BLIND AUDIO SOURCE SEPARATION EXPLOITING PERIODICITY AND SPECTRAL ENVELOPES

Siouar Bensaid, Dirk Slock

EURECOM, Mobile Communications Dept.
2229 Route des Crêtes, BP 193, 06904 Sophia Antipolis Cedex, France
Email: {siouar.bensaid, dirk.slock}@eurecom.fr

ABSTRACT

In this paper we focus on the use of windows in the frequency domain processing of data for the purpose of spectral parameter estimation. Classical frequency domain asymptotics replace linear convolution by circulant convolution leading to approximation errors. We show how the introduction of windows can lead to slightly more complex frequency domain techniques, replacing diagonal matrices by banded matrices, but with controlled approximation error. We focus on the estimation of zero mean Gaussian data with a parametric spectrum model and show the equivalence of three approximation/estimation criteria: Itakura-Saito distance (ISD), Gaussian Maximum Likelihood (GML) and Optimally Weighted Covariance Matching (OWCM). We specialize the discussion to the case of single microphone based separation of quasiperiodic sources with AR spectral envelope.

Index Terms— Gaussian ML, Itakura-Saito, Optimally Weighted Covariance Matching, AR modeling, audio source separation, window, periodogram.

1. INTRODUCTION

Audio signal quasi-periodicity and spectral information have been widely exploited to perform speech enhancement. In fact, in [1], Nehorai et al. propose a sinusoidal model based algorithm for enhancement of speech corrupted by additive white Gaussian noise. The enhancement is achieved by estimating the sinusoidal model parameters which consist of the fundamental frequency (pitch), amplitudes and phases. The fundamental frequency (nonlinear parameter) is estimated using the recursive prediction error adaptive comb filter; amplitudes and phases are estimated using the recursive least squares (RLS) algorithm. In [2], the sinusoidal model, corrupted by additive broadband noise, is used with smoothness constraints imposed on the model parameters. The smoothness condition is induced by the continuous and slow variations with time of the vocal tract transfer function and the pitch. Therefore, this algorithm is restricted only to the voiced speech, while in [3], a more general algorithm is proposed, using two filters jointly, one for enhancing voiced speech exploiting harmonicity, another for unvoiced speech.
In audio source separation, periodicity has been used exhaustively [4, 5, 6, 7, 8, 9, 10, 11].Specifically, in [5, 4], the authors consider a multipitch model for voiced speech (referred to also as the long-term model) and introduce a time-warping function which describes pitch variation with time. The separation is achieved by identifying this function and estimating the ML solution of the other usual parameters (amplitudes, phases, etc..). In [12], the short-term and long-term aspect of speech are jointly modeled. For more references about multipitch modeling and estimation, the reader can refer to [13].

In [14, 15], a joint autoregressive (AR) model (short- plus long-term (ST+LT)) was introduced for quasiperiodic sources. The long-term part allows to capture the quasiperiodicity (with possible imperfect correlation in time), while the short-term part allows to model the spectral envelope. The modeling of the power spectral density is important to allow power splitting between sources at overlapping harmonics in the source extraction operation. In [14, 15] Bayesian approaches were adopted for source and parameter estimation, using EM-Kalman and Variational Bayes techniques resp. In [16], the ST+LT AR models were used for mono-microphone source separation in the frequency domain. Using Gaussian source models, the source extraction is simply Linear MMSE (Wiener) estimation. In the parametric approach, the ST+LT AR parameters need to be estimated also. In [16], three criteria are formulated for the estimation of these parameters on the basis of one frame of data, the Itakura-Saito distance (ISD) and Optimally Weighted Spectrum Matching (OWSM) for matching the parametric observed spectrum and the observations periodogram. The third criterion is Gaussian Maximum Likelihood (GML cf.[17]). The gradients of these three criteria w.r.t. the AR parameters and hence their extrema are shown to be identical. The results in [16] are based on asymptotic frequency domain expressions that are only valid for extremely long frames. In this paper, we extend these results by accounting for the finite window length and by introducing advantageously a non-rectangular window. Non-trivial windows were also introduced in [15], for the different purposes of source extraction and parameter estimation, passing from time to frequency domain. The approach in [15] was based on Variational Bayes, in which sources and their parameters are estimated jointly in an alternating optimization fashion. Here we estimate the parameters separately from the sources (e.g. after elimination of the Gaussian sources from the likelihood function), as in [16]. Due to the introduction of the window, which already limits temporal correlation, we propose to replace the LT AR correlation coefficient by its maximum value 1. We reconsider the equivalence of the three criteria mentioned, but this time based on finite data vectors, for which in frequency domain we can no longer neglect the correlations between different frequencies (the goal of the window design will then be to limit these correlations). The equivalence of multivariate ISD and GML is straightforward [18] as we shall see. In the multivariate case, the OWSM results in Optimally Weighted Covariance Matching (OWCM)[19].OWCM is again shown to be equivalent to ISD and GML in terms of gradients.

