



This is a repository copy of *Nonlinear modeling of cortical responses to mechanical wrist perturbations using the NARMAX method*.

White Rose Research Online URL for this paper:
<https://eprints.whiterose.ac.uk/164370/>

Version: Accepted Version

Article:

Gu, Y., Yang, Y., Dewald, J. et al. (3 more authors) (2021) Nonlinear modeling of cortical responses to mechanical wrist perturbations using the NARMAX method. IEEE Transactions on Biomedical Engineering, 68 (3). pp. 948-958. ISSN 0018-9294

<https://doi.org/10.1109/tbme.2020.3013545>

© 2020 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other users, including reprinting/ republishing this material for advertising or promotional purposes, creating new collective works for resale or redistribution to servers or lists, or reuse of any copyrighted components of this work in other works. Reproduced in accordance with the publisher's self-archiving policy.

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Nonlinear Modeling of Cortical Responses to Mechanical Wrist Perturbations using the NARMAX Method

Yuanlin Gu, Yuan Yang*, Julius P. A. Dewald, *IEEE Member*,
Frans C.T. van der Helm, Alfred C. Schouten, and Hua-Liang Wei*

Abstract— Objective: Nonlinear modeling of cortical responses (EEG) to wrist perturbations allows for the quantification of cortical sensorimotor function in healthy and neurologically impaired individuals. A common model structure reflecting key characteristics shared across healthy individuals may provide a reference for future clinical studies investigating abnormal cortical responses associated with sensorimotor impairments. Thus, the goal of our study is to identify this common model structure and therefore to build a nonlinear dynamic model of cortical responses, using nonlinear autoregressive–moving-average model with exogenous inputs (NARMAX). **Methods:** EEG was recorded from ten participants when receiving continuous wrist perturbations. A common model structure detection method was developed for identifying a common NARMAX model structure across all participants, with individualized parameter values. The results were compared to conventional subject-specific models. **Results:** The proposed method achieved 93.91% variance accounted for (VAF) when implementing a one-step-ahead prediction and around 50% VAF for a k-step ahead prediction ($k = 3$), without a substantial drop of VAF as compare to subject-specific models. The estimated common structure suggests that the measured cortical response is a mixed outcome of the nonlinear transformation of external inputs and local neuronal interactions or inherent neuronal dynamics at the cortex. **Conclusion:** The proposed method well determined the common characteristics across subjects in the cortical responses to wrist perturbations. **Significance:** It provides new insights into the human sensorimotor nervous system in response to somatosensory inputs and paves the way for future translational studies on assessments of sensorimotor impairments using our modeling approach.

Index Terms—EEG, nonlinear system identification, NARMAX, closed-loop system, sensorimotor control.

I. INTRODUCTION

The human nervous system is one of the most complicated systems known. There are at least 10^{11} neurons, 10^{14} interconnections, and many thousand kilometers of cabling in 1.5 kilograms of brain tissue [1, 2]. Our daily activities such as movement and perception are fulfilled by the collective behavior of neural populations. Despite plenty of knowledge of single neuron responses to stimuli, the stimulus-response relation at a system level is not yet fully understood [3]. The system-level neural response is a complex output of collective neuronal activity from neuronal populations and their dynamic interactions, and includes highly nonlinear processes [2, 3]. Assessing the stimulus-response relation between neural populations is essential to a better understanding of the nervous system and could lead to an increased insight of normal and pathological neural functions [4].

A practical way to study a stimulus-response relation in a functionally closed-loop system, like the nervous system, is to use well-designed external perturbations as independent input [3, 5]. Applying periodic mechanical perturbations to the wrist joint of healthy individuals and measuring the associated steady-state cortical responses via electroencephalography (EEG) allows studying the stimulus-response relation in the sensorimotor system [6]. A recent study indicated that over 80%

This work was supported by the Dixon Translational Research Grants Initiative (PI: Y. Yang) from Northwestern Memorial Foundation (NMF) and the Northwestern University Clinical and Translational Sciences (NUCATS) Institute (UL1TR001422), NIH 1R21HD099710 (PIs: Dewald, Yang), USA, and Engineering and Physical Sciences Research Council (EPSRC) under Grant (PI: HLW) EP/I011056/1, and Platform Grant (Co-I: HLW) EP/H00453X/1, U.K. and the Royal Society International Exchanges grant (Co-I: HLW) Ref. IESR3\183107. The research leading to these results has also received funding from the European Research Council under the ERC advanced grant agreement n° 291339 (4D-EEG project, PI: F.C.T. van der Helm).

*Asterisks indicate corresponding authors.

Y. Gu is with the Department of Automatic Control and Systems Engineering, the University of Sheffield, UK. (e-mail: ygu11@sheffield.ac.uk).

*Y. Yang is with the Department of Physical Therapy and Human Movement Sciences, Feinberg School of Medicine, Northwestern University, Chicago, IL 60611 and also with the Stephenson School of Biomedical Engineering, The University of Oklahoma, Tulsa, OK 74135 (e-mail: yuan.yang@northwestern.edu).

J. P. A. Dewald is with the Department of Physical Therapy and Human Movement Sciences, Feinberg School of Medicine, Northwestern University, USA and also with the Department of Biomechanical Engineering, Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, The Netherlands (e-mail: j-dewald@northwestern.edu).

F.C.T. van der Helm is with the Department of Biomechanical Engineering, Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, The Netherlands and also with Department of Physical Therapy and Human Movement Science, Feinberg School of Medicine, Northwestern University, USA (e-mail: f.c.t.vanderhelm@tudelft.nl).

A.C. Schouten is with the Department of Biomechanical Engineering, Faculty of Mechanical, Maritime and Materials Engineering, Delft University of Technology, and Department of Biomechanical Engineering, the MIRA Institute for Biomedical Technology and Technical Medicine, University of Twente, The Netherlands (e-mail: a.c.schouten@tudelft.nl).

*H.-L. Wei is with the Department of Automatic Control and Systems Engineering, the University of Sheffield, S1 3DJ, UK (e-mail: w.hualiang@sheffield.ac.uk).

of cortical responses to mechanical perturbations originates from nonlinear interactions, and a linear model could only explain 10% of cortical response, according to the measured variance accounted for (VAF) [7]. Nonlinear modeling of cortical responses to mechanical perturbations allows for a better understanding of the sensorimotor system and may pave the way for assessments of sensorimotor impairments caused by neurological disorders, such as Parkinson's disease, stroke and cerebral palsy [4, 8].

Our previous studies investigated the cortical responses to mechanical perturbations based on their relative phases [9, 10]. These studies demonstrated the dominance of quadratic nonlinearity in the nervous system. Based on these findings, Vlaar and colleagues modeled cortical responses to wrist perturbations using regularized Volterra series with a second-order nonlinearity [11]. The obtained *subject-specific* models explain around 46% (measured by VAF) of cortical response. This result is better than using a linear model (explaining 10% of cortical response). The study also found that after extensive averaging the recorded cortical response contained around 8% noise, indicating that more advanced methods may be able to model a higher percentage of the response.

Over the past years, many linear modelling techniques have been proposed and successfully applied to neural signal modelling for brain-computer-interface (BCI) research. For example, in [12], a 2-D linear decoupling model was introduced to represent the EEG signals relating to BCI systems. Each of the two sub-models (for the horizontal and vertical velocities of the cursor, respectively) involves a total of $34 \times 11 = 374$ model elements (model terms), which were determined by the number of sensors ($=34$) and the maximum lag ($=10$) for model input variables. In [13], the 2-D model introduced in [12] was extended to 3-D, representing the velocities in the x-, y- and z-axis. Each of the three sub-models involves a total of $64 \times 11 = 704$ model terms. While these models provide a good representation of the relevant EEG signals, they have several limitations, for example, 1) they lack interpretability in that the models include a great number of terms (elements); each of which may just make a trivial contribution to explaining the target signal, and the contributions are unknown. 2) These models cannot reveal the underlying nonlinear relationships between the input and output signals of the systems under study. 3) These models do not answer the question: do subjects participating in experiments share any common features, which are important and useful for future clinical studies e.g. investigating abnormal cortical responses associated with sensorimotor impairments and monitoring the functional changes in the brain after neurological disorders and during the recovery. To overcome the limitations of linear models and obtain more useful information from experimental data, this study proposes a nonlinear modelling approach which can generate parsimonious models. The proposed method will be briefly introduced in the next paragraphs first and described in detail in Section II.

