MODELING NONLINEAR EVOKED HEMODYNAMIC RESPONSES IN FUNCTIONAL ULTRASOUND

Sofia-Eirini Kotti[⋆∗] *Aybüke Erol*[⋆] *Borbála Hunyadi*[⋆]

[⋆] EEMCS, Delft University of Technology, Delft, The Netherlands.

ABSTRACT

Functional ultrasound (fUS) is a high-sensitivity neuroimaging technique that images cerebral blood volume changes, which reflect neuronal activity in the corresponding brain area. fUS measures hemodynamic changes which are typically modeled as the output of a linear time-invariant system, characterized by an impulse response known as the hemodynamic response function (HRF), and a binary representation of the stimulus signal as input. In this work, we quantify the difference between a linear and a nonlinear time-invariant HRF model in terms of data fitting and prediction performance. Our results on fUS data obtained from two mice reveal that: (a) including nonlinearities in the HRF achieves a significantly more precise modeling of the fUS signal compared to the linear assumption under certain stimulus conditions and (b) a second-order Volterra series approximation can be used to characterize the nonlinear model and predict responses to stimuli.

Index Terms— Functional ultrasound, (nonlinear) hemodynamic response, Volterra series.

1. INTRODUCTION

Functional ultrasound (fUS) is an emerging neuroimaging modality that provides an indirect measure of neuronal activity by recording changes in cerebral blood volume. Similarly to functional magnetic resonance imaging (fMRI), this is possible through the phenomenon of neurovascular coupling (NVC), which describes the link between fluctuations in local neuronal activity and the resulting changes in blood flow.

The majority of studies addressing the NVC assume a linear time-invariant (LTI) model for the hemodynamic response, which can be described by an impulse response known as the hemodynamic response function (HRF). Estimating the HRF, and thus uncovering the dynamics of the NVC, is a research topic in hemodynamics-based imaging modalities for decades. Although the hemodynamic response is known to exhibit nonlinear characteristics, linear modeling has maintained its popularity due to its good performance and simplicity. However, under certain circumstances, deviations from the behavior of an LTI system can become too severe to neglect, examples of which include experiments with closely repeated stimuli [1], or under changing duration [2] or strength [3] of stimuli.

Volterra series are high order extensions of the Taylor series. A Volterra series of order one simply describes a one-dimensional convolution operation; higher order Volterra kernels, e.g., second-order kernels that describe pairwise interactions of the input signal at different time instances, were found to sufficiently capture nonlinearities of the fMRI signal caused by interactions between successive stimuli in an auditory experiment [1].

Although research on HRF identification using fMRI data is prolific, the corresponding research using fUS data is increasing only as of recently, see e.g., [4, 5]. The contribution of this paper is twofold. On one hand, fUS measurements on mice during a visual experiment with long duration stimuli were used to show and quantify under which stimulus conditions the fUS data can be better approximated using a nonlinear HRF, as opposed to a linear one. On the other hand, the applicability of Volterra series to characterize the nonlinear aspects of the HRF was shown by estimating the kernels using a series of different stimulus durations and predicting individual responses.

2. MATERIALS AND METHODS

2.1. Theoretical background

The complexity of biological systems makes it difficult to derive analytic equations that describe neuronal dynamics adequately. Related work can be found in fMRI literature, e.g., [6–8]; however, there is no such work specifically for fUS, to the authors' knowledge.

Alternatively, the HRF can be estimated with data-driven approaches. When a causal LTI system with finite memory T is assumed in discrete time, the system model is $y(k) \approx \sum_{k_1=0}^{T-1} h^{(1)}(k_1)u(k-k_1)$, where $y(k)$ is the measured hemodynamic response at time instant k, $h^{(1)}(k)$ is the HRF and the input $u(k)$ is usually the stimulus signal used in the experiment, represented as a binary block-type vector equal to 1 when the stimulus is on and 0 during rest periods [5, 9].

Common approaches in fMRI literature include: a) assuming an a priori shape of the HRF (usually the canonical

[∗]This work is supported by the TU Delft AI Labs programme.

