Aalborg Universitet

AALBORG UNIVERSITY
DENMARK

The Cyclic Matching Pursuit and Its Application to Audio Modeling and Coding

Christensen, Mads Græsbøll; Jensen, Søren Holdt

Published in:
Proc. of Asilomar Conference on Signals, Systems, and Computers

Publication date:
2007

Document Version
Accepted author manuscript, peer reviewed version

Link to publication from Aalborg University

Citation for published version (APA):
Christensen, M. G., & Jensen, S. H. (2007). The Cyclic Matching Pursuit and Its Application to Audio Modeling and Coding. *Proc. of Asilomar Conference on Signals, Systems, and Computers*, 550-554. http://kom.aau.dk/~mgc/publications/asilomar2007b.pdf

THE CYCLIC MATCHING PURSUIT AND ITS APPLICATION TO AUDIO MODELING AND CODING

Mads Græsbøll Christensen[*] *and Søren Holdt Jensen*

Dept. of Electronic Systems
Aalborg University, Denmark
{mgc,shj}@es.aau.dk

ABSTRACT

In this paper, we propose a method, named the cyclic matching pursuit, that outperforms the standard matching pursuit and the perceptual matching pursuit while preserving the computationally efficiency of those methods. We exemplify the application of the method to audio modeling and coding using a perceptual distortion measure and demonstrate using audio signals that the method leads to improved modeling capabilities.

1. INTRODUCTION

The problem of decomposing a segment of data into a linear combination of some bases occurs in many applications. It occurs, for example, in speech and audio applications involving modeling and coding where so-called atomic decompositions have proven to be useful. In such applications it is important that the components are chosen such that they best describe the signal which in audio modeling means that the perceptual distortion is minimized. If the signal is decomposed into L components that are found such that a perceptual distortion measure is minimized, we can claim in audio coding applications that at a given bit-rate (a given number of components), the best possible performance is achieved. In complexity constrained audio modeling, it is likewise desirable the allowable number of components is spent the best way possible. In audio modeling and coding, sinusoidal models have proven to be an efficient representation of stationary, tonal parts. In [1], a framework for perceptual distortion minimization and sinusoidal component selection, i.e. frequency estimation, was presented based on the distortion measure presented in [2]. Within this framework, a number of well-known practical, but suboptimal, methods for decomposing signals into sinusoids, namely the weighted matching pursuit (WMP) [3], the pre-filtering method [4], and the perceptual matching pursuit (PMP) [5] were related to each other and the optimal solution. Many different improvements over the original matching pursuit [6] have been proposed in recent years, like the forward and backward orthogonal matching pursuits (see., e.g., [7, 8, 9, 10]). While these methods improve upon the performance, i.e., achieve lower distortions, of matching pursuit, the improvements usually come at the price of a significant increase in computational complexity. Typically, the elegant and computationally simple calculations associated with finding the inner products in the original matching pursuit are lost in the process. Also, these methods suffer from the problem that once atoms have been chosen, the are fixed in later iterations. In the case of sinusoidal modeling, this means that once the frequency of a sinusoids has been selected, it can no longer be changed. This means that if a biased estimate is obtained in an early iteration of such an algorithm, it is never able to recover. For an overview of these methods and their properties, we refer to [10] and the referenced therein.

In this paper, we propose a new method based on the framework of [1] called the cyclic matching pursuit (CMP) for decomposing signals into a linear combination of bases. Although derived in a particular context, the method may be easily generalized to minimizing any distortion measure that is induced by an inner product and any kind of dictionary. The method retains the simplicity of the original matching pursuit for the 2-norm and the perceptual matching pursuit for the perceptually weighted norm with the computational complexity of the proposed method being proportional to the complexity of those methods. The method is easy to implement and does not require the notoriously difficult multidimensional nonlinear optimization that may otherwise be required. Also, the method is able to overcome an important weakness of the forward and backward orthogonal matching pursuits, namely that it can compensate for biases introduced in early iterations.

The remaining part of this paper is organized as follows: First, we briefly review the framework of [1] in Section 2. In Section 3 we then proceed to present the new method, i.e., the cyclic matching pursuit, before we give an illustrative signal example in 4. Finally, we conclude on the work in Section 5.

