SIViP (2016) 10:389–395
DOI 10.1007/s11760-015-0752-x

ORIGINAL PAPER

Neural network-based adaptive noise cancellation for enhancement of speech auditory brainstem responses

Shiva Gholami-Boroujeny · Anwar Fallatah · Brian P. Heffernan · Hilmi R. Dajani

Received: 11 December 2013 / Revised: 1 February 2015 / Accepted: 2 February 2015 / Published online: 17 February 2015

Abstract The measurement of the speech-evoked auditory brainstem response (speech ABR) is a promising technique for evaluating auditory function. However, the speech ABR is severely contaminated by background noise related to other brain electrical activity. The most commonly used method to enhance the signal-to-noise ratio (SNR) of the response is coherent averaging, while recently adaptive filtering has also been reported. All of the applied methods are based on linear operations, but since the assumption of linearity may not be valid for neural activity, linear methods may not be adequate. In this paper, we present a new nonlinear adaptive noise cancellation (ANC) based on a multilayer perceptron neural network to enhance the speech ABR and compare its performance with a linear ANC algorithm based on least mean squares adaptive filtering. The effectiveness of the methods is tested using speech ABR data and is based on two different types of SNR measures, the local SNR at the fundamental frequency of the response and the overall SNR. The results show that the nonlinear neural network-based ANC can reduce the required recording time and performs better than the linear ANC especially when the SNR of the recorded speech ABR is low.

Keywords Speech auditory brainstem responses · Adaptive filtering · Multilayer perceptron neural network

S. Gholami-Boroujeny (✉) · A. Fallatah · B. P. Heffernan · H. R. Dajani
School of Electrical Engineering and Computer Science,
University of Ottawa, 161 Louis Pasteur, Ottawa,
ON K1N 6N5, Canada
e-mail: sgholami@uottawa.ca

H. R. Dajani
e-mail: hdajani@site.uottawa.ca

1 Introduction

The auditory brainstem response (ABR) is a reliable signal that can be used to evaluate the function of the auditory system. In current clinical practice, the ABR is recorded when the auditory system is stimulated with an artificial sound such as a click, tone burst, or an amplitude-modulated tone [1]. In recent studies [2–4], the ABR generated by a speech stimulus has been investigated because speech is of primary importance in human acoustic communication. Auditory processing of speech can therefore be probed with speech auditory brainstem responses (speech ABR). Speech ABR is a potentially useful diagnostic tool and has been suggested as a marker of central auditory processing disorders in some children with learning disabilities [2–4]. Speech ABR may also be applied to the objective selection and fitting of modern hearing aids, whose performance cannot be easily tested with simple non-speech sounds [5].

Since the ABR is usually recorded using surface scalp electrodes, it is strongly corrupted by background noise and so usually the signal-to-noise ratio (SNR) is very low. In clinical practice, coherent averaging is the most commonly used method which exploits the deterministic nature of the signal to enhance the ABR while suppressing uncorrelated EEG signals, extraneous noise, and artifacts. However, in speech ABR, coherently averaging the responses over multiple presentations of the relatively long duration speech stimulus requires an exceedingly long recording time that ranges from several minutes to tens of minutes with a single speech sample [6–8]. This long recording time has limited the application of speech ABR for clinical use.

Researchers have tried to develop other algorithms to make ABR extraction more effective. Different methods including nonlinear estimation [9], adaptive sinusoidal estimator [10], independent component analysis, and Wiener filter methods [11] have been used with ABR. With speech ABR in particular, coherent averaging is the method that is currently used to improve response SNR. Our group has also investigated the use of adaptive filtering methods [12], but with signals having a higher SNR than some of the signals that are investigated in this work. More importantly, both coherent averaging and the adaptive filtering methods that we tested are based on linear operations. Since the assumption of linearity may not be valid for nervous system activity which is responsible for the speech stimulus-evoked response generation, the linear methods may not be adequate [13–16]. Nonlinear approaches, on the other hand, have been used in various fields of bio-signal extraction. Neural networks are a class of nonlinear filters which can be trained adaptively. They have been successfully applied in many areas of bio-signal applications because of their flexible structure which gives them the ability to approximate nonlinear functions [17–25]. However, to our knowledge, their application in speech ABR enhancement has not been investigated in previous studies.

