Functional Near Infrared Spectroscopy (fNIRS) synthetic data generation

Darren J Leamy, Tomás E Ward and Kevin T Sweeney

Abstract— Accurately modelled computer-generated data can be used in place of real-world signals for the design, test and validation of signal processing techniques in situations where real data is difficult to obtain. Bio-signal processing researchers interested in working with fNIRS data are restricted due to the lack of freely available fNIRS data and by the prohibitively expensive cost of fNIRS systems. We present a simplified mathematical description and associated MATLAB implementation of model-based synthetic fNIRS data which could be used by researchers to develop fNIRS signal processing techniques. The software, which is freely available, allows users to generate fNIRS data with control over a wide range of parameters and allows for fine-tuning of the synthetic data. We demonstrate how the model can be used to generate raw fNIRS data similar to recorded fNIRS signals. Signal processing steps were then applied to both the real and synthetic data. Visual comparisons between the temporal and spectral properties of the real and synthetic data show similarity. This paper demonstrates that our model for generating synthetic fNIRS data can replicate real fNIRS recordings.

I. INTRODUCTION

Functional Near Infrared Spectroscopy (fNIRS) is a non-invasive optical imaging method for monitoring and recording haemodynamic activity in cortical areas of the brain [1], [2]. It is an increasingly popular tool for functional neuroimaging partly due to it's advantages over other modalities such as functional Magnetic Resonance Imaging (fMRI) and Electroencephalography (EEG). The advantages of fNIRS over fMRI include lower cost, portability, higher temporal resolution and lower sensitivity to motion artefact. Advantages of fNIRS over EEG include non-sensitivity to electromagnetic interference, potentially higher spatial resolution and lower sensitivity to motion artefact. fNIRS operates by measuring changes in light intensity transmitted through the scalp and cortex of a subject using two or more wavelengths of near infrared light. From these measurements, changes in concentration of oxygenated haemoglobin (HbO) and deoxygenated haemoglobin (HbR) are calculated. This haemodynamic signal is not unlike that recorded by the BOLD signal in fMRI [3].

Although fNIRS systems are considerably cheaper than fMRI, they are still prohibitively expensive for many researchers. There exists a significant biosignal processing community who, if they had access to representative data, could make major contributions to the area by improving signal processing approaches. Unlike EEG, there are currently no widely available fNIRS data sets. An alternative to prerecorded data sets is synthetic data generation as has been explored for fNIRS [4] and in other physiological measurement domains such as electrocardiography (ECG) [5]. Synthetic signal generation models offer several advantages over prerecorded signal sets. For example, there is no limit to the amount of data that can be generated. Synthetic data generation also allows a user the flexibility to fine tune all parameters of the data to meet their requirements. Additionally, synthetic data can be used as an educational tool to explain the composition of real signals.

This paper describes a simple mathematical model and software implementation for generating synthetic fNIRS data. The solution is capable of generating signals that closely resemble recorded fNIRS signals. The approach is driven through a combination of known physiological models and established light transport relationships.

II. METHODS

A. Acquired data

Generation of synthetic fNIRS data was based on real fNIRS data recorded during an earlier experiment. The real data, representative of typical fNIRS signals, guided our selection of parameters for generating synthetic data. This fNIRS data was recorded using a TechEn CW6 system (TechEn Inc., USA), which used two wavelengths (690 nm and 830 nm) for operation at a sample rate of 25 Hz.

The subject was seated in a comfortable chair in front of a computer monitor. Instructions appeared on the monitor for the subject to either engage in self-paced dominant-hand finger-tapping or to rest. Instruction periods lasted 10 seconds and 40 instruction periods were carried out. 7 channels of fNIRS data were recorded from the subject's dominant motor cortex. We used the best performing channel to guide our synthetic data generation, which sampled location C3 of the 10-20 system.

Typical fNIRS processing steps have been reported in detail elsewhere [6]. In short, the transmitted light intensity measurements are low-pass filtered (0.8 Hz), then normalized, then high-pass filtered (0.01 Hz) and then the negative logarithm is taken of these signals. This produces the Optical Density (OD) signal of the two wavelengths. The OD signals are then used in conjunction with the Modified Beer-Lambert Law to calculate the concentration signals of HbO and HbR [7], [8].

B. Synthetic data process

We utilize a hybrid model of the fNIRS signal process consisting of a physiological model of the vascular response together with a spectrophotometric model which relates haemodynamics to changes in optical properties. A typical optical intensity measurement contains the cortical haemodynamic response signal of interest buried in the much stronger interference and noise signals, due to the nature of fNIRS recordings.

