TRUE DISCRETE CEPSTRUM: AN ACCURATE AND SMOOTH SPECTRAL ENVELOPE ESTIMATION FOR MUSIC PROCESSING

Rémi Mignot and Vesa Välimäki

Aalto University,
Department of Signal Processing and Acoustics
Otakaari 5A, 02150 Espoo, Finland

ABSTRACT

In the tradition of the spectral envelope estimation of periodic sounds, we propose a new accurate method, called *True Discrete Cepstrum*. Solving a constrained optimization problem, it provides a smooth envelope which fits exactly the given peak values. Moreover, based on the auditory masking, we propose a release of the constraint which improves the smoothness, without perceptual change. Contrarily to some other methods, the parametrization of this method is easy, and it gives a complete control of the maximal deviation of the spectral envelope from the peak values. The benefit of this new method is illustrated and an evaluation procedure validates it with a comparison with some other methods.

Index Terms— Acoustic signal processing, Spectral envelope estimation, Music sound analysis-synthesis

1. INTRODUCTION

For the treatment of the sine part of musical sounds, a popular method consists of the factorization of a given spectrum into the product of a source, discrete and spectrally flat, and a continuous spectral envelope, cf. e.g. [1]. With this *Source-Filter* principle the estimation of the spectral envelope is a crucial task because it is generally considered as one of the determining factors of the timbre of sounds.

Using the autocorrelation sequence of the original signal, the *Linear Prediction Coding* is a well-known method. Unfortunately, this method is highly biased with discrete spectra, because the estimated spectral envelope is attracted by the valleys between the peaks, cf. e.g. [2].

In [3], the *Discrete All-Pole* method (DAP) is proposed. Starting from the peak values, frequency positions and magnitudes, the Itakura-Saito distance is iteratively minimized. The obtained spectral envelope is then given by an autoregressive model which roughly passes through the peaks. Although this method is well adapted for speech, the all-pole modeling is not suitable for musical tones in a general case. See Fig. 1 where the order 14 is chosen according to Fig. 5.

This work is funded by the Marie Curie Action project ESUS 299781.

The *True Envelope* of [4, 5] is another approach which consists of an iterative cepstral liftering starting from the DFT spectrum of the original sound. Nevertheless, knowing the peak values, the computation of the input spectrum is a sensitive task because the valleys between the peaks make the convergence slower. An example is given in Fig. 1, where the optimal order 21 is chosen. The obtained spectral envelope is accurate, but we do not control the peak fitting with precision, especially in the case of inharmonic sounds, and some undesirable artifacts can appear as shown at high frequencies.

In [6], Galas and Rodet proposed the *Discrete Cepstrum* which consists of a least mean square approximation to identify the spectral envelope using the peak values and a cepstral model. Nevertheless, as illustrated in Fig. 2, with low orders the estimation is inaccurate, and with higher orders the obtained spectral envelope passes close to the spectral peaks, but some conditioning problems appear as shown at high frequencies. To solve this problem, in [7] Cappé et al. regularized the criterion using a penalty functional which aims to improve the smoothness of the spectral envelope. Even if the smoothness is significantly improved, as shown in Fig. 2, the approximation is less accurate at the spectral peaks, and there is a difficult choice of the regularization factor.

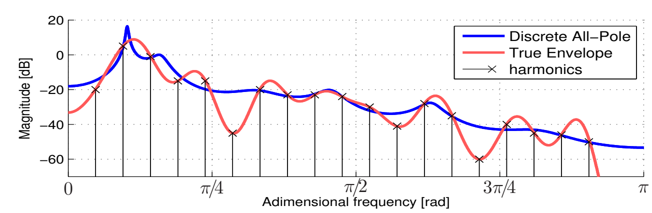


Fig. 1. Illustration of the DAP method and the True Envelope.

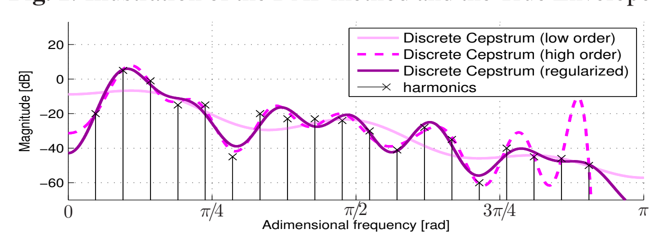


Fig. 2. Illustration of the Discrete Cepstrum.

