frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 16 August 2021
doi: 10.3389/fnhum.2021.655505

Neural Field Theory of Evoked Response Sequences and Mismatch Negativity With Adaptation

Peter A. Robinson[1,2]*, *Natasha C. Gabay*[1,2]* and *Tara Babaie-Janvier*[1,2]

[1] School of Physics, University of Sydney, Sydney, NSW, Australia, [2] Center of Excellence for Integrative Brain Function, University of Sydney, Sydney, NSW, Australia

Physiologically based neural field theory of the corticothalamic system is used to calculate the responses evoked by trains of auditory stimuli that correspond to different cortical locations via the tonotopic map. The results are shown to account for standard and deviant evoked responses to frequent and rare stimuli, respectively, in the auditory oddball paradigms widely used in human cognitive studies, and the so-called mismatch negativity between them. It also reproduces a wide range of other effects and variants, including the mechanism by which a change in standard responses relative to deviants can develop through adaptation, different responses when two deviants are presented in a row or a standard is presented after two deviants, relaxation of standard responses back to deviant form after a stimulus-free period, and more complex sequences. Some cases are identified in which adaptation does not account for the whole difference between standard and deviant responses. The results thus provide a systematic means to determine how much of the response is due to adaptation in the system comprising the primary auditory cortex and medial geniculate nucleus, and how much requires involvement of higher-level processing.

Keywords: evoked response, mismatch negativity, neural field theory, adaptation, modeling

OPEN ACCESS

Edited by:
Christoph Braun,
University of Tübingen, Germany

Reviewed by:
Frank W. Ohl,
Leibniz Institute for Neurobiology (LG), Germany
Kirill Vadimovich Nourski,
The University of Iowa, United States

***Correspondence:**
Peter A. Robinson
peter.robinson@sydney.edu.au
Natasha C. Gabay
natasha.gabay@sydney.edu.au

Specialty section:
This article was submitted to Sensory Neuroscience, a section of the journal Frontiers in Human Neuroscience

Received: *20 January 2021*
Accepted: *20 July 2021*
Published: *16 August 2021*

Citation:
Robinson PA, Gabay NC and Babaie-Janvier T (2021) Neural Field Theory of Evoked Response Sequences and Mismatch Negativity With Adaptation. Front. Hum. Neurosci. 15:655505. doi: 10.3389/fnhum.2021.655505

1. INTRODUCTION

Evoked responses (ERs) to short impulse-like stimuli are commonly used to probe human cognitive processes. These responses are usually measured using electroencephalography (EEG) or magnetoencephalography (MEG), most often in auditory experiments, although other sensory modalities are also used (Luck and Kappenman, 2011; Niedermeyer and Lopes da Silva, 2011; Luck, 2014). Trains of ERs show a rich repertoire of effects in response to any violation of regularity—changes in frequency, location, duration, and intensity (Näätänen, 2003; Luck and Kappenman, 2011; Luck, 2014). Such effects are most often elicited in so-called oddball paradigms in which frequent standard ($\mathcal{S}$) stimuli are interspersed with rarer deviants ($\mathcal{D}$), which elicit very different ERs in general, so long as they are discriminable (Sams et al., 1985; Näätänen, 2003; Garrido et al., 2013), as seen in **Figure 1**. Deviants that are only marginally discriminable give an intermediate response (Sams et al., 1985; Garrido et al., 2009a). Here and throughout this paper we denote stimuli with calligraphic font to distinguish them from responses, which are written in italic font.

Deviant responses D are believed to reflect a response to novelty and, indeed, at the beginning of a stimulus train, all stimuli evoke D responses, but standard (S) responses evolve over a few

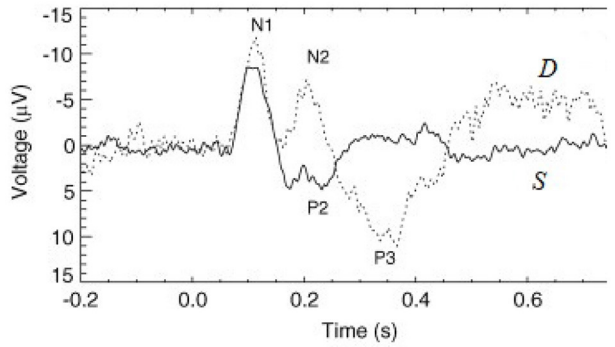

FIGURE 1 | Examples of typical standard (*S*, solid line) and deviant (*D*, dashed line) evoked responses to auditory stimuli in an auditory oddball experiment (Kerr et al., 2009). The traditional phenomenological "components" (peaks and troughs) are labeled, where N1 and P2 indicate a negative deflection around 100 ms and a positive one around 200 ms, respectively, for example.

presentations to their limiting form S_∞ as their preponderance becomes established (Näätänen, 2003; Garrido et al., 2009a, 2013; Luck and Kappenman, 2011; Luck, 2014); since a typical interstimulus interval is $\sim$ 1 s, this sets an adaptation timescale of several seconds. Similarly, after a pause in stimulation, *S* responses return to the *D* form over a few seconds, again indicating that the effects responsible for the difference between the two have a lifetime of seconds (Cowan, 1984; Winkler et al., 1993; Loveless et al., 1996; Näätänen, 2003). Likewise, both *D* and *S* responses differ from their usual forms if two $\mathcal{D}$ stimuli occur consecutively or an $\mathcal{S}$ follows two $\mathcal{D}$s (Sams et al., 1984). This implies that the different responses at least partly reflect recent stimuli rather than their long-term average probabilities.

At a more complex level, a *D* response is seen to a repeated tone in a descending sequence of tones, where there is no repeated $\mathcal{S}$ stimulus (Näätänen et al., 1989a; Tervaniemi et al., 1994; Näätänen, 2003; Garrido et al., 2009b, 2013); likewise, a *D* response occurs after a stimulus that is omitted or changed in duration or intensity (Näätänen et al., 1989b, 2007; Yabe et al., 1997; Näätänen, 2003; Salisbury, 2012); and, more abstract, high-level, irregularities such as violations of grammatical categories or phonemic regularities in a sequence can elicit *D* responses (May et al., 1999; Nelken, 2004; Garrido et al., 2013). Frequency-deviant and random-frequency stimuli elicit an increasing proportion of *D* responses as the overall frequency range of the ensemble of stimuli increases well beyond the discriminability threshold (Sams et al., 1985; Garrido et al., 2013) and it has been argued that the brain can thus encode information about the statistical distribution of stimuli (Garrido et al., 2013).

The difference between *S* and *D* responses is often quantified via the mismatch negativity (MMN), which is mathematically constructed by subtracting the *S* response from the *D* one (Näätänen, 2003; Luck and Kappenman, 2011; Luck, 2014). A longstanding debate is whether the MMN (i) reflects differences in responses due to adaptation of the primary sensory system (relevant thalamic relay nuclei and primary sensory cortex, although few references discuss thalamic participation) to repeated stimuli, with the part of the system that processes $\mathcal{S}$ stimuli being driven further from its starting parameters than the part that processes $\mathcal{D}$ stimuli (Jääskeläinen et al., 2004; Näätänen et al., 2005; Garrido et al., 2009b; May et al., 2015); (ii) a reflection of separate, possibly memory-related or internal-model dependent, stimulus-comparison processes in primary cortex or higher-order cortical areas (Atienza et al., 2001; Näätänen, 2003; Jääskeläinen et al., 2004; Garrido et al., 2009b); or (iii) a combination of both adaptation and stimulus-comparison. The more abstract cases of *D* responses appear to point to the latter interpretation (Näätänen, 2003; Näätänen et al., 2005; Garrido et al., 2009b), but basic biophysics, the evolution of *S* and *D* responses during long trains, the decay of their distinction during a few-second stimulation pause, and the existence of MMN in coma imply a role for the former explanation (Schröger, 1998; Näätänen, 2003; Jääskeläinen et al., 2004; Sussman et al., 2014; May et al., 2015). In this work we take the viewpoint that both types of mechanisms are likely to be simultaneously in play, so the issue we address is which cases can be accounted for by adaptation—potentially the other, more abstract, cases then involve higher cortical areas in further processing and top-down feedback. We note that adaptation is likely to be involved in responses of nonhuman animals without involvement of language processing, as exemplified by stimulus specific adaptation studied in rats Malmierca et al. (2009), Szymanski et al. (2009), Pérez-González and Malmierca (2014); however, we restrict attention to parameters appropriate to humans in the present work.

A weakness of traditional phenomenological analyses of ER time series is that they are usually recorded at hundreds of samples per second, but quantified by noting only the amplitudes and timings of a few peaks and troughs in the waveform, or of underlying "components" that sum to produce them (Luck and Kappenman, 2011; Luck, 2014); each component is asserted to be produced by a particular "generator" with a given location and sign, normally assumed to correspond to a group of excitatory or inhibitory neurons that respond with a fixed post-stimulus delay and temporal profile (Luck and Kappenman, 2011; Luck, 2014). Hence, the MMN is often assumed to correspond to a separate component and corresponding underlying set of MMN neurons. Whilst components have characteristic timings and associated spatial structures (Luck, 2014), this procedure commonly commences analysis by omitting almost all the data points that have been recorded, which is a questionable procedure to include as a key step in any data-analysis pipeline, especially as it makes component timings very sensitive to noise near extremums. Moreover, it has been shown that these timings evolve with age, and the extremums even invert polarity during development (Kerr et al., 2010), so the very use of timings and polarities to designate features is problematic in itself. For example, use of traditional components has tended to limit adaptation theories to qualitative conclusions that particular components are attenuated by adaptation without changing their timing or polarity (Näätänen et al., 2007; Garrido et al., 2009b).

Once it is recognized that the brain is a physical system, whose dynamics generate EEG and MEG signals, including ERs, new analysis and modeling avenues are opened and it is quickly

revealed that components are not fundamental building blocks of the dynamics (Freeman, 1975; Rennie et al., 2002; Kerr et al., 2008, 2011); rather they reflect damped physical oscillations of brain activity in natural modes (Demiralp et al., 1998; Rennie et al., 2002; van Albada et al., 2010; Başar, 2012; Mukta et al., 2019; Babaie-Janvier and Robinson, 2020). In particular, it has long been noted that these signals depend on the average responses of large numbers of neurons that are detected by a given electrode or coil (Nunez, 1995; Nunez and Srinivasan, 2006; Niedermeyer and Lopes da Silva, 2011).

Neural field theory (NFT) has been developed by many authors to predict mean neural activity at scales of tenths of a millimeter and above by starting from physiological and anatomical parameters (Beurle, 1956; Wilson and Cowan, 1972, 1973; Nunez, 1974, 1995; Freeman, 1975; Lopes da Silva et al., 1976; Amari, 1977; Wright and Liley, 1994; Jirsa and Haken, 1996; Steyn-Ross et al., 1999; Robinson et al., 2002, 2004; Deco et al., 2008; Bressloff, 2012; Coombes et al., 2014; Sanz-Leon et al., 2018). In normal regimes of moderate activity, measurable signals have been shown to be approximately linearly related to perturbations of underlying neural activity from its overall mean (Nunez, 1995; Robinson et al., 1997; Deco et al., 2008). In particular, ERs reflect the activity produced by near-impulsive stimuli and the conditions of the brain that generate them can be inferred by fitting model predictions to data (Rennie et al., 2002; Kerr et al., 2011). Most significantly, the strengths of connections, or gains, between excitatory and inhibitory populations in the cortex and thalamus prove to be primary determinants of the forms of ERs and other activity phenomena (Rennie et al., 2002; Kerr et al., 2008, 2011; van Albada et al., 2010; Babaie-Janvier and Robinson, 2020). For reviews of NFT and its use in a wide range of contexts see Deco et al. (2008), Coombes et al. (2014), and Sanz-Leon et al. (2018) for example.

NFT impulse-response models of S and D ERs have been successfully fitted to data from cohorts of up to 1,500 subjects (Kerr et al., 2011). Notably, the inferred prestimulus parameters for S and D responses prove to be quite different from one another, and from those of background EEG (van Albada et al., 2010; Kerr et al., 2011). NFT has since been used to analyze the dynamics of stimulus prediction and automatic attention in the corticothalamic system (Babaie-Janvier and Robinson, 2018, 2019, 2020). This work showed that stimulus-driven gain changes due to a variety of processes such as adaptation and facilitation can improve prediction by increasing the gains that relate to salient stimuli, thereby implementing a form of attention and providing a basis to progress to higher order cognitive processes such as top-down feedback within the cortex (Gazzaley et al., 2005; Friston, 2010, 2011; Babaie-Janvier and Robinson, 2020). Moreover, it has been shown that S and D responses can be separately reproduced as impulse responses from the background EEG state when attentional gain dynamics is taken into account (Babaie-Janvier and Robinson, 2019, 2020). These results have also been interpreted in terms of engineering control theory (Ogata and Yang, 1970; Freeman, 1975; Babaie-Janvier and Robinson, 2020). In these physically based approaches the building blocks of responses are the same damped corticothalamic oscillations that account for ongoing EEG characteristics and other phenomena.

In the present work we develop a unified NFT theory, including adaptation, that can account for S and D responses to sequences of simple stimuli, including development of distinct response characteristics. This both simplifies and reduces the number of parameters required and enables a wide range of experimental conditions to be reproduced from a single model. Moreover, it predicts the entire waveform, not just peaks and troughs, and incorporates changes in amplitudes and timings of oscillations due to changes in corticothalamic parameters. It can thus potentially be fitted to experiment to determine brain-state parameters, as has been done with prior variants (Kerr et al., 2011; Babaie-Janvier and Robinson, 2020).

The structure of this paper is as follows: In section 2, we generalize our prior NFT model to calculate corticothalamic transfer functions and resulting ERs to arbitrary stimuli in the absence of higher-order cognitive processes, but incorporating adaptation effects and stimulus feature dependence. In section 3, we calibrate the model parameters by requiring that it reproduce D responses at the background EEG baseline state, and S responses when repetitively driven by impulsive stimuli that move the system away from the background state via adaptation. This provides the basis to analyze responses to arbitrary stimulus sequences. In the remainder of section 3, we test the theory's predictions for a range of experimental sequences to begin to explore which phenomena can be explained by adaptation and which likely require higher-order processing, but we stress that the literature is too vast to address all possibilities in the present work. Section 4 summarizes the main findings and outlines directions for future work.

2. MATERIALS AND METHODS

In this section we first review how ERs represent impulse responses of a linear approximation to brain dynamics and that these are directly related to system transfer functions (Freeman, 1975; Rennie et al., 2002; Kerr et al., 2008, 2011; Babaie-Janvier and Robinson, 2020). This approach has proved to be successful in the past, and has been extensively tested against experimental results (Kerr et al., 2011). We then briefly review the existing NFT corticothalamic model that is used in the analysis and generalize its dynamics to include slow adaptation processes that reflect a "memory trace." A feature map such as the tonotopic map in auditory cortex is then incorporated.

2.1. ERs as Transfer Functions

Cortical evoked responses (ERs) and magnetoencephalographic (MEG) analogs are generated primarily by perturbations in the activity ϕ_e of pyramidal excitatory cells due to dynamics in the corticothalamic system (Nunez, 1995). In the simplest approximation, we can write

$$\phi_e^{(1)}(t) = \int_{-\infty}^{t} T(t-t')\phi_n^{(1)}(t')dt', \tag{1}$$

for a purely temporal response, where T is the system linear response function, which is zero for $t < t'$ to preserve causality, and $\phi_n^{(1)}$ is the incoming non-corticothalamic stimulus to the corticothalamic system. The form in Equation (1) can be generalized to include spatial aspects but here we focus on the temporal domain in order to bring out the main aspects without undue complexity; generalization to include multiple spatial eigenmodes can be carried out in a similar way (Kerr et al., 2008; Mukta et al., 2019). Equation (1) can be Fourier transformed to yield

$$\phi_e^{(1)}(\omega) = T(\omega)\phi_n^{(1)}(\omega), \tag{2}$$

If the input in (1) is a delta function $\phi_n(t') = \delta(t' - t_0)$, one finds

$$\phi_e^{(1)}(t) = T(t - t_0), \tag{3}$$

whence we see that the ER to a delta input and the transfer function are one and the same. More generally, subsequent physical phenomena such as volume conduction, measurement effects, and post-processing also need to be taken into account in the overall transfer function from stimulus to measurement, but we omit discussion of these issues for simplicity because they do not strongly affect the time course of ERs, which is our main focus here.

In general, the transfer function itself can be changed by the stimulus, owing to a variety of fast and slow dynamical effects (Koch, 1999; Rennie et al., 1999, 2000, 2002; Robinson and Roy, 2015; Babaie-Janvier and Robinson, 2019), but to treat such effects, we must first introduce neural field theory and a model of the corticothalamic system.

