ALL-POLE SPECTRAL ENVELOPE MODELLING WITH ORDER SELECTION FOR HARMONIC SIGNALS

Fernando Villavicencio, Axel Röbel and Xavier Rodet

IRCAM-CNRS-STMS
Analysis-Synthesis team
Place Igor-Stravinsky 75004 Paris France
{villavicencio,roebel,rodet}@ircam.fr

ABSTRACT

We present a study into all-pole spectral envelope estimation for the case of harmonic signals. We address the problem of the selection of the model order and propose to make use of the fact that the spectral envelope is sampled by means of the harmonic structure to derive a reasonable choice for an appropriate model order. The experimental investigation uses synthetic ARMA featured signals with varying fundamental frequency and differing model structure to evaluate the performance of the selected all-pole models. The experimental results confirm the relation between optimal model order and the fundamental frequency.

***Index Terms*—** Envelope detection, Feature extraction, Cepstral analysis, Speech analysis, Speech synthesis,

1. INTRODUCTION

The estimation of the *spectral envelope*, which is a smooth function passing through the prominent peaks of the spectrum, is a very important task in signal processing applications. It is useful for signals that are generated according to the source-filter model, which means that a white excitation signal passes through a resonator filter. The spectral envelope is actually the transfer function of the resonator filter and the task consists of the estimation of this resonator filter from the signal. Spectral envelope estimation can be applied to signal characterization, classification, and modification. While signal characterization and classification applications generally do not require a very precise calculation of the spectral envelope, the quality of voice for timbre conversion systems depends on the quality of the envelope estimate.

In the case of white noise excitation signals, there are reliable and straightforward estimation techniques [1]. For periodic excitation signals, however, as well as for pitched instruments or voiced speech, the estimation is difficult due to the fact that the distinction between the spectral envelope and the excitation signal is ambiguous. In these cases the peaks defining the spectral envelope are the harmonics of the fundamental frequency. Therefore, the spectral envelope should be a transfer function that, if inverted, renders the sequence of spectral peaks of the residual signal as flat as possible, without, extracting the harmonic structure of the excitation signal.

Some problems that hinder the estimation are the proper selection of the filter model (AR, MA, or ARMA) and the proper selection of the model order. The estimation of AR or all-pole models by means of linear prediction (LPC), that has been introduced in [2], is still a technique that is used quite often for the estimation and parametric representation of the spectral envelope of speech signals. While LPC modeling can still be considered to be state of the art if the excitation signal is white noise, for high pitched harmonic excitation signals the LPC technique is known to be biased. For these excitation signals the discrete all-pole (DAP) technique that has been presented in [3] can be used to considerably reduce the bias, however, for the price of increased computational costs and algorithmic complexity. Also, the True-Envelope LPC (TELPC), presented in [4] performs improved all-pole envelope modeling for voiced speech. This technique relies on the computation of an AR model from the spectral envelope estimations provided by the cepstrum based *True Envelope* estimator [5]. Note, that for the order selection problem there is only a physically motivated reasoning [6]. The fact that the filter is observed after having been sampled by the harmonic structure has not yet been taken into account.

In the following article we present an experimental and comparative study of the all-pole based envelope estimation techniques mentioned above. The goal is to derive a simple and effective strategy that allows to select an appropriate model order, and to investigate the limits of the different models with respect to filter properties. With respect to the order selection problem, we will argue that the optimal model order is related directly to the fundamental frequency of the excitation signal.

The article is presented as follows. The selected all-pole models are presented in section two. Cepstrum based order selection is introduced in section three. In the section four we describe the experimental framework. The results are presentend in section five. The works ends in section six with the conclusions.

2. ALL-POLE MODELLING

2.1. Linear Prediction (LPC)

It is a well known fact that LPC can be used to estimate the spectral envelope for white noise excitation signals as long as the order of the model is sufficiently large. For harmonic excitation signals, the selection of the LPC model order is more critical because with increasing order the LPC model will not fit the envelope but the complete spectrum including the harmonic structure. The usual approach to specify the appropriate model order is based solely on the physical properties of the resonator filter [6]. Even though we know that with an increasing pitch the LPC model will degrade, no attempt has been made to connect the model order to the fundamental frequency.