In the present work, inspired by the regularized Discrete Cepstrum, we propose a new and simple method which provides a smooth cepstral model passing exactly through the peak values and which has no sensitive parameter. Moreover, even if the smoothness is naturally better, we can release the equality constraint by controlling the peak deviation. We introduce here a new idea to define an inequality constraint which takes into account the auditory perception.

This article is organized as follows. In Sec. 2, we present some details of the Discrete Cepstrum method [6], and of its regularization. Then, we present the proposed method in Sec. 3. In Sec. 4, an example of interpolation is given and an evaluation procedure is presented. Finally, Sec. 5 concludes this paper and opens some perspectives.

2. STANDARD DISCRETE CEPSTRUM

2.1. Cepstral model and approximation

The desired spectral envelope $|H(\omega)|$ is modeled by its real cepstral model which follows, cf. e.g. [8]:

$$\log(|H(\omega)|) = c_0 + 2\sum_{p=1}^{P-1} c_p \cos(p\omega), \tag{1}$$

where the c_p's are the cepstral coefficients, ω is the adimensional radial frequency in [rad], $\omega = 2\pi f/F_s$ with F_s the sampling rate in [Hz], and P is the order of the cepstrum. Given the M peaks of $X(\omega)$, with frequencies ω_m and magnitudes $a_m = |X(\omega_m)|$, the spectral envelope is known at $\omega = \omega_m$. Then the cepstral representation of $|H(\omega)|$ is obtained by minimizing the criterion

$$\rho = \sum_{m=1}^{M} \Big(\log(|H(\omega_m)|) - \log(a_m)\Big)^2. \tag{2}$$

With $P < M$, more equations than variables, the Discrete Cepstrum method of Galas and Rodet [6] gives the optimal cepstral coefficients by the Least Mean Square solution:

$$C = (\Phi^T\Phi)^{-1}\Phi^T A,$$

where C is the column vector of the cepstral coefficients, $C_p = c_{p-1}$, $\forall p \in [1, P]$, Φ is the $(M \times P)$ matrix of the cepstral model of (1), $\Phi_{m,p} = w_p \cos((p-1)\omega_m)$ with $w_1 = 1$ and $w_p = 2$ for $p > 1$, and A is the vector of the logarithm of the measured values of the spectral envelope, $A_m = \log(a_m)$.

As previously illustrated in Fig. 2, for low orders the estimation is inaccurate, and although it is better for higher orders, undesirable artifacts appear because of the poor conditioning of $\Phi^T\Phi$. Note that for Fig. 2, $M = 19$ harmonics, $P = 6$ for the low order, and $P = 17$ for the high order.

To improve the conditioning of the problem, a simple solution is the regularization of the matrix $\Phi^T\Phi$, adding arbitrary and small values on the diagonal for example. In the next section, an interesting regularization is defined to guarantee the smoothness of the estimated spectral envelope.

2.2. Regularization

In [7], Cappé et al. define the new cost function $\rho_r = \rho + \lambda\mathcal{P}$ where $\mathcal{P}$ is a penalty function measuring the smoothness of the spectral envelope and λ a small regularization factor. A convenient definition of $\mathcal{P}$ is the squared sum of the k-th derivative of $\log|H(\omega)|$, cf. e.g. [9]:

$$\mathcal{P} \triangleq \frac{1}{2\pi}\int_0^{\pi} \left(\frac{d^k}{d\omega^k}\log|H(\omega)|\right)^2 d\omega. \tag{3}$$

Smaller is $\mathcal{P}$ and smoother is $\log(|H(\omega)|)$, then its minimization optimizes the smoothness of the desired spectral envelope. Using the cepstral representation of $|H(\omega)|$, cf. (1), it is easy to prove that $\mathcal{P} = C^T\Delta C$ with $\Delta = \mathrm{diag}([0, 1^{2k}, 2^{2k}, \ldots (P-1)^{2k}])$. Now, the minimization of ρ_r is obtained by

$$C = (\Phi^T\Phi + \lambda\Delta)^{-1}\Phi^T A.$$

As illustrated in Fig. 2, this regularization significantly improves the regularity of the estimated spectral envelope. Nevertheless we can notice that the input peaks are not well fitted anymore; it is not surprising because of the regularization. Moreover, the choice of the regularization factor is a difficult choice. For this estimation, $M = 19$, $P = 17$ and $\lambda = 10^{-4}$. This value of λ has been manually chosen in order to have an obviously good compromise between smoothness and peak fitting; note that this choice is difficult and not automatic.

