

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/317096831>

Long-term test-retest reliability of event-related potential (ERP) recordings during treadmill walking using the mobile...

Article in *Brain research* · May 2017

DOI: 10.1016/j.brainres.2017.05.021

CITATIONS

0

READ

1

6 authors, including:



[John J Foxe](#)

University of Rochester

344 PUBLICATIONS 15,728 CITATIONS

[SEE PROFILE](#)



[John S Butler](#)

Dublin Institute of Technology

78 PUBLICATIONS 641 CITATIONS

[SEE PROFILE](#)



[Sophie Molholm](#)

Albert Einstein College of Medicine

120 PUBLICATIONS 4,205 CITATIONS

[SEE PROFILE](#)



[Pierfilippo De Sanctis](#)

Albert Einstein College of Medicine

35 PUBLICATIONS 494 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



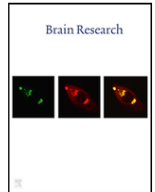
Dystonia & TDT [View project](#)



Go Green Ex: "GO GREEN EX Cubed" Going Outdoors: Gathering Research Evidence on ENvironment: Exercise, Explore, Excite [View project](#)

All content following this page was uploaded by [Pierfilippo De Sanctis](#) on 25 May 2017.

The user has requested enhancement of the downloaded file. All in-text references [underlined in blue](#) are added to the original document and are linked to publications on ResearchGate, letting you access and read them immediately.



Research report

Long-term test-retest reliability of event-related potential (ERP) recordings during treadmill walking using the mobile brain/body imaging (MoBI) approach

Brenda R. Malcolm^{a, b}, John J. Foxe^{a, b, c, d, e, *}, John S. Butler^{a, e, f, g}, Wenzhu B. Mowrey^h, Sophie Molholm^{a, b, c, d}, Pierfilippo De Sanctis^{a, b, i, *}

^a The Sheryl & Daniel R. Tishman Cognitive Neurophysiology Laboratory, Children's Evaluation and Rehabilitation Center (CERC), Department of Pediatrics, Albert Einstein College of Medicine, Bronx, NY 10461, USA

^b Program in Cognitive Neuroscience, The Graduate Center of the City University of New York, New York, NY 10016, USA

^c The Del Monte Institute for Neuroscience, Department of Neuroscience, University of Rochester School of Medicine and Dentistry, Rochester, NY 14642, USA

^d The Dominick P. Purpura Department of Neuroscience, Rose F. Kennedy Intellectual and Developmental Disabilities Research Center, Albert Einstein College of Medicine, Bronx, NY 10461, USA

^e Trinity College Institute of Neuroscience, Dublin, Ireland

^f Trinity College Dublin, Centre for Bioengineering, Trinity Biomedical Sciences Institute, Dublin, Ireland

^g School of Mathematical Sciences, Dublin Institute of Technology, Dublin, Ireland

^h Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY 10461, USA

ⁱ The Saul R. Korey Department of Neurology, Albert Einstein College of Medicine, Bronx, NY 10461, USA

ARTICLE INFO

Keywords:

EEG
Dual-task design
Test-retest reliability
Intraclass correlation coefficient
Response inhibition
Gait
N2/P3
P300
Cognitive performance

ABSTRACT

Advancements in acquisition technology and signal-processing techniques have spurred numerous recent investigations on the electro-cortical signals generated during whole-body motion. This approach, termed Mobile Brain/Body Imaging (MoBI), has the potential to elucidate the neural correlates of perceptual and cognitive processes during real-life activities, such as locomotion. However, as of yet, no one has assessed the long-term stability of event-related potentials (ERPs) recorded under these conditions. Therefore, the objective of the current study was to evaluate the test-retest reliability of cognitive ERPs recorded while walking. High-density EEG was acquired from 12 young adults on two occasions, separated by an average of 2.3 years, as they performed a Go/No-Go response inhibition paradigm. During each testing session, participants performed the task while walking on a treadmill and seated. Using the intraclass correlation coefficient (ICC) as a measure of agreement, we focused on two well-established neurophysiological correlates of cognitive control, the N2 and P3 ERPs. Following ICA-based artifact rejection, the earlier N2 yielded good to excellent levels of reliability for both amplitude and latency, while measurements for the later P3 component were generally less robust but still indicative of adequate to good levels of stability. Interestingly, the N2 was more consistent between walking sessions, compared to sitting, for both hits and correct rejection trials. In contrast, the P3 waveform tended to have a higher degree of consistency during sitting conditions. Overall, these results suggest that the electro-cortical signals obtained during active walking are representative of stable indices of neurophysiological function.

1. Introduction

Event-related potentials (ERPs) are an important tool in neuroscience research, providing temporally-precise electrophysiological correlates of information processing stages relating to perception, cognition and action (Foxe and Simpson, 2002; Kok, 1997). In order to sup-

port the use of ERPs as established indices of neurophysiological function, it is important to determine the degree to which ERP measurement procedures yield consistent results across time. Evidence of intra-individual stability, i.e. reliability, is necessary to establish the validity of results and will add greater weight to conclusions drawn from these studies. The potential use of ERPs as neurobiological markers also depends upon their stability over time (Duncan et al., 2009; Huffmeijer et al., 2014).

* Corresponding authors at: The Sheryl & Daniel R. Tishman Cognitive Neurophysiology Laboratory, Children's Evaluation and Rehabilitation Center (CERC), Department of Pediatrics, Albert Einstein College of Medicine, Bronx, NY 10461, USA.

Email addresses: john_foxe@urmc.rochester.edu (J.J. Foxe); pierfilippo.sanctis@einstein.yu.edu (P. De Sanctis)

For most cognitive neuroscience studies employing electroencephalographic (EEG) recordings, participants engage in what could be considered a minimal behavior approach. This involves being seated in an environment designed to minimize all external stimuli, typically under instructions to limit task responses to simple button presses or saccadic eye-movements. This approach facilitates focus on task-relevant stimulation and minimizes contamination of brain electrophysiological recordings related to motor movements. Numerous studies have evaluated the test-retest reliability of ERP amplitude and latency measures elicited from a variety of paradigms recorded under these typical conditions. Results vary but have typically ranged from moderate to high consistency for both earlier sensory-perceptual components and later-occurring cognitive ones (Brunner et al., 2013; Cassidy et al., 2012; Fallgatter et al., 2002b; Fallgatter et al., 2001). For example, Cassidy et al. (2012) administered four common ERP paradigms to the same group of participants one month apart. Components analyzed included early visual-evoked components (P1, N1), the face-processing associated N170, attentional resource allocation (P3a, P3b), error processing (error-related negativity, error positivity) and memory encoding (P400). Test-retest reliability was assessed with the intraclass correlation coefficient (ICC), a measure of consistency, or stability over time (Shrout and Fleiss, 1979). They found a high level of agreement for all component amplitudes but only latency measures for earlier occurring components (e.g., N170, P1, N1) (Cassidy et al., 2012).

