Research Article

Optimal Multitaper Wigner Spectrum Estimation of a Class of Locally Stationary Processes Using Hermite Functions

Maria Hansson-Sandsten (EURASIP Member)

Mathematical Statistics, Centre for Mathematical Sciences, Lund University, P.O. Box 118, 221 00 Lund, Sweden

Correspondence should be addressed to Maria Hansson-Sandsten, sandsten@maths.lth.se

This paper investigates the time-discrete multitapers that give a mean square error optimal Wigner spectrum estimate for a class of locally stationary processes (LSPs). The accuracy in the estimation of the time-variable Wigner spectrum of the LSP is evaluated and compared with other frequently used methods. The optimal multitapers are also approximated by Hermite functions, which is computationally more efficient, and the errors introduced by this approximation are studied. Additionally, the number of windows included in a multitaper spectrum estimate is often crucial and an investigation of the error caused by limiting this number is made. Finally, the same optimal set of weights can be stored and utilized for different window lengths. As a result, the optimal multitapers are shown to be well approximated by Hermite functions, and a limited number of windows can be used for a mean square error optimal spectrogram estimate.

1. Introduction

A locally stationary process (LSP) has a covariance function which is a multiplication of a covariance function of a stationary process and a time-variable function, [1]. The process is nonstationary with properties suitable for modeling measured signals that, for example, have a transient amplitude behavior, such as evoked or induced potentials arising in the electroencephalogram, [2, 3]. To statistically differentiate between responses from different types of stimuli, choosing a spectral estimator with small bias and low variance is important. Such estimators can be found but additionally important is that the estimators have discrete-time and discrete-frequency properties suitable from implementation aspects, such as few choices of parameters and computational efficiency.

The mean square error optimal kernel for the class of Gaussian harmonizable processes has been obtained by Sayeed and Jones [4], and further, the optimal time-frequency kernel for LSPs, restricted to covariance functions defined by a multiplication of two variable Gaussian functions, is obtained in [5]. The calculation of the two-dimensional convolution between the kernel and the Wigner distribution of a process realization can be simplified using kernel decomposition and calculating multitaper spectrograms, [6, 7]. The time-lag estimation kernel is rotated, and the corresponding eigenvectors and eigenvalues are calculated. The estimate of the Wigner spectrum is given as the weighted sum of the spectrograms of the data with the different eigenvectors as sliding windows and the eigenvalues as weights, [8]. Different approaches to approximate a time-varying spectrum with a few windowed spectrograms have been taken, for example, [9–15]. The sampling properties related to time-frequency analysis are well covered in [16, 17], and differences in the time-discrete case compared to the time-continuous case are considered, for example, in [18, 19].

In this paper, the time-discrete multitapers corresponding to the mean square error optimal time-frequency kernel for a class of LSPs are computed and the performance of the resulting estimator is compared to other well-known algorithms. The approximation with Hermite functions are also evaluated in the discrete-time case and appropriate choices of the scaling of the Hermite functions are given for different window lengths. A reasonable number of computed multitaper spectrograms should be averaged for the final Wigner spectrum estimate, and an evaluation of how this
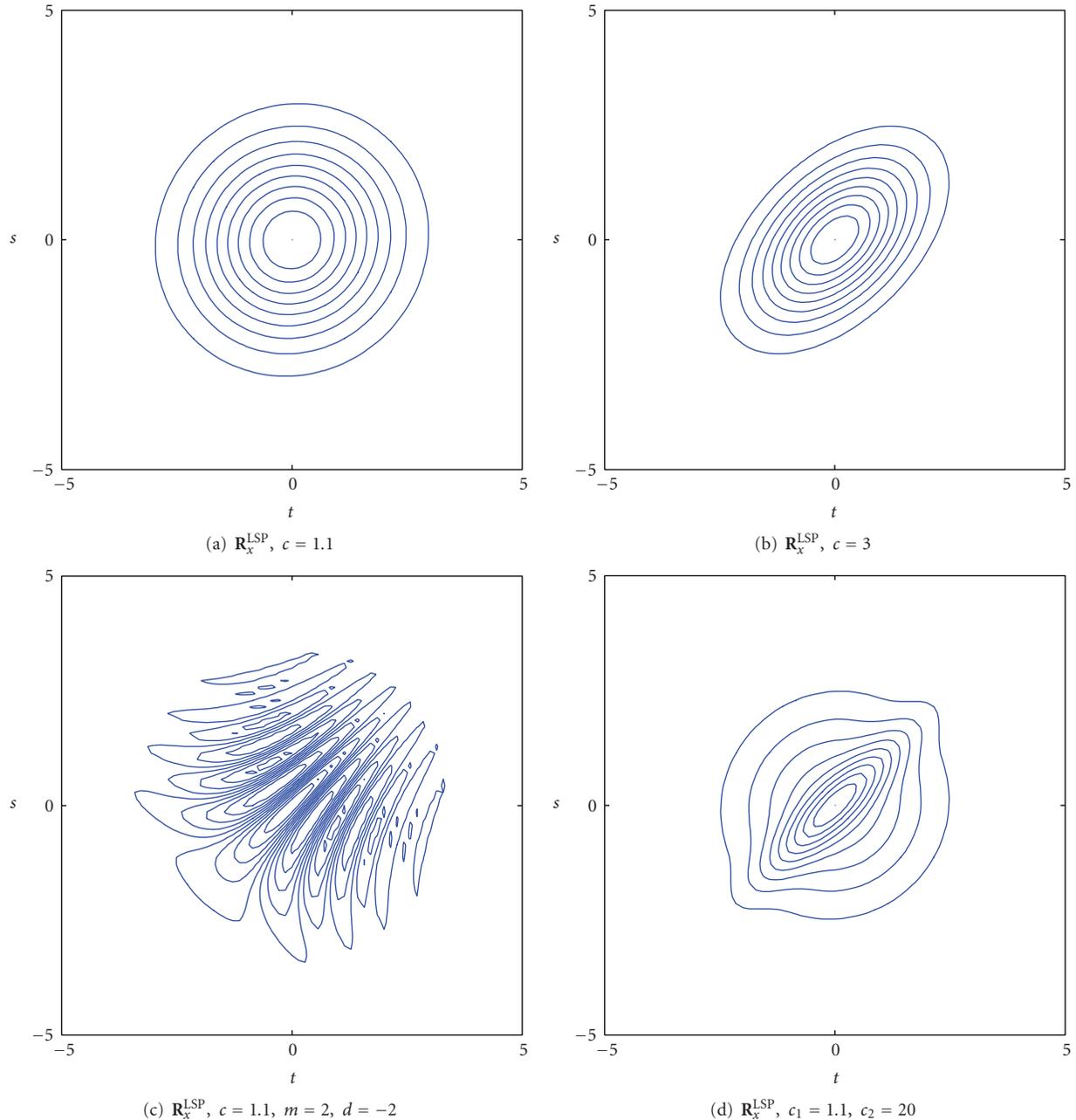
(a) $\mathbf{R}_x^{\mathrm{LSP}}$, $c = 1.1$

(b) $\mathbf{R}_x^{\mathrm{LSP}}$, $c = 3$

(c) $\mathbf{R}_x^{\mathrm{LSP}}$, $c = 1.1$, $m = 2$, $d = -2$

(d) $\mathbf{R}_x^{\mathrm{LSP}}$, $c_1 = 1.1$, $c_2 = 20$

FIGURE 1: Examples of covariance matrices $\mathbf{R}_x$ of LSP for different parameter values with $N = 256$ and $F_s = 7$: (a) LSP: $c = 1.1$, (b) LSP: $c = 3$, (c) LSCP: $c = 1.1$, $m = 2$, $d = -2$ (real-valued part only), and (d) MLSP: $c_1 = 1.1$, $c_2 = 20$.

number could be reduced without degraded performance is done. Many applications where such an algorithm could be useful for implementation can be found. In the present study, examples are shown of the electrophysiological correlates of toddlers' watching of video clips displaying different types of nonlinguistic communication between two actors.

The paper is organized as follows, Section 2 presents examples of the class of LSPs used in the paper. Section 3 summarizes the estimation of the optimal kernel approach and the corresponding multitapers and gives some examples. In Section 4, a comparison of several algorithms for estimation of locally stationary spectra is given. Section 5 evaluates the approximation using Hermite functions, and the reduction of the number of averaged multitaper spectrograms is investigated in Section 6. In Section 7, the electroencephalogram data examples are shown, Section 8 concludes the paper.

