Mean-Square Error in Periodogram Approaches With Adaptive Windowing

Soosan Beheshti, *Senior Member, IEEE*, Maryam Ravan, *Member, IEEE*, James P. Reilly, *Member, IEEE*, and Laurel J. Trainor

***Abstract*—Modified periodogram approaches are nonparametric power spectral density (PSD) estimators. Here, we present a method for estimating the mean-square error (MSE) of these PSD estimators. The proposed approach uses the observed data to estimate not only the PSD but also the associated MSE simultaneously. The MSE estimate from the Blackman–Tukey approach can be utilized for comparison and choice of the optimum window among a set of smoothing windows of possibly different lengths. For Bartlett and Welch methods, the MSE estimate can be used for quality evaluation, and also enables the use of an additional smooth windowing for these modified periodogram approaches. The optimum adaptive windowing improves the performance of these approaches in the MSE sense. Furthermore, the optimally windowed autocorrelation estimate can be used for extrapolation with the maximum entropy method (MEM). Our simulation results confirm that the proposed optimum smooth windowing approach effectively improves the performance of modified periodogram PSD estimates in the MSE sense.**

***Index Terms*—Correlation, estimation, periodogram, spectral analysis.**

I. Introduction

SPECTRAL estimation is by now a mature topic, with applications in non-destructive testing, surveillance, radar and sonar, direction of arrival estimation, diagnostics, and many more. There have been many successful inroads to this problem, including various nonparametric modeling approaches such as the (modified) periodogram approaches of Bartlett and Welch, and the smoothing approach of Blackman–Tukey [7], [14], [15]. The focus of this paper is on improving the performance of these methods in the mean-square error (MSE) sense.

Manuscript received March 28, 2010; revised July 30, 2010, September 28, 2010; accepted November 04, 2010. Date of publication November 22, 2010; date of current version February 09, 2011. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Maria Sabrina Greco.

S. Beheshti is with the Department of Electrical and Computer Engineering, Ryerson University, Toronto, ON M5B 2K3, Canada (e-mail: soosan@ee.ryerson.ca).

M. Ravan is with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON M5S 1A1, Canada (e-mail: mravan@comm.utoronto.ca).

J. P. Reilly is with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON L8S4L8, Canada (e-mail: reilly@mail.ece.mcmaster.ca).

L. J. Trainor is with the Department of Psychology, Neuroscience and Behaviour, Faculty of Science, McMaster University, Hamilton, ON L8S4L8, Canada (e-mail: ljt@mcmaster.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2010.2094192

There are different design elements in each of these approaches. For a method with fixed design elements, evaluation schemes for assessing the performance of power spectral density (PSD) estimation methods often use autocorrelation error and/or PSD frequency error. The asymptotic behavior of these errors as the data length grows, at a particular point in the autocorrelation estimate or at a particular frequency in the PSD estimate, has been well studied [6], [14], [15]. Here, we focus on estimating the MSE of the PSD estimators. This error, also known as integrated MSE (IMSE) [16], is a valuable criterion for performance evaluation of the PSD estimators. In estimation problems, the MSE plays an important role in evaluating the estimator and much research effort has been focused on estimating this error for the purpose of estimator performance assessment [12]. The importance of MSE in PSD estimation is acknowledged in the recent work concerning PSD estimation which provides the cepstral thresholding [16]. In this paper, we propose a novel method for estimating the PSD mean-square error (PMSE) and suggest the use of this criterion to enhance the performance of perodogram-based PSD estimators. Note that, due to the basic properties of the Fourier transform, PMSE in the frequency domain is equal to autocorrelation mean-square error (AMSE) in the time domain.

We show that the AMSE behavior puts the role of additional smooth windowing into a new perspective and that this smoothing can reduce the bias effects due to the leakage in the modified periodograms. The AMSE of tapered versions of the modified periodograms can be estimated and compared. As a result, the tapering window that minimizes the AMSE can be chosen. This study can provide a rigorous justification for using lag windows in spectrum estimation and clarify the ambiguity of the role of this type of PSD tapering [8]. Note that this type of tapering is different from what is denoted by multi-tapering in PSD estimation [18]. In the multi-tapering approaches, the raw data themselves are tapered, whereas here the estimated autocorrelation is tapered. An existing example of such tapering is the Blackman–Tukey approach. It is known that the performance of the basic periodogram method is improved by the additional smooth windowing in the Blackman–Tukey approach. Moreover, it has already been acknowledged that the shape and length of the window affect the estimator variance of this method. Heuristically, the number of lags for this approach is recommended to be around 20% of the number of data samples [10]. However, this window length is not necessarily the optimum one in many practical applications, and the search for the optimum window in particular cases is performed by trial and error [3]. In this paper, we propose a novel method that first estimates the AMSE of the corresponding spectral estimator, and then chooses the optimal window, among competing windows, by minimizing this AMSE estimate. We demonstrate how the additional smooth windowing can improve Bartlett and Welch estimates by comparing and minimizing the associated AMSE. The main contribution of this work is in estimating the AMSE by using the same finite length data that is used for estimating the PSD itself. In addition, the truncated autocorrelation estimate which results from the additional smooth windowing can be used with the maximum entropy method (MEM) [4] for extrapolation and autocorrelation tail estimation. We have demonstrated significant performance improvements resulting from the proposed procedure over that of the classical periodogram approaches, through the use of simulated and real data.

The organization of the rest of the paper is as follows. Section II provides the preliminaries and notations that are used throughout the paper. The effect of smooth windowing in modified periodogram approaches is discussed in Section III. In Section IV the role of AMSE in performance evaluation and in the design of the modified periodograms is discussed. Section V studies the structure of AMSE and provides a method of AMSE estimation using only the available data. Section VI briefly presents the use of MEM for extrapolation and summarizes the resulting AMSE estimator algorithm for optimum windowing. Section VII gives the simulation results and Section VIII contains the concluding remarks.

II. Notations and Preliminaries

Consider a wide-sense stationary zero mean random process X with autocorrelation $r_{xx}[n]$ and power density spectrum $P_{xx}(e^{j\omega})$:

$$r_{xx}[n] = E\left(X[l]X\left[l+|n|\right]\right) \quad (1)$$

$$P_{xx}(e^{j\omega}) = \sum_{n=-\infty}^{\infty} r_{xx}[n]e^{-j\omega n}. \quad (2)$$

Given a sample of X of finite length N, $(x[0], x[1], \cdots, x[N-1])$, we wish to determine an estimate of the autocorrelation and PSD of X.

The modified periodogram approaches are Bartlett method (averaging periodogram), Welch method (averaging modified periodogram), and Blackman–Tukey method (periodogram smoothing) [6], [14]. In this section we set out the related notations for these methods that are used throughout the paper.

Important Notation: While the true autocorrelation $r_{xx}[n]$ is a fixed number for each n, the estimate of this value using the observed data is denoted by $\hat{r}_{xx}[n]$, which is a sample of random variable $\hat{r}_{XX}[n]$.

A. Common Biased and Unbiased Autocorrelation Estimators

In periodogram approaches the autocorrelation estimates are provided by averaging autocorrelation estimates of S segments of the available data. These segments of length L are denoted by $x_{s_i}[n]$ where each segment starts at $s_i = (i-1)D$:

$$x_{s_i}[n] = x\left[n+(i-1)D\right]. \quad 1 \le i \le S,\ 0 \le n \le L-1 \quad (3)$$

For non-overlapped segments $D = L$ and for overlapped segments $D < L$. The windowed version of these segments is used for autocorrelation estimation:

$$y_{s_i}[n] = x_{s_i}[n]w[n] \quad (4)$$

where $w[n]$, known as time window or temporal window, is of length L. Without loss of generality, we assume that the window has a unit power[1]

$$\frac{1}{L}\sum_{n=0}^{L-1} w^2[n] = 1. \quad (5)$$

We also assume the following condition on the time window

$$\sum_{l=0}^{L-|n|-1} w[l]w[l+n] \neq 0, \quad 0 \le |n| \le L-1. \quad (6)$$

The final estimate is in the form of

$$\hat{r}_{xx}[n] = \frac{1}{S}\sum_{i=1}^{S} \hat{r}_{xx}[n, s_i], \quad 0 \le |n| < L-1 \quad (7)$$

where the autocorrelation estimate of each segment is denoted by $\hat{r}_{xx}[n, s_i]$. The segment estimates for an unbiased estimator, for $0 \le |n| < L-1$ and $1 \le i \le S$, are in the form of

$$\hat{r}_{xx}[n, s_i] = \frac{1}{L-|n|}\sum_{l=0}^{L-|n|-1} y_{s_i}[l]y_{s_i}\left[l+|n|\right] \quad (8)$$

and for a biased estimator are in the form of

$$\hat{r}_{xx}[n, s_i] = \frac{1}{L}\sum_{l=0}^{L-|n|-1} y_{s_i}[l]y_{s_i}\left[l+|n|\right]. \quad (9)$$

The biased estimator provides a much lower error variance, which is one reason why the biased estimator is preferred.

B. PSD Estimate

The PSD estimate is the Fourier transform of $\hat{r}_{xx}[n]$ in (7)

$$\hat{P}_{xx}(e^{j\omega}) = \sum_{n=-L+1}^{L-1} \hat{r}_{xx}[n]e^{-j\omega n}. \quad (10)$$

In particular, for the biased estimator in (9), we have

$$\hat{P}_{xx}(e^{j\omega}) = \frac{1}{S}\sum_{i=1}^{S}\frac{1}{L}\left|\sum_{n=0}^{L-1} x_{s_i}[n]w[n]e^{-j\omega n}\right|^2 = \frac{1}{S}\sum_{i=1}^{S}\frac{1}{L}\left|Y_{s_i}(e^{j\omega})\right|^2 \quad (11)$$

where $Y_{s_i}(e^{j\omega})$ is the Fourier transform of y_{s_i} in (4).