3. ACCURATE CEPSTRAL MODELING

3.1. True Discrete Cepstrum

As noted previously, when the cepstral order gets closer to the number of peaks, the Discrete Cepstrum provides an accurate fitting of the peaks. In the critical case $P = M$, if Φ is not singular, we get $C = \Phi^{-1}A$, which provides an exact solution of $A = \Phi C$. Naturally in this case, the observed border effects of conditioning are worse than previously.

Using a cepstral order $P > M$, we get an underdetermined problem which has an infinite number of exact solutions if $\mathrm{rank}(\Phi) = M$. The key idea of this work is to choose the solution that maximizes the smoothness of $H(\omega)$ using (3). Consequently, the proposed *True Discrete Cepstrum* method consists of solving the constrained optimization problem given by

$$\min_{C\in\mathbb{R}^P} \left(C^T\Delta C\right) \quad \text{subject to} \quad \Phi C = A. \tag{4}$$

In other words, if $\mathrm{rank}(\Phi) = M$, the solutions of $A = \Phi C$ live in a subspace with dimension $P - M$ which is the number of degrees of freedom. In Sec. 3.2, using a change of variable, we minimize $\mathcal{P} = C^T\Delta C$ in the subspace of the solutions. First the equality constraint guarantees that the chosen solution perfectly fits the peaks, second the minimization guarantees the smoothness of the estimated spectral envelope.

3.2. Solving method

The optimization problem of (4) is a constrained linear least square problem and has been treated in mathematics, cf. e.g. [10]. We here summarize the solving method.

The solving takes benefit of the QR factorization, cf. e.g. [11], which is commonly implemented in most of scientific computation softwares. With $P > M$, if $\mathrm{rank}(\Phi) = M$ which is validated at least for harmonic signals, the matrix Φ is written as follows

$$\Phi^T = QR = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R_1 \\ \mathbf{0} \end{bmatrix} = Q_1 R_1, \tag{5}$$

with R_1 a $(M \times M)$ upper triangular matrix, $\mathbf{0}$ the $((P-M) \times M)$ null matrix, Q a $(P \times P)$ orthogonal matrix decomposed into $Q = [Q_1,\ Q_2]$ where Q_1 and Q_2 have respectively M and $P - M$ columns. Writing the vector C of the cepstral coefficients into the basis of matrix Q, we get the following change of variables:

$$C = Qx = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} \beta \\ \mu \end{bmatrix} = Q_1\beta + Q_2\mu. \tag{6}$$

Since the columns of Q_2 form a basis of the kernel of Φ, $\ker(\Phi) = \{x \in \mathbb{R}^P \ / \ \Phi x = 0\}$, the $((P-M) \times 1)$ vector μ is the vector of the degrees of freedom. On the contrary, the $(M \times 1)$ vector β is constrained by:

$$\begin{aligned} \Phi C &= R_1^T Q_1^T (Q_1\beta + Q_2\mu) = R_1^T \beta = A \\ &\Rightarrow \beta = (R_1^T)^{-1} A, \end{aligned} \tag{7}$$

with $Q_1^T Q_1 = I_M$, identity matrix of size M, and $Q_1^T Q_2 = \mathbf{0}$ since Q is orthogonal. Now, with the change of variables of (6), we express $\mathcal{P} = C^T \Delta C$ with β and μ. The vector β is fixed by (7) and the vector μ that minimizes $\mathcal{P}$ is given by

$$\frac{d\mathcal{P}}{d\mu} = 0 \ \Rightarrow \ \mu = -(Q_2^T \Delta Q_2)^{-1} Q_2^T \Delta Q_1 \beta. \tag{8}$$

Finally we get the optimal solution of (4) by $C = Q_1\beta + Q_2\mu$.