Commonly used nonlinear modeling methods includes regularization regression, sparse regression (e.g. lasso), basis function expansions, neural networks, and linear and nonlinear

autoregressive moving average. These methods have some advantages and disadvantages. For example, neural networks normally show excellent prediction performance, but they could be very complex and takes a large amount of time for training. The regularization regression and lasso methods are efficient for structure detection and model term selection [14]. In basis function expansions, the basis functions may well capture the temporal dynamics without explicitly considering sampling resolution and number of lags. However, these methods might produce less accurate predictions than neural network models. The nonlinear autoregressive-moving-average with exogenous inputs (NARMAX) model provides parsimonious and transparent representation of nonlinear systems and in general shows excellent prediction performance [16,17]. However, the challenges still remain for building a common model structure. Moreover, one needs to construct transparent and parsimonious models where the role of individual system variables, and their interactions are explicitly known, so as to facilitate future translations to clinically related research.

Recently, we proposed a biologically inspired approach based on the prior knowledge of neuroanatomical connections and corresponding transmission delays in neural pathways [15]. However, this previous method has the limitation to be applied to an unknown "black-box" system. For example, it would be hard to apply method to individuals suffering from a stroke, since the damage to neuroanatomical connections and following neural plasticity will result in an unknown system like a "black" box. Thus, a data-based method that is proposed in this study seeks to address this limitation. Furthermore, a common model structure estimated from different healthy subjects may provide a reference of key characteristics shared across individuals. This reference is important for future clinical studies investigating abnormal cortical responses associated with sensorimotor impairments. This will then provide a potential quantitative tool for monitoring the functional changes in the brain after neurological disorders and during the recovery [4]. However, the common model structure cannot be achieved by the previous subject-specific (nonparametric) Volterra models [11] as well as other system identification techniques as discussed above.

In this study, we modeled the cortical response to mechanical perturbation using a polynomial NARMAX method. Such a NARMAX method is used since 1) the "true" mathematical model of the human sensorimotor system is unknown and 2) most nonlinear functions can be approximately represented by a polynomial series. A common model structure detection (CMSD) method is proposed, which allows for the selection of key model terms from many candidates, to build a common model structure for multiple datasets. The proposed method was applied to the open-access datasets previously recorded by Vlaar and colleagues [7, 11]. The datasets are available in Nonlinear System Identification Benchmarks website (<http://www.nonlinearbenchmark.org/#EEG>)

Results obtained from this study can enhance our understanding of the underlying nonlinear behaviors in the human somatosensory central nervous system. The proposed

method would allow us modeling the human somatosensory system in a more precise way than current state of the art approaches with few key parameters. The common model estimated from different subjects provides a useful reference of key characteristics shared across individuals. This may pave the way for our future research that aims to quantitatively assess the pathological changes in the somatosensory system caused by neurological disorders.

II. NONLINEAR MODELLING USING NARMAX

The NARMAX method [16-18] provides a powerful tool for black-box system identification problems where the true model structure is unknown or hard to obtain. A wide range of nonlinear systems can be represented well using NARMAX modeling. The input-output relationship of a nonlinear dynamic system can be represented using polynomial NARMAX model as follows:

$$y(t) = f(y(t-1), \dots, y(t-n_y), u(t-1), \dots, u(t-n_u), e(t-1), \dots, e(t-n_e)) + e(t) \quad (1)$$

where $y(t)$, $u(t)$ and $e(t)$ are the output, input and prediction error, respectively; n_y , n_u , and n_e are the associated maximum lags, and $f(\cdot)$ is a nonlinear function which is unknown in advance and is identified from experimental data using a model structure detection algorithm.

Most existing model structure detection algorithms focus on identifying a model structure based on one single dataset. The orthogonal forward regression (OFR) algorithm is a commonly used method for such a purpose [17]. The OFR operates in a stepwise manner to produce a parsimonious representation of the input-output relation. It first defines a dictionary consisting of a great number of candidate model terms (e.g. individual variables and their interaction terms). Then, an orthogonalization transformation is performed over the dictionary to generate a subset of model terms. During the orthogonalization procedure, a simple and effective error reduction ratio (ERR) index is used to measure the contribution of each model terms. At each step, the algorithm selects a most important model term from the dictionary. The selection procedure normally generates a small subset of model terms which are used for model building. This algorithm and its variants have been successfully applied to studies in various research fields including ecological [19], environmental [20], geophysical [21], societal [22] and neurophysiological sciences [23-26]. The scenario considered in this study, however, is quite different from previous studies. In this study there are multiple datasets recorded from a series of experiments with different inputs (i.e., seven different multi-sine realizations) and from multiple participants. Thus, the single-dataset based OFR algorithm cannot be used to generate the common model structure that represents all datasets (i.e. within and between participants). Therefore, a new method that can effectively handle multiple-dataset modeling problems is needed. Below, we introduce a Common Model Structure Detection (CMSD) method to address this need.

A. Parsimonious Common Model Structure Detection

The nonlinear autoregressive exogenous (NARX) model, as a special case of NARMAX model, is commonly used in nonlinear system identification. It can be expressed in a linear-in-the-parameters form [16, 17, 27]:

$$y(t) = \theta_1 \varphi_1(t) + \theta_2 \varphi_2(t) + \dots + \theta_M \varphi_M(t) + e(t) \quad (2)$$

where $\theta_1, \dots, \theta_M$ are unknown parameters and M is the total number of candidate regressors, $\varphi_1(t), \dots, \varphi_M(t)$ are model terms (also known as regressors) generated from the regressor vector $[y(t-1), \dots, y(t-n_y), u(t-1), \dots, u(t-n_u)]^T$. For example, for a single input and single output (SISO) system (where $u(t)$ and $y(t)$ are the input and output, respectively), if the nonlinear degree is chosen to be 2, and the time lags of input and output are chosen to be $n_u = 2$ and $n_y = 1$, respectively, then the candidate model terms include the constant term, linear terms $y(t-1), u(t-1), u(t-2)$, and nonlinear terms $y(t-1)y(t-1), y(t-1)u(t-1), y(t-1)u(t-2), u(t-1)u(t-1), u(t-1)u(t-2), u(t-2)u(t-2)$.

Considering the scenario where a total number of K datasets is available, our objective is to find a common model structure in the form of Eq. (2) that summarizes the common characteristics across all datasets. For k -th dataset, the model terms $[\varphi_1^{(k)}(t), \dots, \varphi_M^{(k)}(t)]$ can be generated from the associated regressor vector $[y^{(k)}(t-1), \dots, y^{(k)}(t-n_y), u^{(k)}(t-1), \dots, u^{(k)}(t-n_u)]^T$. Here the superscript is used to index the datasets. For example, $\varphi_1^{(k)}$ indicates that the model term is for the k -th dataset.

If using all the available model terms, the k -th datasets can be represented by a full polynomial NARX model:

$$y^{(k)}(t) = \theta_1 \varphi_1^{(k)}(t) + \dots + \theta_M \varphi_M^{(k)}(t) + e^{(k)}(t) \quad (3)$$

Model (3) can be written in a compact matrix format as:

$$\mathbf{y}^{(k)} = \theta_1 \boldsymbol{\varphi}_1^{(k)} + \dots + \theta_M \boldsymbol{\varphi}_M^{(k)} + \mathbf{e}^{(k)} \quad (4)$$

where $\mathbf{y}^{(k)} = [y^{(k)}(1), \dots, y^{(k)}(N^{(k)})]^T$, $\boldsymbol{\theta}^{(k)} = [\theta_1^{(k)}, \dots, \theta_M^{(k)}]^T$, $\mathbf{e}^{(k)} = [e^{(k)}(1), \dots, e^{(k)}(N^{(k)})]^T$ and $\boldsymbol{\varphi}_m^{(k)} = [\varphi_m^{(k)}(1), \dots, \varphi_m^{(k)}(N^{(k)})]^T$ for $k = 1, 2, \dots, K$ and $m = 1, 2, \dots, M$.

The total number of candidate model terms M depends on the number of input variables n_u (i.e., the length of input history), the maximum time lags of the output n_y and the degree of nonlinearity d . It can be calculated that $M = \frac{(n_u+n_y+d)!}{(n_u+n_y)!d!}$. In order to determine the maximum time lags for both the input and output variables, following the approach described in [25], we have carried out pre-modelling experiments and simulations. In this study, the maximum time lag is chosen to be $n_y = 5$ and $n_u = 20$, and based on the observations of Vlaar [8,11] the second order nonlinearity ($d = 2$) was set. Thus, the total number of candidate terms M is $(20+5+2)!/[(20+5)!2!] =$

351. In practice, a smaller number of the significant model terms could be enough to represent the data [17]. Thus, we proposed a model structure detection algorithm to select the key model terms from these 351 candidates.