Light emitted into the scalp passes through the superficial layers before and after passing through the cortex [9]. These superficial layers contain blood vessels, which affect the intensity of the transmitted light. The largest components of the superficial signal are the cardiac pulse and the Mayer wave. An fNIRS measurement also contains broadband noise arising from environmental optical background and instrumental noise. Lastly, fNIRS signals have an associated offset due to non-varying aspects of the recording channel such as skull thickness, skin thickness, hair and source-detector separation. We therefore model our synthetic raw intensity signals as a sum of these separate components:

$$\Phi_s^\lambda(t) = \Phi_b^\lambda(t) + \Phi_c^\lambda(t) + \Phi_m^\lambda(t) + \Phi_n^\lambda(t) + \Phi_o^\lambda \qquad (1)$$

where $\Phi_s^\lambda(t)$ is our synthetic intensity signal for a given wavelength λ, $\Phi_b^\lambda(t)$ is the raw intensity component as a result of the physiological response of interest, $\Phi_c^\lambda(t)$ is the cardiac pulse component, $\Phi_m^\lambda(t)$ is the Mayer wave component, $\Phi_n^\lambda(t)$ is the broadband noise component and Φ_o^λ is the offset. If desired, additional components can be incorporated into this model.

1) Balloon model simulation: The response due to activation, as measured in fNIRS, is due to the vascular response. There are several models describing the neurophysiological dynamics linking mental activation and the haemodynamics. We chose the well-known Balloon model, which is a biomechanical model of haemodynamic activity at the neuronal level in the cortex during activation [10]. We used a differential equation description as per [11] for the model:

$$E(t) = 1 - (1 - E_0)^{\frac{1}{f_{in}(t)}} \qquad (2)$$

$$\dot{q}(t) = \frac{f_{in}(t)}{\tau_0}\left[\frac{E(t)}{E_0} - \frac{q(t)}{v(t)}\right] + \frac{1}{\tau_v}\left[f_{in}(t) - v^{\frac{1}{\alpha}}\right]\frac{q(t)}{v(t)} \qquad (3)$$

$$\dot{v}(t) = \frac{1}{\tau_v}\left[f_{in}(t) - v^{\frac{1}{\alpha}}\right] \qquad (4)$$

$$\dot{p}(t) = \frac{1}{\tau_v}\left[f_{in}(t) - v^{\frac{1}{\alpha}}\right]\frac{p(t)}{v(t)} \qquad (5)$$

where E, q, v and p denote oxygen extraction factor, normalised dexoxygenated haemoglobin concentration ($\Delta c_{HbR}(t)$), normalised blood volume and normalised total haemoglobin concentration respectively. Normalised oxygenated haemoglobin concentration ($\Delta c_{HbO}(t)$) is obtained by subtracting q from p. The $f_{in}(t)$ waveform (blood flow into the "balloon") is specified by the user and normally reflects the stimulation sequence. Other parameter values of the Balloon model are: $E_0 = 0.4$, $\tau_0 = 2$, $\tau_v = 30$ and $\alpha = 0.4$, similar to [11]. Non-normalised values of concentration can't be found without knowledge of initial concentrations or sample volume. The signals ($\Delta c_{HbR}(t)$) and ($\Delta c_{HbO}(t)$)

are therefore scaled to a typical amplitude before further processing.

2) Conversion to raw intensity signals: Using the modelled HbO and HbR concentration signals, we work backwards through the usual fNIRS processing steps to obtain ideal raw intensity signals. First, the Modified Beer-Lambert law is employed to obtain modelled Optical Density signals:

$$\Delta A_b^\lambda(t) = (\varepsilon_{HbR}^\lambda \Delta c_{HbR}(t) + \varepsilon_{HbO}^\lambda \Delta c_{HbO}(t))BL \qquad (6)$$

where $\Delta A_b^\lambda(t)$ is Optical Density signal of a wavelength (Absorbance Units (AU), a dimensionless unit), $\varepsilon_{HbR}^\lambda$ and $\varepsilon_{HbO}^\lambda$ are the molar extinction coefficients of HbR and HbO of a particular wavelength ($cm^{-1}\,mol^{-1}\,l$), $\Delta c_{HbR}(t)$ and $\Delta c_{HbO}(t)$ are the changes in concentration of HbR and HbO ($mol\,l^{-1}$), B is the differential pathlength factor (unitless) and L is source-detector separation (cm). Values for ε have been tabulated for a wide range of wavelengths [14].