[1]In a more general PSD estimation this window can be replaced by a class of orthonormal tapers of form $w_i[n]$ [18]. While we keep the notation $w[n]$, the work of this paper can be generalized for these multi-tapering windows.

III. Smoothing Window

The general design elements of the modified and smoothing periodograms are as follows.

a) Choice of the segment length L and the overlapping percentage in (3). In the Welch approach the segments can overlap $D \leq L$. However, in the Bartlett method there is no overlap among the segments $D = L$.

b) Choice of the time window $w[n]$ with length L in (4).

In the Blackman–Tukey approach there is no segmentation $(S = 1)$ and we have $D = L = N$. The *only* design element in this approach is the smoothing window. In this case, consider a window of length $2m - 1$ $(m \leq L)$, denoted by $g_m[n]$ ($|g_m[n]| \leq 1$ for all n), which is multiplied by the estimate $\hat{r}_{xx}$

$$\hat{r}_{xx,g_m}[n] = g_m[n]\hat{r}_{xx}[n]. \tag{12}$$

This windowing truncates and shapes the middle part of the autocorrelation estimate. The PSD estimate in this case is

$$\hat{P}_{xx}^m(e^{j\omega}) = \sum_{n=-m+1}^{m-1} \hat{r}_{xx,g_m}[n]e^{-j\omega n}. \tag{13}$$

This PSD estimate is equivalently the convolution of $G(e^{j\omega})$, the Fourier transform of g_m, with the PSD estimate of the periodogram method. The window g_m is known as the smoothing PSD window or the lag window.

We consider the smooth windowing as a design element for all the periodogram approaches. Therefore, in addition to the design parts in a) and b), we have

c) Choice of the smoothing window $g_m[n]$ in (12) with length $2m - 1$ among competing windows of possibly different lengths $(1 \leq m \leq L)$.

IV. Effect of the Smoothing Window on Mean-Square Error

The modified periodograms provide an estimate of $2L - 1$ samples of the true autocorrelation function, which we can denote as a vector r_{xx}^L, given as

$$r_{xx}^L = [r_{xx}[-L+1], \cdots, r_{xx}[0], \cdots, r_{xx}[L-1]]^T \tag{14}$$

where $.^T$ denotes the transpose operation. The autocorrelation estimates without any smooth windowing are in form of

$$\hat{r}_{xx}^L = [\hat{r}_{xx}[-L+1], \cdots, \hat{r}_{xx}[0], \cdots, \hat{r}_{xx}[L-1]]^T \tag{15}$$

and the effect of additional smoothing window in (12) sets some of the values of this estimate to zero and shapes the middle points to[2]

$$\hat{r}_{xx}^m = [0, \cdots, 0, \hat{r}_{xx,g_m}[-m+1], \cdots, \hat{r}_{xx,g_m}[m-1], 0, \cdots, 0]^T. \tag{16}$$

[2]For simplicity and without loss of generality, we eliminate g_m and only keep its length m in this notation.

The AMSE for this estimate is[3]

$$\text{AMSE}[m] = E\left(\left\|r_{xx}^L - \hat{r}_{XX}^m\right\|^2\right). \tag{17}$$

(where The AMSE is equal to the PMSE in the frequency domain due to Parseval's theorem[4]:

$$\text{AMSE}[m] = \text{PMSE}[m]. \tag{18}$$

The MSE error is also denoted as IMSE [16].

The averaging periodogram approaches, Bartlett and Welch, are special cases of the smooth windowing with a uniform window of length $m = L$. Traditionally this window is a design element of the Blackman–Tukey method. Here we use the smooth windowing with all the periodogram methods including the averaging periodograms. Due to finiteness of the available data, the AMSE of the smoothing windowed version of the averaging periodograms may be less than that of the methods alone. This is the motivation for calculating AMSE in the presence of smooth windowing with these approaches. Note that the Blackman–Tukey method is already a smoothing windowed version of the basic periodogram approach and performs much better than the periodogram itself in the sense of MSE. The choice of the optimum window in the Blackman–Tukey approach has been an *ad hoc* process, and it is suggested to choose m, the window length, to be around one fifth of the data length N [15].[5] We propose to find the optimum window of the Blackman–Tukey by minimizing the AMSE among windows of different lengths. To illustrate the effects of the smooth windowing, we first fix the design items a) and b), the number of segments and window $w[n]$, and concentrate on the effects of the smoothing window. We also assume that a class of possible smoothing windows is available and plan to choose

[3]Here $\|y\|$ represents the l^2-norm of vector y. Note that the complete AMSE error between the estimate and the true autocorrelation is in form of

$$\text{AMSE}[m] + \sum_{|j|=L}^{\infty} r_{xx}^2[j]$$

where the second term is constant for variable smoothing windows and a fixed L. When comparing the AMSEs with variable m, this extra term can be ignored.

[4]The PSD mean-square error is

$$\begin{aligned}\text{PMSE}[m] &= \frac{1}{2\pi}\text{E}\left(\int_{\langle 2\pi\rangle}\left|\hat{P}_{XX}^m(e^{j\omega}) - P_{XX}^L(e^{j\omega})\right|^2 d\omega\right)\\ &= \frac{1}{2L-1}\sum_{k=0}^{2L-2}\text{E}\left(\left|\hat{P}_{XX}^m\left(e^{j\frac{2\pi k}{2L-1}}\right) - P_{XX}^L\left(e^{j\frac{2\pi k}{2L-1}}\right)\right|^2\right)\end{aligned}$$

where $\hat{P}_{xx}^m$ is the PSD estimate in (13) and P_{XX}^L is the PSD of the truncated autocorrelation of length $2L - 1$ in (14). The last summation is for discrete Fourier transform (DFT) at $2L - 1$ frequencies, which can also represent the MSE of the FFT of the autocorrelation.

[5]In [11], it is recommended that $m < N/5$.

among these $g_m[n]$s of different length,[6] i.e., of different values of m. The resulting autocorrelation estimate $\hat{r}_{xx}^m$ of length $2m-1$ with this smoothing window is in (12). To choose the optimum window among windows of different lengths, we can comparethe AMSEs in (17) and find the one that minimizes this MSE in the time (and, equivalently, in the frequency) domain:

$$m^* = \arg\min_m \text{AMSE}[m]. \tag{19}$$

A typical performance of AMSE as a function of the window length is illustrated in Fig. 1. In Fig. 1(a), the desired (true) autocorrelation and the estimated autocorrelation of a signal using the Bartlett method with a data segment of length $L = 10$ and uniform smoothing is shown. As the figure shows, as $|m|$ becomes larger than seven, the difference between the desired and estimated autocorrelation becomes noticeable. Therefore, we can not trust $\hat{r}_{xx}^{10}[n]$ when $|n| > 7$ with the same confidence as in $\hat{r}_{xx}^{10}[n]$ when $|n| \le 7$. The logical reason for this behavior is that the closer we get to the tail of the estimate, the less data is available to estimate the desired autocorrelation and the estimation error becomes larger. Fig. 1(b) shows the values of AMSE for different values of m. As the figure confirms, the AMSE has a minimum at $m^* = 7$. Therefore, it indicates that we can only trust $2m^* - 1$ middle estimates of autocorrelation in order to have the smallest AMSE. The main challenge in this paper is estimating the unavailable AMSE by using only the available finite data.[7]

V. Estimating AMSE

A. Biased Estimators

The true autocorrelation is related to the expected value of the autocorrelation estimate as follows:

$$r_{xx}[n] = \frac{1}{\alpha[n]} E\left(\hat{r}_{XX}[n]\right) \tag{20}$$

where for the biased estimator $\alpha[n]$ is[8]

$$\alpha[n] = \frac{1}{L} \sum_{l=0}^{L-|n|-1} w[l]w[l+n] \tag{21}$$

[6]The competing windows can have the same length. However, for notation simplicity and without loss of generality we represent a window with length m with $g_m[n]$.

[7]Note that our smooth windowing is different from the sharpened periodogram that was introduced in [19]. In that work smoothing is done by the use of kernel estimates in the frequency domain. In addition, in [19] the PSD estimates are modeled with an extra assumption about the relationship between the PSD of the random variable and the true PSD. It assumes that the PSD estimate at frequency ω is in the form of $\epsilon(\omega)P_{xx}(e^{j\omega})$, where $\epsilon(\omega)$ itself is a random variable with particular properties. This assumption helps in providing an estimate for the AMSE. However, in our work we avoid any of such extra assumptions on the PSD estimates.

[8]The condition in (6) guarantees $\alpha[n]$ to be nonzero for all m.

