Classification of EEG signals based on mean-square error optimal time-frequency features

Rachele Anderson
Mathematical Statistics
Centre for Mathematical Sciences
Lund University, Sweden
Email: rachele@maths.lth.se

Maria Sandsten
Mathematical Statistics
Centre for Mathematical Sciences
Lund University, Sweden
Email: sandsten@maths.lth.se

***Abstract*—This paper illustrates the improvement in accuracy of classification for electroencephalogram (EEG) signals measured during a memory encoding task, by using features based on a mean square error (MSE) optimal time-frequency estimator. The EEG signals are modelled as Locally Stationary Processes, based on the modulation in time of an ordinary stationary covariance function. After estimating the model parameters, we compute the MSE optimal kernel for the estimation of the Wigner-Ville spectrum. We present a simulation study to evaluate the performance of the derived optimal spectral estimator, compared to the single windowed Hanning spectrogram and the Welch spectrogram. Further, the estimation procedure is applied to the measured EEG and the time-frequency features extracted from the spectral estimates are used to feed a neural network classifier. Consistent improvement in classification accuracy is obtained by using the features from the proposed estimator, compared to the use of existing methods.**

I. Introduction

The analysis of electroencephalography signals (EEG) is one of the main methodological tools in understanding how cognitive functions are supported by the electrical activity of the brain, [1]. The study of a time-frequency image is often the method of choice to address key issues in cognitive electrophysiology. Clearly, the quality of the time-frequency representation is crucial for the extraction of robust and relevant features, [1]–[4], thus leading to the demand for highly performing spectral estimators. In this paper, we present the improvements in classification accuracy for EEG signals measured during a memory encoding task, [5], by using time-frequency features based on a mean square error (MSE) optimal time-frequency estimator.

We consider a stochastic parametric model for the signals, based on the definition of Locally Stationary Processes (LSPs), introduced by Silverman in [6]. LSPs are characterized by a covariance function that is the modulation in time of an ordinary stationary covariance function. The optimal kernel for estimation of the Wigner-Ville spectrum for a certain class of LSPs is obtained in [7]. Based on this result, we derive the MSE optimal time-frequency kernel for our model covariance and we use it to compute an optimal multitaper spectrogram, [8]–[10].

The kernels are parameter dependent and the lack of reliable inference methods has relegated LSPs and their optimal time-frequency representation to a theoretical interest. The inference method, [11], [12], is based on the separation of the two factors of the product defining an LSP covariance function, in order to take advantage of their individual structures. Thanks to the introduced inference method we obtain a complete procedure to achieve a MSE optimal time-frequency kernel from measured data.

Time-frequency features are extracted from the spectral estimates and used to feed a neural network classifier. Such classifiers are suitable for classification based on time-frequency representations of audio signals, such as speech [13] and EEG [14]. In [2] an adaptive and localized time-frequency representation of EEG signals has resulted in improvements in classification accuracy. However, usually the conventional spectrogram is the the time-frequency representation input.

The purpose of this paper is to show how the optimal MSE time-frequency kernel offers a significant improvement in practical applications, leading to a higher classification accuracy thanks to the greater quality of the time-frequency features extracted with the proposed approach. The parameters estimation on a suitable LSP model for EEG signals allows the extraction of improved features for classification.

The paper is structured as follows. In section II, we present the mathematical model of LSPs, the expression for the MSE optimal time-frequency kernel, the specific model used in the simulation study and in the EEG data application, and the classification approach considered. In section III, performance of the derived spectral estimator is evaluated through a simulation study. In section IV we present the results of the classification of the EEG signals, collected within a study on human memory encoding. The paper concludes with some comments and directions for further research in section V.