2. FRAMEWORK

For simplicity, we will make use the complex notation and signals. Consequently, we start out by calculating the N so-called downsampled discrete-time analytic signal samples $x(n)$ from $2N$ real input samples $y(n)$. The analytic signal is defined as $\zeta(n) = y(n) + j\mathcal{H}\{y(n)\}$ where $\mathcal{H}\{\cdot\}$ denotes the Hilbert transform. $x(n)$ is then obtained as $x(n) = \zeta(2n)$ for $n = 0, \ldots, N-1$. This representation is valid for large N.

Using the auditory masking model proposed for sinusoidal audio coding in [2] the distortion D for a particular segment can be written as

$$D = \sum_{k=0}^{K-1} P(k)|E(k)|^2, \qquad (1)$$

where $P(k)$ is a frequency domain real, positive weighting function and $E(k) = \sum_{n=0}^{N-1} w(n)\left[x(n) - \hat{x}(n)\right]\exp(-j2\pi k/Kn)$ is the K point Fourier transform (with $K \geq N$) of the weighted

[*] The work of M. G. Christensen is supported by the Parametric Audio Processing project, Danish Research Council for Technology and Production Sciences grant no. 274–06–0521.

reconstruction error with $\hat{x}(n)$ being the reconstructed signal and $w(n)$ the analysis/synthesis window. In the following discussion, we assume a rectangular window because the relations between the approximate methods and the exact methods to be discussed do not hold otherwise. However, the use of a different window can easily be incorporated in both the PMP and the proposed CMP by applying the window to $x(n)$ and $\hat{x}(n)$. The perceptual weighting function is chosen as the reciprocal of the masking curve which is calculated using the model proposed in [2]. Defining $\mathbf{x} = [\, x(0) \; \cdots \; x(N-1) \,]^T$, $\hat{\mathbf{x}} = [\, \hat{x}(0) \; \cdots \; \hat{x}(N-1) \,]^T$ and assuming that $K > N$, the distortion measure can be put into matrix-vector notation, i.e,

$$D = \left\| \mathbf{H} \begin{bmatrix} \mathbf{e} \\ \mathbf{0} \end{bmatrix} \right\|_2^2 = \left\| \mathbf{H} \left(\begin{bmatrix} \mathbf{x} \\ \mathbf{0} \end{bmatrix} - \begin{bmatrix} \hat{\mathbf{x}} \\ \mathbf{0} \end{bmatrix} \right) \right\|_2^2, \qquad (2)$$

where $\mathbf{e} = \mathbf{x} - \hat{\mathbf{x}}$ is the error signal vector. The matrix $\mathbf{H}$ is the perceptual weighting matrix having the following structure

$$\mathbf{H} = \begin{bmatrix} h(0) & h(K-1) & \cdots & h(1) \\ h(1) & h(0) & \cdots & h(K-1) \\ \vdots & \vdots & \ddots & \vdots \\ h(K-1) & h(K-2) & \cdots & h(0) \end{bmatrix}, \qquad (3)$$

with $h(n) = \frac{1}{K} \sum_{k=0}^{K-1} \sqrt{P(k)} \exp\left(j2\pi kn/K\right)$. The distortion in (2) can also be written as

$$\left\| \mathbf{H} \left(\begin{bmatrix} \mathbf{x} \\ \mathbf{0} \end{bmatrix} - \begin{bmatrix} \hat{\mathbf{x}} \\ \mathbf{0} \end{bmatrix} \right) \right\|_2^2 = \left\| \bar{\mathbf{H}} \left(\mathbf{x} - \hat{\mathbf{x}} \right) \right\|_2^2, \qquad (4)$$

where $\bar{\mathbf{H}}$ is a $K \times N$ matrix containing the N first columns of $\mathbf{H}$. We note in passing that $\bar{\mathbf{H}}$ is still circulant but not square. $\bar{\mathbf{H}}^H \bar{\mathbf{H}}$, on the other hand, is not circulant, unlike $\mathbf{H}^H \mathbf{H}$, but still Toeplitz. For $\bar{\mathbf{H}} = \mathbf{I}$ where $\mathbf{I}$ is the identity matrix, the distortion measure is the usual 2-norm. As can be seen, the perceptual distortion measure can be interpreted as a particular kind of linear transform, namely a linear filter. Interestingly, the eigenvectors of such a matrix are the Fourier basis vectors and asymptotically sinusoids of arbitrary frequency are eigenvectors of this matrix. The circulant matrix $\mathbf{H} \in \mathbb{R}^{K \times K}$, is defined by its first column $\mathbf{h} = [\, h(0) \; \cdots \; h(K-1) \,]^T$. Next, defining the discrete Fourier transform (DFT) matrix as