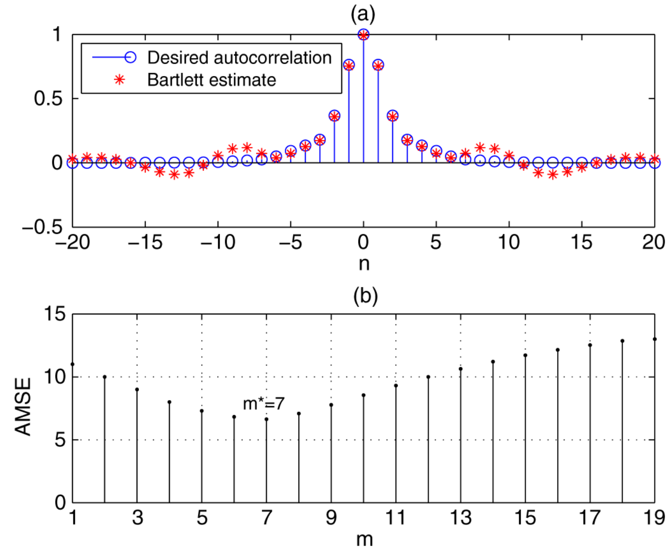

Fig. 1. (a) Desired autocorrelation function and Bartlett estimation of autocorrelation function, (b) AMSE as a function of m (smoothing window length is $2m-1$).

and the detail of this calculation is provided in Appendix I-A. The calculated autocorrelation estimate is used as an estimate of its expected value:

$$\widehat{E\left(\hat{r}_{XX}[n]\right)} = \hat{r}_{xx}[n]. \tag{22}$$

The reliability of this estimate depends on the behavior of the variance of this estimator versus the value of its mean. For example if $E(\hat{r}_{XX}[2])$ is 3 and its variance $\text{var}(\hat{r}_{XX}[2])$ is of order 0.04 (standard deviation (stddev)) of order 0.2), the sample $\hat{r}_{xx}[2]$ can be a valid estimate of the true autocorrelation at 2. However, if the stddev of this random variable is comparable with its mean, then the one sample of the random variable is usually not a good estimate of its mean. So the desired evaluation ratio is the autocorrelation quality factor (AQF)

$$\text{AQF}[n] = \frac{\left\{E\left(\hat{r}_{XX}[n]\right)\right\}^2}{\text{var}\left(\hat{r}_{XX}[n]\right)} \tag{23}$$

that should to be as large as possible. Exact calculation of the variance $\text{var}(\hat{r}_{XX}[n])$ for a Gaussian process is provided in Appendix II. As shown in the Appendix, this variance for the Blackman–Tukey with a biased estimator and a uniform time window, as a function of $g_m[n]$ and $r_{xx}[n]$ is

$$\text{var}\left(\hat{r}_{XX}^m[n]\right) = |g_m[n]|^2 \sum_{k=0}^{N-1-|n|} \frac{N-1-|n|-k}{N^2} \times \left(r_{xx}^2[k] + r_{xx}\left[k+|n|\right] r_{xx}\left[k-|n|\right]\right). \tag{24}$$

For the unbiased estimator the term $1/N^2$ in this variance is increased to $1/(N-n)^2$. This will cause a relatively larger variance, and consequently smaller AQF, for the unbiased estimator at large n. Hence, the biased estimator is preferred over the unbiased one.

B. AMSE Structure and Optimum Window Length

The AMSE$[m]$ in (17) is

$$\text{AMSE}[m] = \Delta[m] + \sum_{n=-m+1}^{m-1} (1 - g_m[n]\alpha[n])^2 \, r_{xx}^2[n] + \text{var}\left(\hat{r}_{XX}^m[n]\right) \quad (25)$$

where

$$\Delta[m] = \sum_{|n|=m}^{L-1} r_{xx}^2[n] \quad (26)$$

and $\alpha[n]$ is provided in (21). Details of this calculation are given in Appendix I. As shown in (24), the variance term $\text{var}(\hat{r}_{XX}[n])$ in (25) is a function of r_{xx} and the smoothing window $g_m[n]$. Therefore, AMSE is a function of $r_{xx}[n]$, time windows $w[n]$, and smoothing window $g_m[n]$ for a range of values of n.

While the first term of AMSE, $\Delta[m]$, is a decreasing function of m, the second term in (25) is an increasing function of m as it is summation of positive values over the range of $-m$ and m. Therefore, the tradeoff between these terms always results in an optimum window length m^* ($m^* \leq L$).

C. Estimating AMSE Using the Observed Data

From (25), it is evident that in order to estimate the AMSE we only need an estimate of the autocorrelation r_{xx}. This estimate can be provided by using (22). This will provide an estimate of the first and second terms of the AMSE. The same autocorrelation estimate can be used in estimating the variance in (24), which is the third term of AMSE. Consequently, the same data that are used for PSD estimation are also used to estimate the associated AMSE. We denote this estimate of AMSE$[m]$ as $\widehat{\text{AMSE}}[m]$

$$\widehat{\text{AMSE}}[m] = \widehat{\Delta[m]} + \sum_{n=-m+1}^{m-1} (1 - g_m[n]\alpha[n])^2 \, \widehat{r_{xx}^2[n]} + \text{var}\left(\widehat{\hat{r}_{XX}^m}[n]\right). \quad (27)$$

The estimate of the optimum smoothing window length is obtained by minimizing the estimate of AMSE

$$\hat{m}^* = \arg\min_m \widehat{\text{AMSE}}[m]. \quad (28)$$

D. AMSE Estimator Behavior

The PSD in (2) is a weighted sum of the autocorrelation values. Similarly, the AMSE$[m]$ in (25) is a weighted sum of products of two autocorrelation values. For both PSD and AMSE the estimates are provided by replacing the autocorrelations with their estimates. Consequently, the validation process for the AMSE estimate is similar to that for the PSD estimate. Mean and variance of AMSE estimate for the Blackman-Tukey approach are provided in Appendix III. The following asymptotic behavior of this AMSE estimator proves that the estimator is consistent:

$$\lim_{N\to\infty} \left(\widehat{\text{AMSE}}[m] - \text{AMSE}[m]\right) = 0 \quad (29)$$

$$\lim_{N\to\infty} \text{var}\left(\widehat{\text{AMSE}}[m]\right) = 0. \quad (30)$$

Details are provided in Appendix III.

E. Smooth Windowing and PSD Quality Factor

It is known that the additional averaging or windowing in modified periodogram approaches reduces the frequency resolution of the periodogram with the advantage of reducing the frequency variance error. In this case, the following PSD quality factor [6]:

$$Q(e^{j\omega}) = \frac{\left\{E\left(\hat{P}_{XX}(e^{j\omega})\right)\right\}^2}{\text{var}\left(\hat{P}_{XX}(e^{j\omega})\right)} \quad (31)$$

is improved by the additional segmentation, averaging, and windowing [14]. The estimate of this quality, as the data length N approaches infinity, is calculated in [14] and [6]. It is illustrated that the smooth windowing improves the quality factor of the basic periodogram in the Blackman–Tukey method. With S segments the PSD estimate is [1]

$$\hat{P}_{XX}(e^{j\omega})) = \frac{1}{S}\sum_{i=1}^{S} \hat{P}_{XX}^{(s_i)}(e^{j\omega}) \quad (32)$$

where $\hat{P}_{XX}^{(s_i)}$ is the PSD estimate of each segment. Due to this relation, the provided expected value and variance of smooth windowed version of periodogram (in Blackman–Tukey) can be generalized for the smooth windowed version of the averaging method[9]

$$E\left(\hat{P}_{XX}(e^{j\omega})\right) \approx G_m(e^{j\omega}) * P_{XX}(e^{j\omega}) \quad (33)$$

$$\text{var}\left(\hat{P}_{XX}(e^{j\omega})\right) \approx P_{XX}(e^{j\omega}) \left[\frac{1}{N}\sum_{n=-m+1}^{m-1} g_m^2[n]\right] \quad (34)$$

where $G_m(e^{j\omega})$ is the Fourier transform of the smoothing window. Details of this calculation are shown in Appendix IV. These values are exactly the expected value and variance for the Blackman–Tukey with no segmentation. Therefore, the PSD quality factor of the smooth windowed version of the averaging periodograms is the same as that of the Blackman–Tukey approach.

VI. Maximum Entropy Method for Autocorrelation Extrapolation

Optimum windowing provides $\hat{m}^*$ the optimum length of autocorrelation estimates that can be trusted based on the observed finite length data. However, many signals of interest have autocorrelations that are nonzero for $|n| > \hat{m}^*$. Therefore, to achieve the proper frequency resolution of PSD, the tail of the autocorrelation needs to be estimated. We use the Maximum Entropy method to extrapolate this tail from the optimally windowed autocorrelation sequence [6]. Given the available windowed autocorrelation, i.e., $\hat{r}_{xx}[n]$ for lags $|n| < \hat{m}^*$, the MEM extrapolates $\hat{r}_{xx}[n]$ for $|n| \geq \hat{m}^*$. Denoting the extrapolated

[9] "$*$" denotes the convolution operation

TABLE I

(S, L): NUMBER OF SEGMENTS AND LENGTH OF SEGMENTS. m^*: LENGTH OF THE OPTIMUM SMOOTHING WINDOW THAT MINIMIZES AMSE. $\hat{m}^*$: ESTIMATE OF m^* USING THE ESTIMATE OF AMSE. AMSE[L]: MSE OF THE APPROACH WITHOUT A SMOOTHING WINDOW. AMSE[m^*], AMSE[$\hat{m}^*$]: MSE OF THE OPTIMALLY WINDOWED PSDE APPROACH USING AMSE AND AMSE ESTIMATE. AMSE[$\hat{m}^*$] & MEM: MSE OF THE EXTRAPOLATED VERSION OF THE OPTIMALLY WINDOWED ESTIMATE

	PSD Estimator (PSDE)	(S, L)	m^*	$\hat{m}^*$	AMSE[m^*]	AMSE[$\hat{m}^*$]	AMSE[$\hat{m}^*$] & MEM	AMSE[L]
x_1	Blackman-Tukey	(1, 1000)	9	7	15	17	14.5	29.20
x_1	Welch	(39,50)	8	7	120	123	102	134
x_3	Bartlett	(10,100)	24	22	8.48	8.56	4.11	13.89
x_2	Blackman-Tukey	(1,1000)	19	16	25	27	21	44.95

value by $r^e_{xx}[n]$, the power spectrum of the extrapolated version is

$$P^e_{xx}(e^{j\omega}) = \sum_{n=-\hat{m}^*+1}^{\hat{m}^*-1} \hat{r}_{xx}[n]e^{-j\omega n} + \sum_{|n|\geq \hat{m}^*} r^e_{xx}[n]e^{-j\omega n}. \tag{35}$$

The tail is added to maximize the entropy of the Gaussian process under the constraint that the $\hat{r}_{xx}[n]$ is available for $|n| < m^*$.