II. Methods

A. Locally Stationary Processes

Let $X(t)$, $t \in [T_0, T_f] \subseteq \mathbb{R}$, be a zero mean stochastic process. We say that $X(t)$ is a Locally Stationary Process (LSP) in the wide sense if its covariance $C(s,t) = \mathbb{E}[X(s)X(t)^*]$ can be written as

$$C(s,t) = q\left(\frac{s+t}{2}\right) \cdot r\left(s-t\right) = q\left(\eta\right) \cdot r\left(\tau\right) \qquad (1)$$

ISBN 978-90-827970-1-5 © EURASIP 2018

with $s,t \in [T_0,T_f] \subseteq \mathbb{R}$, $q(\eta)$ a non-negative function and $r(\tau)$ a normalized ($r(0)=1$) stationary covariance function. When $q(\eta)$ is a constant, (1) reduces to a stationary covariance and this definition therefore includes stationary processes as a special case. The wide range of possibilities for the choice of the functions offers an advantageous flexibility to model time-varying data. Clearly not every choice is suitable, since we recall that a function $C(s,t)$ is a covariance if and only if it is positive semi-definite.

B. Mean Square Error optimal kernel

The Wigner-Ville spectrum of an LSP is defined as

$$W(t,\omega) = \int_{-\infty}^{\infty} \mathbb{E}\left[X\left(t+\frac{\tau}{2}\right)X\left(t-\frac{\tau}{2}\right)\right] e^{-i\tau\omega} d\tau \quad (2)$$
$$= q(t)\cdot \mathcal{F}r(\omega),$$

where $\mathcal{F}f$ denotes the Fourier transform of the function f, [6], [7]. The corresponding ambiguity spectrum is defined as

$$A(\theta,\tau) = \int_{-\infty}^{\infty} \mathbb{E}\left[X\left(t+\frac{\tau}{2}\right)X\left(t-\frac{\tau}{2}\right)\right] e^{-it\theta} dt \quad (3)$$
$$= \mathcal{F}q(\theta)\cdot r(\tau),$$

and any time-frequency representation member of the Cohen's class can be expressed as

$$W_C(t,\omega) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} A(\theta,\tau)\Phi(\theta,\tau)e^{-i(\tau\omega-t\theta)} d\tau d\theta \quad (4)$$

where Φ is an ambiguity kernel, [15]. The general expression for the optimal ambiguity kernel in the mean square error (MSE) sense was derived in [16] as

$$\Phi_0(\theta,\tau) = \frac{|\mathcal{F}q(\theta)|^2|r(\tau)|^2}{|\mathcal{F}q(\theta)|^2|r(\tau)|^2 + (\mathcal{F}|r|^2(\theta))(\mathcal{F}^{-1}|\mathcal{F}q|^2(\tau))}, \quad (5)$$

and in [7] the authors derived the MSE optimal kernel for LSP where the factors of the covariance are Gaussian functions.

Efficient implementation and estimation are based on multitapers, [7], [8], [10], i.e. a weighted sum of windowed spectrograms, as

$$W_C(t,\omega) = \mathbb{E}\left[\sum_{k=1}^{K}\alpha_k \left|\int_{-\infty}^{\infty} X(s)h_k^*(t-s)e^{-i\omega s} ds\right|^2\right], \quad (6)$$

with weights α_k, and windows $h_k(t)$, $k = 1\ldots K$. The weights and windows are derived from the solution of the eigenvalue problem

$$\int_{-\infty}^{\infty} \Psi^{rot}(s,t)h(s)ds = \alpha h(t), \quad (7)$$

where the rotated time-lag kernel is Hermitian and defined as

$$\Psi^{rot}(s,t) = \Psi\left(\frac{s+t}{2}, s-t\right), \quad (8)$$

with

$$\Psi(t,\tau) = \int_{-\infty}^{\infty} \Phi(\theta,\tau)e^{it\theta} d\theta. \quad (9)$$

With a few α_k that differ significantly from zero, the multitaper spectrogram solution is an efficient solution from implementation aspects.

