Pathological responses to single pulse electrical stimuli in epilepsy: the role of feedforward inhibition

SUBMITTED AUTHOR VERSION, ACCEPTED FOR PUBLICATION IN EJN

Jurgen Hebbink\textsuperscript{1,2}, Geertjan Huiskamp\textsuperscript{1}, Stephan A. van Gils\textsuperscript{2}, Frans S. S. Leijten\textsuperscript{1}, and Hil G. E. Meijer\textsuperscript{2}

\textsuperscript{1} Department of Neurology and Neurosurgery, Brain Center Rudolf Magnus, University Medical Centre Utrecht, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands

\textsuperscript{2} Department of Applied Mathematics, MIRA Institute for Biomedical Engineering and Technical Medicine, University of Twente, Drienerlolaan 5, 7500 AE Enschede, The Netherlands

Corresponding author: Jurgen Hebbink, Department of Applied Mathematics, MIRA Institute for Biomedical Engineering and Technical Medicine, University of Twente, Drienerlolaan 5, 7500 AE Enschede, the Netherlands. E-mail: g.j.hebbink@utwente.nl

Short title: Modelling delayed responses to stimulation

Keywords: single pulse electrical stimulation, neural mass model, focal epilepsy, early responses, delayed responses

Abbreviations: DR (delayed response), ECoG (electrocorticography), ER (early response), EZ (epileptogenic zone), HFA (high-frequency activity), IED (interictal epileptiform discharges), NMM (neural mass model), PSP (postsynaptic potential), SOZ (seizure onset zone), SPES (single pulse electrical stimulation)

Number of words abstract: 249
Number of words main text: 6960
Number of pages: 35
Number of figures: 10
Number of tables: 0
Number of supplementary material: 1
Abstract

Delineation of epileptogenic cortex in focal epilepsy patients may profit from single pulse electrical stimulation during intracranial EEG recordings. Single pulse electrical stimulation evokes early and delayed responses. Early responses represent connectivity. Delayed responses are a biomarker for epileptogenic cortex, but up till now, the precise mechanism generating delayed responses remains elusive.

We used a data-driven modelling approach to study early and delayed responses. We hypothesized that delayed responses represent indirect responses triggered by early response activity and investigated this for 11 patients. Using two coupled neural masses we modelled early and delayed responses by combining simulations and bifurcation analysis. An important feature of the model is the inclusion of feedforward inhibitory connections.

The waveform of early responses can be explained by feedforward inhibition. Delayed responses can be viewed as second-order responses in the early response network which appear when input to a neural mass falls below a threshold forcing it temporarily to a spiking state. The combination of the threshold with noisy background input explains the typical stochastic appearance of delayed responses. The intrinsic excitability of a neural mass and the strength of its input influence the probability at which delayed responses to occur.

Our work gives a theoretical basis for the use of delayed responses as a biomarker for the epileptogenic zone, confirming earlier clinical observations. The combination of early responses revealing effective connectivity, and delayed responses showing intrinsic excitability, makes single pulse electrical stimulation an interesting tool to obtain data for computational models of epilepsy surgery.
**Introduction**

Epilepsy surgery may provide a cure for patients with focal epilepsy, especially if treatment with anti-epileptic drugs fails (Noachtar & Borggraefe, 2009; Jobst & Cascino, 2015). Epilepsy surgery aims at removing the epileptogenic zone (EZ), i.e. the smallest area of cortex the removal of which yields seizure freedom (Rosenow & Lüders, 2001; Lüders et al., 2006). Although several non-invasive methods exist to approximate the EZ, the current gold standard remains the seizure onset zone (SOZ) determined via seizures captured during intracranial EEG recordings. The dependence on the occurrence of spontaneous events has the disadvantage that long recording time is therefore needed, typically ranging from a few days up to weeks, which increases the burden of the patient, the risk of complications and costs.

To probe epileptogenicity during electrocorticography (ECoG) registration, single pulse electrical stimulation (SPES) is an alternative for observing spontaneous interictal or ictal changes, with the advantage that it is controlled (Valentín et al., 2002; van ’t Klooster et al., 2011). During SPES, brief electric pulses are applied directly to the cortex using the electrode grids implanted for clinical ECoG recordings. SPES typically evokes two types of responses: early responses (ERs) and delayed responses (DRs).

ERs appear directly and consistently after the stimulation, with a similar timing and shape across stimulation trials. They are well-understood as they represent brain connectivity (Lacruz et al., 2007; Keller et al., 2014; Matsumoto et al., 2017). ERs have, therefore, been used to investigate connectivity in several functional regions like the language area and motor cortex (see (Matsumoto et al., 2017) for an overview). From ERs, directional networks can be constructed (Hebbink et al., 2019). These networks (partly) explain ipsilateral seizure propagation (Enatsu et al., 2012; Mouthaan et al., 2016), in contrast to contralateral seizure spread (Jiménez-Jiménez et al., 2015). Moreover, network measures calculated from these ER
networks exhibit differences between nodes in and outside the resected area and SOZ (Boido et al., 2014; van Blooijs et al., 2018).

DRs appear later than ERs, between 100 ms and 1 s after stimulation (Valentín et al., 2002), have a stochastic occurrence, i.e. they only appear on a subset of the stimulation trials at the same electrode pair, and come with variable timing and shape. Where ERs are physiological responses, linked to the EZ only through network structure, DRs are a direct biomarker for epileptogenic cortex (Valentín et al., 2005b; van ’t Klooster et al., 2011). So far, research on DRs predominantly assessed their clinical value. DRs were observed in different brain regions (Valentín et al., 2002; Valentín et al., 2005a) in both adults and children (Flanagan et al., 2009).

