A Novel, Fast, Reliable and Data-Driven Method for Simultaneous Single-Trial Mining and Amplitude - Latency Estimation Based on Proximity Graphs and Network Analysis

Dimitriadis SI\textsuperscript{1-6*}, Brindley L\textsuperscript{2,3}, Evans L\textsuperscript{2,3}, Linden DE\textsuperscript{2,3,5,6}, Singh K\textsuperscript{1-3}

\textsuperscript{1}Cardiff University Brain Research Imaging Centre, School of Psychology, Cardiff University, Cardiff, United Kingdom

\textsuperscript{2}Neuroinformatics Group, Cardiff University Brain Research Imaging Centre, School of Psychology, Cardiff University, Cardiff, United Kingdom

\textsuperscript{3}Division of Psychological Medicine and Clinical Neurosciences, School of Medicine, Cardiff University, Cardiff, United Kingdom

\textsuperscript{4}School of Psychology, Cardiff University, Cardiff, United Kingdom

\textsuperscript{5}Neuroscience and Mental Health Research Institute, Cardiff University, Cardiff, United Kingdom

\textsuperscript{6}MRC Centre for Neuropsychiatric Genetics and Genomics, School of Medicine, Cardiff University, Cardiff, United Kingdom

\textbf{*Corresponding author:}

Dr. Dimitriadis Stavros
Research Fellow
Institute of Psychological Medicine and Clinical Neurosciences, Cardiff University School of Medicine, Cardiff, United Kingdom

Email: stidimitriadis@gmail.com; DimitriadisS@cardiff.ac.uk
Abstract:

Both amplitude and latency of single-trial EEG/MEG recordings provide valuable information regarding functionality of the human brain. In this article, we provided a data-driven graph and network-based framework for mining information from multi-trial event-related brain recordings. On the first part, we provide the general outline of the proposed methodological approach. In the second part, we provide a more detailed illustration, present the obtained results on every step of the algorithmic procedure. To justify the proposed framework instead of presenting the analytic data mining and graph-based steps, we address the problem of response variability, a perquisite to reliable estimate both the amplitude and latency on specific N/P components linked to the nature of the stimuli.

The major question addressed on this study is the selection of representative single-trials with main scope to uncover a less noised averaged waveform elicited from the nature of the stimuli. This graph and network-based algorithmic procedure increases the signal-to-noise (SNR) of the brain response, a key pre-processing step to reveal significant and reliable amplitude and latency at specific time after the onset of the stimulus and with the right polarity (N or P). We demonstrated the whole approach using EEG auditory mismatch negativity (MMN) recordings from 42 young healthy controls. The method is novel, fast and data-driven succeeding first to reveal the true waveform elicited by MMN on different conditions (frequency, intensity, duration, etc).

The proposed graph-oriented algorithmic pipeline increased the SNR of the characteristic waveforms and the reliability of amplitude and latency within the adopted cohort. We also demonstrated how different EEG reference schemes (REST vs average) can influence amplitude-latency estimation. Simulations results revealed robust amplitude-latency estimations under different SNR and amplitude-latency variations with the proposed algorithm.

Keywords: single-trials, data-mining, proximity graphs, network analysis, amplitude, latency, reliability, signal to noise ration (SNR)
1. Introduction

A prerequisite for the studying of evoked potentials (EPs) is the distinction of the true brain’s response due to a stimulus from the brain ongoing activity. To uncover true brain activity, a large number of single trials (STs) should be collected and averaged to reveal the brain’s response waveform. The assumption that single-trials are time-locked and contaminated by Gaussian noise of zero-mean are both oversimplified (Laskaris et al., 2004). For example, the brain state of each subject changes from time moment to time moment due to shifts of attention and the fatigue level, while habituation during the task across experimental task and/or previous accidentally experience with the nature of the task are significant complemented factors that alter the dynamic behaviour even in short in duration recording responses (Laskaris et l., 2001,2003).

Both single-trial amplitude and latency of EEG/MEG signals contain valuable information regarding brain functionality in various conditions and targeted groups. For example, increased latency variation may be associated with ADHD (De Pascalis et al., 2008), ageing (Fein and Turetsky, 1989 ; Fjell, 2009), with IQ scores (Geurts et al., 2008), in mild cognitive impairment (MCI ;Laskaris et al., 2013) and in psychosis (Bodatsch et al., 2011).

Exploring single-trial differences between groups and/or conditions demands a proper unbiased manipulation of single-trials in order to extract reliable amplitude and latency. This is a non-trivial and demanding task for brain responses given the complexity of both the brain activity and the acquired EEG/MEG recordings due to low signal-to-noise ratio (SNR) of single-trial EEG/MEG responses (SNR, (Fein and Turetsky, 1989 ; Laskaris et al., 2001) and are usually integrated signals derived from multiple brain processes (Gevins,1985).

Only a few exploratory studies attempted to convey information from single trials. To deal with the poor SNR, different methods have been proposed in the past. The basic characteristic of previous techniques to solve the aforementioned issue based on classification and categorization of single-trials (Zouridakis et al., 1997a; Geva, 1998; Lange et al., 2000). The final outcome of this procedure is the categorization of single trials into homogeneous classes. Each of this class may reflect different brain behaviour like spontaneous reaction time, anticipation or reflecting the variability of the regional response dynamics (Laskaris et al., 2003). Complementary, Laskaris et al., proposed a summarization of single trials via Voronoi testellation procedure, minimal spanning tree and BFS procedure in order to reorder prototypical responses (Laskaris et al.,2004). The ordered prototypes reflected the variability of the single-trials while their source localization of neuromagnetic recordings with MFT algorithm succeeded to link this variability with the related sources on different brain areas and time windows. Complementary, they very first mentioned that the ongoing activity before the onset of the stimulus is functionally coupled with the subsequent regional response (Laskaris et al., 2001,2003). Recently, they
demonstrated how the ‘reflex level’ of spontaneous activity at various cognitive subsystems shape the brain activity during cognitive tasks stimulating the same subareas (Cole et al., 2016).

Several methods have been attempted to manipulate properly single trials with main scope to extract the related amplitudes and latencies. These methods can be categorized into two groups: the ones that need an a priori template and those with no waveform constraints. A few methods need the shape of the target signal which should be defined a priori (for example, Woody, 1967; Mayhew et al., 2006). Second, only a few methods allowed the free variability of single trials (for example, Pham et al., 1987; Laskaris et al., 2014), whereas others incorporate to the analytic pathway the constraint of both types of variation (for example, Jaskowski et al., 1999). Third, few methods assume that the data comes from a single signal (for example, Pham et al., 1987; Jaskowski et al., 1999) whereas others allow multi-trials to have their own amplitude and latency (for example, Laskaris et al., 2004; Mayhew et al., 2006; Da Pelo et al., 2018). The latter are on the right place solving the issue of consistent early responses compared to high variable late responses due to the onset of the stimulus. Fourth, others methods provided algorithms that are susceptible to noise (Jaskowski and Verleger, 2000), whereas in others this susceptibility is reduced by incorporating basis functions. Disadvantages of the aforementioned are either the a priori selection of a template waveform derived from grand-averaged time series (Hu et al., 2011) and/or the low performance in low SNR conditions.

In optimal scenarios, principal component analysis could be used for mining electroencephalographic and magnetoencephalographic responses (Friston et al., 1996). Alternatively, independent component analysis (ICA) could be adopted for dimensionality reduction and learning purposes of multi-trial responses (Makeig et al., 1996).

The purpose of the present paper is to demonstrate a fast, reliable and completely data-driven methodology based on data mining, graph and network analysis in order to reveal ‘true’ variability of the single-trials and accurate detection of amplitude and latency linked to responses on specific stimuli. It is more than evident that single-trials are noisy even in the most optimal scenarios and experimental protocols. The motivation of the presented algorithmic steps arose after analysing EEG single trials from the famous mismatch negativity auditory task (Näätänen et al., 2004). Single-trials were completely noisy missing even a clear peak across trials and the multi-feature paradigm. Our analysis combined a member of proximity graph called Gabriel graph (GG) and network analysis to reveal prototypical single-trials covering the whole space of their variability and then mixed them into a combined characteristic single-trial per condition. The proposed analysis is an appropriate tool for geometrical data and vectorial pattern analysis of single-trials.

The scope of our analysis on the adopted MMN paradigm for demonstration of the methodology focused on optimizing the SNR of the selected single-trials under the objective criterion to reveal the best type of filter (IIR/FIR), its order and the degree k of the GG single-trial network that choose the number of selected single-trials.
2. Method
2.1 Feature Extraction

Firstly, we construct a similarity distance matrix between every pair of single trial using the distance correlation estimator (Szekely and Rizzo, 2009). The distance matrix called hereafter DM tabulates the distance between the temporal variability of two time series. The distance correlation is a measure of statistical dependence between two random variables or random vectors.