This paper investigates two different methods for speech ABR signal enhancement. The first one is a linear adaptive noise canceler (ANC) approach. To use this method, we need two input signals: a primary signal and a reference noise. Based on the assumption that a correlation exists between the reference noise and only the additive noise part of the primary path, this method gives the best estimation of the part of the primary signal which is correlated with reference noise [26, 27]. The second method is an ANC system based on neural networks. A three-layer perceptron neural network is used as an adaptive nonlinear controller, trained with the back-propagation algorithm [28,29]. The two methods are tested with speech ABR signals, and the results are illustrated. The results show that in case of recorded signals with lower SNR, the nonlinear method gives a better response. The rest of paper is organized as follows. Section 2 provides a summary of two methods used for extraction of speech ABR signal. The analysis results are given in Sect. 3, and finally Sect. 4 concludes the paper.

2 Methods

2.1 Linear adaptive filtering

The linear adaptive filtering approach is based on the ANC system first proposed by Widrow et al. in 1975 [30]. In an ANC system, two input signals $x(n)$ and $y_0(n)$, referred to as the primary signal and the noise reference signal, are applied to the adaptive filter as shown in Fig. 1. It is assumed that $y_0(n)$ is correlated with only a part of $x(n)$.

The noise reference signal $y_0(n)$ passes through the adaptive filter to generate the best estimate of that signal com-

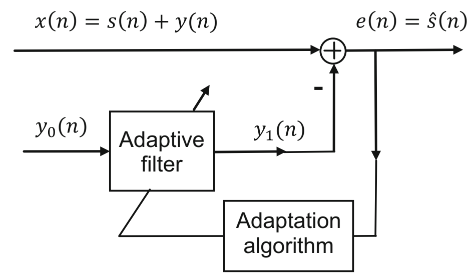

Fig. 1 Schematic diagram of adaptive noise canceler [30]

ponent which is correlated with $y_0(n)$. The error signal is produced by subtracting the best estimate from the primary input. The produced error signal is used as feedback to the adaptive filter for optimizing the tap weights, in the mean square sense. This process is continued to minimize the correlated part at the output. The output signal at this time is an estimate of the un-measurable information signal $s(n)$.

For the application of speech ABR signal enhancement, the least mean squares approach (LMS) is used as an adaptive algorithm to adjust the filter coefficients. The order of the filter was chosen to be $N = 36$, and the step size parameter is $\mu = 0.015$.

2.2 Multilayer perceptron neural network

Linear adaptive controllers have been successfully applied in many areas of noise canceling systems. However, the assumption of linearity may not be valid for nervous system activity. Hence, there is a need for an adaptive nonlinear filtering approach. Soft computing-based approaches are used as a predominant technology in this area [17–25]. In this section, we investigate a nonlinear ANC system based on a multilayer perceptron neural network (NN) [29] for speech ABR enhancement. The network structure is chosen based on three layers, including a hidden layer with a nonlinear activation function as shown in Fig. 2. This structure gives us a nonlinear filter as the adaptive nonlinear controller that replaces the linear adaptive filter shown in Fig. 1.

