MAGNITUDE SPECTRAL ESTIMATION VIA POISSON MOMENTS WITH APPLICATION TO SPEECH RECOGNITION

Samel Çelebi and Jose C. Principe

Computational Neuroengineering Lab. CSE 447, Univ. of Florida, Gainesville FL32611

Email: celebi@synapse.ee.ufl.edu, principe@synapse.ee.ufl.edu

ABSTRACT

We propose to use the Gamma filter as a continuous time spectral feature extractor for the preprocessing of speech signals. The Gamma filter is a simple analog structure which can be implemented as a cascade of identical first order lowpass filters. The filter generates the *Poisson moments* of the input signal at its taps. These moments carry spectral information about the recent history of the input signal and in return they can be used to construct a time-frequency representation alternative to the conventional methods of short-term Fourier transform, cepstrum, etc. The appeal of the proposed method comes from the fact that in the analog domain the Poisson moments are readily available as a continuous time electrical signal and can be physically measured, rather than computed offline by a digital computer. With this convenience, the speed of the discrete time processor following the preprocessor is independent of the highest frequency of the input signal, but is constrained by the stationarity interval of the signal. The moments can be directly fed into artificial neural networks (ANNs) for tasks like classification and identification of timevarying signals like speech.

INTRODUCTION

The Gamma filter is a *generalized feedforward structure* that is composed of a cascade of identical first order lowpass filters. Up to now, this filter has found several application areas in system identification, speech recognition and echo cancellation [1][2]. In this study we show that the taps of the Gamma filter carry valuable information about the frequency spectrum of the input signal in the form of *Poisson moments*. If the filter is implemented as an analog VLSI circuit, the Poisson moments can be physically measured. This convenience significantly lowers the computational cost. With the inclusion of a forgetting factor in the filter, these moments can be used to build a practical time-frequency representation in continuous time. Furthermore, they can be directly fed into ANNs for tasks like classification and identification of timevarying signals.

GAMMA FILTER AND THE POISSON MOMENTS

Fairman and Shen [3] proposed that a distribution f(t) can be expanded in terms of the derivatives of Dirac's delta function as follows

$$f(t) = \sum_{i=0}^{\infty} f_i(t_o)\, e^{-\lambda(t-t_o)}\, \delta^{(i)}(t-t_o) \qquad (1)$$

$f_i(t_0)$ is called the i^{th} *Poisson Moment* [4] of f(t) at $t=t_0$. It is given by

$$f_i(t_o) = f(t) \otimes p_i(t)\big|_{t=t_0}$$
$$p_i(t) = \frac{t^i}{i!} e^{-\lambda t} u(t) \qquad \lambda > 0 \qquad (2)$$

'⊗' stands for the convolution operator, while u(t) stands for the unit step function. $p_i(t)$ can be recognized as the impulse response of a cascade of i+1 identical lowpass filters. This structure is known as the *Gamma filter* [1]. λ is called the *time scale* of the filter and it is responsible for adjusting the region of support of the impulse response $p_i(t)$. Equation (2) suggests that, instead of computing $f_i(t_0)$ offline, one can physically measure it as the value at the $i+1^{st}$ tap of a Gamma filter with input f(t) (Figure 1). This low computational cost makes the moments appealing.

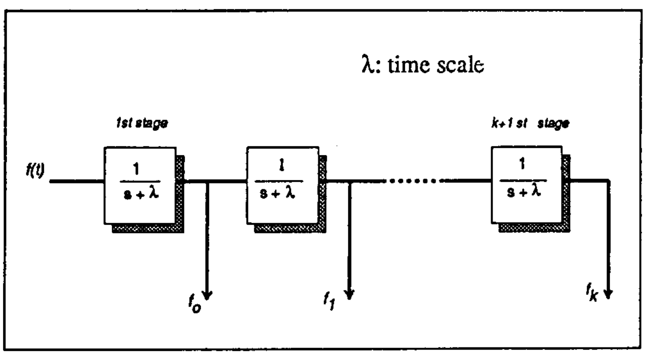

Figure 1 Poisson moments are generated by the Gamma filter

0-7803-2431-5/95 $4.00 © 1995 IEEE

POISSON MOMENTS AND THE INPUT SIGNAL SPECTRUM

Having seen how easily the Poisson moments can be obtained let's examine the relationship between the moments and the input signal spectrum. Assume a causal signal f(t) and expand (2) as