2.2 Embedding in a Feature Space

After constructing the DM derived by the pair-wise estimation of the temporal variability of single trials, we embedded the DM in a 2D feature space. Here, we adopted multidimensional scaling (MDS) a high popular dimensionality technique among neuroscientists. This approach will help us to detect and visualize the variability of single trials within a common embedded feature space.

2.3 Constructing A Proximity Graph on The Embedded Space

A proximity graph is simply a graph in which two vertices are connected by an edge if and only if the vertices satisfy particular geometric requirements. “Proximity” here means spatial distance. Many of these graphs can be formulated with respect to many metrics, but the Euclidean metric is used most frequently. Here, we used the Euclidean distance as a proper spatial distance metric.

Let \( L(p,q) \) be the intersection of the circle about \( p \) with a radius of \( \text{dist}(p,q) \) and the circle about \( q \) with a radius of \( \text{dist}(q,p) \). This is called a lune. The relative neighborhood graph \( \text{RNG}(V) \) of a set of points \( V \), is the graph that has an edge \((p,q)\) if and only if the intersection of \( L(p,q) \) and \( V \) is empty (Fig. 1A).

Let \( C(p,q) \) be the circle centered on the point halfway between \( p \) and \( q \), and with a radius of half the distance between \( p \) and \( q \). The Gabriel graph of a set of points \( V \), \( \text{GG}(V) \), is the graph that has an edge \((p,q)\) if and only if the intersection of \( C(p,q) \) and \( V \) is empty (Fig. 1B).

In mathematics, the Gabriel graph of a set \( S \) of points in the Euclidean plane expresses one notion of proximity or nearness of those points. The Gabriel graph is a subgraph of the Delaunay triangulation (Matula & Sokal 1980). Complementary, the Gabriel graph contains as a subgraph the Euclidean minimum spanning tree, the relative neighborhood graph, and the nearest neighbor graph (Gabriel and Sokal, 1969; see Fig. 1)
**Figure 1. Proximity Graphs.** The relationship between relative neighbourhood graph (RNG), Gabriel graph (GG), Euclidean minimal spanning tree (EMST) and Delaunay triangulation (DT).

If we also consider the Euclidean minimum spanning tree (which is a tree that minimizes the total edge length connecting all points) and the Delaunay triangulation (which maximizes the minimum angle over all triangulations of a set of points), we get the following relationship:

\[ EMST \subseteq RNG \subseteq GG \subseteq DT \]  

Mathematically in GG, two points \( i \) and \( j \) are connected if the square of the distance between them, \( d_{ij}^2 \), is less than the sum of the squared distance between each of these points and any other point \( k \). Under GG main rule, we connect \( i \) and \( j \) if

\[ d_{ij}^2 < d_{ik}^2 + d_{jk}^2 \]  

for all \( k \) points. Schematically, we demonstrated Gabriel’s rule in Fig.2.

**Figure 2. Geometric illustration of Gabriel’s rule regarding the connection of two points.**

Fig.3 illustrates a right (A) and a wrong (B) connections under the main rule of GG. Fig.3C illustrates a GG produced by 100 random points in a 2D plain. The main outcome of this approach is the construction of a GG in the 2D feature space where nodes are the single-trials. Two single trials are
connected if within the circle passes from their 2D coordinates none other single trial is encapsulated. With this approach, GG notion demands to cover the feature space and to sample single-trial variability without over-representing. One can see GG as a denoising procedure to manipulate properly single-trials. GG captures the backbone of single-trials in the embedded space.

![Figure 3](image)

**Figure 3** A) Right, B) wrong connected points in GG and C) the GG of 100 random 2D points.

### 2.4 Network Analysis in Gabriel Graph (GG)

The construction of a connected GG on the 2D embedded space of single-trials opens the window to adopt well-known approached derived from network theory. Here, in order to detect representative prototypical single trials, we used the degree of each node in GG in order to detect the hubs. The degree of a node is a trivial network metric which describes the number of direct neighbours of each node. In simple words, degree counts the number of nodes with which each node is directly connected. Here, we optimized the selection of degree k using as an optimized objective criterion the increment of SNR of grand average single-trial.

### 2.5 Grand Average of Single Trials

After selecting the prototypical single trials that simultaneously capture the variability of single trials, we estimated their grand-averaged. The selection of prototypical single-trials based on the selection of degree k from the GG with main scope to improve the SNR.

### 2.6 Estimation of Amplitude – Latency

Based on the grand average signal constructed by averaging the prototypical single trials, we estimated the amplitude and latency. Both amplitude and latency were extracted completely data-driven by detecting prominent peak from the whole time series as a global maxima.
2.7 Estimating the Variability of Single-Trials

To access the variability of single trials, we estimated the global efficiency (GE) on a network level based on the subgraph defined by the nodes of GG linked to the extracted prototypical single-trials.

**Global efficiency** (GE) for a network W of N x N nodes is the inverse of the harmonic mean of the shortest path length between each pair of nodes and reflects the overall efficiency of parallel information transfer in the network (Latora and Marchiori, 2001).

\[
GE = \frac{1}{N} \sum_{i \neq N} \frac{1}{\sum_{j \neq N} (d_{ij})^{-1}}
\]  

(6)

Here as W, we used the GG while N denotes the prototypical single trials.

2.8 MATLAB Toolbox

A MATLAB toolbox will be released from the author’s website, researchgate and github upon acceptance of the paper (https://github.com/stdimitr/GG_SINGLE_TRIALS_MINING/tree/master). We will demonstrate the pipeline into a few recordings in order to be available to any researcher.

2.9 Influence of Reference on Waveform, Amplitude and Latency Estimation in Single-Trial Analysis

The influence of EEG reference is a critical issue for the electroencephalography (EEG) and event-related potentials (ERPs) studies. It seems that brain connectivity and network analysis is more robust compared to the estimation of single power (Dimitriadis et al., 2010). A recent study proposed the infinity REST reference as an appropriate common reference system for EEG analysis (Yao et al., 2001,2005; Qin and Yao,2010; Chella et al., 2016; Huang et al., 2017). Another study compared different EEG reference systems in different simulation scenarios in both sensor and source level. They demonstrated REST infinity reference is the most preferable system across the highly used reference systems in the literature (Lei and Liao,2017). Here, we adopted also REST reference system in comparison with the average system.

2.10 Alternative Single-Trial Mining Algorithms

To demonstrate the superiority and the simplicity of our method compared to others, we repeated the whole analysis using principal component analysis (PCA), singular value decomposition (SVD) and multi-linear regressor analysis (Hu et al., 2011). Both methodologies have been applied using average and REST reference.
2.11 Simulations

We simulated the original responses using multivariate autoregressive model (MVAR) (Anderson et al., 1998) and optimized the model selection with Akaike criterion (Akaike, 1974). The simulations followed two scenarios: In the first one, both amplitude and latency parameters were the same for the peak while in the second one both amplitude-latency varies independently apart from the peak. Practically, in the first scenario, the peaks were shifted and scales by the same amount while in the second within each trial the peak was shifted and scaled by different values.

We simulated 42 datasets (equals the number of the subjects) each consisting of 128 trials with 205 samples (400 ms). The peak varied over trials in amplitude (lognormal distributed with mean 1 and st.d. 1.2, restricted between low and high values of the empirical dataset), and latency (normally distributed with mean 0 and st.d. 150,170 or 170 ms). The simulation based on recordings derived from the FZ sensor at DIR-L condition and for deviant-standard stimulus.

All simulations were performed using MVAR for the estimated waveform under three signal-to-noise (SNR) conditions (SNR =0.5,1 and 2 ), using correlated noise. Noise was simulated using an AR(5) process with coefficients estimated from baseline trials of the empirical data.

3. Empirical Application in an Auditory MisMatch Negativity (MMN) Protocol

The proposed methodology is demonstrated in an auditory MMN multi-feature paradigm developed by Naatanen et al., 2004. The MMN peaks at about 100–300 ms after change onset but this latency varies slightly according to the specific paradigm or the type of regularity that is violated. According to the adopted protocol, MMN is usually evoked by a change of frequency (Low-High), direction (Low-High), intensity (Low - High), duration and gap, for both standard and deviant stimuli (Näätänen et al., 2004). Two standard tones preceded every deviant tone. Each condition was recorded in 128 trials while the protocol was designed such as to avoid any habituation of the sequence. The total number of trials was 8x128 for standard trials (averaged each pair of standard trials) and 8x128 for deviants.