HRF model [10]) and estimating only its amplitude, b) estimating parameters that define the shape of a template HRF (e.g., time-to-peak and duration) [11] and c) expressing the HRF as a linear combination of several basis functions and estimating the best regression coefficients [11]; this approach also results in physiologically meaningful shapes with carefully chosen basis functions.

2.2. Data acquisition and preprocessing

For our HRF modeling on fUS data, two experiments (on two different mice and brain slices) were conducted at the Center for Ultrasound and Brain imaging at Erasmus University Medical Center (CUBE). During data acquisition, the subject mouse was presented with randomly generated high-contrast images (a rectangular patch of white "speckles" against a black background) on two stimulation screens simultaneously. In the first experiment (Exp. 1), the stimulus was shown with 4 different durations (1, 4, 10 and 20 seconds), 10 repetitions each, in random order. Each repetition was followed by a rest period (completely black screens) of random duration between 20 and 25 seconds. In the second experiment (Exp. 2), the durations were 0.25, 0.5, 1, 2, 3, ..., 10 seconds and there were 11 repetitions. For Exp. 1, we report results on all durations; for Exp. 2, we report results on 1, 4, 8 and 10 seconds.

The time series of a pixel corresponds to its power variation across the power Doppler image (PDI) stream [5]. The final frame rate was 4 Hz in Exp. 1 and 7.44 Hz in Exp. 2. The region of interest (ROI) per experiment was selected as an area of 16 pixels in the left superior colliculus, a region known to be involved in the processing of visual stimuli [12]. Pixels with a Pearson correlation coefficient (PCC) larger than 0.3 with the stimulus are indicated in Fig. 1 for Exp. 1.

The next step of data preprocessing included standardizing each ROI pixel time series and averaging them over the ROI to obtain a single time series. This signal was subsequently subjected to low-pass filtering using a fifth-order Butterworth filter with 0.3 Hz cutoff frequency, in order to remove high frequency noise components [13]. We applied baseline correction to each stimulus repetition separately.

In Fig. 2, the resulting signals per duration are shown for Exp. 1, after averaging over the multiple stimulus repetitions. The response to 1 sec stimulus shows a single peak before it returns to the baseline level. The observed signal takes more intricate shapes for the stimuli of 10 and 20 sec: the responses drop after the initial peak and then increase again. This raised the question of whether such responses can be modeled as the convolution of an HRF with a block-type signal. A delayed component in the response to strong inputs (high concentration odors) has also been observed in [4], where it was modeled using a second linear HRF term that peaks at a late point. However, the authors assume that the neuronal Ca^{2+} signal is the driver of that LTI system, which is not block type.

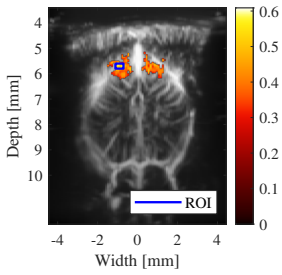

Fig. 1: ROI in experiment 1.

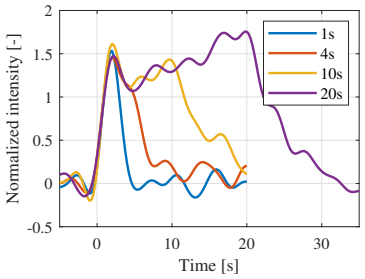

Fig. 2: Average filtered ROI response.

2.3. Volterra series

In this work, we evaluate the LTI assumption and further explore Volterra series [1] as an alternative to perform nonlinear system identification given fUS measurements. The Volterra series is an extension of the Taylor series to cover dynamic systems and can represent any analytic time-invariant system [14]. In our case, the second-order Volterra series approximation is given by

$$y(k) \approx h^{(0)} + \sum_{k_1=0}^{T-1} h^{(1)}(k_1)u(k-k_1) + \sum_{k_1=0}^{T-1}\sum_{k_2=0}^{T-1} h^{(2)}(k_1,k_2)u(k-k_1)u(k-k_2), \quad (1)$$

where $y(k)$ for $k = 1,\dots,K$ is the hemodynamic response captured in the fUS signal, $h^{(m)}(\cdot)$, $m = 0,1,2$, is the m-th order Volterra kernel and $u(k)$ is the block-type stimulus signal.