Further, investigation of the frequency content of DRs revealed that especially high-frequency activity (HFA) in the fast-ripple band (250-500 Hz) is specific for epileptogenic cortex (van ’t Klooster et al., 2011). Although useful for clinical practice, these results do not give a mechanistic explanation for DRs. Such understanding would provide a better basis for the clinical use of DRs.

Computational models offer a tool to investigate the responses observed during SPES. Neural mass models (NMMs) can simulate EEG-like activity of a small patch of neural tissue ranging from a few millimetres to a couple of centimetres. The development of NMMs dates back to the early seventies to explain generation of the α-rhythm in the thalamus (Lopes da Silva et al., 1974). Later NMMs were extended to models for cortical activity (Jansen & Rit, 1995; Wendling et al., 2002; Ursino et al., 2010; Wang & Knösche, 2013). NMMs are able to generate a variety of both healthy and epileptic EEG rhythms (Wendling et al., 2000; Sotero et al., 2007) and also to describe event-related responses (Jansen et al., 1993; Jansen & Rit, 1995; David et al., 2005; Cona et al., 2011; Wang & Knösche, 2013). Multiple neural masses can be coupled to create an interacting network, allowing the study of network mechanisms. Usually, neural masses have been coupled only through purely excitatory connections, although it is known
that multiple connection types exist (Felleman & Van Essen, 1991). Especially feedforward inhibition plays an important role in seizure propagation (Eissa et al., 2017) and responses to transcranial magnetic stimulation (Cona et al., 2011).

In this work we try to explain the two most characteristic properties of DRs, namely their relatively late appearance compared to ERs and their stochastic nature. The first property suggests that DRs might represent indirect, rather than direct, responses to stimulation. These responses may be triggered by ERs, mediated through the ER network. The latter property might be a consequence of noise, which pushes the system only occasionally beyond a threshold for generating DRs. We investigate these hypotheses using both analysis of SPES data and by modelling DRs using a NMM equipped with feedforward connections to inhibitory cells.

First we will summarize properties regarding appearance, timing and amplitude of ERs and DRs. Next, we will investigate if DRs can be seen as indirect responses within the ER network by performing data analysis on SPES data recorded in epilepsy patients. We then present a computational model of DRs which we calibrate by using properties of ERs. Finally, we use bifurcation analysis to reveal a mechanism that explains DRs as second order responses in the ER network including the up to now elusive stochastic appearance of DRs.

**Materials and methods**

**SPES acquisition**

At the UMC Utrecht, SPES is performed as part of clinical routine in the pre-surgical evaluation of epilepsy patients during long-term ECoG monitoring. For the ECoG monitoring intracranial electrode grids, usually consisting of 2-8×8 electrodes and strips (1×8 electrodes), are placed directly on the cortex. Electrodes have a circular shape with a contact area of 4.2 mm² and an inter-electrode distance of 1 cm.
The SPES protocol has been described in detail (van ’t Klooster et al., 2011; van Blooijs et al., 2018). In short, pairs of adjacent electrodes along the length of the grid receive ten monophasic electrical pulses with a duration of 1 ms, at an interval of 5 s and a typical intensity of 8 mA. During SPES, ECoG data is recorded with respect to an extracranial reference located on the contralateral mastoid at a sampling rate of 2048 Hz using a SD LTM express (MicroMed, Veneto, Italy). Stimulations are applied via the LTM stimulator and cause a 9 ms lasting artefact in all signals. Further, the stimulated channels become saturated for about 5 s upon stimulation, so responses can only be observed in the remaining electrodes.

**ERs**

SPES ERs are defined as responses starting within 100 ms after stimulation. ERs are normal, physiological brain responses describing connectivity (Valentín et al., 2002; Lacruz et al., 2007; Keller et al., 2014; Matsumoto et al., 2017). The most common type of ER (Fig. 1A) consists of three peaks, which are in order of appearance: N1, P1 and N2 (Alarcón et al., 2018), while in other ERs (Fig. 1D) the last component is absent. The N1 is a sharp negative peak, occurring roughly between 10 and 50 ms after stimulation (Entz et al., 2014), with the majority around 15 ms, see Figs. 1B, E. The P1 is the positive deflection following the N1. Its maximum lies around baseline level and typically occurs ~35 ms after the N1. The timing of both the N1 and P1 is similar for ERs with and without N2 component. The broad negative slow wave following the P1 is the N2. Its waveform is more variable compared to N1 and P1 peaks as it is shaped by spontaneous activity too. The biggest amplitude is attained some 80 to 160 ms after the N1 with the median around 110 ms, see Fig. 1B. The ratio between N1 and N2 amplitude varies (Fig. 1C). For some stimulation pairs and response electrodes the N2 is smaller than the N1, while for others it is much larger.

In some cases, we observed an additional positive peak preceding the N1 peak, see the inset of Fig. 1A for an example, which we call P0. Such a peak has also been reported in (Matsumoto
et al., 2017; Boyer et al., 2018). In this example the P0 peak is found 11.2 ms after stimulation with an amplitude of 0.4 times that of N1. The P0 peak is difficult to find in the data, due to its relatively low amplitude and because it mixes with the stimulation artefact (up to 9 ms) in most cases. So, only if the N1 peak is relatively late, there is a time window in which such a P0 may be noticed.

To systematically detect ERs in our SPES data we use an automatic detector. This detector is described in detail in the supplementary material of (Hebbink et al., 2019) and has been validated using visually annotated ERs. In short, the responses of an electrode over all ten trials of a stimulation pair are averaged. If the extremum of this signal between 9 to 100 ms after stimulation sufficiently exceeds the standard deviation of the baseline, i.e. the 2 s prior to stimulation, the response is classified as ER.

**DRs**

DRs are responses to SPES appearing at least 100 ms after stimulation (see Fig. 2). DRs occur only in a subset of the trials, e.g. in Fig. 2A DRs occur in trials 1, 2, 5, 6 and 9, around 200-400 ms after stimulation. The exact timing of DRs differs per trial and is even more variable across stimulation pairs and electrodes. For example, DRs in Fig. 2B start around 400 ms after stimulation, while the others in Fig. 2 start around 200 ms.