The number of neurons in the input and the hidden layer is determined experimentally, and the last layer contains only one neuron which is output of the filter. The hyperbolic tangent function is used for the activation of the hidden layer neurons, and the identity function is used for the output layer. The output of the neuron can be expressed by the following equation, Eq. (1):

$$y_k(n) = f\left[\sum_{j=0}^{H} w_{kj}.h\left[\sum_{i=0}^{M} w_{ji}\, y_i(n-i)\right]\right] = f(v_k(n)). \tag{1}$$

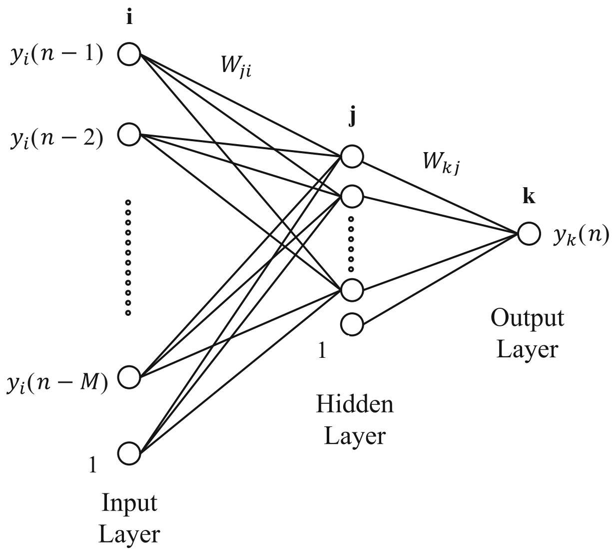

Fig. 2 Three-layer perceptron used for nonlinear filtering [29]

where $y_i(n)$ is the input vector, $y_i(n) = \left[y(n-1), y(n-2), \ldots, y(n-M), 1\right]$. The weights of the connection between the ith neuron in the input layer and the jth neuron in the hidden layer are denoted with w_{ji}. The weights of connection between the jth neuron in the hidden layer and the output one are shown with w_{kj}. The number of input layer neurons and hidden layer neurons is denoted by M and H, respectively. The activation functions of the hidden and output layers are denoted by h and f, respectively.

The back-propagation algorithm [28] is used to adapt the connection weights. In the forward pass, the mean square error for the output layer is determined:

$$e_k(n) = E\left[(y_d(n) - y_k(n))^2\right], \tag{2}$$

in which, $y_d(n)$ and $y_k(n)$ are the desired output and the real output, respectively.

The back-propagation algorithm applies a correction Δw to the synaptic weights w. This adaptation rule at iteration nth is given by the following equation, Eq. (3):

$$w(n) = w(n) - \eta\frac{\partial e_k(n)}{\partial w(n)} \tag{3}$$

where η is the learning-rate parameter of the back-propagation algorithm. The combination of weights that minimizes the error function is considered to be a solution of the learning problem. In our system, for the connection weights between the jth neuron in the hidden layer and the output neuron, the gradient is defined by the following relation, Eq. (4):

$$\frac{\partial e_k(n)}{\partial w_{kj}(n)} = -\delta_k(n).z_j(n). \tag{4}$$

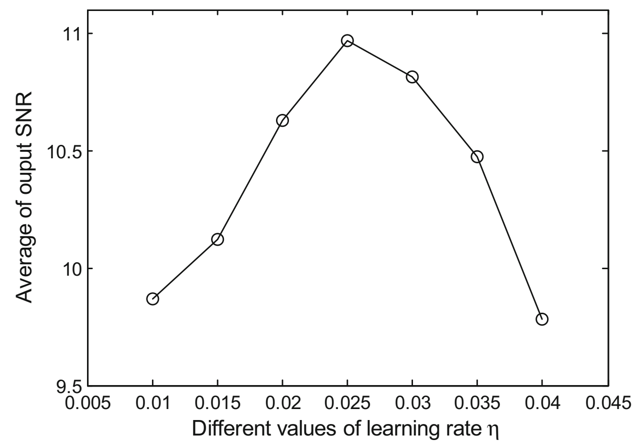

Fig. 3 Average of the SNR values of the denoised signal versus learning rate η

where

$$\delta_k(n) = e_k(n).f'(v_k(n)) = y_d(n) - y_k(n). \tag{5}$$

This is because the activation function of the output layer, f, is the identity function, and,