2.2. Neural Field Theory of the Corticothalamic System

The baseline model that we generalize in the present work has been developed and successfully applied over many years, as mentioned in section 1. The specific formulation used here is the one from Babaie-Janvier and Robinson (2019, 2020), which we outline and generalize. Note that some of the descriptions of model elements in sections 2.2 and 2.3 are identical to those in these prior works to avoid introducing errors and ambiguities by changing the wording simply for the sake of change.

The baseline model, shown in **Figure 2**, incorporates the cortex and thalamus and their connectivities; each includes distinct populations of neurons: cortical excitatory pyramidal (e) and short-range mostly inhibitory (i) neurons, the thalamic reticular nucleus (TRN) (r), thalamic relay neurons (s), and non-corticothalamic neurons that provide external inputs (n). In this study, the relevant relay nucleus is the medial geniculate nucleus, whose projections are to primary auditory cortex. Excitatory projections to the TRN exist from thalamocortical feedforward axons and corticothalamic feedback axons, and there are inhibitory projections from the TRN onto thalamic relay neurons.

The state of each neural population a, is represented by the local mean cell-body potential $V_a(\mathbf{r}, t)$ relative to resting, the mean firing rate $Q_a(\mathbf{r}, t)$, and the outgoing axonal pulse rate field $\phi_a(\mathbf{r}, t)$. NFT averages over spatial scales below a few tenths of a millimeter to obtain equations for evolution of these dynamical variables (Wilson and Cowan, 1973; Freeman, 1975; Deco et al., 2008).

The mean firing rate Q_a has a sigmoidal response to increasing V_a, which can be approximated as (Wilson and Cowan, 1973; Freeman, 1975; Deco et al., 2008)

$$Q_a(\mathbf{r}, t) = S[V_a(\mathbf{r}, t)] = \frac{Q_{\max}}{1 + \exp\left\{-[V_a(\mathbf{r}, t) - \theta]/\sigma'\right\}}, \tag{4}$$

where θ is the mean neural firing threshold and $\sigma'\pi/\sqrt{3}$ is the standard deviation of the difference between the steady state soma voltage of individual neurons and their thresholds.

The potential $V_a(\mathbf{r}, t)$ results from all afferent neural synaptic receptors of types b and is given by

$$\hat{D}_\alpha(t)V_a(\mathbf{r}, t) = \sum_b N_{ab}s_{ab}(\mathbf{r}, t)\phi_b(\mathbf{r}, t - \tau_{ab}), \tag{5}$$

$$\hat{D}_\alpha(t) = \frac{1}{\alpha\beta}\frac{d^2}{dt^2} + \left(\frac{1}{\alpha} + \frac{1}{\beta}\right)\frac{d}{dt} + 1, \tag{6}$$

where the differential operator $\hat{D}_a$ governs the temporal response of V_{ab} to afferent pulse rate fields ϕ_b, encapsulating the rates β and α of the rise and fall, respectively, of the response at the cell body, which are assumed equal for all ab here; N_{ab} is the mean number of synapses on neurons a from neurons of type b; s_{ab} is the mean time-integrated strength of soma response per incoming spike; and $\phi_b(\mathbf{r}, t - \tau_{ab})$ is the mean spike arrival rate from neurons b, delayed by τ_{ab} due to discrete anatomical separations between different populations. The overall connection strength to neural population a from b is

$$\nu_{ab}(\mathbf{r}, t) = N_{ab}s_{ab}(\mathbf{r}, t). \tag{7}$$

Outgoing neural pulses within each population are averaged over short scales to form a field $\phi_a(\mathbf{r}, t)$ whose source is $Q_a(\mathbf{r}, t)$. This field propagates at the characteristic axonal velocity v_a and approximately obeys the damped wave equation (Jirsa and Haken, 1996; Robinson et al., 1997),

$$\hat{\mathcal{D}}_a(\mathbf{r}, t)\phi_a(\mathbf{r}, t) = Q_a(\mathbf{r}, t), \tag{8}$$

$$\hat{\mathcal{D}}_a(\mathbf{r}, t) = \frac{1}{\gamma_a^2}\frac{\partial^2}{\partial t^2} + \frac{2}{\gamma_a}\frac{\partial}{\partial t} + 1 - r_a^2\nabla^2, \tag{9}$$

where the damping rate γ_a satisfies $\gamma_a = v_a/r_a$, where r_a is the characteristic range of axons a. In the corticothalamic system, only axons of excitatory cortical pyramidal neurons are long enough to cause significant propagation effects in Equation (9). In the other populations, we assume the axonal lengths are near zero (i.e., $r_a \approx 0$) so $\mathcal{D}_a \approx 1$ which results in $\phi_a(\mathbf{r}, t) = Q_a(\mathbf{r}, t)$ for these populations.

We set $\nu_{ie} = \nu_{ee}$, $\nu_{ii} = \nu_{ei}$, and $\nu_{is} = \nu_{es}$ because the number of cortical synapses is very nearly proportional to the numbers of source and target neurons (Wright and Liley, 1996; Braitenberg

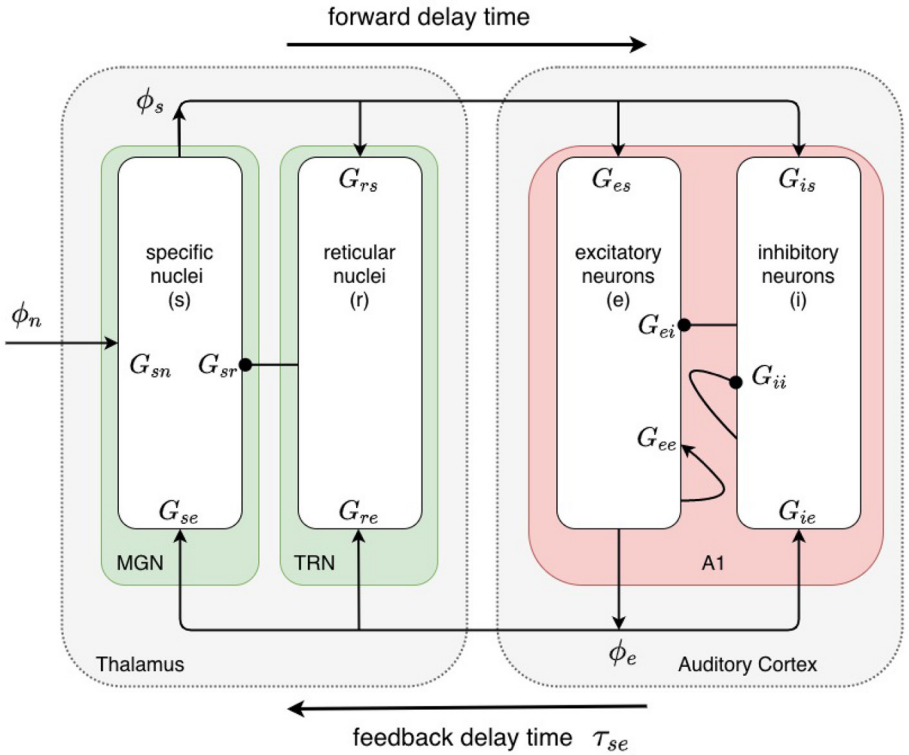

FIGURE 2 | Physiologically based corticothalamic neural field model. Arrowheads represent excitatory effects while circles depict inhibitory ones. The populations are cortical excitatory (*e*) and inhibitory (*i*) neurons, the thalamic reticular nucleus (TRN, *r*), thalamic relay neurons in the medial geniculate nucleus (MGN, *s*) that project to the primary auditory cortex (A1), and non-corticothalamic neurons responsible for external inputs (*n*).

and Schüz, 1998), assuming that synaptic types are determined by the source neurons. Forward time delays are $\tau_{es} = \tau_{is} \approx 20$ ms for thalamocortical signals and feedback delays are $\tau_{se} = \tau_{re} \approx 60$ ms for corticothalamic signals, while the other τ_{ab} are zero; time delays in the long-range excitatory axons in the cortex are included via Equation (9).

Table 1 lists nominal values of model parameters (Robinson et al., 2004) for resting EEG. These values were estimated for normal adults and they have been extensively used in the comparison with experiments, as mentioned in section 1.

2.3. Corticothalamic Transfer Functions

The above NFT equations are nonlinear in general. By setting all spatial and temporal derivatives in these equations to zero, we find spatially uniform steady-states of the system, which are interpreted as characterizing the baseline of normal activity, with firing rates that are in accord with experiments (Robinson et al., 2002, 2004). Linear perturbations from these steady states represent time dependent brain activity by which numerous experimental phenomena have been reproduced, including evoked responses (Robinson et al., 1997, 2002, 2004, 2005; Rennie et al., 2002; O'Connor and Robinson, 2004; Kerr et al., 2008, 2011; van Albada et al., 2010; Roberts and Robinson, 2012; Abeysuriya et al., 2015).

2.3.1. Perturbation Expansion

We expand the equations in section 2.2 to first order in perturbations relative to the steady state, denoting steady-state and perturbed quantities by the superscripts 0 and 1, respectively, and neglecting second- and higher-order terms. We also omit the **r** dependence from this point on, although its retention up to the present point was necessary to correctly account for the parameters γ_a. These steps give

$$Q_a^{(0)} + Q_a^{(1)}(t) = S\left[V_a^{(0)}\right] + \rho_a V_a^{(1)}(t), \tag{10}$$

$$\hat{D}_\alpha(t)\left[V_a^{(0)} + V_a^{(1)}(t)\right] = \sum_b \left[\nu_{ab}^{(0)} + \nu_{ab}^{(1)}(t)\right] \times \left[\phi_b^{(0)} + \phi_b^{(1)}(t - \tau_{ab})\right], \tag{11}$$

$$\hat{\mathcal{D}}_a(t)\left[\phi_a^{(0)} + \phi_a^{(1)}(t)\right] = Q_a^{(0)} + Q_a^{(1)}(t), \tag{12}$$

$$\rho_a = \left.\frac{dQ_a}{dV_a}\right|_{V_a = V_a^{(0)}}, \tag{13}$$

$$\hat{\mathcal{D}}_a(t) = \frac{1}{\gamma_a^2}\frac{\partial^2}{\partial t^2} + \frac{2}{\gamma_a}\frac{\partial}{\partial t} + 1, \tag{14}$$

TABLE 1 | Estimated brain parameters for normal adult humans in the alert, eyes-open state.

Quantity	Description	Resting EEG (P)	ER	Resting EEG	Unit
$Q_{\max}$	Max firing rate	250	250	250	s^{-1}
θ	Firing threshold	15	15	15	mV
σ'	Threshold spread	3.3	3.3	3.3	mV
γ_e	Cortical damping rate	116	200	116	s^{-1}
α	Inverse decay time	80	45	80	s^{-1}
β	Inverse rise time	320	180	320	s^{-1}
τ_{es}	Forward delay time	20	32	20	ms
τ_{se}	Feedback delay time	60	32	60	ms
$\phi_e^{(0)}$	Firing rate *e* neurons	16	16	16	s^{-1}
$\phi_s^{(0)}$	Firing rate *s* neurons	16	16	16	s^{-1}
$\phi_r^{(0)}$	Firing rate *r* neurons	16	16	16	s^{-1}
$\phi_n^{(0)}$	Firing rate *n* neurons	16	16	16	s^{-1}
ρ_e	For *e* neurons	4,200	4,200	4,200	$V^{-1}\ s^{-1}$
ρ_s	For *s* neurons	4,200	4,200	4,200	$V^{-1}\ s^{-1}$
ρ_r	For *r* neurons	6,300	6,300	6,300	$V^{-1}\ s^{-1}$
$G_{ee}^{(0)}$	Gain to *e* from *e*	5.9	3.1	6.8	–
$G_{se}^{(0)}$	Gain to *s* from *e*	2.5	1.18	2.5	–
$G_{ii}^{(0)}$	Gain to *i* from *i*	−8.1	−10.8	−8.1	–
$G_{sr}^{(0)}$	Gain to *s* from *r*	−1.9	−2.8	−1.9	–
$G_{es}^{(0)}$	Gain to *e* from *s*	1.7	0.74	1.7	–
$G_{sn}^{(0)}$	Gain to *s* from *n*	0.8	0.8	0.8	–
$G_{ie}^{(0)}$	Gain to *i* from *e*	5.9	3.1	6.8	–
$G_{re}^{(0)}$	Gain to *r* from *e*	1.3	3.4	1.0	–
$G_{ei}^{(0)}$	Gain to *e* from *i*	−8.1	−10.8	-8.1	–
$G_{rs}^{(0)}$	Gain to *r* from *s*	0.19	0.28	0.19	–
$G_{is}^{(0)}$	Gain to *i* from *s*	1.7	0.74	1.7	–

The first two columns give the symbol and description of each quantity. The third column shows resting-EEG values corresponding to the P state; these are also used as the initial values for ERs in which the gains adjust dynamically as part of the response. The fourth column lists static-gain ER values adapted from Table 1 of Kerr et al. (2008), previously used to match standard ERs using static gains. The fifth columns lists resting-EEG values used in previous work (Babaie-Janvier and Robinson, 2020). The final column gives units.

To zeroth order Equations (10)–(12) yield

$$Q_a^{(0)} = S\left[V_a^{(0)}\right], \tag{15}$$

$$V_a^{(0)} = \sum_b \nu_{ab}^{(0)} \phi_b^{(0)}, \tag{16}$$

$$\phi_a^{(0)} = Q_a^{(0)}. \tag{17}$$

Equations (15) and (17) can be used to eliminate the other variables in favor of the $V_a^{(0)}$, which yields the nonlinear steady-state equation (Robinson et al., 2002, 2004)

$$V_a^{(0)} = \sum_b \nu_{ab}^{(0)} S\left[V_b^{(0)}\right], \tag{18}$$

where b runs over all populations, including n.

The first order terms in Equations (10)–(12) give

$$Q_a^{(1)}(t) = \rho_a V_a^{(1)}(t), \tag{19}$$

$$\hat{D}_\alpha(t) V_a^{(1)}(t) = \sum_b \left[\nu_{ab}^{(0)} \phi_b^{(1)}(t-\tau_{ab}) + \nu_{ab}^{(1)}(t)\phi_b^{(0)}\right], \tag{20}$$

$$\hat{\mathcal{D}}_a(t)\phi_a^{(1)}(t) = Q_a^{(1)}(t), \tag{21}$$

Operation with $\hat{D}_\alpha(t)$ on both sides of Equation (21), and substitution of (19) and (20) into the result, yields

$$\hat{D}_\alpha(t)\hat{\mathcal{D}}_a(t)\left[\phi_a^{(1)}(t)\right] = \rho_a \hat{D}_\alpha V_a^{(1)}(t), \tag{22}$$

$$= \sum_b \left[G_{ab}^{(0)} \phi_b^{(1)}(t-\tau_{ab}) + G_{ab}^{(1)}(t)\phi_b^{(0)}\right], \tag{23}$$

$$G_{ab}^{(0)} = \rho_a \nu_{ab}^{(0)} = \rho_a N_{ab} s_{ab}^{(0)}, \tag{24}$$

$$G_{ab}^{(1)}(t) = \rho_a \nu_{ab}^{(1)}(t) = \rho_a N_{ab} s_{ab}^{(1)}(t)., \tag{25}$$

The gain $G_{ab}(t)$ represents the differential change in output spike rate from neurons a per unit change in input spike rate from neurons b. The net gains of populations of neurons connected serially are denoted by $G_{abc} = G_{ab}G_{bc}$ and $G_{abcd} = G_{ab}G_{bc}G_{cd}$.

2.3.2. Modulation of Synaptic Gains

Many biophysical processes can modulate neuronal coupling strengths, and hence $s_{ab}^{(1)}$ in Equation (25), dependent on current or recent activity, including plasticity, long-term potentiation/depression, adaptation, facilitation, habituation,

and sensitization (Koch, 1999; Rennie et al., 2000; Robinson and Roy, 2015; Babaie-Janvier and Robinson, 2019). We employ a general mathematical form of modulatory processes that can be applied to a broad range of specific mechanisms (Koch, 1999; Rennie et al., 1999; Robinson et al., 2002; Robinson and Roy, 2015), in which presynaptic neuronal activity locally modulates neuronal gains (dynamics driven by postsynaptic firing rate is postponed to future work, but can be treated in a similar way Rennie et al., 1999; Robinson and Roy, 2015), with

$$G_{ab}^{(1)}(t) = [g_{ab}F(t) + h_{ab}H(t)] \otimes \phi_b^{(1)}(t), \tag{26}$$

where the symbol $\otimes$ indicates a temporal convolution. Here $F(t)$ describes the temporal dynamics of the fast gain modulation on timescales of up to a few hundred ms and g_{ab} is its strength, whereas $H(t)$ is a slow adaptation process on timescales of several seconds, and h_{ab} is the corresponding strength.