For the k -th dataset, let $D^{(k)} = \{\varphi_1^{(k)}, \dots, \varphi_M^{(k)}\}$ be the full dictionary of candidate model terms and $\varphi_m^{(k)}$ be the m -th candidate basis vector comprised by the candidate model terms $\delta_m^{(k)}$. Then, the common model structure detection problem is equivalent to finding a common subset $D_n = \{\varphi_{l_1}, \dots, \varphi_{l_n}\}$ with $\{l_1, \dots, l_n\} \in \{1, 2, \dots, M\}$, so that $\mathbf{y}^{(k)}$ ($k = 1, 2, \dots, K$) can be approximated by a linear combination of $\{\varphi_{l_1}, \dots, \varphi_{l_n}\}$, as:

$$\mathbf{y}^{(k)} = \theta_{l_1}^{(k)} \varphi_{l_1} + \dots + \theta_{l_n}^{(k)} \varphi_{l_n} + \mathbf{e}^{(k)} \quad (5)$$

We used a stepwise forward search approach to identify the key model terms in the common model structure. The overall mean absolute error (oMAE) was employed to indicate the significance and contribution of each model term in reducing the modeling error. At the first search step, the oMAE of each candidate model term can be estimated from a MAE matrix:

$$\Psi^{(1)} = \begin{bmatrix} \epsilon_1^{(1)} & \epsilon_2^{(1)} & \dots & \epsilon_M^{(1)} \\ \epsilon_1^{(2)} & \epsilon_2^{(2)} & \dots & \epsilon_M^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ \epsilon_1^{(K)} & \epsilon_2^{(K)} & \dots & \epsilon_M^{(K)} \end{bmatrix} \quad (6)$$

where $\epsilon_m^{(k)}$ is the individual MAE value when the m -th candidate model term is used to predict the k -th output:

$$\epsilon_m^{(k)} = \frac{1}{N_k} \|\mathbf{y}^{(k)} - \alpha_m^{(k)} \varphi_m^{(k)}\|_1 \quad (7)$$

where $\alpha_m^{(k)}$ is the parameter and $\|\cdot\|_1$ represents the L_1 norm. Then, the oMAE associated with the m -th candidate model term $\delta_m^{(k)}$ (i.e., when the m -th candidate basis vector $\varphi_m^{(k)}$ is used to represent all K datasets) is defined as the average of the K MAE values:

$$\bar{\epsilon}_m = \frac{1}{K} (\epsilon_m^{(1)} + \epsilon_m^{(2)} + \dots + \epsilon_m^{(K)}) \quad (8)$$

Define:

$$l_1 = \arg \min_{1 \leq m \leq M} \{\bar{\epsilon}_m\} \quad (9)$$

subsequently, the first significant model term can be selected as $\delta_{l_1}^{(k)}$ and the first associated orthogonal vector can be defined as $\mathbf{q}_1^{(k)} = \varphi_{l_1}^{(k)}$. When $\delta_{l_1}^{(k)}$ is selected by the algorithm, the l_1 -th candidate model term should be removed from the initial dictionary $D^{(k)}$, as well as the corresponding column of matrix $\Phi^{(k)}$ (i.e., the l_1 -th candidate basis vector). After removing $\delta_{l_1}^{(k)}$ from the dictionary, the dictionaries of all K datasets are reduced and consist of only $M - 1$ candidate model terms.

Accordingly, all K matrices associated with these $M - 1$ model terms have $M - 1$ columns.

At a step s ($s \geq 2$), each of the K dictionaries consist of $M - s + 1$ candidate model terms and the bases in each dictionary are first transformed into a new group of orthogonalized bases through the Gram-Schmidt (GS) transformation [13,14,24]. Here, the GS transformation is used to achieve the following objective: to select the most important variables (bases) that are most representative for all the K datasets. The orthogonalization transformation at the step s can be implemented by:

$$\mathbf{q}_j^{(k, s)} = \varphi_j^{(k)} - \sum_{r=1}^{s-1} \frac{(\varphi_j^{(k)})^T \mathbf{q}_r^{(k)}}{(\mathbf{q}_r^{(k)})^T \mathbf{q}_r^{(k)}} \mathbf{q}_r^{(k)} \quad (10)$$

where $\mathbf{q}_r^{(k)}$ ($r = 1, 2, \dots, s - 1$) are orthogonal vectors, $\varphi_j^{(k)}$ ($j = 1, 2, \dots, M - s + 1$) are unselected bases and $\mathbf{q}_j^{(k, s)}$ ($j = 1, 2, \dots, M - s + 1$) are new orthogonalized bases. The MAE matrix at the step s can be then calculated using the new group of K bases $[\mathbf{q}_j^{(1, s)}, \mathbf{q}_j^{(2, s)}, \dots, \mathbf{q}_j^{(K, s)}]$ ($j = 1, 2, \dots, M - s + 1$), and the MAE matrix is:

$$\Psi^{(s)} = \begin{bmatrix} \epsilon_1^{(1)} & \epsilon_2^{(1)} & \dots & \epsilon_{M-s+1}^{(1)} \\ \epsilon_1^{(2)} & \epsilon_2^{(2)} & \dots & \epsilon_{M-s+1}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ \epsilon_1^{(K)} & \epsilon_2^{(K)} & \dots & \epsilon_{M-s+1}^{(K)} \end{bmatrix} \quad (11)$$

The oMAEs of all unselected bases at step s can be calculated and the s -th model term can be selected to be $\varphi_{l_s}^{(k)}$, with:

$$l_s = \arg \min_{1 \leq m \leq M-s+1} \{\bar{\epsilon}_m\} \quad (12)$$

The s -th associated orthogonal vector is defined as $\mathbf{q}_s^{(k)} = \varphi_{l_s}^{(k)}$. Thus, the model terms of the subset $[\varphi_{l_1}^{(k)}, \dots, \varphi_{l_n}^{(k)}]$ can be selected step-by-step to identify the common model structure for all K datasets. The number of model terms is determined by an adjustable prediction sum of squares (APRESS) to achieve a balance between model complexity and model performance. The details of the APRESS can be found in [28].

B. Parameter Estimation

Assume that a total of n model terms are selected, the model parameter vector for the k -th dataset, i.e. $\boldsymbol{\theta}^{(k)} = [\theta_{l_1}^{(k)}, \theta_{l_2}^{(k)}, \dots, \theta_{l_n}^{(k)}]$ can then be estimated from the triangular equation $\mathbf{A}^{(k)} \boldsymbol{\theta}^{(k)} = \mathbf{g}^{(k)}$, where $\mathbf{A}^{(k)}$ is a unity upper triangular matrix, and $\mathbf{g}^{(k)} = \frac{(\mathbf{y}^{(k)})^T \mathbf{q}_j^{(k)}}{(\mathbf{q}_j^{(k)})^T \mathbf{q}_j^{(k)}} (i = 1, \dots, n)$ (see e.g. [26],[27] for details). By averaging the model parameters from the K sub-datasets, we can further obtain a unique model parameter vector $\boldsymbol{\theta} = [\theta_{l_1}, \theta_{l_2}, \dots, \theta_{l_n}]$ for the common model.

C. Noise Modeling

The EEG signal in this study is collected from ten participants, using scalp electrodes. For the data collected from real systems, the noise signal $e(t)$ is usually a colored or correlated noise, which is generally not observed. One of the common approaches to handling noise is to model it using model residuals. A distinctive feature of the NARMAX model, setting it apart from other data based modelling techniques, is that it does not make any assumption on the noise distribution but only treats the noise to be colored.

In this study, noise modelling was incorporated in the NARMAX procedure for each of the individual models. Let $\hat{f}(\cdot)$ represent an estimator for the model $f(\cdot)$, the residuals $\varepsilon(t)$ can then be estimated as

$$\begin{aligned} \varepsilon(t) &= y(t) - \hat{y}(t) \\ &= y(t) - \hat{f}(y(t-1), \dots, y(t-n_y), u(t-1), \dots, u(t-n_u), \varepsilon(t-1), \dots, \varepsilon(t-n_e)) \end{aligned} \quad (13)$$

The algorithm in sections II (A) and (B) includes two extra steps:

- Computing the prediction errors $\varepsilon(t)$,
- Using the value of $\varepsilon(\cdot)$ from the previous iteration to estimate noise model terms in the model $f(\cdot)$.