We recorded a total of 42 subjects (23.75 ± 1.28, 24 females) using a recording BIOSEMI system of 64 channels (10-20 System ; Jasper, 1958). Additional electrodes were placed on the mastoid processes. The electrooculogram (EOG) was recorded from above and below the left eye (vertical (V)EOG) and from the outer canthi (horizontal (H)EOG). The electroencephalogram (EEG; range DC-419 Hz; sampling rate 2048 Hz) was acquired referenced to linked electrodes located midway between POz and PO3/PO4, respectively, and was re-referenced off-line to the average of the signal at the mastoids. Trials containing large EOG artefact were rejected, as were trials containing A/D saturation or baseline drift exceeding 80μV. Prior to any further analysis, we corrected the multichannel recordings.
from artifacts (muscle, blinks, cardiac) using ICA with EEGLAB (Delorme and Makeig, 2004; Dimitriadis et al., 2015, 2016a,b). Data were filtered off-line (0.5–45 Hz) and down-sampled to 512 Hz, resulting to an epoch of 400 ms after the onset of the stimulus or 205 samples. DC offset was removed by subtracting from each channel each low pass filtered component using FIR filter (Roll off 0.001 to 0.05 Hz, butter, (IIR), 6 dB attenuation in the stop band). Afterward, each trial was corrected with the baseline.

Our analysis focused on midline FCZ, FZ and CZ EEG sensors. We accessed the reliability of the proposed techniques in terms of amplitude, latency and signal power analysis.

All the subjects participated on this study were signed a written informed consent. The whole study has been approved by the ethical committee in School of Psychology in Cardiff University as part of a big multi-modal study.

Our analysis on the adopted MMN paradigm for demonstration of the methodology focused on optimizing the SNR of the selected single-trials under the objective criterion to reveal the best type of filter (IIR/FIR), its order and the degree k applied to GG single-trial network for the selection of the representative singe-trials was the SNR (see next section).

### 4. Results

#### 4.1 The Proposed Methodology in Simple Steps

We plot all the trials of a representative condition (high intensity - deviant) from a subject (Fig. 4A). Afterward, we estimated with distance correlation (Székely and Rizzo, 2014), the pair-wise associations of single-trials tabulated in a similarity matrix (Fig. 4B). The, we embedded this similarity matrix in a 2D space with multidimensional scaling (MDS) algorithm in order to visualize the variability of single-trials (Fig. 4C). Using the 2D points of the trials as an input to the Gabriel graph (GG), we constructed the GG demonstrated in Fig. 4D. Green lines represent the connected trials under the notion of GG. To sample the right representative single-trials, we estimated the degree k of each node in GG. We used two criterion to uncover the hubs on this GG based on the degree of each node: degree k = 1-4. In Fig. 4E,G, we demonstrated with red circles the selected trials/nodes in the GG. We selected the hubs nodes/trials based on their degree in the GG network on the assumption that these trials encapsulate the variability of the single-trials out sharing redundant information and also being on the core of variability avoiding the selection of outliers in the periphery of single-trials 2D projection.

Finally, we estimated the grand-averaged trial by averaging the selected single-trials with aforementioned network-based criterion with the objective criterion of improving the SNR of the selected single-trials. The resulting single-trial in both cases is presented in Fig. 4F,H. Based on the example in Fig. 4 where we used a Butterworth filter of order 1 (order since we used zero-phase filtering with filtfilt.m function of Matlab), the best result obtained with degree k >=4 where we selected 15
trials from the 128. We detected a negative peak around 150ms after the onset of the stimulus (Fig.6H). In the next section, we will demonstrate the effect of filter type, order and the selected degree in our network using as an objective criterion the improvement of signal-to-noise ratio across the selected single-trials with our methodology.

It is important to mention here that we applied the methodology independently for standard, deviant and deviant-standard. As a prominent characteristic peak, we revealed the dominant positive for deviant, standard and negative for deviant-standard after 100 ms of the onset of the stimulus.

4.2 The effect of FIR/IIR Filter Settings

In Fig.4-6, we demonstrated the steps of the proposed fast, reliable and data-driven methodology under a graph-based framework. We revealed that both the type of the filter (FIR/IIR) and its order can alter the characteristic waveform for each condition and subject.

We used the eegfilt matlab function provided in EEGLAB for FIR filtering of single-trials (Delorme and Makeig, 2004) and the butter MATLAB function for IIR filtering. We used a zero-phase filter in both cases applied on the concatenated trials separately for each stimulus (standard or deviant) and for each subject. The effect of filter with Butterworth bandpass filter can be seen on the representative time series in Fig.4H (order 1) vs Fig.5H (order 2). Order of one gave the best results for Butterworth bandpass IIR filter. In contrary, the best result for FIR filter using eegfilt function was obtained with order 2 (Fig.6H). In Fig.4H, the characteristic negative was detected around 100 ms after the onset of the stimulus while in Fig.6H, the negative peak located 150 ms after the onset of the stimulus. The effect of type and order of filter was demonstrated in high intensity condition for deviant stimulus from a single subject. Our analysis on the adopted MMN paradigm for demonstration of the methodology revealed as the best option based on SNR for filtering the FIR using eegfilt function and with order 2. The main objective criterion to reveal the best type of filter (IIR/FIR), order and degree k for the selection of the representative single-trials was the SNR (see next section).

4.3 Improvement of Signal-to-Noise Ratio with the proposed methodology

We evaluated the selection of the filter type, its order and the degree k in GG for the selection of single-trials independently for each condition, standard/deviant, subject and recording EEG sensors (FZ, FCZ, CZ). The parameters were filter type (FIR/IIR), order (1, 2, 3) and degree k (1-4). Finally, we estimated the SNR from the selected single-trials via the proposed methodology, adopting a proposed formula previously proposed (Laskaris et al., 2004). We scored to each of the 2x3x4=24 different sets of parameters across conditions (8 deviant and 8 stimuli = 16) and recording EEG sites (3 locations) the number of times where the SNR was maximum across the 24 sets. Formula 7 describes the objective
criterion for the selection of best settings for each subject across the 24 different combinations. Our results demonstrated clearly a maximization of SNR for every subject with FIR filter of order 2 and with most of the cases (39 out of 42) with degree k>=4.

\[
Score = \frac{\sum_{\text{filter}=1}^{2} \sum_{\text{order}=1}^{3} \sum_{\text{degree}=1}^{4} \arg \max \text{SNR (16 stimuli, 3 recording sites)}}{16 \text{ stimuli} \times 3 \text{ stimuli}} \tag{7}
\]

The Score was on group averaged 96.34 with standard deviation 2.31 with best choice for FIR filter (42 out of 42 subjects), order 2 (42 out of 42 subjects) and k=4 (39 out of 42 subjects and 3 with k=3). Table 2 summarizes the group-averaged SNR from each stimulus and EEG sensor location for standard stimuli. We presented results from FZ location where the majority of group-differences in terms of amplitude, latency, signal power and variability were more pronounced compared to FCG and CZ.

Results of SNR for the grand-averaged signal was < 1 and one can see in Fig.3-6, it is completely a bad strategy to estimate a peak for this noisy averaged trial (blue line in Fig4-6.F,H).

Table 2. Group-averaged SNR for each condition across the three selected EEG sensors for standard stimuli.

<table>
<thead>
<tr>
<th></th>
<th>Dir-L</th>
<th>Dir-R</th>
<th>Freq-Hi</th>
<th>Freq-Low</th>
<th>Int-Low</th>
<th>Int-High</th>
<th>Duration</th>
<th>Gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>FZ</td>
<td>6.14 ± 1.01</td>
<td>5.62 ± 0.78</td>
<td>6.23 ± 1.14</td>
<td>6.47 ± 1.07</td>
<td>6.68 ± 1.34</td>
<td>6.39 ± 0.77</td>
<td>6.71 ± 0.91</td>
<td>6.84 ± 1.43</td>
</tr>
<tr>
<td>FCZ</td>
<td>5.94 ± 1.13</td>
<td>5.67 ± 0.92</td>
<td>5.90 ± 0.98</td>
<td>6.07 ± 1.19</td>
<td>6.07 ± 1.31</td>
<td>5.78 ± 1.31</td>
<td>6.39 ± 1.14</td>
<td>6.16 ± 0.91</td>
</tr>
<tr>
<td>CZ</td>
<td>5.47 ± 0.87</td>
<td>5.44 ± 1.12</td>
<td>5.67 ± 1.10</td>
<td>5.61 ± 0.87</td>
<td>5.76 ± 1.45</td>
<td>5.45 ± 1.12</td>
<td>5.76 ± 1.13</td>
<td>5.93 ± 1.14</td>
</tr>
</tbody>
</table>
Figure 4: Butterworth IIR order 1: Outline of the proposed graph-based methodology (High intensity condition - deviant).