$$\mathbf{F} = \frac{1}{\sqrt{K}} \left[\; \mathbf{f}_0 \quad \mathbf{f}_1 \quad \cdots \quad \mathbf{f}_{K-1} \;\right], \qquad (5)$$

with the kth vector $\mathbf{f}_k = \left[\; f_k^0 \; \cdots \; f_k^{K-1} \right]^T$ being composed from

$$f_k = \exp(-j2\pi k/K) \qquad (6)$$

Furthermore, setting $\mathbf{Q} = \mathbf{F}^H$ and $\mathbf{\Lambda} = \sqrt{K}\,\mathrm{diag}(\mathbf{F}\mathbf{h})$, the eigenvalue decomposition (EVD) of $\mathbf{H}$ can then be written as

$$\mathbf{H} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^H. \qquad (7)$$

We see that the function of the perceptual weighting matrix can be seen as an orthogonal transformation followed by a weighting. This is the result that leads to the equivalence of a number of methods in [1] both asymptotically as well as in special cases for a finite number of samples.

The frequency estimation problem can now be stated. Given a real observed signal $x(n)$ for $n = 0, \ldots, N-1$, find the parameters of the signal of interest $\hat{x}(n)$ in additive noise $e(n)$:

$$x(n) = \hat{x}(n) + e(n). \qquad (8)$$

In our case the signal of interest $\hat{x}(n)$ is a sum of sinusoidal components, i.e.,

$$\hat{x}(n) = \sum_{l=1}^{L} a_l \exp\left(j\omega_l n\right), \qquad (9)$$

where $a_l = A_l \exp\left(j\phi_l\right)$ with each component having an amplitude A_l, phase ϕ_l, and frequency ω_l. The difficult part is finding the nonlinear parameters, i.e., the frequencies. The perceptual nonlinear least-squares estimates of the frequencies $\{\omega_l\}_{l=1}^{L}$ are

$$\{\hat{\omega}_l\} = \arg\min_{\{\omega_l\}} \|\bar{\mathbf{H}}(\mathbf{x} - \mathbf{Z}\mathbf{a})\|_2^2. \qquad (10)$$

The matrix $\mathbf{Z}$ is defined from $z_l = \exp(j\omega_l)$ as

$$\mathbf{Z} = \begin{bmatrix} z_1^0 & \cdots & z_L^0 \\ z_1^1 & \cdots & z_L^1 \\ \vdots & & \vdots \\ z_1^{N-1} & \cdots & z_L^{N-1} \end{bmatrix}. \qquad (11)$$

Furthermore, we have that $\mathbf{a} = [\, a_1 \; \cdots \; a_L \,]^T$. For convenience, we introduce $\bar{\mathbf{W}} = \bar{\mathbf{H}}^H \bar{\mathbf{H}}$. The complex amplitudes can be estimated given the frequencies $\{\omega_l\}$ using linear least-squares as

$$\hat{\mathbf{a}} = \left(\mathbf{Z}^H \bar{\mathbf{W}} \mathbf{Z} \right)^{-1} \mathbf{Z}^H \bar{\mathbf{W}} \mathbf{x}. \qquad (12)$$

We can now write the estimates as the set of frequencies that minimize the perceptual distortion, i.e.,

$$\{\hat{\omega}\} = \arg\max_{\boldsymbol{\omega}} \mathbf{x}^H \bar{\mathbf{W}} \mathbf{Z} \left(\mathbf{Z}^H \bar{\mathbf{W}} \mathbf{Z} \right)^{-1} \mathbf{Z}^H \bar{\mathbf{W}} \mathbf{x}. \qquad (13)$$