2.2. Discrete all-pole (DAP)

For harmonic signals, it was shown in [3] that a systematic error performed by the MSE procedure of LPC appears in the fitted envelope as a bias of the spectral peaks towards the pitch harmonics. This is due to the matching of the filter acf with an aliased version of the acf of the signal. It is simple to see that the aliasing effect will increase with increasing fundamental frequency of the excitation signal as well as with decreasing smoothness of the spectral envelope.

The aim of the discrete all-pole model (DAP) [3] is to solve the aliasing problem just described. The basic idea exploited in DAP is to fit the all-pole model using only the finite set of spectral locations that are related to the harmonic positions of the fundamental frequency.

2.3. True-Envelope based LPC (TELPC)

Aside from the aliasing problem metioned above, LPC does not represent the desired spectral information passing through the prominent peaks of the spectrum that we denote as a spectral envelope. Usually, the resulting LPC spectra will overestimate the predominant maximas of the spectrum. In order to reduce this model mistmach, TELPC uses the spectral envelope estimations obtained from the *True-Envelope* estimator as a target spectrum for the autocorrelation matching criteria. This proposal follows the idea introduced in [10] to use interpolated spectrum information for all-pole modeling.

3. ORDER SELECTION FOR CEPSTRUM BASED MODELS

3.1. Order selection

ARMA envelope models are easily obtained through cepstrum based techniques. The cepstrum is a DFT representation of the log amplitude spectrum and it can be shown that ARMA transfer functions can be represented by means of the cepstrum [7]. Also, a major advantage of the cepstral envelope estimation techniques is that a reasonable estimate of the optimal cepstral order can be provided. If the observed signal has fundamental frequency F_0, the harmonic excitation spectrum samples the resonator filter with a samplerate given by F_0. Therefore, one may deduce that the information of the original filter that exceeds the related Nyquist bandlimit in the cepstral domain is lost. Assuming a samplerate of F_s, the related Nyquist quefrency bin number in the discrete cepstrum is $Fs/(2F_0)$. This claim allows us to provide a simple way of selecting a nearly optimal cepstral order given only, that the maximum frequency difference between two spectral peaks that carry envelope information is known. If the difference between those peaks is Δ_F then the cepstral order should be

$$\hat{O} = \frac{F_s}{2\Delta_F} = \alpha\frac{F_s}{\Delta_F}, \quad \alpha_c = 0.5 \qquad (1)$$

While the optimal order, that is the order that provides an envelope estimate with minimum error, depends on the specific properties of the envelope spectrum, the order selection according to (1) is reasonable for a wide range of situations and the resulting error, as we will show in the next section, is generally rather close to the one obtained with the optimal order.

3.2. Pre-smoothed True-Envelope estimation

There are various techniques for cepstrum based envelope estimation. In [8] an attractive cepstrum based spectral envelope estimator named *True-Envelope* (TE) has been presented. This iterative technique allows efficient estimation of the spectral envelope [5] without

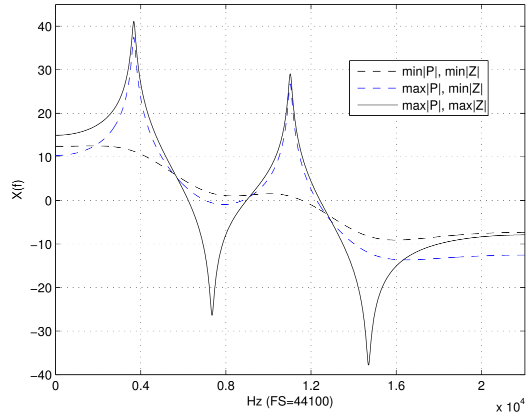

Fig. 1. Evaluation transfer function examples.

the shortcomings of the discrete cepstrum [9]. The resulting estimation can be interpreted as a band limited interpolation of the major spectral peaks.