EURECOM's research is partially supported by its industrial partners: BMW Group, Cisco Systems, Monaco Telecom, Orange, SAP, SFR, STMicroelectronics, Swisscom, Symantec, and also by the French ANR project DIONISOS and the EU FP7 project WHERE2.

2. WINDOWING FOR FRAME-BASED PROCESSING

The audio signals considered are by nature non-stationary. If we can consider the parameters constant during a short time, we can process the signal in frames (time segments), over which the signal can be considered stationary, which corresponds to time-invariant filtering. Many of the signal processing operations (e.g. linear time-invariant filtering and filter computation) could be largely simplified by passing to the frequency domain. However, transforming a frame of signal to the frequency domain directly via the DFT (FFT) leads to approximations due to the periodic extension of the frame assumption inherent in the DFT. We shall see later how we can improve these approximations. Just like the original data signal y_k will be cut into a series of windowed frames of length N, a bit like in the Welch method, a processed signal (e.g. extracted source) will be reconstructed by superposing its reconstructed windowed frame segments. Since the window needs to decay towards its edges, consecutive frames need to overlap. Let M be the hop size (time jump) from one frame to the next, then a perfect reconstruction (PR) window w_n requires

$$\sum_{i=-\infty}^{\infty} w_{n-iM} \;=\; 1\;,\quad \forall n \tag{1}$$

see the top figures in Fig. 1 for the cases of relative overlap of $(N-M)/N \;=\; 50\%, 75\%$ (both the individual windows and their sum are shown for a finite set of windows). Note that one could consider extensions to non-PR windows, in which the superposition of windowed signal frames could be followed by a zero-forcing rescaling with $1/(\sum_{i=-\infty}^{\infty} w_{t-iM})$ or (multi-window) MMSE versions thereof. An example of a PR window is a Hann (or raised cosine) window

$$w_t \;=\; \frac{1}{2}\left[1 - cos\left(2\pi\frac{t}{N}\right)\right]\;,\quad t = 0, 1, \ldots, N-1\,. \tag{2}$$

The continuity of the window at its edges can be expected to be reflected in the continuity of the reconstructed signal and help reduce blocking artifacts (musical noise). The motivations for the window design will be different however in the parameter estimation part as we shall see. In a separate approach for parameter estimation and source extraction, as considered here, different windows could be used in both parts.

3. EQUIVALENCE OF ISD, GML AND OWCM CRITERIA

In what follows we consider a vector of zero mean data Y of length N, with covariance matrix $\mathbf{R}$, and estimation on the basis of the simple sample covariance $\widehat{\mathbf{R}} \;=\; YY^H$. We consider the data Y to be circular complex Gaussian distributed. The covariance matrix $\mathbf{R}$ is parameterized by the vector θ: $\mathbf{R} \;=\; \mathbf{R}(\theta)$. In this paper the superscripts $.^*, .^T, .^H$ denote complex conjugate, transpose and Hermitian transpose respectively.