Additionally, a few previous investigations have assessed ERP test-retest reliability over longer time periods, also with varied results. In one study, electrophysiological measures of cognitive response control during a Continuous Performance Test were recorded an average of 2.7 years apart, yet still an excellent degree of consistency was shown for topographic measures, amplitude and to a lesser extent, latencies, of global field power (Fallgatter et al., 2002a). Another evaluation of long-term (~2 years) test-retest reliability of the error-related negativity (ERN), thought to measure executive control functioning, revealed moderately stable ICCs for amplitude but a lack of reliability over time for latency measures (Weinberg and Hajcak, 2011). Finally, in a paradigm similar to the one reported here, an analysis of the P3 No-Go waveform produced during a visual Go/No-Go task, the authors found good (ICC > 0.75) test-retest reliability for amplitude and excellent (ICC > 0.90) agreement in latency measures over 6–18 months (Brunner et al., 2013).

While ERP studies conducted with the minimal behavior approach have provided invaluable information on cognitive processing, more recently there has been a significant push to develop and refine electro-cortical measurements while participants engage in real-world behaviors (De Vos and Debener, 2014; Debener et al., 2015; Gramann et al., 2010; Gwin et al., 2010; Wagner et al., 2016; Wagner et al., 2012). This approach, termed Mobile Brain/Body Imaging (MoBI), involves the continuous monitoring of cortical activity with lightweight EEG recording systems synchronized with measures of participants' sensory experiences and body motion tracking (Gramann et al., 2014b; Makeig et al., 2009). In this way, the brain dynamics underlying many natural behaviors, such as walking and performing a cognitive task, can be assessed in concert with the behaviors themselves. Studies from other groups (Gramann et al., 2010) as well as our own (De Sanctis et al., 2012) have demonstrated the feasibility of obtaining electrophysiological measures with MoBI. We have previously assessed cognitive and gait performance at varying levels of walking speed (De Sanctis et al., 2014) and characterized differences in behavior, gait and ERPs during dual-task walking that are associated with aging (Malcolm et al., 2015).

Although portable, light-weight EEG-based MoBI systems represent a new frontier for neural investigations of human gross motor behavior, the reliability of such recordings has not been comprehensively assessed over an extended time frame. One study applied advanced signal processing, independent component analysis, to remove gait-related artifacts and assess the signal quality of ERP recordings (Gramann et al.,

2010). They found that the N1 and P3 ERPs for a visual oddball task did not differ between standing, slow walking and fast walking conditions. Component amplitudes were comparable in all three movement conditions as evidenced by significant ICCs of 0.603 for N1 and 0.628 for P3, indicating that speed did not affect the amplitude of early or late visual ERPs (Gramann et al., 2010). More recently, in an effort to compare the suitability and functionality of wireless wet and dry mobile EEG systems, Oliveira et al. (2016) calculated the reliability of pre-stimulus noise, signal-to-noise ratio and P300 amplitude variance for sitting and walking conditions across a time interval of 7–20 days (Oliveira et al., 2016), reporting moderate to good reliability for these variables.

Here, our goal was to evaluate the long-term test-retest reliability of ERPs recorded during ambulation. Participants walked on a treadmill, maintaining a speed of 5.1 km/h, over two testing sessions, separated by an average of 2.3 years. We chose to focus our analysis on a well-established neurophysiological correlate of cognitive control, the N2/P3 ERP complex evoked during a Go/No-Go task. To evaluate the relative consistency of walking ERPs in comparison to those obtained during a traditional seated experiment, we also computed test-retest reliability measures (ICC) for sitting ERPs recorded during the same sessions.

2. Results

2.1. Behavioral results

Participants were asked to perform a speeded, visual Go/No-Go task, responding quickly and accurately to every stimulus presentation by clicking a computer mouse button, while withholding responses to the second instance of any stimulus repeated twice in a row. The probability of Go and No-Go trials was 0.85 and 0.15, respectively. A trial was defined as a Hit if a response was recorded during the time interval 150–800 ms following a Go stimulus. A trial was defined as a Correct Rejection (CR) if no response was registered during the entire time period following a No-Go stimulus, and the previous trial had been classified as a Hit. Incorrect trials including misses (Go trials in which participants failed to respond) and false alarms (No-Go trials which were followed by an incorrect response) were excluded from the analysis. Behavioral results across testing sessions for both sitting and walking conditions are summarized in Table 1. Means, standard deviations and the results of paired comparisons between sessions are shown for the number of artifact-free trials and behavioral performance for Hits and CRs. There were no significant differences between the two recordings for any of the behavioral parameters. Whereas non-significant p values do not provide evidence in favor the null hypothesis, Table 1 also includes JZS Bayes factors for each comparison indicating if there is substantial evidence in support of the null hypothesis, the alternative hypothesis or neither (Dienes, 2014). Three comparisons across testing sessions resulted in JZS values > 3: trial numbers for Hits in the sitting condition, and reaction times for both sitting and walking, signifying that for these comparisons the null hypothesis (i.e., no difference in task performance) is three times more likely than the alternative hypothesis, i.e., change in performance across sessions (Rouder et al., 2009). All other comparisons revealed JZS Bayes factors ranging from approximately 1.1 up to 3.0, indicating a lack of evidence in support of either hypothesis.

2.2. Event-related potential results

Fig. 1 shows the mean and standard error of the mean for Hits and CR waveforms averaged across all 12 participants for both testing sessions. ERPs are displayed over three scalp regions: frontocentral, central and centroparietal. For comparison, ERPs recorded during treadmill walking are displayed adjacent to those recorded during sitting conditions. Not only do the ERPs exhibit a remarkably similar pattern between walking and sitting conditions, but the electrophysiological re-

Table 1

Means (standard deviations), paired *t*-test and JZS Bayes factor comparisons for the number of accepted trials, accuracy performance (percent correct) and reaction times between the two recording sessions.

Number of Trials	Session 1	Session 2	Session 1 vs. Session 2	
Hits Sitting	449 (9)	450 (12)	$p = 0.85$	JZS BF = 3.425
Hits Walking	788 (248)*	697 (44)	$p = 0.21$	JZS BF = 1.682
CRs Sitting	43 (12)	49 (9)	$p = 0.13$	JZS BF = 1.244
CRs Walking	76 (29)*	80 (17)	$p = 0.55$	JZS BF = 2.965
% Correct				
Hits Sitting	96.9 (10)	99.9 (17)	$p = 0.32$	JZS BF = 2.217
Hits Walking	98.6 (3.8)	99.8 (17)	$p = 0.29$	JZS BF = 2.089
CRs Sitting	56.7 (13.2)	66.0 (11.3)	$p = 0.07$	JZS BF = 1.259
CRs Walking	59.6 (12.8)	67.4 (11.1)	$p = 0.08$	JZS BF = 1.123
Reaction Time (ms)				
Hits Sitting	361 (48)	359 (31)	$p = 0.93$	JZS BF = 3.467
Hits Walking	383 (51)	378 (30)	$p = 0.63$	JZS BF = 3.123

Note: CRs = correct rejections.

