A STOCHASTIC SINUSOIDAL MODEL WITH APPLICATION TO SPEECH AND EEG-SLEEP SPINDLE SIGNALS

David Labarre(1), Eric Grivel(1), Yannick Berthoumieu(2) and Mohamed Najim(1)

(1) Equipe Signal & Image, UMR LAP 5131, ENSEIRB BP99
(2) Equipe SACT, Laboratoire IXL UMR 5818,
Talence 33402, France
Tel : +33 5 56 84 61 85 ; Fax : +33 5 56 84 84 06
E-mail : (labarre ; eric.grivel ; najim)@tsi.u-bordeaux.fr ; berthoumieu@ixl.u-bordeaux.fr

ASBTRACT

In this paper, we propose to investigate stochastic sinusoidal models in order to characterise quasi-periodic signals. Indeed, in comparison to the broadly used autoregressive (AR) models, a sinusoidal approach seems to be more efficient to capture quasi-periodic feature. Using AR process as a model for the sine wave magnitudes makes it possible to track the frequential non-stationarity of the signal.

The scheme we propose operates as follows: once the frequency components of the signal are obtained, estimating the magnitudes of each sine component of the model is performed by means of an Expectation-Maximisation (EM) algorithm based on Kalman smoothing.

Results are provided on sleep spindle and speech.

1 INTRODUCTION

Various signals contain quasi-periodic segments. In this paper we will focus our attention on biomedical signal analysis and speech modelling.

In biomedical signal processing, the quasi-periodic feature is exhibited in sleep electro-encephalogram (EEG), where one can detect and extract quasi-periodic patterns called sleep spindles (Cf. figure 1). Once these spindles are detected in the EEG recording by using methods proposed in [1] or [7], the modelling with a sinusoidal approach provides the temporal evolution of their various modes and makes spindle classification possible.

On the other hand, in the framework of speech processing, the AR model has been broadly exploited especially for LPC based synthesis, CELP coding, Kalman filter enhancement [2], [3], [5], etc. However, the AR model does not capture the quasi-periodic nature of voiced speech frames (Cf. figure 2). For this reason, considering an all-pole model in which the excitation may be either noise-like for unvoiced frames, or pulse-like for voiced frames can be considered but requires robust Voiced/Unvoiced decision.

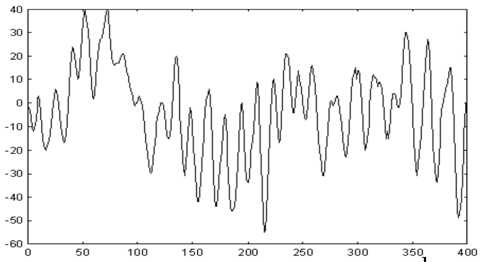

Figure 1: example of a sleep spindle (f_s=100 Hz)

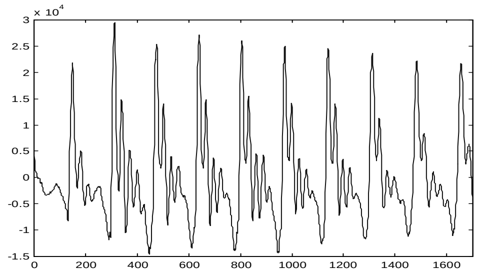

Figure 2: quasi periodic nature of a voiced frame (vowel /A/ with sampling frequency f_s=16KHz)

In this paper, our purpose is to derive a stochastic sinusoidal model defined by a sum of sinusoids, whose magnitudes are stochastic processes. Once the frequency components of the signal are estimated, one can provide the phase of the magnitude of each sine component.

This paper is organised as follows. In part 2, we present the stochastic sinusoidal model. In part 3, we propose an approach to estimate the model parameters. In the last part, the model is tested with synthesized signal, sleep spindles and speech signal.