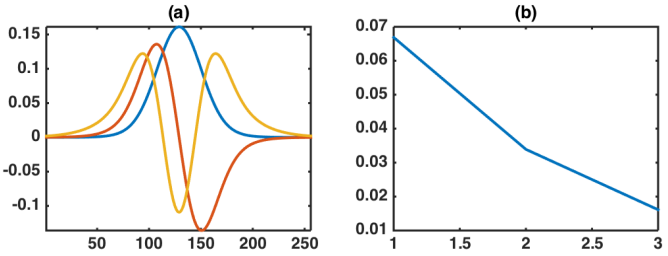

Fig. 1. Example of MSE optimal eigenvectors (a) and eigenvalues (b), corresponding to multitapers and weights of the MSE optimal time-frequency estimator for the model (10) and parameters (L, a_q, b_q, c_q, c_r) =(150, 800, 0.25, 250, 3000).

C. Stochastic model for the simulation study and the EEG data application

In this study we choose the functions $q(\eta)$ and $r(\tau)$ as

$$q(\eta) = L + a_q\cdot\exp\left(-c_q\left(\eta-b_q\right)^2/2\right) \text{ with } \eta = \frac{t+s}{2}$$
$$r(\tau) = \exp\left(-\frac{c_r}{8}\cdot\tau^2\right) \text{ with } \tau = t-s \quad (10)$$

with $b_q \in [T_0,T_f]$, T_0 and T_f initial and final times and $c_r > c_q > 0$. The latter assumption is necessary to assure that the resulting covariance is positive semi-definite. This choice of functions, introduced in [12], is motivated by EEG data application, IV. The parameter L is modeling additive stationary noise on the actual LSP and is useful to model data with heavy disturbances.

Thanks to (5) we are able to compute the parameter dependent optimal kernel $\Phi_0(\theta,\tau)$ for the introduced model (10), as

$$\Phi_0(\theta,\tau) = \frac{|A(\theta,\tau)|^2}{|A(\theta,\tau)|^2 + B(\theta,\tau)} \quad (11)$$

with

$$|A(\theta,\tau)|^2 = |\mathcal{F}q(\theta)|^2|r(\tau)|^2$$
$$= \left(L^2\delta_0(\theta) + \frac{2\pi a_q^2}{c_q}e^{-\frac{\theta^2}{c_q}} + 2a_qL\sqrt{\frac{2\pi}{c_q}}\delta_0(\theta)e^{-\frac{\theta^2}{2c_q}}\right)e^{-\frac{c_r\tau^2}{4}} \quad (12)$$

and

$$B(\theta,\tau) = (\mathcal{F}|r|^2(\theta))(\mathcal{F}^{-1}|\mathcal{F}q|^2(\tau))$$
$$= \left(2\sqrt{\frac{\pi}{c_r}}e^{-\frac{\theta^2}{c_r}}\right)\left(\frac{L^2}{2\pi} + a_q^2\sqrt{\frac{\pi}{c_q}}e^{-\frac{c_q\tau^2}{4}} + a_qL\sqrt{\frac{2\pi}{c_q}}\right) \quad (13)$$

where δ_0 denotes the Dirac delta function.

D. Pattern recognition neural networks

Pattern recognition networks are feed-forward networks that can be trained to classify inputs according to target classes. The input and target vectors are usually divided into three sets: training, validation and testing. After the training of the neural network, the validation phase is necessary to ensure