$$z_j(n) = h\left[\sum_{i=0}^{M} w_{ji}\, y_i(n-i)\right] = h(v_j(n)). \tag{6}$$

in which, h is the hyperbolic tangent function. The adaptation equation for the connection weights between the jth neuron in the hidden layer and the output neuron at iteration nth becomes:

$$w_{kj}(n) = w_{kj}(n) + \eta\delta_k(n).z_j(n). \tag{7}$$

For the connection weights between the ith neuron in the input layer and the jth neuron in the hidden layer, the gradient is defined by the following relation, Eq. (8),

$$\frac{\partial e_k(n)}{\partial w_{ji}(n)} = -\delta_j(n).y_i(n). \tag{8}$$

where

$$\delta_j(n) = \delta_k.w_{kj}.h'\left(v_j(n)\right) = \delta_k.w_{kj}.(1 - z_j(n))(1 + z_j(n)). \tag{9}$$

The connection weights adaptation equation in this case becomes:

$$w_{ji}(n) = w_{ji}(n) + \eta\delta_j(n).y_i(n). \tag{10}$$

The back-propagation algorithm can be applied in many ways. In our case, we used the sequential mode of back-propagation learning in which weight updating is performed after the presentation of each training example [28]. The NN structure consisted of a three-layer perceptron with 12 neurons in the input layer, and six neurons in the hidden layer. It gave us the best average denoising results when the learning rate of the back-propagation algorithm was chosen to be $\eta = 25 \times 10^{-3}$ as shown in Fig. 3 for the average SNR of the output signal.

3 Results

Speech ABRs were measured using three surface scalp electrodes, with the recording electrode placed at the vertex region of the scalp, the reference electrode placed on the right earlobe, and the ground electrode placed on the left earlobe. Five subjects (25–45 years old, two female) participated in this study. All had normal hearing in the right ear with a threshold at or below 15 dB HL at 500, 1000, 2000, and 4000 Hz. A formant synthesizer was used to generate a 300-ms male vowel /a/ speech stimulus. This stimulus has a fundamental frequency (F_0) of 100-Hz, and a first formant (F_1) at 700-Hz. Each speech ABR was recorded synchronously with the stimulus, which was presented to the right ear. The duration of the response recording was 319.8-ms with 1024 points taken, giving a sampling frequency of 3202 Hz. The obtained speech ABR signal passed through an in-line band-

Table 1 Local SNR (LSNR) and overall SNR (OSNR) in dB in the unprocessed recorded signal, after linear ANC, and after NN-based ANC obtained on single blocks, by averaging over three blocks, by averaging over five blocks, and by averaging over 15 blocks. The results shown are averaged over all five subjects

	On one block		By averaging over 3 blocks		By averaging over 5 blocks		By averaging over 15 blocks	
	LSNR	OSNR	LSNR	OSNR	LSNR	OSNR	LSNR	OSNR
Recorded Signal	6.2	2.2	9.9	5.7	11.6	7.2	15.4	11.0
After linear ANC	7.9	2.9	10.9	6.3	12.5	7.7	15.5	11.0
After NN-based ANC	10.6	4.3	13.0	7.5	14.3	9.3	16.6	12.2