2 THE SINUSOIDAL MODEL

The sinusoidal model we propose is defined as follows:

$$y(k) = \sum_{i=1}^{L} \left[a_i(k) \cos(2k\pi \frac{f_i}{f_s}) + b_i(k) \sin(2k\pi \frac{f_i}{f_s}) \right] + v(k) \quad (1)$$

where f_s is the sampling frequency, $v(k)$ is a zero-mean additive Gaussian noise with a variance σ_n^2, L is the number of spectral components.

However, when completing the Wigner-Ville transform of sleep spindles, the frequential non-stationnarity can be exhibited (Cf. figure 3). Such a remark is also valid when we observe voiced speech signal and its pitch evolution (Cf. figure 4).

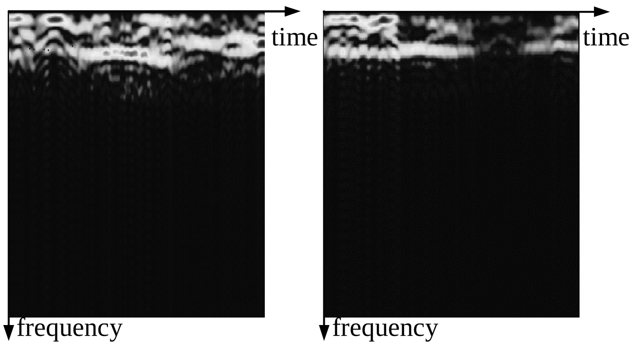

Figure 3: time-frequency representation of two sleep spindles[1]

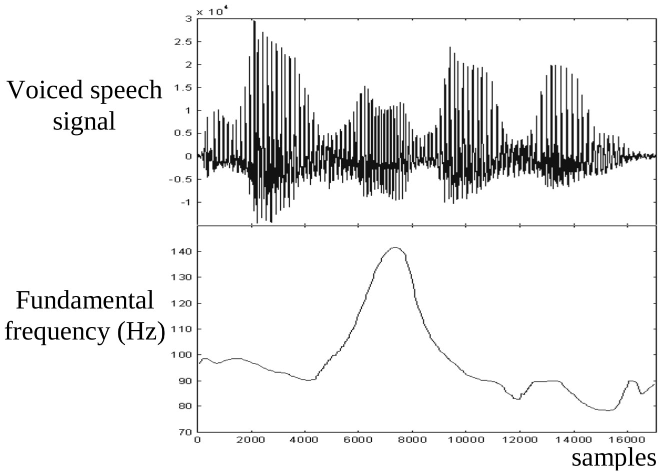

Figure 4: evolution of the pitch of a voiced speech signal

For this reason, we propose to model the magnitudes $a_i(k)$ and $b_i(k)$ of the spectral components by respectively n_a^{th} and n_b^{th} order AR processes.

$$\begin{cases} a_i(k) = \sum_{l=1}^{n_a} \gamma_i^l a_i(k-l) + \varepsilon_{a_i}(k) \\ b_i(k) = \sum_{l=1}^{n_b} \gamma_i^l b_i(k-l) + \varepsilon_{b_i}(k) \end{cases} \quad (2)$$

where $0 \le \gamma_i^l < 1$ will ensure the model stability. $\varepsilon_{b_i}(k)$ and $\varepsilon_{a_i}(k)$ are zero-mean Gaussian sequences, with variance σ_i^2.

[1] The colour (here in grey scale) provides the power of the frequency components (white for high energy ; black for low energy)

In this paper, we consider first order AR models, but the approach can be derived with higher order AR processes.

3 ESTIMATION OF THE MODEL PARAMETERS

We assume that the estimation of the frequencies f_i can be easily completed. Nevertheless, our method depends on the context: for speech signals, as a harmonic model can be considered, robust pitch tracking is completed. When dealing with EEG, one can use High Resolution (HR) spectral analysis, such as Music, Esprit, etc. [8].

Let N observations $y(k)$ be available. Our purpose is to:

- obtain the Maximum Likelihood of the estimates of the following model parameters

$$\theta = \left[\sigma_1^2, \sigma_2^2, ..., \sigma_l^2, \gamma_1, \gamma_2, ..., \gamma_l\right], \quad (3)$$

- estimate the magnitudes $a_i(k)$ and $b_i(k)$.