3.3. Release of the constraint

In some cases, the exact fitting of the peaks is unnecessary, and it is interesting to be able to release the equality constraint in order to improve the smoothness, reduction of $\mathcal{P}$. For example, we can control the maximal deviation between the peaks and the estimated spectral envelope by solving the following optimization problem

$$\min_{C \in \mathbb{R}^P} \left(C^T \Delta C \right) \ \text{ subject to } \ \|\Phi C - A\|_\infty \leq \varepsilon \tag{9}$$

with $\|v\|_\infty = \max_n \{|v_n|\}$ the infinite norm, and ε such that $\varepsilon_{dB} = 20\varepsilon / \log(10)$ is the maximal deviation in [dB].

In this paper, we propose also another interesting idea for the choice of the inequality constraint. It yields a significant improvement of the smoothness without perceptible changes. First we can note that the problems of the undesirable artifacts generally come from some peaks which are lower than their neighboring peaks. Nevertheless, considering the simultaneous auditory masking, cf. e.g. [12], we can expect that these peaks are masked by their neighboring peaks. Then, from the perceptual point of view, it is useless to fit these peaks because they are inaudible. The only important point is that the value of the spectral envelope at these peak frequencies does not exceed the masking threshold. To illustrate this principle, we define the following inequality constraints:

$$A_m - \epsilon_m^- \leq \phi_m C \leq A_m + \epsilon_m^+, \quad \forall m \in [1, M], \tag{10}$$

where ϕ_m is the m-th row of Φ. For the audible peaks, above the masking threshold $f(\omega)$: $\epsilon_m^- = \epsilon_m^+ = \varepsilon$, and for the masked peaks: $\epsilon_m^- = \infty$ and $\epsilon_m^+ = f(\omega_m)$. Here we use the simple model of [13] for the auditory masking threshold.

4. EXAMPLES AND EVALUATION

4.1. Examples

Figure 3 illustrates the True Discrete Cepstrum (TDC) and compares it with the True Envelope of Fig. 1. The order of the True Envelope is its optimal order which is $P = 21$ here, cf. [14]. For the TDC, with $M = 19$ harmonics, we use here the order $P = 25$ and $k = 2$, second derivative cf. (3). We can see that the TDC provides a spectral envelope which exactly fits the peaks and which is more regular than the True Envelope, with an order slightly higher.

In Fig. 4, we illustrate the perceptual constraint release which uses the auditory masking, cf. Sec. 3.3. From the harmonics, the masking threshold is computed, and the values of $\epsilon_m^\pm$ are set according to the "audibility" of the peaks. To prevent some problems at the threshold, it is preferable to use the computed masking threshold with a negative safety margin. First, we can see that the spectral envelope does not fit the inaudible harmonics, but passes above them and below the mask because of the safety margin. Second, as expected the general form of the spectral envelope is slightly more regular.

4.2. Evaluation

To evaluate the performance of the True Discrete Cepstrum with a comparison to the other methods (DAP, Discrete Cepstrum and True Envelope), we realize the evaluation procedure used in [14, 15]. In this section, we only summarize the procedure, more details are given in [14].

The test signals are periodic and have a synthetic spectral envelope computed by an ARMA(4,4) model. The angles of the zeros and the poles in the complex Z-plane are fixed, and the corresponding radii change to provide several different spectral envelopes. The input peak values of the methods DAP, Discrete Cepstrum and TDC are given by the sampling

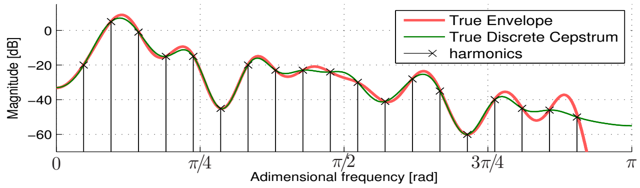


Fig. 3. Illustration of the True Discrete Cepstrum.