In most cases, a linear noise model can be used:

$$\varepsilon(t) = \alpha_1 \varepsilon(t-1) + \dots + \alpha_{n_e} \varepsilon(t-n_e) \quad (14)$$

If this is insufficient, then $\varepsilon(t-p)$ for $p = 1, 2, \dots, n_e$ can be included in model (2), where the basic regressor vector is defined as $y(t-1), \dots, y(t-n_y), u(t-1), \dots, u(t-n_u), \varepsilon(t-1), \dots, \varepsilon(t-n_e)$. This will then increase the computational workload for the modelling task of study due to the huge number of candidate variables for each of the 10 modelling cases (related to the 10 participants). The model validity tests [28-30] were used to determine if the process and noise models are adequate.

D. Model Evaluation

We compared the estimated outputs obtained from one-step-ahead (OSA) and k-step-ahead (3-step ahead in this study) predictions with the measured output to evaluate the model.

- (i) 1-step-ahead model predicted output:

$$\hat{y}(t) = f(y(t-1), \dots, y(t-n_y), u(t-1), \dots, u(t-n_u)) \quad (15)$$

- (ii) 2-step-ahead model predicted output:

$$\hat{y}(t+1) = f(\hat{y}(t), y(t-1), \dots, y(t-n_y+1), u(t), u(t-1), \dots, u(t-n_u+1)) \quad (16)$$

- (iii) 3-step-ahead model predicted output:

$$\hat{y}(t+2) = f(\hat{y}(t+1), \hat{y}(t), y(t-1), \dots, y(t-n_y+2), u(t+1), u(t), \dots, u(t-n_u+2)) \quad (17)$$

where $\hat{y}(t)$ represents the model predicted output, while $y(t)$ is the corresponding measured output. We used 1) the correlation coefficient (Corr), 2) the variance accounted for (VAF) and 3) the normalized root means square error (NRMSE) to determine the model performance (see Table I).

III. DATA AND PREPROCESSING

A. Data

In this proof-of-concept study, the experimental data were recorded from ten healthy participants (age range 22-25 years; 5 women; all right-handed) who received continuously angular position perturbations (i.e., the external input to the nervous system) to their right wrist under passive conditions (i.e., the participant had to relax). The experimental procedure was approved by the Human Research Ethics Committee of the Delft University of Technology. All participants signed informed consent before participating in the experiments.

The experimental setup is shown in Fig.1. Perturbation signals were applied to the participants' wrist as an external input to the nervous system via a wrist joint manipulator (Wristalyzer by MOOG Inc, Nieuw-Venep, The Netherlands). Participants were instructed to relax their wrist muscles and not to voluntarily react to the perturbation signal. The wrist is an ideal joint to study the cortical response to external input, since the wrist (and the hand in general) has a large cortical representation. Furthermore, the wrist joint is relatively lightweight and therefore relatively easier to perturb than other joints. Participants were instructed to relax their wrist muscles. Surface electromyography (EMG) was recorded from the flexor carpi radialis and the extensor carpi radialis muscles and online monitored to ensure no voluntary reaction to the perturbation signal. The perturbation signals were periodic multi-sine signals [31], i.e., the sum of multiple sinusoids with the frequencies of 1, 3, 5, 7, 9, 11, 13, 15, 19, and 23 Hz and a period of 1 s. The multi-sine signals have several advantages over other signals (e.g. white noise, step signal) for system identification of the sensorimotor system [7, 11]: 1) Its excited frequencies cover the frequency band of neural activity in the human sensorimotor system [15]; 2) the periodical characteristics of a multi-sine signal allow for leakage-free

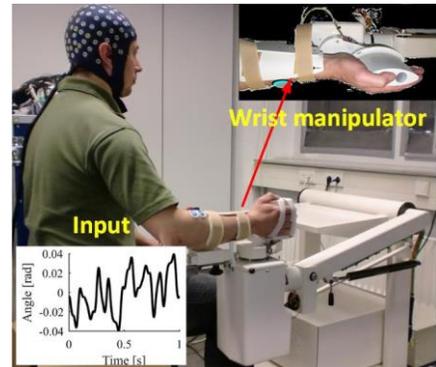


Fig. 1. Experimental setup. Participants were seated with their right forearm attached to an arm support and their hand strapped to the handle of a one-degree freedom wrist joint manipulator. During each realization, a designed multi-sine signal (as shown in the bottom-left for example) was applied as the input to perturb peripheral nervous system sensors via the wrist joint manipulator.

analysis of the steady-state response and averaging to reduce the effect of random signals, e.g. background noise from spontaneous neural oscillations in the brain [32].

In this study, seven different realization of the multi-sine signal (with the same frequencies) were generated using different (random) phase realizations. Our main objective is to evaluate sensory function, therefore we need a small perturbation which is large enough to evoke a cortical response. All perturbation signals had the same root-mean-square of 0.02 radians, with peak-to-peak value less than 0.06 radians. The signals were designed to have the equal power on the first three frequency components (i.e. 1, 3, 5 Hz) and a decaying power spectrum (-20dB/decade slope) for the remaining frequency components, resulting in a flat velocity spectrum for these frequencies. In the time domain, the instantaneous velocity was changing over the time as a multisine with the same frequencies as the perturbation signal, i.e., the first derivative of the perturbation signal. This design is a trade-off between reduced predictability of signal (to prevent the anticipation of participants during the experiment) and the sensitivity of the muscle spindles [7]. Our main objective is to evaluate sensory function. Based on our previous studies, this perturbation is able to evoke a steady-state cortical sensory response [7, 11].

The seven multisine realizations were the identical for all participants but applied in a random order during 49 trials of 36 seconds for each subject. Six seconds were removed from each trial to reduce transient effects, resulting in a total of 1470 recorded periods, i.e. 210 periods for each realization. Cortical responses to the perturbations were recorded by a 126-channel EEG cap (WaveGuard cap, ANT Neuro) according to the 5-10 system with Ag/AgCl electrodes. Both the applied perturbation signal and the recorded EEG signals were sampled at 2048 Hz and stored for offline analyses using a Refa System (TMSi, Oldenzaal, the Netherlands). More details about the datasets can be found in [7].

B. Preprocessing

The preprocessing procedure was in line with previous studies [7, 11] and described below. EEG signals recorded from scalp electrodes have a very poor signal to noise ratio (SNR). Due to the volume conduction EEG signals are “blurred” copies of multiple underlying source activities and noise [33]. We used independent component analysis (ICA) to extract the EEG source activities for modeling purposes. ICA [34] is a widely used preprocessing technique to separate the most important signal contributions from noise by decomposing EEG signals into independent components. Before applying ICA, the continuous EEG signals were filtered by a 1–100 Hz zero-phase shift band-pass filter to remove possible high-frequency noise from neck muscles and slow trends in the data (e.g., blood pressure, heartbeat, breathing and sweat potentials). Notch filters implemented in Fieldtrip toolbox [35] were applied to remove the 50 Hz line power noise and the harmonics. ICA was performed using the Infomax algorithm [36] as implemented in CUDAICA [37]. Subsequently, all signals were resampled to 256 Hz and segmented into 1 s periods, i.e. the period of the perturbation signal. We carefully examined independent

components to identify the components associated with eye movement and muscle artifacts and removed them [38]. The SNRs for rest non-artifact ICA components are calculated using the algorithm from Vlaar et al. (2015) [39]. For each participant, the ICA component with the highest SNR was used as the system output. The contribution (or weights) of this ICA component to EEG channels was projected to the scalp and considered as the spatially distribution of the ICA component in the scalp for building a forward model in the source localization [40]. A dipole fitting algorithm implemented in the Fieldtrip toolbox [35] was used to verify that the sources of all selected components were located in the primary sensorimotor areas in the contralateral hemisphere.