Obtaining an approximation of the raw intensity signal ($\Phi_b^\lambda(t)$) requires an exponential function (described in [6]):

$$\Phi_b^\lambda(t) = exp(-\Delta A_b^\lambda(t)) \qquad (7)$$

Simulation of equations (2) to (7) produces an ideal model of the raw intensity measurement during activation. Interference, noise and offset must be added to this signal to replicate real-world fNIRS raw intensity measurements.

3) Cardiac interference: Cardiac pulse interference of the raw signals was generated based on observed interference patterns in previously recorded data. We had previously recorded a "shallow" NIRS signal from the superficial layers of a subject's scalp. This NIRS recording was dominated by the cardiac signal which gave us a better picture of the temporal features of cardiac pulsation as recorded by an fNIRS system. We modelled cardiac interference as a function approximating the observed cardiac signal. The generated cardiac signal is a scaled function of a dynamic heart rate signal:

$$\Phi_c^\lambda(t) = K_c^\lambda f(k(t), R(t)) \qquad (8)$$

where Φ_c^λ is the cardiac pulse interference for a given wavelength, K_c^λ is the amplitude of the cardiac interference, $k(t)$ is based on a piecewise linear model reflecting a single cardiac pulse and $R(t)$ is a time-varying heart rate signal. $f(.)$ is the non-linear, unity-amplitude, cardiac interference signal which is identical for all wavelengths.

4) Mayer wave interference: Mayer waves are spontaneous changes in arterial blood pressure, oscillating at a frequency near 0.1 Hz in most subjects [12]. Mayer wave interference impacts fNIRS measurements similarly to cardiac interference. We model the Mayer wave as a low-frequency sinusoidal oscillation with time-varying frequency, similar to [13]:

$$\Phi_m^\lambda(t) = K_m^\lambda \sin(2\pi f_m t + \theta_m) \qquad (9)$$

where $\Phi_m^\lambda(t)$ is the Mayer wave interference for a specific wavelength, K_m^λ is the amplitude of the interference, f_m is the time-varying frequency of the Mayer wave and θ_m is the random phase shift.

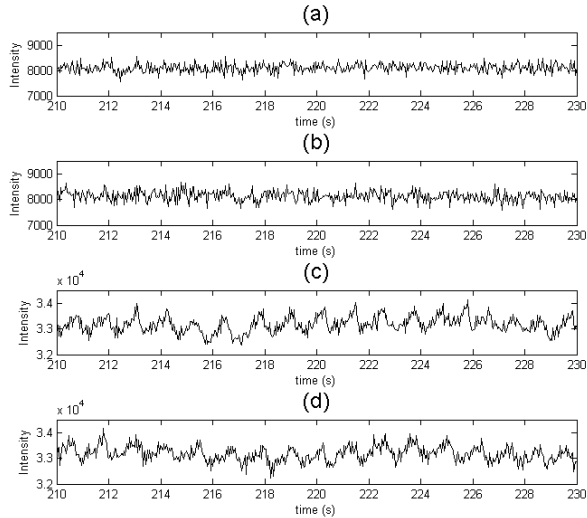Fig. 1. (a) Real 690 nm. (b) Synthetic 690 nm. (c) Real 830 nm. (d) Synthetic 830 nm.

Fig. 2. (a) Real 690 nm. (b) Synthetic 690 nm. (c) Real 830 nm. (d) Synthetic 830 nm.

5) Broadband noise: An fNIRS system records broadband noise due to the environmental optical background and instrumental noise. Based on observations of real fNIRS data we modelled this interference as a normally distributed noise signal:

$$\Phi_n^\lambda(t) = N(0, (\sigma_n^\lambda)^2)$$

where Φ_n^λ is the noise signal for a wavelength and σ_n^λ is the standard deviation of the normal distribution, which is manually selected.

6) Offset: Raw light intensity measurements have an associated offset due to non time-varying aspects of the transmission medium of an fNIRS recording channel such as skin thickness, skull thickness, hair follicles and source-detector separation. Values of offset (Φ_o^λ) are wavelength-dependant and are manually selected.