The waveform of DRs varies substantially across electrodes. The DRs shown in Figs. 2A-C have been recorded from the same electrode. These DRs start with rapid oscillations and are, except for some trials in Fig. 2C, followed by a slow wave. DRs recorded from other electrodes look more like spike-wave discharges as those in trials 1, 3, 5 and 6 of Fig. 2D. Interestingly, trials 8, 9 and 10 of this stimulation exhibit DRs similar to those in Fig. 2A, so DR waveforms may vary even within the same stimulation for the same electrode. Note also that an electrode may show both an ER and DR simultaneously as can be seen in Fig. 2E.
Both the variable timing and varying waveform render averaging responses in the time domain problematic. Typical analysis of DRs employs time-frequency analysis using a Morlet wavelet transformation and averages the resulting time-frequency plots (van ’t Klooster et al., 2011). Time-frequency plots allow classification of DRs with components in the spike (10-80 Hz), ripple (80-250 Hz) and fast-ripple (250-500 Hz) bands (see Fig. 2F). We determine the presence of DRs per frequency band by visual analysis of the time-frequency plots, see (van ’t Klooster et al., 2011) for details. In short, for each channel a time-frequency image is produced per stimulation pair which is independently scored by two observers in the three frequency bands. Only responses on which the two observers agreed are considered. Inter-observer agreement is assessed via Cohen’s κ and is considered to be reasonable if κ>0.4. Here we study DRs that cover at least the spike and ripple band. The presence of a fast-ripple is not required as these occur rarely (van ’t Klooster et al., 2011).

Next, the onset time of DRs is determined by visual analysis of the single trial responses. For each trial a time-frequency image is produced using the same settings as used for the averaged time frequency images. The onset is then marked as the first time point where both increased activity in the time frequency image and a clear onset of the DR in the time signal is visible. We compare the onset time between two groups of DRs, i.e. those which are preceded by an ER, and those that are not. Using the Wilcoxon rank-sum test (function ranksum in Matlab) we test whether the median values of these two groups are equal against the alternative hypothesis that the median onset time of DRs preceded by an ER is later. We reject the null hypothesis for p < 0.05.

Path length of DRs in ER networks

We use the detected ERs to construct a network representing effective connectivity (Keller et al., 2014; Entz et al., 2014; van Blooij et al., 2018; Hebbink et al., 2018). Each electrode
represents a node in this network. An edge from node A to B is present if stimulation involving electrode A evokes an ER at electrode B. The result is an unweighted directional network.

Next, we study the distance from stimulation pairs to the electrodes on which they evoke DRs in the ER network. We define distance as the length of the shortest path between a stimulation pair and an electrode. Starting from a stimulation pair all electrodes on which an ER is evoked are at a distance one from the stimulation pair. Next, a node has distance two from the stimulation pair if it can be reached via some edge from a node with distance one in the ER-network, provided the node itself did not exhibit an ER. Continuing this way, a node is at distance $n$ from the stimulation pair if it is not at distance $n - 1$ or closer and it can be reached via an edge from a node at distance $n - 1$. Nodes that cannot be reached in this way have an infinite distance to the stimulation pair. This does not imply that the tissue under such an electrode is disconnected from the tissue under the stimulated electrodes as the electrode grid only samples a part of the brain and it might be connected via an uncovered part of the brain.

We investigate the distance from stimulation pair to DR electrode in SPES data of 11 patients recorded during long-term ECoG monitoring at the UMC. For each patient we calculate the percentage of DRs at a distance of at most 1, 2, 3 up to $n_{el} - 2$ from the stimulation pair in the ER network, where $n_{el}$ is the number of electrodes. In all patients SPES was performed for clinical reasons with the protocol described above. Patients had been admitted to the intensive epilepsy monitoring unit of the University Medical Centre of Utrecht, the Netherlands. All patients gave prior informed consent which was recorded in the patient’s electronic file and the entire investigation was performed under the approval of the UMC Utrecht’s ethical committee under Dutch law. Data were retrospectively collected and handled coded and anonymously according to the guidelines of the institutional ethical committee following the principles of good clinical practice and adhering to the Declaration of Helsinki.
Mathematical model

To model the observed ECoG responses to SPES we use a system of coupled neural masses. Each of these neural masses can be thought to model the tissue underneath an electrode of the ECoG grid. We consider an extended version of the neural mass proposed by (Wendling et al., 2002), see Fig. 4A. This neural mass models the average membrane potential of four neuronal populations i.e. pyramidal cells (py), local excitatory cells (ex), slow inhibitory cells (is) and fast inhibitory cells (if). As pyramidal cells are the main contributor to EEG signals, their average membrane potential is considered to be proportional to EEG signals (Jansen & Rit, 1995; Wendling et al., 2002; Sotero et al., 2007; Cona et al., 2011) and taken as the output of the model.

The average membrane potential of a population determines the activity of that population, i.e. the mean firing rate of the neurons in the population, via a non-linear function \( S(\mu) \) (see Fig. 4B). Following (Wilson & Cowan, 1972; David et al., 2005; Wang & Knösche, 2013), we shift this function such that \( S(0) = 0 \), to model deviations from the normal activity. Accordingly, negative values are interpreted as activity below baseline.