* The large variability in the number of walking trials during the first session can be attributed to the variable number of blocks performed by different individuals.

sponses during both execution (Hits) and inhibition (CRs) of the cognitive task reveal, for the most part, a high degree of stability over an extended time frame. As an example of the relative degree of between-subject variability (amongst the 12 participants) compared to within-subject variability (between the two testing sessions), single-subject waveforms are shown in Fig. 2 for one walking condition, CRs averaged over central electrode sites.

2.2.1. N2 component

Between-session Intraclass Correlation Coefficients (ICC) for mean amplitude and peak latency measures are presented in Table 2. Interestingly, the N2 tended to exhibit a higher degree of test-retest reliability during walking conditions compared to sitting, for both Hits and CRs. This was especially true for measures of mean amplitude on Hit trials while walking, demonstrating a strong level of agreement between sessions with all ICCs > 0.74 ($ps < 0.005$), and to a slightly lesser extent CR amplitudes, showing an adequate level of consistency over central (ICC = 0.73, $p < 0.05$) and centroparietal scalp (ICC = 0.69, $p < 0.01$). Compared to amplitude measures, ICCs for peak latency revealed less discrepancy between sitting and walking conditions, however walking conditions still showed higher levels of reliability overall. For example, latency for Hits produced excellent measures of consistency between walking sessions (ICC = 0.93, $p < 0.001$) but there was less agreement, though still a good level of reliability for sitting (ICC = 0.81, $p < 0.001$). Averaging across scalp sites, CRs also showed a greater level of reliability in peak latency between the two walking sessions (average ICC = 0.66, all $ps < 0.01$) compared to the slightly less robust level of reliability between sitting sessions (average ICC = 0.62, all $ps < 0.05$). Overall, for both the N2 and P3 components, correlation coefficients for Hit waveforms tended to be more stable compared to CRs.

2.2.2. P3 component

ICCs for mean amplitude and peak latency of the P3 component are shown in Table 3. Generally, test-retest reliability was weaker for the P3 ERP compared to that of the N2, with the exception of a handful of measures assessed when participants performed the task while sitting, notably CR amplitude over frontocentral areas showed a good level of stability with ICC = 0.79, $p < 0.005$. The P3 also exhibited a different pattern of results in that, compared to the N2, only a few measurements revealed a higher degree of reliability for walking conditions compared to sitting, including Hit amplitude over centroparietal scalp

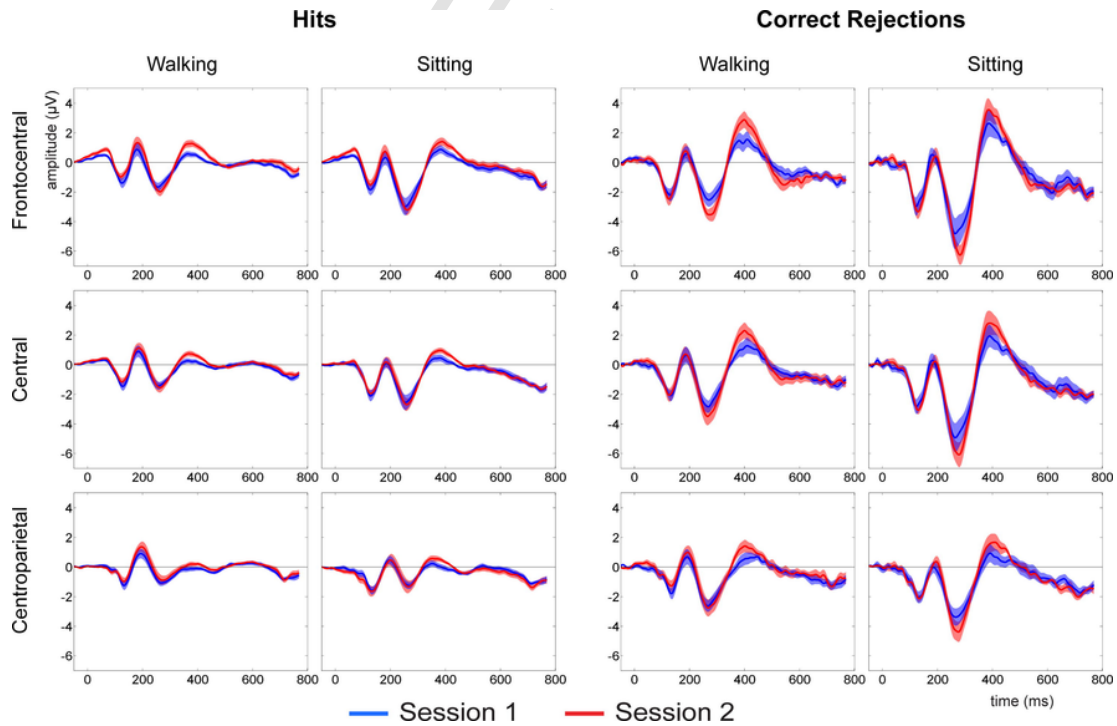


Fig. 1. Grand mean ERPs for Hit and Correct Rejection (CRs) trials during walking and sitting conditions, for Session 1 (blue waveforms) and Session 2 (red waveforms). ERPs are plotted over three scalp regions: frontocentral (top row), central (middle) and centroparietal (bottom row). Stimulus onset occurs at 0 ms. Shading represents standard error. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

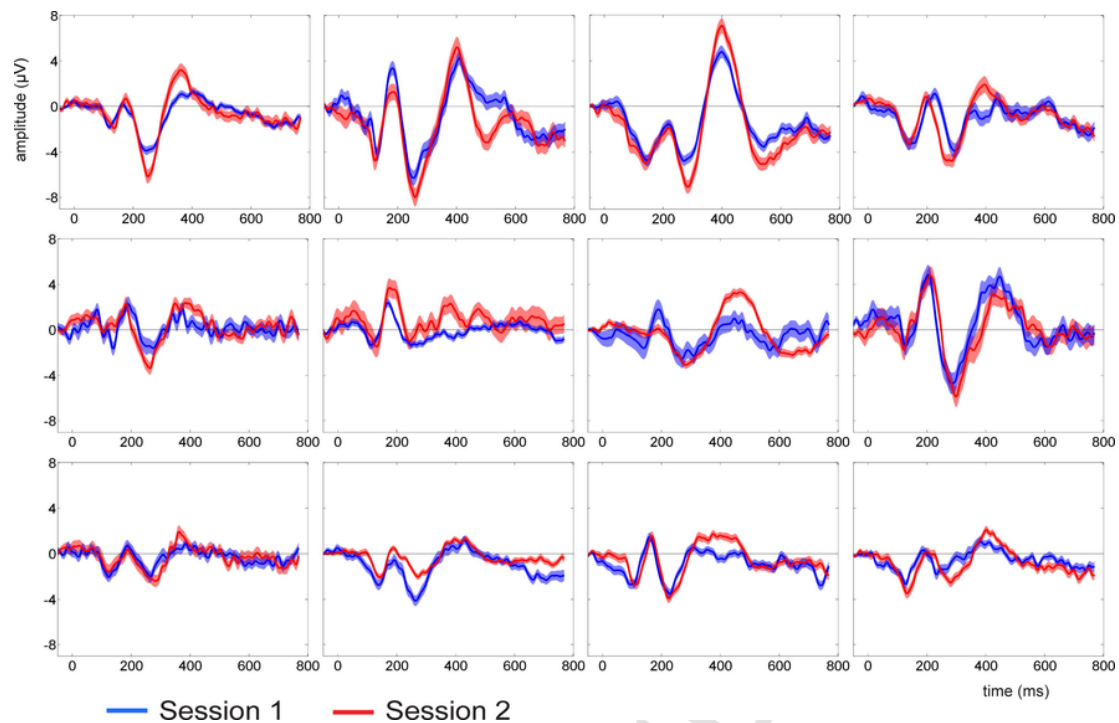


Fig. 2. Correct Rejection waveforms for all individual participants ($n = 12$) recorded during treadmill walking and averaged over central electrode sites C1, Cz and C2. Session 1 is shown in blue and Session 2 in red. Shading represents standard error. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2