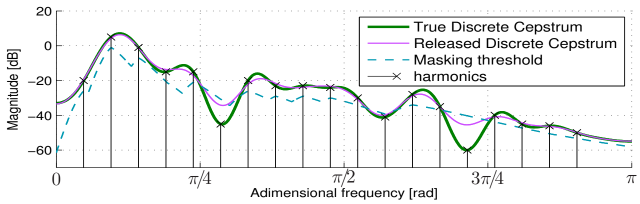


Fig. 4. Illustration of the TDC with perceptual release.

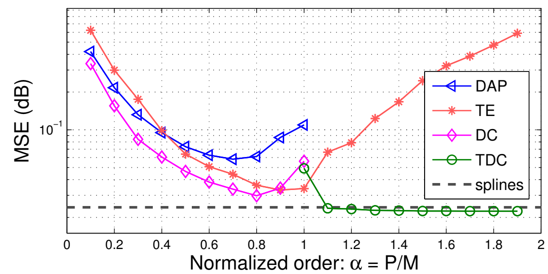
of the synthetic spectral envelope: $|H(\omega_m)|$. For the True Envelope, a sine synthesis gives the time signal and the Discrete Fourier Transform provides the input spectrum of the method, cf. see [16]. Note that the weighting window length and the DFT size are specifically chosen in order to obtain a good frequency resolution and a main lobe width equal to $\omega_0 = 2\pi F_0/F_s$, the fundamental frequency in [rad]. The TDC uses the second derivative for $\mathcal{P}$ ($k = 2$, cf. (3)).

In Fig. 5, the root Mean Square Error (MSE) of the logarithmic amplitude, cf. [14], is presented as a function of the normalized order $\alpha = P/M$. Remind that P is the model order and M the number of peaks. The DAP method and the Discrete Cepstrum are given for $P \leq M$, the TDC is naturally given for $P \geq M$ and the True Envelope is given for all cases. Note that in the case of periodic signals the optimal order of the True Envelope is $P = 0.5F_s/F_0$, which approximately corresponds to $\alpha = 1$. In Fig. 5a, the error is averaged over the test signals with periods $T_0 \in [400, 500]$ in samples, low frequencies. In Fig. 5b, the used periods are $T_0 \in [50, 150]$, higher frequencies. Additionally, the evaluation of the standard spline interpolation is done, cf. e.g. [17]. Note that this modeling is not suitable for the spectral envelope modeling.

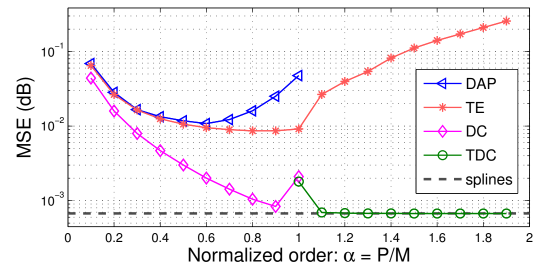
As a general trend, we observe in Fig. 5 that the TDC outperforms all methods when $P > M$, and it is equivalent to the spline interpolation. Moreover we observe a stable behavior with all orders $P > M$. Finally, we notice a great benefit of the TDC for high frequencies, Fig. 5b.

5. CONCLUSION

In this paper we present a new and simple method for the spectral envelope estimation using an accurate and smooth spectral interpolation of the peaks. Contrarily to the regularization of [7], this method has no sensitive parameters and its results are stable for any order $P > M$. Not only the True Discrete Cepstrum exactly fits the peaks, not only it out-



(a) $P_0 \in [400, 500]$, low fundamental frequencies



(b) $P_0 \in [50, 150]$, high fundamental frequencies

Fig. 5. Results of the evaluation procedure.

performs the other methods, cf. Fig. 5, but also it is naturally smoother without conditioning problems, cf. Figs. 1-3. Moreover, when the exact fitting is not necessary, it is possible to release the constraint in order to improve the smoothness, by controlling with precision the maximal deviation. Additionally, we proposed an interesting idea to define an inequality constraint which respects the auditory masking. In this case, the fitting error is not perceptible.

Remark that with the conditioning problem, a small change of the original spectrum can involve a major change in the estimated spectral envelope, then it can give some time discontinuities in the case of a frame-by-frame analysis. Consequently, the smoothness of the estimated spectral envelope may improve the timbre stability for synthesis. Moreover, the smoothness may also improve some sound modifications such as pitch shifting and vibrato, and it may facilitate a low-order ARMA approximation [18] for a low-cost synthesis.

