Multitaper Estimation of Frequency-Warped Cepstra With Application to Speaker Verification

Johan Sandberg, Maria Hansson-Sandsten, Tomi Kinnunen,, Rahim Saeidi, Patrick Flandrin, Pierre Borgnat

Multitaper Estimation of Frequency-Warped Cepstra with Application to Speaker Verification

Johan Sandberg, Maria Hansson-Sandsten, *Member, IEEE,* Tomi Kinnunen, Rahim Saeidi, *Student Member, IEEE,* Patrick Flandrin, *Fellow, IEEE,* and Pierre Borgnat, *Member, IEEE*

***Abstract*—Usually the mel-frequency cepstral coefficients are estimated either from a periodogram or from a windowed periodogram. We state a general estimator which also includes multitaper estimators. We propose approximations of the variance and bias of the estimate of each coefficient. By using Monte Carlo computations, we demonstrate that the approximations are accurate. Using the proposed formulas, the peak matched multitaper estimator is shown to have low mean square error (squared bias + variance) on speech-like processes. It is also shown to perform slightly better in the NIST 2006 speaker verification task as compared to the Hamming window conventionally used in this context.**

***Index Terms*—Cepstral analysis, MFCC, Multiple windows, Multitapers, Speech analysis, Speaker verification**

I. Introduction

THE cepstrum was introduced by Bogert, Healy and Tukey in the early sixties [1]. It is defined as the inverse Fourier transform of the log-spectrum of a stationary random process [2]. The cepstrum has become a fundamental tool in many applications, such as speech and audio processing [3]. In these applications, a psycho-acoustically motivated frequency warping transformation is usually applied to the spectrum before the logarithm and the inverse Fourier transform, such as in the popular mel-frequency cepstral coefficients (MFCCs) [3]. The results presented in this letter will hold for any frequency warped cepstrum, including the ordinary cepstrum, but we will still use the term "MFCC" for brevity.

Typically, the spectrum for MFCC computation is estimated using the periodogram, i.e. squared magnitude of the Fourier transformation of the data. The periodogram suffers from large bias and large variance, altogether causing large estimation errors in the cepstral coefficients. The bias can be reduced by windowing the time series with, for example, a Hamming window [4]. The windowed periodogram has low bias in general, but it still suffers from high variance. Therefore, one may consider using a so-called *multitaper* spectral estimator instead. A multitaper spectral estimator is an average of windowed periodograms using different orthogonal windows (aka *tapers*), e.g. the *Thomson* [5], the *sine* [6], and the *peak matched* multitapers [7]. The multitaper spectrum estimator is known to have low variance, but has not gained much attention in MFCC estimation [8]. One reason may be that the statistical properties of the multitaper MFCC estimator have previously not been investigated. It is our purpose to address this issue in this letter.

In Section II of this letter we state the general form of an MFCC estimator, which will include the MFCC computed from the periodogram, the windowed periodogram, the Bartlett and the Welch method, as well as multitaper spectrum estimators. The statistical properties of the cepstrum computed from the periodogram are well known [2], [9], [10]. However, the bias and variance of the cepstrum or of the MFCCs computed from the multitaper spectrum estimator have not been, to the best of our knowledge, studied so far. In Section III-A, we therefore derive approximate expressions for the bias and variance of our general MFCC estimator. From a statistical viewpoint, this is an important result, since one may argue that it is hazardous to use an estimator without knowing its bias and variance. In Section III-B, we compare the approximations with Monte Carlo computations which show that our approximations are accurate.

The approximate formulas for the bias and variance that we derive are then used in Section IV to compare the mean square error (MSE = squared bias + variance) of different MFCC estimators, including the Hamming window, the Thomson multitapers, the sine multitapers and the peak matched multitapers on speech-like random processes. Our results show that the multitaper MFCC estimators have much lower MSE than the commonly used Hamming window estimator. Finally, we demonstrate the effectiveness of multitaper MFCC estimation over the conventional Hamming window based MFCC extraction, in a speaker verification context. The results, in the framework of NIST 2006 speaker recognition evaluation (SRE), are presented in Section V.