2. Locally Stationary Processes

A zero-mean second-order continuous-time locally stationary process, $\{x(t)\}$ has, per definition in [1], a covariance
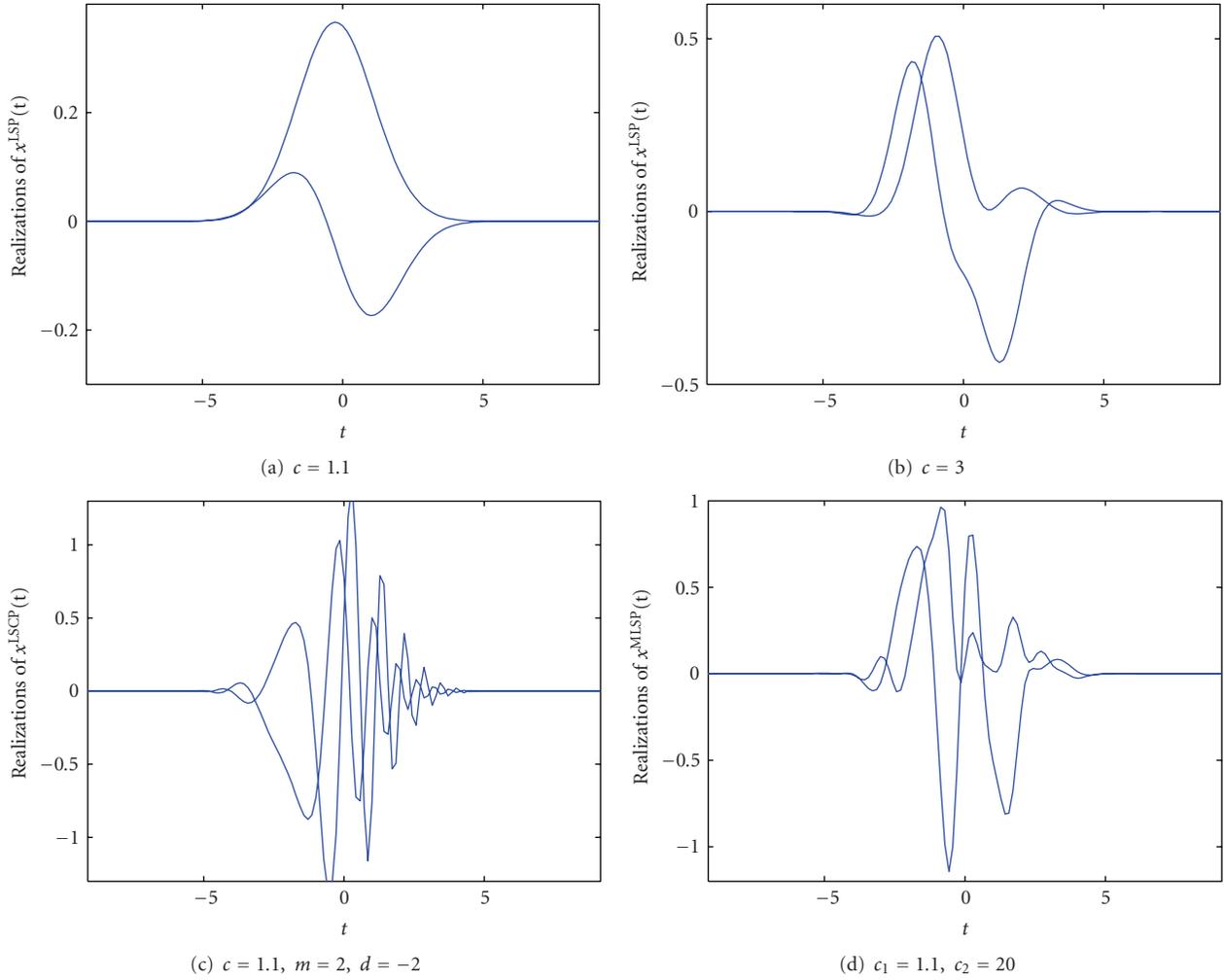
FIGURE 2: Process realizations corresponding to the covariance function $r_x(t,s)$ in Figure 1 for $N = 256$ and $F_s = 7$: (a) LSP: $c = 1.1$, (b) LSP: $c = 3$, (c) LSCP: $c = 1.1$, $m = 2$, $d = -2$ (real-valued part only), and (d) MLSP: $c_1 = 1.1$, $c_2 = 20$.

function determined by two functions according to,

$$r_x^{\text{LSP}}(t,s) = q\left(\frac{t+s}{2}\right) \cdot r(t-s). \qquad (1)$$

In this paper, $q(\tau) = e^{-\tau^2/2}$, is a fixed Gaussian function with a constant sign which we assume to be nonnegative, [5] and the function $r(\tau) = e^{(-c/4)\tau^2/2}$, with $c \geq 1$. When $c \gg 4$, $r(\tau)$ decreases quicker than $q(\tau)$, and we approach a stationary process as $c \to \infty$. The opposite extreme with $c = 1$ gives maximal nonstationarity. A *locally stationary chirp process*, (LSCP), [5], is defined

$$r_x^{\text{LSCP}}(t,s) = q\left(\frac{t+s}{2}\right) \cdot r(t-s)e^{im(t-s)((t+s)/2-d)}, \qquad (2)$$

where m determines the chirp frequency and d the start of the chirp frequency. The definition of LSP is also extended to a sum of locally stationary processes,

$$r_x^{\text{MLSP}}(t,s) = \sum_{j=1}^{J} q\left(\frac{t+s}{2}\right) \cdot r_j(t-s), \qquad (3)$$

a so-called *multicomponent locally stationary process* (MLSP) where the variable Gaussian function $r_j(\tau) = e^{(-c_j/4)\tau^2/2}$ has different c_j for $j = 1, \ldots, J$, [5].

The sampled versions of these definitions are given with $t = s = n/F_s$, $n = -N/2, \ldots, N/2 - 1$, where F_s is chosen for the Gaussian functions to become approximately zero for $t = \pm N/2F_s$. The $N \times N$ covariance matrix $\mathbf{R}_x(t,s)$ is computed, and in Figure 1 some examples of the shape of this matrix for different parameter values are depicted for $N = 256$ and $F_s = 7$.

To illustrate the processes for different parameters, a number of realizations $\mathbf{x} = [\hat{x}(0) \ \ldots \ \hat{x}(N-1)]^T$, are simulated from

$$\mathbf{x} = \mathbf{Hb}, \qquad (4)$$

where $\mathbf{b} = [\hat{b}(0) \ \ldots \ \hat{b}(N-1)]^T$ is a sequence of realizations from the a real-valued white Gaussian zero mean stochastic

process with variance one. The process-generating matrix $\mathbf{H}$ is related to the covariance matrix as

$$\mathbf{R}_x = E[\mathbf{x}\mathbf{x}^*] = \mathbf{H}E[\mathbf{b}\mathbf{b}^*]\mathbf{H}^* = \mathbf{H}\mathbf{H}^*. \tag{5}$$

In Figure 2, examples of the base-band shapes of such processes are shown. Important to remember is that these processes can be translated in frequency as well as in time, and still the Cohen's class estimators will give the same time- and frequency-translated result, [8]. For a small value of $c = 1.1$, the realizations are narrow banded, Figure 2(a), but for a larger c, the realizations are more wide banded with a more irregular form, Figure 2(b). In Figure 2(c), the parameters m and d are chosen for an increasing chirp frequency, but the principal shape is the same as in Figure 2(a) as $c = 1.1$ (note that only the real part of these process realizations are shown). The realizations in Figure 2(d) have the covariance matrix of (3) with $J = 2$, $c_1 = 1.1$, and $c_2 = 20$, giving realizations which are a combination of a narrow-banded and wide-banded shapes.

3. Optimal Kernels and Multitapers

The optimal ambiguity domain kernel in the mean square error sense for zero-mean Gaussian processes was estimated in [4], and based on this, the corresponding kernels for the processes defined in (1), (2), and (3) were derived in [5] as

$$\phi_{\text{opt}}^{\text{LSP}}(\nu, \tau) = \frac{1}{2\pi} \frac{1}{1 + c^{-1/2}e^{(1-1/c)(2\pi\nu)^2 + ((c-1)/4)\tau^2}},$$

$$\phi_{\text{opt}}^{\text{LSCP}}(\nu, \tau) = \frac{1}{2\pi} \frac{1}{1 + c^{-1/2}e^{(1-1/c)(2\pi\nu - m\tau)^2 + ((c-1)/4)\ \tau^2}},$$

$$\phi_{\text{opt}}^{\text{MLSP}}(\nu, \tau)$$

$$= \frac{1}{2\pi}\left(1 \bigg/ \left(1 + \frac{\sum_{i,j}\sqrt{2/(c_i+c_j)}\,e^{(-2/(c_i+c_j))(2\pi\nu)^2 - \tau^2/4}}{\left(\sum_i e^{-(2\pi\nu)^2/2 - ((c_i/8)\tau^2)}\right)^2}\right)\right). \tag{6}$$

The optimal ambiguity domain kernels corresponding to the covariance functions and processes of Figures 1 and 2 are computed and plotted in Figure 3.

Instead of calculating the time-frequency estimate using the ambiguity domain kernel, the calculations can be simplified by using a multitaper spectrogram, [8], where Cohen's class is written as

$$\widehat{W}_C(t, f)$$

$$= \iint A_x(\nu, \tau)\phi_{\text{opt}}(\nu, \tau)e^{-i2\pi(\tau f - t\nu)}d\tau\, d\nu \tag{7}$$

$$= \iint x\left(u + \frac{\tau}{2}\right)x^*\left(u - \frac{\tau}{2}\right)\rho_{\text{opt}}(t - u, \tau)e^{-i2\pi f\tau}du\, d\tau.$$

With $t = (t_1 + t_2)/2$ and $\tau = t_1 - t_2$,

$$\widehat{W}_C(t, f)$$

$$= \iint x(t_1)x^*(t_2)\rho_{\text{opt}}\left(t - \frac{t_1 + t_2}{2}, t_1 - t_2\right)e^{-i2\pi f(t_1 - t_2)}dt_1 dt_2$$

$$= \iint x(t_1)x^*(t_2)\rho_{\text{opt}}^{\text{rot}}(t - t_1, t - t_2)e^{-i2\pi f t_1}e^{i2\pi f t_2}dt_1 dt_2, \tag{8}$$

where

$$\rho_{\text{opt}}^{\text{rot}}(t_1, t_2) = \rho\left(\frac{t_1 + t_2}{2}, t_1 - t_2\right). \tag{9}$$

If the kernel $\rho_{\text{opt}}^{\text{rot}}(t_1, t_2)$ satisfies the Hermitian property $\rho_{\text{opt}}^{\text{rot}}(t_1, t_2) = (\rho_{\text{opt}}^{\text{rot}}(t_2, t_1))^*$, then solving the integral

$$\int \rho_{\text{opt}}^{\text{rot}}(t_1, t_2)q(t_1)dt_1 = \lambda q(t_2) \tag{10}$$

results in eigenvalues λ_k and eigenfunctions q_k which form a complete set. The time-lag kernel can be expressed as

$$\rho_{\text{opt}}^{\text{rot}}(t_1, t_2) = \sum_{k=1}^{\infty}\lambda_k q_k(t_1)q_k^*(t_2). \tag{11}$$

Using the eigenvalues and eigenvectors, (7) is rewritten as a weighted sum of spectrograms,

$$\widehat{W}_C(t, f) = \sum_{k=1}^{\infty}\lambda_k\left|\int x(t_1)e^{-i2\pi f t_1}q_k^*(t - t_1)dt_1\right|^2. \tag{12}$$

With just a few λ_k that differ from zero, the multitaper spectrogram solution is an effective solution from implementation aspects.

Using discrete-time data, the windows and weights are found from the solution of the corresponding matrix equation

$$\mathbf{R}_{\text{opt}}\mathbf{q}_k = \lambda_k \mathbf{q}_k, \quad k = 1, \ldots, N, \tag{13}$$

where the eigenvalues are ordered according to $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$, and $\mathbf{R}_{\text{opt}}$ is the sampled matrix of size $N \times N$ corresponding to $\rho_{\text{opt}}^{\text{rot}}(t_1, t_2)$ in (9) where $t_1, t_2 = n/F_s$, $n = -N/2, \ldots, N/2 - 1$. To extract the corresponding samples of $\rho_{\text{opt}}^{\text{rot}}(t, \tau)$ in an easy way, the time-lag kernel, $\rho_{\text{opt}}(t, \tau)$ should be sampled with $2F_s$, that is, $t = n/2F_s$, with $n = -N, \ldots, N - 1$ and $\tau = n/2F_s$ with $n = -2N, \ldots, 2N - 1$ to cover the required range for t_1 and t_2. The ambiguity domain kernel $\phi_{\text{opt}}(\nu, \tau)$ sampled with $2F_s$ and the discrete fourier transform (DFT) in the first variable will give the sampled $\rho_{\text{opt}}(t, \tau)$ and the matrix $\mathbf{R}_{\text{opt}}$. Examples of the eigenvalues and eigenvectors connected to the corresponding matrices $\mathbf{R}_{\text{opt}}$ are computed for $N = 256$ and $F_s = 7$ and depicted in Figure 4 where the 10 first eigenvalues are plotted in Figure 4(a) and the three first eigenvectors are seen in Figures 4(b),4(c),4(d), and 4(e) for the respective case.