In order to estimate the coefficients of this expansion using linear estimation methods and ensure that the resulting kernel shapes will not be arbitrary, we expand the kernels in terms of L temporal basis functions $b_i(k)$, $i = 1,\dots,L$, following [1]. The kernels are expanded as

$$\begin{aligned} h^{(0)} &= g^{(0)} \\ h^{(1)}(k_1) &= \sum_{i=1}^{L} g_i^{(1)} b_i(k_1) \\ h^{(2)}(k_1,k_2) &= \sum_{i=1}^{L}\sum_{j=1}^{L} g_{ij}^{(2)} b_i(k_1) b_j(k_2). \end{aligned} \quad (2)$$

Using the one-dimensional convolution of the input $u(k)$ and the basis functions $b_i(k)$, we can define a new set of response variables $x_i(k)$ as $x_i(k) = \sum_{k_1=0}^{T-1} b_i(k_1)u(k-k_1)$. Substituting this into Eq. (1) gives

$$y(k) \approx g^{(0)} + \sum_{i=1}^{L} g_i^{(1)} x_i(k) + \sum_{i=1}^{L}\sum_{j=1}^{L} g_{ij}^{(2)} x_i(k)x_j(k). \quad (3)$$

Stacking all $y(k)$ into the observation vector $\mathbf{y} \in \mathbb{R}^{K\times 1}$ and the L' explanatory variables 1, $x_i(k)$ and $x_i(k)x_j(k)$ as

columns of a design matrix $\mathbf{X} \in \mathbb{R}^{K \times L'}$, we obtain the model

$$\mathbf{y} = \mathbf{X}\mathbf{g} + \mathbf{e}, \tag{4}$$

where the error vector $\mathbf{e} \in \mathbb{R}^{K \times 1}$ includes possible measurement noise and uncaptured higher order dynamics. The vector $\mathbf{g} \in \mathbb{R}^{L' \times 1}$ includes all unknown coefficients $g^{(0)}$, $g^{(1)}$, and $g^{(2)}$ from which the kernel coefficients $h^{(m)}(\cdot)$, $m = 0, 1, 2$, can be calculated using Eq. (2).

The basis functions in this work were chosen to follow the linear HRF form assumed in [4, 5], that is

$$b(k; \boldsymbol{\theta}) = \theta_1 (\Gamma(\theta_2))^{-1} \theta_3^{\theta_2} k^{\theta_2 - 1} \mathrm{e}^{-\theta_3 k}, \tag{5}$$

for $k \geq 0$, where $\boldsymbol{\theta} = [\theta_1,\ \theta_2,\ \theta_3]^{\mathrm{T}}$ and $\theta_1, \theta_2, \theta_3 > 0$ control the response height, the delay and the dispersion of the function [5]. The parameters of the basis functions were fixed so that the functions peak during different possible components of the hemodynamic response, i.e., early, intermediate and late components, and were not optimized for.

The following optimization problem is used to estimate $\hat{\mathbf{g}}$

$$\begin{aligned} \underset{\mathbf{g}}{\text{minimize}} \quad & \|\mathbf{X}\mathbf{g} - \mathbf{y}\|_2 + \lambda \|\mathbf{g}\|_1 \\ \text{subject to} \quad & \textstyle\sum_{i=1}^{L} g_i^{(1)} b_i(k) \geq 0 \quad \forall k. \end{aligned} \tag{6}$$

The ℓ_1-norm regularization is used to promote sparsity in the solution, so that interpretable shapes for the HRF can be obtained, given the number of columns in the design matrix. The parameter λ controlling the regularization was set empirically to 10^{-4} (for unit energy $\mathbf{X}$ and $\mathbf{y}$). The constraint requires that the first order kernel $h^{(1)}(k)$ be positive for all k, as the linear fUS HRF is assumed to be always positive [4, 5].

3. RESULTS

3.1. Training per stimulus duration

In order to evaluate and compare the potential of a linear and a nonlinear HRF to describe the obtained fUS measurements, we solve (6) assuming a maximum Volterra order $M = 1$ (linear) and $M = 2$ (nonlinear) per each stimulus duration separately. Given the estimated coefficients $\hat{\mathbf{g}}$, we then reconstructed the observation vector $\mathbf{y}$ as $\hat{\mathbf{y}} = \mathbf{X}\hat{\mathbf{g}}$.