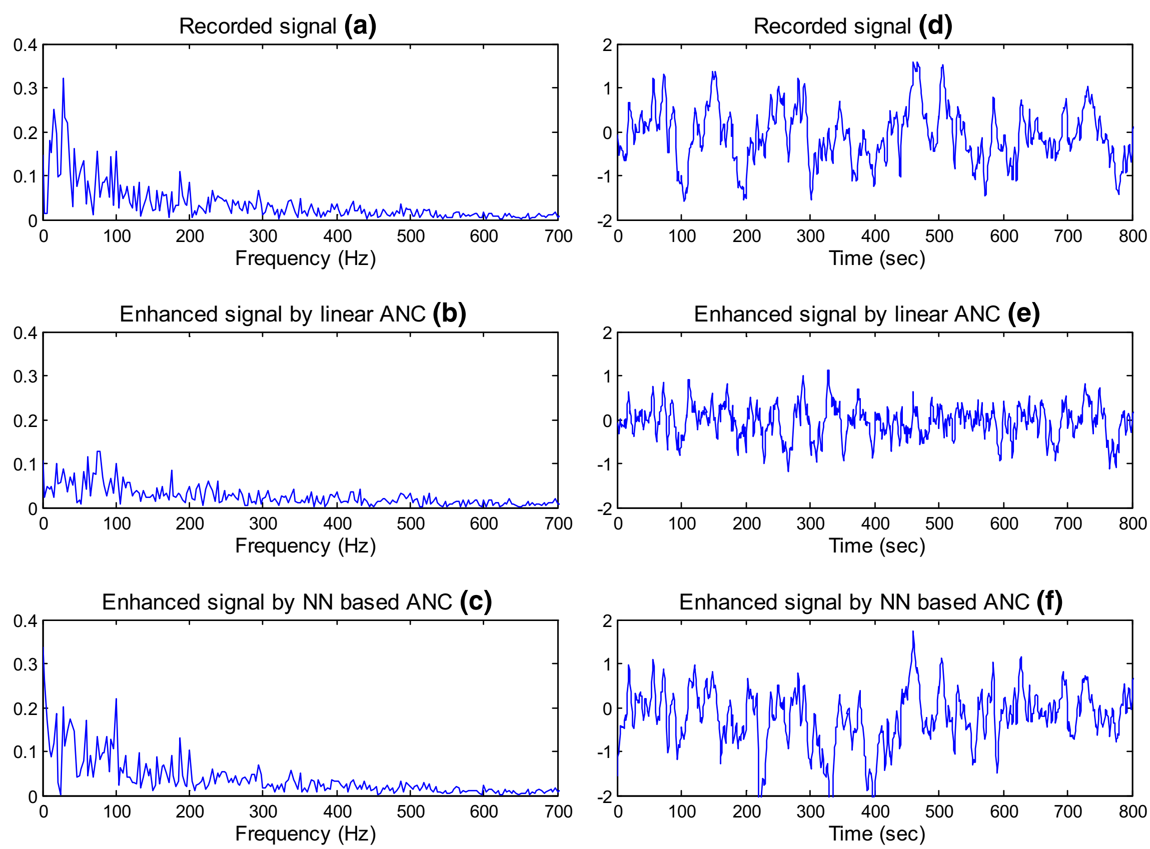

Fig. 4 Results of speech ABR enhancement in one subject in the frequency (**a**–**c**) and time domains (**d**–**f**) by linear ANC and NN-based ANC on one block of the response

pass filter (30–1000 Hz), which suppresses low frequency noise and emphasizes the targeted response components at F_0 and F_1 (100 and 700 Hz). Fifteen blocks of speech ABR signals were gathered, with each block corresponding to the coherent average of responses to 20 repetitions of the stimulus. A sample of electroencephalogram (EEG) noise was also recorded right after recording the response blocks and is therefore correlated with the noise in the response blocks. This EEG noise is used as the reference noise signal in the ANC system.

The performance of the ANC system in enhancing the speech ABR was investigated based on the two methods, namely the LMS-based linear adaptive filtering and the neural network (NN)-based ANC. The assessment of the performance of the methods was done using two different types of SNR measures, the local and the overall SNR. The local SNR (LNSR) was calculated based on the signal power at the fundamental frequency of F_0 (100 Hz) and the noise power over the interval 100–200 Hz and not including the components at 100 and 200 Hz. In the overall SNR (OSNR), the signal power at the fundamental frequency and its harmonics including 100, 200, and up to 700 Hz were combined by adding the square of the root-mean-squared amplitudes at these frequencies, while the remaining signal components up to 700 Hz corresponded to the noise power.

Table 1 shows the performance measures before and after applying the LMS- and NN-based noise cancellation systems. The results are obtained for the recorded speech ABR signals of one block based on 20 response repetitions, by averaging over three blocks and so based on 60 response repetitions, by averaging over five blocks and so based on 100 response repetitions, and one average of all fifteen blocks and so based on 300 response repetitions. The results shown are averaged over all five subjects.