Note that the eigenvalues which are optimal for the LSP with $c = 1.1$ and LSCP with $c = 1.1$, $m = 2$, are equal,

TABLE 1: The mean square error of the LSP-optimal multitapers and weights (LSP$_{\text{opt}}$) compared to optimal choices of parameters of different methods evaluating the class of LSPs when $c = 1.1$, 3, and 20.

Method	$c = 1.1$	$c = 3$	$c = 20$
LSP$_{\text{opt}}$	19.23	11.85	3.27
WIG	19.27	14.45	7.21
C-W	19.25 ($\sigma = 38$)	12.67 ($\sigma = 1.45$)	4.37 ($\sigma = 0.15$)
HANN	20.80 ($N_w = 48$)	11.97 ($N_w = 37$)	4.39 ($N_w = 17$)
WOSA	25.59 ($N_w = 44, K = 2$)	12.69 ($N_w = 35, K = 2$)	3.39 ($N = 25, K = 3$)
TH MW	25.26 ($N_w = 54, K = 2$)	12.51 ($N_w = 42, K = 2$)	3.40 ($N_w = 29, K = 3$)
PM MW	23.07 ($N_w = 36, K = 2$)	11.94 ($N_w = 30, K = 2$)	3.46 ($N_w = 21, K = 8$)

which was shown to be true in [5]. The eigenvectors for LSP with $c = 1.1$ and $c = 3$ have a very similar shape, and it was also shown in [5] that the eigenfunctions approximate Hermite functions and differ with a specific scaling factor. These approximations will be investigated in Section 4.

4. Evaluation

Using the set of windows and weighting factors described in the previous section, a performance comparison of different kernels and multitaper methods is made. For a fair comparison, the optimal parameter values are found for all methods in all evaluations. For the Wigner spectrum (WIG), [20], the time-lag kernel of the Wigner distribution is defined as

$$\rho(t, \tau) = \delta(t), \quad (14)$$

and the Choi-Williams kernel (C-W), [21], has a time-lag kernel defined as

$$\rho(t, \tau) = \frac{\pi\sigma}{|\tau|} e^{-\pi^2 \sigma t^2 / \tau^2}, \quad (15)$$

where σ is optimized for the smallest mean square error in each case. In the computer evaluation, the C-W time-lag kernel for the discrete-time case proposed in [22], Chapter 6, page 271, is used. For the single Hanning window (HANN), the window length N_w is optimized in each case. For the Welch method (WOSA), [23], also a Hanning window is used with an overlap of 50 % between the windows, and the optimal length N_w, and number of windows K is found for each case. In the Thomson multitaper method (TH MW), the lengths N_w of the K multitapers are optimized, and the relation with the resolution bandwidth determined by $B = (K + 3)/N_w$ in each case, [24]. The peak matched multiple windows (PM MW), [25], are designed to give low correlations between subspectra when the spectrum of the random process includes peaks, that is, a spectrum corresponding to a process with large dynamics. It has been shown that this method gives a small mean square error for evoked potentials, [26] as well as for estimation of heart rate variability (HRV), [27]. The window lengths N_w as well as the number of windows K are optimized for the best result, and the corresponding resolution bandwidth is determined by $B = (K + 3)/N_w$, similarly as for TH MW in each case. The weighting factors are chosen as the eigenvalues $\alpha_k = \lambda_k$,

$k = 1, \ldots, K$, as this choice is close to the optimal weighting factors, [28]. The reason for comparing with these methods is that either they are very frequently used various applications or they are designed to be optimal for a similar type of process.

The total mean square error is defined as

$$J = \sum_\nu \sum_\tau E\left[\left| A_x(\nu, \tau) \cdot \gamma \phi_{\text{opt}}(\nu, \tau) - E[A_x(\nu, \tau)] \right|^2 \right], \quad (16)$$

where E denotes expected value. Using a sufficiently large value N and an appropriate sampling frequency F_s, the discrete-time case will produce results close to the continuous-time case. The parameter γ is used to adjust the kernel $\phi_{\text{opt}}(\nu, \tau)$ to give the smallest possible mean square error for every method and choice of parameters. The parameter γ is optimized individually for all methods and parameter cases and is found by derivation as

$$\gamma_{\text{opt}} = \frac{\sum_\nu \sum_\tau E[A_x(\nu, \tau)]^2 \phi_{\text{opt}}(\nu, \tau)}{\sum_\nu \sum_\tau E[A_x^2(\nu, \tau)] \phi_{\text{opt}}^2(\nu, \tau)}. \quad (17)$$

In the first simulation, the resulting total mean square errors (MSE) of the LSP-optimal multitapers (LSP$_{\text{opt}}$) for three different processes defined by (1), with the parameter c as $c = 1.1$, 3, and 20, ranging from nonstationary to more stationary processes, are calculated for $N = 256$ and $F_s = 7$. The results of the different methods are presented in Table 1. The optimal choice is compared with a number of other methods, also optimized for the smallest error. For C-W, the parameter σ is varied in different ranges: $\sigma = 1$–38 for $c = 1.1$, $\sigma = 0.05$–1.9 for $c = 3$, and $\sigma = 0.01$–0.38 for $c = 20$, to find the minimum value in each case. The window lengths N_w that give the optimum result for HANN, WOSA, PM MW, and TH MW are varied between $N_w = 30$ and 140. For WOSA, TH MW, and PM MW, a different number of windows ranging from $K = 2$ to 8 are also tested to find the number of windows that give a minimum error.

The results show that WIG and C-W give a mean square error very close to the optimal error, $J_{\text{opt}} = 19.23$ when $c = 1.1$ and that, out of the spectrogram methods, HANN gives the best result. For $c = 3$, WIG as well as the result of C-W are larger than the spectrogram methods where the PM MW gives an error closer to the optimal mean square error $J_{\text{opt}} = 11.85$. For $c = 20$, the result of the TH MW is closest to the optimal error $J_{\text{opt}} = 3.27$. This example show

(a) $\phi_{\mathrm{opt}}^{\mathrm{LSP}}(\nu,\tau)$, $c = 1.1$

(b) $\phi_{\mathrm{opt}}^{\mathrm{LSP}}(\nu,\tau)$, $c = 3$

(c) $\phi_{\mathrm{opt}}^{\mathrm{LSCP}}(\nu,\tau)$, $c = 1.1$, m $= 2$

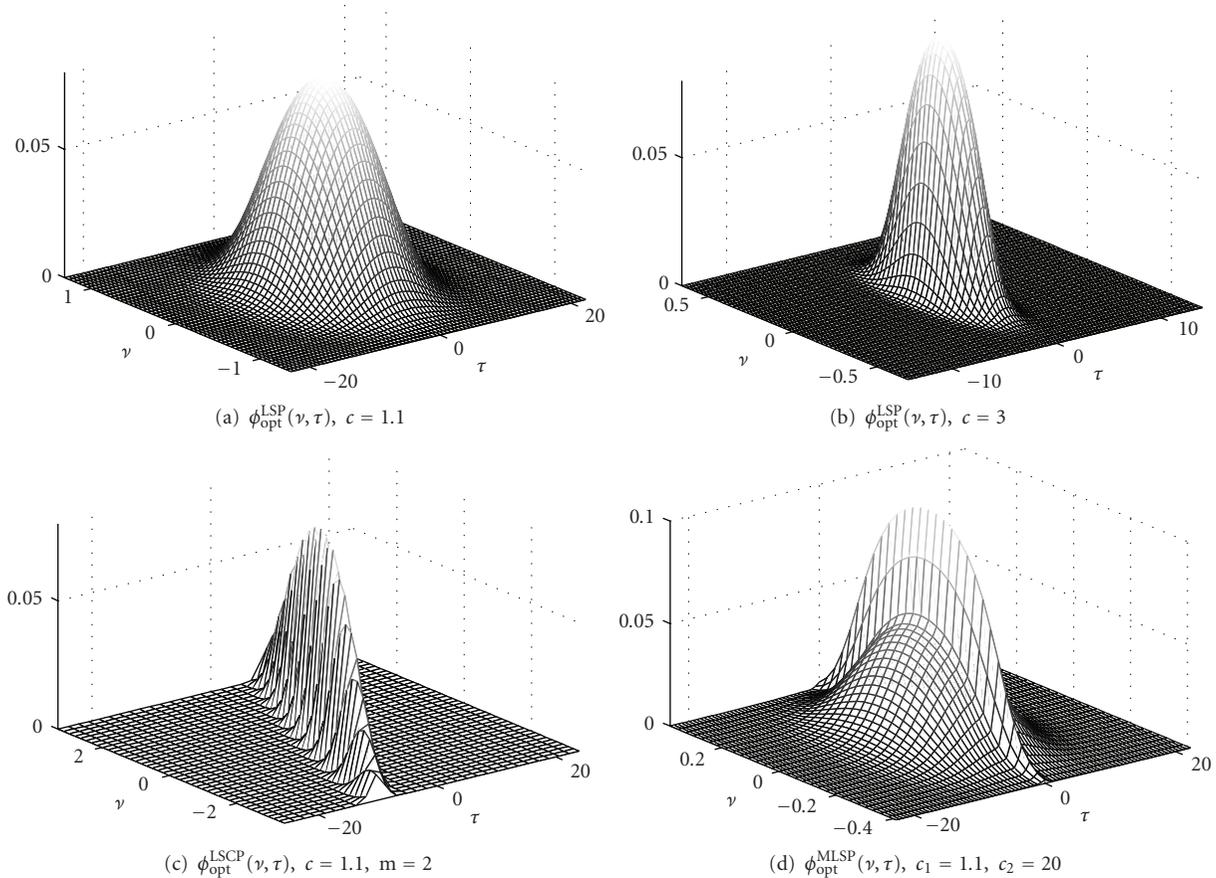(d) $\phi_{\mathrm{opt}}^{\mathrm{MLSP}}(\nu,\tau)$, $c_1 = 1.1$, $c_2 = 20$

FIGURE 3: The mean square error optimal ambiguity domain kernel corresponding to processes with covariance functions in Figure 1 for $N = 256$ and $F_s = 7$: (a) $\phi_{\mathrm{opt}}^{\mathrm{LSP}}$: $c = 1.1$, (b) $\phi_{\mathrm{opt}}^{\mathrm{LSP}}$: $c = 3$, (c) $\phi_{\mathrm{opt}}^{\mathrm{LSCP}}$: $c = 1.1$, $m = 2$, and (d) $\phi_{\mathrm{opt}}^{\mathrm{MLSP}}$: $c_1 = 1.1$, $c_2 = 20$.

