Available online at www.sciencedirect.com

ScienceDirect

Pattern Recognition Letters 28 (2007) 1343–1350

Pattern Recognition Letters

www.elsevier.com/locate/patrec

On cepstral and all-pole based spectral envelope modeling with unknown model order

Axel Röbel, Fernando Villavicencio *, Xavier Rodet

IRCAM – CNRS – UMR STMS, Analysis-Synthesis Team, 1, Place Igor-Stravinsky, 75004 Paris, France

Available online 27 March 2007

Abstract

In this work, we investigate spectral envelope estimation for harmonic signals. We address the issue of model order selection and propose to make use of the fact that the spectral envelope is sampled by means of the harmonic structure of the signal in order to derive upper bounds for the estimator order. An experimental study is performed using synthetic test signals with various fundamental frequencies and different model structures to evaluate the performance of the envelope models. Experimental results confirm the relation between optimal model order and fundamental frequency.

Keywords: Spectral envelope modeling; All-pole modeling; Cepstral analysis; Source-filter deconvolution; Speech analysis; Speaker characterization; Speech synthesis

1. Introduction

Estimation of the *spectral envelope*, which is a smooth function passing through the prominent peaks of the spectrum, is an important task in signal processing applications. The spectral envelope is generally considered as one of the determining factors for the timbre of a sound. In terms of the well known source-filter model, which models sound creation by means of a white excitation signal passing through a filter, the spectral envelope is the transfer function of the filter. Accordingly, the task consists in estimating the resonator filter from the signal. Spectral envelope estimation methods can be used for applications as signal characterization, classification and modification. While signal characterization and classification applications generally do not require a very precise estimation of the spectral envelope, the quality of voice or timbre conversion systems depends on the quality of the envelope estimate.

In the case of white noise excitation signals there are various straightforward estimation techniques (Kay, 1988). If, however, the excitation signal is periodic (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 cases like these 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 including 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 (LP), that was described in (Makhoul, 1975), is a technique that is still used quite often for the estimation and parametric representation of the spectral envelope of speech signals. LP modeling can be considered a state of the art procedure if the excitation signal is white noise. For harmonic

* Corresponding author. Present address: Analysis and Synthesis Team, Institute for Music-Acoustic Research and Coordination, 1, Place Igor Stravinsky, 75004, Paris, France. Tel.: +33 14478 4862; fax: +33 14478 1540.

E-mail addresses: roebel@ircam.fr (A. Röbel), villavicencio@ircam.fr (F. Villavicencio).

URLs: http://www.ircam.fr\anasyin\roebel (A. Röbel), http://www.ircam.fr\anasyin\villavicencio (F. Villavicencio), http://www.ircam.fr\anasyin\rodet (X. Rodet).

0167-8655/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2006.11.021

excitation signals, however, the LP technique is known to be biased. For these excitation signals the discrete all-pole (DAP) technique that was presented in (El-Jaroudi and Makhoul, 1991) can be used to considerably reduce the bias. Note that compared to the LP method the computational costs and the algorithmic complexity of the DAP algorithm are significantly increased. For the order selection problem there exists only a physically motivated reasoning (O'Shaughnessy, 1987). The fact that the filter is observed after having been sampled by the harmonic structure has not yet been taken into account.

ARMA envelope models are most 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 (Smith, 2005). There are different techniques for cepstrum based envelope estimation. In (Imai and Abe, 1979) an attractive cepstrum-based spectral envelope estimator, named *true-envelope* (TE), is presented. This iterative technique allows efficient estimation of the spectral envelope (Roebel and Rodet, 2005) without the shortcomings of the discrete cepstrum (Cappé and Moulines, 1996; Galas and Rodet, 1990). The resulting estimation can be interpreted as a band limited interpolation of the major spectral peaks.