The activity of a population influences the mean membrane potential of other populations via synaptic transmissions. Such synaptic transmission converts the mean firing rate of the sending population into a postsynaptic potential at the receiving population and is modelled by a linear, second order differential equation. The time course of the synaptic transmission is determined by the rising time of the synapse. Following (Jansen & Rit, 1995; Wendling et al., 2002) we set the rising time for excitatory synapses to 10 ms. For the fast inhibition, modelling fast GABA\(_A\) transmission, we take a rising time of 3.3 ms, which is inside the physiological plausible range reported by (Molaee-Ardekani et al., 2010). Finally, we assume that slow inhibition models mainly transmission of slow GABA\(_A\) but also partly transmission of GABA\(_B\). We therefore set this time scale to 100 ms, which is slightly slower than 30 to 70 ms for GABA\(_A\) but faster than
the 200 to 400 ms for GABA\textsubscript{B} as suggested by (Molaee-Ardekani et al., 2010). The impulse responses of the three different synapse types are shown in Fig. 4C.

In each neural mass the pyramidal cells play a key role, as they are reciprocally connected to all other populations. In addition, slow inhibitory cells project to fast inhibitory cells. The relative strength of the connections is set to commonly used values (Wendling et al., 2002) as derived previously from anatomical studies (Jansen & Rit, 1995). To arrive at absolute connection strengths, these relative strengths are multiplied by a constant $C$, governing the general internal connectivity. This $C$ can also be seen as a measure for the excitability of a neural mass as increasing $C$ results in more epileptiform dynamics of the neural mass (Goodfellow et al., 2011; Touboul et al., 2011).

Apart from local interactions neural masses also receive excitatory external input. In the original model this input projects only to the pyramidal cells. Here, also the inhibitory populations receive external input. We model the strength of the feedforward inhibitory input relative to the feedforward excitatory input, using scaling constants $\beta$ and $\gamma$ for slow and fast populations, respectively. External input originates from three sources, i.e. other neural masses, background activity and SPES. Input from other neural masses depends on the activity of their pyramidal cells and connection strength. Background inputs are modelled by independent Gaussian white noise with zero-mean and standard deviation $\sigma$. SPES input to a neural mass is modelled as a short, transient external input (block pulse). This input mimics the activation of the outgoing fibres of a stimulated region which are thought to be activated during SPES rather than the cell bodies (David et al., 2010; Keller et al., 2014). Therefore, SPES input is given simultaneously to all neural masses connected to the stimulated area.

We consider two feedforward coupled neural masses in two different configurations (see Figs. 4D-E). In both configurations the first neural mass receives SPES input and has a connection with strength $k$ to the second neural mass. Depending on the case, SPES input to the second
neural mass might be present or absent. Following our hypothesis, we intend to model an ER on the first and a DR on the second neural mass. Depending on the configuration this DR is preceded by an ER on the second neural mass or not, capturing some of the different cases seen in the data (see Fig. 2). In both cases, the complete system comprises a set of 20 coupled differential equations (see supplementary material 1).

Our first modelling step concerns simulating a realistic ER, both with and without a N2 component. For this part we adapt the input strengths to the two inhibitory populations, $\beta$ and $\gamma$. Once these parameters are set, we proceed with modelling DRs by varying the connectivity strength between the neural masses. We perform bifurcation analysis using Matcont (Dhooge et al., 2008) to infer a mechanism explaining the stochastic occurrence of DRs. Next, we add background noise to the model, which we neglect in the first part, to show a stochastically occurring DR. Finally, we study how the excitability of the second neural mass and the connection strength influence the rate at which DRs occur.

Results

Data analysis

The relatively late appearance of DRs compared to ERs after stimulation suggests that DRs could be indirect responses due to propagation of activity via the ER network. Example data supporting this hypothesis is shown in Fig. 3A. SPES delivered at an electrode pair (yellow nodes) evokes ERs at green nodes. These green nodes correspond to all outgoing connections of the stimulation pair in the ER network. Also, the same stimulation evokes DRs, for instance at the orange-coloured electrode. Next, blue nodes indicate that if that electrode is stimulated then the orange node shows an ER, hence they correspond to all nodes that project to the orange node in the ER network. The overlap is indicated by the green-blue striped pattern. We see that in this network there are multiple, i.e. 5, paths of length two in the ER network from the
stimulation electrodes to the electrode showing a DR. These length-two paths are the shortest routes between the stimulation pair and the orange electrode in the ER network, as the orange electrode does not show an ER for this stimulation pair (see also the response of this electrode in Fig. 2C). Therefore, the orange electrode is at a distance two from the stimulation pair.

Next, we investigate the distance between stimulation pair and DR in 11 patients. The bar chart in Fig. 3D shows the fraction of DRs that can be reached by a path of a certain length. For most patients, a path of length one, i.e. an ER and DR are seen on the same electrode, explains only half of the DRs. With a path of at most length two, one reaches on average 92% of the DRs, which is rather constant over all patients. A path length of three or four gives a slight improvement, but higher path lengths do not add anymore. This suggests to model DRs as second order responses in the ER network. The simplest network architecture satisfying this constraint is a feedforward network consisting of two nodes, which we will consider in this work.

**Modelling ERs**

To model ERs we consider a single neural mass receiving a short transient input (see Fig. 4). An example of a simulated ER is shown in Fig. 5A. Here, the thick line indicates the output of the model, i.e. the simulated potential of the pyramidal cells. Note that the simulated ER has a similar shape as those observed in the data. The latency of the P0, N1, P1 and N2 peaks is also realistic as they are found 6, 23, 50 and 118 ms after stimulation respectively.

The potential of the pyramidal cells is the sum of the contributions from the other populations and the SPES input, which are indicated by the thin coloured lines in Fig. 5A. The initial increase in the potential of the pyramidal cells is a direct consequence of the stimulation. As the stimulation has a similar effect on both inhibitory populations, these populations are activated and start to influence the pyramidal cells. First, input from the fast inhibitory
population comes into play yielding the N1 peak. The effect of the slow inhibitory population appears much later and results in the N2 peak. The local excitatory cells only play a minor role in the generation of ERs as they are not directly activated by the SPES stimulus. The amplitude of the ER depends approximately linearly on the stimulation strength as is shown in Fig. 5B.