A. Plot of single trials from a single subject
B. Similarity matrix that tabulates the pair-wise associations of single-trials with distance correlation metric
C. Embedding the similarity matrix in B in a 2D space with multidimensional scaling (MDS) algorithm where each blue dot refers to a single-trial
D. Construct the Gabriel graph (GG) based on the 2D positions of the single trials. Green lines represent the connections under the notion of GG
E. Detection of hubs-representative single-trials based on their degree k in the GG. Here, we selected k=3 to detect the single trials representing with red circles.
F. Characteristic grand-averaged single-trial derived from the averaging of the selected hubs/single trials (red circles) in E. We selected 41 signals from 128 trials. Blue waveform denotes the grand-average from the whole set of trials while the black from the selected single-trials.
G. Detection of hubs-representative single-trials based on their degree k in the GG. Here, we selected k=4 (compared to 3 in E) to detect the single trials representing with red circles.
H. Characteristic grand-averaged single-trial derived from the averaging of the selected hubs/single trials (red circles) in G. We selected 15 signals from 128 trials. Blue waveform denotes the grand-average from the whole set of trials while the black from the selected single-trials. Amplitude and latency are estimated in the grand-averaged based on the global maxima.
Figure 5: Butterworth IIR order 2: Outline of the proposed graph-based methodology (High intensity condition - deviant).

A. Plot of single trials from a single subject
B. Similarity matrix that tabulates the pair-wise associations of single-trials with distance correlation metric
C. Embedding the similarity matrix in B in a 2D space with multidimensional scaling (MDS) algorithm where every blue dot represents a single-trial
D. Construct the Gabriel graph (GG) based on the 2D positions of the single trials. Green lines represent the connections under the notion of GG
E. Detection of hubs-representative single-trials based on their degree k in the GG. Here, we selected k=3 to detect the single trials representing with red circles.
F. Characteristic grand-averaged single-trial derived from the averaging of the selected hubs/single trials (red circles) in E. We selected 40 signals from 128 trials. Blue waveform denotes the grand-average from the whole set of trials while the black from the selected single-trials.
G. Detection of hubs-representative single-trials based on their degree k in the GG. Here, we selected k=4 (compared to 3 in E) to detect the single trials representing with red circles.
H. Characteristic grand-averaged single-trial derived from the averaging of the selected hubs/single trials (red circles) in G. We selected 25 signals from 128 trials. Blue waveform denotes the grand-average from the whole set of trials while the black from the selected single-trials. Amplitude and latency are estimated in the grand-averaged based on the global maxima.
Figure 6: FIR order 2: Outline of the proposed graph-based methodology (High intensity condition - deviant).

A. Plot of single trials from a single subject
B. Similarity matrix that tabulates the pair-wise associations of single-trials with distance correlation metric
C. Embedding the similarity matrix in B in a 2D space with multidimensional scaling (MDS) algorithm where every blue dot represents a single-trial
D. Construct the Gabriel graph (GG) based on the 2D positions of the single trials. Green lines represent the connections under the notion of GG
E. Detection of hubs-representative single-trials based on their degree $k$ in the GG. Here, we selected $k=3$ to detect the single trials representing with red circles.
F. Characteristic grand-averaged single-trial derived from the averaging of the selected hubs/single trials (red circles) in E. We selected 44 signals from 128 trials. Blue waveform denotes the grand-average from the whole set of trials while the black from the selected single-trials.
G. Detection of hubs-representative single-trials based on their degree $k$ in the GG. Here, we selected $k=4$ (compared to 3 in E) to detect the single trials representing with red circles.
H. Characteristic grand-averaged single-trial derived from the averaging of the selected hubs/single trials (red circles) in G. We selected 15 signals from 128 trials. Blue waveform denotes the grand-average from the whole set of trials while the black from the selected single-trials. Amplitude and latency are estimated in the grand-averaged based on the global maxima.
4.4. Reliability of Amplitude, Latency and Signal Power

We accessed the reliability of amplitude, latency and signal power estimates for each MMN feature, EEG sensors and for standard, deviant and deviant-standard with the coefficient of variation (CV).

The CV was estimated as follow:

\[ CV = \frac{\text{group mean( amplitude, latency )}}{\text{group std( amplitude, latency )} } \]  

(8)

4.5 Amplitude and Latencies

Tables 3-5 demonstrated the group mean amplitude for standard, deviant and deviant – standard for each condition of the MMN experimental protocol and for the three EEG sensors. We estimated the coefficient of variation (CV) across the cohort for every MMN feature for standard, deviant and deviant – standard and for FZ (Table 3) , FCZ (Table 4) and CZ (Table 5) EEG sensors. It is obvious that CV of the amplitude was higher for FZ EEG sensor.

---

### Table 3. FZ EEG sensor: Group-averaged amplitude for each condition and for standard, deviants and their difference (deviant - standard). Within the brackets, we reported the coefficient of variation (CV).

<table>
<thead>
<tr>
<th></th>
<th>Dir-R</th>
<th>Dir-L</th>
<th>Freq-Hi</th>
<th>Freq-Low</th>
<th>Int-Hi</th>
<th>Int-Low</th>
<th>Duration</th>
<th>Gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Std</td>
<td>3.9±1.3 (3)</td>
<td>4.1±1.2 (3.78)</td>
<td>4.0±1.2 (3.5)</td>
<td>4.3±1.4 (3.1)</td>
<td>4.1±0.9 (4.2)</td>
<td>3.7±1.5 (2.4)</td>
<td>2.3±0.7 (3.2)</td>
<td>2.4±0.7 (3.3)</td>
</tr>
<tr>
<td>Dev</td>
<td>–3.5±1.1 (3.2)</td>
<td>–3.5±1.1 (4.4)</td>
<td>–3.9±1.2 (3.4)</td>
<td>–4.1±1.2 (3.6)</td>
<td>–2.4±0.6 (4)</td>
<td>–2.5±0.5 (5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dev-Std</td>
<td>−4.7±1.2 (3.8)</td>
<td>−6.4±1.5 (4.12)</td>
<td>−4.7±1.4 (3.3)</td>
<td>−4.3±1.4 (3.2)</td>
<td>–3.2±0.5 (6.4)</td>
<td>–3.1±0.6 (5.1)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

### Table 4. FCZ EEG sensor: Group-averaged amplitude for each condition and for standard, deviants and their difference (deviant - standard). Within the brackets, we reported the coefficient of variation (CV).

<table>
<thead>
<tr>
<th></th>
<th>Dir-R</th>
<th>Dir-L</th>
<th>Freq-Hi</th>
<th>Freq-Low</th>
<th>Int-Hi</th>
<th>Int-Low</th>
<th>Duration</th>
<th>Gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Std</td>
<td>3.5±1.6 (2.2)</td>
<td>3.8±1.3 (3.9)</td>
<td>3.7±1.3 (2.8)</td>
<td>3.6±1.3 (2.7)</td>
<td>3.8±1.1 (3.4)</td>
<td>3.6±1.3 (2.7)</td>
<td>2.4±0.7 (3.4)</td>
<td>2.2±0.6 (3.75)</td>
</tr>
<tr>
<td>Dev</td>
<td>–3.2±1.6 (2)</td>
<td>–3.9±1.3 (3)</td>
<td>–3.6±1.3 (2.8)</td>
<td>–3.7±1.3 (2.8)</td>
<td>–2.5±1.1 (2.2)</td>
<td>–3.3±1.3 (2.6)</td>
<td>–2.5±0.6 (4.25)</td>
<td>–2.6±0.7 (3.6)</td>
</tr>
<tr>
<td>Dev-Std</td>
<td>−4.7±1.6 (2.9)</td>
<td>−3.7±1.6 (2.3)</td>
<td>−4.1±1.4 (2.9)</td>
<td>−3.9±1.4 (2.8)</td>
<td>−4.1±1.3 (3.2)</td>
<td>−4.0±1.4 (2.8)</td>
<td>−3.3±0.9 (3.5)</td>
<td>−3.2±0.8 (4)</td>
</tr>
</tbody>
</table>

16
Table 5. **CZ EEG sensor**: Group-averaged amplitude for each condition and for standard, deviants and their difference (deviant - standard). Within the brackets, we reported the coefficient of variation (CV).