C. Generating synthetic data

Using our synthetic data generation program, we produced synthetic fNIRS signals and attempted to tune the parameters to match our real data based primarily on visual comparison. For the Balloon model, we modelled the blood in-flow, $f_{in}(t)$, as a trapezoid with a rise time of 5 seconds, plateau time of 5 seconds, fall time of 5 seconds, rest time of 5 seconds and a plateau height of 1.7 for a 10 second active period followed by a 10 second rest period. This was then repeated to match the number of active and rest periods of the real fNIRS data. The HbO signal was scaled to an amplitude of $2e^{-5}$ and HbR was scaled to an amplitude of $0.8e^{-5}$. Our real fNIRS signals used wavelengths of 690 nm and 830 nm so we then obtained the corresponding Optical Density signals for these wavelengths using the modified Beer-Lambert law. For this, we set B to 5.93 in accordance with [14] and set L to 3 cm as this was the source-detector
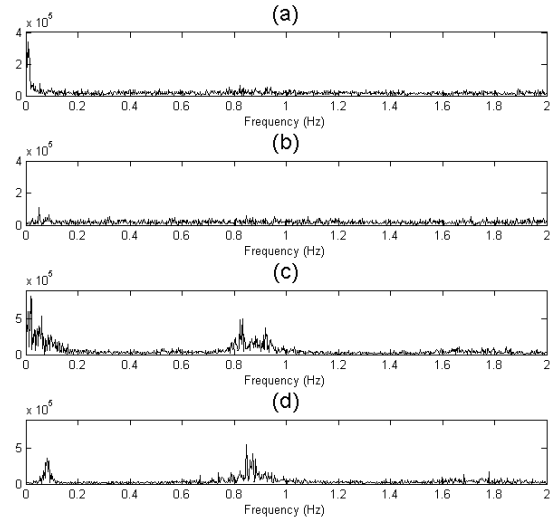
separation of the real data. We then calculated the ideal raw intensity measurements, $\Phi_b^{690}(t)$ and $\Phi_b^{830}(t)$.

To generate the cardiac interference intensity signals, $\Phi_c^{690}(t)$ and $\Phi_c^{830}(t)$, we set the cardiac amplitudes, K_c^{690} and K_c^{830}, to 70 and 700 respectively and set the normally-distributed heart rate parameter $R(t) = N(52, 5)$. To generate the Mayer wave interference intensity signals, $\Phi_m^{690}(t)$ and $\Phi_m^{830}(t)$, we set the Mayer wave amplitudes, A_m^{690} and A_m^{830}, to 25 and 160 respectively and set the normally-distributed oscillation frequency $f_m = N(0.08, 0.01)$Hz. We set both our noise signals, $\Phi_n^{690}(t)$ and $\Phi_n^{830}(t)$, standard deviations, σ_n^{690} and σ_n^{830}, to 199.5. Finally, we set the offset of the raw intensity signals to match the average amplitude of the real data, $\Phi_o^{690} = 8053$ and $\Phi_o^{830} = 33124$.

The modelled raw intensity signals, interference signals, noise signals and offset were summed to produce synthetic raw fNIRS data to match the real fNIRS data.

D. Process synthetic data

The real and synthetic data was visually compared in the temporal and frequency domain to help fine-tune parameters to improve similarity. The amplitude of the haemodynamic signal of interest is too small to be noticed in a temporal or spectral plot. Therefore the typical fNIRS processing steps and averaging over trials are applied to the synthetic data to demonstrate that the HbO/HbR activity is present in the synthetic signal.

III. Results

The aim of this paper is to generate synthetic fNIRS data that resembles real-world fNIRS data as closely as possible, using our model. To demonstrate the results of this work, we present a comparison of the temporal and spectral features of our synthetic data with the real fNIRS data. Presented in Fig
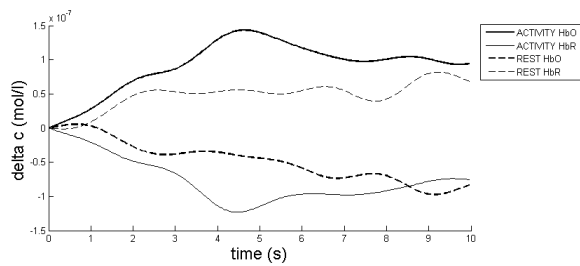
Fig. 3.   Average modelled HbO and HbR responses during rest and activity.

1 is real and synthetic temporal data for both wavelengths. Presented in Fig 2 is real and synthetic spectral data for both wavelengths (offset removed for clarity). Fig 3 displays the average responses of HbO and HbR during both active and rest periods.

IV. DISCUSSION

Visually, the real and synthetic data are very alike. Temporally, both 830 nm signals feature cardiac interference, Mayer wave interference and broadband noise. Spectrally, the center frequencies of cardiac pulsation and the Mayer wave are evident. The synthetic 690 nm data also temporally matches the real signal with low power of cardiac and Mayer wave interference. Due to the low power of cardiac and Mayer wave interference, a spectral comparison between synthetic and real data is difficult. The spectral plots of the real data also feature near-DC components. These components were not from a known physiological source, so we chose to not include them in the model.