The external input to the inhibitory cells is the crucial aspect for the occurrence of the N1 and N2 peaks. Figs. 5C-D show the influence of variations in $\beta$ and $\gamma$ respectively. Variations in $\beta$ only affect the amplitude of the N2 peak, which scales approximately linear with $\beta$. As a consequence, ERs without N2 component can be modelled by setting $\beta = 0$. Variations of $\gamma$ affect mostly the N1 peak. For moderate values of $\gamma$ the N1 peak scales approximately linear.

For low values of $\gamma$ the N1 peak occurs at the same time, but as the amplitude of the peak is small, the neural mass proceeds differently to the N2. For high values of $\gamma$ the amplitude of the N1 peak grows only slowly, while the latency increases slightly.

**Modelling DRs**

We now consider two feedforward coupled neural masses, as depicted in Fig. 4D, where the first neural mass projects to the second with connectivity strength $k$. The first neural mass shows ER activity upon stimulation. Fig. 6A shows the reaction of the second neural mass for various $k$ values. For low values of $k$, the second neural mass shows virtually no response. For $k \gtrsim 20$, a large response, comprising a few oscillations followed by a slow wave, appears. This waveform is qualitatively similar to the DRs in Figs. 2A-B, although the frequency of the oscillations in the limit cycle are slower, i.e. $\sim 28$ Hz compared to the high-frequency activity ($>80$ Hz) at the start of the DRs.

A possible way to model an ER succeeded by a DR is to apply SPES input to both NMMs as in Fig. 4E. For this, $k$ needs to be large, i.e. $k \gtrsim 75$ (see Fig. 6B). The shape of this DR is
similar to the DRs that are not preceded by an ER. For $k$ values below the threshold only ERs are observed on the second neural mass.

Next, we study the influence of the connectivity strength $k$ on the latency of the DRs. For each DR, we consider three time points, i.e. the onset of the oscillations, $t_o$, the transition from oscillations to slow wave, $t_t$ and the time the slow wave reaches its minimum, $t_s$ (see inset of Fig. 6C). Here, we defined $t_o$ and $t_t$ as the first and last time at which the response crosses $u_{py} = 3.5$. As is shown in Fig. 6C, the latency of the DRs decreases rapidly for $k$ slightly above the threshold for which DRs emerge. If $k$ gets bigger, the time the DR shows oscillations grows as the $t_o$ remains decreasing for high $k$, while $t_t$ gets bigger if $k$ is above 60 (in case of a DR not preceded by an ER). The difference between $t_t$ and $t_w$ remains almost constant.

The model predicts that DRs which are preceded by an ER occur later than those which are not. We investigate this in the patient data. Fig. 7A shows a histogram of the onset times of DRs for the two groups in a single patient confirming that DRs preceded by an ER occur in general later than those that are not. Histograms of the other patients are supplied in supplementary material 1. The boxplot in Fig. 7B summarizes the results for all patients. Note that the timing of the simulated DRs without an ER is comparable to the data. DRs preceded by an ER occur rather late in the simulations in comparison with the data. Observe that in some patients a clear difference between the distributions is visible, while for others the median values are close together. A one-sided Wilcoxon rank-sum yields a significant p-value in 8 of the 11 patients for a significance level of 0.05.

**A mechanism for DRs**

So far we have shown that DRs can be modelled as indirect responses to SPES. We observed that DRs emerge all of a sudden, if the connectivity strength passes some critical value. The next step is to understand the mechanism causing this sudden appearance. Using bifurcation
We study the influence of parameters on stationary and periodic solutions, i.e., equilibria and limit cycles, characterizing the activity of the neural mass.

For this analysis we focus solely on the second neural mass and treat the external input, $I$, as a parameter. Note that the input to this neural mass depends on both the connectivity strength $k$, and the activity of the first neural mass. So by studying the input as a bifurcation parameter we aim to get insight in the sudden appearance of DRs when the connectivity strength passes some critical value.

Results of the bifurcation analysis varying $I$ are shown in Fig. 8A. Under normal circumstances the external input $I$, is zero. For this value the neural mass has an equilibrium (stationary solution) with $u_{py} = 0$. This equilibrium is stable, i.e. for any small perturbation of the equilibrium state, the neural mass will converge to the equilibrium if time proceeds. The solid blue line at the bottom of Fig. 8A shows how this equilibrium behaves if $I$ is varied. For positive values of $I$ the equilibrium remains stable. For negative values of $I$ the equilibrium becomes unstable at $I \approx -5.46$, where it undergoes a Hopf bifurcation (Kuznetsov, 2004). The unstable equilibrium curve, indicated by the dashed blue line in Fig. 8A, contains two fold bifurcation at $I \approx -5.58$ and $I \approx 14.20$. At both these points the equilibrium curve changes direction, although the equilibrium remains unstable as it leaves the graph.

For $I \leq -4.58$ the neural mass has a stable limit cycle (periodic solution), with a waveform (Fig. 8B) similar to that of the simulated DRs (Fig. 6A-B). Like for the equilibrium, we continue the limit cycle varying $I$. The minimum and maximum amplitude of the limit cycle are indicated by the solid red lines in the bifurcation diagram. The limit cycle vanishes at $I \approx -4.58$ via a homoclinic bifurcation (Kuznetsov, 2004), which is characterized by an increase in the period of the limit cycle towards infinity (Fig. 6C). In supplementary material 1 we discuss the bifurcation diagram in more detail.
From the bifurcation diagram we can deduce the mechanism responsible for generating DRs in the model. First, note that the input neural mass 2 receives from neural mass 1 showing an ER, is negative, as the average membrane potential of the pyramidal cells of neural mass 1 is negative. Therefore, neural mass 2 will follow the stable equilibrium branch for negative $I$. If $k$ is small, $I$ stays above the Hopf bifurcation at $I \approx -4.58$ and the neural mass will stay on the stable equilibrium branch while the input returns to zero. On the other hand, for larger $k$, $I$ will cross the Hopf bifurcation. As the stable equilibrium is not present here, the neural mass moves to the stable limit cycle. While the neural mass follows the limit cycle for approximately one period, the input $I$ returns to zero and crosses the homoclinic bifurcation where the limit cycle vanishes. As a consequence the neural mass returns to the equilibrium state.