The mean squared error (MSE) achieved by the reconstruction for $M = 1$ and $M = 2$, is given in Tables 1 and 2, for the two experiments respectively. The difference in the reconstruction achieved by the two in the case of stimulus of 1 sec is rather low. The MSE improvement is much larger for longer stimuli. Focusing on the 4 sec stimulus, in Exp. 2 the MSE improvement achieved when incorporating second-order kernels is comparable to that achieved for 1 sec stimulus. In Exp. 1 the improvement factor is much larger. This difference does not come as a surprise, as it is known that the hemodynamic response can differ per subject. These results show that the responses to longer stimuli can be reconstructed more faithfully when a maximum Volterra order $M = 2$, thus a nonlinear HRF, is used. The above observation is justified by the fact that a binary stimulus in combination with an LTI hemodynamic system cannot account for responses that depart from the form of the 1 sec response in Fig. 2.

Table 1: Reconstruction MSE - Exp. 1.

Stimulus duration [sec]	**MSE** $[\times 10^{-5}]$		Improvement factor $[\times 10^3]$
	$M = 1$	$M = 2$	
1	9.014	0.0931	0.097
4	64.672	0.005	12.992
10	31.868	0.007	4.569
20	19.239	0.003	7.746

Table 2: Reconstruction MSE - Exp. 2.

Stimulus duration [sec]	**MSE** $[\times 10^{-6}]$		Improvement factor $[\times 10^3]$
	$M = 1$	$M = 2$	
1	1.845	0.293	0.006
4	31.183	0.081	0.385
8	290.847	0.005	62.371
10	436.196	0.004	98.407

3.2. Training and testing across stimulus durations

After confirming that the Volterra series can be used to model hemodynamic responses to long durations of this visual stimulus, the next step was to investigate whether kernels trained on multiple durations can predict the responses to individual stimuli. Fig. 3 shows the 8 basis functions that were used; their peaks were spaced at 2 seconds.

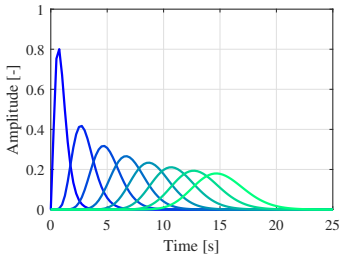

Fig. 3: Basis functions.

Using the data of Exp. 1, the temporally averaged responses to the stimulus durations of 1, 4, 10, and 20 seconds (one time series per duration) were concatenated, as well as the stimuli themselves. It should be noted that the energy of each response was matched to the energy of the corresponding stimulus, to alleviate differences between responses to different durations. During the training phase, a single set of Volterra kernels was obtained by solving 6 using the concatenated signals (70% of the repetitions were averaged) for $M = 1$ and $M = 2$ separately. These kernels were then used on the testing set (30% of the repetitions) to predict

individual responses to stimuli of 1, 4, 10 and 20 seconds. The performance was evaluated against the average of the repetitions in the testing set, and also against the individual repetitions. The prediction results can be found in Fig. 4. In Tables 3 and 4 the PCC between the true response and the predicted response is given, for the average data and each repetition separately.

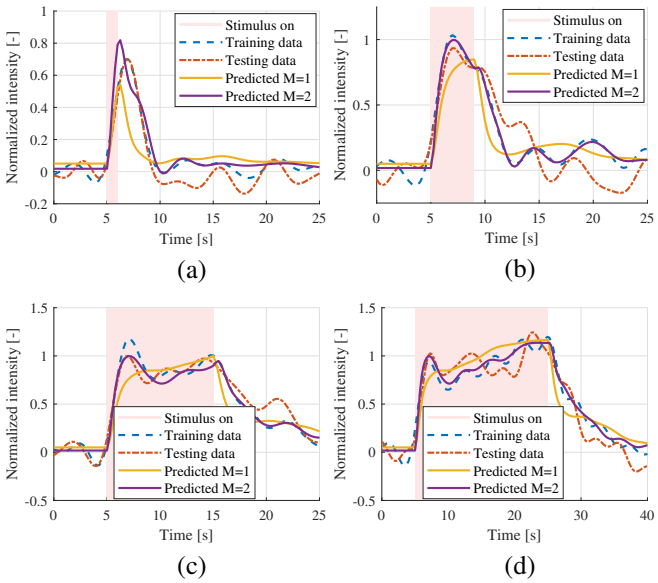

Fig. 4: Training, testing and predicted signals for maximum Volterra order $M = 1$ and $M = 2$ in Exp. 1, for stimulus duration 1 sec (a), 4 sec (b), 10 sec (c) and 20 sec (d), when all durations are used to estimate the kernels. The testing data shown is the average.