Between-session intraclass correlation coefficients (ICC) for mean amplitude and peak latency of the N2 component. ICC values are shown for Hit and Correct Rejection (CR) trials, for either sitting or walking conditions, and over three scalp locations.

Hits	Frontocentral	Central	Centroparietal
Amplitude Sitting	0.64*	0.60*	0.14
Amplitude Walking	0.84****	0.81***	0.74***
Latency Sitting	0.64*	0.78***	0.81****
Latency Walking	0.78***	0.64**	0.93****
Correct Rejections			
Amplitude Sitting	0.19	0.62**	0.42
Amplitude Walking	0.28	0.73***	0.69**
Latency Sitting	0.71***	0.59*	0.55*
Latency Walking	0.67**	0.67***	0.63**

Frontocentral: FC1, FCz, FC2; Central: C1, Cz, C2; Centroparietal: CP1, CPz, CP2.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$, **** $p < 0.001$.

Table 3

Between-session intraclass correlation coefficients (ICC) for mean amplitude and peak latency of the P3 component. ICC values are shown for Hit and Correct Rejection (CR) trials, for either sitting or walking conditions, and over three scalp locations.

Hits	Frontocentral	Central	Centroparietal
Amplitude Sitting	0.42*	0.46*	0.17
Amplitude Walking	0.32*	0.46*	0.60**
Latency Sitting	0.79****	0.60*	0.85****
Latency Walking	0.53*	0.67**	0.80***
Correct Rejections			
Amplitude Sitting	0.79***	0.76****	0.63***
Amplitude Walking	0.18	0.37	0.46*
Latency Sitting	0.22	0.49*	0.51*
Latency Walking	0.56*	0.56*	0.15

Frontocentral: FC1, FCz, FC2; Central: C1, Cz, C2; Centroparietal: CP1, CPz, CP2.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$, **** $p < 0.001$.

(sitting ICC = 0.17, $p > 0.05$; walking ICC = 0.60, $p < 0.01$) and CR latency over frontocentral areas (sitting ICC = 0.22, $p > 0.05$; walking ICC = 0.56, $p < 0.05$). In fact, most P3 ICCs were stronger for sitting

conditions, with the greatest discrepancy between sitting and walking being CR amplitude at all scalp sites (all sitting ICCs > 0.63 (all $ps < 0.005$) and all walking ICCs in the poor range < 0.46), and CR latency at centroparietal sites (sitting ICC = 0.51, $p < 0.05$; walking ICC = 0.15, $p > 0.05$). The weakest measures of consistency for the P3 component were Hit amplitude and CR latency, for both sitting and walking conditions.

3. Discussion

In recent years, innovative studies utilizing the Mobile Brain/Body Imaging (MoBI) approach by recording electro-cortical responses in actively behaving participants have made substantial contributions to further our understanding of brain-behavior interactions in real-life circumstances (De Vos and Debener, 2014; Debener et al., 2012; Gramann et al., 2014a; Petersen et al., 2012; Wagner et al., 2012; Wagner et al., 2014). Here, we assessed the intra-individual stability of ERPs across two recording sessions, evaluating the test-retest reliability of the N2/P3 componentry acquired via a novel MoBI approach while walking, in addition to the traditional so-called minimal behavior approach during sitting.

Our results indicated a strong level of reliability for the earlier N2 component, with most amplitude and latency measurements in the good to excellent range, as assessed by the Intraclass Correlation Coefficient. ICC measures for the later P3 were generally less robust but overall showed adequate to good levels of stability in both amplitude and latency. Remarkably, the N2 exhibited a much greater degree of consistency, in terms of both mean amplitude and peak latency, across the two walking sessions compared to the sitting condition. This was true for both Hits and CR trials. In contrast, the P3 waveform tended to have a higher degree of consistency for sitting ERPs compared to those recorded while walking. The finding of a greater degree of stability in the componentry of the earlier N2 ERP as compared to the later P3 has been reported for other longer latency components as well (e.g., P3a/b, Pe, P400) (Cassidy et al., 2012). This may be a result of earlier components, related to sensory and relatively automatic processing of stimuli,

consisting of a largely stereotyped response function, while later ‘cognitive’ components may be more susceptible to individual variation. Following this line of reasoning, it may also be the case that the potential for increased individual variability in component structure at later latencies is additionally enhanced during active walking. This may have been a factor in our observation that the N2 was overall more reliable for walking and the P3 for sitting.

Many factors are thought to play a role in the test-retest reliability of ERPs, including sample size, age and arousal level of participants (Brunner et al., 2013; Kinoshita et al., 1996) additionally, choices made during the analysis process with regards to averaging over various numbers of trials and electrode sites, as well as of course, the time interval between testing sessions (Huffmeijer et al., 2014). Longer time intervals may allow for significant physiological modulations, even amongst healthy participants, therefore it could very well be the case that for a reliability assessment conducted on a shorter time scale, ICC measures may improve further. Our results indicate that Hit trials tended to produce higher ICC values for both components assessed, compared to CR trials. This may have been due to the fact that Hit waveforms were comprised of many more trials (see Table 1). While a fewer number of trials may suffice for earlier components to reach an adequate level of reliability, it has been recommended that at least 50–60 trials be included for the evaluation of later, broadly distributed components such as the P3 No-Go signal (Brunner et al., 2013; Huffmeijer et al., 2014). In the current investigation, the number of CR trials accepted (following artifact rejection) fell into this range for the sitting condition, while the number of accepted trials for the walking-obtained CRs well exceeded these recommendations.