However, solving this is not computationally feasible due to the nonlinear nature of the frequencies. Instead, an iterative estimation procedure is often used. We will now proceed to describe a number of such methods that have been reported in the literature within the framework of the minimization of the perceptual distortion measure defined in (1). First, we define the residual vector at iteration l as $\mathbf{r}_l = [\, r_l(0) \; \cdots \; r_l(N-1) \,]^T$ with $r_{l+1}(n) = r_l(n) - \hat{a}_l \exp\left(j\hat{\omega}_l n\right)$ which is initialized as $r_1(n) = x(n)$. In the perceptual matching pursuit [11], sinusoids are chosen iteratively one at the time as the minimizer of the perceptual distortion of this residual, i.e.,

$$\hat{\omega}_l = \arg\min_{\omega} \|\bar{\mathbf{H}}[(\mathbf{r}_l - \mathbf{z}a)\|_2^2. \qquad (14)$$

with $\mathbf{z} = [\, \exp(j\omega 0) \; \cdots \; \exp(j\omega(N-1)) \,]^T$. This results in the following frequency estimation criterion[1]:

$$\hat{\omega}_l = \arg\max_{\omega} \frac{|\left\langle \bar{\mathbf{H}}\mathbf{z}, \bar{\mathbf{H}}\mathbf{r}_l \right\rangle|^2}{\|\bar{\mathbf{H}}\mathbf{z}\|_2^2} \qquad (15)$$

with $\langle \mathbf{x}, \mathbf{y} \rangle$ denoting the usual discrete inner product[2] $\mathbf{x}^H\mathbf{y}$ that induces the 2-norm $\|\cdot\|_2$. The estimates can be obtained using two FFTs per iteration. Consider now that we choose the signal model component $\mathbf{z}$ such that it is an eigenvector of the perceptual weighting matrix or at least a good approximation, i.e,

$$\bar{\mathbf{H}}\mathbf{z} = \lambda\mathbf{z}. \qquad (16)$$

[1] We here ignore the amplitude estimates since these can be found from the same inner products that are used in the frequency estimates.

[2] It is also possible to redefine the inner product such that it induces the perceptually weighted norm.

In the following we assume that $K = N$, i.e., $\bar{\mathbf{H}} = \mathbf{H}$ since it cannot have an EVD otherwise. It could be seen from (7) that the perceptual weighting matrix may be seen as a unitary transformation followed by a weighting (the eigenvalues) of the individual directions (the eigenvectors). From this perspective, (16) can be interpreted as the special property of the chosen model that it is invariant to the unitary transformation of the perceptual weighting matrix. Using these observations, the frequency estimation criterion of the PMP can be reduced to the so-called pre-filtering method that has been applied to th perceptual frequency estimation problem and audio coding in [4]. Specifically, the signal is filtered before estimation and quantization, i.e.,

$$\hat{\omega}_l = \arg\min_{\omega} \|\bar{\mathbf{H}}\mathbf{r}_l - \lambda \mathbf{z} a\|_2^2 = \arg\max_{\omega} \frac{\left|\langle \mathbf{z}, \bar{\mathbf{H}}\mathbf{r}_l \rangle\right|^2}{N}. \quad (17)$$

This estimator can obviously be implemented efficiently using an FFT of the pre-filtered signal. The pre-filtering can of course also be implemented this way. We see that the complexity of the method can be greatly reduced this way. Next, we note that the inner product can be written as $\langle \mathbf{z}, \mathbf{H}\mathbf{r}_l \rangle = \lambda^* \mathbf{v}^H \mathbf{r}_l$, whereby the frequency estimation criterion then becomes

$$\hat{\omega}_l = \arg\max_{\omega} |\lambda|^2 \frac{\left|\langle \mathbf{z}, \mathbf{r}_l \rangle\right|^2}{N}. \quad (18)$$

We see that λ can be interpreted as a frequency dependent weighting, and the estimation criterion in (18) is therefore identical to the weighted matching pursuit proposed in [3] which also can be implemented in a simple way.

Comparing the optimal estimator (10) with the iterative, sub-optimal approximations in (15), (17), and (18), we can give some insights into in what cases the estimators may give identical results and in what cases they may differ. For a distinct set of frequencies and a large number of samples, the estimators can be expected to yield similar results since the interactions between the individual components will be come smaller as N grows. Therefore, one would expect that the estimates will differ when N is small or the sinusoids are not well-separated in frequencies. This happens, for example, for transients signals or for complicated mixtures of signals with many harmonics. Similarly, it can be expected that the improvement also will depend on the number of sinusoids that are to be extracted.