The “true” output signal in the studied system is the perturbation “evoked” cortical activity from the primary sensorimotor areas. This “evoked” cortical activity is phase locked to the perturbation, known as a type of event-related potentials (ERPs) [41–43]. The ERPs are mixed with background “spontaneous” neural activity. Thus, we averaged the signal over perturbation periods to remove the “spontaneous” background noise and extract the ERPs [44], leaving 1 s (256 sampled input-output data points) per dataset as shown in Figure 2. There is a scale difference between the

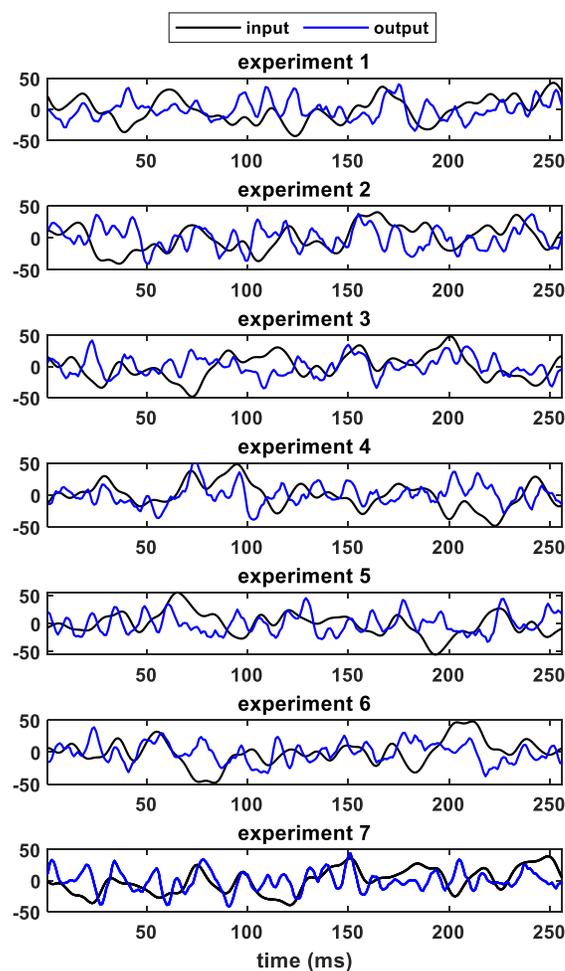


Fig.2. Input-output data pairs of the seven realizations of one representative participant (the input signals were amplified 100 times to make the input and the output in the same scale).

amplitudes of the input (i.e., the mechanical perturbation signal) and output signals (i.e., the IC component of EEG signal) in the original experimental datasets. To avoid the ill-conditioned problem in the relevant procedures (e.g. calculation of designed matrices and associated model parameters), the input signals are scaled up as $u = u' \times 100$, where u is the amplified input signal and u' is the original input signal, so that the amplitude of the input signals used for model identification is at a similar

TABLE II OMAE VALUES AND ERROR REDUCTIONS (ER) OF THE SELECTED 20 COMMON MODEL TERMS (ER= OMAE VALUE OF PREVIOUS TERM - OMAE VALUE OF CURRENT TERM)

| No | MODEL TERMS | OMAE | ER | No | MODEL TERMS | OMAE | ER |
|----|----------------|------|--------|----|----------------|------|--------|
| 1 | y(t-1) | 9.45 | - | 11 | u(t-15)u(t-18) | 5.50 | 0.0291 |
| 2 | y(t-2) | 7.16 | 2.3419 | 12 | u(t-6)u(t-12) | 5.46 | 0.0375 |
| 3 | y(t-3) | 6.37 | 0.7899 | 13 | u(t-1)u(t-8) | 5.43 | 0.0366 |
| 4 | y(t-4) | 6.02 | 0.3456 | 14 | u(t-4)u(t-10) | 5.38 | 0.0411 |
| 5 | y(t-5) | 5.70 | 0.3291 | 15 | u(t-2)u(t-8) | 5.35 | 0.0323 |
| 6 | u(t-7)u(t-14) | 5.65 | 0.0412 | 16 | u(t-4)u(t-5) | 5.30 | 0.0423 |
| 7 | u(t-1)u(t-1) | 5.62 | 0.0311 | 17 | u(t-3)u(t-9) | 5.26 | 0.0455 |
| 8 | u(t-1)u(t-18) | 5.59 | 0.0325 | 18 | constant | 5.23 | 0.0364 |
| 9 | u(t-20)u(t-20) | 5.56 | 0.0312 | 19 | u(t-9)u(t-20) | 5.20 | 0.0317 |
| 10 | y(t-1)y(t-1) | 5.53 | 0.0285 | 20 | u(t-1)u(t-6) | 5.17 | 0.0291 |

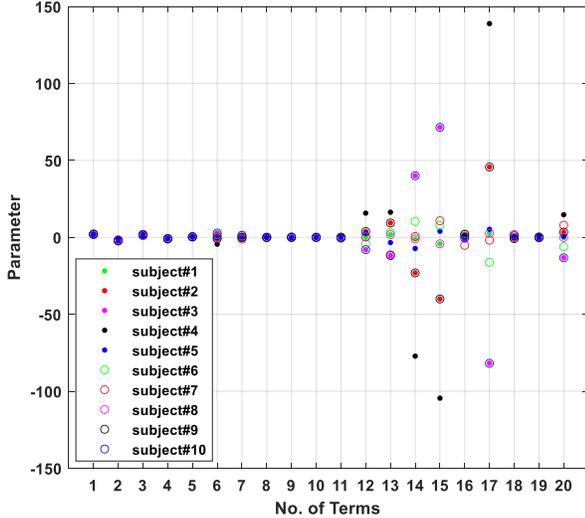


Fig. 3. Estimated parameters of ten models (one per subject) with the common model structure.

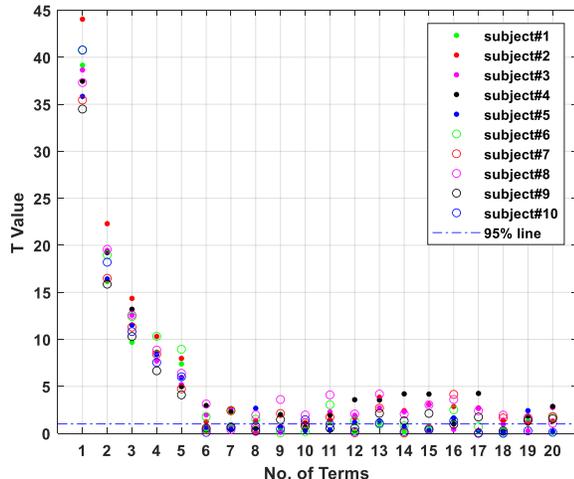


Fig. 4. T values of the 20 terms for the tested ten subjects

scale as that of output signals.

C. Training and Testing Dataset

The cortical responses were recorded from 10 participants. Each participant had 7 datasets according to the 7 different realizations of the multi-sine input signal. Thus, there are 70 datasets in total. The mean SNR across all datasets is 12.5, so the noise is around 8% of the signal. The first six realizations of each participant were used for model identification and the remaining one was used for model evaluation.

In this study, by averaging the model parameters from the 6 estimated datasets for the same participant, we can further obtain a unique model parameter vector for each participant $\theta = [\theta_{l_1}, \theta_{l_2}, \dots, \theta_{l_3}]$.

IV. MODELING RESULTS.

This section presents two identified model for cortical responses to mechanical wrist perturbations, which are the subject-specific structure model and common structure model.

A. Subject-specific Structure Models for Cortical Responses to Mechanical Wrist Perturbations

As a reference, subject-specific NARX models (a NARMAX model specified by Eq. (2)) were first identified for each participant using the OFR algorithm. The number of model terms of each subject-specific model is determined by the APRESS criterion [28], to avoid overfitting. Their performances are shown in Table I, where Corr, VAF and NRMSE represent the ‘correlation coefficient’, ‘variance accounted for’ and ‘normalized root mean square error’, respectively. For Corr and VAF, the higher the value is, the better the performance is. For NRMSE, a lower value indicates a better performance. The model terms for each participant are quite different. The performances of subject-specific models vary from different subjects. The model for the 10th subject has the best performance, while the model for the 3rd subject has the

TABLE I PERFORMANCE STATISTICS OF THE SUBJECT-SPECIFIC STRUCTURE NARX MODELS. NARX: NONLINEAR AUTOREGRESSIVE EXOGENOUS. OSA: ONE-STEP-AHEAD.

| No. OF PARTICIPANT | CORR (NARX OSA) | CORR (NARX K-STEP AHEAD) | VAF (100%) (NARX OSA) | VAF (100%) (NARX K-STEP AHEAD) | NRMSE (NARX OSA) | NRMSE (NARX K-STEP AHEAD) |
|--------------------|-----------------|--------------------------|-----------------------|--------------------------------|------------------|---------------------------|
| P1 | 0.9775 | 0.7976 | 95.55 | 63.57 | 0.0384 | 0.1103 |
| P2 | 0.9749 | 0.6930 | 95.02 | 47.34 | 0.0416 | 0.1359 |
| P3 | 0.9646 | 0.6254 | 93.04 | 34.47 | 0.0458 | 0.1411 |
| P4 | 0.9625 | 0.7030 | 92.57 | 46.97 | 0.0522 | 0.1402 |
| P5 | 0.9725 | 0.8193 | 94.56 | 66.97 | 0.0453 | 0.1109 |
| P6 | 0.9716 | 0.7930 | 94.37 | 62.75 | 0.0441 | 0.1140 |
| P7 | 0.9803 | 0.8397 | 96.08 | 70.11 | 0.0476 | 0.1319 |
| P8 | 0.9607 | 0.6432 | 92.28 | 40.07 | 0.0575 | 0.1601 |
| P9 | 0.9608 | 0.6400 | 92.27 | 39.22 | 0.0521 | 0.1460 |
| P10 | 0.9845 | 0.8772 | 96.92 | 76.92 | 0.0333 | 0.0908 |
| Mean | 0.9710 | 0.7431 | 94.27 | 54.84 | 0.0458 | 0.1281 |
| Std. | 0.0080 | 0.0878 | 1.57 | 14.14 | 0.0067 | 0.0198 |

worst performance. Overall, all the subject-specific models perform well with Corr being over 0.96 and VAF over 93% for one step ahead prediction. For k-step ahead prediction, most of the models perform well, while some of the models can be further improved (e.g. the 3rd model). The person-specific

models will be very interesting in-patient population. However, this study is more focused on identifying a common model structure to reveal the key characteristics shared among all the 10 different healthy subjects, so that a reference can be obtained for future clinical studies. In next section, we will identify the common model to see which terms are shared by all the participants.