Notable is the underestimation of the first peak in the responses for $M = 1$ in Fig. 4, compared to $M = 2$, and the noncapture of the dip after the peak, although the overall duration of the response is correctly predicted. Contrary to [4], where HRF amplitude can vary based on the stimulus, we assume a fixed amplitude for the kernels over all durations. This leads to $M = 2$ being able to capture amplitude differences in the responses better than $M = 1$. When it comes to the PCC results, there is improvement in all durations (averaged or individual data) when kernels are trained with $M = 2$. Training with $M = 1$ using all durations seems to provide rather insufficient kernels for the 1 sec stimulus.

Table 3: PCC between the actual and the predicted responses, $M = 1$.

Stimulus duration [sec]	Testing data average	Testing data 1	Testing data 2	Testing data 3
1	0.773	0.740	0.681	0.830
4	0.826	0.733	0.735	0.858
10	0.876	0.859	0.662	0.783
20	0.914	0.868	0.877	0.839

The estimated kernels for $M = 2$ are shown in Fig. 5. It is clear that the first-order kernel is mostly concentrated in the first 5 seconds, but also shows fluctuations in the tail. This is a result of training on all durations and using basis

Table 4: PCC between the actual and the predicted responses, $M = 2$.

Stimulus duration [sec]	Testing data average	Testing data 1	Testing data 2	Testing data 3
1	0.928	0.898	0.890	0.875
4	0.880	0.773	0.825	0.870
10	0.944	0.901	0.675	0.920
20	0.961	0.897	0.918	0.905

functions that extend in time. This will, in turn, affect the prediction for the shorter durations, which are overfitted in the tail (Fig. 4(a), (b)). The second-order kernel is also mostly concentrated in the first 5 seconds. The negative values in this area suggest that if the stimulus has been on in the last few seconds, the hemodynamic response will be suppressed. The estimated kernels in Exp. 2 are not included, but overall agree that the larger part of the hemodynamic response is contained within the first 5 seconds.

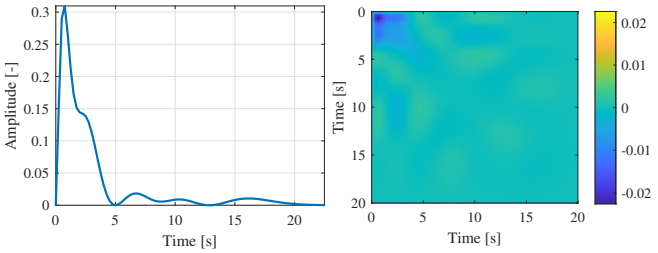

Fig. 5: Estimated kernels $h^{(1)}(\cdot)$ (a) and $h^{(2)}(\cdot)$ (b), for maximum Volterra order $M = 2$ in Exp. 1.

4. CONCLUSIONS

The aim of this work was to complement ongoing modeling efforts in the fUS domain. We drew attention to the fact the nonlinearities in the responses due to stimulus interactions over time can be significant and should be considered when designing experimental paradigms and modeling the hemodynamic system. Including nonlinearities in the model is always expected to better fit the hemodynamic response in general, since the complexity of such a model is higher. In this work, we reported the improvement in the data fitting MSE achieved for longer stimuli compared to shorter stimuli in two experiments. This difference is because the hemodynamic response tends to follow a more complex shape in case of longer stimuli, which cannot be justified by the combination of an LTI system and a binary input signal. Finally, the results show that a Volterra series approximation can be used to model the nonlinear character of the HRF. When trained using the responses to stimuli of different durations, it is possible to predict individual responses to stimuli with good accuracy.

5. ACKNOWLEDGMENTS

The authors would like to thank CUBE for providing the fUS data (www.ultrasoundbrainimaging.com).