Equation (26) assumes that the perturbations are small enough that a linear equation is a reasonable approximation. Furthermore, the modulation is assumed to be local in space, so the g_{ab} and h_{ab} are constant and the functional forms of $F(t)$ and $H(t)$ do not vary with position or time. For the temporal dependence of the modulation we use

$$F(t) = \eta \exp(-\eta t), \tag{27}$$
$$H(t) = \mu \exp(-\mu t), \tag{28}$$

when $t \geq 0$ and $F(t) = H(t) = 0$ for $t < 0$ to enforce causality. The positive rate constants η and μ characterize the timescales of the modulatory processes and the forms (27) and (28) are normalized to unit integral over time. Previous work found that $\eta = 25\ \mathrm{s}^{-1}$ is a reasonable choice (Rennie et al., 1999; Babaie-Janvier and Robinson, 2019), while we set $\mu = 0.65\ \mathrm{s}^{-1}$ because of the several-second timescales over which S response characteristics develop and decay.

2.3.3. Transfer Functions

The transfer function is the ratio of the output of a system to its input in the linear regime. Either the Laplace or Fourier transform can be used to determine transfer functions; we use the former with the definitions

$$\mathcal{L}[g(t)](s) = \int_0^\infty g(t)e^{-st}dt. \tag{29}$$

Application of Equation (29) to Equation (26) gives

$$\begin{aligned}\hat{\mathcal{D}}_b(s)&\left[\phi_a^{(0)} + \phi_a^{(1)}(s)\right] \\ &= L(s)\sum_b \left[G_{ab}^{(0)} + \left\{g_{ab}F(s) + h_{ab}H(s)\right\}\phi_b^{(1)}(s)\right] \\ &\quad \times \left[\phi_b^{(0)} + \phi_b^{(1)}(s)\exp\left(-s\tau_{ab}\right)\right],\end{aligned} \tag{30}$$

$$\hat{\mathcal{D}}_b(s) = (1 + s/\gamma_b)^2, \tag{31}$$
$$L(s) = (1 + s/\alpha)^{-1}(1 + s/\beta)^{-1}. \tag{32}$$

Hence, to first order

$$\begin{aligned}&\hat{\mathcal{D}}_b(s)\phi_a^{(1)}(s) \\ &\quad = L(s)\sum_b \left[G_{ab}^{(0)}e^{-s\tau_{ab}} + \phi_b^{(0)}\left\{g_{ab}F(s) + h_{ab}H(s)\right\}\right]\phi_b^{(1)}(s),\end{aligned} \tag{33}$$

$$F(s) = \eta/(s + \eta), \tag{34}$$
$$S(s) = \mu/(s + \mu). \tag{35}$$

Equation (33) expresses two types of first order responses: the first term in the square brackets represents the part of response that would occur without change to the steady-state gains, while the second term is the response due to stimulus-induced gain changes acting on the steady-state activity.

Equation (33) represents a set of coupled algebraic equations that interrelate the $\phi_a^{(1)}$. It is straightforward to eliminate the other first order quantities to obtain the transfer function to excitatory cortical neurons from retinal signals that reach the thalamus (see Babaie-Janvier and Robinson, 2018 for detailed derivation), giving

$$T_{en}(s) = \frac{\phi_e^{(1)}(s)}{\phi_n^{(1)}(s)}, \tag{36}$$
$$= \frac{A(s)}{q^2(s)r_e^2}, \tag{37}$$
$$A(s) = \frac{X_{esn}}{(1 - X_{ei})(1 - X_{srs})}, \tag{38}$$
$$q^2(s)r_e^2 = \left(1 + \frac{s}{\gamma_e}\right)^2 - \frac{1}{1 - X_{ei}}\left[X_{ee} + \frac{X_{ese} + X_{esre}}{1 - X_{srs}}\right], \tag{39}$$
$$X_{ab} = L(s)\left[G_{ab}^{(0)}e^{-s\tau_{ab}} + \phi_b^{(0)}\left\{g_{ab}F(s) + h_{ab}H(s)\right\}\right]. \tag{40}$$

2.4. Loop-Strength Representation

A useful and compact approximate representation of corticothalamic steady states and dynamics is via the normalized strengths of the corticocortical, corticothalamic, and intrathalamic loops in **Figure 2**. These are defined by (Robinson et al., 2002; Breakspear et al., 2006)

$$X = \frac{G_{ee}}{1 - G_{ei}}, \tag{41}$$
$$Y = \frac{G_{es}(G_{se} + G_{sr}G_{re})}{1 - G_{sr}G_{rs}}, \tag{42}$$
$$Z = -\frac{\alpha\beta G_{sr}G_{rs}}{(\alpha + \beta)^2}, \tag{43}$$

respectively. Originally defined with steady-state values of the G_{ab} on the right, Breakspear et al. (2006) later used instantaneous values of the time-varying $G_{ab}(t)$ to parameterize the orbits of dynamic states. Resonances in these loops are primarily responsible for the dominant frequencies in resting EEG and ERs (Kerr et al., 2008, 2011).

2.5. Control Systems Interpretation

Analysis and interpretation of the transfer function is greatly facilitated by approximating the quotient of exponential polynomials in (37) by a rational function of s. Decomposition into partial fractions then yields

$$T_{ab}(s) = \sum_{j=1}^{n} \frac{r_j}{s + p_j}; \tag{44}$$

where the poles of the system are assumed to be distinct, with

$$p_j = \Gamma_j \pm i\Omega_j, \tag{45}$$

where the damping rate is Γ_j and the frequency is Ω_j; the residues $r_j = r \pm i\Omega_r$ are

$$r_j = \lim_{s \to -p_j} (s + p_j) T_{ab}(s); \tag{46}$$

and n is the number of the poles. Some poles are associated with heavily damped modes and can be neglected, thereby allowing n to be kept small. Indeed, a 6-pole approximation ($n = 6$) has been found to be accurate to within a root-mean-square (rms) fractional error of 0.02 for 0–150 Hz for the parameters in column 3 of **Table 1** (Babaie-Janvier and Robinson, 2018). These partial fractions then are summed in pairs each of which dominates in slow/theta ($f \lesssim 5$ Hz), alpha (5 Hz $\lesssim f \lesssim$ 15 Hz), or beta (15 Hz $\lesssim f$) frequency regimes, respectively. This gives

$$T_{bn}(s) \approx T_{bn}^{\ell}(s) + T_{bn}^{\mathcal{A}} + T_{bn}^{\mathcal{B}}(s), \tag{47}$$

where $b = e, i, r, s$ and T_{bn}^{ℓ}, $T_{bn}^{\mathcal{A}}$, and $T_{bn}^{\mathcal{B}}$ are the sums over the pairs of poles that represent responses in the low, alpha, and beta frequency ranges, respectively. We denote the three corresponding partial transfer functions by $T_{ab}^{\mathcal{F}}(s)$ for $\mathcal{F} = \ell, \mathcal{A}, \mathcal{B}$, with

$$T_{ab}^{\mathcal{F}}(s) = \frac{r_j}{s + p_j} + \frac{r_{j+1}}{s + p_{j+1}}, \tag{48}$$

where the poles j and $j + 1$ form a pair. Note that the poles and residues depend on $\mathcal{F}$, a, and b, but we have not shown this explicitly to avoid unduly cumbersome notation.

Use of the partial fraction representation makes inversion of the Laplace transform straightforward. Two possibilities occur: either the two poles represent damped nonzero-frequency oscillations, and are complex conjugates, or they represent purely damped responses and are both real. In the oscillatory case, $p_{j+1} = p_j^*$ and $r_{j+1} = r_j^*$ so that the time-domain response is real. In this case, for a delta-function stimulus at $t = 0$,

$$T^{\mathcal{F}}(t) = 2|r_j| \exp(-\Gamma_j t) \cos(\Omega_j t - \psi), \tag{49}$$

where we have written $r_j = |r_j| e^{i\psi}$. In the purely damped case, r_j and r_{j+1} are real but not equal and likewise for p_j and p_{j+1}. This gives

$$T_{ab}^{\mathcal{F}}(t) = r_j e^{-\Gamma_j t} + r_{j+1} e^{-\Gamma_{j+1} t} \tag{50}$$

Equation (49), in particular, shows that analyses of ERs in terms of damped sinusoids (Freeman, 1975; Demiralp et al., 1998; Başar, 2012) are not just instances of compact phenomenology, but rest on the dynamics embodied in resonances of the transfer function. At a deeper level, each pair of poles can be interpreted as implementing a control systems data filter—specifically a PID (proportional-integral-derivative) filter—that predicts incoming signals based on signal value, rate of change, and integrated time history (Ogata and Yang, 1970; Babaie-Janvier and Robinson, 2018, 2019). Using this formulation, dynamic gain changes have been interpreted as improving prediction by implementing attention to salient features (Babaie-Janvier and Robinson, 2019).

2.6. Feature Map

The final generalization we require to the model is to incorporate different stimulus features, such as frequency. These are mapped to slightly different locations in the auditory cortex, which will adapt differently, so we need to distinguish them by a label σ. Here we assume that the ER measuring electrode responds equally to all relevant locations, although this is an assumption that could later be relaxed. In place of Equation (1) we write

$$\phi_e^{(1)}(\sigma, t) = \int_{-\infty}^{t} T(t - t') \int w(\sigma - \sigma') \phi_n^{(1)}(\sigma', t') d\sigma' dt', \tag{51}$$

where the weight function w quantifies the discriminability of stimuli; a suitable form is

$$w(\sigma - \sigma') = \exp\left[-\frac{(\sigma - \sigma')^2}{2(\Delta\sigma)^2} \right]. \tag{52}$$

In Equation (51) we have assumed that T does not depend explicitly on σ, but such a dependence could easily be included. Aside from the issue of discriminability, and the inclusion of σ, the bulk of the above analysis is unchanged. However, the weight function w implies that stimuli σ' within $\sim \Delta\sigma$ influence the dynamics at σ.

In the case of auditory stimuli, σ can be viewed as the frequency and a very short sinusoidal stimulus at t_0 with frequency corresponding to σ_A has

$$\phi_n^{(1)}(\sigma', t') \approx \delta(\sigma' - \sigma_A) \delta(t' - t_0). \tag{53}$$

There is no sinusoidal time dependence in Equation (53) because the input pathway via the cochlea and superior colliculus translates each frequency to a point in the tonotopic map, without retaining its waveform. Using Equation (53), Equation (51) simplifies to

$$\phi_e^{(1)}(\sigma, t) \approx w(\sigma - \sigma_A) T(t - t_0). \tag{54}$$

If $\mathcal{S}$ and $\mathcal{D}$ stimuli are fully discriminable the only relevant values of $w(\sigma - \sigma')$ are 1 and 0; i.e., there is no cross-talk.

Note that, although we have assumed that σ is a scalar here, corresponding to stimulus frequency and the tonotopic auditory map, more generally it could be a vector of feature attributes, including frequency, interaural delay, intensity, and other quantities, with different sensoricortical maps (Herdener et al., 2013).

3. RESULTS

We are now in a position to analyze the different responses to frequent and rare stimuli, which will ultimately evoke S and D responses, respectively, in a long sequence. In this section we assume that these two stimulus types are fully discriminable so we do not need to include the parameter $\Delta\sigma$ of section 2.5 and simply denote the frequent and rare stimuli by $\sigma = \mathcal{S}$ and $\sigma = \mathcal{D}$, respectively. We optionally denote the nth consecutive stimulus of a given type by the subscript n; i.e., $\mathcal{S}_n$ describes the nth consecutive $\mathcal{S}$ stimulus and $\mathcal{D}_n$ denotes the nth consecutive $\mathcal{D}$ stimulus. We write the corresponding system responses, which can differ between presentations of the same stimulus type, due to adaptation, as $S_n(t)$ and $D_n(t)$ but omit the argument t when referring to the entire response. Standard responses rapidly approach a limiting form $S_\infty(t)$ after a few (typically about $n = 5$) presentations of $\mathcal{S}$, with little change thereafter (Cowan, 1984; Winkler et al., 1993; Loveless et al., 1996). The MMN between the nth D response and the mth S response, for example, is defined to be the difference

$$\mathrm{MMN}(D_n, S_m, t) = D_n(t) - S_m(t), \tag{55}$$

with analogous definitions for other pairs of responses. Most commonly the MMN is defined to be $\mathrm{MMN}(D_1, S_\infty, t)$ the version obtained by subtracting the limiting form $S_\infty(t)$ of the standard response from $D_1(t)$. Note that the first responses to both stimuli are identical because both are novel: $S_1(t) = D_1(t)$.

In this section we first explain the formulation of differential adaptation to $\mathcal{S}$ and $\mathcal{D}$ stimuli. We then calibrate the parameters g_{ab} and h_{ab} of the model by matching their predictions for $\phi_e^{(1)}(t)$ to typical oddball data before applying the results to a variety of other stimulus sequences in later subsections. These results allow exploration of which MMN effects can be accounted for by adaptation in the primary auditory cortex and the medial geniculate nucleus of the thalamus.

3.1. Contribution of Adaptation to Responses

The central idea used here is that the system initially occupies the same prestimulation state as the one corresponding to background EEG in the period before any stimuli have been presented. We label this P in the schematic space of gains in **Figure 3**. In the absence of adaptation, each stimulus causes transient gain changes due to the term $F(t)$ in Equation (26). However, because the time constant of $H(t)$ is roughly 40 times larger, resulting changes due to that process may not have fully relaxed by the time the next stimulus arrives. Hence, the system will be pushed to the location corresponding to S_∞ by a long series of $\mathcal{S}$ stimuli (blue curve in **Figure 3**), eventually oscillating around a point where

$$\Delta G_{ab} \sim g_{ab} \left\langle \phi_b^{(1)}(t) \right\rangle_\eta + h_{ab} \left\langle \phi_b^{(1)}(t) \right\rangle_\mu , \tag{56}$$

where the angle brackets denote an average over the most recent time interval of order $1/\eta$ or $1/\mu$, as indicated by the subscript; these averages are nonzero in general because incoming delta-function stimuli have a nonzero mean. The first average decays within tens of ms, and can usually be neglected by the time the next stimulus arrives, but the second can be significant in a train of $\mathcal{S}$ stimuli. In contrast, in the case of $\mathcal{D}$ stimuli, which come more rarely, the system gains will have time to relax almost to P in the interim (orange curve in **Figure 3**). Hence, we argue that deviant responses occur from near-P conditions, whereas standard responses occur from a location in parameter space that shifts gradually toward the parameters of S_∞ over several stimuli. We see from **Figure 3** that the S gain response doesn't get a chance to relax back to P due to the shortly-spaced stimuli whereas the D gain response is triggered relatively rarely and decays back almost to P between stimuli. Both responses to the first stimulus are the same.

One key point above is that each ER starts from the relevant time-evolving gains after the previous stimulus, or else all responses would simply be added with the same functional form. It might be objected that this amounts to retention of a second order term in the perturbation expansion and that we should therefore retain all second order terms. However, although this point of view is formally correct, it is not necessary to retain the other second-order terms because the long time constant of $H(t)$ effectively "promotes" its formally first-order effects by integrating over several seconds to produce changes that are comparable with zeroth-order terms. It is only after times $\gtrsim 5 - 10$ s without stimuli that these changes decay and the system again approaches the state P. Indeed, Equations (27) and (28) show that terms arising from $H(t)$ are of order $\exp[(\eta - \mu)t]$ larger than those arising from $F(t)$ a time t after a stimulus; this can be a very large factor and these terms cannot generally be neglected relative to zeroth-order gains.

3.2. Parameter Calibration

A comparison of the present pre-stimulus gain parameters $G_{ab}^{(0)}$ corresponding to the baseline EEG state (P) with those used in previous work that reproduced Standard ERs with static-gains (Kerr et al., 2008) and with modified gains (Babaie-Janvier and Robinson, 2020) can be done by comparing the third column of **Table 1** with the fourth and fifth columns, respectively. The present P parameters are identical to those of Babaie-Janvier and Robinson (2020) except for slight changes in $G_{ee}^{(0)}$ and $G_{re}^{(0)}$, and are mostly larger than those used by Kerr et al. (2008) except for G_{rs} and G_{re}. Of course, we do not expect exact correspondence because the previous studies did not include slow adaptation via $H(t)$.