3.1. Itakura-Saito Distance (ISD)

The multivariate Itakura-Saito distance is based on the observation that for a nonnegative definite matrix $\mathbf{A}$, the tangent hyperplane to $\ln\det\mathbf{A}$ at $\mathbf{A} \;=\; \mathbf{I}_N$ is $\mathrm{tr}\,\{\mathbf{A} - \mathbf{I}_N\}$, where tr denotes trace and $\mathbf{I}_N$ is the identity matrix of size N. The concavity of $\ln\det(.)$ then leads to

$$\mathrm{tr}\,\{\mathbf{A} - \mathbf{I}_N\} - \ln\det\mathbf{A} \geq 0\;.$$

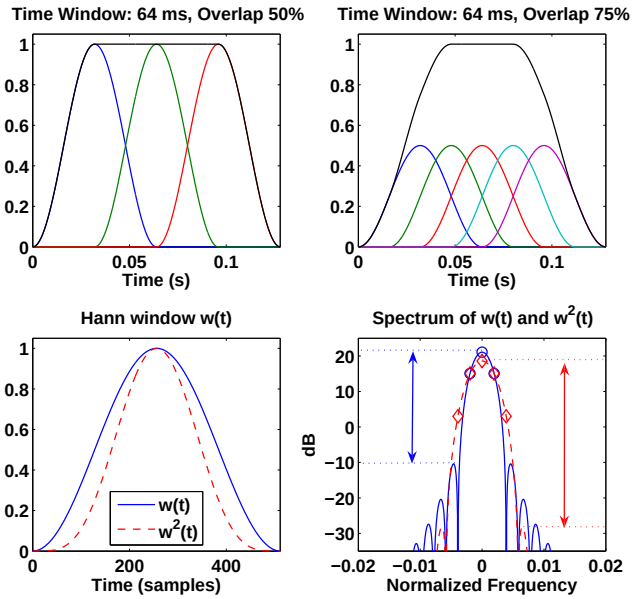

Fig. 1. Perfect reconstruction windowing.

The Itakura-Saito distance is obtained by taking the ratio of the two matrices to be compared $\mathbf{A} = \widehat{\mathbf{R}}\,\mathbf{R}^{-1}$:

$$ISD(\theta) = \mathrm{tr}\,\{\widehat{\mathbf{R}}\,\mathbf{R}^{-1} - \mathbf{I}_N\} - \ln\det(\widehat{\mathbf{R}}\,\mathbf{R}^{-1})\;. \tag{3}$$

3.2. Gaussian Maximum Likelihood (GML)

Assuming a circular complex Gaussian distribution, the negative log-likelihood becomes (apart from constants)

$$GML(\theta) = \ln\det(\mathbf{R}) + Y^H\mathbf{R}^{-1}Y\;. \tag{4}$$

Now note that using a property of the trace operator, $Y^H\mathbf{R}^{-1}Y \;=\; \mathrm{tr}\,\{Y^H\mathbf{R}^{-1}Y\} \;=\; \mathrm{tr}\,\{YY^H\mathbf{R}^{-1}\} \;=\; \mathrm{tr}\,\{\widehat{\mathbf{R}}\,\mathbf{R}^{-1}\}$. On the other hand, $\ln\det(\widehat{\mathbf{R}}\,\mathbf{R}^{-1}) = \ln\det(\widehat{\mathbf{R}}) - \ln\det(\mathbf{R})$. Hence, apart from constants ($\ln\det(\widehat{\mathbf{R}})$ being one of them), the IS and GML criteria are identical (in their dependence on θ). Note that the GML criterion only has an estimation motivation, whereas the IS (and hence GML also) performs jointly approximation and estimation. The approximation part refers to the fact that the true covariance matrix of Y may not be of the form $\mathbf{R}(\theta)$ for some θ, in which case minimizing the ISD will lead to a θ that best approximates the data.

3.3. Optimally Weighted Covariance Matching (OWCM)

OWCM is in fact optimally weighted least-squares applied to a sample covariance. Consider the $vec(.)$ operator which stacks the consecutive columns of a matrix into a vector. Then $vec(\widehat{\mathbf{R}}) \;=\; vec(YY^H) \;=\; Y^* \otimes Y$ where $\otimes$ denotes the Kronecker product. The mean of $Y^* \otimes Y$ is of course $vec(\mathbf{R})$. Using expressions for fourth moments of complex Gaussians, we get for its covariance matrix $\mathbf{R}^* \otimes \mathbf{R}$. The OWCM criterion is then

$$\begin{aligned} &OWCM(\theta) \\ &= (Y^* \otimes Y - vec(\mathbf{R}))^H(\mathbf{R}^* \otimes \mathbf{R})^{-1}(Y^* \otimes Y - vec(\mathbf{R})) \\ &= \mathrm{tr}\,\{(\widehat{\mathbf{R}} - \mathbf{R})\mathbf{R}^{-1}(\widehat{\mathbf{R}} - \mathbf{R})\mathbf{R}^{-1}\} \end{aligned} \tag{5}$$