TABLE 2: The mean square error of the LSCP-optimal multitapers and weights (LSCP$_{\mathrm{opt}}$) compared to optimal choices of parameters of different methods evaluating the class of LSCPs when $c = 1.1$, 3, and 20, and $m = 2$, $d = -2$.

Method	$c = 1.1$	$c = 3$	$c = 20$
LSCP$_{\mathrm{opt}}$	19.23	11.85	3.27
WIG	19.27	14.45	7.21
C-W	19.33 ($\sigma = 950$)	13.23 ($\sigma = 37.5$)	4.88 ($\sigma = 0.275$)
HANN	25.49 ($N_w = 24$)	13.35 ($N_w = 24$)	4.44 ($N_w = 17$)
WOSA	30.64 ($N_w = 21, K = 2$)	15.76 ($N_w = 21, K = 2$)	3.81 ($N_w = 21, K = 3$)
TH MW	31.34 ($N_w = 27, K = 2$)	15.94 ($N_w = 27, K = 2$)	3.82 ($N_w = 25, K = 3$)
PM MW	26.65 ($N_w = 20, K = 2$)	13.76 ($N_w = 20, K = 2$)	3.77 ($N_w = 19, K = 8$)

the great variation in the properties of the class of LSPs when c changes. When c increases from $c = 1.1$ to larger values the process varies from being narrow banded to wide banded. This is why WIG gives a closer optimum value for $c = 1.1$, but TH MW gives the MSE closer to the optimal value for $c = 20$. None of the established methods are able to give small errors in all three cases, even if all possible parameters of the methods are optimized.

The next simulation evaluates the performance for an LSCP, (2) and the corresponding optimal windows, (LSCP$_{\mathrm{opt}}$). In this case, we use $m = 2$ and $d = -2$ for the simulation of all three processes using the same c-values as

in the preceding example. The evaluation is done using the same methods with the C-W parameter varied in the range, $\sigma = 25$–950 for $c = 1.1$, $\sigma = 2.5$–95 for $c = 3$, and $\sigma = 0.025$–0.95 for $c = 20$. In Table 2, the results show that WIG and the C-W are almost as optimal as LSCP$_{\mathrm{opt}}$ for $c = 1.1$. The optimal value for $c = 3$, $J_{\mathrm{opt}} = 11.85$ is not reached by any other method. For $c = 20$, the optimal result is $J_{\mathrm{opt}} = 3.27$ and the three multitaper methods give the closest result.

The last simulation with results shown in Table 3 evaluates the performance for the optimal window and weights of MLSP, (3), and (MLSP$_{\mathrm{opt}}$), using two different processes in the sum for each of three examples, the first case with
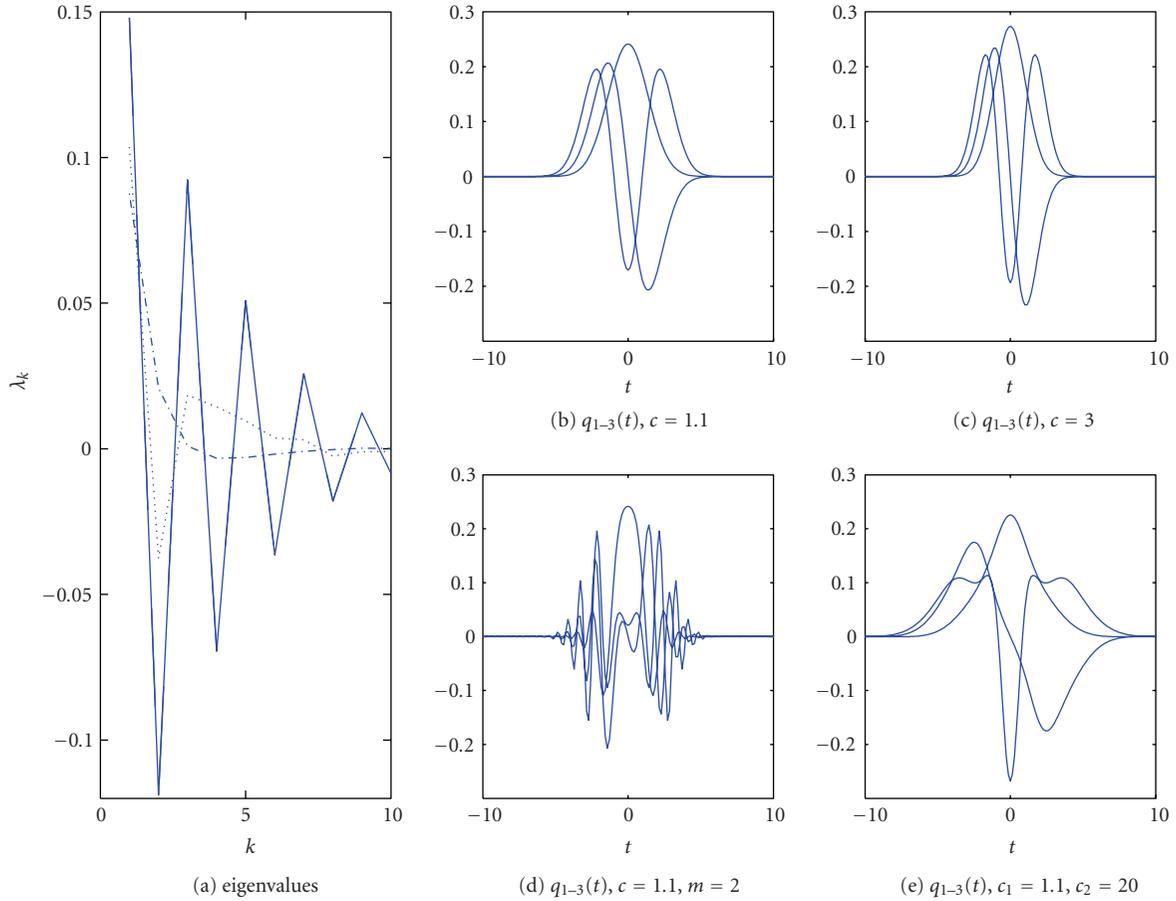
FIGURE 4: Optimal eigenvalues and eigenvectors corresponding to the ambiguity domain kernels in Figure 3 for $N = 256$ and $F_s = 7$: (a) the 10 largest eigenvalues, (solid line = LSP, $c = 1.1$ and LSCP, $c = 1.1$, $m = 2$, $d = -2$, dash-dotted line = LSP, $c = 3$, dotted line = MLSP, $c_1 = 1.1$, and $c_2 = 20$). Optimal eigenvectors: (b) $c = 1.1$; (c) $c = 3$; (d) $c = 1.1$, $m = 2$, $d = -2$ (real-valued part only); (e) $c_1 = 1.1$ and $c_2 = 20$.

TABLE 3: The mean square error of the MLSP-optimal multitapers and weights (MLSP$_{opt}$) compared to optimal choices of parameters for different methods evaluating the class of MLSPs in three cases: $c_1 = 1.1$, $c_2 = 20$; $c_1 = 2$, $c_2 = 10$; $c_1 = 4$, $c_2 = 8$.

Method	$c_1 = 1.1$, $c_2 = 20$	$c_1 = 2$, $c_2 = 10$	$c_1 = 4$, $c_2 = 8$
MLSP$_{opt}$	37.09	36.88	31.72
WIG	48.87	49.75	46.51
C-W	40.91 ($\sigma = 1.1$)	40.53 ($\sigma = 0.8$)	36.00 ($\sigma = 0.55$)
HANN	38.32 ($N_w = 41$)	38.10 ($N_w = 35$)	33.92 ($N_w = 31$)
WOSA	41.45 ($N_w = 35, K = 2$)	38.72 ($N_w = 32, K = 2$)	32.68 ($N_w = 32, K = 2$)
TH MW	41.44 ($N_w = 41, K = 2$)	38.33 ($N_w = 39, K = 2$)	32.30 ($N_w = 35, K = 2$)
PM MW	38.46 ($N_w = 31, K = 2$)	37.23 ($N_w = 29, K = 2$)	32.34 ($N_w = 27, K = 3$)

$c_1 = 1.1$, $c_2 = 20$, the second case with $c_1 = 2$, $c_2 = 10$, and the third case with $c_1 = 4$, $c_2 = 8$. For the C-W, the evaluation is done using σ that varies from $\sigma = 0.05$–1.9 in the first and second case, and from $\sigma = 0.025$–0.95 in the third case. It is seen that none of the methods reach the optimal mean square error for this more complicated process. The HANN and the PM MW, however, give small errors comparable to the error of the MLSP$_{opt}$ multitapers in the first case. In the last case, all the multitaper methods are able to reach a small mean square error. The worst results are produced from the WIG and C-W in all three cases.