II. The General Non-Parametric MFCC estimator

In speech applications, one frame of data (~ 30 ms) can be modeled as a stationary Gaussian random process. Thus, let $\mathbf{x} = [x(0) \ \ldots \ x(n-1)]^T$ be a part of a real-valued Gaussian zero-mean stationary random process in discrete time with a strictly positive spectrum $s(f)$, $0 \leq f < 1$. It is assumed that the covariance function of the process is zero for time-lags greater than n. Our aim is to estimate the MFCC, $\mathbf{c_M} \in \mathbb{R}^m$,

J. Sandberg and M. Hansson-Sandsten are with the Mathematical Statistics, at the Centre for Mathematical Sciences, Lund University, Box 118, SE-221 00 Lund, Sweden (e-mail: {sandberg, sandsten}@maths.lth.se).

T. Kinnunen and R. Saeidi are with the Speech and Image Processing Unit, Department of Computer Science and Statistics, University of Eastern Finland, FIN-80101 Joensuu, Finland (e-mail: {tkinnu, rahim}@cs.joensuu.fi).

P. Flandrin and P. Borgnat are with Laboratoire de Physique, École Normale Supérieure de Lyon, UMR 5672 CNRS, 46 allée d'Italie 69364 Lyon Cedex 07 France (e-mail: {flandrin, pborgnat}@ens-lyon.fr).

The first author would like to thank ENSL for great hospitality during his stay there. This work was in part supported by the Swedish Research Council and by The Royal Physiographic Society in Lund.

of this process, which is defined as:

$$\mathbf{c_M} \triangleq \frac{1}{m}\mathbf{\Phi}^H \log(\mathbf{M s}) \tag{1}$$

where log operates element-wise, $\mathbf{M} \in \mathbb{R}^{m\times n}$ is a frequency warping filter bank, the superscript H denotes conjugate transpose, $\mathbf{\Phi}$ is the m-by-m Fourier matrix with the (a,b):th element:

$$\mathbf{\Phi} \triangleq \left\{ \mathrm{e}^{-i2\pi(a-1)(b-1)/m} \right\}_{ab},$$

and $\mathbf{s} = [s(0/n)\ s(1/n)\ s((n-1)/n)]^T$ is the spectrum vector, which is symmetrical, i.e. $s(k/n) = s((n-k)/n)$. The filter bank $\mathbf{M}$ is chosen such that $\mathbf{Ms}$ possesses the same symmetry as the spectrum vector $\mathbf{s}$. Due to this symmetry, (1) is real-valued and can efficiently be computed using the discrete cosine transform (DCT). Note that $\mathbf{c_M}$ reduces to the ordinary cepstrum if $\mathbf{M}$ is chosen to be the n-by-n identity matrix.

In this letter, we will consider the following MFCC estimator:

$$\hat{\mathbf{c}}_\mathbf{M} = \frac{1}{m}\mathbf{\Phi}^H \log(\mathbf{M}\hat{\mathbf{s}}) \tag{2}$$

where $\hat{\mathbf{s}}$ is the multitaper spectrum estimator [4], [5], given by:

$$\hat{\mathbf{s}} = [\hat{s}(0)\ \hat{s}(1)\ \dots\ \hat{s}(n-1)]^T \quad \text{with} \tag{3}$$

$$\begin{aligned}\hat{s}(p) &= \sum_{j=1}^{k} \lambda(j) \left| \sum_{t=0}^{n-1} w_j(t) x(t) \mathrm{e}^{-i2\pi tp/n} \right|^2 \\ &= \sum_{j=1}^{k} \lambda(j) \left| \mathbf{w}_j^T \mathbf{\Psi}_p \mathbf{x} \right|^2,\ p = 0, \dots, n-1 \end{aligned} \tag{4}$$

where k multitapers, $\mathbf{w}_j = [w_j(0) \dots w_j(n-1)]^T$, $j = 1, \dots, n$, are used with corresponding weights $\lambda(j)$, and where $\mathbf{\Psi}_p$ is the n-by-n diagonal Fourier matrix defined by:

$$\mathbf{\Psi}_p \triangleq \mathsf{diag}\left(\left[\mathrm{e}^{-i2\pi p\frac{0}{n}}\ \mathrm{e}^{-i2\pi p\frac{1}{n}}\ \dots\ \mathrm{e}^{-i2\pi p\frac{n-1}{n}} \right]^T \right).$$

