picture. This was a photograph depicting the target word on white background, if possible. Target words describing landscapes such as *forest*, or *mountain* were shown as full-screen photographs. Each target word was the last word of one constrained and one unconstrained sentence. As such, each target word had one corresponding sentence per condition, yielding 448 experimental sentences. All linguistic material was in Dutch and taken from previous studies (Piai et al., 2014, 2015, 2017a, 2017b). Pictures were collected from the BOSS database (Brodeur, Dionne-Dostie, Montreuil & Lepage, 2010) and via online search. The length of the target words varied from 2 to 11 phonemes (mean length = 5). Sentence length varied between 4 and 13 words including the target word (mean length = 7) and was kept as similar as possible for both sentences associated with the same target word.

Design

The stimuli were presented in three main lists, uniquely divided in halves controlled for frequency, word length, and initial letter. Each half was pseudorandomized using Mix (van Casteren & Davis, 2006) so that there were at least 20 trials between the first and the second appearance of the same target word, and a maximum of five repetitions of trials with the same condition. Participants were randomly assigned to one of the three main lists. Since the study consisted of a test and a re-test session scheduled between 13 and 28 days apart (*Mdn* = 20), each participant was presented with half of the target words per session, alternating the order of which half was presented first. Thus, one session consisted of 112 target words presented as pictures, once preceded by a constrained sentence and once by an unconstrained sentence, yielding 224 trials per session.

Procedure

Each session started with informing the participants about the task and the scanning session and clarifying possible questions. Then, participants signed the consent forms and were screened according to the employed scanning method. Before entering the scanner room, they were familiarized with the pictures of the experiment and the corresponding target words. These were presented in a slide show with four pictures on one slide and the target words printed below. Each session started with four practice trials in the scanner, so that participants knew what to expect and had the chance to clarify remaining doubts before the start of the experiment. Stimuli were presented with Presentation software (Neurobehavioral Systems, Inc.) and projected on a screen in front of the participants in the scanner. Figure 1 shows a trial overview for both sentences for the same target word with the experiment-specific presentation times highlighting the time window of interest. Each trial started with a 500 ms fixation cross and consisted of a word-by-word presentation of the sentence in the center of the screen. Each word was presented for 300 ms followed by a 200 ms blank screen. Words were presented in black on a grey-scaled background. The last word of each sentence was the target word. This was presented as a picture on screen for 1000 ms. The task was to silently read the sentences attentively and name the pictures with the words that participants were familiarized with. Also, participants were asked to keep fixation to the center of the screen and move their jaw and head as little as possible.

In the following sections, the procedure for the behavioral data with participants' reaction times for naming the pictures and error coding is reported first. Then, acquisition, preprocessing, and analysis steps of each method are reported separately, first for the MEG experiment and then for the fMRI experiment. Last, the procedure for calculating the Dice coefficients for both methods is stated.

Behavioral Analysis.

In both scanners, trials were recorded to monitor participants' responses for picture naming. Recordings started simultaneously with picture onset

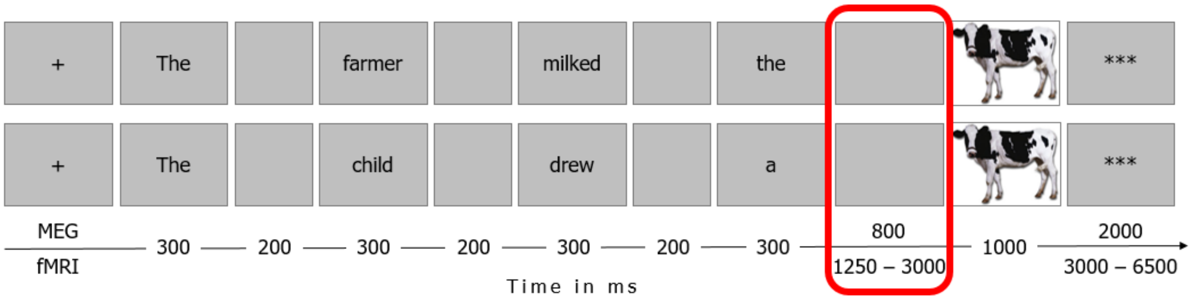


Fig. 1. Overview of the constrained (top) and unconstrained (bottom) trial for the target word cow. Boxes represent the screens that participants saw in the scanner, with experiment-specific presentation times below and the analyzed time window of interest circled in red

and lasted 2500 ms. Trials in which participants hesitated, stuttered, responded either with more than one word or later than 2500 ms after picture onset were considered as errors and not included in the analyses. Trials in which the response was a synonym to the original target word that makes sense in the sentence context of the corresponding trial were marked as correct. If participants' speech onset started prior to picture presentation and recording onset no reaction time could be measured and the trial was discarded from this analysis. Reaction times were calculated using the speech editor Praat (Boersma & Weenink, 2017), blind for condition, and statistically analyzed in R (Team, 2017). The mean reaction time per condition was calculated for each participant and the behavioral effect was evaluated by means of an analysis of variance with condition and session as within-participant variables at an alpha-level of 0.05.

MEG Experiment.

MEG Acquisition.

Participants taking part in the MEG study had to change into the non-magnetic scanner clothing provided in the MEG laboratory. Then they were prepared with electrodes attached to their face and body to measure the vertical and horizontal electrooculogram, the electromyogram, and the electrocardiogram. Electrode impedance was kept below 20 kOhm. Before and during the experiment participants were instructed to restrict blinking to the inserted blinking intervals at the end of each trial showing three asterisks (***). MEG data were acquired with a 270 axial gradiometer system (CTF Systems Inc., VSM MedTech Ltd.) at a sampling rate of 1200 Hz. Participants were positioned in the MEG chair with pillows as they preferred. Localization coils were attached to the left and right ear canal, and the nasion. Head localization was performed in real-time (Stolk, Todorovic, Schoffelen & Oostenveld, 2013) and the head position relative to the sensors at the start of session 1 was stored. This was used at the start of session 2 to place participants in the same position as in session 1. Then this position was updated to the real position at the start of session 2, to keep an overview of participants' head movement within the session. The head position was kept as constant as possible across trials and sessions to minimize noise deriving from head position variance. If participants moved more than 8 mm away from their initial position, they were relocalized in the breaks after every block of 28 trials. The scanning for one MEG session lasted approximately 30 minutes and participants were in the laboratory for one hour, including preparation time.

If not yet available, structural T1-weighted MRI scans of participants' heads were acquired either after one of the two MEG sessions or on a third day.

MEG Preprocessing.

MEG data preprocessing was performed in Matlab using the FieldTrip toolbox (Oostenveld, Fries, Maris & Schoffelen, 2011). The data were demeaned to take out drifts and each trial was cut down to the time window of interest of 800 ms before picture onset. Incorrect trials were not considered, which led to a loss of 0 to 37 trials per session (*M* = 6, *SD* = 7). Subsequently, the data were down-sampled to 600 Hz and blinking trials were discarded by means of the vertical electrooculogram channel. This led to a loss of 0 to 26 trials per session (*M* = 5, *SD* = 5). Finally, remaining noisy trials and sensors were marked by means of a trial and sensor overview summary and not considered for further analyses. During preprocessing, 8 to 20 sensors (*M* = 15, *SD* = 3) were removed per session and each session consisted of 170 to 213 trials for analysis (*M* = 198, *SD* = 10), including 82 to 110 (*M* = 99, *SD* = 6) unconstrained trials and 81 to 106 (*M* = 98, *SD* = 6) constrained trials.

MEG Analysis.

The MEG analysis is based on the differences in brain activity between the constrained and the unconstrained condition in the specified time window of interest (see Figure 1). This is the 800 ms interval between the last presented word and the onset of the picture in each trial, during which the screen was blank.

Source Localization.

To identify the sources of the obtained brain activity, the MRI scans of participants' heads were realigned with the coordinates of the MEG data by marking the fiducials in both ear canals and the nasion. Then they were resliced and segmented into brain, scalp, and skull using SPM8, to obtain individual volume conduction models of participants' heads. Next, the realigned MRI scans were warped to the Montreal Neurological Institute (MNI) space template to obtain subject-specific source model grids in normalized space, so they can be compared across participants. Using the volume conduction models, lead field matrices were computed for each grid point per participant. Then, the beamforming technique was applied to estimate the activity at the source-level. The cross-spectral density matrix of the sensor-level data for both conditions combined was computed at 15 Hz. Spectral smoothing of 5 Hz yielded a cross-spectral density matrix between 10 and 20 Hz. This frequency range is based on previous findings resulting from the same task and analysis technique (Piai et al., 2015). As the transition from alpha to beta activity is usually considered around 12 to 15 Hz, this frequency range is referred to as alpha-beta power for the present report. Together with the lead field matrices, the cross-spectral density matrices were used to calculate a common spatial filter for each grid point. These filters were then applied to the Fourier transformed sensor data per condition to estimate source-level power for each grid point. Then, the power estimates for constrained and unconstrained trials were averaged for each participant. The power change was calculated as the difference between power in the constrained and the unconstrained condition divided by the common average. The effect of the power differences from the constrained to the unconstrained condition over all participants was evaluated by means of a non-parametric cluster-based permutation test. A dependent-samples t-test thresholded at an alpha-level of 0.05 served to identify the biggest cluster of neighboring voxels for the effect on the group-level. The p-values of this cluster were calculated as the amount out of 5000 random permutations that yield a larger effect than the observed one, employing a Monte Carlo method with 5000 random permutations.

MEG First Half Effect Size.

To estimate the effect size of our paradigm for an MEG experiment of shorter duration, the first 112 trials (out of 224 trials in total) were selected from each session. This served as a representation of half a session and was analyzed in the same way as the full sessions specified above. Importantly, the trial selection for half a session was performed only after the preprocessing step, meaning that incorrect, noisy, and blinking trials were discarded previously. Thus, every representation of half a session consisted of 112 clean and correct trials for all participants, with an average of 55 (*SD* = 4) unconstrained trials and 57 (*SD* = 4) constrained trials.

fMRI Experiment.

fMRI Acquisition.