B. Common Structure Models for Cortical Responses to Mechanical Wrist Perturbations

A common model structure was built to characterize the cortical response behavior of the 10 participants. The common model structure was identified using the proposed CMSD method based on 60 datasets (first 6 realizations of each of the 10 participant). The duration of each time lag is 3.9 ms. According to the APRESS criterion [28], the optimal number of model terms should be 20. The common model structure includes the most important 20 model terms (regressors) selected from a great number of candidates (i.e. 351 candidates) (see Table II). Although the same model structure was obtained for all participants, subject-specific parameters were estimated to indicate the individual differences (see Fig. 3). As shown in Fig. 3, the variations of estimated parameters of the first 11 and the 18th, 19th model terms are quite small; this means that the system components are similar for all the 10 participants. The variations of the other terms are large; this indicates that the contributions of these terms vary for different participants.

The significance of each model terms is assessed by the proposed oMAE (see Section II-A). The oMAE values of all selected model terms in the common structure are presented in Table II. As shown, the inclusion of each model term progressively reduced the prediction error. Additionally, the t -statistics (with 95% confidence) of each selected model terms are presented in Fig. 4. The t -statistics indicate that the selected model terms are significant for most of the participants. It is worth mentioning that the significance of each model term varies for different participants. While the treatment for each participant should be mainly determined by its most significant model terms, the less significant terms should also be considered. As shown in Table II, the first 5 autoregressive terms are important in reducing the prediction error. However, this does not indicate that a linear auto-regressive (AR) model

is sufficient to describe the system. The VAF of the linear AR model with only the 5 AR terms $y(t-1) \dots y(t-5)$ is only 36.83% in the 3-step ahead prediction. These results show that the inclusion of AR terms has significantly improved the model performance. The reason is that the AR terms are very important in neuron systems. However, the previous models [e.g. [9-11]] do not include the auto-regressive (AR) part. The new NARMAX model can better capture the system feedback components and thus help improve the performance.

We compared the OSA prediction as well as k-step ahead ($k = 3$) model predicted outputs with the measured output using correlation coefficient, VAF and NRMSE to evaluate the models (see Table III). Comparisons of the NARX model predicted output (obtained from the k-step ahead prediction) and the corresponding measured cortical responses are shown in Fig. 5 for the ten participants. As shown in Fig.5, waveforms of predicted outputs and measured cortical responses look very

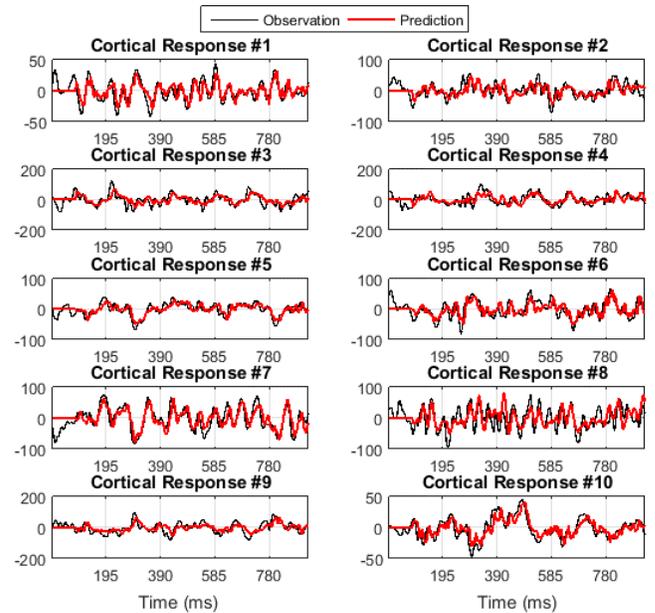


Fig.5. Comparisons of the model predicted outputs (k-step ahead prediction) and the corresponding measurements of cortical responses for the ten participants.

TABLE III PERFORMANCE STATISTICS OF NARX MODELS WITH THE COMMON STRUCTURE. NARX: NONLINEAR AUTOREGRESSIVE EXOGENOUS, OSA: ONE-STEP-AHEAD, CORR: CORRELATION COEFFICIENT, VAF: VARIANCE ACCOUNT FOR, NRMSE: NORMALIZED ROOT MEAN SQUARE ERROR.

| No. of participant | CORR (NARX-OSA) | CORR (NARX-K-STEP AHEAD) | VAF (100%) (NARX-OSA) | VAF (100%) (NARX-K-STEP AHEAD) | NRMSE (NARX-OSA) | NRMSE (NARX-K-STEP AHEAD) |
|--------------------|-----------------|--------------------------|-----------------------|--------------------------------|------------------|---------------------------|
| P1 | 0.9773 | 0.7556 | 95.52 | 57.08 | 0.0397 | 0.1224 |
| P2 | 0.9735 | 0.6366 | 94.74 | 39.53 | 0.0435 | 0.1459 |
| P3 | 0.9642 | 0.5750 | 92.95 | 31.17 | 0.0467 | 0.1437 |
| P4 | 0.9591 | 0.5891 | 91.94 | 32.26 | 0.0543 | 0.1563 |
| P5 | 0.9698 | 0.7848 | 94.04 | 61.57 | 0.0468 | 0.1191 |
| P6 | 0.9681 | 0.7028 | 93.72 | 49.18 | 0.0464 | 0.1323 |
| P7 | 0.9784 | 0.8084 | 95.73 | 65.35 | 0.0487 | 0.1398 |
| P8 | 0.9587 | 0.5952 | 91.90 | 32.57 | 0.0584 | 0.1689 |
| P9 | 0.9607 | 0.6164 | 92.24 | 37.98 | 0.0515 | 0.1461 |
| P10 | 0.9813 | 0.8024 | 96.28 | 64.21 | 0.0362 | 0.1126 |
| Mean | 0.9691 | 0.6866 | 93.91 | 47.09 | 0.0472 | 0.1387 |
| Std. | 0.0079 | 0.0898 | 1.54 | 13.28 | 0.0062 | 0.0165 |

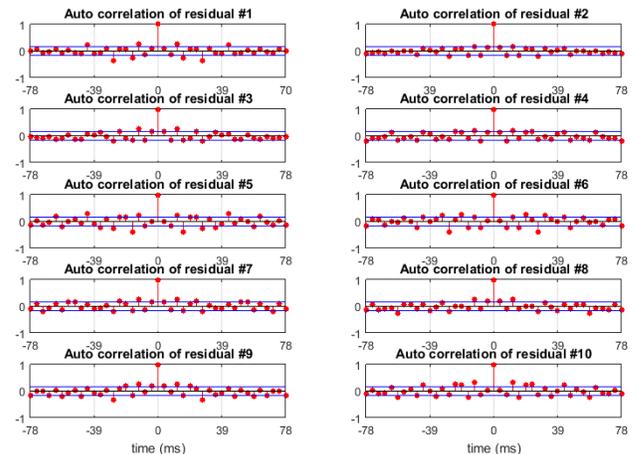


Fig.6. Auto-correlations of the model residuals for the ten participants (blue lines indicate 99% confidence bounds)

similar across participants. We applied Kolmogorov-Smirnov test on the model residual. The results show that the residuals do not follow the normal standard distribution. Since the common model estimation requires that the model fits different data realizations, the model residual may not be a perfect white noise. The Kolmogorov-Smirnov test might be too sensitive for this real data modelling problem. Thus, we used autocorrelation to evaluate if the model can approximately fit the data, with the results shown in Fig. 6. For most participants, the statistically significant non-zero auto-correlation values rarely occur with very small magnitudes, indicating that the estimated NARX models describe the inherent dynamics of the cortical responses well.

V. DISCUSSION

This study focused on modeling the cortical responses to position perturbations applied at the wrist joint. Our results indicated that the cortical response can be well explained by the NARMAX method using a common model structure for all participants.