The following summarizes the optimum windowing algorithm using the proposed AMSE estimation:

Optimum Windowing Algorithm

1) The autocorrelation of the available finite length data is calculated with (7).
2) The windowed version of the autocorrelation estimate is calculated in (12). The AMSE estimate (27) is calculated for each windowed version of length m, $1 \leq m \leq L$, as it is described in Section V-C by using the same available finite length data.
3) Among the smoothing windows of length m, $1 \leq m \leq L$, the one that minimizes the AMSE estimate in (28) is chosen. The adaptive window with length $\hat{m}^*$ is a function of the observed data itself.
4) To recover the truncated tail of the autocorrelation, the maximum entropy method can be implemented on the optimum autocorrelation estimate of length $\hat{m}^*$, and AMSE is estimated for the extrapolated version. Between the estimate without MEM (from step 4) and the extrapolated one, the algorithm retains the one with smaller AMSE estimate.

VII. SIMULATION AND EXPERIMENTAL RESULTS

In the following sections we illustrate the performance of the proposed method in both simulation and experimental situations.

A. Simulation Results

Consider a wide-sense stationary (WSS) random process X with the following structure

$$X = h * W \tag{36}$$

where W is a unit variance white Gaussian process and h is a filter with real coefficients ('$*$' denotes the convolution operator). The true autocorrelation $r_{xx}[n]$ and power density spectrum $P_{xx}(e^{j\omega})$ of this random process are

$$r_{xx}[n] = \sum_n h[l]h[l+n] \tag{37}$$

$$P_{xx}(e^{j\omega}) = \sum_{n=-\infty}^{\infty} r_{xx}[n]e^{-j\omega n}. \tag{38}$$

To simulate the method, five filter sequences $h_1, \cdots, h_5$ were generated. The sequence h_1 has a single pole that generates an infinite length autocorrelation and filter h_2 is finite length, given by

$$h_1[n] = 2.5 \times 0.7^n u[n] \tag{39}$$

$$h_2[n] = \text{Kaiser filter of length } 40 \tag{40}$$

where $u[n]$ is a unit step function. The filters h_3, h_4, and h_5 with the structure of (36) have the following form:

$$H(z) = k\frac{A(z)}{B(z)} \tag{41}$$

where the values for h_3, h_4, and h_5 are[10]

$$h_3: k = 2.33, A = [1, -.51], B = [1, -1.5, 0.54] \tag{42}$$

$$h_4: k = 100, A = [1, -1, 1], B = [36, -60, 37, -10, 1] \tag{43}$$

$$h_5: k = 10, A = [1, -.8], B = [1, -1, 0.5] \tag{44}$$

while h_3 provides a PSD with dominant low frequencies, h_4 has a zero at frequency 0.33π rad/s, and h_5 has a peak at 0.25π rad/s. The data length in the following experiments is 1000. Fig. 2 shows the behavior of AMSE as a function of the length of the smoothing window for x_3 and with the Bartlett method. It also shows the estimate of AMSE using the available data. The resulting autocorrelation estimate is shown in Fig. 3. As the figure shows, the optimally windowed autocorrelation estimate suppresses ripples far away from the true autocorrelation and set them to zero. On the other hand, the extrapolation of this truncated estimate with MEM provides an improved estimate of the tail closer to the true autocorrelation. Table I shows the results of using the PSDE methods with AMSE adaptive

[10]Filter h_3 has two poles at 0.6 and 0.9 and a zero at 0.51; filter $h_4[n]$ has two poles at 1/2, two poles at 1/3 and two zeros at $e^{\pm j\pi/3} = 0.5 \pm j0.866$ (on the unit circle); and filter $h_5[h]$ has a zero at 0.8 and two poles at $0.5\sqrt{2}e^{\pm j\pi/4} = .5 \pm .5j$.

TABLE II
AMSE FOR DIFFERENT PSD ESTIMATION METHODS (AVERAGED OVER 50 TRIALS)

	Average length of optimum window $\hat{m}^*$	Blackman-Tukey N/5=200	Blackman-Tukey $\hat{m}^*$	Blackman-Tukey $\hat{m}^*$ & MEM	MTM
x_1	9	28	18	14	31
x_2	18	43	27	25	43
x_3	53	1.1e3	0.91e3	1.52e3	1.15e3
x_4	30	1.58e3	0.88e3	1.76e3	1.8e3
x_5	32	1.95e3	1.01e3	1.6e3	2.2e3

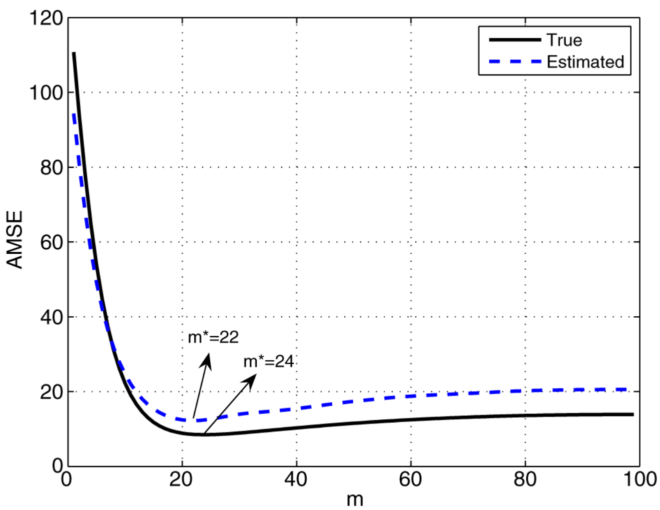

Fig. 2. **Bartlett Windowing**: AMSE (solid line) and $\widehat{\text{AMSE}}$ (dotted line) for 10 segments of x_3 as a function of m (data length N is 1000).

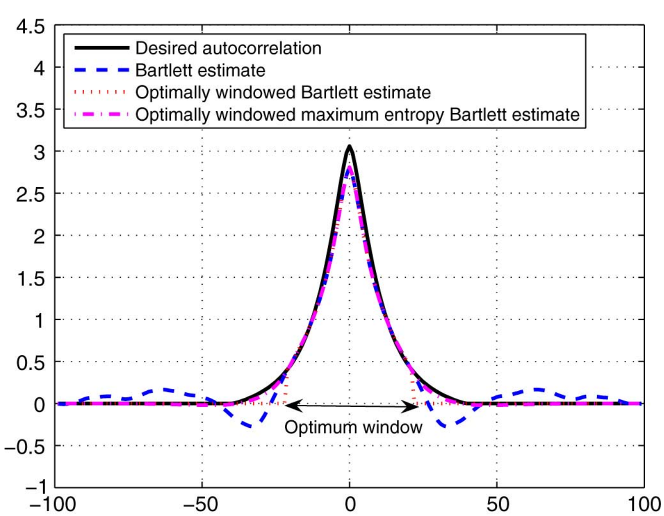

Fig. 3. Autocorrelation of random process x_3 (solid line), Bartlett method (dashed line), Bartlett optimum windowing $\hat{m}^* = 22$ (dotted line) and maximum entropy Bartlett optimum windowing (dashed–dotted line).

windowing. For example, in the case of x_1, the periodogram error (no windowing) is as large as 29.20. Using the smoothing window of Blackman-Tukey reduces this error, and for the optimal window length of $2m^* - 1 = 17$ the error is as small as 15. Using our method, the estimate of the optimum window length is $2\hat{m}^* - 1 = 13$ that results in an error of 17. Similar results for x_2 are also shown in the table. As the table shows, the MEM extrapolation can reduce the amount of the MSE error. To check whether the MEM improves the PSD estimate, the algorithm compares the AMSE estimates and decides.

Table II Compares the performance of the Blackman-Tukey with the conventional window length of one fifth of the data length and the method with optimum window along with the Multi-tapering method (MTM) [17].[11]

Fig. 4 shows the PSD estimates for x_4 and x_5. As the figures show, if the purpose of PSD estimation is to provide estimates at all frequencies that on average are as close as possible to the true PSD, the proposed method outperforms MTM. The average MSE errors are smaller for the proposed method (numbers are provided in Table II). On the other hand, if the goal of the PSD estimator is to capture the PSD behavior close to zeros, the plot in decibels shows that the MTM performs better. This happens at the cost of increasing the MSE of the PSD estimate due to the large errors at frequencies with high PSD values.

So far we have shown the simulation results for the optimum windowing algorithm. In the following, the method is used with a real set of data.