The multitaper estimate is thus computed as a weighted average of k sub-spectra, $\left|\mathbf{w}_j^T \mathbf{\Psi}_p \mathbf{x}\right|^2$, $j = 1, \dots, k$. This will reduce the variance of the estimate, since the multitapers are designed such that the different sub-spectra are approximately uncorrelated with each other [4], [5], [7].

The estimator reduces to the windowed periodogram if $k = 1$ and $\lambda = 1$ and if additionally, $w_1(t) = \frac{1}{\sqrt{n}}$, it reduces to the periodogram. It will also account for the Bartlett and the Welch method by appropriate choice of $\mathbf{w}_j$ and $\lambda(j)$, $j = 1, \dots, k$. By selecting the frequency warping matrix $\mathbf{M}$, this estimator will transform to any frequency warped cepstrum, including MFCC and the ordinary non-warped cepstrum.

III. Bias and variance of the estimator

A. Proposed approximation using Taylor expansion

From a statistical perspective, it is of great interest to compute the bias and variance of a proposed estimator. In this section we will, for the first time, derive approximate formulas for the bias, $\mathsf{bias}\left[\hat{\mathbf{c}}_\mathbf{M}\right]$, and the covariance matrix, $\mathsf{V}\left[\hat{\mathbf{c}}_\mathbf{M}\right]$, of the MFCC estimator. From a practical perspective, it seems reasonable to prefer the estimator with the smallest MSE for all coefficients:

$$\mathsf{MSE}(\hat{\mathbf{c}}_\mathbf{M}) \triangleq \mathsf{E}\left[(\mathbf{c_M} - \hat{\mathbf{c}}_\mathbf{M})^2\right] = \mathsf{bias}\left[\hat{\mathbf{c}}_\mathbf{M}\right]^2 + \mathsf{diag}\left(\mathsf{V}\left[\hat{\mathbf{c}}_\mathbf{M}\right]\right) \tag{5}$$

where the square operates element-wise.

To derive the bias and variance of our general MFCC estimator, we start with the expectation of the multitaper spectrum estimator, which is given in [4]:

$$\mathsf{E}\left[\hat{\mathbf{s}}\right] = \begin{bmatrix} \boldsymbol{\lambda}^T \mathrm{diag}\left(\mathbf{W}^T \mathbf{\Psi}_0 \mathbf{R} \mathbf{\Psi}_0^H \mathbf{W}\right) \\ \vdots \\ \boldsymbol{\lambda}^T \mathrm{diag}\left(\mathbf{W}^T \mathbf{\Psi}_{n-1} \mathbf{R} \mathbf{\Psi}_{n-1}^H \mathbf{W}\right) \end{bmatrix} \tag{6}$$

where $\mathbf{R}$ denotes the covariance matrix of $\mathbf{x}$, the multitaper vectors are the columns of $\mathbf{W} \in \mathbb{R}^{n\times k}$ with weights $\boldsymbol{\lambda} = [\lambda(1)\ \dots\ \lambda(k)]^T$. The (a,b):th element of the covariance matrix of the multitaper spectrum estimator can be found, similar to what is done in [4, page 229], as:

$$\begin{aligned}\mathsf{V}\left[\hat{\mathbf{s}}\right] = \Big\{ &\boldsymbol{\lambda}^T \left|\mathbf{W}^T \mathbf{\Psi}_{a-1} \mathbf{R} \mathbf{\Psi}_{b-1} \mathbf{W}\right|^2 \boldsymbol{\lambda} + \\ &\boldsymbol{\lambda}^T \left|\mathbf{W}^T \mathbf{\Psi}_{a-1} \mathbf{R} \mathbf{\Psi}_{b-1}^H \mathbf{W}\right|^2 \boldsymbol{\lambda} \Big\}_{ab} \end{aligned} \tag{7}$$

where the absolute and square operator are defined element-wise.