Now, it is well-known that the weighting matrices $\mathbf{R}^{-1}$ can be replaced by consistent estimates without modifying the asymptotic covariance matrix of the estimation errors resulting from minimizing the OWCM criterion. Once the $\mathbf{R}^{-1}$ are replaced by a consistent estimate, they are no longer a function of θ. Now, taking the gradient of OWCM w.r.t. a parameter θ_i by only considering the $\mathbf{R}(\theta)$ appearing in the quadratic "numerator", we get

$$\frac{\partial\, OWCM(\theta)}{\partial \theta_i} = -2\,\mathrm{tr}\,\{\frac{\partial \mathbf{R}}{\partial \theta_i}\mathbf{R}^{-1}(\widehat{\mathbf{R}}-\mathbf{R})\mathbf{R}^{-1}\}\;. \tag{6}$$

On the other hand we get for $GML(\theta) = \ln\det(\mathbf{R}) + \mathrm{tr}\,\{\widehat{\mathbf{R}}\mathbf{R}^{-1}\}$ that

$$\begin{array}{rcl}\frac{\partial\, GML(\theta)}{\partial \theta_i} & = & \mathrm{tr}\,\{\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \theta_i}\} - \mathrm{tr}\,\{\widehat{\mathbf{R}}\,\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \theta_i}\mathbf{R}^{-1}\} \\ & = & -\,\mathrm{tr}\,\{\frac{\partial \mathbf{R}}{\partial \theta_i}\mathbf{R}^{-1}(\widehat{\mathbf{R}}-\mathbf{R})\mathbf{R}^{-1}\}\;.\end{array} \tag{7}$$

Comparing (6) and (7), we see that the extrema of $OWCM(\theta)$ and $GML(\theta)$ coincide.

4. GML APPLIED TO THE DATA DFT

Working in the time domain, we have a full covariance $\mathbf{R}$ to work with. By going to the frequency domain, one typically assumes to be able to work with a diagonal $\mathbf{R}$ because asymptotically, different frequency components are uncorrelated. We shall analyze more precisely the nonasymptotic regime. Because of the correspondence of the three criteria above, we shall henceforth only consider the GML criterion. Now, let the current frame of N samples be $\mathbf{y} = [y_0\, y_1 \cdots y_{N-1}]^T$ and w.l.o.g. we assumed that the first sample starts at time zero. Before applying the DFT, the data get windowed. Let $\mathbf{W} = \mathrm{diag}\,\{w_0, w_1, \ldots, w_{N-1}\}$ and $\mathbf{F}$ is the $N \times N$ discrete Fourier transform (DFT) matrix, with inverse DFT $\frac{1}{N}F^* = \frac{1}{N}F^H$. Then we shall work with the transformed windowed data vector

$$Y = \mathbf{F}\,\mathbf{W}\,\mathbf{y}\;. \tag{8}$$

The data are assumed to have zero mean so that covariance and correlation matrices are equal. Note now that $\mathbf{y}$ is real, but Y is complex due to the DFT. Y is strictly speaking non-circular as both $\mathbf{R} = \mathrm{E}\,YY^H$ and $\mathrm{E}\,YY^T$ are nonzero. However, Y is not a genuine complex random vector as only the real vector $\mathbf{y}$ is random and the complex aspect is due to a deterministic transformation. As a result we can continue as if Y has a circular complex Gaussian distribution (which corresponds to a real Gaussian distribution with transposes replaced by Hermitian transposes). Now, all we need for GML is $\mathbf{R}$. Note that component Y_k of $Y = [Y_0\, Y_1 \cdots Y_{N-1}]^T$ is in fact the discrete-time Fourier transform $\mathcal{F}$ (DTFT) $Y^w(f)$ of the windowed signal evaluated at frequency $f = k/N$. To constitute $\mathbf{R}$, we shall need the correlations between different frequencies $\mathrm{E}\,Y^w(f_1)\,Y^{w*}(f_2)$. For this consider