For such a purpose, we propose to develop an Expectation-Maximisation (EM) algorithm [8] based on Kalman smoothing [4]. It is an iterative algorithm. Each iteration operates in two steps. The M-step consists in providing an updated estimation of γ_i and σ_i^2, which requires conditional expectation quantities concerning $a_i(k)$ and $b_i(k)$. These quantities can be obtained during the E-step.

3 . 1 M-step

At the $r+1^{th}$ iteration, one obtains:

$$\begin{cases} \gamma_i^{r+1} = \dfrac{\sum_{k=2}^{N} E\left\{a_i(k)a_i(k-1) + b_i(k)b_i(k-1)/Y\right\}}{\sum_{k=2}^{N} E\left\{a_i(k-1)^2 + b_i(k-1)^2/Y\right\}} \\ \left(\sigma_i^2\right)^{r+1} = \dfrac{1}{2(N-1)} \sum_{k=2}^{N} E\left\{ \begin{array}{l} \left[a_i(k) - \gamma_i^{r+1} a_i(k-1)\right]^2 \\ + \left[b_i(k) - \gamma_i^{r+1} b_i(k-1)\right]^2 / Y \end{array} \right\} \end{cases} \quad (4)$$

where $E\{./Y\}$ represents the conditional expectation based on the N samples of data. Such quantities can be obtained by mean of Kalman smoothing, which corresponds to the E-step.

3 . 2 E-step

In order to carry out a Kalman smoothing, let us consider the following state-space representation:

$$\begin{cases} \underline{x}(k+1) = \Phi \underline{x}(k) + \underline{u}(k+1) \\ y(k) = \underline{H}(k)\underline{x}(k) + v(k) \end{cases}, \quad (5)$$

where $\underline{x}(k) = [a_1(k),..., a_L(k), b_1(k),..., b_L(k)]^T$ is the state vector, Φ the diagonal transition matrix (filled with the parameters γ_i), $\underline{u}(k)$ the vector of the driving processes and $\underline{H}(k)$ the transfer matrix.

Completing Kalman smoothing requires the estimation of both γ_i and σ_i^2 (obtained during the M-step), and σ_n^2 the variance of the noise $v(k)$ (obtained, either during the HR spectral analysis or during silent period of speech).

4 APPLICATION TO SPINDLES CHARACTERISATION

4 . 1 High resolution spectral analysis for the frequency components estimation

The ESPRIT TLS [8] method is used to estimate the frequency components of the spindle. First, the autocorrelation matrix of the observed signal is estimated as follows:

$$\hat{R}_{yy} = \frac{1}{N-M+1} YY^t \quad (6)$$

where N is the length of the observed signal, M the size of the autocorrelation matrix, and Y the observation Hankel matrix, which both rows and columns are sequences of M successively observed samples.

The second step consists in smoothing the diagonals of $\hat{R}_{yy}$, i.e. assigning the average of the terms of a diagonal to all the coefficients of this diagonal. Thus, the resulting $\hat{R}_{yy}^s$ matrix has a Toeplitz structure.

Then, we complete an eigenvalue decomposition of $\hat{R}_{yy}^s$ which leads to the estimation of the frequency components of the spindle and the variance σ_n^2 of the additive Gaussian noise $v(k)$.

4 . 2 Application to synthetic spindles

We first exploit the approach we have derived with synthetic data. In the previous step, we estimate the frequency components. Here, we focus our attention on the magnitude estimation.

Tests have been completed on various synthetic spindles. We present here some examples. The algorithm proves to be efficient to follow the synthetic magnitudes (Cf. figure 5), even when a spectral component disappears for a while (Cf. figure 6).