ISBN 978-90-827970-1-5 © EURASIP 2018

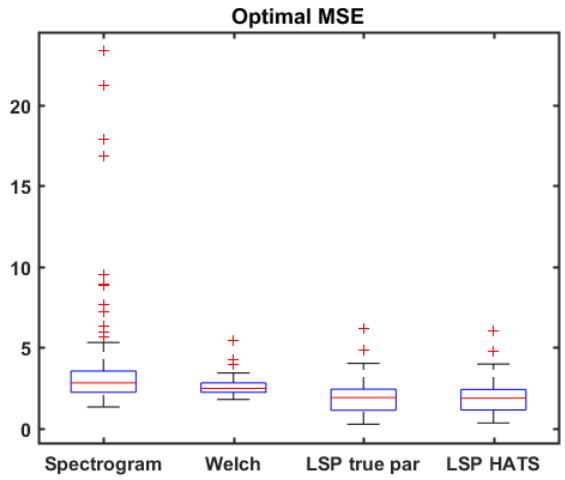

Fig. 2. Boxplots of the MSE on 100 simulation for the spectral estimators considered, all with parameters optimized. Average MSE are 3.8, 2.6, 1.9 and 1.9, for HANN ($N_w = 32$), WOSA ($K = 10$), the MSE optimal time-frequency estimator derived from LSP true and LSP HATS parameters respectively.

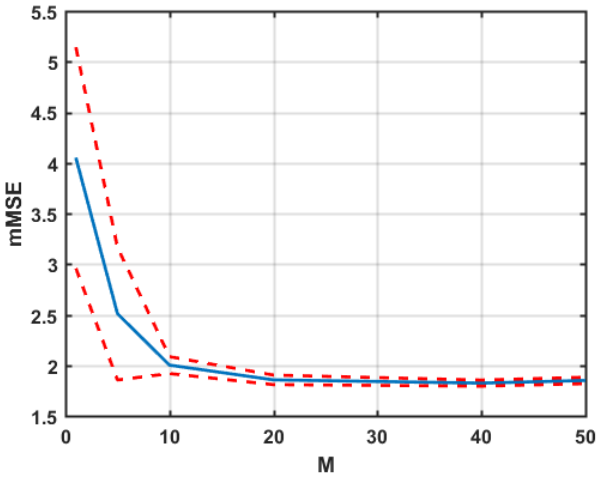

Fig. 3. Mean square error for the MSE optimal time-frequency estimator as function of the number of realizations used to produce the model parameter estimates with HATS. Red lines are 95 % confidence intervals.

that the network is generalizing and to stop training before over-fitting, and the testing phase consists of a completely independent test of the network. As in standard networks used for pattern recognition, [17], in this study we consider a two-layer feed-forward network, with a sigmoid transfer function in the hidden layer, 20 hidden neurons, and a softmax transfer function in the output layer.

III. Evaluation in simulation study

We present a simulation study to evaluate the method performance in terms of MSE of the derived optimal spectral estimator. We consider 60 realizations of a LSP with covariance function (10), sampled in 256 equidistant points during the time interval $[T_0,T_f]$=[0,0.5] seconds. The vector of true parameters used to simulate the data is (L, a_q, b_q, c_q, c_r)=(100, 600, 0.2, 1000, 10000). The inference method, HAnkel-Toeplitz Separation (HATS), [11], [12], is used to estimate the parameters $\lambda = (L, a_q, b_q, c_q)$ and $\rho = (c_r)$.

Based on the parameter estimates, the MSE optimal kernel and corresponding multitapers are calculated as described in section II-B and II-C. An example of resulting multitapers and weights are presented in Figure 1. Two other classical estimators are considered for comparison: the single Hanning window spectrogram (HANN) and the Welch method, with 50 % overlapping Hanning windows (WOSA), [18]–[21]. For a fair comparison of the performance of the different estimators, these two methods are optimized to give the smallest possible total MSE. For HANN, the window length $N_w \in \{16, 32, 64, 128, 256\}$ is optimized, while for WOSA the optimized parameter is the number of windows used $K \in \{1, 2, ..., 16\}$, where the total length of all included windows is 256.

The expected value of the MSE (mMSE) is computed as the average of 100 independent realizations. In Figure 2 we present boxplots of the MSE achieved with the different methods in the 100 simulations. The optimal mMSE for HANN and WOSA are 3.8 and 2.6 respectively, obtained respectively with $N_w = 32$ for HANN and $K = 10$ for WOSA. The mMSE value for the MSE optimal estimator with the true parameters or with parameters estimated with HATS is 1.9. Notice that not only the spectral estimate obtained using MSE optimal kernels achieves the best mMSE as expected, but using the true parameters or those estimated with HATS leads to the same result.