We consider the following Taylor expansion around the mean, m_z, of a random variable z:

$$\log(z) \approx \log(m_z) + \frac{1}{m_z}(z - m_z) - \frac{1}{2m_z^2}(m_z - z)^2, \tag{8}$$

which is an extension of the commonly used formulas for propagation of uncertainty. Since $\mathsf{E}\left[\frac{1}{m_z}(z - m_z)\right] = 0$, this approximation gives us

$$\mathsf{E}\left[\log(z)\right] \approx \log(m_z) - \frac{\mathsf{V}\left[z\right]}{2m_z^2}.$$

Applying this element-wise on a random vector $\mathbf{z}$ gives:

$$\mathsf{E}\left[\log(\mathbf{z})\right] \approx \log(\mathsf{E}\left[\mathbf{z}\right]) - \frac{\mathsf{diag}\left(\mathsf{V}\left[\mathbf{z}\right]\right)}{2\mathsf{E}\left[\mathbf{z}\right]^2}$$

where the logarithm, the division and the square operate element-wise on vectors. This gives us the following approximation of the expectation of the multitaper cepstrum estimator:

$$\begin{aligned}\mathsf{E}\left[\hat{\mathbf{c}}_\mathbf{M}\right] &= \frac{1}{m}\mathbf{\Phi}^H \mathsf{E}\left[\log(\mathbf{M}\hat{\mathbf{s}})\right] \\ &\approx \frac{1}{m}\mathbf{\Phi}^H \left(\log\left(\mathbf{M}\mathsf{E}\left[\hat{\mathbf{s}}\right]\right) - \frac{\mathsf{diag}\left(\mathbf{M}\mathsf{V}\left[\hat{\mathbf{s}}\right]\mathbf{M}^T\right)}{2\left(\mathbf{M}\mathsf{E}\left[\hat{\mathbf{s}}\right]\right)^2} \right). \end{aligned} \tag{9}$$

The bias of the estimator is:

$$\mathsf{bias}\left[\hat{\mathbf{c}}_\mathbf{M}\right] \approx \frac{1}{m}\mathbf{\Phi}^H \left(\log\left(\frac{\mathbf{M}\mathsf{E}\left[\hat{\mathbf{s}}\right]}{\mathbf{Ms}}\right) - \frac{\mathsf{diag}\left(\mathbf{M}\mathsf{V}\left[\hat{\mathbf{s}}\right]\mathbf{M}^T\right)}{2\left(\mathbf{M}\mathsf{E}\left[\hat{\mathbf{s}}\right]\right)^2} \right). \tag{10}$$

Based on comparisons with Monte Carlo computations as described in Section III-B, it is our experience that the last term in the Taylor expansion in (8) significantly improves the

approximation of the bias. The approximation of the variance, however, is sufficiently accurate even when omitting the last term. Thus, after dropping the last term in (8), we find

$$\mathsf{V}\left[\log(z)\right] \approx \frac{\mathsf{V}\left[z\right]}{m_z^2}.$$

For a random vector $\mathbf{z}$, the above expression generalizes into the following approximation of the covariance matrix:

$$\mathsf{V}\left[\log(\mathbf{z})\right] \approx \frac{\mathsf{V}\left[\mathbf{z}\right]}{\mathsf{E}\left[\mathbf{z}\right]\mathsf{E}\left[\mathbf{z}\right]^T} \tag{11}$$

where the division is element-wise. Consequently we can approximate the covariance matrix of the MFCC estimator as:

$$\begin{aligned}\mathsf{V}\left[\hat{\mathbf{c}}_\mathbf{M}\right] &= \mathsf{V}\left[\frac{1}{m}\mathbf{\Phi}^H \log\left(\mathbf{M}\hat{\mathbf{s}}\right)\right] \\ &\approx \frac{1}{m^2}\mathbf{\Phi}^H \frac{\mathbf{M}\mathsf{V}\left[\hat{\mathbf{s}}\right]\mathbf{M}^T}{\mathbf{M}\mathsf{E}\left[\hat{\mathbf{s}}\right]\mathsf{E}\left[\hat{\mathbf{s}}\right]^T\mathbf{M}^T}\mathbf{\Phi}. \end{aligned} \tag{12}$$

Using (10) and (12), we can approximate the MSE of each of the different coefficients for any given Gaussian random process and for any given set of multitapers in our general MFCC estimator by the equation given in (5).