Participants taking part in the fMRI study had to wear metal-free clothing on their upper body and change into scanner clothing if necessary. Then they were taken into the scanner room and positioned on the scanner bed with cushions underneath their knees and elbows. Their forehead was taped to the lower part of the head coil with crepe tape to minimize their head movement, and an emergency button was placed on their belly. All functional scans were acquired on a 3T Siemens PrismaFit scanner with a 32-channel head coil using echo-planar imaging (EPI) to minimize movement artefacts, employing a multiband sequence (multiband acceleration factor 6, 2 mm isotropic voxels, 66 slices, TR = 1000 ms, TE = 34 ms, FoV = 210 x 210 x 132 mm, flip angle = 60°). Localizer and head scout scans were performed before the start of the experiment to obtain the location of participants' brains. The experiment started with a pulse countdown from nine to zero followed by all 224 trials consecutively. Field maps were acquired at the end of the trials (TR = 620 ms, TE 1 = 4.7 ms, 64 slices, voxel size 2.4 x 2.4 x 2 mm, FoV = 210 x 210 x 132 mm, flip angle

= 60°) to calculate voxel displacement maps (VDM) for each session. Structural T1-weighted MPRAGE images (TR = 2300 ms, TE = 3.03 ms, 192 slices, FoV = 256 x 256 x 192 mm, 1 mm isotropic voxels) for anatomical reference were acquired after session 1. The fMRI experiment had a jittered design to capture the BOLD response at different stages. The interval before picture onset was randomly jittered between 1250 ms and 3000 ms, and the fixation cross between trials was randomly jittered between 3000 ms and 6500 ms. This prolonged the fMRI sessions such that the experimental scanning took approximately 45 minutes.

fMRI Preprocessing.

fMRI preprocessing was performed session-individually using Matlab and SPM12. The first nine volumes of each session were discarded as dummy scans to allow the magnetization to reach a steady state. All other images were realigned with reference to the 10th volume and unwarped by applying the calculated session-specific VDM to reduce movement artefacts of the functional EPI scans. After estimation of coregistration the images underwent segmentation into different tissue types based on probability maps such as grey matter, white matter, cerebrospinal fluid, bone, and soft tissue. Then the images were normalized to MNI space and resliced, and a smoothing Gaussian kernel of twice the voxel size (FWHM = 4 mm) was applied.

fMRI Analysis.

As for the MEG experiment, the fMRI analysis focuses on the difference in brain activity between both conditions in the time window of interest before picture presentation (see Figure 1). This interval from the last word of the sentence to the picture was jittered between 1250 ms and 3000 ms.

General Linear Model.

The fMRI analysis was performed by means of a general linear model (GLM) for each participant. This included five regressors per session containing the onsets of task-specific events as well as the six motion parameters outputted from the session-specific preprocessing to account for further movement artefacts of participants in the scanner. A high-pass filter removed slow signal drifts with periods longer than 128 seconds. The task-specific regressors contained the time-locked onsets of each trial for the first word, the pre-picture interval per condition, and the picture. The fifth regressor modelled all incorrect trials per session (*M* = 3, *SD* = 4) by means of the onsets of the pre-picture intervals of incorrect trials. Not considering errors, each session consisted of an average of 110 trials per condition (*SD* = 2). The onsets were modelled as delta (stick) functions (duration = 0) and convolved with the canonical HRF, i.e., the BOLD response. The design matrix for each participant consisted of the GLM for the first and second session. The contrast of interest was the increase of the BOLD response for constrained over unconstrained trials. Therefore, the BOLD signal of the whole brain at the onset of the pre-picture interval for unconstrained trials was subtracted from that of constrained trials for each participant. These individual t-contrasts

Table 1.
Amount of total errors and errors per condition for each session. Mean reaction time effect per session and average reaction times per condition.

Experiment	MEG		fMRI	
Session	1	2	1	2
Errors in %				
Total	2.26	2.74	1.49	1.55
unconstrained	2.08	3.33	1.85	1.79
constrained	2.44	2.14	1.13	1.31
Reaction Times in ms				
Effect	143 (*SD* = 83)	177 (*SD* = 72)	201 (*SD* = 59)	203 (*SD* = 64)
unconstrained	738 (*SD* = 92)	759 (*SD* = 84)	896 (*SD* = 225)	926 (*SD* = 230)
constrained	595 (*SD* = 89)	583 (*SD* = 129)	694 (*SD* = 269)	723 (*SD* = 269)

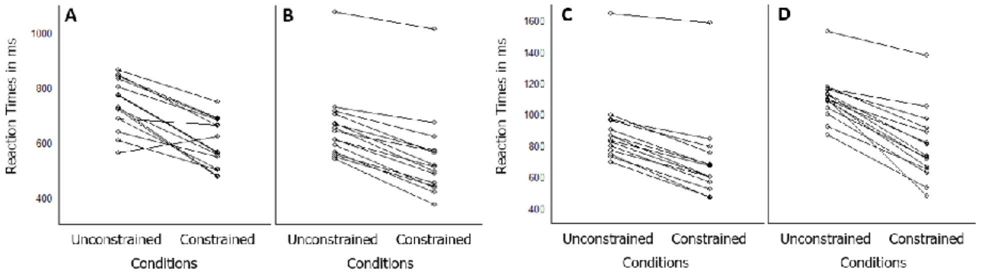


Fig. 2. Mean reaction times per condition in session 1 (A) and session 2 (B) of the MEG experiment, and session 1 (C) and session 2 (D) of the fMRI experiment. Each line connects both conditions for one participant

then entered the second-level analysis to obtain t-contrasts of the group-level BOLD increases in the whole brain per session.

fMRI First Half Effect Size.

To analyze the first half of each fMRI session, a cutoff point was determined by means of the picture onset of trial 112 (out of 224) of each session. To capture the full BOLD response for this trial, 20 additional volumes were included. Then, the motion parameters from the preprocessing step were modified to match the length of each individual half session. A new GLM was constructed per participant with all onsets for the first 112 trials and the same regressors as specified above. This included an average of 54 unconstrained (*SD* = 4) and 57 constrained (*SD* = 4) trials as well as errors (*M* = 1, *SD* = 2). The individual and group-level t-contrasts were also constructed in the same manner as described above.

Dice Coefficients.

To quantify the extent of overlap of the measured brain activity between session 1 and session 2 and compare it for both experiments, the corresponding Dice coefficients were computed. This measure is calculated by twice the number of overlapping voxels (*overlap*) divided by the sum of activated voxels in session 1 (*ses1*) and session 2 (*ses2*). Here, an outcome of 0 indicates no overlap and an outcome of 1 indicates a perfect overlap of activation between both sessions:

For both methods, this calculation was based on the significant voxels resulting from the analyses. For MEG the statistical voxel threshold was set to 0.05. For fMRI the p-threshold on voxel-level was 0.001 uncorrected, only including Family-wise error (FWE) corrected clusters with $p < 0.05$ (as reported in Table 3).

Results

In this section, the behavioral results for reaction time measurements are reported first, followed by the results for the MEG and the fMRI experiment respectively. Finally, the Dice coefficients as the chosen measure of overlap are reported for both experiments in comparison.

Figure 2 A and B shows the mean reaction time effects for picture naming of each participant in the MEG experiment per session. This effect is the difference between the mean reaction times for unconstrained and constrained sentences. The mean reaction time for unconstrained sentences was 738 ms (*SD* = 92) in session 1 and 759 ms (*SD* = 84) in session 2. For constrained sentences this was 595 ms (*SD* = 89) in session 1 and 583 ms (*SD* = 129) in session 2. Thus, participants named the pictures in constrained sentences faster than in unconstrained sentences. More exact, the mean effect of sentence constraint was 143 ms (*SD* = 83) in session 1, and 177 ms (*SD* = 72) in session 2. Overall, this yields a main effect of condition, $F(1, 14) = 81.15$, $p < 0.001$, no significant effect of session, $F(1, 14) = 0.02$, $p = 0.9$, and no interaction effect, $F(1, 14) = 3.25$, $p = 0.1$.

Figure 2 C and D shows the mean effects for picture naming of all fMRI participants per session. Here, the mean reaction time for unconstrained sentences was 896 ms (*SD* = 225) in session 1 and 926 ms (*SD* = 230) in session 2, and for constrained sentences 694 ms (*SD* = 269) in session 1 and 723 ms (*SD* = 269) in session 2. As in the MEG experiment, participants reacted faster to constrained than unconstrained sentences, with a mean effect of 201 ms (*SD* = 59) in session 1 and 203 ms (*SD* = 64) in session 2. This also yields a main effect of condition, $F(1, 14) = 223.11$, $p < 0.001$, no significant effect of session, $F(1, 14) = 2.32$, $p = 0.15$, and no interaction effect $F(1, 14) = 0.01$, $p = 0.93$.

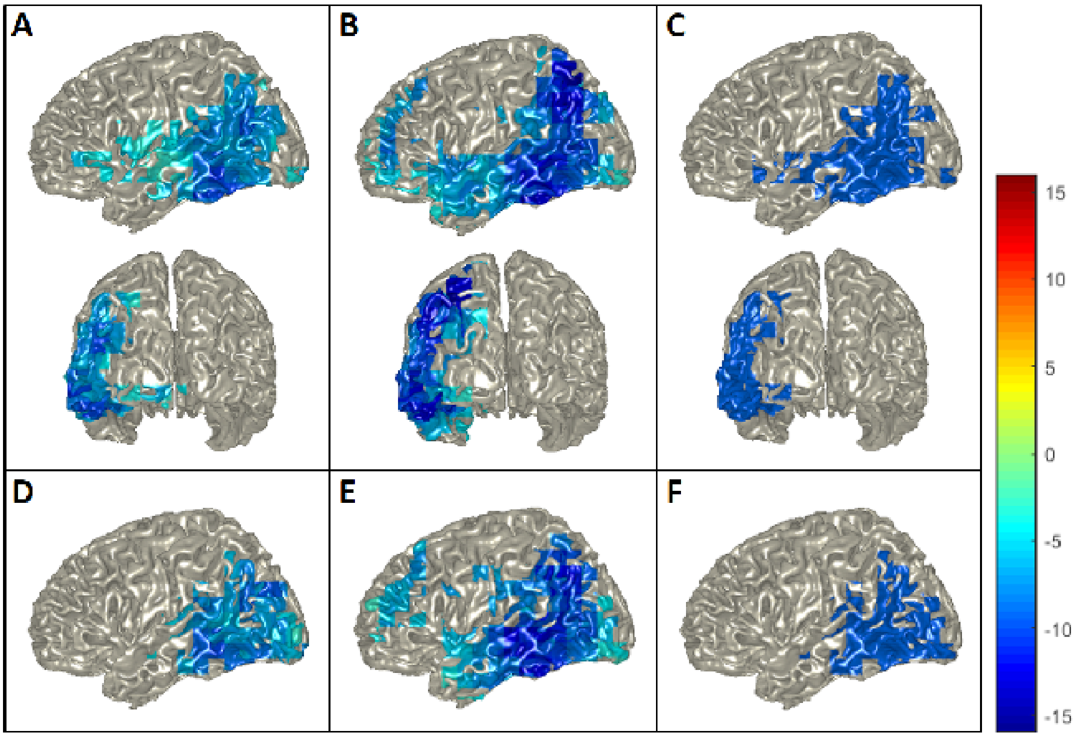


Fig. 3. Group-level source localization power decreases for constrained relative to unconstrained condition at 10 - 20 Hz for full session 1 (A) and full session 2 (B) in left and posterior view. Power decreases for first half of session 1 (D), and first half of session 2 (E) in the left hemisphere. Depicted areas are masked by statistically significant clusters, color bars show relative power change in percentage. Overlap map of group-level source localization showing areas of power changes common to both sessions for the full sessions (C) and the first half of both sessions (F)