<table>
<thead>
<tr>
<th></th>
<th>Dir-R</th>
<th>Dir-L</th>
<th>Freq-Hi</th>
<th>Freq-Low</th>
<th>Int-Hi</th>
<th>Int-Low</th>
<th>Duration</th>
<th>Gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Std</td>
<td>3.2 ± 1.3 (2.5)</td>
<td>2.4 ± 1.0 (2.4)</td>
<td>2.3 ± 1.0 (2.3)</td>
<td>2.3 ± 0.9 (2.5)</td>
<td>2.4 ± 0.8 (3)</td>
<td>1.9 ± 0.7 (2.4)</td>
<td>1.9 ± 0.6 (3.1)</td>
<td>2.5 ± 0.7 (3.4)</td>
</tr>
<tr>
<td>Dev</td>
<td>−2.2 ± 1.0 (2.2)</td>
<td>−2.3 ± 1.1 (2.89)</td>
<td>−2.3 ± 0.9 (2.6)</td>
<td>−2.2 ± 0.8 (2.7)</td>
<td>−2.5 ± 1.0 (2.5)</td>
<td>−2.0 ± 0.8 (2.5)</td>
<td>−2.1 ± 0.7 (2.7)</td>
<td>−2.6 ± 0.6 (4.3)</td>
</tr>
<tr>
<td>Dev-Std</td>
<td>−3.5 ± 1.2 (2.9)</td>
<td>−3.4 ± 1.3 (4.12)</td>
<td>−3.1 ± 1.1 (2.8)</td>
<td>−3.3 ± 1.4 (2.4)</td>
<td>−3.2 ± 1.3 (2.5)</td>
<td>−2.4 ± 1.1 (2.2)</td>
<td>−2.3 ± 0.9 (2.6)</td>
<td>−3.3 ± 0.7 (4.6)</td>
</tr>
</tbody>
</table>

Tables 6-8 demonstrated the group mean latencies for standard, deviant and deviant – standard for each condition of the MMN experimental protocol and for the three EEG sensors. We estimated the coefficient of variation (CV) across the cohort for every MMN feature for standard, deviant and deviant – standard and for FZ (Table 6), FCZ (Table 7) and CZ (Table 8) EEG sensors. It is obvious that CV of the latency was higher for FZ EEG sensor.

Table 6. **FZ EEG sensor**: Group-averaged latency for each condition and for standard, deviants and their difference (deviant - standard). Within the brackets, we reported the coefficient of variation (CV).

<table>
<thead>
<tr>
<th></th>
<th>Dir-R</th>
<th>Dir-L</th>
<th>Freq-Hi</th>
<th>Freq-Low</th>
<th>Int-Hi</th>
<th>Int-Low</th>
<th>Duration</th>
<th>Gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Std</td>
<td>0.18 ± 0.03 (6)</td>
<td>0.19 ± 0.03 (6.3)</td>
<td>0.18 ± 0.02 (9)</td>
<td>0.17 ± 0.03 (5.6)</td>
<td>0.21 ± 0.03 (7)</td>
<td>0.18 ± 0.02 (9)</td>
<td>0.19 ± 0.03 (6.3)</td>
<td>0.17 ± 0.03 (5.6)</td>
</tr>
<tr>
<td>Dev</td>
<td>0.19 ± 0.03 (6.3)</td>
<td>0.18 ± 0.03 (6)</td>
<td>0.17 ± 0.03 (5.6)</td>
<td>0.18 ± 0.02 (9)</td>
<td>0.19 ± 0.02 (9.5)</td>
<td>0.16 ± 0.03 (5.3)</td>
<td>0.18 ± 0.03 (6)</td>
<td>0.17 ± 0.03 (5.6)</td>
</tr>
<tr>
<td>Dev-Std</td>
<td>0.17 ± 0.02 (8.5)</td>
<td>0.16 ± 0.02 (8)</td>
<td>0.16 ± 0.02 (8)</td>
<td>0.17 ± 0.03 (5.6)</td>
<td>0.17 ± 0.03 (5.6)</td>
<td>0.16 ± 0.02 (8)</td>
<td>0.16 ± 0.02 (5.3)</td>
<td>0.16 ± 0.02 (8)</td>
</tr>
</tbody>
</table>

Table 7. **FCZ EEG sensor**: Group-averaged latency for each condition and for standard, deviants and their difference (deviant - standard). Within the brackets, we reported the coefficient of variation (CV).

<table>
<thead>
<tr>
<th></th>
<th>Dir-R</th>
<th>Dir-L</th>
<th>Freq-Hi</th>
<th>Freq-Low</th>
<th>Int-Hi</th>
<th>Int-Low</th>
<th>Duration</th>
<th>Gap</th>
</tr>
</thead>
</table>

17
Table 8. CZ EEG sensor: Group-averaged latency for each condition and for standard, deviants and their difference (deviant - standard). Within the brackets, we reported the coefficient of variation (CV).

<table>
<thead>
<tr>
<th></th>
<th>Dir-R</th>
<th>Dir-L</th>
<th>Freq-Hi</th>
<th>Freq-Low</th>
<th>Int-Hi</th>
<th>Int-Low</th>
<th>Duration</th>
<th>Gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Std</td>
<td>0.17±0.02(8.5)</td>
<td>0.18±0.03(6)</td>
<td>0.17±0.02(8.5)</td>
<td>0.17±0.03(6.3)</td>
<td>0.17±0.03(6.3)</td>
<td>0.17±0.02(8.5)</td>
<td>0.18±0.03(6)</td>
<td>0.19±0.03(6.3)</td>
</tr>
<tr>
<td>Dev</td>
<td>0.18±0.03(6)</td>
<td>0.21±0.03(7)</td>
<td>0.19±0.03(6.3)</td>
<td>0.16±0.02(8)</td>
<td>0.17±0.02(8.5)</td>
<td>0.18±0.03(6)</td>
<td>0.21±0.03(7)</td>
<td>0.18±0.03(6)</td>
</tr>
<tr>
<td>Dev-Std</td>
<td>0.19±0.02(9.5)</td>
<td>0.18±0.02(9)</td>
<td>0.18±0.02(9)</td>
<td>0.18±0.03(6)</td>
<td>0.16±0.03(5.3)</td>
<td>0.19±0.02(9.5)</td>
<td>0.20±0.03(6.6)</td>
<td>0.21±0.02(10.5)</td>
</tr>
</tbody>
</table>

We repeated the whole analysis by selecting a subset of single-trials from each condition and subject starting from the first 20% of the trials per condition till 100% with a step of 5% in order to explore how the number of trials affect amplitude/latency estimations. We revealed that the CV of amplitude/latency reached high values close to the ones tabulated in Tables 3 – 8 when the number of trials ranged between [85% - 95%] of the total amount of single trials. This practically means that the design in MMN which is state-of-the-art is more than significant to detect true amplitude/latency estimations.

4.6 Signal Power

Figures 7-9 demonstrated the group mean signal power for standard, deviant and deviant – standard for each condition of the MMN experimental protocol and for the three EEG sensors. We estimated the signal power for every MMN feature for standard, deviant and deviant – standard and for FZ (Figure 7), FCZ (Figure 8) and CZ (Figure 9) EEG sensors. In Figs 7-9, we demonstrated the signal power for each condition and std, dev and std-dev for the whole set.
of trials and also for the selection of a subset of trials. CV of signal power was higher for the subset of trials compared to the whole set of trials.

Fig.7. FZ - Group-averaged signal power for each condition, for standard, deviants and their difference (deviant - standard) in both groups and across seven frequency bands. We demonstrated the signal power for the whole set of trials and for the selection of a subset of trials.
Fig. 8. FCZ - Group-averaged signal power for each condition, for standard, deviants and their difference (deviant - standard) in both groups and across seven frequency bands. We demonstrated the signal power for the whole set of trials and for the selection of a subset of trials.
Fig. 9. CZ - Group-averaged signal power for each condition, for standard, deviants and their difference (deviant - standard) in both groups and across seven frequency bands. We demonstrated the signal power for the whole set of trials and for the selection of a subset of trials.
4.7 Variability of Single-Trials via Global Efficiency Estimations

To access the variability of single trials, we estimated the global efficiency (GE) on a network level based on the subgraph defined by the nodes of GG linked to the extracted prototypical single-trials (red nodes in Fig.4,5,6,E,G). We constructed the weighted graph by estimating the Euclidean distance between every pair of hub areas. Then, we normalized by the maximum value and we inversed the normalized weights in order to express functionality. This practically means that the higher the distance between two 2D points, the less efficient are communicated. In a global level, the higher the GE the more efficient is the communication of the nodes within network which further means that the nodes are closer in 2D points. So higher values of GE can be linked directly to less variability.

Table 9 tabulates the group-averaged GE for each condition in the three midline located EEG sensors for standard) stimuli.

### Table 9. Group-averaged GE for each condition across the three selected EEG sensors for (deviant - standard) stimuli. Within the brackets, we presented the coefficient of variation.