In the following article an experimental comparative study of envelope estimation techniques is presented. The goal of this investigation is to derive a simple and effective strategy allowing us to select an appropriate model order, and to investigate the performance of different models with respect to the filter properties. For experimental investigation the LP, DAP and TE techniques will be used. The experimental setup is especially relevant for tasks that require the estimation of the residual or excitation signal of pitched signals, such as voice morphing or timbre modification. For these tasks, in contrast to formant detection, a uniform approximation of the envelope is generally advantageous because an error in the excitation signal, whether due to a formant or an anti-formant, may become perceptually important once the envelope has been modified. With respect to the order selection problem we will demonstrate that for the DAP and TE estimators, a reasonable model order can generally be derived from the fundamental frequency of the excitation signal.

The article is organized as follows. The cepstrum based *True-Envelope* algorithm is introduced in Section 2. LP and DAP all-pole based models are described in Section 3. In Section 4, we present the experimental framework and we describe the results in Section 5. Section 6 summarizes the article.

2. Efficient cepstrum-based spectral envelope estimation

2.1. The True-Envelope estimator

There are a number of approaches for estimating the spectral envelope by means of cepstral smoothing. The discrete cepstrum is the most well know, but, is rather demanding computationally. It requires a pre-selection of the spectral peaks. The *True-Envelope* (TE) estimator was originally proposed in (Imai and Abe, 1979). Recently, a procedure has been proposed that allows significant reduction of computational costs to a level comparable with the Levinson recursion such that real time processing can be achieved (Roebel and Rodet, 2005). Note however, that reduction of the computational cost comes with slightly reduced precision. Therefore, we will not use the real time version of the TE estimator for the following experiments. The true-envelope estimator will be used as representative of the cepstrum based spectral estimators.

TE estimation is based on cepstral smoothing of the amplitude spectrum. Let $X(\omega_k)$ be the K-point DFT of the signal frame $x(n)$ and $C_i(\omega_k)$ the cepstrally smoothed spectrum at iteration i. The algorithm then iteratively updates the smoothed input spectrum $A_i(\omega_k)$ with the maximum of the original spectrum and the current cepstral representation

$$A_i(\omega_k) = \max(\log(|X(\omega_k)|), C_{i-1}(\omega_k)) \quad (1)$$

and applies the cepstral smoothing to $A_i(\omega_k)$ to obtain $C_i(\omega_k)$. The procedure is initialized setting $A_0(\omega_k) = \log(|X(\omega_k)|)$ and starting the cepstral smoothing to obtain $C_0(\omega_k)$. As depicted in Fig. 1 the estimated envelope grows steadily. The algorithm stops if for all ω_k the relation $A_i(\omega_k) < C_i(\omega_k) + \theta$ is true with θ being a user supplied threshold. For the current experiments $\theta = 0.01$ dB was used. Given the fact that the cepstral order is limited the TE estimator creates a band limited function that passes through the prominent spectral peaks. The peaks that are considered prominent are automatically selected according to the cepstral order. The explicit peak selection that is necessary for the DAP estimator as well as for the discrete cepstrum, is not required.

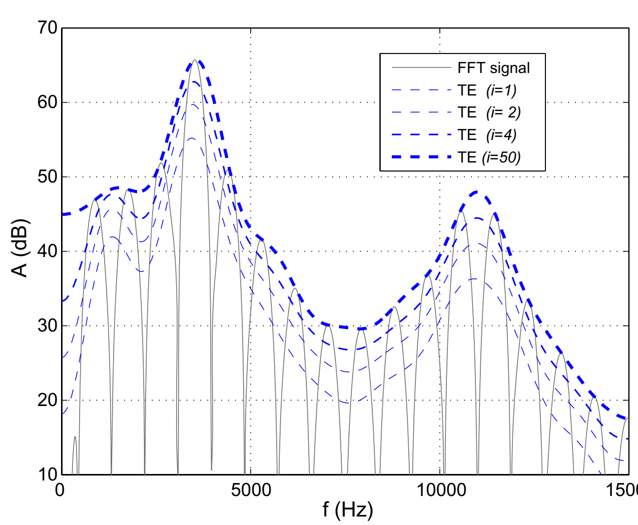

Fig. 1. True-Envelope estimator iteratively approaching the ARMA test spectrum (model order $O = 25$, sample rate $F_s = 44{,}100$ Hz, fundamental period $P_0 = F_s/F_0 = 50$).