6. REFERENCES

[1] K. J. Friston, O. Josephs, G. Rees, and R. Turner, "Nonlinear event-related responses in fMRI," *Magnetic Resonance in Medicine*, vol. 39, no. 1, pp. 41–52, 1998.

[2] Nikos K. Logothetis, "The underpinnings of the BOLD functional magnetic resonance imaging signal," *The Journal of Neuroscience*, vol. 23, pp. 3963 – 3971, 2003.

[3] Geoffrey M. Boynton, Stephen A. Engel, Gary H. Glover, and David J. Heeger, "Linear systems analysis of functional magnetic resonance imaging in human V1," *Journal of Neuroscience*, vol. 16, no. 13, pp. 4207–4221, 1996.

[4] Ali-Kemal Aydin, William Haselden, Yannick Houssen, Christophe Pouzat, Ravi Rungta, Charlie Demené, Mickaël Tanter, Patrick Drew, Serge Charpak, and Davide Boido, "Transfer functions linking neural calcium to single voxel functional ultrasound signal," *Nature Communications*, vol. 11, 06 2020.

[5] Aybuke Erol, Chagajeg Soloukey, Bastian Generowicz, Nikki van Dorp, Sebastiaan K. E. Koekkoek, Pieter Kruizinga, and Borbála Hunyadi, "Deconvolution of the functional ultrasound response in the mouse visual pathway using block-term decomposition.," *Neuroinformatics*, 2022.

[6] K.J. Friston, A. Mechelli, R. Turner, and C.J. Price, "Nonlinear responses in fMRI: The Balloon model, Volterra kernels, and other hemodynamics," *NeuroImage*, vol. 12, no. 4, pp. 466–477, 2000.

[7] Ying Zheng, Yi Pan, Sam Harris, Steve Billings, Daniel Coca, Jason Berwick, Myles Jones, Aneurin Kennerley, David Johnston, Chris Martin, Ian M. Devonshire, and John Mayhew, "A dynamic model of neurovascular coupling: Implications for blood vessel dilation and constriction," *NeuroImage*, vol. 52, no. 3, pp. 1135–1147, 2010, Computational Models of the Brain.

[8] Theodore Huppert, Monica Allen, Heval Benav, Phill Jones, and David Boas, "A multicompartment vascular model for inferring baseline and functional changes in cerebral oxygen metabolism and arterial dilation," *Journal of Cerebral Blood Flow and Metabolism*, vol. 27, pp. 1262–79, 07 2007.

[9] Kevin M Aquino, Mark M Schira, Peter A Robinson, Peter M Drysdale, and Michael Breakspear, "Hemodynamic traveling waves in human visual cortex," *PLoS Computational Biology*, vol. 8, no. 3, pp. e1002435, 2012.

[10] K.J. Friston, P. Fletcher, O. Josephs, A. Holmes, M.D. Rugg, and R. Turner, "Event-related fMRI: Characterizing differential responses," *NeuroImage*, vol. 7, no. 1, pp. 30–40, 1998.

[11] Martin A. Lindquist, Ji Meng Loh, Lauren Y. Atlas, and Tor D. Wager, "Modeling the hemodynamic response function in fMRI: Efficiency, bias and mis-modeling," *NeuroImage*, vol. 45, no. 1, Supplement 1, pp. S187–S198, 2009, Mathematics in Brain Imaging.

[12] Ana F Oliveira and Keisuke Yonehara, "The mouse superior colliculus as a model system for investigating cell type-based mechanisms of visual motor transformation," *Frontiers in Neural Circuits*, vol. 12, pp. 59, 2018.

[13] Anwar O. Nunez-Elizalde, Michael Krumin, Charu Bai Reddy, Gabriel Montaldo, Alan Urban, Kenneth D. Harris, and Matteo Carandini, "Neural correlates of blood flow measured by ultrasound," *Neuron*, vol. 110, no. 10, pp. 1631–1640.e4, 2022.

[14] Jonathan Wray and Gary G. R. Green, "Calculation of the Volterra kernels of non-linear dynamic systems using an artificial neural network," *Biological Cybernetics*, vol. 71, pp. 187–195, 2004.