5. Approximation of Windows Using Hermite Functions

In [5], the optimal multitapers of the proposed class of LSPs are shown to approximate a set of Hermite functions.
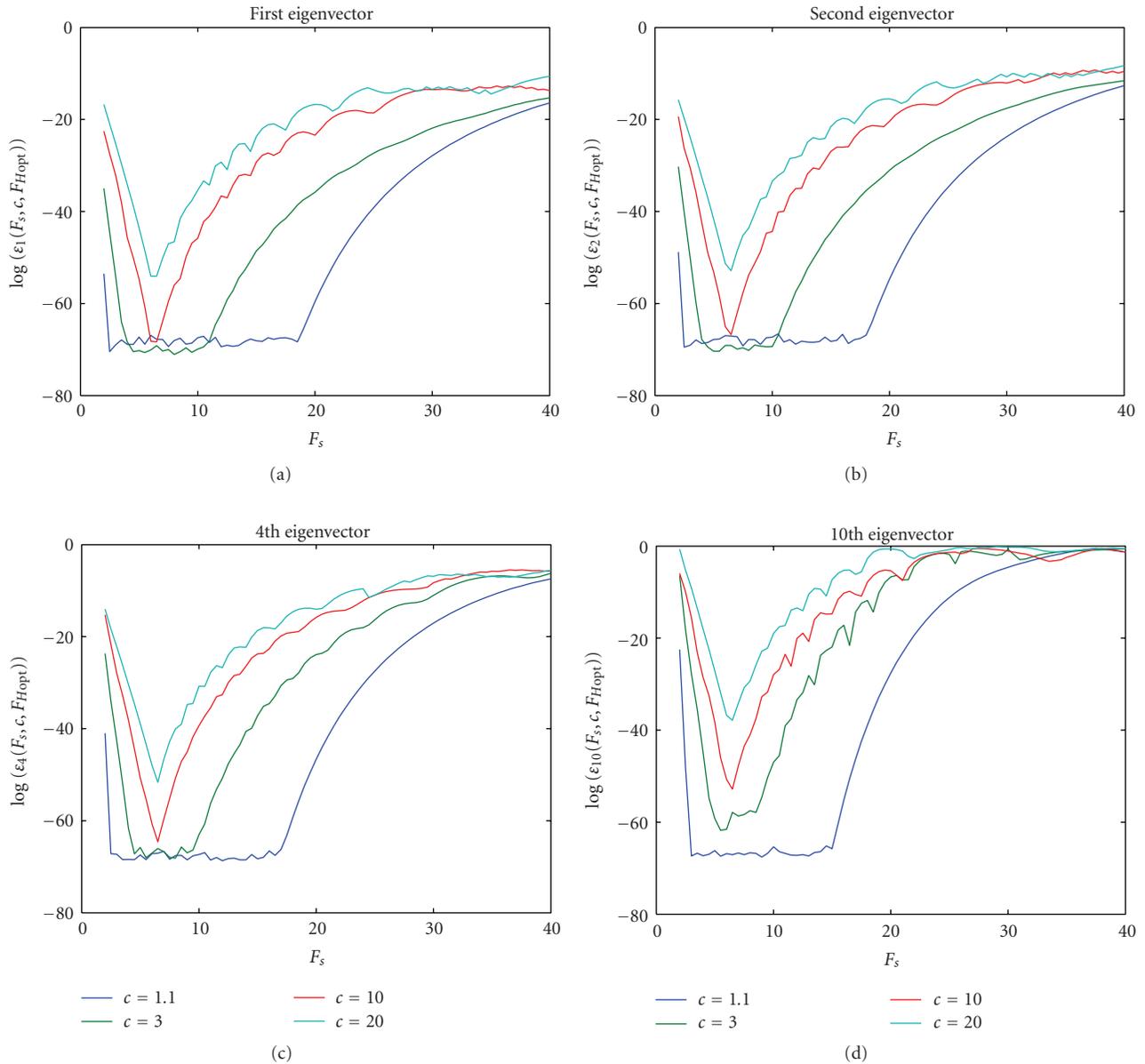
(a)

(b)

(c)

(d)

FIGURE 5: The dependence on F_s of the errors between the eigenvector optimal to LSP and the corresponding hermite function for different c-values. The 1st, 2nd, 4th, and 10th eigenvectors are compared. The eigenvector length is $N = 256$.

This approximation is appropriate in the continuous infinite-length case, but in the finite-length discrete-time case, a set of Hermite functions might be far from the true eigenvectors. How far, depends on the scaling of the matrix $\mathbf{R}_{\text{opt}}$ with use of the sampling frequency parameter F_s. If F_s is very small, the covariance values are large at the edges of the matrix $\mathbf{R}_{\text{opt}}$ and the Hermite functions certainly are inappropriate as eigenvectors. With F_s too large, the larger part of the power of $\mathbf{R}_{\text{opt}}$ will be located in just a few samples in the middle of the matrix, and the eigenvectors will consist of a few values that differ from zero. A proper choice between these extremes should be carefully considered.

Another result from [5] is that the Hermite functions can be dilated, where the dilation factors depend on c according to $c^{1/4}/\sqrt{2}$. If these results transfer to the time-discrete case, a very efficient implementation of the method is possible. Such an approximation is advantageous when it comes to calculation, since only a limited number of Hermite functions need to be calculated instead of solving an possibly large-scale eigenvalue problem which is different for all parameters c in the class. If, as well, *the same* set of Hermite functions can be used for different c-values, an even more efficient implementation can be done as the windowed spectrograms can be computed once for all c-values and weighted together using the different sets of eigenvalues for

each c. Thereby, the whole range of c-values are easily tuned by a new weighting of the set of computed spectrograms. A scaling of the sampling frequency of the optimal kernel and corresponding time-lag matrix $\mathbf{R}_{opt}$ is done according to

$$F_s^c = \frac{F_s c^{1/4}}{\sqrt{2}}. \tag{18}$$

Instead of dilating the Hermite functions, the optimal kernel is dilated for different c-values. Using this c-dependent sampling frequency we investigate a set of Hermite functions,

$$h_1(t) = e^{-t^2/2},$$

$$h_2(t) = 2te^{-t^2/2}, \tag{19}$$

$$h_k(t) = 2th_{k-1}(t) - 2(k-2)h_{k-2}(t), \quad k = 3, \ldots, K,$$

with $t = n/F_H$ for $n = -N/2, \ldots, N/2 - 1$. The resulting set of vectors is

$$\mathbf{h}_k^{F_H} = \left[h_k\left(-\frac{N}{2F_H}\right) \quad \cdots \quad h_k\left(\frac{N}{2F_H}\right) \right]^T, \quad k = 1, \ldots, K. \tag{20}$$

The square error

$$\varepsilon_k(F_s, c, F_H) = \left(\mathbf{h}_k^{F_H} - \mathbf{q}_k^{F_s^c}\right)^T \left(\mathbf{h}_k^{F_H} - \mathbf{q}_k^{F_s^c}\right) \tag{21}$$

is computed for all sets of Hermite functions with F_H varying from 0.1 to 50 with a step of $\Delta F_H = 0.1$. The best set of Hermite functions is chosen as the one where the sum of square errors between the first eigenvector and the first Hermite function is minimized, that is,

$$F_{H_{opt}} = \arg \min_{F_H} \varepsilon_1(F_s, c, F_H). \tag{22}$$

Thereby, the whole set of optimal Hermite functions $h_k^{F_{H_{opt}}}$, $k = 1, \ldots, K$, is determined for a specific F_s^c. However, the optimal set of Hermite functions found from this criterion does not necessarily have $F_{H_{opt}} = F_s$. The resulting square error $\varepsilon_k(F_s, c, F_{H_{opt}})$ is studied for a number of different F_s and c-values, where the sampling frequency F_s^c according to (18) is used for the computation of the optimal kernel and corresponding multitapers. The results are given in Figure 5 for $N = 256$ and in Figure 6 for $N = 32$. The logarithm of the different error vectors for c-values $c = 1.1, 3, 10$, and 20 are depicted in the same figure as a function of F_s and the results are presented for the 1st, 2nd, 4th, and 10th eigenvectors. The results in Figure 5 for $N = 256$ show that for the 1st eigenvector, the sampling F_s is less sensitive than that of the 10th eigenvector. The range of F_s, where the error is small, is between 2.5 and 20 for all $\mathbf{q}_1^{F_s^c}$ where it is just between 4 and 15 for $\mathbf{q}_{10}^{F_s^c}$, see Figures 5(a) and 5(d). However, with a proper choice of F_s, the Hermite functions approximate the first 10 eigenvectors in an appropriate way for $N = 256$. Within this range, it can also be verified that $F_s = F_{H_{opt}}$, and thereby the same set of Hermite functions can be used independently of the value of c.

TABLE 4: The appropriate range for $F_s = F_{H_{opt}}$ for small errors when using Hermite functions as multitapers optimal to LSP for different values of N.

N	512	256	128	64	32
interval F_s	7–30	4–12	4–6	3–3.5	2–2.5

In Figure 6, the square error $\varepsilon_k(F_s, c, F_{H_{opt}})$ is shown for $N = 32$, where the error is small in the range of F_s between 2 and 2.5 with an maximum error of 10^{-7} (1.5 and 3.4 if we allow a larger maximum error of 10^{-5}). Also within these ranges, it can be verified that $F_s = F_{H_{opt}}$. The error increases for the 2nd and 4th eigenvectors. The range of the minimum errors is, however, about the same as that of $\mathbf{q}_1^{F_s^c}$. For $\mathbf{q}_{10}^{F_s^c}$ the error is much larger independently of F_s (note the larger scale on the Y-axis). This means that for small N-values, the higher order eigenvectors are not possible to approximate as Hermite functions with a small error. The corresponding ranges for $N = 512, 128$, and 64 have also been studied, and the results are summarized in Table 4.

6. Reduction of the Number of Averaged Spectrograms

The only variables left to adapt the optimal set of multitapers to a certain process are the eigenvalues. In Figure 7, the eigenvalues for different c are depicted where the calculations are made for $N = 512$ and $F_s = 15$. Note the logarithmic X-axis. Using other values of N with appropriate corresponding F_s will give equal sets of eigenvalues. Changing F_s and N do not alter the eigenvalues, as N and F_s just change the sampling and size of the matrix $\mathbf{R}_{opt}$. This is verified to be true if $\mathbf{R}_{opt}$ is still sampled in an appropriate way, with the choice of F_s in the intervals specified in Table 4. From Figure 7, it is seen how the eigenvalues vary with c, for example, the λ_1 decreases slowly where λ_2 is first negative and then positive. For large values of c, several eigenvalues differ from zero where, for example, for $c = 2$ there are just a few eigenvalues that differ from zero.