As expected, both the local and overall SNR are higher when a greater number of blocks are averaged. However, the speech ABR enhancement obtained using the nonlinear ANC system based on the multilayer perceptron performed better than the linear method, especially when the recorded signal based on one block is used as the input to the ANC system. In this case, the improvement in LSNR with linear ANC is 1.7 dB, while it is 4.4 dB with NN-based ANC. In contrast, when the ANC is applied on the average of 15 blocks, the improvement in LSNR with linear ANC is 0.1 dB, while it is 1.2 dB with the NN-based ANC. Moreover, the LSNR achieved with the NN-based ANC on one block

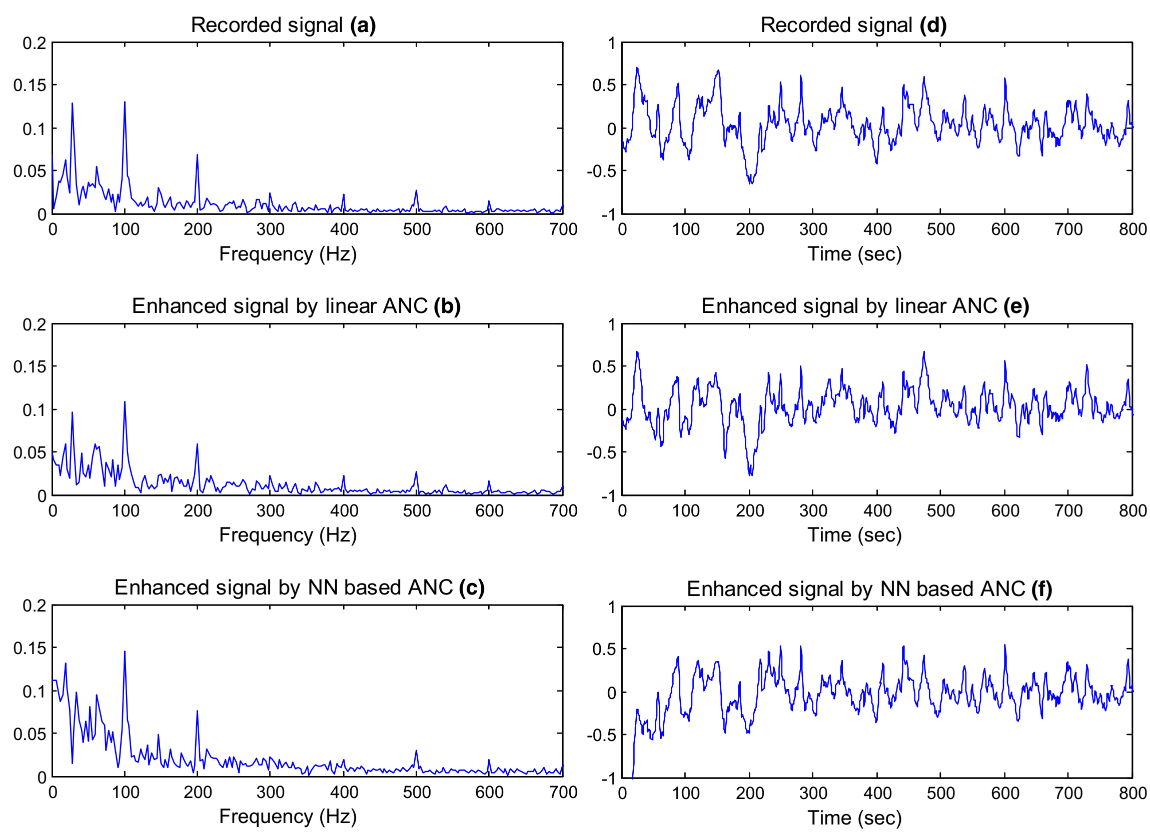

Fig. 5 Results of speech ABR enhancement in one subject in the frequency (**a–c**) and time domains (**d–f**) by linear ANC and NN-based ANC on the average of 15 blocks