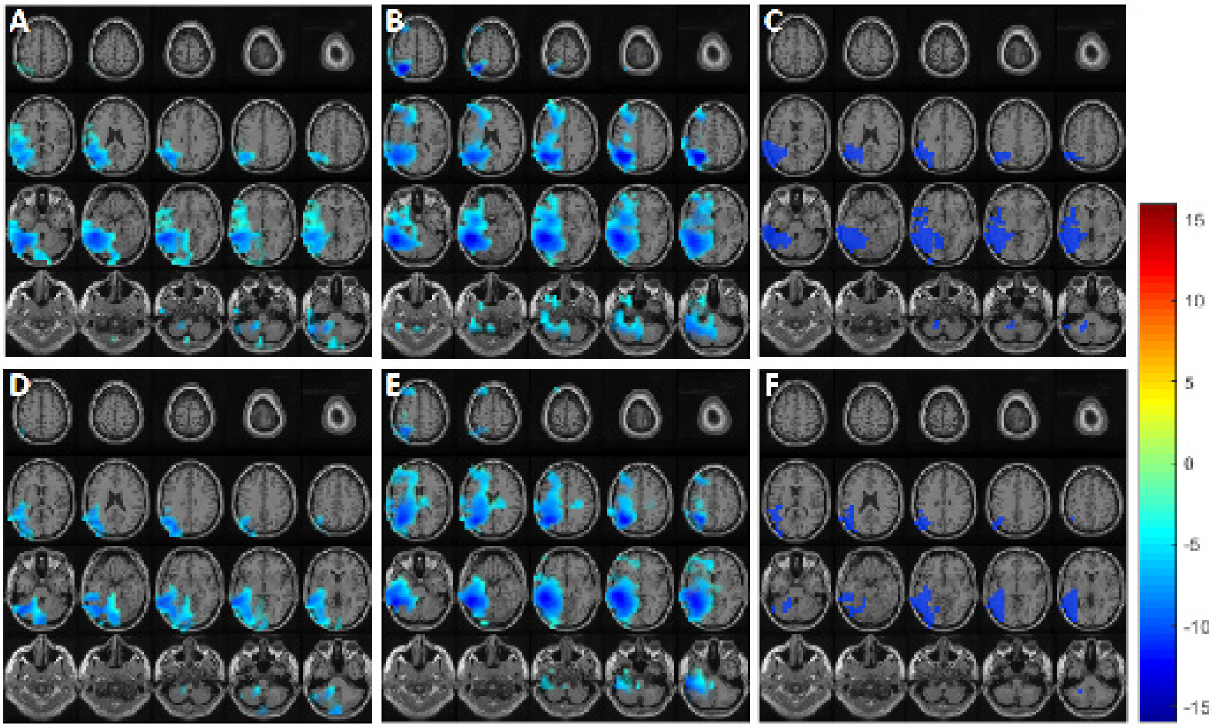


Fig. 4. Group-level source localization power decreases between conditions at 10 - 20 Hz for full session 1 (A), full session 2 (B), first half of session 1 (D), and first half of session 2 (E) in axial brain slices. Depicted areas are masked by statistically significant clusters, color bars show relative power change in percentage. Overlap map of group-level source localization showing areas of power changes common to both sessions for the full sessions (C) and the first half of both sessions (F)

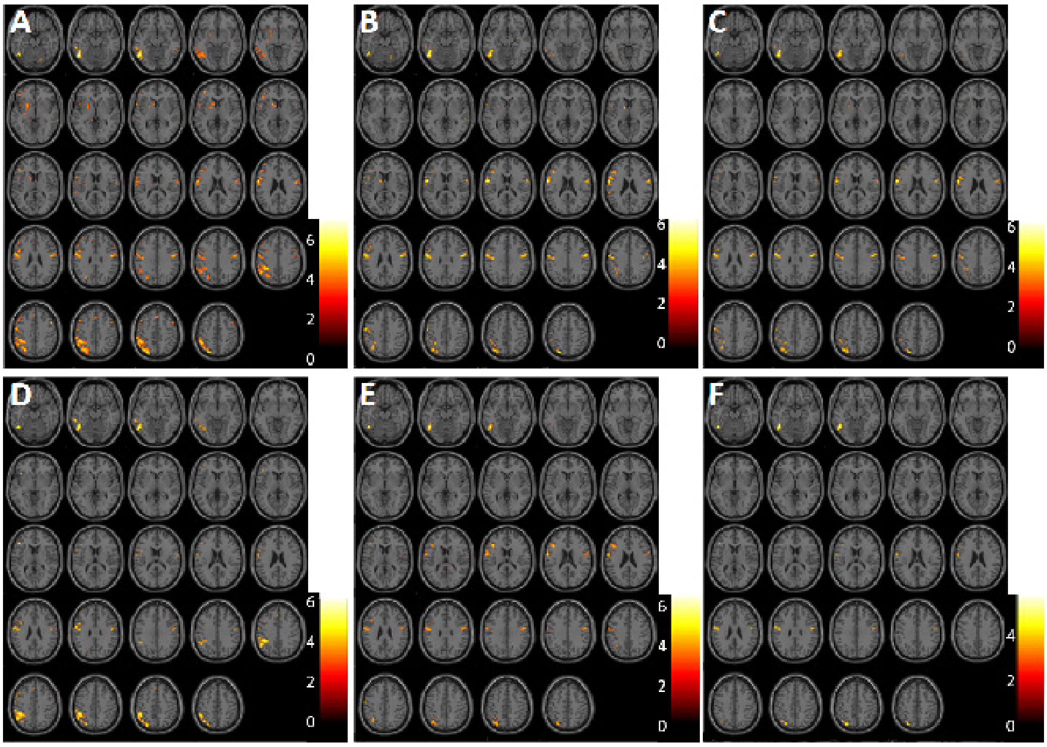


Fig. 5. Clusters of group-level BOLD increases for constrained over unconstrained trials for full session 1 (A), full session 2 (B), first half of session 1 (D), and first half of session 2 (E) in axial slices (MNI space -18 to 54 mm, steps of 3 mm). Conjunction showing areas of group-level BOLD increases common to both sessions for the full sessions (C) and the first half of both sessions (F). Color bars show t-values, $p < 0.001$ on voxel-level

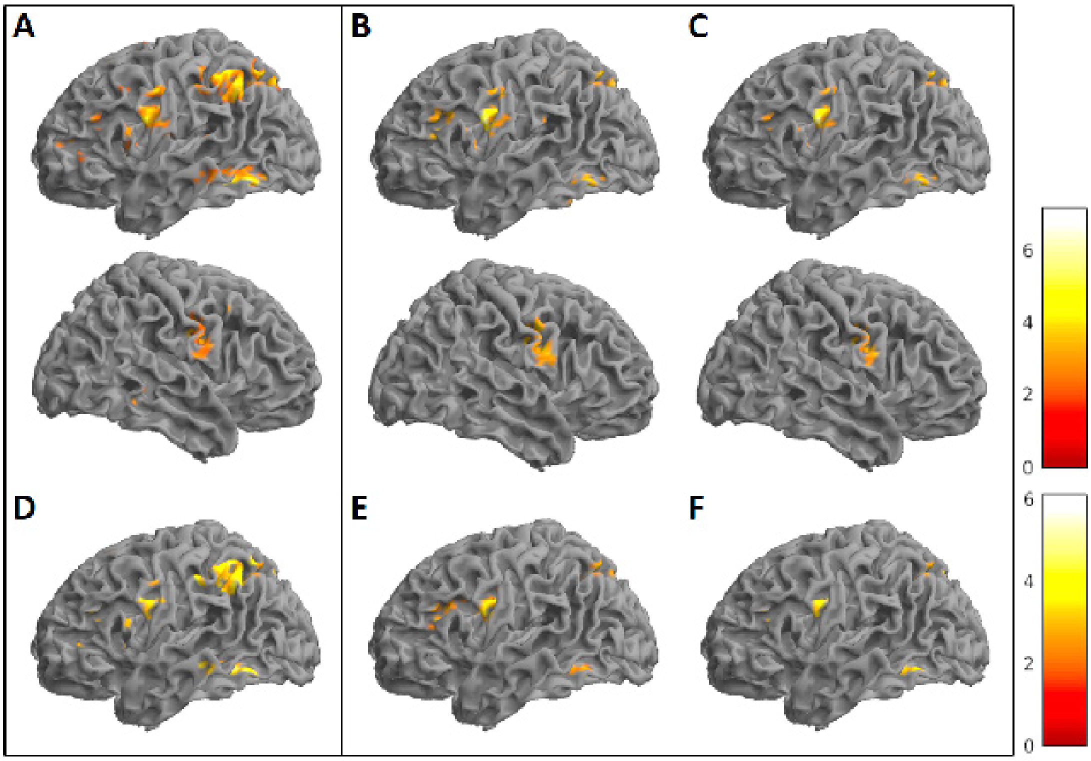


Fig. 6. Group-level BOLD increases for full session 1 (A) and full session 2 (B) projected to the surface of the left (top) and right (bottom) hemisphere. BOLD increases in the first half of session 1 (D), and first half of session 2 (E) projected to the surface of the left hemisphere. Conjunction showing areas of group-level BOLD increases common to both sessions for the full sessions (C) and the first half of both sessions (F). Color bars show t-values, $p < 0.001$ on voxel-level

MEG Results

The top row of Figure 3 shows the group-level results for the source localization of the power differences for both sessions on the brain surface as well as their overlap, masked by the statistically significant clusters.

The top row of Figure 4 depicts the obtained clusters and their spatial distribution inside the brain in axial slices. However, caution should be taken when interpreting these slices, as the spatial resolution as well as sensitivity of MEG for subcortical sources is reduced (Hillebrand et al., 2016). In both figures, color scales show the percentage of power change in session 1 and session 2.

The power decreases of constrained relative to unconstrained sentences in the alpha-beta frequency range from 10 to 20 Hz were statistically significant for both session 1 (p = 0.0092) and session 2 (p = 0.0052). The significant clusters of power changes were exclusively lateralized to the left hemisphere in both sessions. In session 1, the strongest power decreases around 10% were observed in the posterior temporal lobe, as shown in Figure 3 A. More precisely, they extended from the inferior to the superior temporal gyrus, and then posteriorly to the inferior parietal lobe. Anteriorly a weaker decrease around 5% extended until the inferior frontal gyrus.