<table>
<thead>
<tr>
<th></th>
<th>Dir-L</th>
<th>Dir-R</th>
<th>Freq-Hi</th>
<th>Freq-Low</th>
<th>Int-Low</th>
<th>Int-High</th>
<th>Duration</th>
<th>Gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>FZ NC</td>
<td>0.37±0.05 (7)</td>
<td>0.39±0.07 (5.6)</td>
<td>0.41±0.04 (10.25)</td>
<td>0.36±0.04 (9)</td>
<td>0.42±0.03 (14)</td>
<td>0.38±0.06 (6.3)</td>
<td>0.36±0.05 (7.2)</td>
<td>0.37±0.04 (9.25)</td>
</tr>
<tr>
<td>FCZNC</td>
<td>0.34±0.06 (5.6)</td>
<td>0.39±0.04 (9.8)</td>
<td>0.37±0.04 (9.25)</td>
<td>0.35±0.04 (8.75)</td>
<td>0.37±0.03 (12.3)</td>
<td>0.34±0.03 (11.3)</td>
<td>0.38±0.06 (6.33)</td>
<td>0.39±0.06 (6.5)</td>
</tr>
<tr>
<td>CZ NC</td>
<td>0.39±0.03 (13)</td>
<td>0.37±0.04 (9.25)</td>
<td>0.36±0.04 (9)</td>
<td>0.38±0.04 (9.5)</td>
<td>0.38±0.07 (5.4)</td>
<td>0.36±0.04 (9)</td>
<td>0.38±0.04 (9.5)</td>
<td>0.38±0.06 (6.33)</td>
</tr>
</tbody>
</table>

4.8 Influence of EEG Reference System to Amplitude and Latency Estimation

We compared average to REST reference system in terms of amplitude and latency estimation across EEG sensor locations and in standard, deviant and standard-deviant stimuli. Comparing Tables 3-5 vs Tables 10-12 for amplitude and Tables 6-8 vs Tables 13-15 for latency, we revealed alterations of the group mean amplitude and latency. In both cases, the CV was too high while in some cases especially in the amplitude of (deviant - standard), the CV was higher for REST reference. For more detailed results, you can see section 2 in supp.material.
Table 10. FZ EEG sensor: Group-averaged amplitude for each condition and for standard, deviants and their difference (deviant - standard). Within the brackets, we reported the coefficient of variation (CV).

<table>
<thead>
<tr>
<th></th>
<th>Dir-R</th>
<th>Dir-L</th>
<th>Freq-Hi</th>
<th>Freq-Low</th>
<th>Int-Hi</th>
<th>Int-Low</th>
<th>Duration</th>
<th>Gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Std</td>
<td>4.3 ± 0.9(4.7)</td>
<td>4.3 ± 1.2(3.5)</td>
<td>4.0 ± 1.2(3.3)</td>
<td>4.6 ± 1.3(3.5)</td>
<td>4.7 ± 0.7(6.7)</td>
<td>4.5 ± 1.6(2.8)</td>
<td>2.9 ± 0.8(3.6)</td>
<td>3.4 ± 0.9(3.7)</td>
</tr>
<tr>
<td>Dev</td>
<td>5.1 ± 0.9(3.2)</td>
<td>6.7 ± 1.0(5.1)</td>
<td>5.1 ± 1.2(3.4)</td>
<td>4.5 ± 1.3(3.5)</td>
<td>4.9 ± 0.9(5.4)</td>
<td>4.8 ± 1.1(4.3)</td>
<td>2.8 ± 0.5(5.6)</td>
<td>2.9 ± 0.5(5.8)</td>
</tr>
<tr>
<td>Dev-Std</td>
<td>5.1 ± 0.9(3.2)</td>
<td>6.7 ± 1.0(5.1)</td>
<td>5.1 ± 1.2(3.4)</td>
<td>4.5 ± 1.3(3.5)</td>
<td>4.9 ± 0.9(5.4)</td>
<td>4.8 ± 1.1(4.3)</td>
<td>2.8 ± 0.5(5.6)</td>
<td>2.9 ± 0.5(5.8)</td>
</tr>
</tbody>
</table>

Table 11. FCZ EEG sensor: Group-averaged amplitude for each condition and for standard, deviants and their difference (deviant - standard). Within the brackets, we reported the coefficient of variation (CV).

<table>
<thead>
<tr>
<th></th>
<th>Dir-R</th>
<th>Dir-L</th>
<th>Freq-Hi</th>
<th>Freq-Low</th>
<th>Int-Hi</th>
<th>Int-Low</th>
<th>Duration</th>
<th>Gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Std</td>
<td>3.7 ± 1.4(2.6)</td>
<td>3.7 ± 1.0(3.7)</td>
<td>3.8 ± 0.9(4.2)</td>
<td>3.7 ± 0.9(4.2)</td>
<td>3.7 ± 1.2(3.1)</td>
<td>2.7 ± 0.6(4.5)</td>
<td>2.5 ± 0.5 (5)</td>
<td></td>
</tr>
<tr>
<td>Dev</td>
<td>3.4 ± 1.3 (2.6)</td>
<td>3.8 ± 1.1(3.4)</td>
<td>3.7 ± 1.0(3.7)</td>
<td>3.9 ± 1.1(3.5)</td>
<td>2.8 ± 0.9(4.2)</td>
<td>3.5 ± 1.1(3.1)</td>
<td>2.8 ± 0.5(5.8)</td>
<td>2.8 ± 0.6(4.6)</td>
</tr>
<tr>
<td>Dev-Std</td>
<td>4.6 ± 1.2(3.8)</td>
<td>3.9 ± 0.8(4.8)</td>
<td>4.2 ± 0.9(4.6)</td>
<td>4.2 ± 1.2(3.5)</td>
<td>4.3 ± 0.9(4.7)</td>
<td>4.5 ± 1.1(4.1)</td>
<td>3.5 ± 0.7 (5)</td>
<td>3.5 ± 0.7(5)</td>
</tr>
</tbody>
</table>

Table 12. CZ EEG sensor: Group-averaged amplitude for each condition and for standard, deviants and their difference (deviant - standard). Within the brackets, we reported the coefficient of variation (CV).

<table>
<thead>
<tr>
<th></th>
<th>Dir-R</th>
<th>Dir-L</th>
<th>Freq-Hi</th>
<th>Freq-Low</th>
<th>Int-Hi</th>
<th>Int-Low</th>
<th>Duration</th>
<th>Gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Std</td>
<td>3.7 ± 1.1(3.3)</td>
<td>2.6 ± 0.9(2.8)</td>
<td>2.8 ± 0.8(3.5)</td>
<td>2.7 ± 0.8(3.3)</td>
<td>2.8 ± 0.9(3.1)</td>
<td>2.3 ± 0.8(2.8)</td>
<td>2.3 ± 0.5(6.4)</td>
<td>2.6 ± 0.4(6)</td>
</tr>
<tr>
<td>Dev</td>
<td>2.5 ± 0.9(2.7)</td>
<td>2.7 ± 1.0(2.7)</td>
<td>2.7 ± 0.7(3.8)</td>
<td>2.8 ± 0.9 (3.1)</td>
<td>2.9 ± 1.1(2.6)</td>
<td>2.5 ± 0.9(2.7)</td>
<td>2.6 ± 0.5(5.2)</td>
<td>2.7 ± 0.5(5.4)</td>
</tr>
<tr>
<td>Dev-Std</td>
<td>3.8 ± 1.1(3.4)</td>
<td>3.7 ± 0.9(4.1)</td>
<td>3.5 ± 0.8(4.3)</td>
<td>3.7 ± 1.1(3.3)</td>
<td>3.9 ± 0.8(4.8)</td>
<td>2.9 ± 0.9(3.2)</td>
<td>2.8 ± 0.7(4)</td>
<td>3.3 ± 0.5(6.6)</td>
</tr>
</tbody>
</table>

Tables 13-15 demonstrated the group mean latencies for standard, deviant and deviant – standard for each condition of the MMN experimental protocol and for the three EEG sensors. We estimated the coefficient of variation (CV) across the cohort for every MMN feature for standard, deviant and deviant – standard and for FZ (Table 6), FCZ (Table 7) and CZ (Table 8) EEG sensors. It is obvious that CV of the latency was higher for FZ EEG sensor.
**Table 13. FZ EEG sensor:** Group-averaged latency for each condition and for standard, deviants and their difference (deviant - standard). Within the brackets, we reported the coefficient of variation (CV).

<table>
<thead>
<tr>
<th></th>
<th>Dir-R</th>
<th>Dir-L</th>
<th>Freq-Hi</th>
<th>Freq-Low</th>
<th>Int-Hi</th>
<th>Int-Low</th>
<th>Duration</th>
<th>Gap</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Std</strong></td>
<td>0.17 ± 0.03(5.6)</td>
<td>0.18 ± 0.03(6)</td>
<td>0.18 ± 0.02 (9)</td>
<td>0.18 ± 0.03 (6)</td>
<td>0.18 ± 0.03 (5.6)</td>
<td>0.18 ± 0.03</td>
<td>0.17 ± 0.03 (5.6)</td>
<td>0.17 ± 0.03</td>
</tr>
<tr>
<td><strong>Dev</strong></td>
<td>0.18 ± 0.03 (6)</td>
<td>0.17 ± 0.03(5.6)</td>
<td>0.16 ± 0.03(6)</td>
<td>0.17 ± 0.03 (5.6)</td>
<td>0.16 ± 0.03 (5.3)</td>
<td>0.17 ± 0.03 (5.6)</td>
<td>0.16 ± 0.03(5.3)</td>
<td>0.16 ± 0.03</td>
</tr>
<tr>
<td><strong>Dev-Std</strong></td>
<td>0.16 ± 0.02 (8)</td>
<td>0.15 ± 0.02 (7.5)</td>
<td>0.15 ± 0.02 (7.5)</td>
<td>0.16 ± 0.02 (7.5)</td>
<td>0.15 ± 0.03 (5)</td>
<td>0.16 ± 0.02 (8)</td>
<td>0.16 ± 0.02 (8)</td>
<td></td>
</tr>
</tbody>
</table>

**Table 14. FCZ EEG sensor:** Group-averaged latency for each condition and for standard, deviants and their difference (deviant - standard). Within the brackets, we reported the coefficient of variation (CV).