**Stochastic DRs and local excitability**

We can use the mechanism described above to model DRs that have a stochastic appearance. Consider again two feedforward coupled neural masses as in Fig. 4D and set the connectivity strength around the critical value for which DRs emerge, i.e. $k = 20$. By adding noisy background input to the model, the second neural mass will pass the Hopf bifurcation occasionally and hence DRs will appear stochastically as show in Fig. 9.

Next, we investigate how the probability of evoking a DR depends on both the connectivity strength $k$ and the intrinsic excitability of the second neural mass $C$. Fig. 10 shows the result of simulating one hundred SPES stimulations of the model for various values of $C$ and $k$. Both increasing $k$ and $C$ increases the probability of evoking DRs. If $C$ gets too large the probability decreases slightly, because the neural mass also shows spontaneous interictal discharges in this case. The relation between $k$ and $C$ agrees with results found using bifurcation analysis varying $I$ and $C$ (see supplementary material 1).
Discussion

With our data-driven modelling study, we have revealed how network connectivity and intrinsic excitability lead to pathological delayed responses upon stimulation. This sheds new light on their role as a clinical biomarker for epileptogenic tissue.

Starting with direct responses to stimulation, i.e. ERs, we showed that the two common peaks, i.e. the N1 and N2, can be explained by triggering not only the pyramidal cells but also both fast and slow inhibitory cells. This is in line with a modelling study by (Alarcón et al., 2018) using linear response theory. The NMM predicted an additional P0 peak which we subsequently identified in the data.

Next, based on connectivity analysis of ER networks (van Blooijs et al., 2018; Hebbink et al., 2019), we concluded that DRs are indirect responses to the stimulation. Model simulations show that they may arise from feedforward propagation of ER-activity, and notably projections to inhibitory populations that normally restrain neural activity. Moreover, the model correctly predicted that DRs preceded by an ER occur later than DRs that are not. We find that DRs may appear when the input falls below a threshold, temporarily allowing the neural mass to enter a spiking state. This threshold also explains the stochastic appearance of DRs through noisy background activity. The probability for a DR to occur, depends on both the intrinsic excitability of a neural mass and the strength of the input from other regions with ERs. A higher intrinsic excitability moves this threshold closer to the normal state, so that less input is needed to pass the threshold, while stronger connections amplify input deviations to this region.

Limitations of modelling

In this work we considered Wendling's neural mass that can generate several activity types as observed in ECoG signals (Wendling et al., 2002; Blenkinsop et al., 2012) and explains transitions in these activities seen during epileptic seizures (Wendling et al., 2016). However,
an important feature which cannot be modelled is high-frequency activity. Model extensions with self-feedback for the fast inhibitory population (Ursino et al., 2010; Chehelcheraghi et al., 2016) allow to simulate responses with frequencies up to approximately 70 Hz. Another study (Molaee-Ardekani et al., 2010) considered a NMM consisting of only pyramidal and fast inhibitory cells to produce HFA. Their analysis shows that the connectivity constants in the loop between the fast inhibitory and pyramidal cells are important for the maximum frequency that the neural mass can generate. Adjusting these constants, the systems' activity will resonate with this frequency allowing frequencies up to 120 Hz. Thus, it may be possible to simulate such high frequencies with a Wendling-type NMM extended with self-feedback for the fast inhibitory population, which may allow to increase the frequency of the fast oscillations at the start of the simulated DRs. Simulating higher frequencies with neural masses poses a challenge, as these models only take synaptic transmissions into account and not the fast electric transmissions via gap junctions which are, among other factors, hypothesized to play an important role in generation of fast-ripples (Traub et al., 2001; Wendling et al., 2016).

Following other neural mass studies (Jansen & Rit, 1995; Wendling et al., 2002; Cona et al., 2011; Wang & Knösche, 2013) we compare the average membrane potential of the pyramidal cells to measured EEG signals. As the pyramidal cells are the main contributor of EEG signals, due to the parallel alignment of their dendrites, their average membrane potential is a reasonable first approximation of the EEG signal. More specifically, the average membrane potential is a weighted sum of the postsynaptic potentials (PSPs) on the pyramidal cells. The currents induced by these PSPs are the main source of EEG signals (Goodfellow, 2011). The precise relation between the PSPs and the observed EEG however, depends on multiple factors, e.g. the relative location of the electrode with respect to the gyrus (Goodfellow, 2011). Therefore, differences in the waveforms of observed responses, such as the relative amplitude of the N1 and N2
responses, are not necessarily explained by different model parameters, but may originate from the precise measurement setup.

As a first step in understanding the mechanisms triggering early and delayed responses to SPES we investigated the most simple network configuration, i.e. we considered two feedforward coupled neural masses. These two neural masses model only a small part of neuronal tissue, while in reality also recurrent connections will be present. These factors will influence the precise waveforms of the ERs and especially the DRs. We think this could also explain the difference between the latency of DRs preceded by an ER in the model and in the data. Moreover, depending on the network topography, the connectivity strength and the intrinsic parameters of the neural masses, complex transient dynamics may arise (Goodfellow et al., 2012). However, within the current model the mechanisms responsible for the ERs (direct activation due to stimulation) and DRs (limit cycle) are robust with respect to small parameter changes.