Based on previous NFT analysis of standard and deviant responses (Kerr et al., 2008, 2011) and recent work which analyzed the role the different gains play in determining ER features (Babaie-Janvier and Robinson, 2019), we derived static-gain ERs above by adjusting the $G_{ab}^{(0)}$ in the transfer function component of Equation (40) and setting $g_{ab} = 0$ and $h_{ab} = 0$. These served as benchmarks for the S_∞ and D responses shown in **Figure 4**. The key difference between these two curves is the relative increase of the corticothalamic loop gains $G_{es}^{(0)}$ and $G_{se}^{(0)}$ as well as the top-down pathway $G_{re}^{(0)}$ for D relative to S,

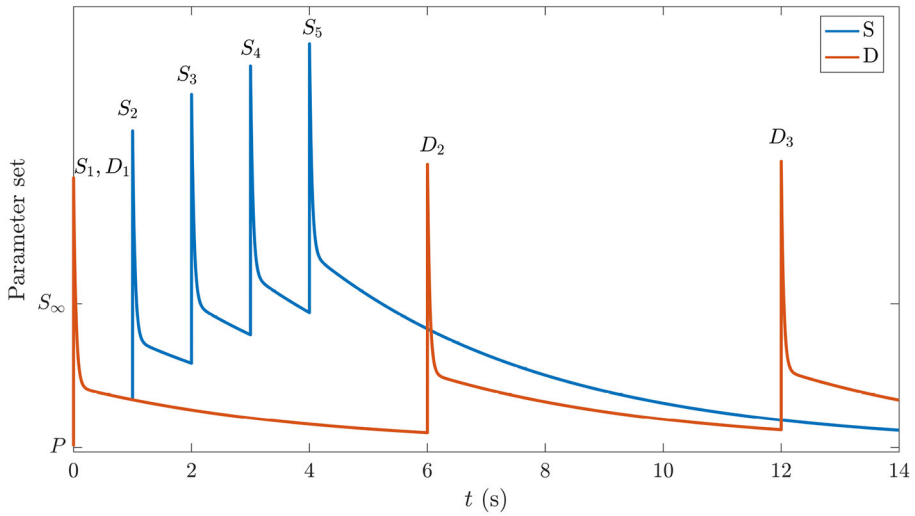

FIGURE 3 | Schematic of gain responses of the pre-stimulation (P) state and the trajectories followed by the different parts of the corticothalamic system that process the $\mathcal{S}$ and $\mathcal{D}$ stimuli. The vertical axis schematically represents the response of the set of system gains, with starting point P and the asymptotic value in the S_∞ response labeled. The horizontal axis indicates time and is labeled with stimulus types and numbers. The blue curve corresponds to stimuli $\mathcal{S}_1, \ldots, \mathcal{S}_5$ and the orange curve corresponds to $\mathcal{D}_1, \ldots, \mathcal{D}_3$.

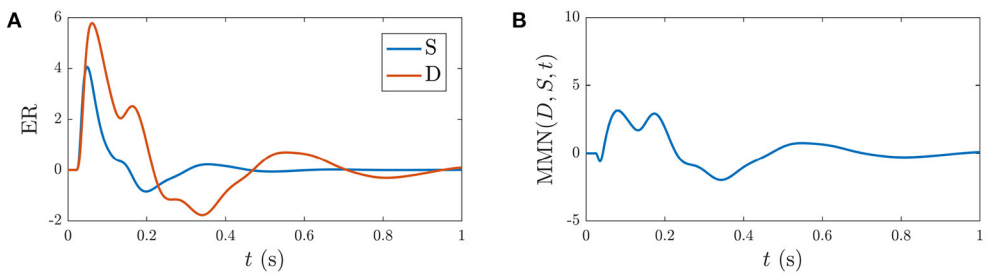

FIGURE 4 | Static-gain ERs calibrated to closely approximate typical *S* and *D* responses from Kerr et al. (2008), Kerr et al. (2011), and Babaie-Janvier and Robinson (2019), as indicated in the legend, which were used as benchmarks for the S_∞ and *D* responses (outlined in section 3.2). **(A)** Benchmark *S* and *D* curves used to derive gain modulations in **Table 2**. **(B)** Corresponding MMN(D, S, t).

resulting in the presence of the N2 and P2 features in *D*. It is worth noting that a variety of standard and deviant responses are found in the literature, due in part to slightly varied experimental conditions, and that our present aim is not to reproduce a particular set of response curves exactly, but rather to incorporate common features such as the standard response being reduced in amplitude and lacking the widely established N2 deflection that is often seen and interpreted as contributing to the MMN (Näätänen et al., 1978, 1989a; Sams et al., 1984).

Building on findings and recent estimates of local feedback modulation in g_{ab} that give rise to ERs (Babaie-Janvier and Robinson, 2019), we calibrate the model parameters g_{ab} and h_{ab} by minimizing the error between the benchmark curves and the model-calculated curves such that the resulting activity resembled the *S* benchmark upon a number of closely-spaced, consecutive stimulation, and resembled the *D* benchmark upon less-frequent stimulation. These calibrated parameters are shown in **Table 2** and discussed with respect to their contributions to the activity responses in the next section.

Here we analyze the local feedback strengths g_{ab} and h_{ab}, presented in **Table 2**, that give rise to successive S_n and D_n responses. The slow adaptation contributions, parameterized by the h_{ab}, determine the gradual evolution of the baseline of responses due to a series of stimuli over several seconds, while fast gain changes parameterized by the g_{ab} primarily determine the shape of the responses on the few-hundred ms scale, with differences between S_∞ and D_1 resulting from their different starting points.

The present work allows us to distinguish the parts of S_∞ and *D* responses that are specifically due to fast gain modulations and slower adaptation. **Figure 5A** shows the static gain baseline ER (starting at P) along with the fast and slow gain modulation contributions to the *S* response,

$$\Delta_g S_\infty(t) \;=\; S_\infty(t) - S_\infty(t)\big|_{h_{ab}=0}\,, \tag{57}$$

$$\Delta_h S_\infty(t) \;=\; S_\infty(t) - S_\infty(t)\big|_{g_{ab}=0}\,, \tag{58}$$

TABLE 2 | Characteristic fast and slow gain response parameters and their percentage change relative to the static baseline parameters $G_{ab}^{(0)}$ (which are dimensionless).

Parameter	Value	$\Delta\lvert G_{ab}^{(0)}\rvert$ (%)
FAST CHANGE CONTRIBUTION		
ηg_{ee}	−0.1157	−2
ηg_{ei}	0.7684	9
ηg_{es}	0.7419	44
ηg_{se}	0.1047	4
ηg_{sr}	−0.1390	7
ηg_{rs}	0.1149	60
ηg_{re}	0.2822	22
ADAPTATION CONTRIBUTION		
μh_{ee}	0.4390	7
μh_{ei}	− 1.1180	14
μh_{es}	−0.0890	−5
μh_{se}	−0.3299	−13
μh_{sr}	0.0053	0.3
μh_{rs}	0.0018	1
μh_{re}	−0.0969	−7

In each case, the parameter is multiplied by its inverse timescale to obtain a characteristic contribution, as seen from Equations (27) and (28).

which are the differences between the total S_∞ response and the S_∞ responses due to setting all $h_{ab} = 0$ and $g_{ab} = 0$, respectively. The combined effect of these individual processes (green broken line) is calculated according to

$$\Delta_{g+h} S_\infty(t) \;=\; S_\infty(t) - S_\infty(t)\big|_{g_{ab}=0, h_{ab}=0}\,, \tag{59}$$

and combines with the baseline ER (blue line) to give the S_∞ response (orange line). We note that fast gain modulations act to decrease the N1 and N2 deflections in the baseline ER, whereas the adaptation contribution is smaller in magnitude and predominantly increases the N1 deflection. The overall local gain modulation contribution decreases the N1 and N2 deflections in the baseline ER and produces a deflection at ≈200–300 ms, giving the S_∞ response.

Analogously, **Figure 5B** shows the static gain baseline ER (starting at P) alongside the fast gain and adaptation contributions $\Delta_g D$ and $\Delta_h D$ which sum with the baseline ER to give the D response. We note that fast gain modulations exhibit a P2 deflection at a slightly earlier latency compared to the D response, whereas the adaptation contribution exhibits small N1 and N2 deflections. The overall local gain modulation contribution thus slightly increases the N1 amplitude, reduces the N2 amplitude, and introduces a large P2 deflection at ≈200–300 ms in the baseline ER state to give rise to the D response.

Figure 5C explores how these distinct gain contributions affect the MMN. Calculating $\mathrm{MMN}(D - \Delta_g D, S_\infty - \Delta_g S_\infty, t)$ isolates the part of the MMN that is caused by fast gain modulations (broken magenta line) and calculating $\mathrm{MMN}(D - \Delta_h D, S_\infty - \Delta_h S_\infty, t)$ exposes the part of the MMN that is caused by adaptation (broken black line). As expected, without adaptation there is negligible distinction between the two responses at the short timescales shown, so the MMN is zero. The difference between the $\mathrm{MMN}(D, S_\infty, t)$ curve (blue line) and the adaptation contribution (black broken line) reveals how turning on the fast gain modulations affects the resultant MMN shape; the fast gain modulations act to slightly increase the amplitude of N1 and N2 features of the MMN.

As can be seen in **Table 2**, the dominant percentage changes in fast gain dynamics occur in the bottom-up pathways g_{es} and g_{rs}, and to a lesser extent in the top-down pathway g_{re} whereas the dominant percentage changes in slow adaptation occur in the cortical and top-down pathways h_{ei} and h_{se}, and to a lesser extent in the cortical, bottom-up, and top-down pathways h_{ee}, h_{es}, and h_{re}. These findings suggest that cortical and top-down pathways play enhanced roles in adaptation to produce S responses.

These results generalize recent work that only considered fast change contributions to local feedback modulations g_{ab} (Babaie-Janvier and Robinson, 2019). In agreement with that work, we find a decrease of the inhibitory cortical and intrathalamic gains g_{ee} and g_{sr} and an increase in top-down corticothalamic gain g_{re}. Although, in contrast, we found increases in the cortical inhibitory, top-down corticothalamic, and bottom-up thalamocortical gains g_{ei}, g_{se}, and g_{es} rather than decreases.

3.3. Development and Decay of Responses

Here we analyze the development and decay of the S_∞ and D responses during long trains that are distinguishable by their different ISIs. **Figure 6A** shows the evolution of $S_n(t)$ toward the $S_\infty(t)$ response for $n = 1, \ldots 5$ with 1 s ISI. As expected, the initial response to the first standard stimulus $\mathcal{S}_1$ is a deviant response, $S_1 = D_1$, which does not exhibit as strong an N2 deflection as the benchmark in **Figure 4A** and the latency of the N2 peak is slightly earlier (≈ 10 ms) than the benchmark due to the effects of the g_{ab} and h_{ab}. Importantly, D_1 contains the N1 and P2 deflections, as seen in experiments (Garrido et al., 2009b). The gradual adaptation of S_n with n is also seen; little further change is found to occur after $n = 5$. The response $S_5(t) \approx S_\infty(t)$ shows a reduction in N1 and N2 amplitudes relative to D_1, which has also been experimentally observed (Sams et al., 1984; Cowan et al., 1993; Garrido et al., 2009b).

When the ISI is increased to the typical value for $\mathcal{D}$ stimuli in auditory oddball paradigms almost identical $D_n(t)$ responses emerge, independent of n, as seen in **Figure 6B**, which shows the system response to five stimuli with 6 s ISI. This occurs because there is sufficient time for the parameters to relax very nearly to the prestimulation state P between stimuli. We thus refer to deviant responses as $D(t)$ without subscript from now on unless otherwise specified.

The development of $\mathrm{MMN}(D_1, S_n, t)$ vs. n, defined in Equation (55), is seen in **Figure 6C**. We see that the MMN is zero at $n = 1$ and grows with n as adaptation occurs in response to successive $\mathcal{S}$ stimuli. This MMN is positive in the 20–180 ms range, in agreement with experimental findings (Cowan et al., 1993; Garrido et al., 2007, 2009b; Näätänen et al., 2007).

The gain dynamics corresponding to the stimulus sequences in **Figures 6A,B**, followed by a stimulus-free interval, are illustrated in the first two columns of **Figure 7**, which further

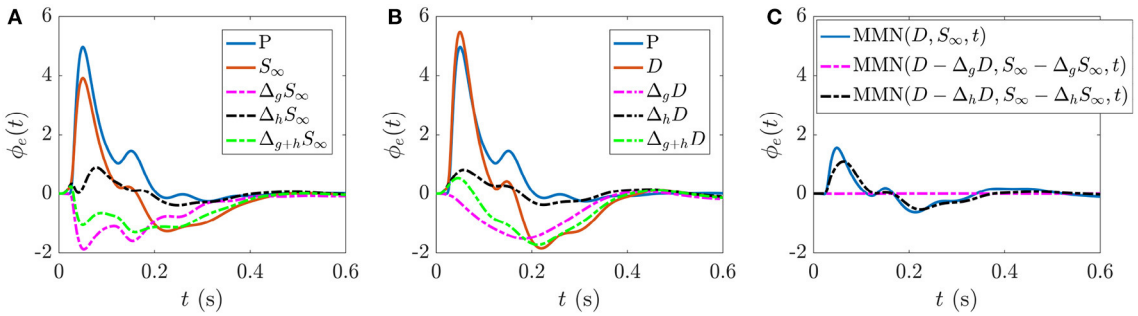

FIGURE 5 | ERs and gain contributions corresponding to different responses. **(A)** Static gain baseline ER starting from P, fast gain modulation $\Delta_g S_\infty$ and adaptation $\Delta_h S_\infty$ contributions which, when added together ($\Delta_{g+h} S_\infty$) and combined with the baseline ER give the S_∞ response. **(B)** Static gain baseline ER starting from P, fast gain modulation $\Delta_g D$, and adaptation $\Delta_h D$ contributions which, when added together ($\Delta_{g+h} D$) and combined with the baseline ER give the D response. **(C)** MMN(D, S_∞, t) and the corresponding isolated fast gain modulation (magenta broken line) and adaptation (black broken line) contributions.

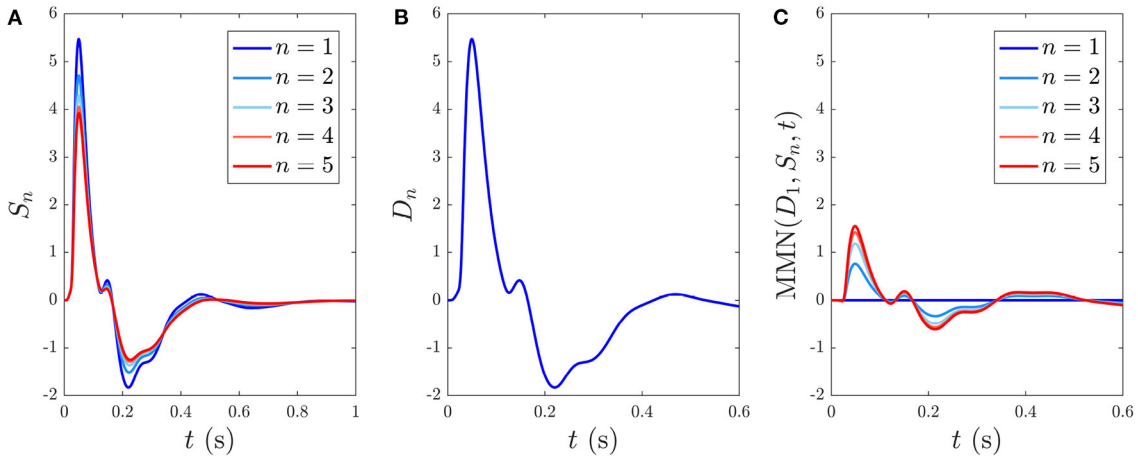

FIGURE 6 | Development of responses $S_\infty(t)$ and $D(t)$ over multiple stimuli due to their different ISIs, with stimulus numbers indicated in the legends. **(A)** $S_n(t)$ for $n = 1, \ldots, 5$ with a 1 s ISI. **(B)** $D_n(t)$ for $n = 1$, which is indistinguishable on this scale from cases with $n > 1$ with a 6 s ISI. **(C)** The MMN(D_1, S_n, t) corresponding to the sequential S_n responses from **(A)**.

underlines how the development of $S_\infty(t)$ involves the gains approaching a new baseline as adaptation occurs. In the stimulus-free interval after $t = 5$ s the gains decay back to their P-values on a timescale of roughly 5 s.

Our expectation that D responses should occur from near-P conditions while S_∞ responses occur from a shifted starting point in parameter space is confirmed by examining prestimulus gains at the moment of each successive stimulus in the $\mathcal{S}$ and $\mathcal{D}$ streams above. **Figure 7** (third column) shows these gains just before each stimulus vs. stimulus number for the $\mathcal{S}$ and $\mathcal{D}$ streams; it is evident that the $\mathcal{S}$ stream pushes the system further from P while the $\mathcal{D}$ stream allows it to relax back to near P at the time of the next such stimulus.