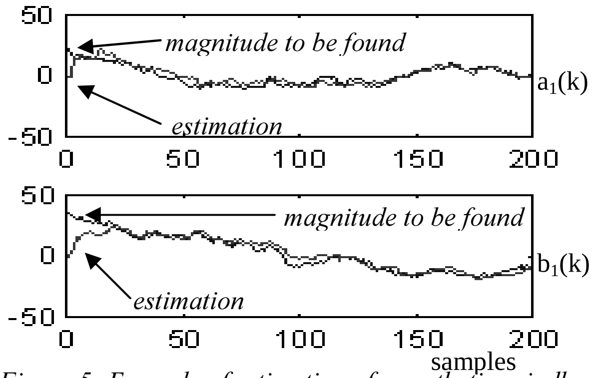

Figure 5: Example of estimation of a synthetic spindle component

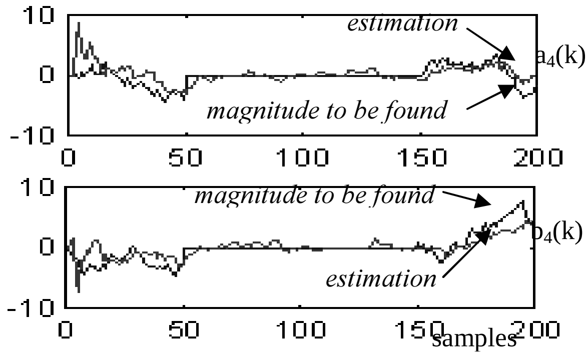

Figure 6: Estimation of the magnitudes of a component which disappears for a while

When overestimating the model order, i.e. when taking into account too-a-higher number of frequency components, the temporal evolution of the spurious estimated component is close to zero. Cf. figure 7: a 5th order model is chosen whereas the synthetic spindle has only four sine components.

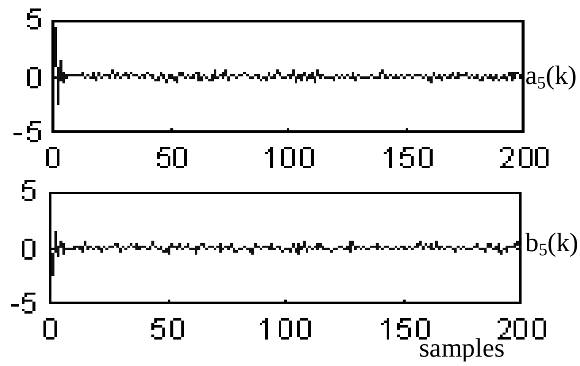

Figure 7: Estimation of a superfluous component

4 . 3 Application to real spindles recording

In this section, results based on a real spindle (Cf. figure 2) are provided.

frequency	f_1	f_2	f_3	f_4
Hz	0.5426	2.3538	4.2355	6.2066
frequency	f_5	f_6	f_7	f_8
Hz	7.2466	8.8793	11.0779	13.0858

Table 1: estimation of the frequency components of a real spindle

Table 1 provides the estimation of the frequency components obtained by HR spectral analysis.

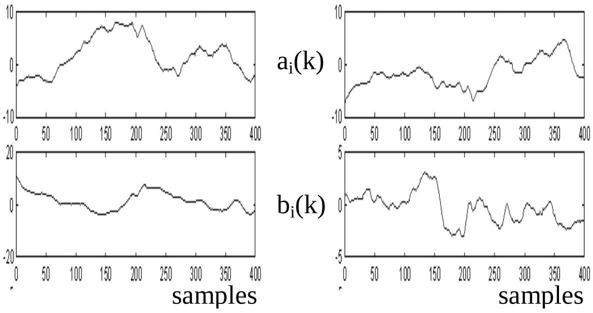

Figure 8: estimation of two sine components of a real spindle

For each spindle, the algorithm converges and estimates a model which permits to reconstruct the signal.

5 APPLICATION TO SPEECH PROCESSING: PRELIMINARY RESULTS

We have computed the presented approach in the context of voiced speech enhancement. The quality of the reconstructed voiced speech signal depends on both the number of harmonics L and the initial values of γ_i and σ_i^2.

Here, L is equal to 8. The initial values for σ_i^2 and γ_i are respectively 100000 and 0.98.