We focused on modelling the two most common types of SPES responses, i.e. ERs and DRs. Besides these, two other responses have been observed, i.e. stable and repetitive responses (Valentín et al., 2005a; Flanagan et al., 2009). Stable responses are small, consistently appearing spikes or sharp waves, mostly superimposed on the N2 component of an ER (Flanagan et al., 2009). They appear at the same electrodes for a variety of stimulation pairs (Flanagan et al., 2009), suggesting it to be a local phenomenon. Stable responses might therefore be modelled by slightly altering the parameters of a neural mass, or if needed, considering a small extension to the default neural mass. However, as stable responses are not related to the epileptic zone (Flanagan et al., 2009), such a change must not make the neural mass pathological.

Repetitive responses are observed exclusively in the frontal lobe and consist of an early response component which is repeated at least once (Valentín et al., 2005a). They seem to be
induced by stimulation of the epileptic zone and appear simultaneously in a spatially extended area (Valentín et al., 2005a). The former suggests that it might be necessary to model the effect of the stimulation in detail. The latter might be explained by cortico-subcortical loops (Valentín et al., 2005a), which would require a spatially extended network of neural masses (Goodfellow et al., 2012).

**ERs reveal feedforward inhibition**

A crucial aspect in our model is that external input projects not only onto pyramidal cells but also to both slow and fast inhibitory populations. This is in line with the seminal work by (Felleman & Van Essen, 1991) and has been considered in multiple computational modelling studies (David et al., 2005; Babajani-Feremi & Soltanian-Zadeh, 2010; Spiegler et al., 2010; Cona et al., 2011; Eissa et al., 2017; Shamas et al., 2018). These projections lead to direct activation of the inhibitory populations upon stimulation and as a consequence to the N1 and N2 peaks of ERs in the model. One might also consider external input projecting to the local excitatory cells of the neural mass. In supplementary material 1 we show that adding this connection does not quantitatively change our results, provided it is not too strong.

Single neuron measurements have revealed some distinct neuronal firing patterns after SPES, i.e. burst-suppression, burst only, suppression only or unchanged (Alarcón et al., 2012), where suppression typically lasts 5-10 times as long as bursting activity. In our model the different populations also show different firing patterns during an ER. Activity of the slow inhibitory cells is increased upon stimulation. Pyramidal cells and fast inhibitory cells first show increased activity followed by a long lasting period of decreased activity. The local excitatory population mainly shows a long lasting decrease in activity. It has also been suggested that only N1 peaks reflect direct activation of fibres, while the N2 might arise from cortico-cortical or cortico-subcortico-cortical reverberation circuits (Matsumoto et al., 2017). Our study shows that physiologically plausible values for the synaptic time constants, as reported in (Molae-
Ardekani et al., 2010), can naturally explain the N1 and N2 peaks as a direct consequence of the stimulation.

Our model is able to reproduce common properties of ERs, like the linear relation between stimulation strength and the amplitude of ERs (Donos et al., 2016). It has been reported that the majority of the waveforms contains either a N1 or N2 component, or both (Alarcón et al., 2018). Therefore, we considered two types of ERs, those with only a N1 component and those with a N1 and N2 component. Simulations of ERs without N2 are easily achieved by weakening the projection to the slow inhibitory population.

**Clinical role of DRs**

Several studies have revealed the clinical value of DRs for supporting hypotheses when delineating the epileptogenic zone (Valentín et al., 2002; Valentín et al., 2005a; van ’t Klooster et al., 2011). Our study provides a theoretical explanation for DRs and their use. DRs indicate that, already under normal conditions, the underlying neural mass is close to a state of epileptiform activity.

In our centre, DRs are visually classified based on spikes, ripples and fast-ripples using time-frequency images obtained from wavelet-analysis (van ’t Klooster et al., 2011). The majority of the DRs exhibits activity in at least the spike and ripple bands. The waveforms of the simulated DRs show a qualitative match, i.e. they consist of fast oscillations followed by a spike, although the frequency of the fast oscillations is below the ripple band. The morphology of DRs has similarities with interictal epileptiform discharges (IEDs). Moreover, (Nayak et al., 2014) found that DRs were similar to at least one IED pattern for every patient, while in single neuron measurements similar firing patterns were found during IEDs and after SPES (Alarcón et al., 2012). NMMs are also capable of simulating IEDs (Wendling et al., 2000). In fact, the mechanism responsible for these IEDs (Blenkinsop et al., 2012) is exactly the same as the
mechanism responsible for DRs, i.e. both arise from temporarily escaping to the limit cycle corresponding to epileptiform activity. So, in patients with few spontaneous IEDs, SPES-triggered DRs may be a useful addition.

Computational modelling of epilepsy surgery has received considerable attention in recent years with promising results (Terry et al., 2012; Goodfellow et al., 2016; Hebbink et al., 2017; Jirsa et al., 2017; Lopes et al., 2017; Sinha et al., 2017). In this framework, neural masses or similar models are coupled through a patient-specific connectivity. Subsequently, the effect of surgery is predicted by removing nodes from the network and the results are compared against a base-line simulation of the model, as well as actual clinical outcomes. Typically for these studies, each node is equally excitable. Our finding that DRs depend on both local excitability and network effects suggests that both these factors play an important role in the ability of a node to start a seizure. Hence, SPES may improve the prediction of such frameworks in two ways: a patient-specific network can be obtained from ERs (van Blooijs et al., 2018; Hebbink et al., 2019) and DRs can be used to differentiate excitability of the nodes.

Acknowledgements

The authors would like to thank especially Marij Tijssen and Dorien van Blooijs for their help with the visual analysis of delayed responses. J.H. is supported by ZonMW/Dutch Epilepsy Foundation Translational Research grant 95104015.

Competing interests

The authors declare that they have no conflict of interest.