We now explore the S and D responses from **Figure 6** in XYZ space, illuminating how cortical, corticothalamic, and intrathalamic feedback loops contribute to such dynamics. We first analyze the D response from **Figure 6B** and then explore the development of $S_\infty(t)$ from **Figure 6A**. **Figure 8** shows the sequence of evoked responses and their corresponding $X(t)$, $Y(t)$, and $Z(t)$ (first column) alongside the trajectory they traverse in XYZ space, and the XY, YZ, and XZ planes (second column). A video of this activity is provided in **Supplementary Video 1**. As can be most easily seen in the video but also evident in this figure, each D response follows almost the same path in XYZ space. Each loop of the trajectory is characterized by an initial sharp increase of Z from $\approx$ 0.05 to 0.3 until $t \approx 25$ ms post-stimulus, followed by a decrease of all coordinates. Then Y starts to increase at $t \approx 70$ ms while X and Z continue to decrease until $t \approx 100$ ms, at which point there is a short-lived rise in Z. Then X increases substantially, taking the system back to its starting point; Y also increases during this phase, becoming briefly positive before peaking at $t \approx 380$ ms and returning to the starting value of $Y \approx 0$. Around $t = 600$ ms the trajectory displays a small excursion from near its starting point as X travels further in the positive X direction (from $X \approx 0.65 - 0.66$, most evident in the XY and XZ plane plots) before decaying back to

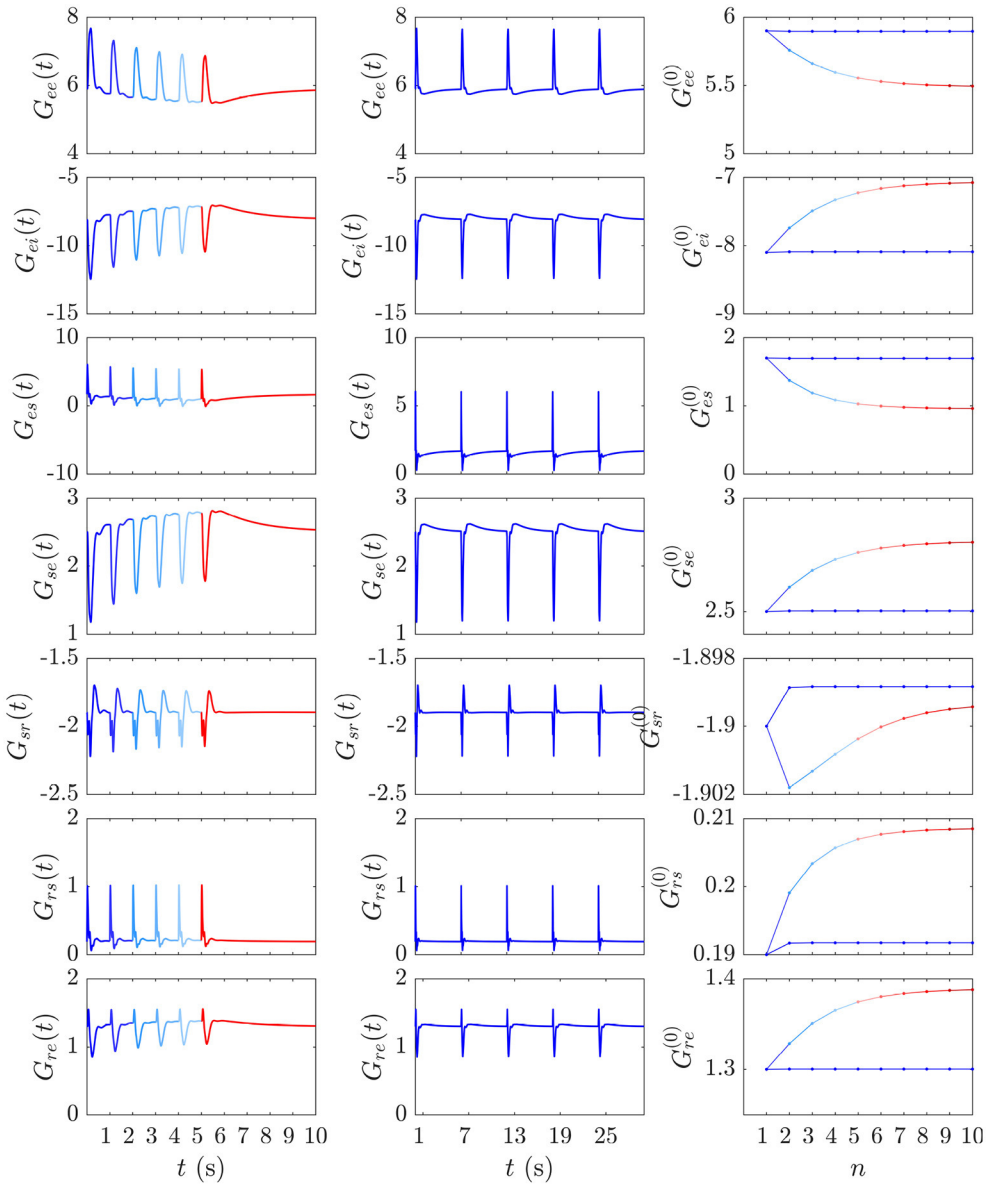

FIGURE 7 | Evolution of the instantaneous $G_{ab}(t)$ and prestimulus $G_{ab}^{(0)}$ gains corresponding to the two cases from **Figures 6A,B**. The first and second columns show the gain evolution $G_{ab}(t)$ during the development of the S_∞ response (blue to red) and the successive D responses (blue), respectively. The third column depicts the corresponding pre-stimulus gains $G_{ab}^{(0)}$ immediately prior to the successive $\mathcal{S}$ and $\mathcal{D}$ stimuli vs. stimulus number n.

its starting point by $t \approx 4$ s. This is because of the relatively long-lasting shifts in $G_{ee}(t)$ and $G_{ei}(t)$ [which determine $X(t)$] in the second column of **Figure 7**. This suggests that during ERs, shifts in intracortical feedback strengths take longer to return to baseline than corticothalamic and intrathalamic ones.

We now examine the development of the S_∞ response from **Figure 6A**. As can be seen from the third column of **Figure 7**, the development of this response is accompanied by a 7% decrease in G_{ee}, 13% decrease in $|G_{ei}|$, 44% decrease in G_{es}, 12% increase in G_{se}, $\approx$ 0% change in G_{sr}, 10% increase in G_{rs}, and a 7% increase in G_{re}. **Figure 9** shows the evoked response corresponding to the development of $S_\infty(t)$, the corresponding $X(t)$, $Y(t)$, and $Z(t)$ (first column) alongside the trajectory it traverses in XYZ space, and the XY, YZ, and XZ planes (second column), with arrows indicating the direction of motion. A video of this activity is provided in **Supplementary Video 2**. The first thing to note is that, as expected, the first orbit (deep blue) is identical to $D(t)$ from **Figure 8**. As adaptation occurs, the starting point for activity moves roughly in the positive X direction, as seen in **Figure 9**, with smaller shifts in Y and Z. The dominant shift in X is a consequence of the abovementioned fact that X takes the longest to decay back to its baseline value; i.e., the

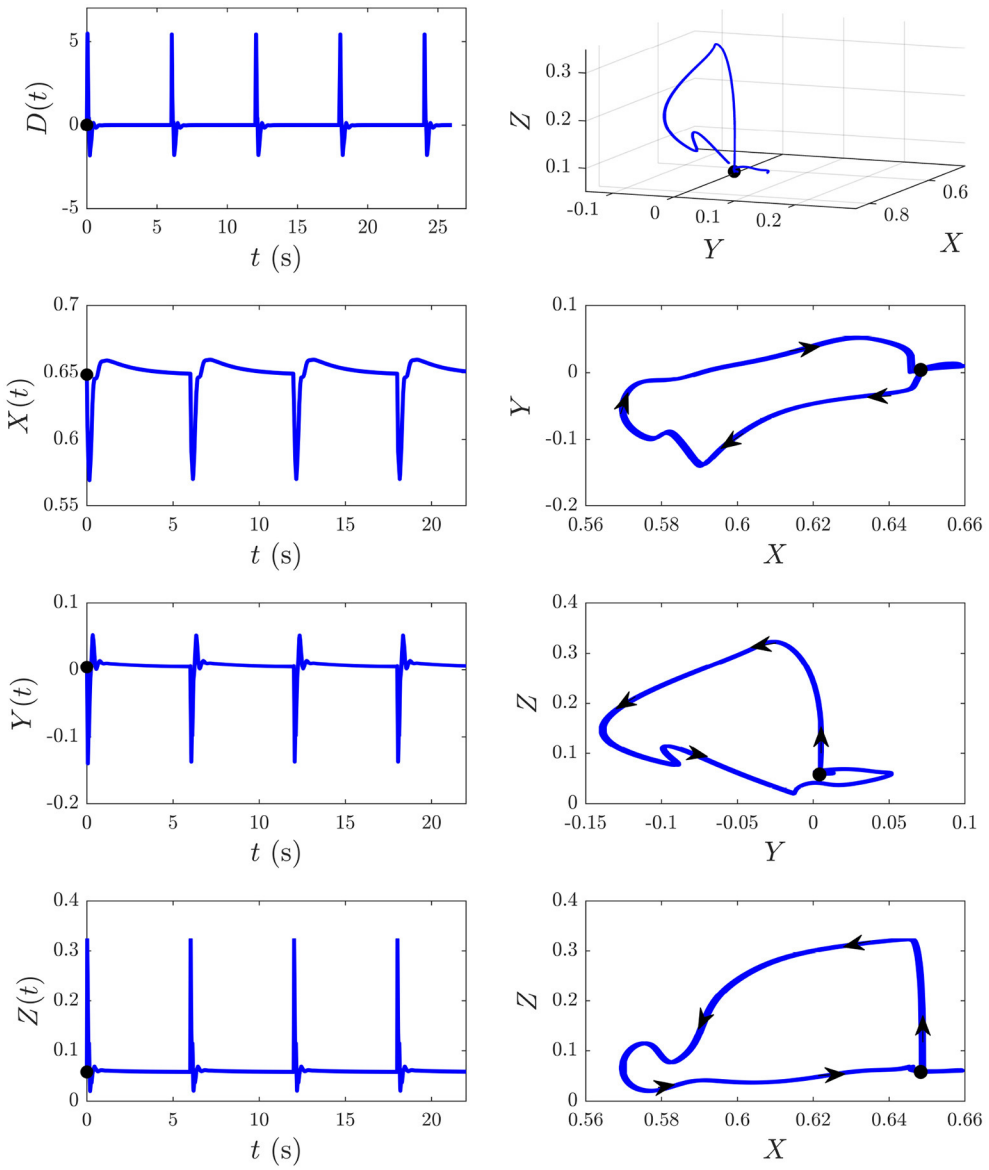

FIGURE 8 | The D responses from **Figure 6B** and their associated trajectories in XYZ space (X, Y, and Z are intracortical, corticothalamic, and intrathalamic loop strengths, respectively). The black dot indicates the start point $t = 0$. First column shows the sequence of evoked responses at 6 s ISI and corresponding $X(t)$, $Y(t)$, and $Z(t)$. Second column shows this trajectory in XYZ space and in the XY, YZ, and XZ planes, where the arrows indicate the direction of motion.

starting points for successive stimuli shift further along the X axis. It also implies that the dominant changes in brain dynamics occurring during adaptation to $\mathcal{S}$ stimuli involve increased intracortical feedback followed by increased corticothalamic and intrathalamic feedback, respectively.

3.4. Sequences of Stimuli

Here we move on from investigating the S and D responses individually and analyze a variety of different sequences of $\mathcal{S}$ and $\mathcal{D}$ stimuli that have been implemented experimentally. Recall that in the present analysis $\mathcal{S}$ and $\mathcal{D}$ stimuli are fully distinguishable and there is no cross-talk in Equation (54) so we can model each response separately. After several consecutive identical stimuli are interrupted by an occurrence of the other type of stimulus, n is reset to 1 here, so a given n can correspond to different situations, depending on what has occurred in the previous 5–10 s. We also introduce the following additional notation for brevity: a stream of m consecutive $\mathcal{S}$ stimuli is written as $(m\mathcal{S})$. It is important to note that the responses to individual stimuli in $(m\mathcal{S})$ are not identical and depend on the history. For example, in the sequence $(m\mathcal{S})\mathcal{D}_1(m\mathcal{S})$, the first set of S responses are not identical to the corresponding members of the second set because the latter start from a more adapted corticothalamic region than the first. (The occurrence of the single $\mathcal{D}_1$ stimulus does not give enough time

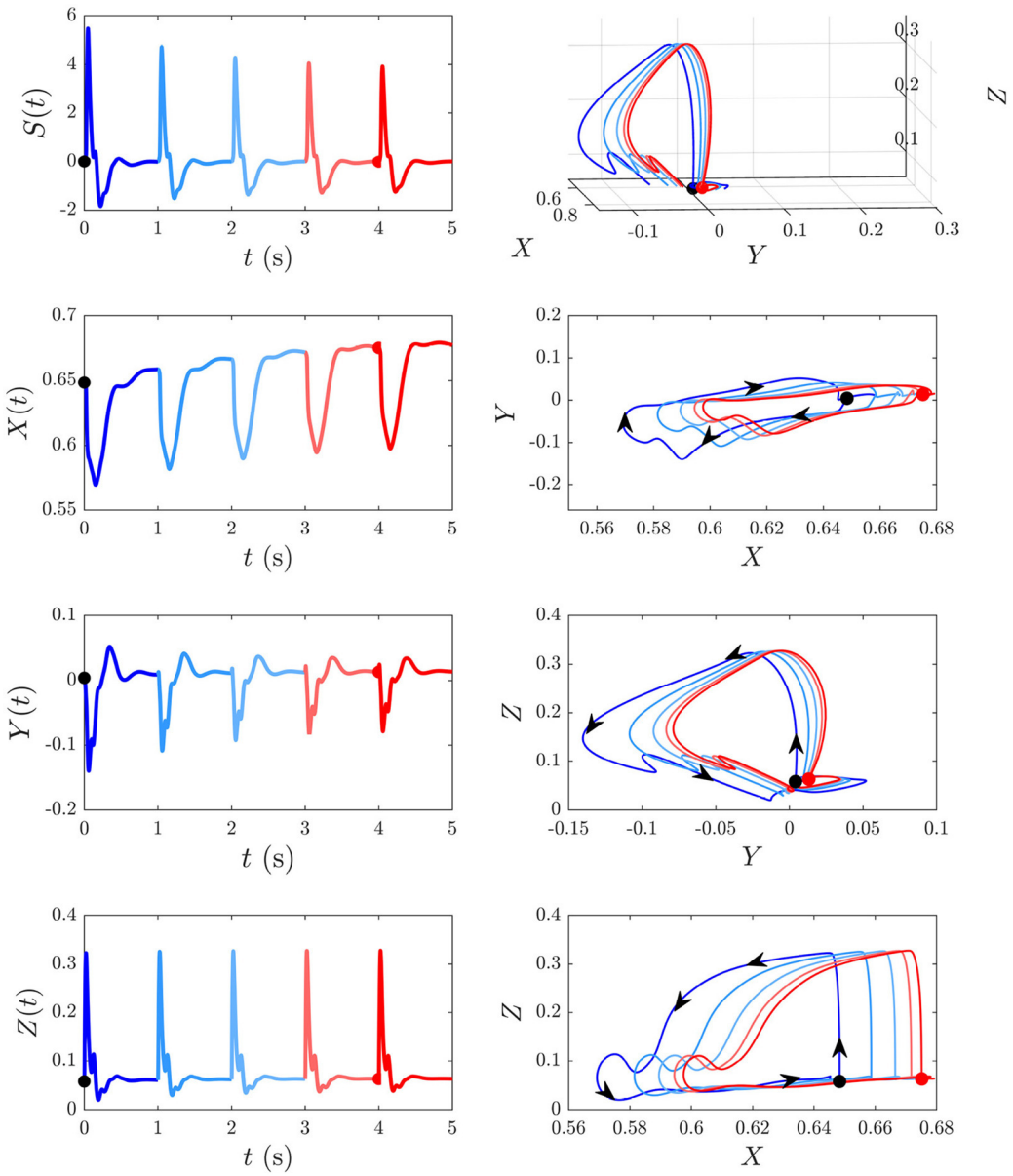

FIGURE 9 | Development of the S_∞ response from **Figure 6B** and its associated trajectories in *XYZ* space. The black dot indicates the start point at baseline P and the red dot indicates the start point S_∞. First column shows the evoked response from **Figure 6**, and the corresponding $X(t)$, $Y(t)$, and $Z(t)$ curves (X, Y, and Z are intracortical, corticothalamic, and intrathalamic loop strengths, respectively). Second column shows this trajectory in *XYZ* space, and in the *XY*, *YZ*, and *XZ* planes where the arrows indicate the direction of motion.

for the next $\mathcal{S}$ stimulus to start from the same baseline as the very first stimulus in the sequence, so the residual adaptation arising from the first group of $\mathcal{S}$ stimuli is still significant when the second set commences.) The present analysis provides a firm, physiologically-based footing from which to analyze to what extent adaptation plays a role in the many different features of the MMN, and in which case higher-order processes are involved. Because digitized experimental data are not available from the literature we cannot fully calibrate our model to individuals, so the comparisons presented are necessarily semiquantitative.