B. *Experimental Application in a Classification Problem*

We consider a classification problem for a specific subject age and corresponding state of brain development based on the electroencephalography (EEG) signal in response to an auditory stimulus. EEGs were recorded using a HydroCel GSN (HCGSN) (Electrical Geodesics, Inc., Eugene, OR) from 128 locations on the scalp. The classification problem under consideration is the assignment of subjects to one of the three predetermined age groups, which are 6-month-old and 12-month-old infants and adults. The EEG signals of 68 healthy subjects, each with length 475, consisting of 29 6-month-olds, 19 12-month-olds, and 20 adults with no known hearing deficits, were used for the classification. For these types of classification problems, the wavelet transform of the autocorrelation sequences gives improved features for the discrimination of age group. In particular, the smoothing feature inherent in the Daubechies wavelet of order 2 made it most suitable for this application. Since, the total number of wavelet coefficients (candidate features) is very large, we applied the maximum mutual information and minimum redundancy criteria to select the most relevant features [13]. The classifier is trained by using these features to predict the age group of each subject. Among different classification methods, the fuzzy c-means (FCM)

[11]We use the Thomson multi-tapers with seven tapers. (The time-bandwidth product in MTM method has the suggested default value of 4.) For the available training data, the estimate becomes smoother if this value is changed through supervision. However, this type of tuning is not applicable in real application as the true PSD is not available for comparison purposes.

TABLE III
COMPARISON OF THE CLASSIFICATION PERFORMANCE IN PREDICTING THE AGE OF SUBJECTS USING DIFFERENT AUTOCORRELATION ESTIMATION METHODS

Method	Classes	6-months	12-months	Adults	Sensitivity	Specificity	TCA
Periodogram	6-months	28	1	0	96.6%	82%	
	12-months	5	14	0	73.7%	87.8%	80.9%
	Adults	2	5	13	65%	100%	
MTM	6-months	27	2	0	93.1%	82%	
	12-months	4	15	0	78.9%	85.7%	79.4%
	Adults	3	5	12	60%	100%	
BT & Adaptive Windowing	6-months	27	2	0	93.1%	89.7%	
	12-months	2	17	0	89.5%	91.8%	88.2%
	Adults	2	2	16	80%	100%	
BT & Window length L/5	6-months	27	2	0	93.1%	87.2%	
	12-months	2	17	0	89.5%	89.8%	85.3%
	Adults	3	3	14	70%	100%	

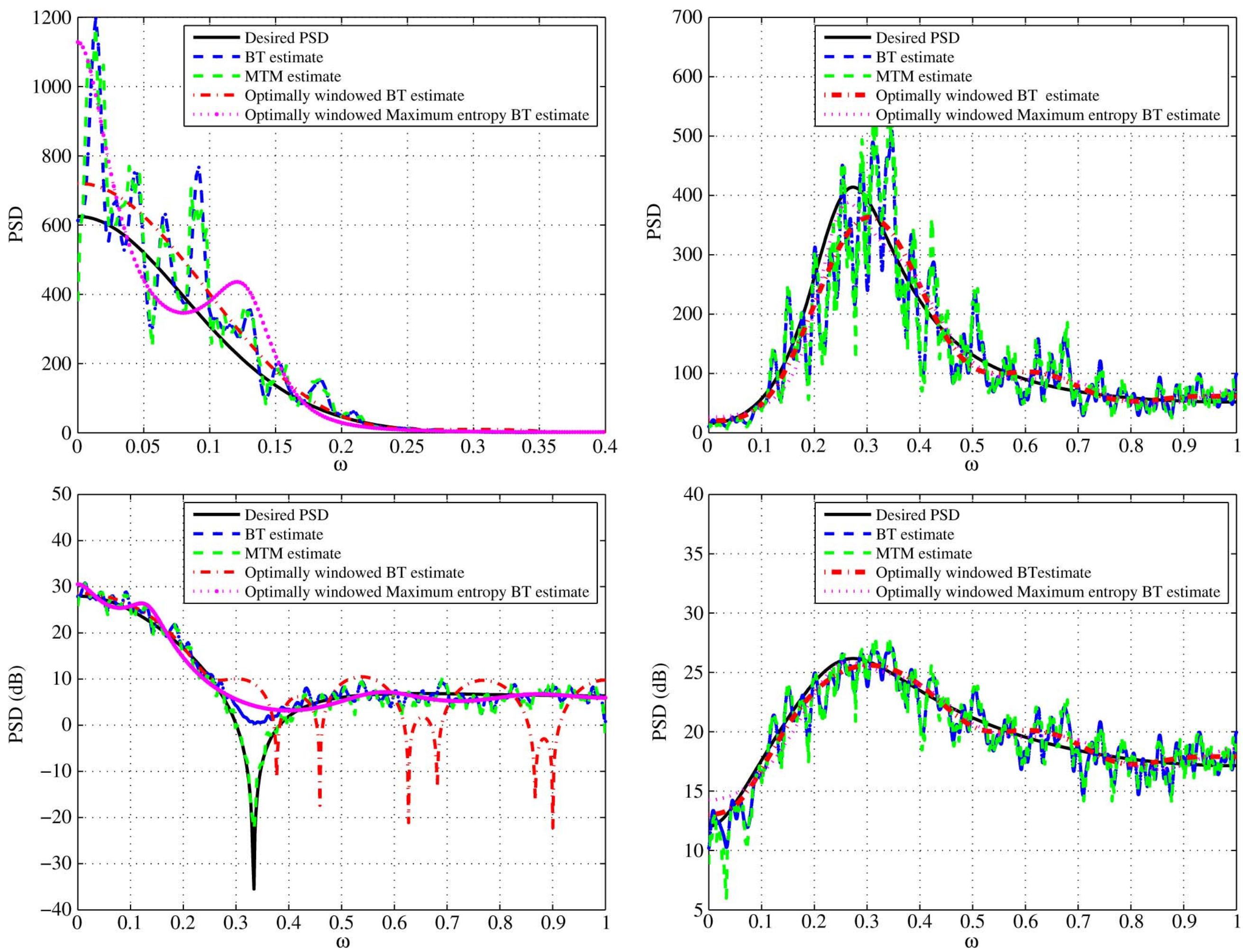

Fig. 4. PSD estimates of random process x_4 (left) and x_5 (right). The lower figures are the same PSDs in decibels.

classifier is found to provide the optimum performance for predicting the age group of the subjects [5]. Fig. 5 shows the estimated autocorrelation for three subjects, selected from three different age groups, using Bartlett with an optimum window length. Autocorrelation estimates for the MTM method are also shown. As the figures show, the optimally windowed autocorrelation estimates are more discriminative than the multi-taper ones. Table III compares the prediction performance of the FCM classifier using three autocorrelation estimation methods. In this table the parameters that are indicative of the performance of the resulting classification structure are computed.[12] The table indicates that the three classes of age groups are separated more accurately using an optimally windowed estimate of the autocorrelation in comparison to the case where multi-taper estimation is used. Note that the optimum length for the 68

[12]The indicative parameters are defined as follows: *Sensitivity*: number of subjects that are identified to be in one class divided by the number of subjects that are actually in that class. *Specificity*: number of true subjects that are identified not to be in a particular class divided by the total number of subjects that are actually not in that class. *Total classification accuracy (TCA)*: number of correct identifications in all classes divided by the total number of subjects.

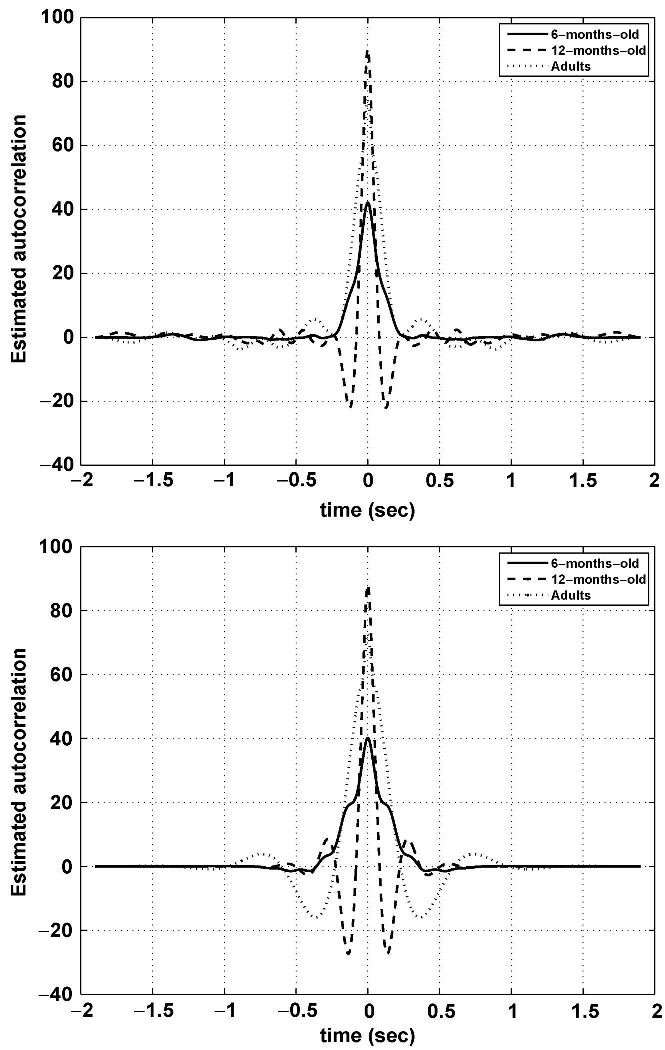

Fig. 5. Estimated autocorrelation of the frontal left part of the brain for three selected subjects from three age groups using MTM (above) and optimal smooth windowing methods (below).

signals varies between 9 to 30 which is much smaller than the conventional window length of Blackman–Tukey, that in this case is 95.