To test how the number of realizations used to estimate the model parameters with HATS affects the results in the time-frequency domain, we study the variation of the MSE of the spectral estimator based on parameter estimates obtained using a different number of realizations $N \in \{1, 5, 10, 25, 50\}$. In Figure 3 we present the resulting mMSE, computed as average on 100 independent simulations, as function of the number of realizations used, with corresponding 95 % confidence interval.

IV. Classification of EEG signals

The data considered has been collected within a study on human memory retrieval, conducted at the department of Psychology of Lund University, Sweden, during the spring of 2015. The EEG signals have been measured from one subject participating in the experiment, during 180 trials of a memory recognition task, in which the subject had to associate a presented word with a target picture. Each picture presented belongs to one of three categories: ”Faces”, ”Landmarks”, ”Objects”. For each category, 60 different trials were performed and we restrict to the coding phase. The measurements were recorded from channel O1 (International 10-20 system), as primary visual areas can be found below the occipital

ISBN 978-90-827970-1-5 © EURASIP 2018

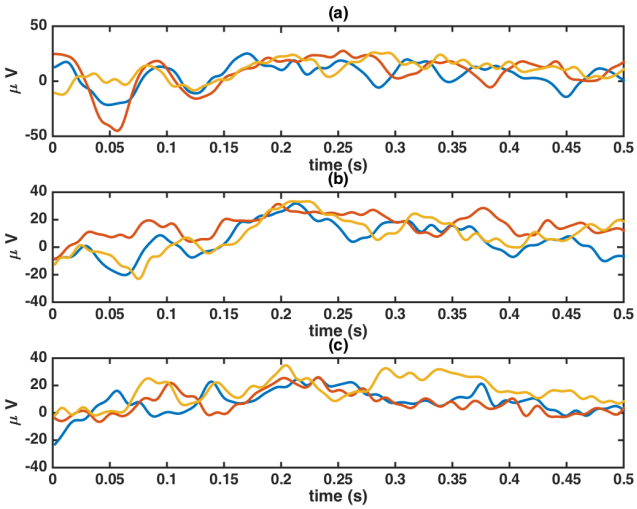

Fig. 4. Three random EEG signals, after 70 Hz low-pass filtering, corresponding to three different trials of a memory task, from each category: (a) 'Faces'; (b) 'Objects'; (c) 'Landmarks'.

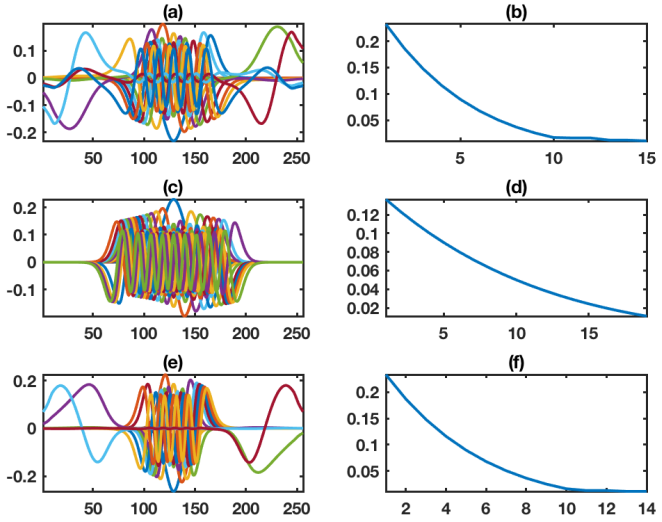

Fig. 5. Example of MSE optimal multitapers and weights of the MSE optimal time-frequency estimator for a random set of realizations from the three categories: (a) multitapers and (b) weights for 'Faces'; (c) multitapers and (d) weights for 'Objects'; (e) multitapers and (f) weights for 'Landmarks'.

lobes, and downsampled to frequency 512 Hz. Each time series considered then has 256 equidistant samples during the time interval [0, 0.5] seconds (Figure 4).