In session 2, the strongest power decreases were obtained around 15% and extended over the posterior temporal lobe, and further over the angular and supramarginal gyrus up to the superior parietal gyrus, as shown in Figure 3 B. Anteriorly a weaker decrease of 5 - 10% extended over the anterior temporal lobe and the orbital inferior frontal gyrus to the middle frontal gyrus.

MEG Across-Session Consistency.

To depict the overlap of brain areas exhibiting power changes in both sessions, a new mask was created. This was based on the significant clusters resulting from the source localization for both sessions and only included voxels that were active in both session 1 and session 2. This overlap is shown in Figure 3 C. In line with the strongest power decreases in session 1 and 2, the area of overlap extends over the posterior temporal lobe to the inferior parietal gyrus, and anteriorly over the superior temporal gyrus slightly into the inferior frontal gyrus.

MEG First Half Effect Size.

The bottom rows of Figure 3 and Figure 4 show the results for the source localization of the power changes for the first half of both sessions. Here, the decreases in power between conditions were also significant for both session 1 (p = 0.0036) and session 2 (p = 0.0024). Compared to the full session, the power decreases of the first half of session 1 are restricted to the posterior temporal lobe and inferior parietal gyrus, not extending to the inferior frontal gyrus, as shown in Figure 3 D. The power decreases in the first half of session 2 are quite consistent with the full session. They extend from the posterior temporal lobe to the inferior parietal gyrus, and anteriorly over the anterior temporal lobe to the middle frontal gyrus, as shown in Figure 3 E.

Likewise, the overlap of the source localization results for the first half of session 1 and the first half of session 2 is depicted by means of the same masking procedure as described above and shown in Figure 3 F. This area of overlap extends over the posterior temporal lobe to the inferior parietal gyrus, presenting a similar overlap as for the full sessions, but less anterior and continuous.

fMRI Results

The upper part of Table 3 lists significant clusters of neighboring voxels showing a BOLD increase for the contrast of interest per session, meaning areas that show a stronger BOLD response for constrained than unconstrained trials. The corresponding brain regions are based on the MNI space voxel coordinates obtained with SPM12 and determined by using the Automated Anatomical Labeling (AAL) atlas (Tzourio-Mazoyer et al., 2002). The top row of Figure 5 shows the whole-brain BOLD increases in session 1 and session 2 as well as their overlap by means of the same selected axial slices. The top row of Figure 6 depicts these clusters of BOLD increases in session 1 and session 2 and their overlap projected to the brain surface for both hemispheres.

The BOLD increases for session 1 are shown in Figure 5 A and Figure 6 A. Here, left-hemisphere BOLD increases were detected in a cluster in the inferior occipital lobe and inferior temporal gyrus, a large cluster in the superior and inferior parietal gyrus, a cluster in the post- and precentral gyrus, followed by clusters in the superior frontal gyrus, the middle and triangular inferior frontal gyrus, the pallidum, the caudate, and the supplementary motor area and medial superior frontal gyrus. Clusters of BOLD increase in the right hemisphere were detected in the precentral and middle frontal gyrus, the cerebellum (crus 2), and the postcentral gyrus.

BOLD increases for session 2 are shown in Figure 5 B and in Figure 6 B. Left-hemisphere clusters showing BOLD increases were located in the fusiform and inferior temporal gyrus, the Rolandic operculum and postcentral gyrus, the superior and inferior parietal and angular gyrus, the inferior parietal gyrus, and the triangular inferior frontal gyrus. For session 2 only one cluster in the postcentral gyrus was significant in the right hemisphere.

fMRI Across-Session Consistency.

To look at the overlap of voxel clusters exhibiting an increase in the BOLD response for constrained over unconstrained trials for both sessions, the contrasts of interest from session 1 and 2 were entered in a conjunction analysis in SPM12. The significant clusters of overlap from both sessions are listed under *Full Conjunction* in Table 3 and shown for the same slices as depicted for the individual sessions in Figure 5 C. Clusters of overlapping BOLD increase in the left hemisphere were located in the fusiform and inferior temporal gyrus, the postcentral gyrus, the superior parietal and angular gyrus, and the inferior parietal gyrus. The overlap of significant BOLD increases in the right hemisphere was limited to the postcentral gyrus.

fMRI First Half Effect Size.

The lower part of Table 3 lists all brain areas with significant clusters of BOLD increases for the contrast of interest per session and their overlap, when analyzing only the first half of each session. These are shown in the bottom row of Figure 5 and Figure 6 for session 1, session 2, and their overlap.

The first half of session 1 only yielded significant clusters in the left hemisphere, shown in Figure 5 D and Figure 6 D. A large cluster was detected in the inferior parietal gyrus, followed by clusters in the fusiform gyrus and inferior occipital lobe, the inferior and middle temporal gyrus, and the precentral gyrus.

As shown in Figure 5 E and Figure 6 E, in the first half of session 2 the significant left-hemisphere clusters were located in the fusiform and inferior temporal gyrus, the post- and precentral gyrus, the inferior and superior parietal gyrus, and the triangular inferior frontal gyrus. In the right hemisphere there was one significant cluster in the postcentral gyrus.

As for the full session data, a conjunction analysis of the contrasts of interest was conducted with SPM12 for the first half of both sessions to look at the overlapping voxel clusters that exhibit a BOLD increase for constrained over unconstrained trials.

Table 2.

Dice coefficient values for full and half session calculation per experiment and average amount of input trials per participant.

Experiment	MEG	fMRI
FULL SESSION		
Dice Coefficient	0.49	0.43
Trials	198 (*SD* = 10)	220 (*SD* = 4)
HALF SESSION		
Dice Coefficient	0.35	0.31
Trials	112 (*SD* = 0)	111 (*SD* = 2)

The locations of these overlaps are listed under *Half Conjunction* in the lower part of Table 3 and are shown in Figure 5 F and Figure 6 F. Overlapping clusters were located in the fusiform and inferior temporal gyrus, the superior and inferior parietal gyrus, and the pre- and postcentral gyrus, all in the left hemisphere.

Table 2 shows the Dice coefficients per experiment and the average amount of trials per participant that the respective calculation is based on. Dice coefficients were calculated as described above for the full and half sessions of both experiments, by means of the respective group-level results. Based on the partition thresholds of Cohen's *d* (Cohen, 1988) as a similar effect size measure, Dice coefficients are classified as *small* (0.20-0.49), *medium* (0.50-0.79), and *large* (0.80-1). The highest Dice coefficient was obtained for the full MEG experiment (0.49), followed by the full fMRI experiment (0.43). The first half analysis resulted in lower Dice coefficients for MEG (0.35) as well as fMRI (0.31).

Discussion

The present study aimed to determine the most suitable imaging method, the across-session consistency, and the effect size for short testing sessions for the present paradigm to develop a clinical tool to track the recovery of language functions in patients. Therefore, participants were tested with MEG and fMRI while performing the selected sentence completion task to be employed in the clinical tool. In both experiments, the behavioral differences in terms of reaction times per condition demonstrate faster word retrieval for constrained than unconstrained sentences, approving that the

Table 3.
Significant clusters of whole-brain BOLD increases for constrained over unconstrained trials. Cluster size given in number of voxels (2 mm isotropic) in the cluster. Coordinates given for maximally activated voxel and up to two local maxima more than 8 mm apart. Voxel-threshold at p = 0.001; cor: Family-wise error (FWE) corrected; unc: unocorrected; l: left; r: right, tr.: triangular; *cases where obtained coordinates deviate 1 mm from AAL region (-47).

Cluster		Voxel			MNI space	Brain structure (AAL)
p(cor)	size	t value	z value	p(unc)	x, y, z (mm)	
FULL SESSION 1						
0.000	531	7.12	5.73	0.000	-48, -58, -14	l inferior occipital lobe
		6.48	5.37	0.000	-46, -50, -16	l inferior temporal gyrus
		5.92	5.02	0.000	-58, -62, -12	l inferior temporal gyrus
0.000	1728	5.87	4.99	0.000	-28, -74, 50	l superior parietal gyrus
		5.71	4.89	0.000	-46, -48, 52	l inferior parietal gyrus
		5.57	4.79	0.000	-50, -54, 42	l inferior parietal gyrus
0.000	717	5.50	4.75	0.000	-58, -2, 22	l postcentral gyrus
		5.22	4.55	0.000	-56, -6, 42	l postcentral gyrus
		5.10	4.48	0.000	-46*, -10, 30	l precentral gyrus
0.015	79	5.47	4.72	0.000	52, 12, 44	r precentral gyrus
		4.46	4.01	0.000	40, 6, 60	r middle frontal gyrus
		3.81	3.51	0.000	46, 12, 52	r middle frontal gyrus
0.042	64	5.28	4.59	0.000	8, -78, -30	r cerebellum (crus2)
0.025	72	5.06	4.45	0.000	-16, 18, 62	l superior frontal gyrus
		3.73	3.44	0.000	-18, 8, 66	l superior frontal gyrus
0.000	321	5.02	4.42	0.000	50, -8, 32	r postcentral gyrus
		4.14	3.77	0.000	58, 2, 20	r postcentral gyrus
		4.01	3.67	0.000	64, -6, 18	r postcentral gyrus
0.000	154	5.01	4.41	0.000	-50, 14, 44	l middle frontal gyrus
		4.67	4.17	0.000	-48, 30, 26	l tr. inferior frontal gyrus
		3.69	3.42	0.000	-50, 22, 28	l tr. inferior frontal gyrus
0.001	134	4.95	4.37	0.000	-24, -4, -4	l pallidum
		4.84	4.29	0.000	-22, 10, -4	l pallidum
		4.81	4.27	0.000	-18, 2, 8	l pallidum
0.021	74	4.82	4.28	0.000	-14, 12, 10	l caudate nucleus
0.034	67	4.56	4.09	0.000	0, 24, 46	l supplementary motor area
		3.61	3.35	0.000	0, 36, 48	l medial superior frontal gyrus
FULL SESSION 2						
0.000	216	6.19	5.19	0.000	-48, -56, -18	l fusiform gyrus
		5.58	4.80	0.000	-46, -48, -16	l inferior temporal gyrus
0.000	611	5.97	5.05	0.000	-54, 0, 16	l Rolandic operculum
		4.93	4.35	0.000	-56, -6, 44	l postcentral gyrus
		4.81	4.27	0.000	-54, -8, 30	l postcentral gyrus
0.000	240	4.85	4.30	0.000	-28, -74, 52	l superior parietal gyrus
		4.68	4.18	0.000	-30, -48, 44	l inferior parietal gyrus