3.4.1. Sequence of Standards With Occasional Single Deviants

We first model an early study (Sams et al., 1984) which presented standard tones (1,000 Hz) 90% of the time and deviant tones (1,250 Hz) 10% of the time in random order and calculated the MMN corresponding to the first deviant tone after at least four standard tones. Each block contained 500 tones with 1 s ISI and the ERs to standard and deviant tones were separately averaged. We simulate a block by generating sequences of random $\mathcal{S}$ and $\mathcal{D}$ stimuli with the above probability distribution, only considering

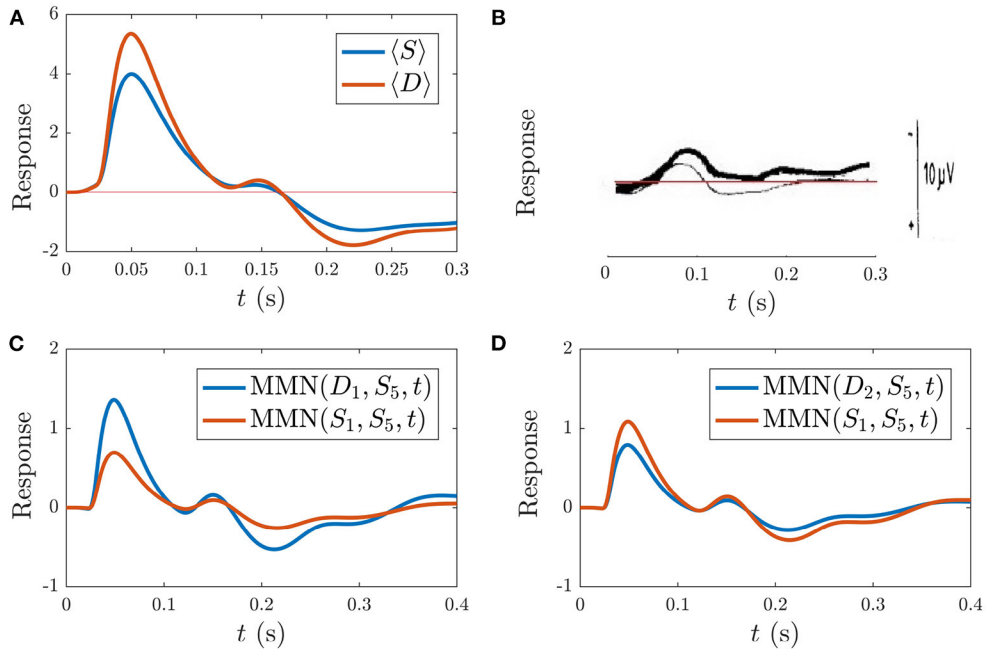

FIGURE 10 | Simulations vs. experiment (Sams et al., 1984) for the sequences discussed in section 3.4.1. **(A)** Model *S* and *D* (see legend), each averaged over all presentations during the typical sequence of standards with occasional deviants. **(B)** Corresponding experimental findings from this same sequence, adapted from Sams et al. (1984), where *D* is indicated with the thin line and *D* is indicated by the bold line. The experimental scale bar at right is 10 μV in length. **(C)** Model findings for the MMN(D_1, S_5, t) and MMN(S_1, S_5, t) (see legend) corresponding to the sequence $(4\mathcal{S})\mathcal{S}_5\mathcal{D}_1\mathcal{S}_1$. **(D)** Model findings for the MMN(D_2, S_5, t) and MMN(S_1, S_5, t) corresponding to the sequence $(4\mathcal{S})\mathcal{S}_5\mathcal{D}_1\mathcal{D}_2\mathcal{S}_1$.

cases where a $\mathcal{D}$ stimulus immediately follows at least four $\mathcal{S}$ stimuli.

Figure 10 shows the model simulation comparison to the experimental ER findings. **Figures 10A,B** display the resultant model averaged *S* and *D* responses over all instances of each stimulus as well as the experimental ERs from Sams et al. (1984), respectively. The model *S* and *D* responses reproduce the main features of the experimental responses: the *S* response displays prominent N1 and P2 deflections with similar latencies to the experimental *S* response, while *D* has a larger N1 peak than *S* and an N2 peak of lower amplitude than N1, which is also the case experimentally. The model findings for *D* differ from experiment in there being a prominent P2 deflection at $t > 170$ ms which isn't seen in the experimental response. This difference could signify the presence of higher-level feedbacks, or it may merely indicate that we have not adjusted our parameters to optimize the fit to this specific experiment. The fact that we see reasonable agreement between model simulations and experiment without further adjustments is evidence that adaptation plays a significant role in the development of the MMN. In future, fits to high-quality data for multiple experiments done on the same subjects should resolve this issue. Note that the timings of deflections in the model responses change depending on model parameters, unlike the fixed timings of traditional ER components.

In addition to calculating the MMN corresponding to the first $\mathcal{D}$ after four or more $\mathcal{S}$ as defined above, the experimental study tested whether an $\mathcal{S}$ stimulus directly following the $\mathcal{D}$ stimulus also caused a MMN with respect to the $\mathcal{S}$ preceding the $\mathcal{D}$. Specifically, they implemented the following sequence of stimuli: $(4\mathcal{S})\mathcal{S}_5\mathcal{D}_1\mathcal{S}_1$ and calculated MMN(S_1, S_5, t). They found that $\mathcal{S}_1$ did indeed yield a nonzero MMN(S_1, S_5, t), albeit smaller in amplitude than the MMN(D_1, S_5, t) (Sams et al., 1984). Simulation of this sequence yields results in agreement with these findings. **Figure 10C** shows MMN(S_1, S_5, t) alongside the MMN(D_1, S_5, t), revealing that the latter is larger in amplitude, in agreement with experiment.

The extent to which our model reproduces the above experiments sheds light on the role adaptation plays in these particular scenarios. Historically, the experimental observation of the MMN(S_1, S_5, t) was interpreted via an argument that each stimulus is associated with its own "neuronal model" such that $\mathcal{D}_1$ not only causes a mismatch process relative to the neuronal model corresponding to $\mathcal{S}$ stimuli, but also initiates a neuronal model of its own (Sams et al., 1984). The present findings suggest that adaptation is able to account for the MMN(S_1, S_5, t) being smaller in amplitude than MMN(D_1, S_5, t) in this experiment, without needing to invoke such higher-order neuronal models or representations.

3.4.2. Sequence of Standards With Occasional Double Deviants

The authors from the study examined in the previous section also explored sequences of the form: $(4\mathcal{S})\mathcal{S}_5\mathcal{D}_1\mathcal{D}_2\mathcal{S}_1$, whereby a second deviant immediately following the first was presented. They subsequently calculated MMN(D_2, S_5, t) and found that it had smaller amplitude than MMN(D_1, S_5, t). In addition,

they found that MMN(S_1, S_5, t) in this sequence had larger amplitude than MMN(S_1, S_5, t) from the sequence with only isolated single $\mathcal{D}$ stimuli. Our simulations agree with these experimental findings, as illustrated in **Figure 10D**, where the simulated MMN(D_2, S_5, t) is smaller than MMN(D_1, S_5, t) from **Figure 10C**, and the MMN(S_1, S_5, t) is larger in amplitude than MMN(S_1, S_5, t) from the previous sequence with occasional isolated deviants, shown in **Figure 10C**.

The authors of the experimental study interpreted the reduced-amplitude MMN(D_2, S_5, t) and the increased-amplitude MMN(S_1, S_5, t), relative to the corresponding cases with a single deviant, as evidence of the involvement of "neuronal models." However, the agreement seen between their findings and the model simulations in this section suggest that adaptation can account for these findings. Again we suggest that the "neuronal models" posited in these early studies correspond to adaptation of the relevant cortical regions associated with each stimulus, at least to a first approximation. In favor of this is the way that these "neuronal models" appear to be strengthened by repeated stimuli (Sams et al., 1984). The present adaptation process naturally accounts for this by the fact that repeated stimuli drive greater adaptation, resulting in a larger mismatch when followed by a different stimulus.

3.4.3. Sequences of Identical Stimuli With Different ISIs

Here we consider streams of identical stimuli in order to probe how adaptation and the resulting S_∞ and MMN depend on the ISI. We also compare the model with an early study that looked at the ERs to tone-only sequences of infrequent stimuli (Näätänen et al., 1989a). Due to the development of the S_∞ and D responses illustrated in section 3.3, an ISI on the order of $\approx 0.5-1$ s should give rise to an adapted S response, where the number of repeated stimuli required for S to approach its limiting form S_∞ depends on the ISI. As shown in section 3.3, for an ISI of 1 s, approximately 5 stimuli are required to approximate S_∞.

Reducing the ISI increases the number of stimuli that occur during the $\sim$ 5 s window before S_∞ is reached, as illustrated in **Figure 11**, which shows the development of S_∞ due to multiple successive stimuli at varying ISIs; the S_∞ parameters are thereby pushed further from the P state and its profile is modified accordingly. **Figure 11A** shows how S_∞ is reached only after 11 stimuli spaced at an ISI of 0.5 ms, whereas **Figures 11B,C** show that S_∞ is reached only after 9 and 6 stimuli spaced at ISIs of 0.6 and 0.7 ms, respectively. The adapted S_∞ response therefore depends on the ISI. This is illustrated by simulating sequences of responses at variable ISIs and plotting the limiting form S_∞ of each sequence vs. ISI, shown in **Figure 12A** along with the corresponding MMN(D, S_∞, t) in **Figure 12B**. We see that as the ISI decreases from 1 to 0.5 s, the S_∞ response curve exhibits smaller N1 and P2 amplitudes and the N2 component, which was small relative to the N1 peak in the 1 s ISI case, vanishes completely. Furthermore, as the ISI decreases the MMN amplitude increases, which agrees with experimental findings of increased MMN amplitudes at shorter ISIs (Ford and Hillyard, 1981; Nordby et al., 1988; Näätänen et al., 1993).

3.4.4. Tone-Only Sequence

In our model completely discriminable stimuli do not affect one another's responses via adaptation, as expressed via Equation (51). As a result, streams of identical $\mathcal{D}$ stimuli result in identical streams of D responses, resembling those shown in **Figure 6B**, regardless of whether any fully discriminable $\mathcal{S}$ stimuli are presented in between.

In contrast to our pure-adaptation prediction, experiments have found that ERs to the infrequent stimuli alone (tone-only sequence) differed from the case where more frequent $\mathcal{S}$ stimuli occurred in between. Specifically, the tone-only responses did not exhibit a negative deflection overlapping the N1 and P2 features, but rather exhibited a larger N1 deflection (Näätänen et al., 1989a). This implies that effects other than adaptation of distinct cortical regions, such as higher-order processes and memory effects, play an important role in the distinction between these two cases. The present work enables the effects of adaptation to be separated from the other contributions to allow the latter to be focused on more specifically.

3.4.5. Well-Separated Trains of Stimuli

We now investigate an experiment that specifically invoked higher-order processing as a contributing factor to the MMN, in order to tease apart how much can be explained by adaptation alone. The experiment (Cowan et al., 1993) investigated the links between MMN and memory representation by setting out to measure if a long-term or "silent" memory representation (longer than the typical decay rate of the S_∞ response) of a given $\mathcal{S}$ stimulus persists over a long interval between two well-separated trains of $\mathcal{S}$ stimuli such that a $\mathcal{D}$ stimulus occurring in the second position of the second train elicits an MMN. Because the inter-train interval was longer than the MMN decay time, they concluded that any MMN associated with the second-position $\mathcal{D}$ reflects the reactivation of a memory representation that became dormant during the inter-train interval and was reactivated by the first $\mathcal{S}$ stimulus of the new train (Cowan et al., 1993). Their findings confirmed the presence of an MMN associated with the second-position $\mathcal{D}$ and led them to interpret this as evidence for such memory formation, inactivation, and reactivation.

The experimental procedure involved placing a $\mathcal{D}$ stimulus in position 1, 2, 4, 6, or 8 of a nine-item train of standards at an ISI of 610 ms between tones within a single train and an inter-train interval of $11-15$ s (Cowan et al., 1993). We use a superscript n to label the position of the $\mathcal{D}$ stimuli within the 9-element train such that the above cases can be distinguished as $\mathcal{D}^1$, $\mathcal{D}^2$, $\mathcal{D}^4$, $\mathcal{D}^8$, and $\mathcal{D}^8$. The authors found that stimuli at positions $\mathcal{D}^{n\geq 2}$ were sufficient to yield an MMN (Cowan et al., 1993).

To model this experiment we simulate the following blocks of trains: an $\mathcal{S}$-only train: $(9\mathcal{S})$, and nine-element trains of mostly $\mathcal{S}$ stimuli with $\mathcal{D}$ stimuli placed at positions listed above: $\mathcal{D}(8\mathcal{S})$, $(1\mathcal{S})\mathcal{D}(7\mathcal{S})$, $(3\mathcal{S})\mathcal{D}(5\mathcal{S})$, $(5\mathcal{S})\mathcal{D}(3\mathcal{S})$, and $(7\mathcal{S})\mathcal{D}(1\mathcal{S})$. The long inter-train intervals mean that there are no cumulative effects from the adaptation occurring in each train that last until the next train and thus that such effects can be disregarded, allowing each train to be studied in isolation. To follow what was measured experimentally, the MMN corresponding to the D responses at the positions listed above with respect to the

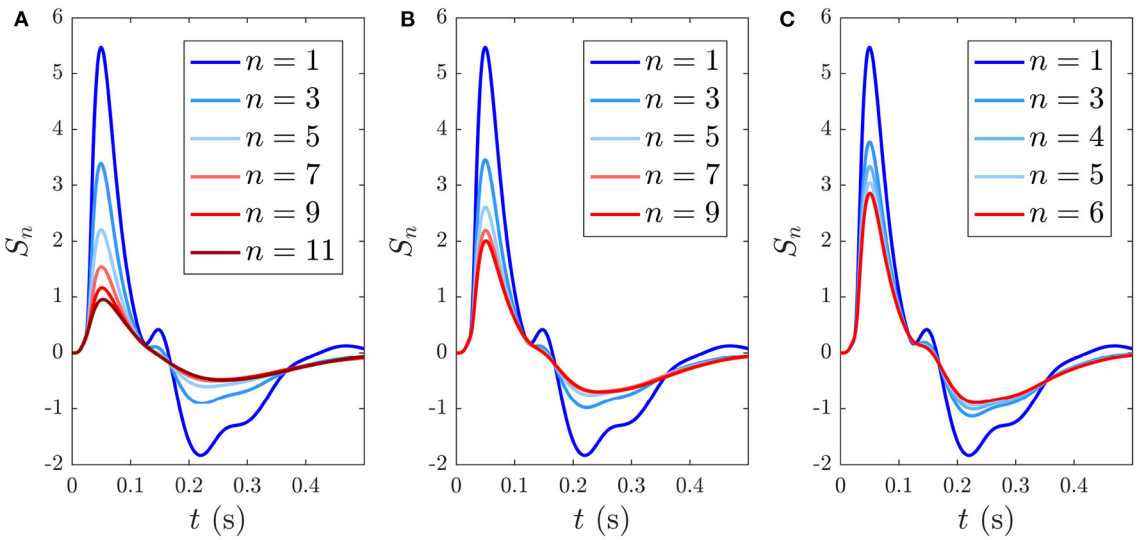

FIGURE 11 | Development of the S_∞ response during consecutive stimuli spaced at varying ISIs, with stimulus number indicated in the legends. **(A)** $S_n(t)$ for $n = 1, \ldots, 11$ with a 0.5 s ISI. **(B)** $S_n(t)$ for $n = 1, \ldots, 9$ with a 0.6 s ISI. **(C)** $S_n(t)$ for $n = 1, \ldots, 6$ with a 0.7 s ISI.