For each class, 40 out of the total 60 realizations are used to infer on the parameters L, a_q, b_q, c_q, c_r of the LSP model covariance (10). The estimated parameters are used to compute the optimal kernels and multitapers for each class. The spectral estimates obtained with the MSE optimal estimator, HANN ($N_w = 32$) and WOSA ($K = 10$), are used to extract time-frequency features, where each feature is the spectral power at each time-frequency point in the time interval [0, 0.5] seconds and frequency up to 40 Hz.

The remaining 20 realizations are used for independent testing in the network. The model parameters are re-estimated from these independent data-sets and the consequent spectral estimates computed. Classification accuracy is based on these data-sets. Since the accuracy of the network varies with the initial parameters, e.g. the subsets of the data chosen for training and validation, each network is retrained 10 different times. The random partition in 40 realizations for training and 20 for testing is repeated 10 times as well, and the test is repeated with the different random sets of testing realizations. In figure 5 we present an example of MSE optimal multitapers and weights of the MSE optimal time-frequency estimator for a random set of realizations from the three categories. In tables I, II, III we report the confusion matrices resulting from the classification of the EEG signals, using the time-frequency features from the three spectral estimators compared.

Clearly, accuracy of the classification is a feedback for how good are the features used. As expected, the use of better time-frequency features results in improvement in classification accuracy, with a total classification accuracy of 80 % for the proposed MSE optimal time-frequency estimator, compared to 57.4 % and 50.3 % of HANN and WOSA respectively.

TABLE I
CONFUSION MATRIX USING FEATURES OBTAINED WITH THE SINGLE WINDOWED SPECTROGRAM (HANN). TOTAL CLASSIFICATION ACCURACY 57.4 %.

		Actual class		
		Faces	Objects	Landmarks
Predicted class	Faces	1408	420	370
	Objects	241	957	550
	Landmarks	351	623	1080

V. CONCLUSION

In this paper we have illustrated an approach leading to improvement in classification accuracy of EEG signals measured during a memory encoding task. This result has been achieved by estimating an MSE optimal time-frequency estimator and extract features to feed a pattern recognition neural network classifier. The proposed approach leads to a total classification accuracy of 80 %, compared to 57.4 % and 50.3 % obtained with the classical single Hanning windowed spectrogram and the Welch spectrogram respectively.

Future extensions of this research will consider a multi-dimensional model to deal with correlated signals and an extended classification study, including several subjects.

TABLE II
CONFUSION MATRIX USING FEATURES OBTAINED WITH WOSA. TOTAL CLASSIFICATION ACCURACY 50.3 %.

		Actual class		
		Faces	Objects	Landmarks
Predicted class	Faces	1054	581	527
	Objects	485	961	470
	Landmarks	461	458	1003

ISBN 978-90-827970-1-5 © EURASIP 2018

TABLE III
CONFUSION MATRIX USING FEATURES OBTAINED WITH MSE OPTIMAL ESTIMATOR. TOTAL CLASSIFICATION ACCURACY 80 %.

		Actual class		
		Faces	Objects	Landmarks
Predicted class	Faces	1092	37	149
	Objects	232	1912	53
	Landmarks	676	51	1798

ACKNOWLEDGMENT

The authors would like to thank the eSSENCE Academy for funding and the department of Psychology, Lund University, for data collection.

REFERENCES

[1] M. X. Cohen and R. Gulbinaite, "Review: Five methodological challenges in cognitive electrophysiology," *NeuroImage*, vol. 85, no. Part 2, pp. 702 – 710, 2014.

[2] K. Samiee, P. Kovacs, and M. Gabbouj, "Epileptic seizure classification of EEG time-series using rational discrete short-time Fourier transform," *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 2, pp. 541 – 552, 2015.