3. THE CYCLIC MATCHING PURSUIT

We will now proceed to propose a new method, called cyclic matching pursuit (CMP), for minimization of the perceptual distortion measure defined in (1). Although derived in this particular context, the method may be easily generalized to minimizing any distortion measure that is induced by an inner product. The methods retains the simplicity of the original matching pursuit for the 2-norm and the perceptual matching pursuit for the perceptually weighted norm and is thus easy to implement. We note in passing that the approach introduced next is conceptually reminiscent of iterative techniques found in estimation theory such as those in [12, 13]. For more on such iterative methods, we refer the interested reader to [14] and the references therein.

First we will describe the new algorithm in general terms before showing the exact equations. Let $\theta_l = (\omega_l, a_l)$ denote the parameters associated with the lth complex sinusoid and let

$$D(\theta_1, \dots, \theta_L) \quad (19)$$

denote the distortion that results from synthesizing the signal using the parameters $\boldsymbol{\Theta} = \{\theta_l\}_{l=1}^{L}$. Using this notation, we can write the iterations of the matching pursuit as

$$\begin{aligned} \hat{\theta}_1 &= \arg\min_{\theta_1} D\left(\theta_1\right) \\ \hat{\theta}_2 &= \arg\min_{\theta_2} D\left(\hat{\theta}_1, \theta_2\right) \\ &\vdots \\ \hat{\theta}_L &= \arg\min_{\theta_L} D\left(\hat{\theta}_1, \dots, \hat{\theta}_{L-1}, \theta_L\right). \end{aligned} \quad (20)$$

Since the distortion $D(\cdot)$ is minimized in each step, it follows that the distortion is a non-increasing function of the model order L. The efficient implementation of matching pursuit stems from each minimization being simpler than finding all parameters in $\{\theta_l\}_{l=1}^{L}$ simultaneously. The approach proposed here is to optimize the parameter set iteratively given an initial set of parameters $\{\theta_l^{(1)}\}_{l=1}^{L}$ as

$$\begin{aligned} \hat{\theta}_1^{(i+1)} &= \arg\min_{\theta_1} D\left(\theta_1, \hat{\theta}_2^{(i)}, \dots, \hat{\theta}_L^{(i)}\right) \\ \hat{\theta}_2^{(i+1)} &= \arg\min_{\theta_2} D\left(\hat{\theta}_1^{(i+1)}, \theta_2, \hat{\theta}_3^{(i)}, \dots, \hat{\theta}_L^{(i)}\right) \\ &\vdots \\ \hat{\theta}_L^{(i+1)} &= \arg\min_{\theta_L} D\left(\hat{\theta}_1^{(i+1)}, \dots, \hat{\theta}_{L-1}^{(i+1)}, \theta_L\right) \end{aligned}, \quad (21)$$

with i being the iteration index. Again, since the distortion is minimized in each step, it can be seen that the distortion is a non-increasing function. Similarly, it can easily be verified that the distortion is a non-increasing function across iterations i since

$$D\left(\hat{\theta}_1^{(i+1)}, \hat{\theta}_2^{(i)}, \dots, \hat{\theta}_L^{(i)}\right) \leq D\left(\hat{\theta}_1^{(i)}, \dots, \hat{\theta}_L^{(i)}\right). \quad (22)$$

We propose to build the model iteratively by increasing the model order as follows. Given the set of parameters $\{\hat{\theta}_l^{(1)}\}_{l=1}^{K-1}$ that were estimated for the $K-1$th order model, the model order is incremented by one and the parameters associated with the new sinusoid $\hat{\theta}_K$ are found as

$$\hat{\theta}_K = \arg\min_{\theta_k} D\left(\{\hat{\theta}_l\}_{l=1}^{K-1}, \theta_K\right). \quad (23)$$

Then the parameter set for the Kth order model $\{\hat{\theta}_l^{(1)}\}_{l=1}^{K}$ are optimized in a cyclic manner for all k and $i = 1, \dots, I$ as

$$\hat{\theta}_k^{(i+1)} = \arg\min_{\theta_k} D\left(\{\hat{\theta}_l^{(i+1)}\}_{l=1}^{k-1}, \theta_k, \{\hat{\theta}_l^{(i)}\}_{l=k+1}^{K}\right), \quad (24)$$

with the estimates being initialized as $\hat{\theta}_l^{(1)} = \hat{\theta}_l$, $\forall l$. The whole process is then repeated by incrementing the model order by one and finding the new parameters using (23) by using the parameters obtained in (24), i.e. by setting $\hat{\theta}_l = \hat{\theta}_l^{(I)}$, $\forall l$. The step in (23) can also be written in the form of (23) by setting the amplitude in $\{\hat{\theta}_l^{(1)}\}_{l=K+1}^{L}$ with $K \leq L$ to zero. We refer to the step in (23) as the augmentation step and the step in (24) as the optimization step. It is of course possible to skip the optimization step until the desired number of sinusoids have been reached. This corresponds to initializing the optimization step by estimates obtained using matching pursuit.