There are mixed results from the ERP reliability literature as to whether ERP amplitude or latency measures tend to be more consistent across time. This issue may obviously be impacted by the precise method used to calculate these measures, for example Cassidy et al., found that absolute peak amplitude was more reliable than mean amplitude for shorter latency components (Cassidy et al., 2012). Our results show generally more stable measures of peak latency across time compared to component mean amplitude (an exception being the more consistent P3 amplitude measures recorded for CRs in the sitting condition), most prominently for the P3 on Hit trials, but also to a lesser extent, the N2. Similar results have been reported elsewhere for the P3 (Brunner et al., 2013) and have attributed this outcome to latency measures tending to be less susceptible to individual variations in physiological activity (Brunner et al., 2013). However, others have reported weaker and more variable ICCs for P3 peak latency compared to mean amplitude, using a modified flanker task (Huffmeijer et al., 2014), an auditory oddball paradigm assessed over eight sessions (Kinoshita et al., 1996) and one study comparable to ours in terms of time between assessments (~2 years) in which ICC values for amplitude (either peak or area) ranged from 0.56 to 0.67 (moderately stable neural measures) but peak latency measures were much weaker and would not be considered reliable (0.29) (Weinberg and Hajcak, 2011).

Our findings have important implications as there is significant clinical interest in obtaining clean electrophysiological recordings during walking, and our data speaks to the quality of procedures to correct for mechanical and movement artifacts, an issue that remains a matter of vigorous debate (Gwin et al., 2010; Kline et al., 2015; Nathan and Contreras-Vidal, 2015). Alternatively, it might be argued that gait-related artifacts did not affect the ERP components of interest. In other words, averaging across Go/No-Go trials time-locked to the stimulus onset might have eliminated mechanical or movement artifacts as they are randomly distributed in relation to the event of interest. In previous studies we have characterized modulations in gait pattern and electro-cortical responses (N2/P3) specific to dual-task walking demands in young adults (De Sanctis et al., 2014), while in older adults increased load (i.e., walking while performing the cognitive task) resulted in behavioral performance costs but minimal changes to gait and to the neural correlates of inhibitory processing (Malcolm et al., 2015). These

findings were interpreted as an age-related decline in the ability to flexibly allocate attentional resources across multiple domains. Furthermore, other investigations have reported kinematic EEG signals coinciding with specific gait phases (Gwin et al., 2011; Presacco et al., 2012; Seeber et al., 2014; Wagner et al., 2016, 2012) as well as preceding imminent loss of balance control (Sipp et al., 2013). Studies are underway to determine the neural underpinnings of gait adaptation in an effort to probe motor deficits related to stroke, multiple sclerosis and Parkinson’s disease. Another crucial line of research for mobile EEG is the deployment of brain-computer interfaces (BCI) in rehabilitation therapies (Kranzioch et al., 2014). Demonstrating a sufficient level of consistency in the ERP measures obtained between testing sessions will add more weight to prior outcomes and advance the basic and translational utility of the MoBI technique.

In conclusion, this study characterized the test-retest reliability of two common ERP measures, the N2 and the P3, recorded during active walking conditions and for comparison, while seated. Intraclass correlation coefficients showed an adequate to excellent level of reproducibility over an average two-year period, indicating the potential for MoBI studies to access stable indices of noninvasive brain markers over time.

4. Methods and materials

4.1. Participants

Twelve healthy adults (four females, eight males) took part in the experiment across two separate sessions. The average time between the two recordings was 2.3 years (range: 1.1–2.9 years). All individuals reported normal or corrected-to-normal vision and were free from any neurological or psychiatric disorders. Participants were recruited from the lab’s existing subject pool and from flyers posted at the Albert Einstein College of Medicine. The mean age of the participants at the first recording was 24.2 years (SD = 3.0) while mean age at the second recording was 26.5 years (SD = 3.4). The Institutional Review Board of the Albert Einstein College of Medicine approved the experimental procedures and all participants provided their written informed consent. All procedures were compliant with the principles laid out in the Declaration of Helsinki for the responsible conduct of research.

4.2. Procedure

Participants performed a visual Go/No-Go response inhibition task while either seated or walking on a treadmill. Visual stimuli consisted of 168 affectively positive or neutral images from the International Affective Picture System (IAPS) (Lang et al., 2008). Images were projected centrally (InFocus XS1 DLP, 1024 × 768 pixel) onto a black wall approximately 1.5 m in front of the participant. Stimulus duration was 600 ms with a random stimulus-onset-asynchrony (SOA) ranging from 800 to 1000 ms. Participants were instructed to quickly and accurately click a wireless computer mouse button in response to the presentation of each image (Go trials), while withholding button presses to the second instance of any picture repeated twice in a row (No-Go trials). Probability of Go and No-Go trials was 0.85 and 0.15, respectively. Stimulus display was programmed with Presentation software version 14.4 (Neurobehavioral Systems, Albany, CA, USA). On average, images subtended 28° horizontally by 28° vertically.

The task was presented in blocks, each lasting approximately 4 min. During walking blocks all participants performed the experiment while walking at 5.1 km/h, approximating the average gait speed of young adults (Silva et al., 2014). During the first recording session subjects participated in a variable number of walking blocks ranging from four to ten (mean = 5.75 blocks) and three sitting blocks. In the second session all participants performed five walking blocks and three sitting blocks. All blocks were conducted in a pseudo-random order and a

practice block was performed before undertaking the main experiment. No specific task prioritization instructions (i.e., walking versus cognitive task) were given.

In addition to maintaining a constant walking speed, temporal parameters of the gait cycle were assessed to confirm that consistent gait patterns were upheld across sessions. Participants were equipped with three foot force sensors (Tekscan FlexiForce A201 transducers) on the sole of each foot to measure stride time and stride time variability (De Sanctis et al., 2014; Malcolm et al., 2015). No differences between testing sessions were observed for either mean stride time (Session 1: 1066 ms, Session 2: 1063 ms, $p = 0.74$) or stride time variability (Session 1: 45 ms, Session 2: 57 ms, $p = 0.46$).

4.3. Electrophysiological recording and analysis

Continuous EEG was recorded with a 72-channel BioSemi ActiveTwo system array (digitized at 512 Hz; 0.05–100 Hz pass-band). Pre-processing and analysis was performed using custom MATLAB scripts (MathWorks, Natick, MA) and EEGLAB (Delorme and Makeig, 2004). Individual participant data was subjected to an independent component analysis (ICA)-based artifact identification and removal procedure (Gramann et al., 2010; Jung et al., 2000). ICA is a technique that attempts to separate the multiple generators contributing to task-evoked ERPs, including neural as well as artifactual activity (Makeig et al., 1996). Previous literature has demonstrated the feasibility of employing an ICA approach to detect and remove artifacts while preserving data that would otherwise be entirely rejected (Delorme et al., 2007; Jung et al., 2000). In MoBI paradigms especially, ICA is beneficial to detect activity arising from blinks, neck muscles, cable sway, and the force of footfalls on the treadmill, which may otherwise obscure task-related neural signals.