An important aspect is that a multitaper spectrogram should be calculated with a reasonable number of windows. This is possible if a major part of the eigenvalues is close to zero. Here, four different levels for the eigenvalues to contribute to the final estimate are suggested based on

$$\lambda_k^{\delta} = \begin{cases} \lambda_k & \text{if } \left|\dfrac{\lambda_k}{\lambda_1}\right| \geq \delta, \\ 0 & \text{otherwise,} \end{cases} \tag{23}$$

where $\delta = 0$ (i.e., all eigenvalues are included), $\delta = 0.01$, $\delta = 0.05$, and $\delta = 0.2$. With this choice, the number of averaged multiple spectrograms reduces according to Figure 8. With the level set to $\delta = 0.01$, the number of spectrograms vary from $K = 3$ for $c = 1.5$ to $K = 23$ when $c = 30$, solid line of Figure 8. As many as 23 averaged spectrograms are rather impractical, and using the level $\delta = 0.05$, K reduces to be between 2 and 7, dashed line. In the last case with $\delta = 0.2$, the number of included eigenvalues might be too small to give an

First eigenvector

Second eigenvector

4th eigenvector

10th eigenvector
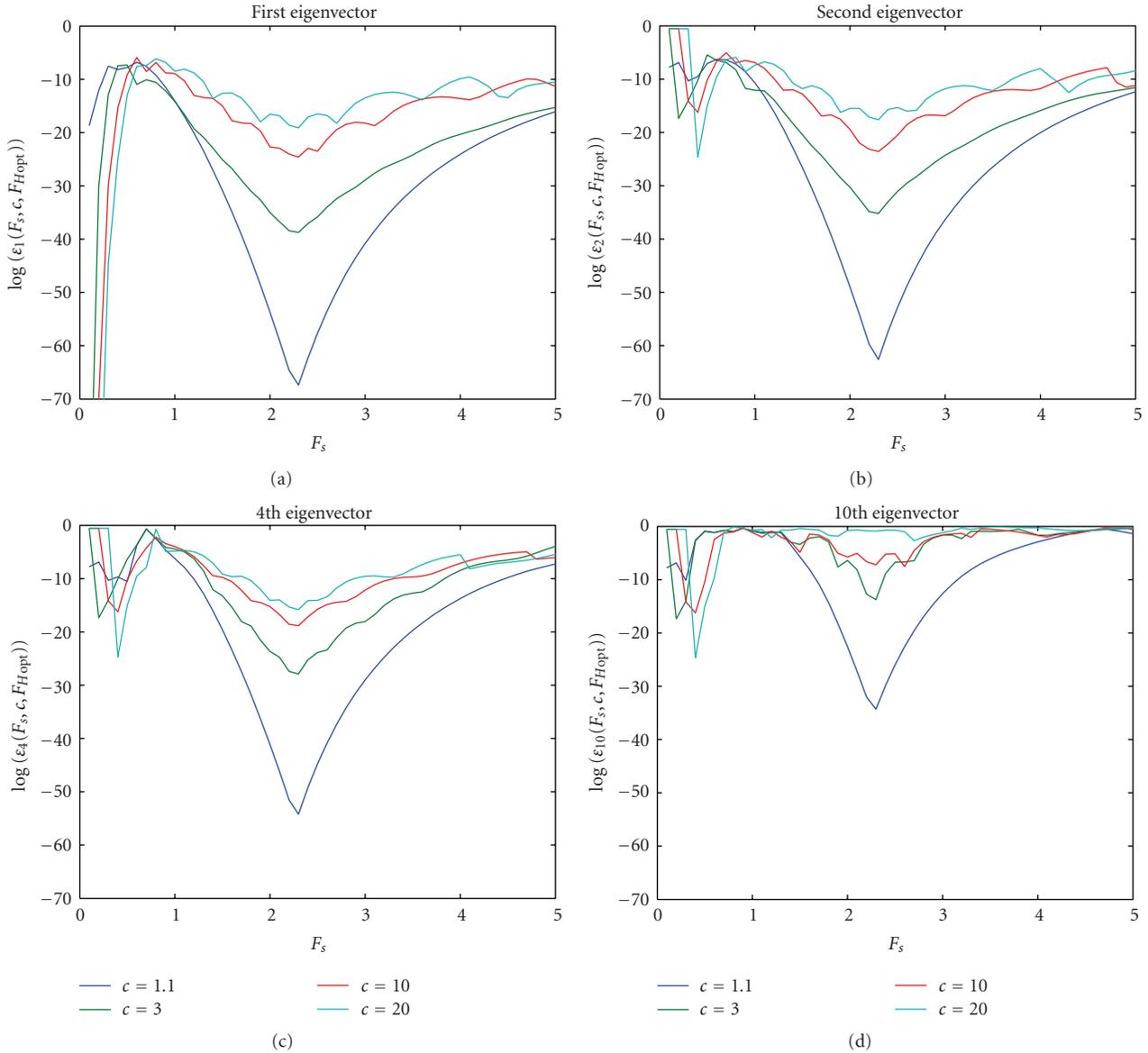
$c = 1.1$
$c = 3$
$c = 10$
$c = 20$

(a) (b) (c) (d)

FIGURE 6: The dependence on F_s of the errors between the eigenvector optimal to LSP and the corresponding hermite function for different c-values. The 1st, 2nd, 4th, and 10th eigenvector are compared. The eigenvector length is $N = 32$.

appropriate estimate, varying between $K = 1$–5, dash-dotted line.

A set of Hermite functions $\mathbf{h}_k^{F_{H\mathrm{opt}}}$, $k = 1, \ldots, K$, is easy to compute, and as the set can also be used as corresponding eigenvectors independently of the parameter c the advantage is even larger. If we, by this approximation of the eigenvectors, avoid solving an eigenvalue problem from a matrix of size $N \times N$, we should also use the advantage that the eigenvalues can be computed once and for all and stored to be utilized independently of the value of N. We use the computed eigenvalues in Figure 7 and name them $\lambda_k^{\mathrm{store}}$. Using this set of windows and eigenvalues, the mean square error J_{MW}^{δ} is computed where δ is determining the limit for the number of averages

in the spectrograms according to Figure 8. The result is compared with J_{opt}. In Table 5, the results using $N = 256$ and $F_s = 7$ is shown. It is apparent that the use of the Hermite functions does not degrade the result, and also that using $\delta = 0.01$ is equally optimal. With fewer windows (larger δ), a small difference in the mean square error can be seen.

In the next example, the window length is much shorter, $N = 32$ and $F_s = 2.5$. The results are seen in Table 6. For this short window length, the Hermite functions are not that appropriate as approximations of the eigenvectors, at least not for higher values of k, compare with Figure 6. This affects the case $c = 20$ the most, as K is larger in this case. The effect of reducing the number of windows is not that significant,
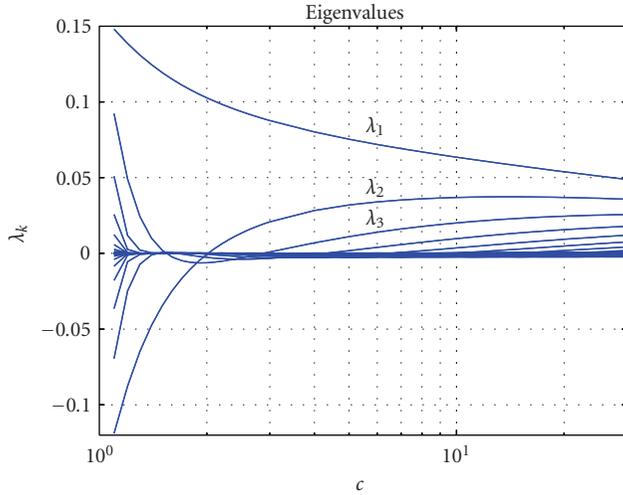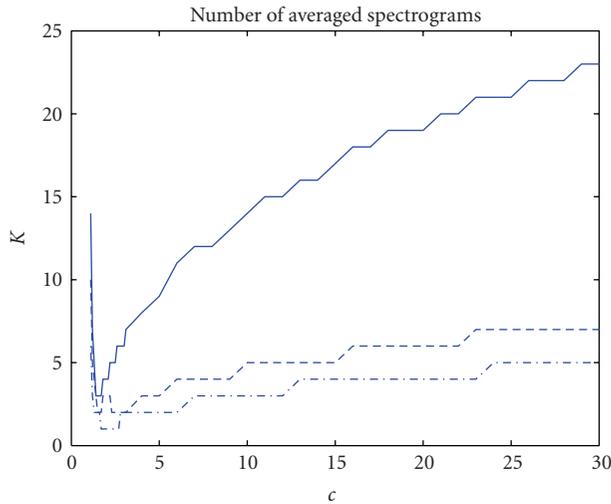
FIGURE 7: Eigenvalues, dependence on the parameter c.

FIGURE 8: Number of averaged multitaper spectrograms K for different values of c; solid line = $\delta = 0.01$, dashed line = $\delta = 0.05$, dash-dotted line = $\delta = 0.2$.

TABLE 5: Results of the MSE of the LSP for different parameter choices using Hermite functions as windows and a decreasing number of eigenvalues, $N = 256$ and $F_s = 7$.

Method	$c = 1.1$	$c = 3$	$c = 20$
J_{opt}	19.23	11.85	3.27
J_{MW}^{0}	19.23	11.85	3.27
$J_{MW}^{0.01}$	19.23	11.85	3.27
$J_{MW}^{0.05}$	19.29	11.87	3.30
$J_{MW}^{0.2}$	19.30	11.87	3.31

and the level $\delta = 0.01$ gives the same result as using all windows, ($\delta = 0$).

The last example in Table 7 shows the result when $N = 32$, but F_s is now altered to be outside the appropriate range, to $F_s = 5$. This choice of F_s gives a very large error when all windows are included in the estimate, ($\delta = 0$). The reason of

TABLE 6: Results of the MSE of the LSP for different parameter choices using Hermite functions as windows and a decreasing number of eigenvalues, $N = 32$ and $F_s = 2.5$.

Method	$c = 1.1$	$c = 3$	$c = 20$
J_{opt}	19.23	11.85	3.27
J_{MW}^{0}	19.23	11.86	3.32
$J_{MW}^{0.01}$	19.23	11.86	3.32
$J_{MW}^{0.05}$	19.23	11.87	3.34
$J_{MW}^{0.2}$	19.29	11.87	3.34

TABLE 7: Results of the MSE of the LSP for different parameter choices using Hermite functions as windows and a decreasing number of eigenvalues, $N = 32$ and $F_s = 5$.

Method	$c = 1.1$	$c = 3$	$c = 20$
J_{opt}	19.23	11.85	3.27
J_{MW}^{0}	37.62	22.79	8.83
$J_{MW}^{0.01}$	19.36	11.85	6.41
$J_{MW}^{0.05}$	19.23	11.87	3.30
$J_{MW}^{0.2}$	19.30	11.87	3.31

course is that the Hermite functions are not at all appropriate to use for larger K, as they are no longer similar to the eigenvectors. Reducing the number of windows, however, $\delta \geq 0.01$, gives a mean square error sufficiently close to the true J_{opt}. This shows that the Hermite functions together with limiting the number of windows give an appropriate estimate even if F_s is chosen inappropriately.

7. Estimation of Induced Potential Power in the Electroencephalogram

To show the performance of the proposed method and especially show the possibility to view the multitaper spectrograms for different model assumptions, that is, different c-values, examples are shown of the electrophysiological correlates of toddlers' processing of imperative and declarative gestures. Neural responses to video clips displaying different types of nonlinguistic communication between two actors are recorded. Results from investigations of language learning, indicate that there may be important developments in this domain in the second half of the second year, [29]. The hypothesis was that if the comprehension of declarative, but not imperative, communication acts is specific to humans, the processing of these two types of acts should be associated with different neural activity. There is not much research on what young children understand as onlookers of communication between other people. It is important to find suitable models for the time-varying power estimation of these signals.