Next, we will re-write the optimization step in CMP in (24) into the notation of the framework presented in Section 2. First, we calculate the residual used for obtaining the parameters of the kth sinusoid in the ith iteration from the parameter set $\mathbf{\Theta} \setminus \theta_l$ as

$$\begin{aligned} r_k^{(i)}(n) = x(n) - \sum_{l=1}^{k-1} \hat{a}_l^{(i+1)} \exp\left(j\hat{\omega}_l^{(i+1)} n\right) \\ - \sum_{l=k+1}^{L} \hat{a}_l^{(i)} \exp\left(j\hat{\omega}_l^{(i)} n\right) \end{aligned} \tag{25}$$

and construct the corresponding residual vector $\mathbf{r}_k^{(i)}$. We can now write the frequency estimation criterion as the minimization of the perceptual distortion of this residual, i.e.,

$$\hat{\omega}_k^{(i+1)} = \arg\min_{\omega} \|\mathbf{H}(\mathbf{r}_k^{(i)} - \mathbf{z}a)\|_2^2 \tag{26}$$

$$= \arg\max_{\omega} \frac{\left|\left\langle \mathbf{Hz}, \mathbf{Hr}_k^{(i)} \right\rangle\right|^2}{\|\mathbf{Hz}\|_2^2}, \tag{27}$$

and the associated optimal complex amplitude estimate as

$$\hat{a}_k^{(i+1)} = \arg\min_{a} \|\mathbf{H}(\mathbf{r}_k^{(i)} - \mathbf{z}a)\|_2^2 \tag{28}$$

$$= \frac{\left\langle \mathbf{Hz}, \mathbf{Hr}_k^{(i)} \right\rangle}{\|\mathbf{Hz}\|_2^2}. \tag{29}$$

Since equations (27) and (29) have the same form as the PMP in (15), the CMP too can be implemented efficiently using two FFTs per iteration [11]. If very accurate estimates are desired, numerical nonlinear optimization can easily be applied to the maximization of (29) given the initial coarse estimates obtained using the FFT method. Since such an optimization still would be one-dimensional, it is computationally much simpler than the approach presented in [15] where all L sinusoids are optimized using Newton's method. Note that the augmentation step in (23) too can be described this framework by setting $a_l = 0$ for $l > k$ in the calculation of the residual in (25). The approximations used in deriving the WMP and the pre-filtering method also can be applied in this framework. However, since these methods do not minimize an explicit distortion measure, it is unclear how the convergence properties would be.

Regarding the number of iterations used in the optimization step, we here regard it as a parameter that is chosen by the user in accordance with theavailable resources. It is of course possible to use termination criteria based on the convergence of the distortion.

4. AN EXAMPLE

We will now illustrate the application of the proposed method for audio modeling using the perceptual distortion measure. We will here focus on the common case of a sinusoidal model using a 30 ms Hanning analysis-synthesis window and an FFT size of 4096. As has already been discussed, the CMP can be expected to outperform the PMP whenever there is significant interaction between the model components. This can be expected whenever we are dealing with complicated signals containing closely spaced sinusoids, modulated sinusoids or many sinusoids. Therefore, we will use a segment of the castanet signal from the EBU SQAM disc which is samples at 44.1 kHz. The signal is shown in Figure 1. Matching pursuit will tend to model this signal with many closely spaced sinusoids to capture the strong modulation. Figures 2 and 3 show the sinusoidal signal models that are obtained using the PMP and CMP, respectively. In both cases 100 sinusoids are used and 10 the CMP is set to perform 10 iterations in the optimization step. From the figure, it can be seen that the resulting signal models do not match the signal well. The distortion as a function of the number of sinusoids are plotted for the two methods, CMP (dashed) and PMP (solid), in Figure 4. It is important to note that the CMP both results in a higher rate of convergence for a low number of components and, it would seem, a lower saturation level for a high number of sinusoids. As can be seen from the figures, the CMP is better at modeling the complicated signal with the same number of sinusoids and it is also better for all possible numbers of sinusoids. It should be stressed that we here only optimize over a discrete set of frequency points, here 4096, i.e., we do not employ any kind of numerical optimization technique once the frequencies of sinusoids have been selected using the FFT-based implementation of the PMP and CMP. We remark that the interaction effects between components depends on the segment length N.