		4.52	4.06	0.000	-36, -64, 44	l angular gyrus
0.000	344	4.69	4.18	0.000	50, -6, 32	r postcentral gyrus
		4.36	3.94	0.000	58, 2, 22	r postcentral gyrus
		4.21	3.82	0.000	64, -2, 18	r postcentral gyrus
0.020	75	4.29	3.89	0.000	-50, -44, 56	l inferior parietal gyrus
		3.88	3.57	0.000	-48, -38, 50	l inferior parietal gyrus
		3.87	3.56	0.000	-38, -40, 42	l inferior parietal gyrus
0.001	135	4.23	3.84	0.000	-42, 22, 24	l tr. inferior frontal gyrus
		4.03	3.68	0.000	-40, 28, 18	l tr. inferior frontal gyrus
		3.84	3.54	0.000	-44, 34, 12	l tr. inferior frontal gyrus
FULL CONJUNCTION						
0.000	209	6.19	5.19	0.000	-48, -56, -18	l fusiform gyrus
		5.58	4.80	0.000	-46, -48, -16	l inferior temporal gyrus
0.000	475	5.50	4.75	0.000	-58, -2, 22	l postcentral gyrus
		4.86	4.30	0.000	-56, -6, 42	l postcentral gyrus
		4.60	4.12	0.000	-48, -10, 30	l postcentral gyrus
0.000	187	4.85	4.30	0.000	-28, -74, 52	l superior parietal gyrus
		4.44	4.00	0.000	-36, -64, 44	l angular gyrus
		3.81	3.51	0.000	-32, -68, 56	l superior parietal gyrus
0.000	244	4.69	4.18	0.000	50, -6, 32	r postcentral gyrus
		4.14	3.77	0.000	58, 2, 20	r postcentral gyrus
		3.83	3.53	0.000	62, -4, 26	r postcentral gyrus
0.020	75	4.29	3.89	0.000	-50, -44, 56	l inferior parietal gyrus
		3.88	3.57	0.000	-48, -38, 50	l inferior parietal gyrus
		3.87	3.56	0.000	-38, -40, 42	l inferior parietal gyrus
HALF SESSION 1						
0.000	1043	6.08	5.12	0.000	-42, -44, 40	l inferior parietal gyrus
		5.08	4.46	0.000	-50, -54, 42	l inferior parietal gyrus
		4.91	4.34	0.000	-54, -32, 44	l inferior parietal gyrus
0.000	238	5.70	4.88	0.000	-46, -56, -14	l fusiform gyrus
		5.29	4.61	0.000	-52, -62, -12	l inferior occipital lobe
		4.81	4.27	0.000	-46, -48, -16	l inferior temporal gyrus
0.049	57	4.60	4.11	0.000	-60, -38, -14	l inferior temporal gyrus
		3.44	3.21	0.000	-52, -46, -6	l inferior temporal gyrus
		3.31	3.10	0.000	-62, -36, -4	l middle temporal gyrus
0.000	160	4.55	4.08	0.000	-46*, -10, 30	l precentral gyrus
		4.46	4.01	0.000	-56, 0, 30	l precentral gyrus
		4.17	3.79	0.000	-60, -6, 26	l precentral gyrus
HALF SESSION 2						
0.000	203	6.79	5.55	0.000	-48, -56, -18	l fusiform gyrus
		4.52	4.06	0.000	-46, -48, -16	l inferior temporal gyrus
0.000	321	5.14	4.50	0.000	-58, -2, 22	l postcentral gyrus
		4.33	3.92	0.000	-46*, -10, 30	l precentral gyrus
		4.23	3.84	0.000	-58, -6, 42	l postcentral gyrus

0.000	220	4.95	4.37	0.000	-34, -68, 46	l inferior parietal gyrus
		4.80	4.26	0.000	-28, -72, 50	l superior parietal gyrus
		4.27	3.87	0.000	-34, -56, 42	l inferior parietal gyrus
0.000	205	4.93	4.35	0.000	-40, 24, 20	l tr. inferior frontal gyrus
		4.54	4.07	0.000	-46, 30, 20	l tr. inferior frontal gyrus
		3.88	3.57	0.000	-42, 32, 12	l tr. inferior frontal gyrus
0.000	179	4.49	4.03	0.000	54, -4, 34	r postcentral gyrus
		4.06	3.70	0.000	60, 2, 24	r postcentral gyrus
HALF CONJUNCTION						
0.000	171	5.70	4.88	0.000	-46, -56, -14	l fusiform gyrus
		4.52	4.06	0.000	-46, -48, -16	l inferior temporal gyrus
		3.60	3.34	0.000	-42, -52, -8	l inferior temporal gyrus
0.001	120	4.45	4.01	0.000	-28, -74, 48	l superior parietal gyrus
		4.14	3.77	0.000	-32, -68, 56	l superior parietal gyrus
		4.00	3.66	0.000	-32, -66, 44	l inferior parietal gyrus
0.001	117	4.33	3.92	0.000	-46*, -10, 30	l precentral gyrus
		4.13	3.76	0.000	-56, -4, 28	l postcentral gyrus

context of constrained sentences provides the necessary information to retrieve the target word.

MEG Discussion

The MEG results show the expected alpha-beta power decreases in accordance with previous findings by Piai et al. (2014, 2015, 2017a, 2017b). These decreases have previously been argued to reflect context-dependent aspects as well as motor preparation of word production, especially in the beta frequency range. However, the analysis of the present data only yielded left-lateralized significant clusters of alpha-beta decreases, whereas motor preparation for speaking activates speech motor areas in both hemispheres. This suggests that the motor activity in the right-hemisphere was not significant enough to be captured in the present study. Further, none of the obtained clusters in the left hemisphere clearly involved left speech motor areas, meaning that also the captured activity most likely did not reflect any motor preparation. Thus, the present findings indicate that motor preparation is not such a robust aspect of alpha-beta power decreases as it has previously been argued, but that the alpha-beta effect mostly reflects the context-driven retrieval of concept and word information.

Comparing the results from both sessions, session 2 shows a stronger and spatially more distributed alpha-beta power decrease that is more significant than that of session 1. Especially the involvement of the middle frontal gyrus in session 2 is surprising, as there was no such frontal activity obtained in session 1. This could possibly represent a familiarization effect with the whole experimental procedure on the side of the participants, as session 2 was an analogue replication of session 1 and only differed in the stimulus materials. This could be further investigated by looking at the individual power decrease maps for each participant to see whether this is a common effect across the group. Alternatively, frontal MEG activation for picture naming has also been found to vary between participants (Liljeström et al., 2009).

In terms of consistency, the activation maps of session 1 and session 2 share a high amount of overlap, yielding a marginally small Dice coefficient of 0.49.

When analyzing only the first half of the MEG sessions, the core areas of the strongest power decreases are the same as for the full sessions, but not as spatially spread out. Accordingly, the area of overlapping activity from the half sessions is smaller than the overlap of the full sessions. Also, the Dice coefficient decreased from 0.49 to 0.35. Interestingly, the robustness of the source localization clusters obtained with the non-parametric Monte Carlo permutation test increased from full to half analysis for both sessions. In other words, the likelihood of accidentally finding a stronger result than the observed one out of 1000 permutations reduced by more than one half. This shows that the quality of the acquired data decreases with the duration of the session, probably partly due to movement and fatigue of the participants. In general, for MEG this suggests that acquiring more data is not always better and does not necessarily increase the significance of the investigated effect.

fMRI Discussion

The brain activity captured with BOLD-fMRI revealed significant clusters, mostly but not exclusively in the left hemisphere. Left-hemisphere BOLD increases were spread over the frontal, parietal, and temporal lobe. A strong aspect of both fMRI sessions as well as their overlap is the motor activity in the right and left postcentral gyrus. Obtaining this in the contrast of constrained over unconstrained trials but not in the reverse direction suggests that motor preparation activity is closely linked to concept and word retrieval and depends on participants' knowledge of which word will be articulated. Possibly, participants already imagine to pronounce the word while waiting for the picture to appear. Thus, the motor preparation most likely includes speech planning in terms of adjusting the speech organs for articulation and pronunciation. This could be further investigated with the existing data of the present study by comparing the motor preparation activity of both conditions at the time point when participants could retrieve the concept of the target word. For constrained trials this would be at the same time point as analyzed here, the pre-picture interval. For unconstrained trials, however, the concept is only presented by the appearance of the picture. Thus, an adequate timepoint for this would be after picture onset, but before participants' speech onset to prevent increased scanning artefacts through speech-related motion.

The results of session 1 reveal some clearly visible left-lateralized clusters of BOLD activity. However, this activity majorly diminished from session 1 to session 2, meaning that participants showed less BOLD increases when performing the experiment a second time. This diminishing activity from session 1 to session 2 also largely impairs across-session consistency. The conjunction analysis therefore only captures few spatially overlapping areas of BOLD increases and the corresponding Dice coefficient quantifying the overlap of the full fMRI sessions results in 0.43.

An alternative approach to improve the power of the fMRI experiment and possibly capture stronger BOLD signals in session 2 would be a block design. As the BOLD response also reflects long-lasting processes it is quite sensitive to variations from trial to trial (Liljeström et al., 2009). Therefore, trials could be presented in blocks of sentences in the same condition, instead of in a randomized trial order as it was in the present experiment. That way, participants would constantly activate the same brain regions for the duration of one block. Especially with the sensitive contrast between the two conditions in our paradigm, that only differed in the context of the sentence preceding the picture, an event-related design might be too subtle and noisy to capture strong BOLD increases between conditions. Another possibility to increase power for fMRI is to restrict the analysis to a region of interest. But for the purpose of the present project, to investigate the reorganization of the brain, it is important that all analyses are conducted on a whole-brain level, without any prior spatial restrictions.

The analysis of the first half of fMRI session 1 only yielded about one third of the BOLD increase clusters that resulted from the full session analysis. Also, the number of significant voxels composing the clusters, meaning the cluster size, reduced vastly. This also holds for session 2, but here the half session analysis still revealed all brain areas that were detected to show BOLD increases in the full session analysis. Whereas for session 1 many detected areas of BOLD increases are not significant anymore for the first half of session 1. In line with this, the Dice coefficient of overlap for the first half of both sessions also decreased to 0.31. Especially considering the loss of clusters when only analyzing the first half of session 1 suggests that the power of fMRI data generally increases with the amount of data collected. However, the session duration and the fact that fatigue and motion of the participants over time increasingly introduce noise also need to be considered. These aspects constitute limitations regarding fMRI session duration.

MEG vs. fMRI

The obtained activation maps of both methods do not completely converge across experiments, showing that both methods capture different aspects of neuronal activity. MEG is sensitive to the magnetic fields induced by electrical currents in the brain on a sub-second time scale, whereas fMRI depends on the blood oxygen level that changes in form of a 6 to 12 second response curve. This time constraint of the fMRI signal constitutes a plausible explanation of the divergence of activity detected per method. Liljeström et al. (2009) for example obtained visual activation for 300 ms stimuli with

MEG, but not with fMRI. They argue that very short stimuli might not be detected with fMRI, as the BOLD response is rather slow and may also reflect a summation of long-term processes. Other studies agree with this and further discuss that also MEG might fail to capture activity that shows significant BOLD responses in fMRI, in cases where neurons are not activated in synchrony (Kujala et al., 2014; Vartiainen et al., 2011). Also, MEG is less sensitive to tangential than radial components of neuronal currents (Kujala et al., 2014), meaning that the magnetic fields of neuronal currents that flow radial to the scalp are not detected well with MEG.