EEG data was first high-pass filtered at 1.5 Hz using a zero phase FIR filter (order 5632) (Winkler et al., 2015). Then all blocks from each condition type (sitting or walking) were appended into one dataset (separately for each of the two recording sessions). Noisy channels were identified and removed by visual inspection and by automatic detection of channels with signals more than five times the standard deviation of the mean across all channels. The remaining channels were re-referenced to a common average reference. EEG was then epoched into 1000 ms intervals time-locked to image presentation, with 200 ms before stimulus onset and 800 ms following. Epochs were subjected to a manual visual inspection resulting in the rejection of any epoch that contained large or non-stereotypical artifacts. The aforementioned extended ICA decomposition was performed on the remaining epochs of interest using default training mode parameters (Makeig et al., 1996). Next, independent components (ICs) that appeared to exclusively represent non-brain or artifactual activity including blinks, line noise, walking artifacts, mechanical noise associated with cable sway and (especially neck) muscles were manually identified and rejected. Note that artifact rejection via ICs was performed for each condition (sitting, walking) and session (Run 1, Run 2) independently. Typically, these artifactual components were noticeably distinct from brain-derived activity based on their activation time course, scalp topography, spectra and single trial ERP epochs (Castermans et al., 2014; Kline et al., 2015; Onton and Makeig, 2006). After artifact-free EEG signals were back projected and summed onto the scalp electrodes (Jung et al., 2000), data was again visually inspected and any epochs with remaining artifacts were rejected. This resulted in the rejection of very few epochs, indicating that the IC rejection process was quite effective. Lastly, previously discarded data channels were replaced using spherical interpolation, and a zero phase low-pass butterworth filter of 45 Hz (24 dB/octave) was applied.

4.4. Event-related potential analysis

Following preprocessing, 1000 ms ERPs with a 50 ms pre-stimulus baseline were computed for the two conditions of interest: Hits (Go trials followed by a correct response) and Correct Rejections (No-Go trials in which a response was correctly withheld) separately for the sitting and walking blocks. The time windows and electrode sites considered for N2 and P3 components were chosen a priori based on previous literature (Bokura et al., 2001; Donkers and van Boxtel, 2004; Eimer, 1993; Falkenstein et al., 2002; Nieuwenhuis et al., 2003) and confirmed by visual inspection of grand average waveforms. Thus, test-retest reliability of task-evoked N2/P3 focused on three regions along the midline, each composed of the averaged signal of three individual electrodes: fronto-central (FC1, FCz, FC2), central (C1, Cz, C2) and centroparietal (CP1, CPz, CP2). For each task condition and recording session, we used the grand average waveform's peak amplitude to encapsulate a 100 ms time window for the N2 and a 150 ms time window for the P3, which were then used to compute mean amplitude and detect peak latency across the respective time periods. These ERP features were then used to assess reliability across the two recording sessions.

4.5. Statistical analysis

Statistical analysis on all behavioral and ERP data was performed with IBM SPSS 21.0. Two-tailed paired-sample t-tests were computed to evaluate differences in behavioral performance between the two recording sessions. Additionally, a Bayes factor analysis was conducted to investigate evidence for the null hypothesis (that there is no difference between the first and second session) or the alternative hypothesis (that there is a difference between the sessions). The Bayes factor analysis is an alternative to a post hoc power analysis but has the benefit that it takes into account the sensitivity of the data to distinguish between the null and alternative hypothesis (Butler et al., 2016; Dienes, 2014, 2016). The Jeffreys, Zellner and Siow (JZS) Bayes factor was computed using the default effect size of 0.707 (Rouder et al., 2009). A JZS Bayes factor can be read such that a value greater than three favors the null hypothesis three times more than the alternative hypothesis, while a value less than one third favors the alternative three times more than the null, values between one third and three suggest that there is not enough evidence to favor either.

Test-retest reliability of mean amplitude and peak latency of the N2 and P3 ERP components was assessed with the ICC using a two-way mixed effect model with absolute agreement (Bartko, 1966; McGraw and Wong, 1996; Shrout and Fleiss, 1979). ICC is the most appropriate statistic to evaluate agreement in measurements over time, whereas a procedure such as Pearson's product moment correlation (r) provides only a measure of association (Bartko, 1991). ICCs were computed for single values of amplitude and latency and interpreted in terms of consistency. The higher the ICC value, the more stable the signal is over time. ICC values below 0.50 are generally considered a poor level of reliability, moderate reliability ranges from 0.50 to 0.75, good from 0.75 to 0.90 and excellent when higher than 0.90 (Portney and Watkins, 2009).

In addition, in response to a reviewer's suggestion we also performed a time-frequency analysis of single trial data over a central electrode site (Cz), in order to examine changes in power (event-related spectral perturbations, ERSPs) and inter-trial coherence (ITC) across sessions. ITC represents the consistency of the phase of the evoked response and functions as a measure of inter-trial reliability (Butler et al., 2016; Delorme and Makeig, 2004). These data are reported as Supplemental Materials and also lend support to the stability of the evoked response signal over an extended time frame.

Acknowledgments

The primary source of funding for this work was provided by a pilot grant from the Einstein-Montefiore Institute for Clinical and Translational Research (UL1-TR000086) and the Sheryl & Daniel R. Tishman Charitable Foundation. Participant recruitment and scheduling were performed by the Human Clinical Phenotyping Core at Einstein, a facility of the Rose F. Kennedy Intellectual and Developmental Disabilities Research Center (RFK-IDDRC) which is funded by a center grant from the Eunice Kennedy Shriver National Institute of Child Health & Human Development (NICHD P30 HD071593). We would like to express our sincere gratitude to the participants for giving their time to this effort.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.brainres.2017.05.021>.