Nevertheless, in the special case of real-time processing, the True Envelope of [5] remains the most suitable method. First, while the TDC needs the peak analysis, cf. e.g. [19], the True Envelope only needs an FFT computation for the input spectrum. Second, whereas the computation time of the TDC depends on the matrix sizes, number of peaks M and order P, the computation time of the True Envelope weakly depends on the fundamental frequency.

As a consequence, the proposed TDC method provides an interesting alternative method which has great benefits for off-line analysis with a full control of the precision.

6. REFERENCES

[1] U. Zölzer, *DAFX: Digital Audio Effects*, Wiley, 2nd edition, 2011, 624 pages.

[2] J. Makhoul, "Linear prediction: A tutorial review," *Proceedings of the IEEE*, vol. 63, no. 4, pp. 561–580, 1975.

[3] A. El-Jaroudi and J. Makhoul, "Discrete all-pole modeling," *IEEE Trans. Signal Process.*, vol. 39, no. 2, pp. 411–423, 1991.

[4] S. Imai and Y. Abe, "Spectral envelope extraction by improved cepstral method," *Electronics and Communication*, vol. 62-A, no. 4, pp. 10–17, 1979, in Japanese.

[5] A. Röbel and X. Rodet, "Efficient spectral envelope estimation and its application to pitch shifting and envelope preservation," in *Proc. Int. Conf. on Digital Audio Effects (DAFx'05)*, 2005, pp. 30–35.

[6] T. Galas and X. Rodet, "An improved cepstral method for deconvolution of source-filter systems with discrete spectra: Application to musical sounds," in *Proc. of the Int. Computer Music Conf. (ICMC'90)*, 1990, pp. 82–84.

[7] O. Cappé, J. Laroche, and E. Moulines, "Regularized estimation of cepstrum envelope from discrete frequency points," in *Proc. IEEE Workshop on Applicat. of Signal Process. Audio Acoust. (WASPAA'95)*, 1995, pp. 213–216.

[8] A. Oppenheim, D. Johnson, and K. Steiglitz, "Computation of spectra with unequal resolution using the fast Fourier transform," *Proceedings of the IEEE*, vol. 59, no. 2, pp. 299–301, 1971.

[9] M. Campedel-Oudot, O. Cappé, and E. Moulines, "Estimation of the spectral envelope of voiced sounds using a penalized likelihood approach," *IEEE Trans. Speech and Audio Proc.*, vol. 9, no. 5, pp. 469–481, 2001.

[10] Å. Björck, *Numerical Methods for Least Squares Problems*, SIAM, 1996, 408 pages.

[11] G.H. Golub and C.F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, 3rd edition, 1996, 728 pages.

[12] S.A. Gelfand, *Hearing: An Introduction to Psychological and Physiological Acoustics*, Taylor & Francis, 4th edition, 2004, 512 pages.

[13] G. Garcia and J. Pampin, "Data compression of sinusoidal modeling parameters based on psychoacoustic masking," in *Proc. of the Int. Computer Music Conf. (ICMC'99)*, 1999, pp. 40–43.

[14] A. Röbel, F. Villavicencio, and X. Rodet, "On cepstral and all-pole based spectral envelope modeling with unknown model order," *Pattern Recognition Letters*, vol. 28, no. 11, 2007.

[15] F. Villavicencio, A. Röbel, and X. Rodet, "All-pole spectral envelope modelling with order selection for harmonic signals," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP'07)*, 2007, pp. I49–I52.

[16] X. Rodet and P. Depalle, "Spectral envelopes and inverse FFT synthesis," *93nd AES Convention*, 1992.

[17] C. De Boor, *A Practical Guide to Splines*, vol. 27, Springer-Verlag New York, 1978.

[18] M.H. Hayes, *Statistical Digital Signal Processing and Modeling*, Wiley, 1999, 624 pages.

[19] R. McAulay and T. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Trans. Acoustics, Speech and Signal Process.*, vol. 34, no. 4, pp. 744–754, 1986.