VIII. Conclusion

We proposed a new approach to estimate the unavailable autocorrelation (equivalently PSD) MSE using the observed data only. It was shown that the bias-variance tradeoff of the autocorrelation MSE leads to the existence of an optimum window length that minimizes the MSE amongst a set of smoothing windows. The MSE estimate was used to choose the optimum Blackman-Tukey smoothing window. The analysis of the behavior of the MSE mean and MSE variance proved the consistency of the MSE estimate for the Blackman–Tukey approach. Smooth windowing can also be applied to the averaging periodogram (Welch and Bartlett) methods. These methods with the optimum smoothing window are guaranteed to perform better than the methods with no additional smooth windowing in the sense of MSE. It was shown that maximum entropy extrapolation may improve performance in some cases. Whether or not there has been improvement can be determined by comparing the respective MSE estimates before and after the extrapolation.

Appendix I
Structure of AMSE

The AMSE in (17) is

$$\mathrm{AMSE}[m] = E\left(\left\|r_{xx}^{L} - \hat{r}_{XX}^{m}\right\|^{2}\right) \tag{45}$$

$$= \Delta[m] + \sum_{n=-m+1}^{m-1} E\left(r_{xx}[n] - \hat{r}_{XX}^{m}[n]\right)^{2} \tag{46}$$

where $\Delta[m] = \sum_{|n|=m}^{L-1} |r_{xx}[n]|^2$. By adding and subtracting an extra term $E(\hat{r}_{XX}^{m}[n])$ inside the second term we have

$$\mathrm{AMSE}[m] = \Delta[m] + \sum_{n=-m+1}^{m-1} E\left(\left(r_{xx}[n] - E\left(\hat{r}_{XX}^{m}[n]\right)\right) + \left(E\left(\hat{r}_{XX}^{m}[n]\right) - \hat{r}_{XX}^{m}[n]\right)\right)^{2}. \tag{47}$$

On the other hand, the extra term $E(\hat{r}_{XX}^{m}[n])$ has the following relationship with the true autocorrelation (see the following section, Appendix I-A, for details)

$$E\left(\hat{r}_{XX}^{m}[n]\right) = g_m[n]\alpha[n]r_{xx}[n]. \tag{48}$$

Therefore, we have

$$\begin{aligned}\mathrm{AMSE}[m] &= \Delta[m] + \sum_{n=-m+1}^{m-1} E\left(\left(1 - g_m[n]\alpha[n]\right) r_{xx}[n] + \left(E\left(\hat{r}_{XX}^{m}[n]\right) - \hat{r}_{XX}^{m}[n]\right)\right)^{2} \\ &= \Delta[m] + \sum_{n=-m+1}^{m-1} \left(1 - g_m[n]\alpha[n]\right)^{2} r_{xx}^{2}[n] + \mathrm{var}\left(\hat{r}_{XX}^{m}[n]\right) + 2\sum_{n=-m+1}^{m-1}\left(1 - g_m[n]\alpha[n]\right) \\ &\quad \times r_{xx}[n] E\left(E\left(\hat{r}_{XX}^{m}[n]\right) - \hat{r}_{XX}^{m}[n]\right)\end{aligned} \tag{49}$$

where the last summation term in (49) is zero.

A. *Calculation of the Expected Value* $E(\hat{r}_{XX}^{m}[n])$

From (12) we have

$$E\left(\hat{r}_{XX}^{m}[n]\right) = g_m[n] E\left(\hat{r}_{XX}[n]\right) \tag{50}$$

and from (7) we have

$$E\left(\hat{r}_{XX}[n]\right) = \frac{1}{S}\sum_{i=1}^{S} E\left(\hat{r}_{XX}[n, s_i]\right), \quad 0 \le |n| < L, \tag{51}$$

and due to wide sense stationarity,

$$E\left(\hat{r}_{XX}[n]\right) = E\left(\hat{r}_{XX}[n, s_1]\right). \tag{52}$$

From (9) and (4), for the biased estimator we have

$$\begin{aligned} E\left(\hat{r}_{XX}[n, s_1]\right) &= \frac{1}{L} \sum_{l=0}^{L-|n|-1} E\left(Y_{s_1}[l] Y_{s_1}[l+|n|]\right) \quad (53)\\ &= \frac{1}{L} \sum_{l=0}^{L-|n|-1} E\left(X_{s_1}[l] X_{s_1}[l+|n|]\right) \\ &\quad \times w[l]w[l+n] \\ &= r_{xx}[n] \left(\frac{1}{L} \sum_{l=0}^{L-|n|-1} w[l]w[l+n] \right). \quad (54) \end{aligned}$$

Therefore, this equality can be written in the form of (48) where

$$\alpha[n] = \frac{1}{L} \sum_{l=0}^{L-|n|-1} w[l]w[l+n]. \quad (55)$$

Similarly for the unbiased estimator, we have

$$E\left(\hat{r}^m_{XX}[n]\right) = r_{xx}[n] \left(\frac{g_m[n]}{L-|n|} \sum_{l=0}^{L-|n|-1} w[l]w[l+n] \right). \quad (56)$$

In this case $\alpha[n]$ in (48) is

$$\alpha[n] = \frac{1}{L-|n|} \sum_{l=0}^{L-|n|-1} w[l]w[l+n] \quad (57)$$

APPENDIX II
CALCULATION OF $\mathrm{var}(\hat{r}_{XX}[n])$

In the following calculation we show how the variance is a function of the autocorrelation. From (12), we have

$$\mathrm{var}\left(\hat{r}^m_{XX}[n]\right) = |g_m[n]|^2 \mathrm{var}\left(\hat{r}_{XX}[n]\right) \quad (58)$$

and $\mathrm{var}(\hat{r}_{XX}[n])$ is

$$\begin{aligned} \mathrm{var}\left(\hat{r}_{XX}[n]\right) &= \mathrm{var}\left(\frac{1}{S} \sum_{i=1}^{S} \hat{r}_{XX}[n, s_i] \right) \quad (59)\\ &= \frac{1}{S^2} \sum_{i=1}^{S} \mathrm{var}\left(\hat{r}_{XX}[n, s_i]\right) + J[n] \quad (60) \end{aligned}$$

where $J[n]$ is the covariance terms between the segments,

$$J[n] = \frac{2}{S^2} \sum_{i=1, j=1, i\neq j}^{S} \mathrm{Cov}\left(\hat{r}_{XX}[n, s_i], \hat{r}_{XX}[n, s_j]\right) \quad (61)$$

and $\hat{r}_{XX}[n, s_i]$, defined in (8), is the autocorrelation estimate of segment s_i at point n. In order to calculate the desired variance in (60), it is enough to calculate the covariance between $\hat{r}_{XX}[n_1, s_{k_1}]$ and $\hat{r}_{XX}[n_2, s_{k_2}]$. The first term of (60) is the covariance when $n_1 = n_2 = n$ for each segment, $k_1 = k_2$, and the second term $J(n)$ is the covariance when $n_1 = n_2 = n$ and the segments are different, $k_1 \neq k_2$. The desired covariance is

$$\begin{aligned} &\mathrm{Cov}\left(\hat{r}_{XX}[n_1, s_{k_1}], \hat{r}_{XX}[n_2, s_{k_2}]\right) \\ &= E\left(\hat{r}_{XX}[n_1, s_{k_1}] \hat{r}_{XX}[n_2, s_{k_2}]\right) \\ &\quad - E\left(\hat{r}_{XX}[n_1, s_{k_1}]\right) E\left(\hat{r}_{XX}[n_2, s_{k_2}]\right). \quad (62) \end{aligned}$$

The value of the second term is provided in (54) and the first term is calculated as follows[13]:

$$\begin{aligned} &E\left(\hat{r}_{XX}[n_1, s_{k_1}] \hat{r}_{XX}[n_2, s_{k_2}]\right) \quad (63)\\ &= E\Bigg(A_1 \sum_{i=0}^{L-1-|n_1|} X_{s_{k_1}}[i] w[i] X_{s_{k_1}}[i+|n_1|] w[i+|n_1|] \\ &\qquad \times A_2 \sum_{j=0}^{L-1-|n_2|} X_{s_{k_2}}[j] w[j] X_{s_{k_2}}[j+|n_2|] \\ &\qquad \times w[j+|n_2|] \Bigg) \\ &= A_1 A_2 \sum_{i=0}^{L-1-|n_1|} \sum_{j=0}^{L-1-|n_2|} w[i] w[i+|n_1|] w[j] w[j+|n_2|] \\ &\quad \times E\left(X[i+(k_1-1)D] X[i+|n_1|+(k_1-1)D] X\right. \\ &\qquad \left. \times [j+(k_2-1)D] X[j+|n_2|+(k_2-1)D]\right) \quad (64)\\ &= A_1 A_2 \sum_{i=0}^{L-1-|n_1|} \sum_{j=0}^{L-1-|n_2|} w[i] w[i+|n_1|] w[j] w[j+|n_2|] \\ &\quad \times \left(r_{xx}[|n_1|] r_{xx}[|n_2|] + r_{xx}[i-j-D(k_1-k_2)] r_{xx}\right. \\ &\qquad \times [i-j+|n_1|-|n_2|-D(k_1-k_2)] \\ &\qquad + r_{xx}[i-j-|n_2|-D(k_1-k_2)] r_{xx} \\ &\qquad \left. \times [i-j+|n_1|-D(k_1-k_2)]\right) \quad (65) \end{aligned}$$

where from (64) to (65) the Gaussian moment factoring theorem is used [6], and A_1 and A_2 for the unbiased estimator are

$$A_1 = \frac{1}{L-|n_1|}, \quad A_2 = \frac{1}{L-|n_2|} \quad (66)$$

and for the biased estimator is

$$A_1 = A_2 = \frac{1}{L}. \quad (67)$$

Variance in Blackman-Tukey: In the Blackman Tukey approach there is only one segment ($L = N$). Therefore, the cross-terms in (60), $J[n]$, are zero. Also the time window is rectangular. Hence, $E(\hat{r}^2_{XX}[n])$ is calculated by setting $k_1 = k_2 = 1$ and $n_1 = n_2 = n$ in (63)

$$\begin{aligned} &E\left(\hat{r}^2_{XX}[n]\right) \\ &= \frac{(N-1-|n|)^2}{N^2} r^2_{xx}[n] \\ &\quad + \frac{1}{N^2} \sum_{i,j=0}^{N-1-|n|} \left(r^2_{xx}[i-j] + r_{xx}[i-j+|n|] r_{xx}\right. \\ &\qquad \left. \times [i-j-|n|]\right) \quad (68) \end{aligned}$$