B. Confirmation of the proposed approximate formulas

In this section, we will demonstrate the accuracy of the proposed approximate formulas (10) and (12) by comparing them with Monte Carlo computations. The bias and variance of the MFCC estimator can be Monte Carlo computed for any given random process that we can simulate realizations of, and for any given set of multitapers. The number of simulations is chosen large enough (100 000) for the Monte Carlo error to be negligible. We choose a Gaussian AR(10)-process with parameters estimated from a recorded /a/ in the Swedish word "Hallå", $n = 240$, with sampling frequency $f_s = 8$ kHz, and we choose to use the ordinary mel-scale filterbank with $m = 27$ [3]. Computations are made for a Hamming window and for a peak matched multitaper with $k = 12$ windows [7]. The result is shown in Fig. 1. In this example, the approximation is very accurate. Indeed, the approximations are so close to the true values that it is even difficult to separate the lines. Similar observations were made also for other AR-processes using Hanning, rectangular, Thomson, sine, and peak matched multitapers with different number of tapers and with and without the mel-scale filterbank. One also notes that although the bias is, in general, larger for the peak matched multitaper, the variance is smaller, resulting in a smaller MSE.

The bias and variance depend on the random process, the multitapers, the mel-filter bank and the coefficient number. This may be compared with the rough approximations in [2], where it is stated that the bias of the cepstrum is asymptotically zero and the variance of the cepstrum is $\frac{\pi^2}{6n} \approx 0.007$ for all coefficients, where $n = 240$ is the frame length.

IV. Performance on Speech-like random processes

Using the proposed formulas derived in Section III-A, we can approximate the bias, variance and MSE of each coefficient in the MFCC estimator for a given random process

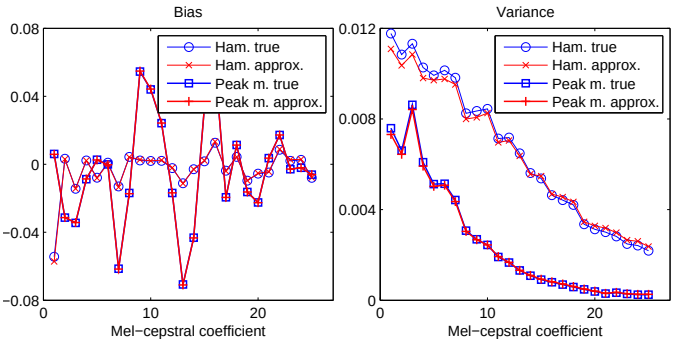

Fig. 1. This figure shows how accurate the proposed approximate formulas of bias and variance are. In the left plot, the approximation of the bias and the true bias (Monte Carlo computed) is shown when the MFCC is estimated using a Hamming window and peak matched multitapers with $k = 12$ for a Gaussian AR(10) process. In the right plot, the approximate variance of the MFCC estimated using a Hamming window and peak matched multitapers with $k = 12$ for a Gaussian AR(10) process is compared to the true variance. As seen the approximations are accurate.

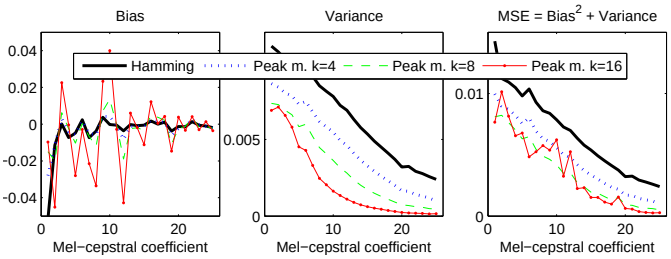

Fig. 2. Bias, variance and MSE (squared bias plus variance) of the MFCC estimator on speech-like random processes when Hamming window and peak matched multitapers ($k = 4,\ 8,\ 16$) are used.