Author Contributions

JH designed the study, collected data, analysed data and model, and wrote the manuscript. GH: designed the study, collected and analysed data, is data curator, and wrote the manuscript. FL
designed the study, collected data and wrote the manuscript. SvG designed the study and wrote the manuscript. HM designed the study, analysed data and model, and wrote the manuscript.

Data Accessibility

The computational model and simulation scripts are available via the 4TU centre for research data under DOI 10.4121/uuid:a0b0dece-99a2-455a-af89-b21b7fb0e307. Datasets will be made available upon reasonable request to g.j.m.huiskamp@umcutrecht.nl, following the brain imaging data structure standard.

Supplementary material

1. Detailed model description, additional bifurcation analysis and supplementary data.

References


Figure 1: Shape and properties of early responses (ERs) in vivo. (A) and (D) Examples of an ER with and without N2, respectively. Thin coloured lines are multiple trials for the same stimulation. The thick black line indicates the average response. The inset in (A) shows a magnification of the first part of the response, here the P0 peak is visible. (B) and (E) Distribution of the N1, P1 and N2 times for ERs with and without N2, respectively. Time of N1 peaks is the time after stimulation, P1 and N2 times are relative to N1. (C) and (F) Size of the amplitude A of the P1 and N2 peaks relative to that of the N1. Amplitudes are taken as the deviation from the mean activity 50 ms before stimulation, hence N1 and N2 peaks usually have a negative value.
Figure 2: Examples of delayed responses (DRs) in vivo. (A)-(E) Response of a single electrode to ten subsequent trials of the same stimulation pair. Red-coloured trial numbers indicate DRs. Responses at (A)-(C) were recorded at the same electrode, the same hold for the responses in (D)-(E). (F) Averaged time-frequency response of the trials in (C).
Figure 3: Delayed responses (DRs) as indirect response in a patient. (A) Example of a DR as second order response in a schematic overview of the electrode grid. The stimulation pair (yellow) induces a DR on the orange electrode (time trace shown in Fig. 2C) and early responses (ERs) on all green (solid + striped) electrodes. Blue (solid + striped) electrodes are part of stimulation pairs that induce ERs on the magenta electrode. The arrows indicate length two paths between stimulation pair and DR electrode. (B) Responses of the green-blue striped electrodes on stimulation of the yellow stimulation pair. (C) Responses of the orange electrode on stimulation pairs that consist of the green-blue striped electrodes. (D) Fraction of DRs that can be reached by a path of a certain length in the ER network, where order $\infty$ means that the DR can be reached by a path of arbitrary length. For each path order the 11 bars represent the results of the 11 patients, while the red dashed line indicates the mean over all patients.
Figure 4: Overview of the neural mass model. (A) Architecture of a single neural mass. The circles represent the four populations, i.e. pyramidal cells (py), local excitatory cells (ex), slow inhibitory cells (is) and fast inhibitory cells (if). Arrows represent excitatory connections, circles and squares are slow and fast inhibitory connections. (B) Graph of the sigmoid function used to convert the average membrane potential of a population into a firing rate. (C) Impulse response of the different synapse types. (D, E) The two configurations of feedforward coupled neural masses considered in this work.
Figure 5: Simulated early responses (ERs). (A) Simulated ER using default parameters (thick black line). Thin coloured lines indicate the 4 different sources, i.e. external input (ext) and input from excitatory (ex), slow inhibitory (is) and fast inhibitory (if) populations, adding up to the potential of the pyramidal cells (py). (B) Variation of stimulation strength, where $P_0 = 1500$ is the default strength (indicated with the tick black line). (C) Influence of varying $\beta$, the fraction of external input added to slow inhibitory cells. (D) Influence of varying $\gamma$, the fraction of external input added to the fast inhibitory cells.
**Figure 6: Simulating delayed responses (DRs).** (A) Results for the configuration in Fig. 4D. Neural mass 1 shows an early response (ER) upon stimulation (green line). Depending on the coupling strength $k$, the second neural mass shows a DR (other colours). (B) Results for the configuration in Fig. 4E. Both neural masses show an ER. For sufficiently high values of the coupling strength $k$ also a DR is simulated on the second neural mass. (C) Latency of the simulated DRs against the coupling strength $k$ for the situation in A (green) and B (magenta) for three reference points. The inset shows the location of the three reference points on the wave shape of a DR.
Figure 7: Onset time of delayed responses (DRs) in vivo. (A) Distribution of the onset time of DRs preceded by an ER (magenta) and DRs not preceded by an ER (green) in patient 2. (B) Boxplots for all patients showing the onset of the DRs, where colours refer to the same groups as in (A). Values between brackets indicate p-values of the statistical test and significant results (significance level 0.05) are marked with *. If $p < 0.001$ results are marked with **.
Figure 8: Bifurcation analysis. (A) Bifurcation diagram for varying external input $I$. Blue lines indicate equilibria, red lines are minima and maxima of limit cycles. Solid lines indicate stable solutions, while dashed lines indicate unstable ones. Fold (F), Hopf (H) and homoclinic (Hom) bifurcations are indicated by yellow circles. (B) Time profile of the stable limit cycle for $I = -7$. (C) Period of the stable limit cycle. The period approaches infinity near the homoclinic bifurcation.
**Figure 9:** Simulating stochastic DRs. Simulation of ten stimulation trials for the configuration in Fig. 4D where both neural masses receive additional background noise ($\sigma=1$) for a coupling strength $k=20$. An ER appears consistently on neural mass 1, while a DR appears stochastically on the second neural mass.
Figure 10: Probability of evoking delayed responses for varying $k$ and $C$. The connectivity strength $k$ and the intrinsic excitability $C$ are varied with steps of 0.5 and 2 respectively. For each combination of $k$ and $C$ 100 stimulations were simulated with noise ($\sigma = 1$) and the number of evoked delayed responses was counted.