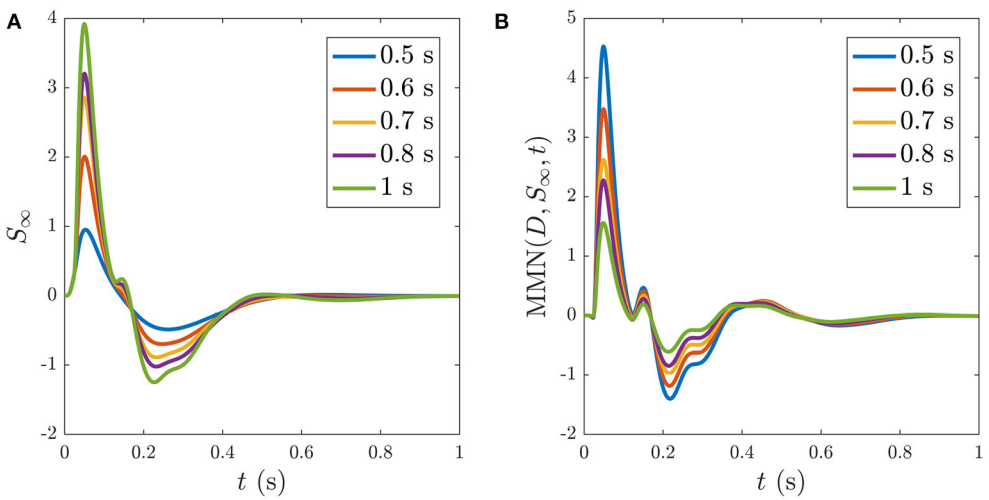

FIGURE 12 | Dependence of S_∞ on the ISI, with ISI indicated in the legends. **(A)** S_∞ for ISIs of: 0.5, 0.6, 0.7, 0.8, and 1 s, as indicated in the legend. **(B)** The corresponding $\mathrm{MMN}(D, S_\infty, t)$ for each of the cases in **(A)**.

corresponding S response at the same position in the $\mathcal{S}$-only train, $\mathrm{MMN}(D^n, S_n, t)$ was calculated and is shown **Figure 13** alongside the corresponding experimental MMN adapted from Cowan et al. (1993).

We find that $\mathrm{MMN}(D^n, S_n, t)$ is zero at $n = 1$, because $D_1 = S_1$, and that it increases with n because adaptation effects alone are enough for $\mathcal{D}^2$ to elicit an MMN in the above trains without the need to invoke memory formation and representation processes. The model detection of a nonzero $\mathrm{MMN}(D^2, S_2, t)$ reflects the fact that the S_2 response in a stream of standards has already adapted significantly enough to cause a mismatch with the subsequent D response. The extent of the adaptation depends on the ISI, as shown in section 3.4.3. Other similarities between the model $\mathrm{MMN}(D^n, S_n, t)$ and experiment include the presence of N1 and N2 deflections as well as P2 and P3 deflections emerging as n increases, as well as an increase of the N1 deflection as n increases. Experimentally, as n increases the N2 deflection approaches that of N1, whereas their relative amplitudes remain unchanged in the model simulations. This could reflect an effect of the memory representation process that was conjectured to be involved (Cowan et al., 1993), especially because considerable experimentation with changing the model adaptation parameters found the N2 peak to consistently remain smaller in amplitude compared to the N1 peak.

Overall, the level of agreement between model simulations and experiment for the above sequences suggests that adaptation

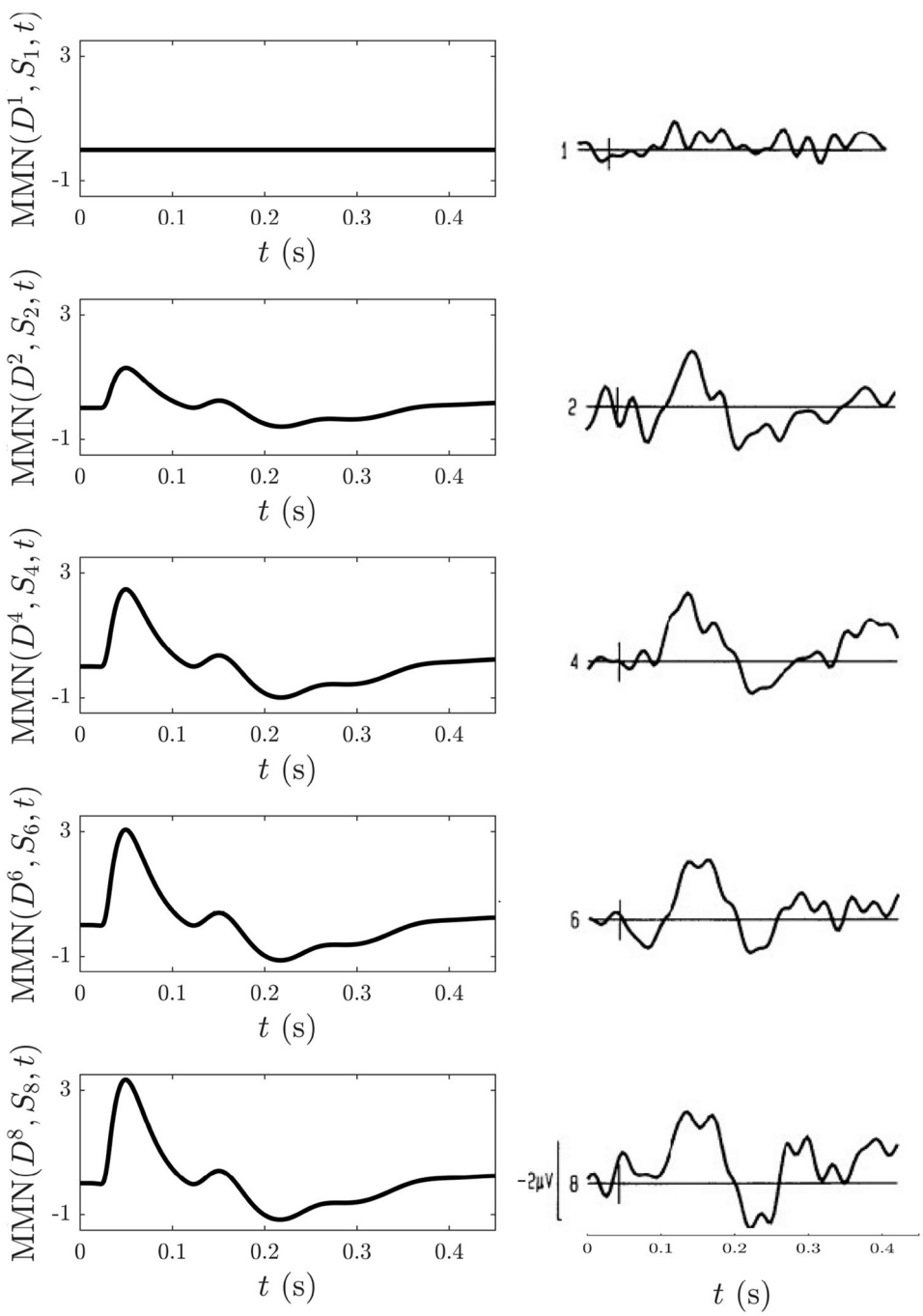

FIGURE 13 | Comparison of model simulations and experimental findings (Cowan et al., 1993) discussed in section 3.4.5. Simulation of the MMN(D^n, S_n, t) for $n = 1, 2, 4, 6, 8$ is shown on the left and the corresponding experimental MMN curves adapted from Cowan et al. (1993) on the right. The small vertical mark in each experimental frame indicates the stimulus time and the first curve gives an indication of noise levels.

can explain key features of the experiment that were previously assumed to be the result of higher order processes (Cowan et al., 1993). However, such processes are likely to be needed to account for the remaining differences between the model and experimental findings concerning the relative amplitudes of the N1 and N2 deflections.

4. SUMMARY AND DISCUSSION

We have modeled and analyzed sequences of auditory evoked responses, used in human cognitive studies, by means of a physiologically based neural field theory whose predictions have previously reproduced a wide range of experimental

data on brain activity and connectivity, as mentioned in the Introduction. To do so, the theory was generalized to include corticothalamic adaptation to repeated stimuli arriving at a given point on the tonotopic map. Repeated stimulation within the 5–10 s lifetime of adaptation leads to greater movement of corticothalamic gains away from the prestimulation baseline and contributes to standard responses evolving away from deviants.

The central aim of the work is to provide a means of calculating the response to arbitrary sequences of discriminable stimuli in order to determine how much of the dynamics can be accounted for by adaptation and how much might be ascribable to higher-order top-down memory-related stimulus-comparison processes—a long-running controversy in the field. This accords with Occam's Razor, which dictates that one should first determine how much can be explained by low-level processes such as adaptation in thalamus and primary auditory cortex before invoking higher-order aspects. However, we stress that the latter processes are certainly relevant in many contexts, especially those involving long-term memory or comparison of abstract stimulus features.

This work provides a quantitative bridge between biophysical analysis of brain activity and electrophysiological measurements of human cognitive processes, in more detail than has previously been possible. Tools based on this approach should help to clarify the relative roles of low-level adaptive processes and high-level feedbacks in determining evoked responses in various situations.

The main outcomes are:

(i) A corticothalamic NFT model of the medial geniculate nucleus of the thalamus and the primary auditory cortex was formulated in which ERs are viewed as impulse responses, with evoked changes occurring both directly in the activity and indirectly via the system gains, resulting in a bilinear response that we treated via perturbation theory. Long-timescale adaptation of gains was also incorporated for the first time.

(ii) The dynamical building blocks of ERs are damped oscillations at natural resonant frequencies, usually with deflections of both polarities. This is in contrast to the traditional notion of ER components with fixed polarities and timings and in accord with temporal shifts and inversions of some features during development (Kerr et al., 2010).

(iii) Adaptation changes both amplitudes and timings of ER waveforms. This invalidates common assertions that adaptation can change only amplitudes of traditional ER components with fixed timings.

(iv) The MMN is a mathematical entity that is constructed by subtracting one response from another. Most commonly the response to a common standard is subtracted from the response to a rare deviant, but there is no unique definition. However, if the parts of ERs that are due to adaptation can be identified and shown to be insufficient to account for the differences between the two responses (i.e., for their MMN), the remainder may well be attributable to higher-level processes.

(v) Our generalized notation for the MMN between two responses—e.g., $\mathrm{MMN}(S_3, S_5, t)$ for the third standard relative to the fifth—highlights the implausibility of there being a separate dedicated set of neurons that generate an MMN for every possible comparison that might be conceived of by experimenters. This is thrown into stark relief when one notes that any pair of responses whatsoever can be used to define an MMN—even responses in different sensory areas at very different times.

(vi) Identification of higher-order processing and other effects is facilitated by the model. This is because one model must be able to account for a given subject's responses to arbitrary stimulus sequences—ideally with little or no change in parameters so long as the subject's physiological state is unchanged. Hence, once parameters have been calibrated on sequences of identical stimuli at various ISIs, for example, they should yield the responses to arbitrary stimulus sequences, as far as adaptive changes go. Further differences can then be explored as potentially being due to other mechanisms.

(vii) Deviant ERs start from a point nearer the corticothalamic baseline than standard ERs, which begin from a point that adaptation has driven away from the pre-stimulation state. Corticothalamic feedforwards and feedbacks change in strength with adaptation, with the largest changes found to be in gains involving the cortex, as summarized in **Table 2**. This is broadly consistent with theories such as predictive coding in which top-down predictions are compared with bottom-up signals and the system adapts to reduce the discrepancy (Garrido et al., 2009b).

(viii) In oddball paradigms, the model accounts naturally for (a) the difference between S and D responses, (b) the effects of consecutive $\mathcal{D}$ stimuli on subsequent S and D responses, (c) the effect of the position of the stimulus in a long train, and (d) the development of S responses (starting as identical with D ones) with repeated presentation, and their decay after a stimulus-free interval.

(ix) Some aspects of ERs have not been accounted for by adaptation alone, which points to their likely dependence on higher-order processes and feedbacks. These include tone-only sequences which provoke the same model D responses regardless of whether discriminable S responses occur in between, which is not in accord with experimentally observed differences between the two cases. Likewise, differences between well-separated stimulus trains with deviants in different positions can be partly explained by adaptation effects, but late structure has not been fully reproduced and needs further investigation. In this context, we again stress that our aim is not to account for all ER features by adaptation, but to determine which features can be explained in this manner so as to focus attention more sharply on those that are produced by other mechanisms.

Overall, we have shown that adaptation can account for many but not all features of ERs in various stimulus sequences. This both highlights the role of such processes in the initial corticothalamic

stages of signal processing and cognition and points the way to focus on higher-order aspects, especially in humans. The formulation in terms of resonances and gains makes immediate links to control-systems interpretations that tie the results to the dynamics of prediction and unconscious attention behind many cognitive processes (Babaie-Janvier and Robinson, 2018, 2019). More generally, the NFT used has accounted for a wide variety of normal and abnormal brain activity and connectivity phenomena, as mentioned in the Introduction, so ER dynamics is thereby also integrated into this broader landscape.

The present work provides a starting point for quantitative exploration of the role of adaptation in the host of ER sequences that have been studied in the literature discussed in the Introduction. This would include analysis of ERs to sequences of stimuli that are not fully distinguishable, omitted tones, tones of variable frequency, duration, or amplitude, and other variants. For optimal outcomes the model should be calibrated for individual subjects on simple oddball sequences, then used to predict their responses to more complex stimulus sequences—something that has previously been done when applying NFT to a range of other phenomena mentioned in the Introduction. A key advantage of NFT is that its parameters are closely tied the physiology, so links to underlying biophysics are more direct and easier to make than via phenomenological component analysis.

Many further extensions and applications of the model can be made. A key generalization needed to better probe ERs is to include spatial aspects of the tonotopic map and the responses to enable comparison with observations of ER topography. Some such work has been done on evoked responses using NFT, albeit without adaptation (Mukta et al., 2019; Robinson et al., 2019) and the work here will enable it to be generalized by appropriately modifying the response functions and including the spatial structure of natural modes of brain activity. Application to ERs that involve other auditory features (e.g., interaural delays and directionality), or other sensory modalities, is also an obvious direction for future work because the present formulation is certainly not limited to auditory systems. Similarly, one could apply this approach to evoked responses in nonhuman animals, although the parameters would need to be recalibrated in that case. It is also worth noting that the stimuli used do not have to be impulsive—replacement of a delta-function input by a periodic drive enables steady-state evoked responses to be studied, as has previously been done in the absence of adaptation (Robinson et al., 2008).

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

PR conceived the project and performed the analytic work. NG and TB-J performed the coding, the numerical calculations, and their analysis. All authors drafted the respective sections of the MS and collaborated to produce the final version.

FUNDING

The Australian Research Council supported this work under Laureate Fellowship grant FL1401000225 and Center of Excellence Grant CE140100007.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2021.655505/full#supplementary-material

REFERENCES

Abeysuriya, R. G., Rennie, C. J., and Robinson, P. A. (2015). Physiologically based arousal state estimation and dynamics. *J. Neurosci. Methods* 253, 55–69. doi: 10.1016/j.jneumeth.2015.06.002

Amari, S. I. (1977). Dynamics of pattern formation in lateral-inhibition type neural fields. *Biol. Cybern.* 27, 77–87. doi: 10.1007/BF00337259

Atienza, M., Cantero, J. L., and Escera, C. (2001). Auditory information processing during human sleep as revealed by event-related brain potentials. *Clin. Neurophysiol.* 112, 2031–2045. doi: 10.1016/S1388-2457(01)00650-2

Babaie-Janvier, T., and Robinson, P. A. (2018). Neural field theory of corticothalamic prediction with control systems analysis. *Front. Hum. Neurosci.* 12:334. doi: 10.3389/fnhum.2018.00334

Babaie-Janvier, T., and Robinson, P. A. (2019). Neural field theory of corticothalamic attention with control systems analysis. *Front. Neurosci.* 13:1240. doi: 10.3389/fnins.2019.01240

Babaie-Janvier, T., and Robinson, P. A. (2020). Neural field theory of evoked response potentials with attentional gain dynamics. *Front. Hum. Neurosci.* 14:293. doi: 10.3389/fnhum.2020.00293

Başar, E. (2012). *Brain Function and Oscillations: Vol. I: Brain Oscillations.* Principles and approaches. Berlin: Springer Science & Business Media.