<table>
<thead>
<tr>
<th></th>
<th>Dir-R</th>
<th>Dir-L</th>
<th>Freq-Hi</th>
<th>Freq-Low</th>
<th>Int-Hi</th>
<th>Int-Low</th>
<th>Duration</th>
<th>Gap</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Std</strong></td>
<td>0.18 ± 0.03(6)</td>
<td>0.17 ± 0.03(5.6)</td>
<td>0.18 ± 0.03 (6)</td>
<td>0.17 ± 0.03 (5.6)</td>
<td>0.17 ± 0.03 (5.6)</td>
<td>0.17 ± 0.03</td>
<td>0.18 ± 0.03 (6)</td>
<td></td>
</tr>
<tr>
<td><strong>Dev</strong></td>
<td>0.17 ± 0.02 (8.5)</td>
<td>0.18 ± 0.03(6)</td>
<td>0.17 ± 0.02 (5.6)</td>
<td>0.15 ± 0.02 (7.5)</td>
<td>0.18 ± 0.03(5.6)</td>
<td>0.17 ± 0.03 (5.6)</td>
<td>0.19 ± 0.03(5.6)</td>
<td>0.19 ± 0.03</td>
</tr>
<tr>
<td><strong>Dev-Std</strong></td>
<td>0.17 ± 0.02 (8.5)</td>
<td>0.18 ± 0.02 (9)</td>
<td>0.18 ± 0.03 (6)</td>
<td>0.17 ± 0.03 (5.6)</td>
<td>0.17 ± 0.02 (8.5)</td>
<td>0.17 ± 0.03 (5.6)</td>
<td>0.19 ± 0.02 (8.5)</td>
<td></td>
</tr>
</tbody>
</table>

**Table 15. CZ EEG sensor:** Group-averaged latency for each condition and for standard, deviants and their difference (deviant - standard). Within the brackets, we reported the coefficient of variation (CV).

<table>
<thead>
<tr>
<th></th>
<th>Dir-R</th>
<th>Dir-L</th>
<th>Freq-Hi</th>
<th>Freq-Low</th>
<th>Int-Hi</th>
<th>Int-Low</th>
<th>Duration</th>
<th>Gap</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Std</strong></td>
<td>0.17 ± 0.02(8.5)</td>
<td>0.18 ± 0.03(6)</td>
<td>0.18 ± 0.03 (6)</td>
<td>0.17 ± 0.03 (5.6)</td>
<td>0.18 ± 0.03(5.6)</td>
<td>0.17 ± 0.03</td>
<td>0.17 ± 0.03 (5.6)</td>
<td></td>
</tr>
<tr>
<td><strong>Dev</strong></td>
<td>0.18 ± 0.02 (8.5)</td>
<td>0.19 ± 0.03 (6.3)</td>
<td>0.18 ± 0.02 (6)</td>
<td>0.15 ± 0.02 (7.5)</td>
<td>0.17 ± 0.03(5.6)</td>
<td>0.18 ± 0.03(6)</td>
<td>0.19 ± 0.03(6.3)</td>
<td>0.18 ± 0.03(6)</td>
</tr>
<tr>
<td><strong>Dev-Std</strong></td>
<td>0.16 ± 0.02 (8)</td>
<td>0.17 ± 0.02 (8.5)</td>
<td>0.17 ± 0.03 (5.6)</td>
<td>0.16 ± 0.03 (5.6)</td>
<td>0.16 ± 0.03 (9.5)</td>
<td>0.18 ± 0.02 (9)</td>
<td>0.18 ± 0.03 (6)</td>
<td>0.19 ± 0.02 (9.5)</td>
</tr>
</tbody>
</table>
4.9 Comparison with Alternative Mining Algorithms

We compared our methodology with PCA and multi-linear regressor analysis (Hu et al., 2011). The first one proposed a multiple linear regression (MLR) and multiple linear regression with dispersion term (MLRd) to estimate the single-trial latency and amplitude of event-related potential (ERP) peaks. Regressors (an average and its temporal derivative) for each ERP peak are calculated from the average ERP waveform within a given post-stimulus interval (in this case, 0 to 0.3 s) for each subject. These regressors are then applied against each single trial within the same post-stimulus interval and used to model each single-trial ERP peak. In MLR d, variability matrices that capture the variations of latency and morphology of each ERP peak are generated by simultaneously shifting and compressing the average ERP waveform (step 1). These variability matrices, whose order of trials (with the latency shifted and the morphology varied simultaneously) is of no importance, are fed to a principal component analysis (PCA; step 2). The resulting 3 main principal components (PCs) are used to define 3 regressors for each peak within a given post-stimulus interval (in this case, 0 to 0.5 s; step 3). These regressors are then applied against each single trial within the same post-stimulus interval and used to model each single-trial ERP peak (step 4). The methodology is explained in details in Hu et al., 2011.

Since the original methodology focused on the estimation of amplitude-latency per single-trial, we grand-averaged the single-trials after first applying the regressors.

The second one is principal component analysis (PCA) where we kept the first PCs that explained more than 95% of the variance of single-trials.

In Fig.10, we illustrated the resulting grand-averaged time series from subject 1 and stimulus DIR-L for standard, deviant and deviant-standard using the multi-linear regressor algorithm and the average reference system. Complementary, Fig.11 demonstrates the effect of REST reference on the grand-averaged time series illustrated in Fig.16. Both grand-averaged time series were extracted from FZ EEG sensor.
Fig. 10. Grand-averaged response for DIR-L from subject 1 using multi-linear regressor analysis and average reference system (FZ-Sensor).

A. Standard stimulus
B. Deviant stimulus
C. Deviant-Standard stimulus

Fig. 11. Grand-averaged response for DIR-L from subject 1 using multi-linear regressor analysis and REST reference system (FZ-Sensor).

A. Standard stimulus
B. Deviant stimulus
C. Deviant-Standard stimulus

In Fig. 12, we illustrated the resulting grand-averaged time series from subject 1 and stimulus DIR-L for standard, deviant and deviant-standard using PCA algorithm and the average reference system. On this exemplar, we demonstrated the 4th and 5th principal component per case. Complementary, Fig. 13 demonstrates the effect of REST reference on the grand-averaged time series illustrated in Fig. 12. Both grand-averaged time series were extracted from FZ EEG sensor. We adopted the same stimulus, sensor location and subject with
multi-linear regressor analysis for comparison purposes between multi-linear regressor analysis and PCA. For further details, see section 3 in supp.material.

Fig.12. Grand-averaged response for DIR-L from subject 1 using PCA analysis and average reference system (FZ-Sensor).

A.Standard stimulus
B.Deviant stimulus
C.Deviant-Standard stimulus. We plotted the characteristic time series derived from PCA analysis, the plain average from the whole set of time series and the time series derived from the proposed method.

Fig.13. Grand-averaged response for DIR-L from subject 1 using PCA analysis and REST reference system (FZ-Sensor).

A.Standard stimulus
B.Deviant stimulus
C. Deviant-Standard stimulus. We plotted the characteristic time series derived from PCA analysis, the plain average from the whole set of time series and the time series derived from the proposed method.

In Fig. 14, we illustrated the resulting grand-averaged 4th and 5th singular time series from subject 1 and stimulus DIR-L for standard, deviant and deviant-standard using SVD algorithm and the average reference system. Complementary, Fig. 15 demonstrates the effect of REST reference on the grand-averaged first two right singular time series illustrated in Fig. 14. Both grand-averaged time series were extracted from FZ EEG sensor. We adopted the same stimulus, sensor location and subject with multi-linear regressor analysis for comparison purposes between multi-linear regressor analysis and PCA.

**Fig. 14.** Grand-averaged of the 4th and 5th singular-values for DIR-L from subject 1 using SVD analysis and average reference system (FZ-Sensor).

A. Standard stimulus  
B. Deviant stimulus  
C. Deviant-Standard stimulus
Fig.15. Grand-averaged of the 4th and 5th singular values for DIR-L from subject 1 using PCA analysis and REST reference system (FZ-Sensor).