and for a given multitaper and mel-filterbank. For further investigation, we take a set of 50 different recordings of the /a/ in the Swedish word "Hallå", $n = 240$ and $f_s = 8$ kHz, and model each of these recordings both as an AR(10) process and an AR(20) process. Similarly, we choose a set of 50 different AR(10) processes and a set of 50 different AR(20) processes with parameters estimated from the /l/ in the same word. For these four sets of processes, we compute the average bias, variance and MSE of each coefficient in the MFCC estimator for the following methods: Hamming window, Thomson multitapers, sine multitapers and peak matched multitapers, all with $k = 2, 4, \ldots, 16$. The Thomson and sine multitapers are commonly used, and peak matched multitapers are designed for peaked spectra, which may be suitable for speech analysis.

Fig. 2 shows the results of the peak matched multitapers with $k = 4, 8$ and 16, and the Hamming window, averaged over the set of 50 different AR(10) models of the /a/. Generally, the bias is larger and the variance lower when more multitapers are used. This is expected since averaging over more subspectra corresponds to more smoothing of the spectrum estimate [5]. Even though it is possible to use different estimators for different cepstral coefficients, it seems that peak matched multitapers with k between 8 and 16, represent a good trade-off between bias and variance for most cepstral coefficients. We got similar results for the Thomson and sine multitapers and also for the AR(20) models and for

the models of the /l/.

V. Speaker Verification Experiments

In speaker verification, the MFCCs are usually estimated using a Hamming-windowed periodogram, although multitapers provide smaller MSE on speech-like random process as seen above. To study whether this advantage carries on to a full recognition system, we consider the core task of the NIST 2006 SRE corpus[1]. It contains conversational telephony speech from 816 target speakers with 5077 genuine and 48,889 impostor verification trials. The length of speech data for training and testing is about 2.5 minutes.

Based on the result from Section IV, we chose to compare the conventionally used Hamming window with peak matched multitapers, $k = 12$. A more thorough evaluation of different multitapers and databases will be the topic of future work. MFCCs are extracted from 30 msec frames ($f_s = 8$ kHz, $n = 240$). Depending on the method, the frame is first processed either by a single Hamming window or by $k = 12$ peak matched multitapers, followed by 27-channel mel-frequency warped filterbank, log-compression and DCT. Twelve cepstral coefficients are retained. *RelAtive SpecTrAl* (RASTA) filtering is used for reducing channel effects. Delta and double delta coefficients are then added followed by voice activity detection (VAD) and utterance level cepstral mean and variance normalization (CMVN).

We use a standard *Gaussian mixture model* with universal background model (GMM-UBM) [11] and a *generalized linear discriminant sequence kernel support vector machine* (GLDS-SVM) [12] for classification. The background modeling data were taken from the NIST 2004 corpus. For more details of the system setup, refer to [13], [14].

We use two standard metrics to assess recognition accuracy: equal error rate (EER) and minimum detection cost function value (MinDCF). EER corresponds to the threshold at which the miss rate (P_{miss}) and false alarm rate (P_{fa}) are equal; MinDCF is the minimum value of a weighted cost function given by $0.1 \times P_{\text{miss}} + 0.99 \times P_{\text{fa}}$. In addition, we plot detection error tradeoff (DET) curves which show the full trade-off curve between false alarms and misses in a normal deviate scale, see Fig. 3. The accuracies for the Hamming window based MFCCs estimator and peak matched multitaper estimator are close to each other in the case of GMM-UBM system. For the support vector classifier, however, the peak matched multitaper estimator outperforms the Hamming window based estimator.

VI. Conclusions

The MFCC can be estimated from a Hamming-windowed periodogram or by using a multitaper spectrum estimator. We proposed new approximate formulas for the bias and variance of these MFCC estimators. Moreover, we demonstrated that these approximations are accurate. On a set of processes similar to the phoneme /a/ we showed that the peak matched MFCC estimate has lower MSE than the popular Hamming window. The result was the same for the phoneme /l/, indicating the robustness of the multitaper estimator for speech-like processes. We also demonstrated that the peak matched MFCC performs slightly better than the Hamming window MFCC in the NIST 2006 SRE.