The stimulus material consisted of different 13-second video clips merged to a 2×15-minute continuous sequence, intended to show the three kinds of social behavior; communicative (protodeclarative or protoimperative gestures) and communicatively neutral parallel play. Silver-silver chloride electrodes (EasyCap, Falk Minow) were placed according to
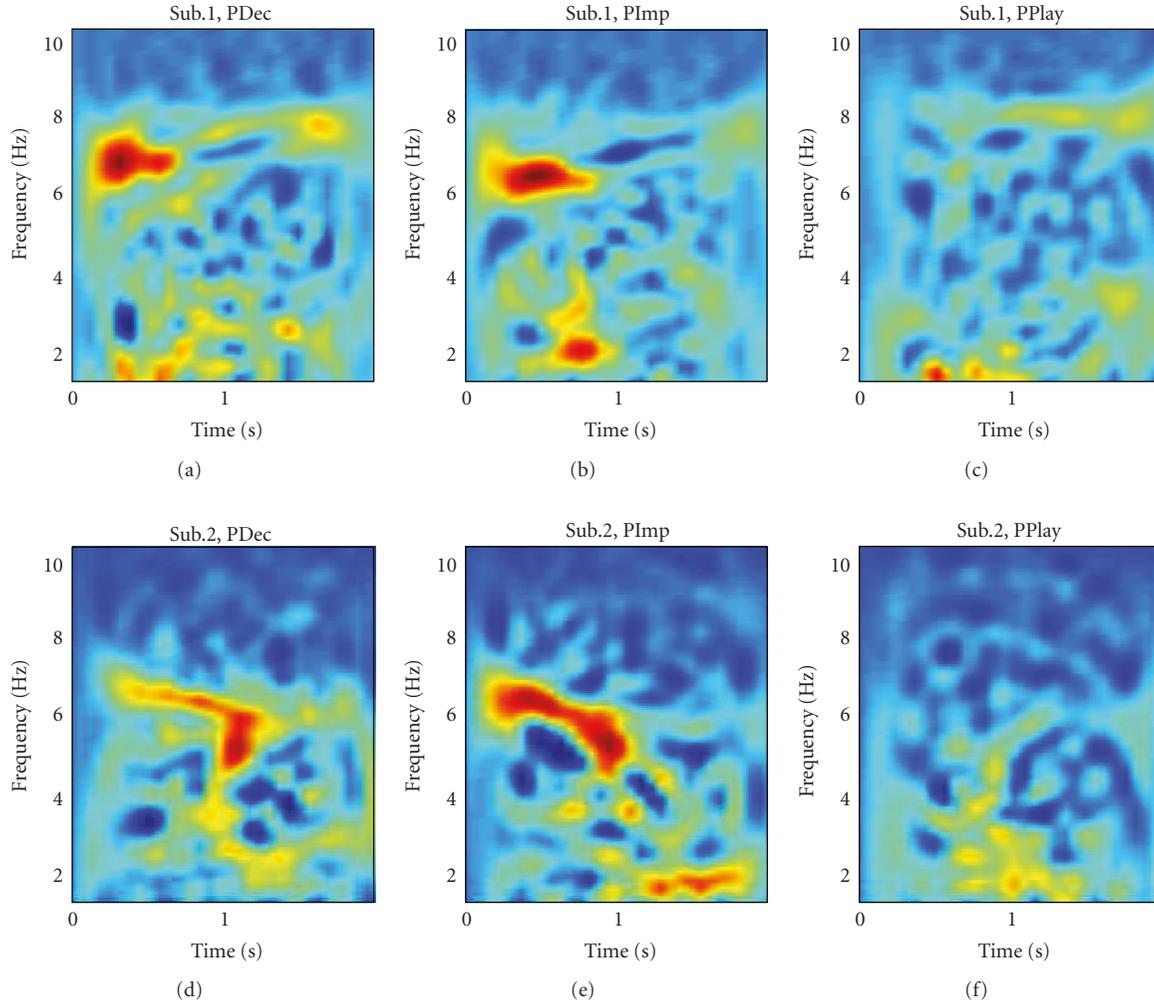
FIGURE 9: Multitaper spectrogram estimation of induced potentials from Protodeclarative (PDec), Protoimperative (PImp) and neutral parallel play (PPlay) shows, using $N = 256$, $F_s = 12$, $c = 1.1$, and $\delta = 0.05$. The number of averaged power responses are for subject 1, 20 and for subject 2, 29.

the international 10–20 system. The vertical electrooculogram (VEOG) was recorded from electrodes placed above and below the right eye, and the horizontal electrooculogram (HEOG) was recorded from electrodes placed lateral to the left and the right eye. All electrodes were referenced to the average of the left and the right mastoids. Impedances were kept below 5 kΩ for all electrodes. The EEG was recorded with a 0.1/70 Hz band-pass filter at a sampling rate of 500 Hz, and amplified with a Neuroscan NuAmps amplifier. The sampling frequency is reduced to 62.5 Hz by decimation in Matlab and before and after the 2s epochs, zeros are added to allow for the total data window length. The included data epochs are limited so that the same number of epochs is used for the declarative, imperative, and parallel play acts. For subject 1, the number of epochs were 20 in each of the three cases, and for subject 2, the number of epochs were 29.

The averaged power from two different subjects is estimated using a set of Hermite functions with $N = 256$ and $F_s = F_{H_{opt}} = 12$. Different sets of eigenvalues are used as weighting factors, corresponding to $c = 1.1$, $c = 3$,

and $c = 20$. Using $\delta = 0.05$, the estimation of the spectrogram is then done using a reasonable number of averaged spectrograms, $K = 10$, $K = 2$ and, $K = 6$, respectively. Remember, that with this procedure only the averaging of the spectrograms need to be redone for a new value of the parameter c, the individual single-window periodograms with the set of chosen Hermite functions are the same. The results of two subjects are shown in Figures 9, 10, and 11. For each subject, the averaged estimates for all epochs are shown. The colormap used is the same for each row (each subject) in all the figures. The results show that the induced responses are stronger around 6 Hz and between 0 and 1 s for the protodeclarative as well as the protoimperative act compared to the parallel play act, for both subjects. Similar performance is found for other subjects. In the different figures, the resolution varies from being similar to the Wigner distribution ($c = 1.1$) in Figure 9, to almost a single-window spectrogram estimator ($c = 3$) in Figure 10 (just $K = 2$ windows) and finally similar to the usual multitaper estimator, for example, the Thomson
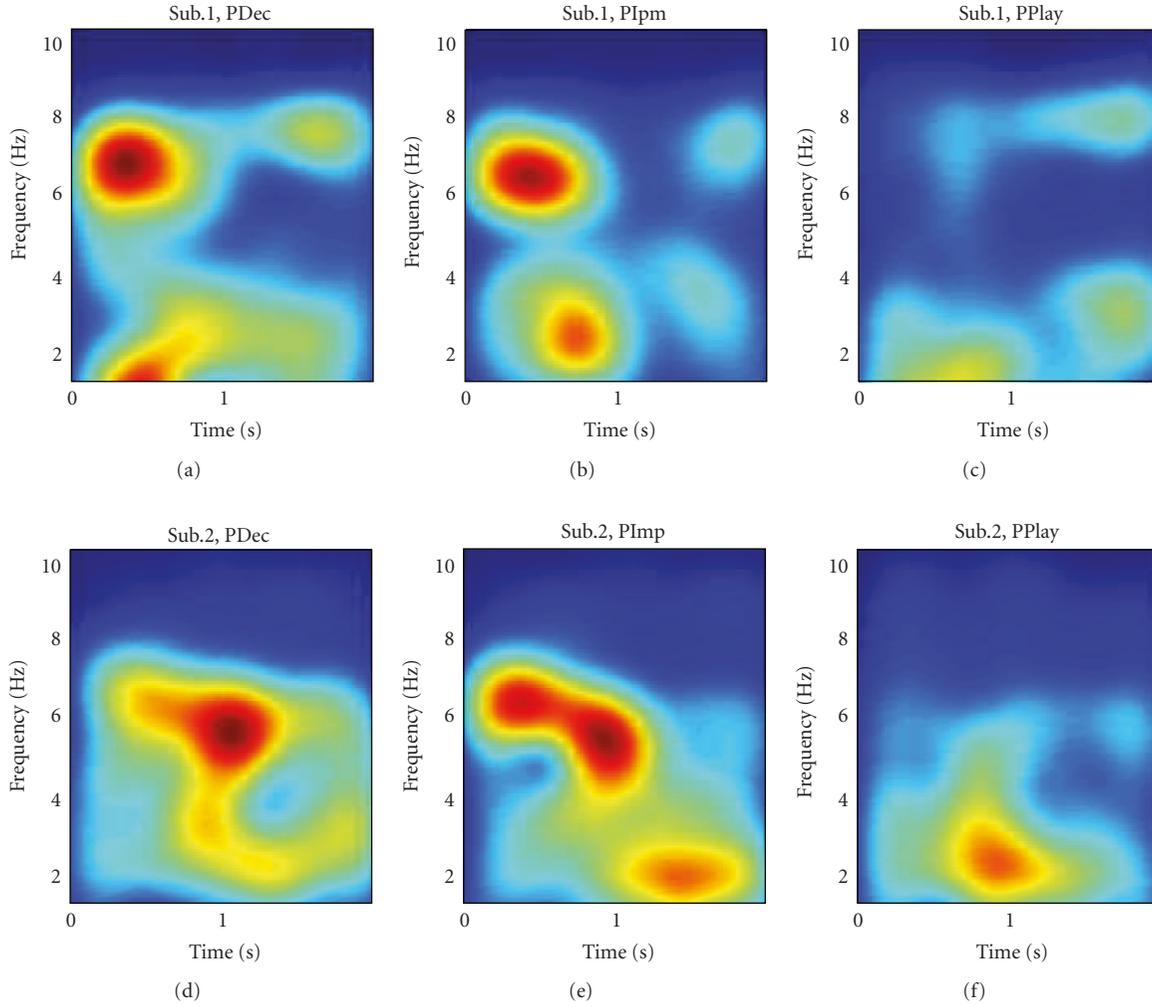
FIGURE 10: Multitaper spectrogram estimation of induced potentials from Protodeclarative (PDec), Protoimperative (PImp) and neutral parallel play (PPlay) shows, using $N = 256$, $F_s = 12$, $c = 3$, and $\delta = 0.05$. The number of averaged power responses are for subject 1, 20 and for subject 2, 29.

estimator, ($c = 20$) in Figure 11. This shows the possibility to easily investigate a whole range of estimators when the data model is not well defined.