For the findings of the present study, this indicates that the large activation clusters detected with MEG are mostly based on computations with millisecond time scales and thus do not evoke a wide-spread corresponding BOLD response. Further, the fMRI results of the present study reveal consistent motor activity that was not reflected in the MEG results, but has previously been detected with the same paradigm in MEG (Piai et al., 2015). This suggests that the lack of captured motor activity with MEG in the present study is due to analysis thresholds, rather than unsynchronized or radially oriented neurons. Although, for this direction of comparing the results across methods, it is important to keep in mind that the MEG analysis in the present study was based on prior knowledge of the frequency of interest (Piai et al., 2014, 2015, 2017a, 2017b). In this sense, the obtained MEG results already diverge from the fMRI results because they are priorly restricted to a frequency range.

As depicted in models of word retrieval, the process of word production undergoes several stages before the word can be produced (Roelofs, 1992). In the present study, the concept of the target word is either presented by sentence context or picture presentation. In constrained trials, participants retrieve the concept based on the information that is given in the sentence context. In unconstrained trials, the concept retrieval depends on identifying the object that is presented in the picture. When the concept is accessed, further information about the word and its phonology are retrieved. Only then, at the later stages of word retrieval, the preparation for articulation takes place. Relating this to the results of the present study suggests that fMRI better reflects the motor preparation stages of word planning, but not necessarily the early stages of conceptual retrieval. In line with the reasons for diverging findings discussed above, this indicates that the early stages of concept retrieval might occur too fast for a measurable BOLD response to establish. MEG, in contrast, captures the computations underling the early stages of concept retrieval.

The brain regions that seem to be most crucial for concept and word retrieval in contextually constrained sentences in both experiments are the left temporal and inferior parietal lobe. These regions were commonly activated across sessions and captured by both methods in the experiments. The inferior temporal gyrus has previously been shown to be a crucial area to access lexical semantic concepts in object recognition and word production (Price, 2012; Roelofs, 2014). Additionally, the inferior parietal lobe has been argued to be essential in predicting and integrating semantic knowledge (Binder, Desai, Graves & Conant, 2009; Price, 2012). This further corresponds to the findings of Piai et al. (2017b) that established a causal link between these brain regions and context-driven word retrieval. In this study, left-hemisphere stroke patients performed the same task that was employed in the present study. Patients whose left middle and superior temporal gyri and inferior parietal lobe were damaged by the lesion did not show a behavioral effect of sentence context. In addition, their EEG revealed a reduced alpha-beta power decrease. Thus, the results of the preset study and the fact that the left temporal and inferior parietal lobe were captured with MEG as well as fMRI support the claim that these brain regions are essential for context-dependent word retrieval.

To further investigate the divergence between the fMRI and the MEG results, the MEG data of the present study could be analyzed without a priorly specified frequency range and different threshold. This would reveal whether the activity captured with fMRI but not with MEG was not robust enough for the analysis strategies of the present study, or whether this absence is due to undetectable neuronal activity for MEG, as discussed above. However, as the patient study serves to investigate language processing rather than speech preparation, the motor activity is not a major aspect of interest. Especially since the aim is to assess whether brain activity lateralizes from the left to the right hemisphere, it is more suitable to employ a paradigm that only elicits left-lateralized effects in neurotypical control subjects.

When comparing the results for both experiments with one another, the across-session familiarization seems to have a reversed effect across experiments. While MEG results show more and stronger power decreases in session 2 than in session 1, fMRI results show less BOLD increases in session 2 than in session 1. As a next step to further investigate this finding, the data of the present study should be analyzed on the subject-level. This will reveal whether this familiarization effect is consistent across participants, or else only over-represented in some of the participants. Alternatively, this could also be examined by including a potential third session of the same paradigm for both experiments. This would allow to investigate the activity captured at another time point following session 2 and help to determine the direction of this potential familiarization effect.

Finally, the Dice coefficients of both experiments are higher for the full sessions and decrease when the calculation only includes the first half of the sessions. Notably, the full MEG session overlap yields the highest Dice coefficient, even though the calculation is based on an average of 22 trials less than the full fMRI session overlap. For the half session Dice coefficient the number of included trials is equal, but MEG still yields a higher coefficient than fMRI. Overall, this indicates that MEG trials result in a higher Dice coefficient than fMRI trials, which is further supported by the fact that the half MEG session overlap reveals the highest ratio of Dice coefficient and average number of trials per participant that the calculation is based on. As a further analysis step with the existing data of the present study, the individual Dice coefficients for the activation overlap from session 1 to session 2 for each participant should be calculated. As the patient study will be investigating individual cases with different lesion size and location per patient, the process of recovery will also be evaluated on an individual level. Thus, Dice coefficients on subject-level are ultimately more informative for the purpose of establishing the clinical tool than group-level Dice coefficients.

Conclusion

Recalling the main goal of the project to develop a clinical tool to investigate language recovery in patients, the results of the present study serve to determine the optimal parameters for this purpose. The aim was to find a suitable imaging method that provides spatially reliable effect profiles over time that can be measured within short testing sessions. The obtained results reveal a different focus of the reflected aspect of neuronal activity per method. Particularly, they vary in terms of spatiality and significance for short and long sessions. Also, the Dice coefficients reflect different relationships of consistency per included trials for both methods. By means of the performed analyses of the present study, the results convincingly determine MEG to be a- more suitable method for the clinical tool than fMRI. This does not derive from a direct statistical comparison between the two methods, but rather from a more practical origin. Aspects such as the laterality of the effect or the duration of the testing session are crucial points that should not be neglected when it comes to the development of a clinical tool for patient testing.

Acknowledgements

I want to thank everyone who supported me and my project in the past year in any way, including all members of the *Language Function and Dysfunction* group and the students in the trainee room. A special thanks to Nathalie Meyer for helpful discussions and reviewing my thesis, as well as Mauricio Saldivar for helping me with programming and the layout. Further, thanks to Atsuko Takashima for all her support and guidance for the fMRI part of the project. Finally, I want to thank my supervisor Vitória Piai for giving me the opportunity to work on such a great project during my master internship, as well as for everything I have learned from her in the past year and all the support, suggestions, and guidance I received.

References

Binder, J. R., Desai, R. H., Graves, W. W., & Conant, L. L. (2009). Where is the semantic system? a critical review and meta-analysis of 120 functional neuroimaging studies. Cerebral Cortex, 19(12), 2767–2796.

Boersma, P., & Weenink, D. (2017). Praat, software for speech analysis and synthesis.

Bradshaw, A. R., Thompson, P. A., Wilson, A. C., Bishop, D. V., & Woodhead, Z. V. (2017). Measuring
language lateralisation with different language tasks: a systematic review. PeerJ, 5, e3929.

Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The bank of standardized stimuli (boss), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. PloS one, 5(5), e10773.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. 1988, hillsdale, nj: L. Lawrence Earlbaum Associates, 2.

Conner, C. R., Ellmore, T. M., Pieters, T. A., DiSano, M. A., & Tandon, N. (2011). Variability of the relationship between electrophysiology and bold-fmri across cortical regions in humans. Journal of Neuroscience, 31(36), 12855–12865.

Dronkers, N. F., & Baldo, J. (2010). Language: aphasia. In

Encyclopedia of neuroscience. Elsevier
Ltd.

Duffau, H. (2005). Lessons from brain mapping in surgery for low-grade glioma: insights into associations between tumour and brain plasticity. The Lancet Neurology, 4(8), 476–486.

Findlay, A. M., Ambrose, J. B., Cahn-Weiner, D. A., Houde, J. F., Honma, S., Hinkley, L. B., … Kirsch, H. E. (2012). Dynamics of hemispheric dominance for language assessed by magnetoencephalographic imaging. Annals of neurology, 71(5), 668–686.

Hillebrand, A., Nissen, I., Ris-Hilgersom, I., Sijsma, N., Ronner, H., Van Dijk, B., & Stam, C. (2016). Detecting epileptiform activity from deeper brain regions in spatially filtered meg data. Clinical Neurophysiology, 127(8), 2766–2769.

Kim, M. J., Holodny, A. I., Hou, B. L., Peck, K. K., Moskowitz, C. S., Bogomolny, D. L., & Gutin, P. H. (2005). The effect of prior surgery on blood oxygen level–dependent functional mr imaging in the preoperative assessment of brain tumors. American journal of neuroradiology, 26(8), 1980–1985.

Kujala, J., Sudre, G., Vartiainen, J., Liljeström, M., Mitchell, T., & Salmelin, R. (2014). Multivariate analysis of correlation between electrophysiological and hemodynamic responses during cognitive processing. NeuroImage, 92, 207–216.

Liljeström, M., Hulten, A., Parkkonen, L., & Salmelin, R. (2009). Comparing meg and fmri views to naming actions and objects. Human brain mapping, 30(6), 1845–1856.

Liljeström, M., Stevenson, C., Kujala, J., & Salmelin, R. (2015). Task-and stimulus-related cortical networks in language production: Exploring similarity of meg-and fmri-derived functional connectivity. Neuroimage, 120, 75–87.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). Fieldtrip: open source software for advanced analysis of meg, eeg, and invasive electrophysiological data. Computational intelligence and neuroscience, 2011, 1.

Piai, V., Meyer, L., Dronkers, N. F., & Knight, R. T. (2017a). Neuroplasticity of language in left-hemisphere stroke: Evidence linking subsecond electrophysiology and structural connections. Human brain mapping, 38(6), 3151–3162.

Piai, V., Roelofs, A., & Maris, E. (2014). Oscillatory brain responses in spoken word production reflect lexical frequency and sentential constraint. Neuropsychologia, 53, 146–156.

Piai, V., Roelofs, A., Rommers, J., & Maris, E. (2015). Beta oscillations reflect memory and motor aspects of spoken word production. Human brain mapping, 36(7), 2767–2780.

Piai, V., Rommers, J., & Knight, R. T. (2017b). Lesion evidence for a critical role of left posterior but not frontal areas in alpha–beta power decreases during context-driven word production. European Journal of Neuroscience, 48(7), 2622-2629.

Price, C. J. (2012). A review and synthesis of the first 20 years of pet and fmri studies of heard speech, spoken language and reading. Neuroimage, 62(2), 816–847.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. Cognition, 42(1-3), 107–142.