$$\begin{array}{l} Y^w(f_1) = \sum_{n=0}^{N-1} w_n\, y_n e^{-j2\pi f_1 n} = \sum_{n=-\infty}^{\infty} w_n e^{-j2\pi f_1 n}\, y_n \\ = \sum_{n=-\infty}^{\infty} h_{-n}\, y_n = h_n * y_n|_{n=0} \\ = \int_{-\frac{1}{2}}^{\frac{1}{2}} H(f)\, Y(f)\, df = \int W(f_1 - f)\, Y(f)\, df \end{array} \tag{9}$$

where we zeropadded the finite window to infinity. Now we get

$$\begin{array}{l} \mathrm{E}\, Y^w(f_1)\, Y^{w*}(f_2) \\ = \mathrm{E} \int W(f_1 - f)\, Y(f)\, df \int W^*(f_2 - f_0)\, Y^*(f_0)\, df_0 \\ = \int df\, W(f_1 - f) \int df_0 W^*(f_2 - f_0)\; \mathrm{E}\, Y(f) Y^*(f_0) \\ = \int df\, W(f_1 - f) \int df_0 W^*(f_2 - f_0)\, S_{yy}(f)\, \delta_1(f - f_0) \\ = \int df\, W(f_1 - f)\, W^*(f_2 - f)\, S_{yy}(f) \end{array} \tag{10}$$

where $Y(f) = \sum_{k=-\infty}^{\infty} y_k e^{-j2\pi fk}$ is the DTFT of the stationary random process y_n with spectrum $S_{yy}(f)$, $W(f)$ is the DTFT of the window w_n, and $\delta_1(f) = \sum_{k=-\infty}^{\infty} \delta(f-k)$ is the periodicized delta function. Now let us introduce the vector of DFT frequencies $\underline{f} = [0\, 1 \cdots N-1]^T/N$ and the $N \times 1$ vector of ones $\underline{1}$, let $W(\underline{f})$ denote the column vector of $W(.)$ evaluated at the components of $\underline{f}$, then we can write for

$$\mathbf{R} = \int df\, W(\underline{f} - f\underline{1})\, W^H(\underline{f} - f\underline{1})\, S_{yy}(f)\;. \tag{11}$$

We get in particular for the diagonal elements $\mathbf{R}_{kk} = \int df\, |W(\frac{k-1}{N} - f)|^2\, S_{yy}(f)$ which is the well-known spectrum smearing appearing in the mean of the periodogram. Now, to limit complexity in the frequency domain based methods, one should sparsify $\mathbf{R}$ as much as possible. Here is where the window design comes in. For a properly designed window, $W(f)$ can be neglected outside of its main lobe (see e.g. the lower right corner in Fig. 1). Note that from this point of view, a rectangular window is (again) not a very good choice since the sidelobes are not much attenuated. If Δf is the doublesided width of the main lobe of $W(f)$, then $\int df\, W(f_1 - f)\, W^*(f_2 - f)\, S_{yy}(f)$ can be approximated to zero for $|f_1 - f_2| > \Delta f$. This means that $\mathbf{R}$ can be approximated by a banded matrix with only $\lceil N\,\Delta f \rceil$ non-zero diagonals. E.g. the inversion of $\mathbf{R}$ can then be done efficiently using the LDU triangular factorization of $\mathbf{R}$ in which the triangular factors will also be banded. Compared to classical frequency-domain asymptotics, the spectrum gets smeared on the diagonal and spills onto the main sub- and super-diagonals, leading to correlations between neighboring frequencies (only). In those classical asymptotics, the smearing effect of $W(f)$ gets neglected, leading to $\mathbf{R} = \mathrm{diag}\,\{S_{yy}(\underline{f})\}$.

If $S_{yy}(f)$ is sufficiently smooth, the integral in (11) can be approximated by a sum over frequencies spaced more densely at $\underline{f}'$, containing multiples of $1/N'$, where $N' > N$. This can be obtained by zeropadding the signal from N to N' samples and applying the DFT of size N'. We then get $\mathbf{R}'$ of the form

$$\mathbf{R}' = \mathcal{C}(W(\underline{f}'))\;\mathrm{diag}\,\{S_{yy}(\underline{f}')\}\,\mathcal{C}^H(W(\underline{f}')) \tag{12}$$

where $\mathcal{C}$ denotes a circulant matrix constructed from the vector argument. The entries of $\mathbf{R}'$ can be downsampled to obtain $\mathbf{R}$ if desired.