In modeling, the performance of a common structure model (and using individualized model parameter values) is slightly lower than subject-specific structure models. However, a subject-specific model structure could not summarize common characteristics across subjects. A common model structure attempts to capture the common characteristics shared by and buried in all datasets, by sacrificing local properties hidden in individual datasets. A key advantage of the common model structure for the cortical response is that the model structure reveals the most important inherent features that can explain all data from different participants. Nevertheless, the parameter values may differ from subject to subject when the common model structure is used (see Fig. 3). This result is consistent with previous EEG studies demonstrating individual differences in estimated parameter [45, 46]. The common model structure approach may especially be useful for future pathophysiological research to detect abnormalities after neurological dysfunction.

Following the same procedure as used in the NARX modelling, we investigated the performance of the Volterra models. The subject specific Volterra models achieve average correlation of 0.6625 and VAF of 42.84%. A common structure Volterra models with 20 model terms are also built. The mean correlation coefficient, VAF, and NRMSE are 0.4893, 23.27% and 0.1690, respectively. From these results and that reported in Table III, the NARX models outperform the Volterra models. These indicate that the inclusion of autoregressive terms, as with a NARX model, improves the model prediction performance substantially. It is because that the NARX model structure captures inherent dynamics in the nervous system using autoregressive variables. The OSA yielded much better performances than the k -step ahead for both subject-specific models as well as the common model. The k -step ahead prediction for brain activity is still a recognized challenge in the specific field of brain signal modeling due to the complexity of brain dynamics [47], as well as the poor signal to noise ratio and the non-stationary properties of EEG signals [48]. In this study, the sampling rate of EEG signal is 256 Hz, then each sample time lag is approximately 4 milliseconds (ms). Thus, k -

step ahead prediction actually estimates brain activity based on the measured brain “state”, i.e. the output, around 12 ms ago (in case k is 3 steps).

As shown in Fig. 3, all model terms (except the constant term) are dynamic components with specific time lags. Our dynamic modeling is in line with dynamic properties of the human nervous system summarized in a recent review article [47]. Multiple nonlinear terms and time lags in the common model structure revealed that the processing of somatosensory information in the human nervous system involves multiple neuronal circuitries with different neural transmission delays. These results provide new evidence to support our previous theoretical explanations on neurophysiological mechanisms underlying nonlinear processing of somatosensory information in the human nervous system [4].

The human nervous system receives the mechanical perturbation to the wrist via mechanoreceptors including muscle spindles, Golgi tendon organs, and cutaneous afferents. There are two kinds of sensory fibers in muscle spindles: type Ia primarily sensing muscle stretch velocity and type II primarily sensing muscle stretch. Golgi tendon organ (Ib fibers) detects the tendon strain and as such the force in the muscle-tendon complex. The transmission delays for type Ia fibers are much shorter than those for type II and Ib fibers. Finally, cutaneous afferents (A β fibers) conduct the activity of skin sensors resulting from the mechanical perturbation. When the participants are subjected to the mechanical perturbations, all these sensory fibers are active and sense different modalities with different transmission delays. Nonlinear terms with input signal u are likely associated with nonlinear encoding and processing of external inputs in the nervous system. Different time lags in these nonlinear terms (e.g. $u(t-2)*u(t-8)$) may be related to different transmission delays in the sensory input pathways from the mechanoreceptors to the brain. The individual differences are reflected on the subject-specific parameters (see Fig. 3.). However, invasive recording or animal models will be needed in the future to further interpret the relation between specific fibers and the terms in the common model structure, as well as the individual difference.

In the model, we also found (AR) terms with output signal y , both linear (e.g., $y(t-5)$) and nonlinear (e.g. $y(t-1)y(t-1)$). These output related terms indicate that both linear and nonlinear neuronal interactions occur at the cortex, presumably caused by cortical neural networks or the inherent dynamics of the cortical processes. Nevertheless, the linear terms have much large weights than the nonlinear terms (see Fig. 3), indicating the dominance of the linear terms in the AR part of the model. This result is in line with our recent brain network modeling study, showing that the local neuronal interaction at the cortex may be dominated by linear interactions [49].

In this study, we used EEG source component obtained by independent component analysis (ICA) instead of raw EEG in our modeling as we explained in III.B Preprocessing. The raw, single-trial EEG data has strong background noise, so it is not suitable for a modeling study. We used a series preprocessing steps (as detailed in III.B Preprocessing) to improve its SNR, so as to avoid overfitting in the modeling. The proposed modeling method includes the history of the output signal in the prediction. That allows us to capture system dynamics, which is important in the modeling. The proposed method has the

potential to advance brain signal modeling. It may have clinical value in assessing sensorimotor impairments, since previous studies have indicated clinical relevance of cortical response to somatosensory input in stroke rehabilitation [49, 50].

However, we acknowledge that the multi-step ahead prediction is still a recognized challenge in time series forecasting, especially for cortical activity. In the future, we will work on improving the long-term prediction performance of the common structure models.

VI. CONCLUSION

This study modeled the nonlinear cortical responses to wrist position perturbations using the NARMAX method. Different from previous studies, we used a common model structure, with individualized parameter values, to describe the data for all participants. The identified common model generates good model predictions (OSA and k-step-ahead) for the cortical responses and reveals the most important model terms which can explain system behaviors of all participants. Our results suggest that the measured cortical response is a mixed outcome of the nonlinear transformation of the external input and local neuronal interaction or inherent neuronal dynamics at the cortex. This proof-of-concept study may bring us with a useful tool to improve our understanding of the human sensorimotor system.

VII. ACKNOWLEDGMENT

Authors would like to thank Dr. Martijn Vlaar for collecting the experimental data. Authors also thank the associate editor and anonymous reviewers for their constructive comments