(10.6 dB) is slightly higher than that achieved by coherently averaging three blocks (9.9 dB). This implies that by using the NN-based ANC, a significantly shorter recording time would be needed to achieve a similar SNR. The ABR peak at the fundamental frequency of 100 Hz, which is a distinctive feature for clinical diagnosis, can be clearly discerned after processing the recorded signal. Figure 4 shows an example of speech ABR enhancement with the linear ANC and NN-based ANC for the input recorded signal based on one block, while Fig. 5 shows an example of enhancement on the average of 15 blocks. The results in both frequency and time domains are shown.

4 Conclusion

Recordings of individual speech auditory brainstem responses (speech ABRs) usually have very low SNR, and coherent averaging is the standard method used to enhance these responses. This paper presented an application of adaptive noise cancellation (ANC) for speech ABR enhancement based on two methods, a linear method based on LMS adaptive filtering and a nonlinear one based on multilayer perceptron neural network. The tested signals were based on averages of 20 to 300 responses. Since the assumption of linearity may not be valid for nervous system activity, the linear methods may not work well for it. Therefore, a nonlinear neural network-based ANC system is proposed for the first time in this application. Results obtained with speech ABR data show the superiority of the nonlinear NN-based ANC compared to the linear ANC system, particularly for the signals with lower SNR. The signal enhancement that is achieved also implies that a significantly shorter recording time would be needed in order to achieve the same SNR, which is important for clinical applications of the technique of measuring brain responses evoked by speech stimuli. The processing time of NNs is usually higher than that of linear filters depending on the complexity of their design. However, real-time performance is not critical for our application because in speech ABR measurements, the number of stimulus repetitions is usually set beforehand and the processing is done after all the responses are recorded.

Acknowledgments This study was funded in part by the Natural Sciences and Engineering Research Council of Canada. Ethics approval for the experimental work was given by the University of Ottawa Research Ethics Board. There are no competing interests to declare.