References

- Bartko, J.J., 1966. The intraclass correlation coefficient as a measure of reliability. *Psychol. Rep.* 19 (1), 3–11. <http://dx.doi.org/10.2466/pr0.1966.19.1.3>.
- Bartko, J.J., 1991. Measurement and reliability: statistical thinking considerations. *Schizophr. Bull.* 17 (3), 483–489.
- Bokura, H., Yamaguchi, S., Kobayashi, S., 2001. Electrophysiological correlates for response inhibition in a Go/NoGo task. *Clin. Neurophysiol.* 112 (12), 2224–2232.
- Brunner, J.F., Hansen, T.L., Olsen, A., Skandsen, T., Haberg, A., Kropotov, J., 2013. Long-term test-retest reliability of the P3 NoGo wave and two independent components decomposed from the P3 NoGo wave in a visual Go/NoGo task. *Int. J. Psychophysiol.* 89 (1), 106–114. <http://dx.doi.org/10.1016/j.ijpsycho.2013.06.005>.
- Butler, J.S., Molholm, S., Andrade, G.N., Foxe, J.J., 2016. An examination of the neural unreliability thesis of autism. *Cereb. Cortex* <http://dx.doi.org/10.1093/cercor/bhw375>.
- Cassidy, S.M., Robertson, L.H., O'Connell, R.G., 2012. Retest reliability of event-related potentials: evidence from a variety of paradigms. *Psychophysiology* 49 (5), 659–664. <http://dx.doi.org/10.1111/j.1469-8986.2011.01349.x>.
- Castermans, T., Duvinage, M., Cheron, G., Dutoit, T., 2014. About the cortical origin of the low-delta and high-gamma rhythms observed in EEG signals during treadmill walking. *Neurosci. Lett.* 561, 166–170. <http://dx.doi.org/10.1016/j.neulet.2013.12.059>.
- De Sanctis, P., Butler, J.S., Green, J.M., Snyder, A.C., Foxe, J.J., 2012. Mobile brain/body imaging (MoBI): high-density electrical mapping of inhibitory processes during walking. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2012, 1542–1545. <http://dx.doi.org/10.1109/EMBC.2012.6346236>.
- De Sanctis, P., Butler, J.S., Malcolm, B.R., Foxe, J.J., 2014. Recalibration of inhibitory control systems during walking-related dual-task interference: a mobile brain/body imaging (MoBI) study. *Neuroimage* 94, 55–64. <http://dx.doi.org/10.1016/j.neuroimage.2014.03.016>.
- De Vos, M., Debener, S., 2014. Mobile EEG: towards brain activity monitoring during natural action and cognition. *Int. J. Psychophysiol.* 91 (1), 1–2. <http://dx.doi.org/10.1016/j.ijpsycho.2013.10.008>.
- Debener, S., Emkes, R., De Vos, M., Bleichner, M., 2015. Unobtrusive ambulatory EEG using a smartphone and flexible printed electrodes around the ear. *Sci. Rep.* 5, 16743. <http://dx.doi.org/10.1038/srep16743>.
- Debener, S., Minow, F., Emkes, R., Gandras, K., de Vos, M., 2012. How about taking a low-cost, small, and wireless EEG for a walk? *Psychophysiology* 49 (11), 1617–1621. <http://dx.doi.org/10.1111/j.1469-8986.2012.01471.x>.
- Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134 (1), 9–21. <http://dx.doi.org/10.1016/j.jneumeth.2003.10.009>.
- Delorme, A., Sejnowski, T., Makeig, S., 2007. Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis. *Neuroimage* 34 (4), 1443–1449. <http://dx.doi.org/10.1016/j.neuroimage.2006.11.004>.
- Dienes, Z., 2014. Using Bayes to get the most out of non-significant results. *Front. Psychol.* 5, 781. <http://dx.doi.org/10.3389/psyg.2014.00781>.
- Dienes, Z., 2016. How Bayes factors change scientific practice. *J. Math. Psychol.* 72, 78–89. <http://dx.doi.org/10.1016/j.jmp.2015.10.003>.
- Donkers, F.C., van Boxtel, G.J., 2004. The N2 in go/no-go tasks reflects conflict monitoring not response inhibition. *Brain Cogn.* 56 (2), 165–176. <http://dx.doi.org/10.1016/j.bandc.2004.04.005>.
- Duncan, C.C., Barry, R.J., Connolly, J.F., Fischer, C., Michie, P.T., Naatanen, R., Van Petten, C., 2009. Event-related potentials in clinical research: guidelines for eliciting, recording, and quantifying mismatch negativity, P300, and N400. *Clin. Neurophysiol.* 120 (11), 1883–1908. <http://dx.doi.org/10.1016/j.clinph.2009.07.045>.
- Eimer, M., 1993. Effects of attention and stimulus probability on ERPs in a Go/Nogo task. *Biol. Psychol.* 35 (2), 123–138.
- Falkenstein, M., Hoormann, J., Hohnsbein, J., 2002. Inhibition-related ERP components: variation with modality, age, and time-on-task. *J. Psychophysiol.* 16 (3), 167–175. <http://dx.doi.org/10.1027//0269-8803.16.3.167>.
- Fallgatter, A.J., Aranda, D.R., Bartsch, A.J., Herrmann, M.J., 2002. Long-term reliability of electrophysiological response control parameters. *J. Clin. Neurophysiol.* 19 (1), 61–66.
- Fallgatter, A.J., Bartsch, A.J., Herrmann, M.J., 2002. Electrophysiological measurements of anterior cingulate function. *J. Neural Transm. (Vienna)* 109 (5–6), 977–988. <http://dx.doi.org/10.1007/s007020200080>.
- Fallgatter, A.J., Bartsch, A.J., Strik, W.K., Mueller, T.J., Eisenack, S.S., Neuhauser, B., et al., 2001. Test-retest reliability of electrophysiological parameters related to cognitive motor control. *Clin. Neurophysiol.* 112 (1), 198–204. [http://dx.doi.org/10.1016/S1388-2457\(00\)00505-8](http://dx.doi.org/10.1016/S1388-2457(00)00505-8).
- Foxe, J.J., Simpson, G.V., 2002. Flow of activation from V1 to frontal cortex in humans. A framework for defining “early” visual processing. *Exp. Brain Res.* 142 (1), 139–150. <http://dx.doi.org/10.1007/s00221-001-0906-7>.
- Gramann, K., Ferris, D.P., Gwin, J., Makeig, S., 2014. Imaging natural cognition in action. *Int. J. Psychophysiol.* 91 (1), 22–29. <http://dx.doi.org/10.1016/j.ijpsycho.2013.09.003>.
- Gramann, K., Gwin, J.T., Bigdely-Shamlo, N., Ferris, D.P., Makeig, S., 2010. Visual evoked responses during standing and walking. *Front. Hum. Neurosci.* 4, 202. <http://dx.doi.org/10.3389/fnhum.2010.00202>.
- Gramann, K., Jung, T.P., Ferris, D.P., Lin, C.T., Makeig, S., 2014. Toward a new cognitive neuroscience: modeling natural brain dynamics. *Front. Hum. Neurosci.* 8, 444. <http://dx.doi.org/10.3389/fnhum.2014.00444>.
- Gwin, J.T., Gramann, K., Makeig, S., Ferris, D.P., 2010. Removal of movement artifact from high-density EEG recorded during walking and running. *J. Neurophysiol.* 103 (6), 3526–3534. <http://dx.doi.org/10.1152/jn.00105.2010>.
- Gwin, J.T., Gramann, K., Makeig, S., Ferris, D.P., 2011. Electro-cortical activity is coupled to gait cycle phase during treadmill walking. *Neuroimage* 54 (2), 1289–1296. <http://dx.doi.org/10.1016/j.neuroimage.2010.08.066>.
- Huffmeijer, R., Bakermans-Kranenburg, M.J., Alink, L.R., van Ijzendoorn, M.H., 2014. Reliability of event-related potentials: the influence of number of trials and electrodes. *Physiol. Behav.* 130, 13–22. <http://dx.doi.org/10.1016/j.physbeh.2014.03.008>.
- Jung, T.P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., Sejnowski, T.J., 2000. Removal of eye activity artifacts from visual event-related potentials in normal and clinical subjects. *Clin. Neurophysiol.* 111 (10), 1745–1758. [http://dx.doi.org/10.1016/S1388-2457\(00\)00386-2](http://dx.doi.org/10.1016/S1388-2457(00)00386-2).
- Kinoshita, S., Inoue, M., Maeda, H., Nakamura, J., Morita, K., 1996. Long-term patterns of change in ERPs across repeated measurements. *Physiol. Behav.* 60 (4), 1087–1092. [http://dx.doi.org/10.1016/0031-9384\(96\)00130-8](http://dx.doi.org/10.1016/0031-9384(96)00130-8).
- Kline, J.E., Huang, H.J., Snyder, D.L., Ferris, D.P., 2015. Isolating gait-related movement artifacts in electroencephalography during human walking. *J. Neural Eng.* 12 (4), 046022. <http://dx.doi.org/10.1088/1741-2560/12/4/046022>.
- Kok, A., 1997. Event-related-potential (ERP) reflections of mental resources: a review and synthesis. *Biol. Psychol.* 45 (1–3), 19–56.
- Kranzich, C., Zich, C., Schierholz, I., Sterr, A., 2014. Mobile EEG and its potential to promote the theory and application of imagery-based motor rehabilitation. *Int. J. Psychophysiol.* 91 (1), 10–15. <http://dx.doi.org/10.1016/j.ijpsycho.2013.10.004>.
- Lang, P.J., Bradley, M.M., Cuthbert, B.N., 2008. International affective picture system (IAPS): affective ratings of pictures and instructional manual. Technical report A-8. University of Florida, Gainesville, FL, 2008.
- Makeig, S., Bell, A.J., Jung, T.P., Sejnowski, T.J., 1996. Independent component analysis of electroencephalographic data. *Adv. Neural Inf. Process. Syst.* 8, 145–151.
- Makeig, S., Gramann, K., Jung, T.P., Sejnowski, T.J., Poizner, H., 2009. Linking brain, mind and behavior. *Int. J. Psychophysiol.* 73 (2), 95–100. <http://dx.doi.org/10.1016/j.ijpsycho.2008.11.008>.
- Malcolm, B.R., Foxe, J.J., Butler, J.S., De Sanctis, P., 2015. The aging brain shows less flexible reallocation of cognitive resources during dual-task walking: a mobile brain/body imaging (MoBI) study. *Neuroimage* 117, 230–242. <http://dx.doi.org/10.1016/j.neuroimage.2015.05.028>.
- McGraw, K.O., Wong, S.P., 1996. Forming inferences about some intraclass correlation coefficients. *Psychol. Methods* 1 (1), 30–46. <http://dx.doi.org/10.1037/1082-989x.1.4.390>.
- Nathan, K., Contreras-Vidal, J.L., 2015. Negligible motion artifacts in scalp electroencephalography (EEG) during treadmill walking. *Front. Hum. Neurosci.* 9, 708. <http://dx.doi.org/10.3389/fnhum.2015.00708>.
- Nieuwenhuis, S., Yeung, N., van den Wildenberg, W., Ridderinkhof, K.R., 2003. Electrophysiological correlates of anterior cingulate function in a go/no-go task: effects of response conflict and trial type frequency. *Cogn. Affect Behav. Neurosci.* 3 (1), 17–26.
- Oliveira, A.S., Schlink, B.R., Hairston, W.D., Konig, P., Ferris, D.P., 2016. Proposing metrics for benchmarking novel EEG technologies towards real-world measurements. *Front. Hum. Neurosci.* 10, 188. <http://dx.doi.org/10.3389/fnhum.2016.00188>.
- Onton, J., Makeig, S., 2006. Information-based modeling of event-related brain dynamics. *Prog. Brain Res.* 159, 99–120. [http://dx.doi.org/10.1016/S0079-6123\(06\)59007-7](http://dx.doi.org/10.1016/S0079-6123(06)59007-7).
- Petersen, T.H., Willerslev-Olsen, M., Conway, B.A., Nielsen, J.B., 2012. The motor cortex drives the muscles during walking in human subjects. *J. Physiol.* 590 (Pt 10), 2443–2452. <http://dx.doi.org/10.1113/jphysiol.2012.227397>.
- Portney, L.G., Watkins, M.P., 2009. Foundations of Clinical Research: Applications to Practice, 3rd ed. Pearson/Prentice Hall, Upper Saddle River, N.J.
- Presacco, A., Forrester, L.W., Contreras-Vidal, J.L., 2012. Decoding intra-limb and inter-limb kinematics during treadmill walking from scalp electroencephalographic (EEG) signals. *IEEE Trans. Neural Syst. Rehabil. Eng.* 20 (2), 212–219. <http://dx.doi.org/10.1109/TNSRE.2012.2188304>.