[3] J. Meng, L. M. Merino, N. B. Shamlo, S. Makeig, and K. Robbins, "Characterization and robust classification of eeg signal from image rsvp events with independent time-frequency features," *PLoS ONE*, vol. 7, no. 9, 2012, doi:10.1371/journal.pone.0044464.

[4] Fu Kai, Qu Jianfeng, Chai Yi, and Dong Yong, "Classification of seizure based on the time-frequency image of eeg signals using hht and svm.," *Biomedical Signal Processing and Control*, vol. 13, pp. 15 – 22, 2014.

[5] R. Hellerstedt and M. Johansson, "Competitive semantic memory retrieval: Temporal dynamics revealed by event-related potentials," *PLOS ONE*, vol. 11, no. 2, pp. e0150091, 2016.

[6] R. Silverman, "Locally stationary random processes," *IRE Transactions on Information Theory*, vol. 3, no. 3, pp. 182, 1957.

[7] P. Wahlberg and M. Hansson, "Kernels and multiple windows for estimation of the Wigner-Ville spectrum of Gaussian locally stationary processes," *IEEE Transactions on Signal Processing*, vol. 55, no. 1, pp. 73 – 84, 2007.

[8] G. S. Cunningham and W. J. Williams, "Kernel decomposition of time-frequency distributions," *IEEE Transactions on Signal Processing*, vol. 42, pp. 1425–1442, June 1994.

[9] F. Cakrak and P. J. Loughlin, "Multiple window time-varying spectral analysis," *IEEE Transactions on Signal Processing*, vol. 49, no. 2, pp. 448–453, 2001.

[10] M. Hansson-Sandsten, "Optimal multitaper Wigner spectrum estimation of a class of locally stationary processes using Hermite functions," *EURASIP Journal on Advances in Signal Processing*, p. 980805, 2011.

[11] R. Anderson and M. Sandsten, "Inference for time-varying signals using locally stationary processes," *submitted*, 2017.

[12] R. Anderson and M. Sandsten, "Stochastic modelling and optimal spectral estimation of EEG signals," in *EMBEC & NBC 2017*. 2018, pp. 908–911, Springer Singapore.

[13] O. Abdel-Hamid et. al., "Convolutional neural networks for speech recognition.," *IEEE-ACM Trans. on audio and speech and language processing*, vol. 22, no. 10, pp. 1533 – 1545, 2014.

[14] Y. R. Tabar and U. Halici, "A novel deep learning approach for classification of EEG motor imagery signals.," *Journal of Neural Engineering*, vol. 14, no. 1, 2017.

[15] L.Cohen, *Time-Frequency Analysis*, Prentice-Hall, 1995.

[16] A.M. Sayeed and D.L. Jones, "Optimal kernels for nonstationary spectral estimation," *IEEE Transactions on Signal Processing*, vol. 43, no. 2, pp. 478 – 491, 1995.

[17] The MathWorks Inc., "Neural network toolbox, version r2017b," Natick, Massachusetts, United States.

[18] P. D. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Transactions on Audio Electroacoustics*, vol. AU-15, no. 2, pp. 70–73, June 1967.

[19] Sarah N. Carvalho, Thiago B.S. Costa, Luisa F.S. Uribe, Diogo C. Soriano, Glauco F.G. Yared, Luis C. Coradine, and Romis Attux, "Comparative analysis of strategies for feature extraction and classification in SSVEP BCIs," *Biomedical Signal Processing and Control*, vol. 21, pp. 34 – 42, 2015.

[20] Maria Hansson-Sandsten, "Evaluation of the optimal lengths and number of multiple windows for spectrogram estimation of SSVEP," *Medical Engineering and Physics*, vol. 32, pp. 372 – 383, 2010.

[21] Thilakavathi Bose, Shenbaga Devi Sivakumar, and Bhanu Kesavamurthy, "Identification of schizophrenia using EEG alpha band power during hyperventilation and post-hyperventilation," *Journal of Medical and Biological Engineering*, vol. 36, no. 6, pp. 901, 2016.