Roelofs, A. (2014). A dorsal-pathway account of aphasic language production: The weaver++/arc model. Cortex, 59, 33–48.

Skipper-Kallal, L. M., Lacey, E. H., Xing, S., & Turkeltaub, P. E. (2017). Right hemisphere remapping of naming functions depends on lesion size and location in poststroke aphasia. Neural plasticity, 2017.

Stolk, A., Todorovic, A., Schoffelen, J.-M., & Oostenveld, R. (2013). Online and offline tools for head movement compensation in meg. Neuroimage, 68, 39–48.

Team, R. C. (2017). R: A language and environment for statistical computing. vienna, austria: R foundation for statistical computing; 2017. ISBN3-900051-07-0 https://www. Rproject.org.

Turkeltaub, P. E., Messing, S., Norise, C., & Hamilton, R. H. (2011). Are networks for residual language function and recovery consistent across aphasic patients? Neurology, 76(20), 1726–1734.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., … Joliot, M. (2002). Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri singlesubject brain. Neuroimage, 15(1), 273–289.

van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization. Behavior research methods, 38(4), 584–589.

Vartiainen, J., Liljeström, M., Koskinen, M., Renvall, H., & Salmelin, R. (2011). Functional magnetic resonance imaging blood oxygenation level-dependent signal and magnetoencephalography

evoked responses yield different neural functionality in reading. Journal of Neuroscience, 31(3), 1048–1058.

Wilson, S. M., Bautista, A., Yen, M., Lauderdale, S., & Eriksson, D. K. (2017). Validity and reliability of four language mapping paradigms. NeuroImage: Clinical, 16, 399–408.

Wilson, S. M., Lam, D., Babiak, M. C., Perry, D. W., Shih, T., Hess, C. P., … Chang, E. F. (2015). Transient aphasias after left hemisphere resective surgery. Journal of neurosurgery, 123(3), 581–593.

Abstracts

Proceedings of the Master's Programme Cognitive Neuroscience is a platform for CNS students to publish their Master theses. Given the number of submissions, we select the articles that received the best reviews, under recommendation of our editors, for the printed edition of the journal. The abstracts of the other articles are provided below, and for interested readers a full version is available on our website: www.ru.nl/master/cns/journal.

A Bayesian Account for Estimating the Number of Neurons during Spike Sorting

Kees van Rooijen, Bernhard Englitz

Extracellular recordings have long been an invaluable tool for understanding neural population activity. Spike sorting is the process of unmixing the contributing sources in a recording to obtain the spiking activity of individual neurons. Identifying the correct number of neurons is an error-prone process involving a considerable amount of intrinsic uncertainty. However, most spike sorting algorithms do not account for this uncertainty, but instead use a single point estimate. Using a fully probabilistic approach, we demonstrate that the point estimate leads to systematic misestimation of the number of neurons. We estimate the number of neurons present in the data by sampling from the actual posterior distribution using reversible jump Markov chain Monte Carlo, in the context of realistic ground truth data. The expected value of the probabilistic estimate is then compared to the widely used maximum a posteriori (MAP) estimate of the number of neurons. We find that even in the absence of incorrect modelling assumptions, using a point estimate leads to a systematic underestimation of the number of present neurons. This effect is visible for a wide range of values for the recording time and the noise available in the recording. More specifically, we find that decreasing noise leads to a decrease in this bias only for high sorting accuracy. If the sorting accuracy is low, this effect is reversed. Furthermore, we find that the size of the bias can initially be decreased by increasing the recording time, but for longer recordings this effect comes to a halt. Misestimating the number of neurons contributes to errors in dividing spikes into clusters, and thus impacts the clarity of the results, e.g. by fusing different neurons, or splitting single neurons. As a consequence, correlations and other estimated properties would be affected. The present results provide an analytical guide to correct for this error.

Spectrotemporal Modulation Sensitivity in Developmental Dyslexia

Yingdi Xie, Kiki Van der Heijden, Elia Formisano, Milene Bonte

Previous research has suggested various general auditory processing deficits which may underlie the reduced phonological awareness in developmental dyslexia. However, despite the importance of spectrotemporal modulations for speech processing, there is no study to date which systematically examined auditory processing of the modulation components characteristic of speech in dyslexia. Thus, the present study aims to address whether dyslexic and normal readers differ in perceptual sensitivity to these spectrotemporal modulations. We predict a reduced sensitivity in dyslexic readers. We used adaptive transformed up-down procedure (Chi et al., 1999; Levitt, 1971) to estimate detection thresholds of dyslexic and normal readers for different combinations of spectrotemporal modulations in dynamic ripples and AM broadband noises. Contrary to our prediction multilevel modeling revealed that there was no significant group difference, indicating comparable modulation sensitivity between dyslexic and normal readers. It opposes all present hypothesized auditory deficits. Moreover, we found a significant interaction between the effects of temporal modulations and those of spectral modulations. It implies a dependency of these two processing mechanisms. Future research is needed to further inspect the auditory processing of speech as well as other natural sounds in dyslexia.

The Role of Prediction Disconfirmation in Language Comprehension and its Consequences for Memory

Laura Giglio, Joost Rommers

The consequences of prediction disconfirmation in language comprehension are still unclear, as well as the functional role of EEG signatures underlying prediction disconfirmation. Here we investigated the consequences of prediction disconfirmation for later recognition memory. We ran a sentence comprehension study with sentences ending with plausible unexpected words that either disconfirmed a prediction or followed an unconstraining context. This was followed by a surprise memory test where recognition memory for the previously read words was probed. We found that recognition memory performance was better for items that previously disconfirmed a prediction than when no strong predictions could be made. By back-sorting items based on later memory responses, we further characterized the EEG signal at sentence comprehension and memory retrieval on the basis of processes underlying successful encoding. We found that a late parietal positivity was predictive of subsequent recognition and was enhanced in items disconfirming a prediction. In addition, time-frequency analysis showed stronger beta decreases in items disconfirming a prediction. Effects of contextual constraint were also found prior to the presentation of the disconfirming word with a beta decrease possibly involved in word retrieval. At retrieval, items disconfirming a prediction were characterized by an old/new effect in the form of an N400 and LPC suggesting that prediction disconfirmation enhanced both familiarity and recollection memory processes. Overall, these findings show that prediction disconfirmation has beneficial consequences on memory encoding and retrieval which supports models of prediction error as a driver of memory encoding.

Luminance Contrast Modulation of Spatial Frequency VEPs in 5-Year-Old Children

Myrte Druyvesteyn, Tessa van Leeuwen

Different components of visual perception such as spatial frequency (SF) detection, luminance contrast sensitivity and colour vision develop at different rates. At 5-6 years of age, High SF (HSF) sensitivity is nearly adult-like and luminance contrast is also nearing the end of its maturation around 5-7 years. Low SF (LSF), on the other hand, still has a substantial development to undergo till the age of 12. Additionally, there is a shift in selective processing of SF from the N2 component of visual evoked potential (VEP) in 3-6 year olds to the earlier N80 component of VEP at ages 7-8. Adult literature has revealed a complex interaction between SF and luminance contrast. How this interaction takes form in children however is relatively unknown. The aim of this study is to investigate the interaction between spatial frequency and luminance contrast in 5-6 year olds when these different components of visual perception are at different developmental stages to gain more insight into how this interaction develops. VEP modulations of HSF/LSF gratings at high and low luminance contrast levels were measured with electroencephalography (EEG) in 22 children aged 5. Results show that the P1 peak shows an adult-like pattern in 5 year olds, with an increase in amplitude as contrast increases, irrespective of SF. The N80 peak however shows only partial resemblance to the adult VEPs, showing the same selective enhanced activity for HSF but not LSF, yet is still lacking the interaction with contrast. Contrarily the N2 peak closely resembles the interactions patterns found for the N80 peak in adults. Taken together, the argument can be made that the complex interaction between SF and contrast is characteristic for the selective processing of HSF and LSF stimuli and that the interaction between spatial frequency and contrast depends on the maturity of the underlying systems.

Computational-level analysis of insight problem solving

Stefano Gentili, Iris Van Rooij, Mark Blokpoel, Todd Wareham

Having an insight is a central aspect of the human ability to do problem-solving. When trying to reach a solution for a problem, having an insight is what lets us formulate a problem in a way that we can come up with an answer. However, there is no scientific consensus about the cognitive mechanisms of reaching insight. In this thesis we will briefly present the current literature about insight problem solving and show some of the shortcomings that have been affecting it. In particular, we will ask if existing models of insight problem solving are able to generalize and be applicable to all insight problems or only to a selected few they are specifically designed for. We will focus on an existing model of insight problem solving, and give a computationally-backed argument about how the existing model could theoretically be used to encode any type of problem. We will also give an example of an encoding of an insight problem in the existing model and from there we will point out that the model seems to construct its input so that the solution is easily found, almost built into the model. We will argue about the risks of this in-building and then we will consider what could be a minimum to be built-in. This will lead to a reformulation of a more useful model capable of insight. Finally, with another result we will show that an important aspect of problem solving is often overlooked, namely the amount of time (or steps) necessary before finding the solution. Indeed we will show that pre-specifying this amount is actually necessary for a cognitive model of insight. These results will give important theoretical constraints to future theories of insight problem solving. Furthermore, the thesis will suggest specific future research approaches for advancing our knowledge in this field.

Investigating the Glial Subtype Contributing to Sleep Disturbances in CHARGE Syndrome and Autism Spectrum Disorders

Isabel Terwindt , Mireia Coll-Tané , Annette Schenck

CHARGE Syndrome (CS) and a specific subtype of Autism Spectrum Disorder (ASD) are caused by mutations in the chromatin remodelers CHD7 and CHD8, respectively. Both disorders share common features, one of them is the presence of sleep disturbances. Notably, yeast two-hybrid assays showed a physical interaction between the CHD7 and CHD8 proteins, suggesting their presence in a multisubunit complex in humans. CHD7 and CHD8 have a common ortholog in Drosophila melanogaster, kismet, which has been implicated in circadian regulation. Unpublished data from our lab show that the pan-glial knockdown of kismet results in sleep fragmentation, resembling the sleep deficits observed in CHD7/CHD8 patients. Our aim was to further identify which glial subtype underlies the sleep dysfunction, by specifically knocking down kismet in isolated subtypes and monitoring the resulting activity and sleep. Additionally, we investigated the role of kismet in glia during development at different developmental stages. We identified that the two types of glial cells forming for the blood-brain barrier, subperineurial glia (SPG) and perineurial glia (PG), are the underlying subtypes causing sleep disturbances upon kismet knockdown. A decrease of the mean duration of the sleep episodes, specifically during the dark period was observed with a subsequent increase in the number of sleep episodes. Kismet knockdown in the SPG and PG cells resulted in such fragmented sleep patterns. Our results identify a novel role of the blood brain barrier in the regulation of sleep.