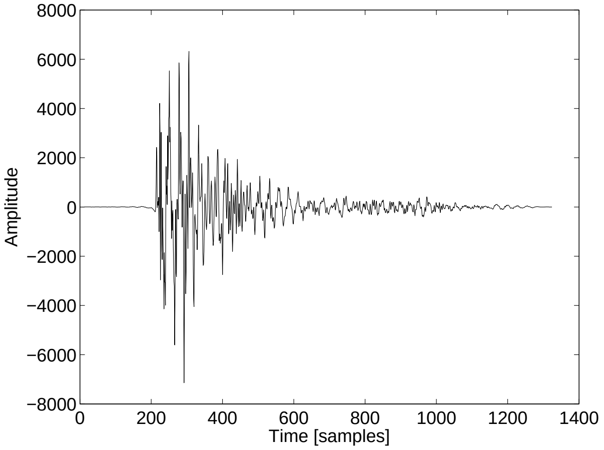

Fig. 1. Original signal, an excerpt of the castanet signal on the EBU SQAM disc.

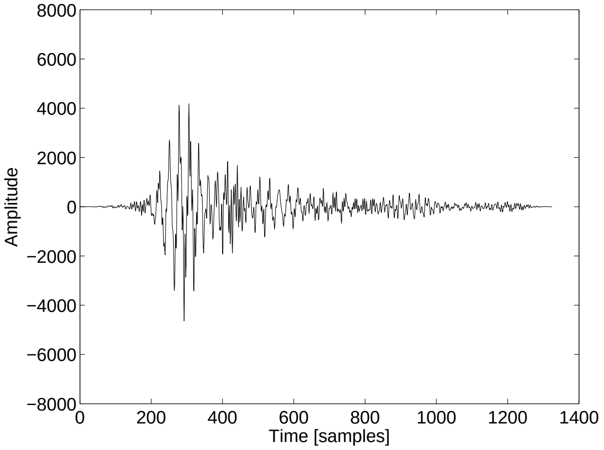

Fig. 2. Signal model obtained for 100 sinusoids using the PMP.

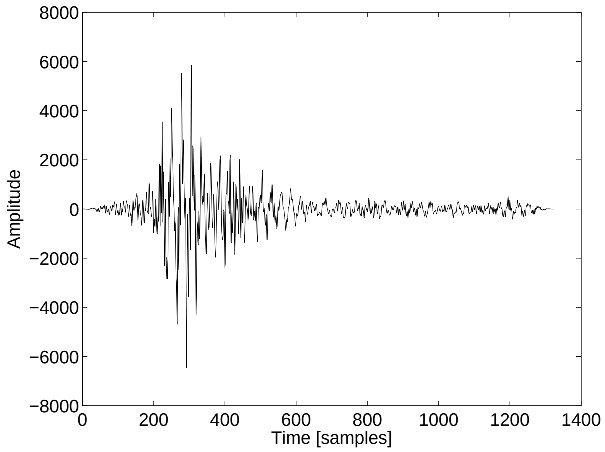

Fig. 3. Signal model obtained for 100 sinusoids using the proposed method, the CMP.

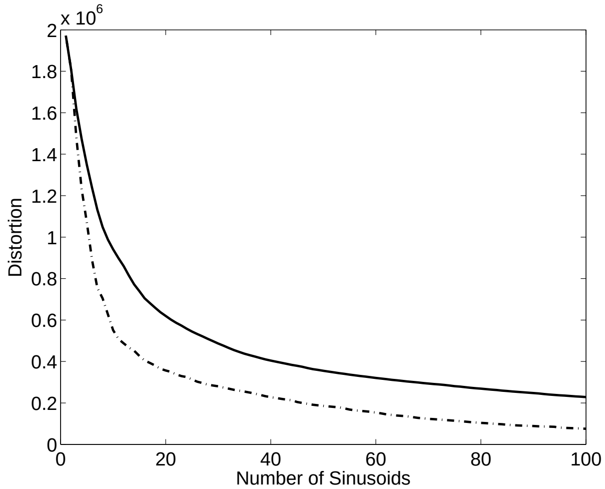

Fig. 4. Perceptual distortion as a function of the number of sinusoids for the two methods, CMP (dashed) and PMP (solid).