[1] http://www.itl.nist.gov/iad/mig/tests/sre/2006/index.html

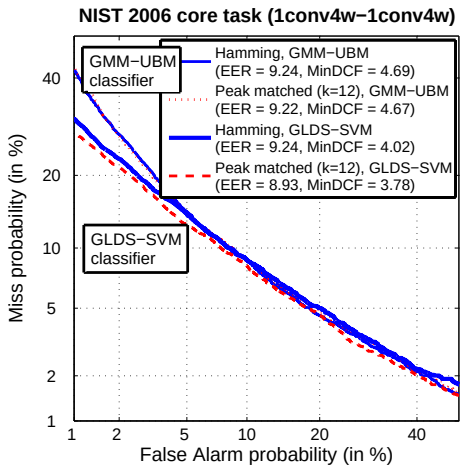

Fig. 3. Recognition accuracy on the NIST 2006 speaker recognition corpus.

References

[1] B. P. Bogert, M. J. R. Healy, and J. W. Tukey, "The quefrency alanysis of time series for echoes: Cepstrum, pseudo-autocovariance, cross-cepstrum and saphe cracking," in *Proc. of the Symposium on Time Series Analysis*, M. Rosenblatt, Ed., chapter 15. John Wiley, 1963.

[2] P. Stoica and N. Sandgren, "Smoothed nonparametric spectral estimation via cepsturm thresholding - introduction of a method for smoothed nonparametric spectral estimation," *IEEE Signal Proc. Magazine*, vol. 23, no. 6, pp. 34–45, Nov 2006.

[3] T. F. Quatieri, *Discrete-Time Speech Signal Processing*, Prentice Hall, 2002.

[4] D. B. Percival and A. T. Walden, *Spectral Analysis for Physical Applications*, Cambridge University Press, 1993.

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

[6] K. S. Riedel and A. Sidorenko, "Minimum bias multiple taper spectral estimation," *IEEE Trans. on Signal Proc.*, vol. 43, no. 1, pp. 188–195, Jan 1995.

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

[8] L. P. Ricotti, "Multitapering and a wavelet variant of MFCC in speech recognition," *IEE Proc. Vision, Image and Signal Proc.*, vol. 152, no. 1, pp. 29–35, Feb 2005.

[9] T. Gerkmann and R. Martin, "On the statistics of spectral amplitudes after variance reduction by temporal cepstrum smoothing and cepstral nulling," *IEEE Trans. on Signal Proc.*, vol. 57, no. 11, pp. 4165–4174, Nov 2009.

[10] Y. Ephraim and M. Rahim, "On second-order statistics and linear estimation of cepstral coefficients," *IEEE Trans. on Speech and Audio Proc.*, vol. 7, no. 2, pp. 162–176, Mar 1999.

[11] D.A. Reynolds, T.F. Quatieri, and R.B. Dunn, "Speaker verification using adapted Gaussian mixture models," *Digital Signal Proc.*, vol. 10, no. 1, pp. 19–41, Jan 2000.

[12] W.M. Campbell, J.P. Campbell, D.A. Reynolds, E. Singer, and P.A. Torres-Carrasquillo, "Support vector machines for speaker and language recognition," *Computer Speech and Language*, vol. 20, no. 2-3, pp. 210–229, April 2006.

[13] T. Kinnunen, J. Saastamoinen, V. Hautamäki, M. Vinni, and P. Fränti, "Comparative evaluation of maximum *a Posteriori* vector quantization and Gaussian mixture models in speaker verification," *Pattern Recognition Letters*, vol. 30, no. 4, pp. 341–347, March 2009.

[14] R. Saeidi, H.R.S. Mohammadi, T. Ganchev, and R.D. Rodman, "Particle swarm optimization for sorted adapted Gaussian mixture models," *IEEE Trans. Audio, Speech and Lang. Proc.*, vol. 17, no. 2, pp. 344–353, Feb 2009.