References

1. Boston, J.R., Moller, A.R.: Brainstem auditory-evoked potentials. Crit. Rev. Biomed. Eng. **13**(2), 97–123 (1985)
2. Johnson, K.L., Nicol, T.G., Kraus, N.: Brain stem response to speech: a biological marker of auditory processing. Ear Hear. **26**(5), 424–434 (2005)
3. Vander-Werff, K.R., Burns, K.S.: Brainstem responses to speech in younger and older adults. Ear Hear. **32**(2), 168–180 (2011)
4. Russo, N., Nicol, T., Musacchia, G., Kraus, N.: Brainstem responses to speech syllables. Clin. Neurophysiol. **115**(9), 2021–2030 (2004)
5. Aiken, S.J., Picton, T.W.: Envelope and spectral frequency-following responses to vowel sounds. Hear. Res. **245**(1–2), 35–47 (2008)
6. Dajani, H.R., Purcell, D., Wong, W., Kunov, H., Picton, T.W.: Recording human evoked potentials that follow the pitch contour of a natural vowel. IEEE Trans. Biomed. Eng. **52**(9), 1614–1618 (2005)
7. Laroche, M.H., Dajani, H.R., Prévost, F., Marcoux, A.M.: Brainstem auditory responses to resolved and unresolved harmonics of a synthetic vowel in quiet and noise. Ear Hear. **34**(1), 63–74 (2012)
8. Prévost, F., Laroche, M., Marcoux, A.M., Dajani, H.R.: Objective measurement of physiological signal-to-noise gain in the brainstem response to a synthetic vowel. Clin. Neurophysiol. **124**(1), 52–60 (2013)
9. McNamara, D.M., Ziarani, A.K.: A new adaptive technique of estimation of steady state auditory evoked potentials. In: Conference on proceedings of IEEE engineering in medicine and biology society 26, pp. 4544–4547 (2004)
10. Cheah, L.A., Hou, M.: Real-time detection of auditory steady-state responses. In: Conference on proceedings of IEEE engineering in medicine and biology society, pp. 1382–1385 (2010)
11. Van Dun, B., Wouters, J., Moonen, M.: Improving auditory steady-state response detection using independent component analysis on multichannel EEG data. IEEE Trans. Biomed. Eng. **54**(7), 1220–1230 (2007)
12. Fallatah, A.: Enhancement of speech auditory brainstem responses using adaptive filters. M.A.Sc Thesis, University of Ottawa (2012)
13. Stam, C.J.: Nonlinear dynamical analysis of EEG and MEG: Review of an emerging field. Clin. Neurophysiol. **116**(10), 2266–2301 (2005)
14. Gholami-Boroujeny, S., Eshghi, M.: Active noise control using an adaptive bacterial foraging optimization algorithm. Signal Image Video Process., 1–10 (2012). doi:10.1007/s11760-012-0385-2
15. Mihandoost, S., Chehel Amirani, M.: EEG signal analysis using spectral correlation function and GARCH model. Signal Image Video Process., 1–12 (2014). doi:10.1007/s11760-013-0600-9
16. Kumar, Y., Dewal, M.L., Anand, R.S.: Epileptic seizures detection in EEG using DWT-based ApEn and artificial neural network. Signal Image Video Process., 1–12 (2012). doi:10.1007/s11760-012-0362-9
17. Zeng, Y., Zhang, J., Yin, H., Pan, Y.: Visual evoked potential estimation by adaptive noise cancellation with neural-network-based fuzzy inference system. J. Med. Eng. Technol. **31**(3), 185–190 (2007)
18. Hu, Y.H., Tompkins, W.J., Urrusti, J.L., Afonso, V.X.: Application of artificial neural networks for ECG signal detection and classification. J. Electrocardiol. **26**, 66–73 (1993)
19. Camps, G., Martinez, M., Soria, E.: Fetal ECG extraction using an FIR neural network. In: Conference on proceedings of IEEE computers in cardiology, pp. 249–252 (2001)
20. Selvan, S., Srinivasan, R.: Removal of ocular artifacts from EEG using an efficient neural network based adaptive filtering technique. IEEE Signal Process. Lett. **6**(12), 330–332 (1999)
21. Hamzaoui, E., Regragui, F., Himmi, M.M., Mghari, A.: Visual evoked potentials' non linear adaptive filtering based on three layers perceptron. J. Appl. Math. Sci. **4**(56), 2759–2772 (2010)
22. Nair, M.S., Shanker, V.: Predictive-based adaptive switching median filter for impulse noise removal using neural network-

based noise detector. Signal Image Video Process. **7**(6), 1041–1070 (2013)
23. Tian, J., Juhola, M., Grönfors, T.: Latency estimation of auditory brainstem response by neural networks. Artif. Intell. Med. **10**(2), 115–128 (1997)
24. Subasi, A., Erçelebi, E.: Classification of EEG signals using neural network and logistic regression. Comput. Methods Progr. Biomed. **78**(2), 87–99 (2005)
25. Majumdar, K.: Human scalp EEG processing: various soft computing approaches. Appl. Soft Comput. **11**(8), 4433–4447 (2011)
26. Haykin, S.: Adaptive Filter Theory. Prentice-Hall, Upper Saddle River (2002)
27. Widrow, B., Stearns, S.: Adaptive Signal Processing. Prentice-Hall, Upper Saddle River (1985)
28. Hecht-Nielsen, R.: Theory of the backpropagation neural network. In: Conference on proceedings of IEEE neural networks, IJCNN, pp. 593–605 (1989)
29. Haykin, S.: Neural Networks: A Comprehensive Foundation. Prentice-Hall, Upper Saddle River (1999)
30. Widrow, B., Glover, J.R., McCool, J.M.: Adaptive noise cancelling: principles and applications. Proc. IEEE **63**(12), 1692–1716 (1975)