5. FREQUENCY DOMAIN CRAMER-RAO BOUNDS (CRBS)

For a Gaussian process with zero mean, the element (i,j) (pertaining to θ_i and θ_j) of the Fisher Information Matrix (FIM) are obtained as

$$FIM_{i,j} = \mathrm{tr}\,\{\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \theta_i}\mathbf{R}^{-1}\frac{\partial \mathbf{R}}{\partial \theta_j}\}\;. \tag{13}$$

Here, $\mathbf{R}$ is given in (11) and we get for the derivatives

$$\frac{\partial \mathbf{R}}{\partial \theta_i} = \int df\, W(\underline{f} - f\underline{1})\, W^H(\underline{f} - f\underline{1})\, \frac{\partial S_{yy}(f)}{\partial \theta_i}\;. \tag{14}$$

In the classical asymptotics, the FIM gets then approximated as

$$\begin{aligned} FIM_{i,j} &= \int df\, S_{yy}^{-2} \frac{\partial S_{yy}(f)}{\partial \theta_i} \frac{\partial S_{yy}(f)}{\partial \theta_j} \\ &= \int df\, \frac{\partial \ln S_{yy}(f)}{\partial \theta_i} \frac{\partial \ln S_{yy}(f)}{\partial \theta_j} \ . \end{aligned} \tag{15}$$

6. PERIODIC SOURCES WITH ST AR SPECTRAL ENVELOPE

The single microphone measurement signal y_n is considered to be composed of K quasiperiodic sources $s_{k,n}$ plus noise v_k. Assuming stationarity, the spectrum $S(f)$ of y_n can be written as

$$S(f) = S_0(f) + \sum_{k=1}^{K} S_k(f) \ . \tag{16}$$

In the case of white noise, $S_0(f) = \sigma_v^2$. For quasiperiodic sources, which are observed over a limited time frame which is furthermore windowed with reduced weight towards the edges, we can neglect possible limited long-term correlation and model the source as a Gaussian periodic signal with ST AR spectral envelope, leading to a spectrum of the form

$$\begin{aligned} S_k(f) &= \frac{\sigma_k}{|A_k(f)|^2} \sum_{m=-\lfloor \frac{1}{2f_k} \rfloor}^{\lfloor \frac{1}{2f_k} \rfloor} \delta(f - m\, f_k) \\ &= \sigma_k^2 \sum_m \frac{1}{|A_k(m\, f_k)|^2} \delta(f - m\, f_k) \end{aligned} \tag{17}$$

where σ_k^2 adjusts the source power and f_k if the source pitch. We have for the AR spectral envelope the ST filter

$$A_k(f) = \sum_{i=0}^{L_k} a_{k,i} e^{j2\pi f\, i} \ , \ \text{ with } a_{k,0} = 1 \tag{18}$$

where L_k is the AR order of source k.

With the above signal model the parameters θ are $\{\sigma_v^2, a_{ki}, i = 1, \ldots, L_k,\ k = 1, \ldots, K\}$ and we get for $\mathbf{R}(\theta)$

$$\begin{aligned} &\mathbf{R}(\theta) = \sigma_v^2 \int df\, W(\underline{f} - f\underline{1})\, W^H(\underline{f} - f\underline{1}) \\ &+ \textstyle\sum_{k=1}^{K} \sum_m \frac{1}{|A_k(m\, f_k)|^2} W(\underline{f} - m\, f_k \underline{1})\, W^H(\underline{f} - m\, f_k \underline{1}) \ . \end{aligned} \tag{19}$$

7. REFERENCES

[1] A. Nehorai and B. Porat, "Adaptive comb filtering for harmonic signal enhancement," *Acoustics, Speech and Signal Processing, IEEE Trans. on*, Oct. 1986.

[2] J. Jensen and J.H.L. Hansen, "Speech enhancement using a constrained iterative sinusoidal model," *Speech and Audio Processing, IEEE Transactions on*, Oct. 2001.