[13] This covariance calculation is the modified version of covariance calculation in [9], where the exact effect of the finiteness of the available data on the boundaries of the summations is ignored.

$$= \frac{(N-1-|n|)^2}{N^2} r_{xx}^2[n] + \frac{1}{N^2} \sum_{j=0}^{N-1-|n|} \sum_{k=-j}^{N-1-|n|-j} \left(r_{xx}^2[k] + r_{xx}[k+|n|] \times r_{xx}[k-|n|]\right) \quad (69)$$

$$= \frac{(N-1-|n|)^2}{N^2} r_{xx}^2[n] + \sum_{k=0}^{N-1-|n|} \frac{N-1-|n|-k}{N^2} \times \left(r_{xx}^2[k] + r_{xx}[k+|n|]\, r_{xx}[k-|n|]\right). \quad (70)$$

From (68) to (69), index i is replaced by new index $k = i - j$. On the other hand, using (54) to find $(E(\hat{r}_{XX}[n]))^2$ we have

$$(E(\hat{r}_{XX}[n]))^2 = \frac{(N-1-|n|)^2}{N^2} r_{xx}^2[n] \quad (71)$$

Therefore, the variance of the tapered version Blackman-Tukey approach is

$$\begin{aligned} \operatorname{var}(\hat{r}_{XX}^m[n]) &= |g_m[n]|^2 \times \left(E\left(\hat{r}_{XX}^2[n]\right) - \frac{(N-1-|n|)^2}{N^2} r_{xx}^2[n]\right) \quad (72) \\ &= |g_m[n]|^2 \sum_{k=0}^{N-1-|n|} \frac{N-1-|n|-k}{N^2} \times \left(r_{xx}^2[k] + r_{xx}[k+|n|]\, r_{xx}[k-|n|]\right) \quad (73) \end{aligned}$$

where based on (58), the extra term $|g_m[n]|^2$ has the effects of tapering.

Appendix III
Mean and Variance of the AMSE estimate in Blackman-Tukey

The AMSE in (25) is a linear combination of $r_{xx}^2[n]$ and $\operatorname{var}(\hat{r}_{XX}[n])$. It is known that the biased estimate of the autocorrelation is consistent. We can confirm this fact by checking the exact behavior of the variance provided in (73):

$$\lim_{N\to\infty} \operatorname{var}(\hat{r}_{XX}[n]) = 0. \quad (74)$$

Therefore, for the AMSE, (25), we have

$$\lim_{N\to\infty} \text{AMSE} = \lim_{N\to\infty} \sum_{|n|=m}^{N-1} r_{xx}^2[n] + \sum_{n=-m+1}^{m-1} r_{xx}^2[n] \lim_{N\to\infty} (1 - g_m[n]\alpha[n])^2 \quad (75)$$

and the AMSE estimate in (27) is

$$E\left(\widehat{\text{AMSE}}[m]\right) = \sum_{|n|=m}^{N-1} E\left(\hat{r}_{XX}^2[n]\right) + \sum_{n=-m+1}^{m-1} (1 - g_m[n]\alpha[n])^2 E\left(\hat{r}_{XX}^2[n]\right) + E\left(\widehat{\operatorname{var}(\hat{r}_{XX}[n])}\right). \quad (76)$$

Using (70), we have

$$\begin{aligned} \lim_{N\to\infty} E\left(\hat{r}_{XX}^2[n]\right) &= \lim_{N\to\infty} \frac{(N-1-|n|)^2}{N^2} r_{xx}^2[n] \quad (77) \\ &= r_{xx}^2[n] \quad .(78) \end{aligned}$$

On the other hand, $\widehat{\operatorname{var}(\hat{r}_{XX}[n])}$ is the variance in (73) when the autocorrelations are replaced by their estimates:

$$\widehat{\operatorname{var}(\hat{r}_{XX}[n])} = |g_m[n]|^2 \sum_{k=0}^{N-1-|n|} \frac{N-1-|n|-k}{N^2} \times \left(\hat{r}_{xx}^2[k] + \hat{r}_{xx}[k+|n|]\, \hat{r}_{xx}[k-|n|]\right) \quad (79)$$

and for its expected value we have

$$E\left(\widehat{\operatorname{var}(\hat{r}_{XX}[n])}\right) = |g_m[n]|^2 \sum_{k=0}^{N-1-|n|} \frac{N-1-|n|-k}{N^2} \times \left(E\left(\hat{r}_{XX}^2[k] + \hat{r}_{XX}[k+|n|]\, \hat{r}_{XX}[k-|n|]\right)\right). \quad (80)$$

Since $E(\hat{r}_{XX}^2[k] + \hat{r}_{XX}[k+|n|]\hat{r}_{XX}[k-|n|])$ is a finite value for each k, in the limit we have

$$\lim_{N\to\infty} E\left(\widehat{\operatorname{var}(\hat{r}_{XX}[n])}\right) = 0. \quad (81)$$

Therefore, from (81) and (78), we conclude that the limit of the AMSE estimate in (76) is the same as the limit of AMSE (75)

$$\lim_{N\to\infty} \widehat{\text{AMSE}}[m] = \lim_{N\to\infty} \text{AMSE}[m] \quad (82)$$

and the estimator is unbiased in the limit.

Variance of the AMSE Estimate: The structure of the AMSE estimate in (27) and the structure of the autocorrelation variance estimate in (79) result in the following AMSE variance:

$$\begin{aligned} \operatorname{var}(\text{AMSE[m]}) &= E\left(\text{AMSE[m]} - E(\text{AMSE[m]})\right)^2 \\ &= E\Bigg(\left(\hat{r}_{XX}^2[n] - E\left(\hat{r}_{XX}^2[n]\right)\right) + \sum_{n=-m+1}^{m-1} (1 - g_m[n]\alpha[n])^2 \left(\hat{r}_{XX}^2[n] - E\left(\hat{r}_{XX}^2[n]\right)\right) \\ &\quad + \sum_{n=-m+1}^{m-1} \left(\widehat{\operatorname{var}(\hat{r}_{XX}[n])} - E\left(\widehat{\operatorname{var}(\hat{r}_{XX}[n])}\right)\right)\Bigg)^2. \quad (83) \end{aligned}$$

Any term of the AMSE variance with $\mathrm{var}(\widehat{\hat{r}_{XX}[n]})$ and its expected value is in form of $(1/N)\gamma[n,N]$ where $\gamma[n,N]$ is a finite value. This is due to the structure of $\mathrm{var}(\widehat{\hat{r}_{XX}[n]})$ and its expected value in (79) and (80). The rest of the terms in the AMSE variance are in the form of $E((\hat{r}^2_{XX}[i]-E(\hat{r}^2_{XX}[i]))(\hat{r}^2_{XX}[j]-E(\hat{r}^2_{XX}[j])))$ for a range of i,j. To show that the AMSE variance in the limit is zero, it is enough to show that these terms are zero in the limit. Since $E(ZQ)\le\sqrt{E(|Z|^2)E(|Q|^2)}$, it is enough to show that

$$\mathrm{var}\left(\hat{r}^2_{XX}[n]\right)=E\left(\left(\hat{r}^2_{XX}[n]-E\left(\hat{r}^2_{XX}[n]\right)\right)^2\right) \quad (84)$$
$$=E\left(\hat{r}^4_{XX}[n]\right)-\left(E\left(\hat{r}^2_{XX}[n]\right)\right)^2 \quad (85)$$

is zero in the limit. From (70), we have

$$\lim_{N\to\infty}\left(E\left(\hat{r}^2_{XX}[n]\right)\right)^2=r^4_{xx}[n]. \quad (86)$$

For $E(\hat{r}^4_{XX}[n])$, we have

$$E\left(\hat{r}^4_{XX}[n]\right) \quad (87)$$
$$=\frac{1}{N^4}E\left(\sum_{i=0}^{N-1-|n|}X[i]X(i+n]\sum_{j=0}^{N-1-|n|}X[j]X[j+n]\times\sum_{k=0}^{N-1-|n|}X[k]X(k+n]\times\sum_{v=0}^{N-1-|n|}X[v]X[v+n]\right) \quad (88)$$
$$=\frac{1}{N^4}\sum_{i=0}^{N-1-|n|}\sum_{j=0}^{N-1-|n|}\sum_{k=0}^{N-1-|n|}\sum_{v=0}^{N-1-|n|}E\left(X[i]X(i+n]X[j]X[j+n]\times X[k]X(k+n]X[v]X[v+n]\right) \quad (89)$$
$$=\frac{1}{N^4}\sum_{i=0}^{N-1-|n|}\sum_{j=0}^{N-1-|n|}\sum_{k=0}^{N-1-|n|}\sum_{v=0}^{N-1-|n|}(105\text{ terms of form }E\left(X[\cdot]X[\cdot]\right)\times E\left(X[\cdot]X[\cdot]\right)E\left(X[\cdot]X[\cdot]\right)E\left(X[\cdot]X[\cdot]\right) \quad (90)$$
$$=\frac{(N-1-|n|)^4}{N^4}r^4_{xx}[n]+\frac{(N-1-|n|)^3}{N^4}\beta[n,N] \quad (91)$$

where $\beta[n,N]$ is such that $\lim_{N\to\infty}\beta[n,N]=B_0[n]$ is a finite value. Note that from (89) to (90) the Gaussian moments of order eight is used.[14] As a result, $E(\hat{r}^4_{XX}[n])$ in the limit is

$$\lim_{N\to\infty}E\left(\hat{r}^4_{XX}[n]\right)=r^4_{xx}[n]. \quad (92)$$

From (92) and (86), we have

$$\lim_{N\to\infty}\mathrm{var}\left(\mathrm{AMSE}[m]\right)=0 \quad (93)$$

and consistency of the AMSE estimate is proven.