While very high performance estimation can be achieved with the TE method, the necessary sub-sampling in the spectral domain will result in a suboptimal performance with respect to envelope estimation errors. Therefore, in the following experiments we do not use the frequency domain sub-sampling described in [5]. Initial experiments with the TE estimator revealed a problem with the described optimal order. This problem is due to the fact that in real world signals the spectral envelope is not sampled regularly. The main problem here is the fact that the spectral peak at 0Hz is generally missing such that for harmonic excitation with fundamental frequency F_0 the maximal frequency difference between the supporting peaks will be $\Delta_F = 2F_0$. To be able to increase the model order we propose a two step estimation. First a TE model with order $O = F_s/(4F_0)$ is estimated. From this estimate we derive an estimate of the envelope at position 0Hz and create an artificial DC spectral peak that has the proper amplitude. In the second step, according to (1), we select $\hat{O}_{TE} = \alpha_c(Fs/F_0)$, so that all available details in the envelope spectrum can be resolved. The two step estimation procedure is what will be denoted as *True-Envelope* method below.

4. EXPERIMENTAL WORK

4.1. Synthetic ARMA featured signals

For the experimental evaluation of the algorithms that have been briefly presented above, synthetic signals with a small number of poles and zeros will be used. The transfer functions consist of 2 pairs of poles and 2 pairs of complex zeros. Because the smoothness of the spectral envelope is mostly related to the pole and zero radius we select fixed angular locations of poles and zeros. The first pole pair has angle $\pm\pi/6$ and the second $\pm\pi/2$. For the zeros the angles $\pm 2\pi/6$ and $\pm 4\pi/6$ are used. The set of radii used to form the pole r_p and zero r_z locations are given by

$$r = \log(2.013 + k0.04) \qquad \text{with} \quad k = 0, 1, \ldots, 17. \qquad (2)$$

The radii that are used sample the interval $r = [0.7, 0.99]$. Using a samplerate of 44100kHz they represent 3dB-bandwidth in

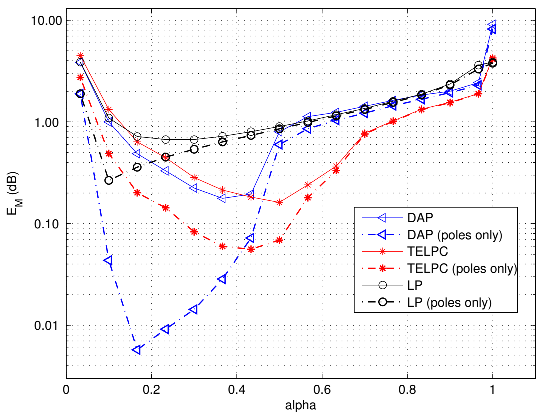

Fig. 2. Average modelling error ($P_0 = 150$).

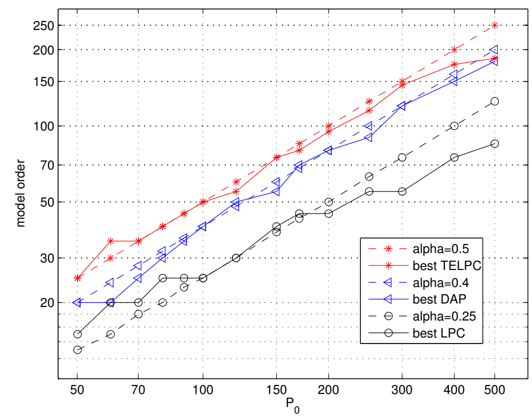

Fig. 3. Experimental and proposed order comparison.

the range 132Hz-5014Hz. These are values that cover the range of formant bandwidths that are common for the spectral envelopes of speech and musical instruments. 14 values for P_O were selected in the interval [50, 500] ($F_0 = [88.2Hz, 882Hz]$). For all experiments we use a Hanning window that covers exactly 4 periods of the fundamental frequency of the excitation signal. To prevent systematic errors that may arise due to the fact that the spectral bins will not sample the harmonic peaks exactly at the local maximum we use a DFT size that is a power of 2 at least 8 times longer than the analysis window. To restrict the dimensionality of the problem only a 2-dimensional grid of radii is considered. The first dimension controls the radius of the 4 poles while the second dimension controls the radius of the zeros. The complete 2-dimensional grid allows us to study transfer functions that are dominated by AR or MA filter characteristics, as well as an important number of ARMA filters. In Fig. 1 we show 3 cases of special interest: ARMA (radii close to 1), AR dominatad and ARMA(lowers radii).