[3] J.R. Jensen, J. Benesty, M.G. Christensen, and S.H. Jensen, "Enhancement of single-channel periodic signals in the time-domain," *Audio, Speech, and Language Processing, IEEE Trans. on*, Sept. 2012.

[4] Y. Stettiner, D. Malah, and D. Chazan, "Estimation of the parameters of a long-term model for accurate representation of voiced speech," in *Acoustics, Speech, and Signal Processing, 1993. ICASSP-93., 1993 IEEE Int'l Conf. on*, April 1993.

[5] D. Chazan, Y. Stettiner, and D. Malah, "Optimal multi-pitch estimation using the em algorithm for co-channel speech separation," in *Acoustics, Speech, and Signal Processing, 1993. ICASSP-93., 1993 IEEE Int'l Conf. on*, April 1993.

[6] T. Virtanen and A. Klapuri, "Separation of harmonic sound sources using sinusoidal modeling," in *Acoustics, Speech, and Signal Processing, 2000. ICASSP '00. Proceedings. 2000 IEEE Int'l Conf. on*, 2000, vol. 2.

[7] P. Mowlaee, M.G. Christensen, and S.H. Jensen, "Improved single-channel speech separation using sinusoidal modeling," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE Int'l Conf. on*, March 2010.

[8] P. Mowlaee, R. Saeidi, Z.-H. Tan, M.G. Christensen, P. Fränti, and S.H. Jensen, "Joint single-channel speech separation and speaker identification," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE Int'l Conf. on*, March 2010.

[9] M.G. Christensen and A. Jakobsson, "Optimal filter designs for separating and enhancing periodic signals," *Signal Processing, IEEE Trans. on*, Dec. 2010.

[10] J.R. Jensen, M.G. Christensen, and S.H. Jensen, "An optimal spatio-temporal filter for extraction and enhancement of multi-channel periodic signals," in *Signals, Systems and Computers (ASILOMAR), 2010 Conference Record of the Forty Fourth Asilomar Conf. on*, Nov. 2010.

[11] Pejman Mowlaee, Rahim Saeidi, Zheng-Hua Tan, Mads Græsbøll Christensen, Tomi Kinnunen, Pasi Fränti, and Søren Holdt Jensen, "Sinusoidal approach for the single-channel speech separation and recognition challenge," in *INTERSPEECH*, 2011.

[12] D. Giacobello, M.G. Christensen, J. Dahl, S.H. Jensen, and M. Moonen, "Joint estimation of short-term and long-term predictors in speech coders," in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE Int'l Conf. on*, April 2009.

[13] Mads Christensen and Andreas Jakobsson, *Multi-Pitch Estimation*, Morgan and Claypool Publishers, 2009.

[14] Siouar Bensaid, Antony Schutz, and Dirk T M Slock, "Single Microphone Blind Audio Source Separation Using EM-Kalman Filter and Short+Long Term AR Modeling," in *LVA ICA 2010, 9th Int'l Conf. on Latent Variable Analysis and Signal Separation, September 27-30, Saint-Malo, France / Also published in LNCS, 2010*, 2010.

[15] Antony Schutz and Dirk T M Slock, "Single-microphone blind audio source separation via Gaussian Short+Long Term AR Models," in *ISCCSP 2010, 4th Int'l Symposium on Communications, Control and Signal Processing, March 3-5, 2010, Limassol, Cyprus*, 2010.

[16] Antony Schutz and Dirk T M Slock, "Blind audio source separation using Short+Long Term AR source models and spectrum matching," in *DSP/SPE 2011, 14th IEEE Digital Signal Processing ; 6th Signal Processing Education Workshop, January 4-7, 2011, Sedona, Arizona, USA*, 2011.

[17] E. de Carvalho and D.T.M. Slock, "A fast gaussian maximum-likelihood method for blind multichannel estimation," in *Signal Processing Advances in Wireless Communications, 1999. SPAWC '99. 2nd IEEE Workshop on*, 1999.

[18] B.A. Carlson and M.A. Clements, "A computationally compact divergence measure for speech processing," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1991.

[19] B Ottersten, P Stoica, and R Roy, "Covariance matching estimation techniques for array signal processing applications," *Digital Signal Processing*, , no. 3, 1998.