[14]For the 8th order moment of a zero mean Gaussian WSS random process, we have $E(X_1\cdots X_8)=E(X_1X_2)E(X_3X4)E(X_5X_6)E(X_7X_8)+\cdots$ which includes all possible $7\times 5\times 3=105$ combinations of products of four autocorrelations among X_1 up to X_8.

Appendix IV
PSD Mean and Variance With Smoothing Window

By generalizing the expected value and variance of the Blackman-Tukey in [14], we have

$$E\left(\hat{P}_{XX}(e^{j\omega})\right)=\frac{1}{S}\sum_{i=1}^{S}E\left(\hat{P}^{(s_i)}_{XX}(e^{j\omega})\right) \quad (94)$$
$$\approx\frac{1}{S}\sum_{i=1}^{S}G(e^{j\omega})*P_{XX}(e^{j\omega}) \quad (95)$$
$$\approx G(e^{j\omega})*P_{XX}(e^{j\omega}) \quad (96)$$

and

$$\mathrm{var}\left(\hat{P}_{XX}(e^{j\omega})\right)\approx\frac{1}{S^2}\sum_{i=1}^{S}\mathrm{var}\left(\hat{P}^{(s_i)}_{XX}(e^{j\omega})\right) \quad (97)$$
$$\approx\frac{1}{S^2}\sum_{i=1}^{S}P_{XX}(e^{j\omega})\left[\frac{1}{L}\sum_{n=-m+1}^{m-1}g^2_m[n]\right]$$
$$\approx\frac{1}{S}P_{XX}(e^{j\omega})\left[\frac{1}{L}\sum_{n=-m+1}^{m-1}g^2_m[n]\right] \quad (98)$$
$$\approx P_{XX}(e^{j\omega})\left[\frac{1}{N}\sum_{n=-m+1}^{m-1}g^2_m[n]\right] \quad (99)$$

as $SL=N$.

Acknowledgment

The authors would like to thank Prof. J. G. Proakis and the reviewers for their constructive comments.

References

[1] T. W. Anderson, *The Statistical Analysis of Time Series*. New York: Wiley, 1971.
[2] S. Beheshti and M. Ravan, "Adaptive windowing in nonparametric power spectral density estimation," in *Proc. 21st IEEE Canad. Conf. Electr. Comput. Eng. (CCECE)*, Apr. 2008, pp. 1183–1186.
[3] R. E. Bekka and D. Chikouche, "Effect of the window length on the EMG spectral estimation through the Blackman–Tukey method," in *Proc. 7th Int. Symp. Signal Process. Its Appl.*, Jul. 1–4, 2003, vol. 2, pp. 17–20.
[4] J. P. Burg, "Maximum entropy spectral analysis," Ph.D. dissertation, Stanford Univ., Stanford, CA, 1975.
[5] J. C. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters," *J. Cybern.*, vol. 3, no. 3, pp. 32–57, 1973.
[6] M. H. Hayes, *Statistical Digital Signal Processing and Modeling*. New York: Wiley, 1996.
[7] S. Haykin, Ed., *Proc. IEEE (Special issue on Spectral Estimation)*, vol. 70, no. 9, Sep. 1982.
[8] E. T. Jaynes, "On the rationale of maximum-entropy methods," *Proc. IEEE*, vol. 70, no. 9, pp. 939–952, Sep. 1982.
[9] G. M. Jenkins and D. G. Watts, *Spectral Analysis and Its Applications*. San Francisco, CA: Holden-Day, 1968.
[10] S. M. Kay and S. L. Marple, "Spectrum analysis-A modern perspective," *Proc. IEEE*, vol. 69, no. 11, pp. 1380–1419, Nov. 1981.
[11] S. M. Kay, "Noise compensation for autoregressive spectral estimates," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-28, no. 3, pp. 292–303, 1980.
[12] H. Krim, D. Tucker, S. Mallat, and D. Donoho, "On denoising and best signal representation," *IEEE Trans. Inf. Theory*, vol. 45, no. 7, pp. 2225–2238, Nov. 1999.

[13] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.

[14] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing, Principles. Algorithms, and Applications*. Englewood Cliffs, NJ: Prentice-Hall, 2002.

[15] P. Stoica and R. L. Moses, *Spectral Analysis of Signals*. Englewood Cliffs, NJ: Prentice-Hall, 2005.

[16] P. Stoica and N. Sandgren, "Smoothed nonparametric spectral estimation via cepsturm thresholding—Introduction of a method for smoothed nonparametric spectral estimation," *IEEE Signal Process. Mag.*, vol. 23, no. 6, pp. 34–45, Nov. 2006.

[17] D. J. Thomson, "Spectrum estimation and harmonic analysis," *Proc. IEEE*, vol. 70, pp. 1055–1096, 1982.

[18] A. T. Walden, D. B. Percival, and E. J. McCoy, "Spectrum estimation by wavelet thresholding of multitaper estimators," *IEEE Trans. Signal Process.*, vol. 46, no. 12, pp. 3153–3165, Dec. 1998.

[19] F. Yao and T. C. M. Lee, "Spectral density estimation using sharpened periodograms," *IEEE Trans. Signal Process.*, vol. 55, no. 9, pp. 4711–4716, Sep. 2007.

Soosan Beheshti (M'03–SM'06) received the B.S. degree from Isfahan University of Technology, Isfahan, Iran, and the M.S. and Ph.D. degrees from the Massachusetts Institute of Technology (MIT), Cambridge, in 1996 and 2002, respectively, all in electrical engineering.

From September 2002 to June 2005, she was a Postdoctoral Associate and a Lecturer at MIT. Since July 2005, she has been with the Department of Electrical and Computer Engineering, Ryerson University, Toronto, ON, Canada, where she is currently an Assistant Professor and Director of the Signal and Information Processing (SIP) Laboratory. Her research interests include statistical signal processing, hyperspectral imaging, and system dynamics and modeling.

Maryam Ravan (M'10) received the Ph.D. degree from Amirkabir University of Technology, Tehran, Iran, in 2007.

From May 2007 to December 2007, she was a Postdoctoral Fellow with the Department of Electrical and Computer Engineering, University of Toronto, ON, Canada, and a Research Assistant with the Department of Electrical and Computer Engineering, Ryerson University, Canada. From January 2008 to August 2009, she was a Postdoctoral Fellow with the Department of Electrical and Computer Engineering, McMaster University, Montreal, QC, Canada. Since August 2009, she has been a Postdoctoral Fellow with the Department of Electrical Engineering, University of Toronto, and a Lecturer with the School of Computational Engineering and Science, McMaster University. Her research interests include biomedical signal and image processing, neural and wavelet networks, machine learning, optimization techniques, MIMO radar systems and space time adaptive processing, power spectrum estimation, and nondestructive testing. She has authored and coauthored over 30 journal and conference papers and a book chapter.

Dr. Ravan is the recipient of the NSERC Engage Grant program with Raytheon Canada Limited, Waterloo, ON, Canada.

James P. Reilly (S'76–M'80) received the B.A.Sc. degree in electrical engineering from the University of Waterloo, ON, Canada, in 1973 and the M.Eng and Ph.D. degrees from McMaster University, Hamilton, ON, Canada, in 1977 and 1980, respectively, both in electrical engineering.

He worked from 1973 to 1975 in the CATV industry designing RF distribution equipment, and from 1980 to 1985 at then Bell-Northern Research, Ottawa, Canada, where he was involved in the development of terrestrial microwave radio and multiplexing systems for digital transmission systems. He joined the faculty at McMaster University, Montreal, QC, Canada, in 1985 as an Associate Professor and was promoted to the rank of Full Professor in 1992. His research interests are in several aspects of signal processing, specifically machine learning for biomedical applications, Bayesian methods, blind source separation, blind identification, and array signal processing.

Prof. Reilly is a registered Professional Engineer in the province of Ontario, Canada.

Laurel J. Trainor received the Ph.D. degree in psychology from the University of Toronto, ON, Canada.

She is a Professor in the Department of Psychology, Neuroscience and Behaviour at McMaster University, Montreal, QC, Canada; a Research Scientist at the Rotman Research Institute at Baycrest Hospital, Toronto, ON, Canada; and the Director of the McMaster Institute for Music and the Mind. She has published over 100 research articles and book chapters on the neuroscience of auditory development and the perception of music in journal, including *Science*, *Nature*, the *Journal of Neuroscience*, *Signal Processing*, and *Psychological Science*.

Dr. Trainor is a fellow of the America Psychological Society, and an Innovator of Distinction. She holds major grants from the Canadian Foundation for Innovation, The Canadian Institutes of Health Research, the Natural Science and Engineering Research Council of Canada, the Social Science Research Council of Canada, and the Grammy Foundation.