8. Conclusions

The mean square error optimal multitaper spectrogram estimator, for a class of locally stationary processes is evaluated and compared to other frequently used methods: the Wigner distribution, the Choi-Williams distribution, the single Hanning spectrogram, the Welch method, the Thomson multitapers, and the peak matched multitapers. The results show that it is possible to find other methods that give a result more or less close to the result of the optimal multitapers in mean square error sense, but it is different methods for different processes of the class. An evaluation is also made for the classes of locally stationary chirp processes and sum of locally stationary processes.

An investigation of the error shows that with a proper sampling frequency of the windows, the windows are well approximated by a set of Hermite functions where the same set of functions is valid for the whole class if an appropriate dilation factor is used. Another advantage is that the optimal set of weighting factors can be stored and utilized for all different window lengths. It is also investigated if the number of included spectrograms in the average can be reduced, and the results show that only a small number of windows, 1–10, are needed for an appropriate approximation of the true kernel.

The advantage of the usage of the Hermite functions is that the multitapers as well as the corresponding spectrograms can be calculated and stored, and then a tuning process using just one parameter, which determines the weights, will give the final weighted multitaper spectrogram estimate for a certain parameter of the whole class. Power estimation of induced potentials of the brain from toddlers'
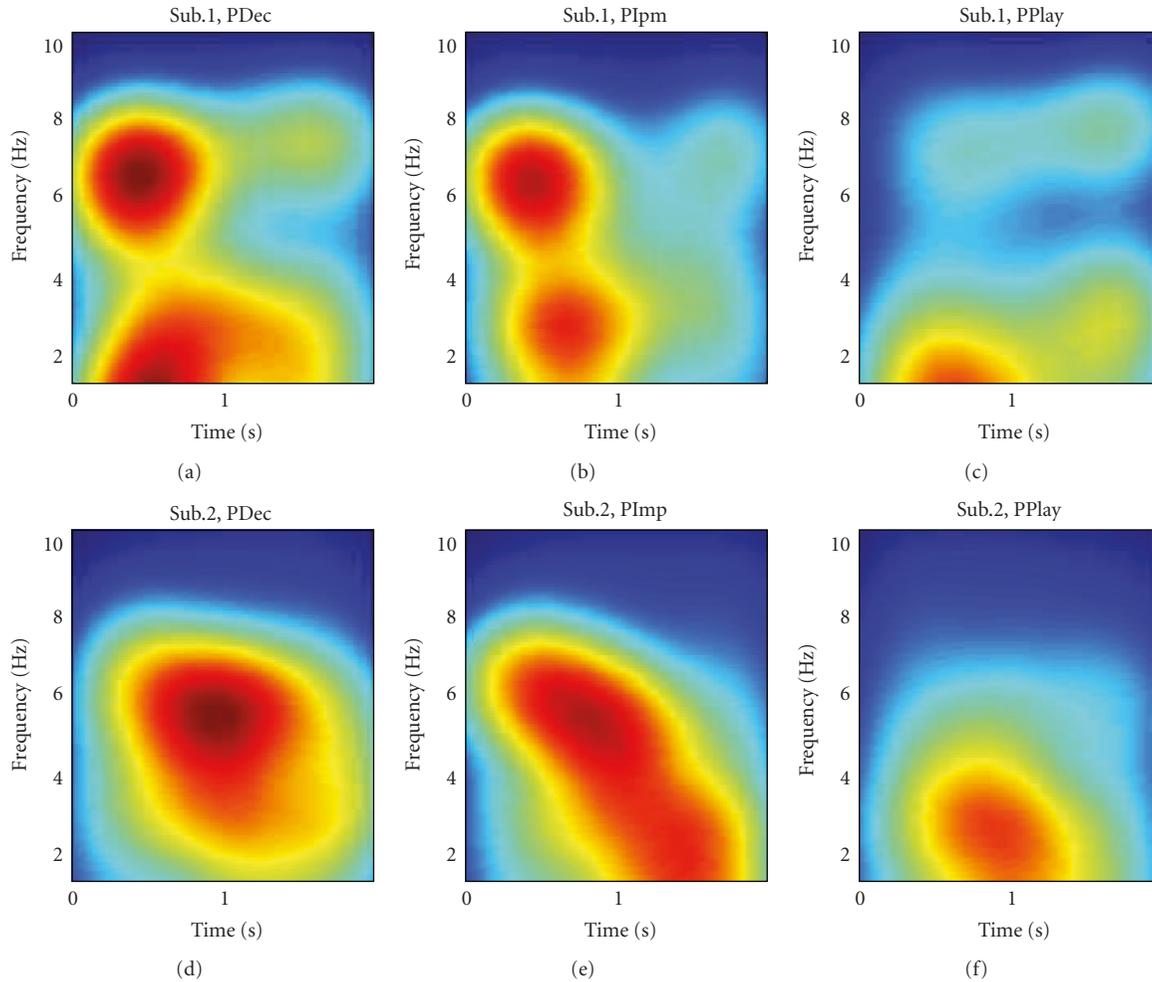
Figure 11: Multitaper spectrogram estimation of induced potentials from Protodeclarative (PDec), Protoimperative (PImp) and Neutral Parallel Play (PPlay) shows, using $N = 256$, $F_s = 12$, $c = 20$ and $\delta = 0.05$. The number of averaged power responses are for subject 1, 20 and for subject 2, 29.

processing of imperative and declarative gestures, are shown as example of the use of the proposed method.

Acknowledgments

References

[1] R. A. Silverman, "Locally stationary random processes," *IRE Transactions on Information Theory*, vol. 3, pp. 182–187, 1957.

[2] M. Hansson-Sandsten and J. Sandberg, "Optimization of weighting factors for multiple window spectrogram of event-related potentials," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, Article ID 391798, 10 pages, 2010.

[3] M. Hansson-Sandsten, "Evaluation of the optimal lengths and number of multiple windows for spectrogram estimation of SSVEP," *Medical Engineering and Physics*, vol. 32, no. 4, pp. 372–383, 2010.

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

[5] 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.

[6] G. S. Cunnigham and W. J. Williams, "Kernel decomposition of time-frequency distributions," *IEEE Transactions on Signal Processing*, vol. 42, no. 6, pp. 1425–1442, 1994.

[7] M. G. Amin, "Spectral decomposition of time-frequency distribution kernels," *IEEE Transactions on Signal Processing*, vol. 42, no. 5, pp. 1156–1165, 1994.

[8] L. Cohen, *Time-Frequency Analysis*, Prentice-Hall, New York, NY, USA, 1995.

[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] L. L. Scharf and B. Friedlander, "Toeplitz and Hankel kernels for estimating time-varying spectra of discrete-time random processes," *IEEE Transactions on Signal Processing*, vol. 49, no. 1, pp. 179–189, 2001.

[11] M. Bayram and R. G. Baraniuk, "Multiple window time-frequency analysis," in *Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, pp. 511–514, June 1996.

[12] S. Aviyente and W. J. Williams, "A centrosymmetric kernel decomposition for time-frequency distribution computation," *IEEE Transactions on Signal Processing*, vol. 52, no. 6, pp. 1574–1584, 2004.

[13] S. Aviyente and W. J. Williams, "Multitaper marginal time-frequency distributions," *Signal Processing*, vol. 86, no. 2, pp. 279–295, 2006.

[14] M. Hansson, "Multiple window decomposition of time-frequency kernels using a penalty function for suppressed sidelobes," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '06)*, vol. 3, pp. 432–435, Toulouse, France, May 2006.

[15] M. Hansson-Sandsten, "Multi taper wigner distribution with predetermined doppler-lag bandwidth and sidelobe suppression," in *Proceedings of the European Signal Processing Conference (EUSIPCO '10)*, Aalborg, Denmark, 2010.

[16] M. S. Richman, T. W. Parks, and R. G. Shenoy, "Discrete-time, discrete-frequency, time-frequency analysis," *IEEE Transactions on Signal Processing*, vol. 46, no. 6, pp. 1517–1527, 1998.

[17] A. H. Costa and G. Faye Boudreaux-Bartels, "An overview of aliasing errors in discrete-time formulations of time-frequency representations," *IEEE Transactions on Signal Processing*, vol. 47, no. 5, pp. 1463–1474, 1999.

[18] J. Sandberg and M. Hansson-Sandsten, "A comparison between different discrete ambiguity domain definitions in stochastic time-frequency analysis," *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 868–877, 2009.

[19] C. Richard, "Time-frequency-based detection using discrete-time discrete-frequency Wigner distributions," *IEEE Transactions on Signal Processing*, vol. 50, no. 9, pp. 2170–2176, 2002.

[20] E. Wigner, "On the quantum correction for thermodynamic equilibrium," *Physical Review*, vol. 40, no. 5, pp. 749–759, 1932.

[21] H. I. Choi and W. J. Williams, "Improved time-frequency representation of multicomponent signals using exponential kernels," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 6, pp. 862–871, 1989.

[22] B. Boashash, Ed., *Time Frequency Signal Analysis and Processing*, Elsevier, New York, NY, USA, 1st edition, 2003.

[23] 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. 15, no. 2, pp. 70–73, 1967.

[24] D. J. Thomson, "Spectrum estimation and harmonic analysis," *Proceedings of the IEEE*, vol. 70, no. 9, pp. 1055–1096, 1982.

[25] M. Hansson and G. Salomonsson, "A multiple window method for estimation of peaked spectra," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 778–781, 1997.

[26] M. Hansson and M. Lindgren, "Multiple-window spectrogram of peaks due to transients in the electroencephalogram," *IEEE Transactions on Biomedical Engineering*, vol. 48, no. 3, pp. 284–293, 2001.

[27] M. Hansson-Sandsten and P. Jönsson, "Multiple window correlation analysis of HRV power and respiratory frequency," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 10, pp. 1770–1779, 2007.

[28] M. Hansson, "Optimized weighted averaging of peak matched multiple window spectrum estimates," *IEEE Transactions on Signal Processing*, vol. 47, no. 4, pp. 1141–1146, 1999.

[29] J. v. K. Torkildsen, H. Friis Hansen, J. M. Svangstu et al., "Brain dynamics of word familiarization in 20-month-olds: effects of productive vocabulary size," *Brain and Language*, vol. 108, no. 2, pp. 73–88, 2009.