REFERENCES

- [1] C. Koch and G. Laurent, "Complexity and the nervous system," *Science*, vol. 284, pp. 96-98, 1999.
- [2] E. R. Kandel, J. H. Schwartz, T. M. Jessell, D. o. Biochemistry, M. B. T. Jessell, S. Siegelbaum, *et al.*, *Principles of neural science* vol. 4: McGraw-hill New York, 2000.
- [3] A. J. Langdon, T. W. Boonstra, and M. Breakspear, "Multi-frequency phase locking in human somatosensory cortex," *Progress in biophysics and molecular biology*, vol. 105, pp. 58-66, 2011.
- [4] Y. Yang, J. Dewald, F. C. van der Helm, and A. C. Schouten, "Unveiling neural coupling within the sensorimotor system: directionality and nonlinearity," *European Journal of Neuroscience*, vol. 48, pp. 2407-2415, 2018.
- [5] C. S. Herrmann, "Human EEG responses to 1–100 Hz flicker: resonance phenomena in visual cortex and their potential correlation to cognitive phenomena," *Experimental brain research*, vol. 137, pp. 346-353, 2001.
- [6] S. F. Campfens, A. C. Schouten, M. J. van Putten, and H. van der Kooij, "Quantifying connectivity via efferent and afferent pathways in motor control using coherence measures and joint position perturbations," *Experimental brain research*, vol. 228, pp. 141-153, 2013.
- [7] M. P. Vlaar, T. Solis-Escalante, A. N. Vardy, F. C. van der Helm, and A. C. Schouten, "Quantifying nonlinear contributions to cortical responses evoked by continuous wrist manipulation," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, pp. 481-491, 2017.
- [8] Y. Yang, B. Guliyev, and A. C. Schouten, "Dynamic causal modeling of the cortical responses to wrist perturbations," *Frontiers in neuroscience*, vol. 11, p. 518, 2017.
- [9] Y. Yang, T. Solis-Escalante, F. C. van der Helm, and A. C. Schouten, "A generalized coherence framework for detecting and characterizing nonlinear interactions in the nervous system," *IEEE Transactions on Biomedical Engineering*, vol. 63, pp. 2629-2637, 2016.
- [10] Y. Yang, T. Solis-Escalante, J. Yao, A. Daffertshofer, A. C. Schouten, and F. C. van der Helm, "A general approach for quantifying nonlinear connectivity in the nervous system based on phase coupling," *International Journal of Neural Systems*, vol. 26, p. 1550031, 2016.
- [11] M. P. Vlaar, G. Birpoutsoukis, J. Lataire, M. Schoukens, A. C. Schouten, J. Schoukens, *et al.*, "Modeling the nonlinear cortical response in EEG evoked by wrist joint manipulation," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2017.
- [12] T. J. Bradberry, R. J. Gentili, and J. L. Contreras-Vidal, "Fast attainment of computer cursor control with noninvasively acquired brain signals," *Journal of Neural Engineering*, vol. 8, p. 036010, 2011.
- [13] J.-H. Kim, F. Bießmann, and S.-W. Lee, "Decoding Three-Dimensional Trajectory of Executed and Imagined Arm Movements From Electroencephalogram Signals," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 23, pp. 867 - 876, 2015.
- [14] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 58, pp. 267-288, 1996.
- [15] R. Tian, Y. Yang, F. C. van der Helm, and J. Dewald, "A novel approach for modeling neural responses to joint perturbations using the NARMAX method and a hierarchical neural network," *Frontiers in computational neuroscience*, vol. 12, p. 96, 2018.
- [16] S. Chen and S. Billings, "Representations of non-linear systems: the NARMAX model," *International Journal of Control*, vol. 49, pp. 1013-1032, 1989.
- [17] S. Chen, S. A. Billings, and W. Luo, "Orthogonal least squares methods and their application to non-linear system identification," *International Journal of control*, vol. 50, pp. 1873-1896, 1989.
- [18] S. Billings and I. Leontaritis, "Input-output parametric models for non-linear systems Part I: deterministic non-linear systems," *International Journal of Control*, vol. 41, pp. 303-328, 1985.
- [19] A. M. Marshall, G. R. Bigg, S. M. Leeuwen, J. K. Pinnegar, H. L. Wei, T. J. Webb, *et al.*, "Quantifying heterogeneous responses of fish community size structure using novel combined statistical techniques," *Global change biology*, vol. 22, pp. 1755-1768, 2016.
- [20] G. Bigg, H. Wei, D. Wilton, Y. Zhao, S. Billings, E. Hanna, *et al.*, "A century of variation in the dependence of Greenland iceberg calving on ice sheet surface mass balance and regional climate change," in *Proc. R. Soc. A*, 2014, p. 20130662.
- [21] M. Balikhin, R. Boynton, S. Walker, J. Borovsky, S. Billings, and H. Wei, "Using the NARMAX approach to model the evolution of energetic electrons fluxes at geostationary orbit," *Geophysical Research Letters*, vol. 38, 2011.
- [22] Y. Gu and H.-L. Wei, "Significant Indicators and Determinants of Happiness: Evidence from a UK Survey and Revealed by a Data-Driven Systems Modelling Approach," *Social Sciences*, vol. 7, p. 53, 2018.
- [23] Y. Zhao, S. A. Billings, H. L. Wei, and P. Sarrigiannis, "A parametric method to measure time-varying linear and nonlinear causality with applications to EEG data," *IEEE Transactions on Biomedical Engineering*, vol. 60, pp. 3141-3148, 2013.
- [24] F. He, H. L. Wei, S. A. Billings, and P. Sarrigiannis, "A nonlinear generalization of spectral granger causality," *IEEE Transactions on Biomedical Engineering*, vol. 6, pp. 1693-1701, 2014.
- [25] Y. Li, H. L. Wei, S. A. Billings, and P. Sarrigiannis, "Identification of nonlinear time-varying systems using an online sliding-window and common model structure selection (CMSS) approach with applications to EEG," *International Journal of Systems Science*, vol. 47, pp. 2671-2681, 2015.
- [26] J. R. Ayala Solares, H. L. Wei, and S. A. Billings, "A novel logistic-NARX model as a classifier for dynamic binary classification," *Neural Computing and Applications*, vol. 31, pp. 11-25, 2019.
- [27] H. L. Wei and S. A. Billings, "Model structure selection using an integrated forward orthogonal search algorithm assisted by squared correlation and mutual information," *International Journal of Modelling, Identification and Control*, vol. 3, pp. 341-256, 2008.
- [28] S. Billings and H. Wei, "An adaptive orthogonal search algorithm for model subset selection and non-linear system identification," *International Journal of Control*, vol. 81, pp. 714-724, 2008.

- [29] S. Billings and Q. Zhu, "Nonlinear model validation using correlation tests," *International Journal of Control*, vol. 60, pp. 1107-1120, 1994.
- [30] S. Billings and W. Voon, "Correlation based model validity tests for nonlinear models," *International Journal of Control*, vol. 44, pp. 235-244, 1986.
- [31] R. Pintelon and J. Schoukens, *System identification: a frequency domain approach*: John Wiley & Sons, 2012.
- [32] M. Bach and T. Meigen, "Do's and don'ts in Fourier analysis of steady-state potentials," *Documenta Ophthalmologica*, vol. 99, pp. 69-82, 1999.
- [33] B. Graimann and G. Pfurtscheller, "Quantification and visualization of event-related changes in oscillatory brain activity in the time-frequency domain," *Progress in brain research*, vol. 159, pp. 79-97, 2006.
- [34] S. Makeig, A. J. Bell, T.-P. Jung, and T. J. Sejnowski, "Independent component analysis of electroencephalographic data," in *Advances in neural information processing systems*, 1996, pp. 145-151.
- [35] R. Oostenveld, P. Fries, E. Maris, and J.-M. Schoffelen, "FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data," *Computational intelligence and neuroscience*, vol. 2011, p. 1, 2011.
- [36] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural computation*, vol. 7, pp. 1129-1159, 1995.
- [37] F. Raimondo, J. E. Kamienskowski, M. Sigman, and D. F. Slezak, "CUDAICA: GPU optimization of infomax-ICA EEG analysis," *Computational intelligence and neuroscience*, vol. 2012, p. 2, 2012.
- [38] J. Tzzy-Ping, C. Humphries, T.-W. Lee, S. Makeig, M. J. McKeown, V. Iragui, *et al.*, "Extended ICA removes artifacts from electroencephalographic recordings," *Advances in neural information processing systems*, pp. 894-900, 1998.
- [39] M. P. Vlaar, F. C. van der Helm, and A. C. Schouten, "Frequency domain characterization of the somatosensory steady state response in electroencephalography," *IFAC-PapersOnLine*, vol. 48, pp. 1391-1396, 2015.
- [40] K. E. Hild II and S. S. Nagarajan, "Source localization of EEG/MEG data by correlating columns of ICA and lead field matrices," *IEEE Transactions on Biomedical Engineering*, vol. 56, pp. 2619-2626, 2009.
- [41] U. R. Acharya, E. C.-P. Chua, K. C. Chua, L. C. Min, and T. Tamura, "Analysis and automatic identification of sleep stages using higher order spectra," *International journal of neural systems*, vol. 20, pp. 509-521, 2010.
- [42] M. P. Vlaar, T. Solis-Escalante, A. N. Vardy, F. C. T. van der Helm, and A. C. Schouten, "Quantifying Nonlinear Contributions to Cortical Responses Evoked by Continuous Wrist Manipulation," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, pp. 481-491, 2016.
- [43] A. Starr, B. McKeon, N. Skuse, and D. Burke, "Cerebral potentials evoked by muscle stretch in man," *Brain: a journal of neurology*, vol. 104, pp. 149-166, 1981.
- [44] S. J. Luck, *An introduction to the event-related potential technique*: MIT press, 2014.
- [45] Y. Yang, S. Chevallier, J. Wiart, and I. Bloch, "Time-frequency optimization for discrimination between imagination of right and left hand movements based on two bipolar electroencephalography channels," *EURASIP Journal on Advances in Signal Processing*, vol. 2014, p. 38, 2014.
- [46] Y. Yang, S. Chevallier, J. Wiart, and I. Bloch, "Subject-specific time-frequency selection for multi-class motor imagery-based BCIs using few Laplacian EEG channels," *Biomedical Signal Processing and Control*, vol. 38, pp. 302-311, 2017.
- [47] M. Breakspear, "Dynamic models of large-scale brain activity," *Nature neuroscience*, vol. 20, pp. 340-352, 2017.
- [48] K. F. K. Wong, A. Galka, O. Yamashita, and T. Ozaki, "Modelling non-stationary variance in EEG time series by state space GARCH model," *Computers in biology and medicine*, vol. 36, pp. 1327-1335, 2006.
- [49] O. G. Filatova, Y. Yang, J. P. Dewald, R. Tian, P. Maceira-Elvira, Y. Takeda, *et al.*, "Dynamic information flow based on EEG and diffusion MRI in stroke: a proof-of-principle study," *Frontiers in neural circuits*, vol. 12, p. 79, 2018.
- [50] S. F. Campfens, C. G. Meskers, A. C. Schouten, M. J. van Putten, and H. van der Kooij, "Stretch evoked potentials in healthy subjects and after stroke: a potential measure for proprioceptive sensorimotor function," *IEEE transactions on neural systems and rehabilitation engineering*, vol. 23, pp. 643-654, 2015.