4.2. Evaluation procedure

LPC, DAP and TELPC algorithms are used to obtain estimates of the spectral envelope of the synthesized signals using a grid of orders covering the range from $p = [5, F_S/F_0]$ and using an order increment of 5. To evaluate the estimation error we use the root mean square error of the log amplitude

$$E_M = \sqrt{\frac{1}{K}\sum_{k=0}^{K-1}(\log|S(k)| - \log|\hat{S}(k)|)^2}, \qquad (3)$$

where $S(k)$ and $\hat{S}(k)$ are the K-point DFT of the filter transfer function and the estimated envelope respectively. Note that the use of the magnitude spectrum does not exclude the phases because for minimum phase filters the phases are determined by the log amplitude spectrum such that the phase spectra do not add any information. The error measure (3) appears especially useful for algorithms that try to achieve timbre modification by means of deconvolution of the spectral envelope, because for these applications any error, whether it is underestimation or overestimation of the envelope is equally important. For applications that try to achieve formant location, another error measure would be preferred.

5. RESULTS

In Fig. 2 is presented an example of the average error described by eq. (3) displayed as a function of the parameter α, which represents the model order relative to the number of samples in the fundamental period $P_0 = 1 = F_0$. These shapes were commonly found in the selected P_0 values. Two sets of transfer functions are used to calculate the averages of the estimation error (3). The first set contains all transfer functions described above, and the second set fixes the zero radius to the lowest value ($r_z = 0.7$), such that only the subset of transfer functions that is closest to an AR model is taken into account.

Looking at the position of the resulting best order provides us with interesting observations. As expected, the TELPC method obtains optimal performance for the order described by (1) that represents the Nyquist quefrency limit. A similar trend can be observed for the DAP method. Here, however, the optimal position is slightly below the theoretical limit at about $\alpha_d = 0.4$. It could be due to the fact that the optimization criterion used in DAP differs from (3) combined with the lack of the uniqueness property in the convergence procedure. On the other hand, we can't state a theoretical interpretion for the value found for LPC ($\alpha_l = 0.25$).

We observed that for envelopes that are dominated by AR characteristics the resulting best order of the DAP method is much lower as well as the resulting average error. This can be related to the fact that for smaller order the errors in the spectral samples that are due to the sampling of the envelope according to the peak positions can be suppressed if a strong structure is imposed by the model. In the real world we usually do not make definite assumptions about the number of poles, and therefore it would be advantageous to select the DAP model order according to (1) and α_d.

In Fig 3, we compare the resulting best order for each model with the orders selected according to (1) and the values mentioned above. Clearly, the experimental best orders for TELPC and DAP follows closely the proposed order selection. Fig. 4, reveals that the error (averaged over the whole pole-zero grid) decreases with increasing P_0. This is related to the fact that with a larger period more harmonics are used to sample the envelope which is an advantage for the estimation. Note that for all models, the performance provided by the order selection follows closely the experimental optimal results. In Fig. 5 we show the same comparison limited to the second

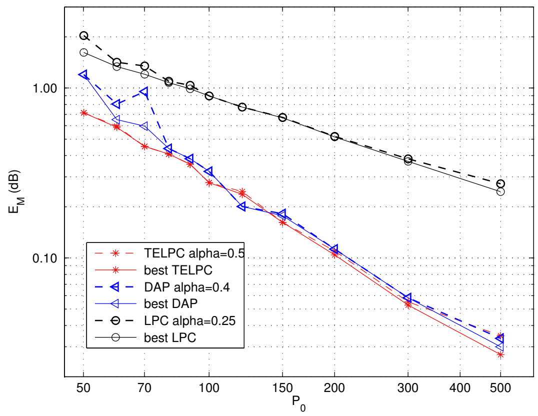

Fig. 4. Average modelling error using experimental and proposed order

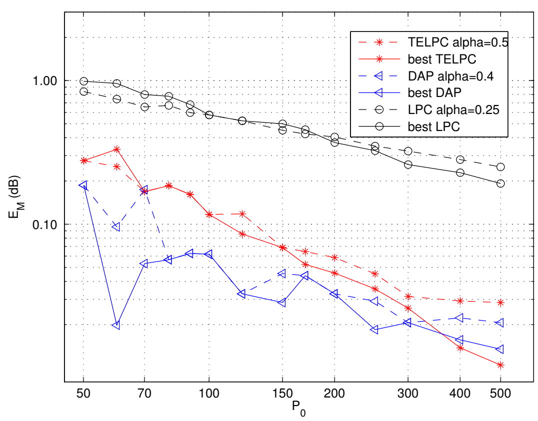

Fig. 5. Average modelling error (AR dominated transfer functions)

set of transfer functions (AR dominated). For both experiments, LPC shows clearly the worst modelling performance.