Neurocognitive Mechanisms of Smoking cue-reactivity and Inhibitory Control in Novice Smoking Adolescents.

Kelly van Egmond, Maartje Luijten, Joyce Dieleman, Esther Aarts

The developmental period of adolescence is strongly associated with an increased sensitivity towards motivational cues and an impaired inhibitory control, rendering adolescents especially susceptible for future tobacco dependence. Indeed, smoking cue-reactivity and inhibitory control deficits have been linked to tobacco dependence. It is still unclear whether these deficits are a consequence of tobacco abuse or may act as risk factors to develop tobacco dependence. Therefore, it is important to investigate underlying neurocognitive mechanisms of smoking cue-reactivity and inhibitory control in individuals at risk for tobacco dependence. For the current study, novice (n=75) and non-smoking (n=25) adolescents aged 14-19 performed both a smoking cue-reactivity paradigm and a GO/NO-GO task during fMRI. Participants were matched regarding age, sex, educational level (low, middle, and high) and AUDIT scores. Both behavioral and fMRI data were analyzed for group differences. Two sample t-tests of the fMRI data revealed no significant group differences in smoking cue- reactivity and inhibitory control-related BOLD activation. The same applied for the behavioral analysis, as the groups did not significantly differ in performance. Furthermore, ROI analyses revealed no group differences in smoking cue-related VMPFC and inhibitory control-related rIFG activity. Current results led to the conclusion that novice and non-smoking adolescents elicit similar smoking cue and inhibitory control-related BOLD activation. Tobacco does not seem to influence the adolescent' brain during the early and experimental stages of smoking. Speculating that tobacco induced brain changes are a consequence of severe tobacco use, rather than representing a priori deficits rendering individuals susceptible to tobacco use/dependence.

Psychobiological Mechanisms of Costly Avoidance Behaviour

Anneloes Hulsman, Floris Klumpers

Excessive avoidance behaviour is a cardinal symptom of anxiety and depressive disorders. Recent findings show that excessive avoidance is a better predictor of poor disease outcome than current levels of anxiety and depression. Even though avoidance plays an important role in anxiety and depression, little is known about the underlying psychobiological mechanisms. Moreover, there is a lack of previous studies that assess approach-avoidance behaviour with ecologically-valid paradigms. Therefore, we developed a novel fearful avoidance task (FAT) in which multiple reward and threat levels are integrated under high arousal. Concomitantly, we measured startle responses (study 1, N=343) and neural responses (fMRI, study 2, N=29) to identify the psychobiological mechanisms that are involved in approach-avoidance decisions. In both studies our task was successful in creating an approach avoidance conflict. In the first study we did not find a strong link between avoidance behaviour and defensive responses, suggesting that there are also other mechanisms involved in avoidance. Behavioural results suggested a role for appetitive processing, which is further investigated in the second study. In the second study we found that both appetitive and defensive processes are present during presentation of the offer and anticipation of the outcome. Taken together, our results suggest that avoidance behaviour is likely driven by a combination of appetitive and defensive mechanisms. Future analyses will have to demonstrate whether the found activity in appetitive and defensive regions is indeed related to avoidance behaviour.

Neuronal activity patterns associated with a traumatic experience in PTSD-vulnerable and -resilient mice

L.V.J. van Melis, B.C.J. Dirven, T. Kozicz, M.J.A.G. Henckens

Posttraumatic stress disorder (PTSD) is a clinical condition that can develop when an individual is exposed to a traumatic event, leading to significant social, occupational and interpersonal impairment. Although many individuals experience at least one traumatic event in their lifetime, only a subset of around 10 to 20% of these develops PTSD. This suggests the existence of inter-individual differences in vulnerability to developing psychopathology after trauma exposure. Discovering the exact neurobiological nature of these differences may be key to understanding the factors that constitute PTSD resilience, ultimately enabling the development of new treatment options. In this study we use an established mouse model for PTSD induction that employs a battery of behavioral tests to differentiate between subgroups of mice that show varying degrees of PTSD symptomatology after trauma exposure. Since re-experiencing of traumatic memories forms a core feature of PTSD, it is interesting to study the neuronal activity patterns that are induced by a specific traumatic experience and that are likely to be involved in the formation of a trauma memory trace. By coupling a fluorescent molecular label to the promoter of the immediate-early gene Arc, neurons that are active in these mice during trauma exposure can be labeled. In our study, we try to investigate 1) if neuronal activation patterns during trauma exposure differ between the PTSD-like and resilient mice, 2) if the neurons that are activated during trauma exposure are re-activated during subsequent re-exposure to the trauma environment and 3) whether both subgroups of mice differed from each other in neuronal type and distribution in the hippocampus. Neuronal activation during trauma exposure did not differ between PTSD-like and resilient mice, but during re-exposure to the trauma environment PTSD-like mice showed significantly less neuronal activation in both the dorsal and ventral hippocampus. A small percentage of neurons that were activated during trauma exposure were re-activated during re-exposure to the trauma environment, although this only significantly differed between both subgroups of mice in the ventral inferior dentate gyrus. This suggests that PTSD-like and resilient mice did not differ in fear memory encoding, but did so in memory recall. To investigate whether PTSD-like and resilient mice differed from each other in the type of interneurons that were present, brain slices of these mice were stained for the GABAergic cell markers parvalbumin and somatostatin. PTSD-like mice showed significantly more parvalbumin positive cells in the ventral CA1, while the number of somatostatin positive cells in the dorsal dentate gyrus and dorsal CA1 was significantly reduced when compared to resilient mice, suggesting an altered inhibitory network between these subgroups of mice.

White matter changes in the perforant path in ALS Providing evidence for ALS as a multisystem disease

M. Hiemstra, J. Mollink, M. Pallabage-Gamarallage, I.N. Huszar, K.L. Miller, M. Jenkinson,
Ansorge, A.M. van Cappellen van Walsum

Amyotrophic lateral sclerosis (ALS) is a severe, progressive and incurable motor disease. Roughly 20% of the ALS patients are affected by a level of cognitive decline that meets the criteria for behavioral frontotemporal lobe dementia (bvFTD). ALS and bvFTD share some clinical and pathological features, for example the deposition of TAR DNA binding protein 43(pTDP-43) in several brain regions that are part of the circuit of Papez. Previous literature suggests involvement of the perforant path, a white matter tract in the hippocampus that is part of the circuit of Papez in patients with both ALS and bvFTD. We hypothesize that white matter degeneration in the perforant pathway is a key feature of ALS, providing a neuronal correlate for ALS as a multisystem disorder. To verify our hypothesis we studied white matter changes in ex-vivo hippocampal blocks from patients with known ALS (n=13) and controls (n=5) using diffusion MRI. The dMRI results were evaluated using polarised light imaging (PLI), a microscopy technique sensitive to density and orientation of myelinated axons. From the same hippocampal blocks, sections were cut and stained for myelin, pTDP-43, neurofilaments and activated microglia. The dMRI results show a significant decrease in fractional anisotropy (p=0.018) and an increase in mean diffusivity (p=0.0017), axial diffusivity (p=0.023) and radial diffusivity (p=0.028) in the perforant path in ALS patients compared to controls, likely indicating a loss of fibres. The PLI retardance values within the perforant path were lower in cases compared to controls, however not significantly (p=0.16). The retardance correlates with the fractional anisotropy (p=0.04). Furthermore, an increase in dispersion was observed in ALS specimens(p=0.04), implying a less organised axonal structure. Histology data showed a (non-significant) increase in myelin (PLP) (p=0.11) and an increase in neurofilaments (SMI-312) (p=0.03) in ALS cases compared to controls. No differences were found in the amount of inflammation and two out of the 13 ALS cases exhibited pTDP-43 pathology in the hippocampus. These results demonstrate degradation of the perforant path in ALS patients, providing a potential neuronal correlate for the cognitive symptoms observed in ALS and substantiating the hypothesis that ALS and bvFTD are part of the same spectrum of diseases. Future research should focus on correlating the degree of clinically observed cognitive decline to the amount of white matter atrophy in the perforant path.

Elucidating the diagnostic, therapeutic and mechanistic implications of stroke in Alzheimer's disease

R.Cansu Egitimci, Amanda Kiliaan, Maximilian Wiesmann

By 2040, Alzheimer's disease (AD) will affect approximately 81 million people worldwide. Studies show that hypertension, diabetes, atherosclerosis, and obesity are risk factors for both vascular disorders such as stroke and AD. Risk of both stroke and AD increases with age. Emerging evidence shows that stroke increases the risk of developing AD and in return, AD is a risk factor for stroke. But exact mechanisms behind this correlation are unknown and remain to be investigated. To understand underlying mechanisms of stroke on AD pathophysiology and sex-specific differences, we investigated the effect of ischemic stroke on female and male double transgenic APP SWE /PS1 ΔE9 (AD) and C57BI/6 wild type (WT) mice from 3- months until 12- months of age. Mice were subjected to transient occlusion of the right middle cerebral artery (tMCAo) to induce an ischemic stroke. Before the stroke induction baseline measurement of general health parameters (e.g, body weight, blood pressure) and motor skills (e.g. activity, strength, coordination) were measured. After stroke induction, these measurements were repeated at several time points along with MRI measurements (e,g, rsfMRI, DTI, FAIR-ASL) to assess the effect of stroke on brain structure, function, and connectivity. APP SWE /PS1 ΔE9 mice and stroke mice showed impairments in physiological parameters, motor function, locomotion, and cognition. Our data show that stroke not only contributes significantly to AD but also exacerbate the symptoms. APP SWE /PS1 ΔE9 stroke mice were more hyperactive and anxious than APP SWE /PS1 ΔE9 sham mice. Furthermore, male mice showed greater surgery effects than female mice, indicating strong sex differences in stroke and AD pathophysiology. Our findings suggest that there is a strong correlation between vascular risk factors, stroke and AD. We further showed that there is a great difference between sexes and genotypes, indicating any preventative or therapeutic approach should be personalized.

Institutes associated with the Master's Programme Cognitive Neuroscience

Donders Institute for Brain, Cognition
and Behaviour:
Centre for Cognitive Neuroimaging
Kapittelweg 29
6525 EN Nijmegen

P.O. Box 9101
6500 HB Nijmegen
http://www.ru.nl/donders/

Max Planck Institute for Psycholinguistics
Wundtlaan 1
6525 XD Nijmegen

P.O. Box 310
6500 AH Nijmegen
http://www.mpi.nl

Radboudumc
Geert Grooteplein-Zuid 10
6525 GA Nijmegen

P.O. Box 9101
6500 HB Nijmegen
http://www.umcn.nl/

Baby Research Center
Montessorilaan 3
6525 HR Nijmegen

P.O. Box 9101
6500 HB Nijmegen
http://www.babyresearchcenter.nl