A.Standard stimulus  
B.Deviant stimulus  
C.Deviant-Standard stimulus

The three algorithms independently of the EEG reference system failed to detect an accurate amplitude and latency. The main reason is that both algorithms are sensitive to the grand-averaged response which in many cases like in MMN experimental paradigm are too noisy to get a nice waveform that can be used as representative time series of brain response. The proposed data mining scheme worked better compared to the three comparable techniques and also it is a parameter free method that can easily be used in any experimental multi-trial paradigm.

4.10 Simulations

Based on the results derived from the simulations, we revealed that both amplitude and latency are within acceptable limits. Simulations have shown that estimates of amplitude and latency are within acceptable limits (Fig.16 and Fig.17). Only if SNR is low and latency variation is low, estimates become unreliable. Fig.16-17 illustrate the simulation based on recordings derived from the FZ sensor at DIR-L condition and for deviant-standard stimulus using the two simulated scenarios.
Fig. 16 The simulation based on recordings derived from the FZ sensor at DIR-L condition and for deviant-standard stimulus. Real versus estimated amplitude estimates for different amounts of SNR and latency variation.
5. Discussion

The proposed methodology can reliable sample the representative single-trials in order to simultaneously express their variability and also to reconstruct a grand-average with high SNR. A high SNR was detected across subjects, conditions and recording EEG locations which secure a reliable estimation of the amplitude and latency of the characteristic peak elicited from the whole set of MMN features. The whole approach relies on graph theory by constructing a
distance network from the 2D projected single trials in a common feature space. The construction of the distance matrix based on members/nodes of proximity graph called Gabriel Graph (GG). Then, we detected from GG, the hubs nodes/single-trials using the degree k of each node as an input. These hubs can clearly describe the variability of single-trials and also reconstruct a waveform with a high SNR which clearly demonstrates a characteristic peak. We presented also how the final reconstructed waveform changed due to different type of filters and the related order. To get the best combination of filter type, order and selection of degree k directly linked to the selected hubs, we employed SNR as a true objective function. Our results can be summarized below:

- FIR filter with eegfilt function of order 2 and degree=4 demonstrates the highest SNR across conditions and subjects
- The reliability of amplitude and latency was higher for FZ EEG sensor compared to FCZ and CZ based on coefficient of variation (CV)
- CV of signal power was higher for the subset of trials compared to the whole set of trials.
- Amplitude and latency are sensitive to EEG reference system
- REST reference system improved the CV of amplitude in the deviant-standard stimuli
- Similar amplitude/latency estimations were revealed with the 85-95 % of the total amount of single-trials
- PCA, SVD and multi-linear regressor manipulation of single-trials failed to retrieve a robust waveform, latency and amplitude estimation.
- Amplitude and latency estimations with the proposed method are reliable unless SNR and latency variation is too low.

There are several extensions that can be applied to the present methodology in specific steps of the analytic algorithm. First of all, we demonstrated the effect of different filtering schemes where alternative adaptive filters can be used (Mandic and Goh, 2009). One can select different metrics to estimate the pair-wise correlation between single-trials. Complementary, one can use different members of proximity graphs like minimal spanning tree alone or in combination with voronoi testellation (Laskaris et al., 2001,2004) and relative neighborhood graph (RNG). Here, we tested both of them but we revealed best results with Gabriel graph (GG). Additionally, it would be very interesting to apply source-localization algorithms on the
representative single-trials to connect single-trials with sources and the timing of related activity (Laskaris et al., 2001, 2003). For example, one can localize the early segment of activity and the later one, in order to demonstrate the early activation of auditory cortex and the later activation in frontal lobe (Rinne et al., 2000).

Regarding the adopted experimental paradigm to demonstrate this methodology, the MMN mechanism consists of an auditory-based frontal lobe network. After the pre-processing of the content of MMN by the targeted sensory system here the auditory cortex, frontal areas are activated playing a significant role in the elicitation of a reflex (Naatanen and Michie, 1979). The MMN generators come from temporal and frontal lobes and the related activity is captured mainly by fronto-central EEG sensors (FZ, FCZ, CZ) and also from temporal electrodes (T3, T4) (Rinne et al., 2000). Two studies focused to reveal both with EEG and MEG recordings the origin of the elicited activity linked to MMN. Dipole modelling techniques applied to MMN (Scherg et al., 1989) and its magnetic counterpart (MMNm) (Hari et al., 1984) were found as generators in the auditory cortex and in the temporal lobes. Complementary, the analysis of scalp-potential distribution revealed a right-hemispheric MMN source, which mainly was located over the frontal lobe (Giard et al., 1990; Deouell et al., 1998). A more recent paper compared to the aforementioned with simultaneously EEG-MEG recording set up, source-localized both EEG and MEG activity in an auditory MMN (Rinne et al., 2000). They validated the hypothesis that frontal MMN generators are activated later than generator of the auditory cortex. For a review of MMN generators in both healthy and disease groups and various settings, an interested reader can refer to a detailed review (Garrido et al., 2009).

The mismatch negativity (MMN) is an ERP elicited by the occurrence of a rare event (deviance) in a regular acoustic environment, and is assumed to reflect a pre-attentive mechanism for change detection. Cortical generators of MMN are located in the superior temporal planes bilaterally which are responsible for the sensory memory part of change detection and frontal lobe sources responsible for triggering an attention shift upon change detection (for a review see Deouel, 2007). These bilaterally temporal-frontal generators of MMN can be better detected with EEG compared to MEG while the combination of both modalities was suggested (Hamalainen et al., 1993). Apart from bilateral auditory-cortex activation which underlines a pre-perceptual change detection with a short time-delay (Rinne et al., 2000), a predominant right hemispheric frontal process could be detected linked to involuntary attention switch to auditory change (Rinne et al., 2005, 2006). The dominant hemisphere of MMN response due to acoustic changes is the right hemisphere (Levanen et al.,
1996). For that reason, it is important in a next study to further explore amplitude-latency estimations also in bilateral frontal electrode sites complementary to EEG sensors located in the midline.

In the present study, we focused on the presentation of a data-driven methodology for a proper manipulation of single-trials. We demonstrated high reliability in amplitude, latency, variability and signal power for the whole cohort of young adults. Additionally, the majority of the fronto-central EEG channels should be studied for uncover any significant asymmetries of the brain activity between the two groups. Complementary, the main focus of this study was to enhance the reliability of the proposed methodology to unfold high SNR grand-averaged trials in various MMN conditions and reliable estimates of amplitude, latency and variability in a healthy group. The REST reference system improved the CV of amplitude in the deviant-standard stimuli while PCA, SVD and multi-linear regressor manipulation of single-trials failed to retrieve a robust waveform, latency and amplitude estimation. The proposed data-driven scheme worked better compared to the three well-known comparable methodologies. Moreover, it is parameter free method that can easily be adjusted to any multi-trial experimental paradigm using EEG-MEG recordings in both sensor and source level. Finally, amplitude and latency estimations with the proposed method are reliable unless SNR and latency variation is too low.

The whole methodology will be valuable for neuroscientists particularly interest on defining reliable biomarker based on ERP studies in various cognitive states (Horvath et al., 2018) and also in disease brain states such as the Alzheimer’s Disease (Tsolaki et al., 2017).

6. Conclusions

We presented a fast, reliable and data-driven methodology for simultaneously data-mining single-trials and amplitude-latency estimation. The method relies on graph and network analysis as appropriate tools of geometrical data analysis and vectorial pattern analytic tools of single-trials. We demonstrated the effect of filtering settings on the grand-averaged trial and the related amplitude-latency estimates. Additionally, the whole methodology was presented in an auditory EEG MMN task with main scope to detect reliable amplitude, latency and signal power derived from the appropriate preselection of single-trials. Based on the data-driven approach of the current methodology, the whole analysis could be of high value for various evoked/event-related potentials in various neuroimaging studies including EEG, MEG and fMRI.
Acknowledgments
SID and DEJL were supported by MRC grant MR/K004360/1 (Behavioural and Neurophysiological Effects of Schizophrenia Risk Genes: A Multi-locus, Pathway Based Approach. SID is supported by a MARIE-CURIE COFUND EU-UK research fellowship. We would like to acknowledge Cardiff RCUK funding scheme.

Conflict of Interest Statement
Authors declare that there is no conflict of interest.

Author Contributions
Conception of the research: SD; Methods and design: SD; Data analysis (SD); Drafting the manuscript: SD; Critical revision of the manuscript: (LB, LE, DL, KS); Every author read and approved the final version of the manuscript.
References


Deouell, L. Y. The frontal generator of the mismatch negativity revisited J. Psychophysiol., 21 (3–4) (2007), pp. 188-203,

De Pascalis V, Varriale V, Matteoli A. Intelligence and p3 components of the event-related potential elicited during an auditory discrimination task with masking. Intelligence. 2008;36:35–47.


Rinne T, Degerman A, Alho K. Superior temporal and inferior frontal cortices are activated by infrequent sound duration decrements: an fMRI study, NeuroImage , 2005, vol. 26 (pg. 66-72)