- Rouder, J.N., Speckman, P.L., Sun, D., Morey, R.D., Iverson, G., 2009. Bayesian t tests for accepting and rejecting the null hypothesis. *Psychon. Bull. Rev.* 16 (2), 225–237. <http://dx.doi.org/10.3758/PBR.16.2.225>.
- Seeber, M., Scherer, R., Wagner, J., Solis-Escalante, T., Muller-Putz, G.R., 2014. EEG beta suppression and low gamma modulation are different elements of human upright walking. *Front. Hum. Neurosci.* 8, 485. <http://dx.doi.org/10.3389/fnhum.2014.00485>.
- Shrout, P.E., Fleiss, J.L., 1979. Intraclass correlations: uses in assessing rater reliability. *Psychol. Bull.* 86 (2), 420–428.
- Silva, A.M.C.B., da Cunha, J.R.R., da Silva, J.P.C., 2014. Estimation of pedestrian walking speeds on footways. *Proc. Inst. Civ. Eng.-Municipal Eng.* 167 (1), 32–43. <http://dx.doi.org/10.1680/muen.12.00048>.
- Sipp, A.R., Gwin, J.T., Makeig, S., Ferris, D.P., 2013. Loss of balance during balance beam walking elicits a multifocal theta band electrocortical response. *J. Neurophysiol.* 110 (9), 2050–2060. <http://dx.doi.org/10.1152/jn.00744.2012>.
- Wagner, J., Makeig, S., Gola, M., Neuper, C., Muller-Putz, G., 2016. Distinct beta band oscillatory networks subserving motor and cognitive control during gait adaptation. *J. Neurosci.* 36 (7), 2212–2226. <http://dx.doi.org/10.1523/JNEUROSCI.3543-15.2016>.
- Wagner, J., Solis-Escalante, T., Grieshofer, P., Neuper, C., Muller-Putz, G., Scherer, R., 2012. Level of participation in robotic-assisted treadmill walking modulates midline sensorimotor EEG rhythms in able-bodied subjects. *Neuroimage* 63 (3), 1203–1211. <http://dx.doi.org/10.1016/j.neuroimage.2012.08.019>.
- Wagner, J., Solis-Escalante, T., Scherer, R., Neuper, C., Muller-Putz, G., 2014. It's how you get there: walking down a virtual alley activates premotor and parietal areas. *Front. Hum. Neurosci.* 8, 93. <http://dx.doi.org/10.3389/fnhum.2014.00093>.
- Weinberg, A., Hajcak, G., 2011. Longer term test-retest reliability of error-related brain activity. *Psychophysiology* 48 (10), 1420–1425. <http://dx.doi.org/10.1111/j.1469-8986.2011.01206.x>.
- Winkler, I., Debener, S., Muller, K.R., Tangermann, M., 2015. On the influence of high-pass filtering on ICA-based artifact reduction in EEG-ERP. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2015, 4101–4105. <http://dx.doi.org/10.1109/EMBC.2015.7319296>.