The most useful feature of this model for synthetic fNIRS data generation is the flexibility afforded when constructing the data. All parameters can be fine-tuned to replicate a real fNIRS signal. This may be useful for signal processing researchers who wish to investigate the effect of varying a single parameter on an fNIRS recording. A model for synthetic fNIRS generation such as the one presented here would mean that a researcher could save time and effort by not having to run an experiment many times over with different parameters. For example, the function describing the flow of blood in the Balloon model ($f_{in}(t)$) could be modulated over time to observe the effect.

Movement artefact could be incorporated into this model by adjusting the L parameter of the modified Beer-Lambert law over time. Although fNIRS systems are less sensitive to movement, this type of artefact does affect fNIRS recordings. Incorporating movement into our model could be very useful for biomedical signal researchers attempting to improve signal quality.

V. CONCLUSIONS

We have described here a method for generating synthetic raw light intensity measurements with typical characteristics of those produced by an fNIRS system during adult motor activation stimulus. The synthetic signals are based on observations of recorded fNIRS signals. All parameters of the synthetic signals can be tuned by a user to replicate other fNIRS signals. Synthetic data such as this may be used for *in silico* experiments, research and also as an educational tool. We believe that this method for generating synthetic fNIRS data can be expanded upon and improved to more closely model fNIRS data and be of more use to researchers in this field.

VI. ACKNOWLEDGMENTS

REFERENCES

[1] Y. Hoshi, Functional near-infrared spectroscopy: current status and future prospects., *J. Biomed. Opt.*, vol. 12, 2007, pp 062106.

[2] M. Wolf and M. Ferrari and V. Quaresima, Progress of near-infrared spectroscopy and topography for brain and muscle clinical applications. *J. Biomed. Opt.*, vol. 12, 2007, pp 062104.

[3] T. J. Huppert and R. D. Hoge and S. G. Diamond and M. A. Franceschini and D. A. Boas, A temporal comparison of BOLD, ASL, and NIRS hemodynamic responses to motor stimuli in adult humans., *Neuroimage*, vol. 29, 2006, pp 368–382.

[4] U. E. Emir and C. B. Akgül and A. Akin and A. Ertüzün and B. Sankur and K. Harmanci, Wavelet denoising vs ICA denoising for functional optical imaging., *Proceedings of First International IEEE EMBS Conference on Neural Engineering, 2003*, 2003, pp 384–387.

[5] P. E. McSharry and G. D. Clifford and L. Tarassenko and L. A. Smith, A dynamical model for generating synthetic electrocardiogram signals., *IEEE Transactions on Biomedical Engineering*, vol. 50, 2003, 289–294.

[6] T. J. Huppert and S. G. Diamond and M. A. Franceschini and D. A. Boas, HomER: a review of time-series analysis methods for near-infrared spectroscopy of the brain., *Appl Opt*, vol. 48, 2009, pp D280–D298.

[7] D. J. Leamy and T. E. Ward, A novel co-locational and concurrent fNIRS/EEG measurement system: Design and initial results, *Proceedings of Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, 2010, 4230–4233.

[8] F. Matthews and B. A. Pearlmutter and T. E. Ward and C. Soraghan and C. Markham, Hemodynamics for Brain-Computer Interfaces., *IEEE Signal Processing Magazine*, vol. 25, 2008, pp 87–94.

[9] E. Okada and M. Firbank and M. Schweiger and S. R. Arridge and M. Cope and D. T. Delpy, Theoretical and experimental investigation of near-infrared light propagation in a model of the adult head., *Appl. Opt.*, vol. 36, 1997, pp 21–31.

[10] R. B. Buxton and E. C. Wong and L. R. Frank, Dynamics of blood flow and oxygenation changes during brain activation: the balloon model., *Magn. Reson. Med.*, vol. 39, 1998, pp 855–864.

[11] X. Cui and S. Bray and A. L. Reiss, Functional near infrared spectroscopy (NIRS) signal improvement based on negative correlation between oxygenated and deoxygenated hemoglobin dynamics., *Neuroimage*, vol. 49, 2010, pp 3039–3046.

[12] C. Julien, The enigma of Mayer waves: Facts and models., *Cardiovasc. Res.*, vol. 70, 2006, pp 12–21.

[13] S. Coyle and T. Ward and C. Markham, Physiological noise in near-infrared spectroscopy: implications for optical brain computer interfacing., *Conf Proc IEEE Eng Med Biol Soc.*, 2004, pp 4540–4543.

[14] M. Cope, The application of near infrared spectroscopy to non invasive monitoring of cerebral oxygenation in the newborn infant., PhD. Thesis, University College London, 1991.