We provide results for an input Signal to Noise Ratio (SNR) equal to 15 dB. The SNR improvement is equal to 6 dB. The temporal evolutions of the original signal, the noisy signal and the enhanced signal are given in figure 9.

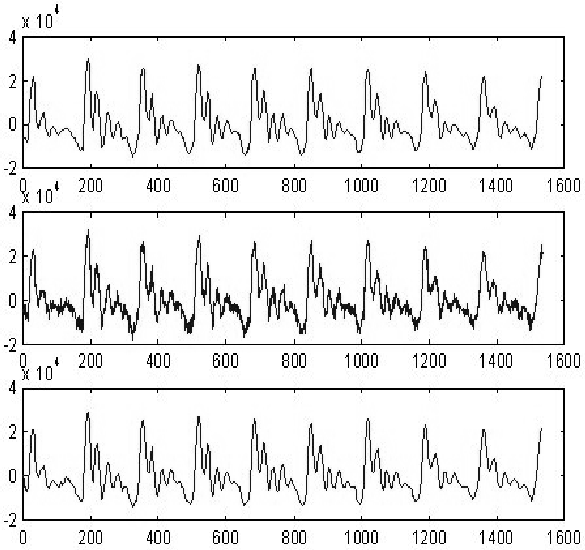

Figure 9: original signal, noisy signal and reconstructed signal

6 CONCLUSION

In this paper, we have proposed a stochastic sinusoidal model, where the sinusoidal magnitudes are modelled by AR processes. An EM-based algorithm is then used to estimate the model parameters and makes it possible to obtain the magnitudes using Kalman smoothing. In comparison to commonly derived AR models, the proposed approach provides promising results to capture the quasi-periodic feature of signals.

The algorithm has been exercised first on sleep spindle and provides a time/frequency analysis of the signal.

In the field of speech enhancement, a harmonic model is considered. The next step consists in defining and deriving an algorithm able to find for each frame both the adapted number of sine components and the initial values of the variance of the driving processes. Those preliminary results prefigure a global EM-based algorithm with a voiced/unvoiced decision. When the unvoiced decision is made, a stochastic AR model [3] is considered whereas harmonic model is used for voiced decision.

REFERENCES

[1] T. Akgül, M. Sun, R. J. Sclabassi and A. Enis Cetin, "*Characterization of Sleep Spindles Using Higher Order Statistics and Spectra*", IEEE Trans.on Biomedical Engineering, Vol. 47, n°8, Août 2000.

[2] M. Gabrea, E. Grivel and M. Najim, "*A Single Microphone Kalman Filter-Based Noise Canceller*", IEEE Signal Processing Letters, march 1999.

[3] S. Gannot, D. Burchtein and E. Weinstein, "*Iterative and Sequential Kalman Filter-Based Speech Enhancement Algorithms*", IEEE Trans. On Speech and Audio Processing, pp. 373-385, July 1998.

[4] J. M. Mendel, "Lessons in Estimation Theory for Signal Processing, Communication and Control", Prentice Hall 1995.

[5] K. K. Paliwal and A. Basu, "*A Speech Enhancement Method Based on Kalman Filtering*", ICASSP 87.

[6] J. Silva, "*Electrocardiographic Signal Analysis Using Genetic Algorithms"*, Computers in Cardiology, 1996, IEEE Press, pp. 445-448.

[7] K. Suarez, J. Silva and M. Najim, "*A Genetic Algorithm Approach For Pattern Recognition In Biomedical Signals"*, Gretsi 2001.

[8] C. W. Therrien, *"Discrete Random Signals and Statistical Signal Processing"*, Prentice Hall, 1992.

AKNOWLEDGEMENT

We would like to acknowledge O. Cappe (ENST/CNRS) and Marc Tomczak, (Université Henri Poincare de Nancy) who respectively provide us voiced speech signals and real sleep spindles base. This work has been partially supported by the CNRS and ISIS Research Network (GdR-PRC ISIS: Information, Signal, Images and ViSion).