Beurle, R. L. (1956). Properties of a mass of cells capable of regenerating pulses. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 240, 55–94. doi: 10.1098/rstb.1956.0012

Braitenberg, V., and Schüz, A. (1998). *Cortex: Statistics and Geometry of Neuronal Connectivity, 2nd Edn.* Heidelberg: Springer-Verlag. doi: 10.1007/978-3-662-03733-1

Breakspear, M., Roberts, J. A., Terry, J. R., Rodrigues, S., Mahant, N., and Robinson, P. A. (2006). A unifying explanation of primary generalized seizures through nonlinear brain modeling and bifurcation analysis. *Cereb. Cortex* 16, 1296–1313. doi: 10.1093/cercor/bhj072

Bressloff, P. C. (2012). Spatiotemporal dynamics of continuum neural fields. *J. Phys. A Math. Theor.* 45:033001. doi: 10.1088/1751-8113/45/3/033001

Coombes, S., beim Graben, P., Potthast, R., and Wright, J. (2014). *Neural Fields: Theory and Applications.* Berlin: Springer. doi: 10.1007/978-3-642-54593-1

Cowan, N. (1984). On short and long auditory stores. *Psychol. Bull.* 96:341. doi: 10.1037/0033-2909.96.2.341

Cowan, N., Winkler, I., Teder, W., and Näätänen, R. (1993). Memory prerequisites of mismatch negativity in the auditory event-related potential (ERP). *J. Exp. Psychol.* 19:909–21. doi: 10.1037/0278-7393.19.4.909

Deco, G., Jirsa, V. K., Robinson, P. A., Breakspear, M., and Friston, K. (2008). The dynamic brain: From spiking neurons to neural masses and cortical fields. *PLoS Comput. Biol.* 4:e1000092. doi: 10.1371/journal.pcbi.1000092

Demiralp, T., Ademoglu, A., Istefanopulos, Y., and Gülçür, H. Ö. (1998). Analysis of event-related potentials (ERP) by damped sinusoids. *Biol. Cybern.* 78, 487–493. doi: 10.1007/s004220050452

Ford, J. M., and Hillyard, S. A. (1981). Event-related potentials (ERPs) to interruptions of a steady rhythm. *Psychophysiology* 18, 322–330. doi: 10.1111/j.1469-8986.1981.tb03043.x

Freeman, W. (1975). *Mass Action in the Nervous System*. New York, NY: Academic Press.

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787

Friston, K. J. (2011). Functional and effective connectivity: a review. *Brain Connect.* 1, 13–36. doi: 10.1089/brain.2011.0008

Garrido, M. I., Kilner, J. M., Kiebel, S. J., and Friston, K. J. (2007). Evoked brain responses are generated by feedback loops. *Proc. Natl. Acad. Sci. U.S.A.* 104, 20961–20966. doi: 10.1073/pnas.0706274105

Garrido, M. I., Kilner, J. M., Kiebel, S. J., Stephan, K. E., Baldeweg, T., and Friston, K. J. (2009a). Repetition suppression and plasticity in the human brain. *NeuroImage* 48, 269–279. doi: 10.1016/j.neuroimage.2009.06.034

Garrido, M. I., Kilner, J. M., Stephan, K. E., and Friston, K. J. (2009b). The mismatch negativity: a review of underlying mechanisms. *Clin. Neurophysiol.* 120, 453–463. doi: 10.1016/j.clinph.2008.11.029

Garrido, M. I., Sahani, M., and Dolan, R. J. (2013). Outlier responses reflect sensitivity to statistical structure in the human brain. *PLoS Comput. Biol.* 9:e1002999. doi: 10.1371/journal.pcbi.1002999

Gazzaley, A., Cooney, J. W., McEvoy, K., Knight, R. T., and D'Esposito, M. (2005). Top-down enhancement and suppression of the magnitude and speed of neural activity. *J. Cogn. Neurosci.* 17, 507–517. doi: 10.1162/0898929053279522

Herdener, M., Esposito, F., Scheffler, K., Schneider, P., Logothetis, N. K., Uludag, K., et al. (2013). Spatial representations of temporal and spectral sound cues in human auditory cortex. *Cortex* 49, 2822–2833. doi: 10.1016/j.cortex.2013.04.003

Jääskeläinen, I. P., Ahveninen, J., Bonmassar, G., Dale, A. M., Ilmoniemi, R. J., Levänen, S., et al. (2004). Human posterior auditory cortex gates novel sounds to consciousness. *Proc. Natl. Acad. Sci. U.S.A.* 101, 6809–6814. doi: 10.1073/pnas.0303760101

Jirsa, V. K., and Haken, H. (1996). Field theory of electromagnetic brain activity. *Phys. Rev. Lett.* 77, 960–963. doi: 10.1103/PhysRevLett.77.960

Kerr, C. C., Rennie, C. J., and Robinson, P. A. (2008). Physiology-based modeling of cortical auditory evoked potentials. *Biol. Cybern.* 98, 171–184. doi: 10.1007/s00422-007-0201-1

Kerr, C. C., Rennie, C. J., and Robinson, P. A. (2009). Deconvolution analysis of target evoked potentials. *J. Neurosci. Methods* 179, 101–110. doi: 10.1016/j.jneumeth.2009.01.003

Kerr, C. C., Rennie, C. J., and Robinson, P. A. (2011). Model-based analysis and quantification of age trends in auditory evoked potentials. *Clin. Neurophysiol.* 122, 134–147. doi: 10.1016/j.clinph.2010.05.030

Kerr, C. C., van Albada, S. J., Rennie, C. J., and Robinson, P. A. (2010). Age trends in auditory oddball evoked potentials via component scoring and deconvolution. *Clin. Neurophysiol.* 121, 962–976. doi: 10.1016/j.clinph.2009.11.077

Koch, C. (1999). *Biophysics of Computation*. Oxford: Oxford University Press.

Lopes da Silva, F. H., Van Rotterdam, A., Barts, P., Van Heusden, E., and Burr, B. (1976). Models of neuronal populations: the basic mechanisms of rhythmicity. *Prog. Brain Res.* 45, 281–308. doi: 10.1016/S0079-6123(08)60995-4

Loveless, N., Levänen, S., Jousmäki, V., Sams, M., and Hari, R. (1996). Temporal integration in auditory sensory memory: neuromagnetic evidence. *Electroencephalogr. Clin. Neurophysiol.* 100, 220–228. doi: 10.1016/0168-5597(95)00271-5

Luck, S. J. (2014). *An Introduction to the Event-Related Potential Technique*. Cambridge, MA: MIT Press.

Luck, S. J., and Kappenman, E. S. (2011). *The Oxford Handbook of Event-Related Potential Components*. New York, NY: Oxford University Press. doi: 10.1093/oxfordhb/9780195374148.001.0001

Malmierca, M. S., Cristaudo, S., Pérez-González, D., and Covey, E. (2009). Stimulus-specific adaptation in the inferior colliculus of the anesthetized rat. *J. Neurosci.* 29, 5483–5493. doi: 10.1523/JNEUROSCI.4153-08.2009

May, P., Tiitinen, H., Ilmoniemi, R. J., Nyman, G., Taylor, J. G., and Näätänen, R. (1999). Frequency change detection in human auditory cortex. *J. Comput. Neurosci.* 6, 99–120. doi: 10.1023/A:1008896417606

May, P. J. C., Westö, J., and Tiitinen, H. (2015). Computational modelling suggests that temporal integration results from synaptic adaptation in auditory cortex. *Eur. J. Neurosci.* 41, 615–630. doi: 10.1111/ejn.12820

Mukta, K. N., Gao, X., and Robinson, P. A. (2019). Neural field theory of evoked response potentials in a spherical brain geometry. *Phys. Rev. E* 99:062304. doi: 10.1103/PhysRevE.99.062304

Näätänen, R. (2003). Mismatch negativity: clinical research and possible applications. *Int. J. Psychophysiol.* 48, 179–188. doi: 10.1016/S0167-8760(03)00053-9

Näätänen, R., Gaillard, A. W., and Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychol.* 42, 313–329. doi: 10.1016/0001-6918(78)90006-9

Näätänen, R., Jacobsen, T., and Winkler, I. (2005). Memory-based or afferent processes in mismatch negativity (MMN): a review of the evidence. *Psychophysiology* 42, 25–32. doi: 10.1111/j.1469-8986.2005.00256.x

Näätänen, R., Jiang, D., Lavikainen, J., Reinikainen, K., and Paavilainen, P. (1993). Event-related potentials reveal a memory trace for temporal features. *Neuroreport* 5, 310–312. doi: 10.1097/00001756-199312000-00033

Näätänen, R., Paavilainen, P., Alho, K., Reinikainen, K., and Sams, M. (1989a). Do event-related potentials reveal the mechanism of the auditory sensory memory in the human brain? *Neurosci. Lett.* 98, 217–221. doi: 10.1016/0304-3940(89)90513-2

Näätänen, R., Paavilainen, P., and Reinikainen, K. (1989b). Do event-related potentials to infrequent decrements in duration of auditory stimuli demonstrate a memory trace in man? *Neurosci. Lett.* 107, 347–352. doi: 10.1016/0304-3940(89)90844-6

Näätänen, R., Paavilainen, P., Rinne, T., and Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clin. Neurophysiol.* 118, 2544–2590. doi: 10.1016/j.clinph.2007.04.026

Nelken, I. (2004). Processing of complex stimuli and natural scenes in the auditory cortex. *Curr. Opin. Neurobiol.* 14, 474–480. doi: 10.1016/j.conb.2004.06.005

Niedermeyer, E., and Lopes da Silva, F. H. (2011). *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Baltimore, MD: Lippincott Williams & Wilkins.

Nordby, H., Roth, W. T., and A, P. (1988). Event-related potentials to breaks in sequences of alternating pitches or interstimulus intervals. *Psychophysiology* 25, 262–268. doi: 10.1111/j.1469-8986.1988.tb01239.x

Nunez, P. L. (1974). The brain wave equation: a model for EEG. *Math. Biosci.* 21, 279–297. doi: 10.1016/0025-5564(74)90020-0

Nunez, P. L. (1995). *Neocortical Dynamics and Human EEG Rhythms*. Oxford: Oxford University Press.

Nunez, P. L., and Srinivasan, R. (2006). Electric fields of the brain: the neurophysics of EEG. Oxford: Oxford University Press.

O'Connor, S. C, and Robinson, P. A. (2004). Spatially uniform and nonuniform analyses of electroencephalographic dynamics, with application to the topography of the alpha rhythm. *Phys. Rev. E* 70:011911. doi: 10.1103/PhysRevE.70.011911

Ogata, K., and Yang, Y. (1970). *Modern Control Engineering*. Englewood Cliffs, NJ: Prentice-Hall.

Pérez-González, D., and Malmierca, M. S. (2014). Adaptation in the auditory system: an overview. *Front. Integr. Neurosci.* 8:19. doi: 10.3389/fnint.2014.00019

Rennie, C. J., Robinson, P. A., and Wright, J. J. (1999). Effects of local feedback on dispersion of electrical waves in the cerebral cortex. *Phys. Rev. E* 59, 3320–3329. doi: 10.1103/PhysRevE.59.3320

Rennie, C. J., Robinson, P. A., and Wright, J. J. (2002). Unified neurophysical model of EEG spectra and evoked potentials. *Biol. Cybern.* 86, 457–471. doi: 10.1007/s00422-002-0310-9

Rennie, C. J., Wright, J. J., and Robinson, P. A. (2000). Mechanisms of cortical electrical activity and emergence of gamma rhythm. *J. Theor. Biol.* 205, 17–35. doi: 10.1006/jtbi.2000.2040

Roberts, J. A., and Robinson, P. A. (2012). Quantitative theory of driven nonlinear brain dynamics. *NeuroImage* 62, 1947–1955. doi: 10.1016/j.neuroimage.2012.05.054

Robinson, P. A., Chen, P.-C., and Yang, L. (2008). Neural field theory of perceptual echo and implications for estimating brain connectivity. *Biol. Cybern.* 98, 1–10. doi: 10.1007/s00422-007-0191-z

Robinson, P. A., Pagés, J., Gabay, N. C., Babaie-Janvier, T., and Mukta, K. N. (2019). Neural field theory of perceptual echo and implications for estimating brain connectivity. *Phys. Rev. E* 97:042418. doi: 10.1103/PhysRevE.97.042418

Robinson, P. A., Rennie, C. J., and Rowe, D. L. (2002). Dynamics of large-scale brain activity in normal arousal states and epileptic seizures. *Phys. Rev. E* 65:041924. doi: 10.1103/PhysRevE.65.041924

Robinson, P. A., Rennie, C. J., Rowe, D. L., and O'Connor, S. C. (2004). Estimation of multiscale neurophysiologic parameters by electroencephalographic means. *Hum. Brain Mapp*. 23, 53–72. doi: 10.1002/hbm.20032

Robinson, P. A., Rennie, C. J., Rowe, D. L., O'Connor, S. C., and Gordon, E. (2005). Multiscale brain modelling. *Philos. Trans. R. Soc. Lond. B Biol. Sci*. 360, 1043–1050. doi: 10.1098/rstb.2005.1638

Robinson, P. A., Rennie, C. J., and Wright, J. J. (1997). Propagation and stability of waves of electrical activity in the cerebral cortex. *Phys. Rev. E* 56:826. doi: 10.1103/PhysRevE.56.826

Robinson, P. A., and Roy, N. (2015). Neural field theory of nonlinear wave-wave and wave-neuron processes. *Phys. Rev. E* 91:062719. doi: 10.1103/PhysRevE.91.062719

Salisbury, D. F. (2012). Finding the missing stimulus mismatch negativity (MMN): Emitted MMN to violations of an auditory gestalt. *Psychophysiology* 49, 544–548. doi: 10.1111/j.1469-8986.2011.01336.x

Sams, M., Alho, K., and Näätänen, R. (1984). Short-term habituation and dishabituation of the mismatch negativity of the ERP. *Psychophysiology* 21, 434–441. doi: 10.1111/j.1469-8986.1984.tb00223.x

Sams, M., Paavilainen, P., Alho, K., and Näätänen, R. (1985). Auditory frequency discrimination and event-related potentials. *Electroencephalogr. Clin. Neurophysiol*. 62, 437–448. doi: 10.1016/0168-5597(85)90054-1

Sanz-Leon, P., Robinson, P. A., Knock, S. A., Drysdale, P. M., Abeysuriya, R. G., Fung, F. K., et al. (2018). NFTsim: theory and simulation of multiscale neural field dynamics. *PLoS Comput. Biol*. 14:e1006387. doi: 10.1371/journal.pcbi.1006387

Schröger, E. (1998). Measurement and interpretation of the mismatch negativity. *Behav. Res. Methods Instrum. Comput*. 30, 131–145. doi: 10.3758/BF03209423

Steyn-Ross, M. L., Steyn-Ross, D. A., Sleigh, J. W., and Liley, D. T. J. (1999). Theoretical electroencephalogram stationary spectrum for a white-noise-driven cortex: evidence for a general anesthetic-induced phase transition. *Phys. Rev. E* 60, 7299–7311. doi: 10.1103/PhysRevE.60.7299

Sussman, E. S., Chen, S., Sussman-Fort, J., and Dinces, E. (2014). The five myths of MMN: redefining how to use MMN in basic and clinical research. *Brain Topogr*. 27, 553–564. doi: 10.1007/s10548-013-0326-6

Szymanski, F. D., Garcia-Lazaro, J. A., and Schnupp, J. W. H. (2009). Current source density profiles of stimulus-specific adaptation in rat auditory cortex. *J. Neurophysiol*. 102, 1483–1490. doi: 10.1152/jn.00240.2009

Tervaniemi, M., Maury, S., and Näätänen, R. (1994). Neural representations of abstract stimulus features in the human brain as reflected by the mismatch negativity. *Neuroreport* 5, 844–846. doi: 10.1097/00001756-199403000-00027

van Albada, S. J., Kerr, C. C., Chiang, A. K. I., Rennie, C. J., and Robinson, P. A. (2010). Neurophysiological changes with age probed by inverse modeling of EEG spectra. *Clin. Neurophysiol*. 121, 21–38. doi: 10.1016/j.clinph.2009.09.021

Wilson, H. R., and Cowan, J. D. (1972). Excitatory and inhibitory interactions in localized populations of model neurons. *Biophys. J*. 12, 1–24. doi: 10.1016/S0006-3495(72)86068-5

Wilson, H. R., and Cowan, J. D. (1973). A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Biol. Cybern*. 13, 55–80. doi: 10.1007/BF00288786

Winkler, I., Reinikainen, K., and Näätänen, R. (1993). Event-related brain potentials reflect traces of echoic memory in humans. *Percept. Psychophys*. 53, 443–449. doi: 10.3758/BF03206788

Wright, J. J., and Liley, D. T. J. (1994). A millimetric-scale simulation of electrocortical wave dynamics based on anatomical estimates of cortical synaptic density. *Netw. Comput. Neural Syst*. 5, 191–202. doi: 10.1088/0954-898X_5_2_005

Wright, J. J., and Liley, D. T. J. (1996). Dynamics of the brain at global and microscopic scales: neural networks and the EEG. *Behav. Brain Sci*. 19, 285–295. doi: 10.1017/S0140525X00042679

Yabe, H., Tervaniemi, M., Reinikainen, K., and Näätänen, R. (1997). Temporal window of integration revealed by MMN to sound omission. *Neuroreport* 8, 1971–1974. doi: 10.1097/00001756-199705260-00035

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Robinson, Gabay and Babaie-Janvier. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.