As a result of this first experimental section we conclude that the model order for TELPC and DAP estimation should be selected as a function of the fundamental frequency using α_c and α_d in eq. (1). For the LPC model no conclusive result can be obtained besides the advantages experimentally found using α_l. After having obtained a simple method of selecting proper orders for DAP and TELPC we investigate the relation between the envelope characteristics and the best model to use. We computed, for each fundamental period, the percentage of cases on the pole-zero grid for which each model achieved the smallest estimation error. Firstly, the experimental best orders were used. Then, the proposed order selection was performed. In the first case, TELPC performed better than DAP in almost all the P_0 cases. Using the order selection, we observed that for a medium period the TELPC method is the best, while for a large period (low F_0) DAP is clearly advantageous. For a small period, the best model depends on the system type. The LPC model could never beat the other methods in both experiments.

6. CONCLUSIONS

The article presented an experimental comparison of envelope estimation techniques for pitched excitation signals.

The main goal of the investigation was to establish experimental evidence for a simple scheme that derives a proper model order from the fundamental frequency of the observed signal. A slight modification of the true envelope estimator was required in order to be able to achieve optimal performance with the suggested order selection. The experiments indicate that for the modified TELPC estimator the Nyquist quefrency is a proper indicator for model order selection. For the DAP estimator the model order should be limited to 0.4 times the number of samples per period of the fundamental frequency. The order selection method that has been proposed has the fundamental advantage of being able to relate the model order to the information that will be present in the observed spectrum. Therefore, we suppose that the derived limits are not only valid for the experimental setup, but they can be used for real world applications as well.

The direct comparison of the estimators demonstrate that the LPC model is clearly the worse estimator. The comparison of DAP and TELPC estimators did reveal that the choice of the optimal estimator does depend on the fundamental frequency and the envelope charac- teristics. The TELPC envelope estimator has the advantage to allow an efficient real time implementation.

7. REFERENCES

[1] S. Kay, *Modern Spectral Estimation: Theory and Application*, Prentice-Hall Signal Processing Series, USA, 1988.

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

[3] A. El-Jaroudi and J. Makhoul, "Discrete All-Pole Modeling," *IEEE Transactions on Signal Processing*, vol. 39, no. 2, pp. 411–423, 1991.

[4] A. Roebel F. Villavicencio and X. Rodet, "Improving LPC Spectral Envelope Extraction of Voiced Speech by True-Envelope Estimation," in *Proc. of the ICASSP'06*, France, 2006.

[5] A. Roebel and X. Rodet, "Efficiente Spectral Envelope Estimation and its Application to Pitch Shifting and Envelope Preservation," in *Proc. of the DAF'x'05*, Spain, 2005.

[6] Douglas O'Shaughnessy, *Speech Communication*, Addison-Wesley Series in Electrical Engineering: Digital Signal Processing. Addison-Wesley Publishing Company, 1987.

[7] Julius O. Smith, *Introduction to Digital Filters, September 2005 Draft*, Prentice-Hall, USA, 2005.

[8] S. Imai and Y. Abe, "Spectral Envelope Extraction by Improved Cepstral Method," *Electron. and Comm. (in Japan)*, vol. 62, no. 4, pp. 10–17, 1979.

[9] T. Galas and X. Rodet, "An improved cepstral method for deconvolution of source filter systems with discrete spectra: Application to musical sound signals," in *Proceedings of the International Computer Music Conference (ICMC)*, 1990, pp. 82–84.

[10] H. Fujisaki H. Hermansky and Y. Sato, "Spectral Envelope Sampling and Interpolation in Linear Predictive Analysis of Speech," in *Proc. of the ICASSP'84*, 1984, pp. 2.2.1–2.2.4.