5. CONCLUSION

A new algorithm based on the principles of matching pursuit has been proposed. The new method, called the cyclic matching pursuit, is based on a iterative, cyclic minimization of a distortion measure where the parameters of each model component are refined in each iteration according to a distortion measure. The method is conceptually simple and easy to implement and does not require any multidimensional nonlinear optimization. The method has been derived in the context of sinusoidal audio modeling based on a perceptually relevant distortion measure, and an illustrative audio example showing its performance has been given. The example shows that the method is superior to the usual matching pursuit algorithms when modeling complicated signals or using many components. As a special case, the method reduces to the usual matching pursuit or the perceptual matching pursuit, depending on the distortion measure.

6. REFERENCES

[1] M. G. Christensen and S. H. Jensen, “On perceptual distortion minimization and nonlinear least-squares frequency estimation,” *IEEE Trans. on Audio, Speech and Language Processing*, vol. 14(1), pp. 99–109, Jan. 2006.

[2] S. van de Par, A. Kohlrausch, R. Heusdens, J. Jensen, and S. H. Jensen, “A perceptual model for sinusoidal audio coding based on spectral integration,” *EURASIP J. on Applied Signal Processing*, vol. 9, pp. 1292–1304, 2005.

[3] T. S. Verma and T. H. Y. Meng, “Sinusoidal modeling using frame-based perceptually weighted matching pursuits,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Mar. 1999, vol. 2, pp. 981–984.

[4] G. D. T. Schuller, B. Yu, D. Huang, and B. Edler, “Perceptual audio coding using adaptive pre- and post-filters and lossless compression,” in *IEEE Trans. Speech and Audio Processing*, Sept. 2002, vol. 10(6), pp. 379–390.

[5] R. Heusdens, R. Vafin, and W. B. Kleijn, “Sinusoidal modeling using psychoacoustic-adaptive matching pursuits,” *IEEE Signal Processing Lett.*, vol. 9(8), pp. 262–265, Aug. 2002.

[6] S. Mallat and Z. Zhang, “Matching pursuit with time-frequency dictionaries,” *IEEE Trans. Signal Processing*, vol. 41(12), pp. 3397–3415, Dec. 1993.

[7] J. Adler, B. Rao, and K. Kreutz-Delgado, “Comparison of basis selection methods,” in *Rec. Asilomar Conf. Signals, Systems, and Computers*, Nov. 1996, vol. 1, pp. 252–257.

[8] Y. Pati, R. Rezaiifar, and P. Krishnaprasad, “Orthogonal matching pursuit: Recursive function approximation with application to wavelet decomposition,” in *Rec. Asilomar Conf. Signals, Systems, and Computers*, Nov. 1993, vol. 1, pp. 40–44.

[9] H. Feichtinger, A. Turk, and T. Strohmer, “Hierarchical parallel matching pursuit,” *Proc. SPIE: Image Reconstruction and Restoration*, vol. 2302, pp. 222–232, July 1994.

[10] M. M. Goodwin, *Adaptive Signal Models: Theory, Algorithms, and Audio Applications*, Ph.D. thesis, University of California, Berkeley, 1997.

[11] R. Heusdens and S. van de Par, “Rate-distortion optimal sinusoidal modeling of audio using psychoacoustical matching pursuits,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, May 2002, pp. 1809–1812.

[12] J. Li and P. Stoica, “Efficient mixed-spectrum estimation with application to target feature extraction,” *IEEE Trans. Signal Processing*, vol. 44(2), pp. 281–295, Feb. 1996.

[13] M. Feder and E. Weinstein, “Parameter estimation of superimposed signals using the EM algorithm,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36(4), pp. 477–489, Apr. 1988.

[14] P. Stoica and Y. Selen, “Cyclic minimizers, majorization techniques, and the expectation-maximization algorithm: a refresher,” *IEEE SP Mag.*, vol. 21(1), pp. 112–114, 2004.

[15] D. Kloosterman, R. Heusdens, and J. Jensen, “Estimation of sinusoidal model parameters using newton optimization and a perceptual distortion measure,” in *Proc. IEEE Benelux Signal Processing Symposium*, 2004, pp. 199–202.