$$f_i(t_o) = \int_{-\infty}^{\infty} f(t_o - t)\, rect\left(\frac{t - t_o/2}{t_o}\right) e^{-\lambda t} \frac{t^i}{i!} dt \qquad (3)$$

Next, introduce $f_W(t)$, the decaying exponential and rectangular windowed version of the delayed and inverted f(t), such that

$$f_i(t_o) = \int_{-\infty}^{\infty} f_W(t) \frac{t^i}{i!} dt \qquad (4)$$

Manipulating (3), it can be shown that the i^{th} Poisson moment at time t_o is related to $F_W(\Omega)$ as follows

$$f_i(t_o) = \frac{1}{i!} \frac{d^i}{d\Omega^i} F_W(\Omega) |_{\Omega = 0} \qquad f_W(t) \xrightarrow{F} F_W(\Omega) (5)$$

As far as the magnitude spectrum goes (which is generally the main concern in speech recognition) delay and time inversion have no effect on the magnitude spectrum. The decaying exponential windowing however helps to achieve locality in time at the cost of blurring in the frequency domain. There is also a rectangular windowing, but for $t_o > 1/\lambda$ its effect can be safely ignored. Therefore, we can state that $|F_W(\Omega)|$ approximates the magnitude spectrum of the recent history of the original signal f(t) with a frequency resolution λ.

Having established the relationship between $|F_W(\Omega)|$ and the magnitude spectrum of the original signal f(t), let's go back to (5). It is apparent from this equation that the i^{th} Poisson moment of f(t) is the i^{th} Taylor's series approximation coefficient of $F_W(\Omega)$ around $\Omega=0$. Using these coefficients a polynomial approximation of the magnitude spectrum can be easily constructed. A similar argument can be brought out for the Poisson moments, too. They can be easily obtained by measuring the tap outputs of the Gamma filter and can be used to form a vector that alone represents the recent input signal spectrum. The moment vector can be directly fed into an artificial neural network (ANN) for tasks like prediction, identification and classification [2][7][8].

PIECEWISE SPECTRAL APPROXIMATION

Evidently, the Poisson moments carry valuable information about the blurred spectrum of the input signal in the form of Taylor's series coefficients at the pivot point $\Omega=0$. Given the moments, construction of a Taylor series approximation of the spectrum becomes a trivial problem. One shortcoming of the Taylor's series expansion however, is its locality, i.e. approximation diverges at points away from the pivot. Hence, it is not possible to characterize a wide bandwidth of frequencies using a finite number of Poisson moments. As a cure for that shortcoming, one can partition the frequencies of interest into several small bands and do the approximations for each band seperately, thereby obtaining piecewise representation of a wide bandwidth of frequencies. In practice, bandpass filters tuned to different center frequencies followed by mixers can be used to shift each band to the origin. The baseband signals can be fed into the Gamma filter to obtain the corresponding moment vectors. Consequently, these moment vectors can be concatenated together to give the overall spectral picture (Figure 2).

Tracey and Principe [2] used a similar scheme in their simple word recognition task using artificial neural networks. A group of constant Q bandpass filters were used to model the *cochlea* [5]. The authors also replaced the mixer with a cascade of a square device and an envelope detector, thereby roughly approximating the power spectrum rather than the magnitude spectrum. As described above the baseband signals were fed into Gamma filters, outputs of which are the Poisson moments. These moments are further fed into an ANN for word classification.

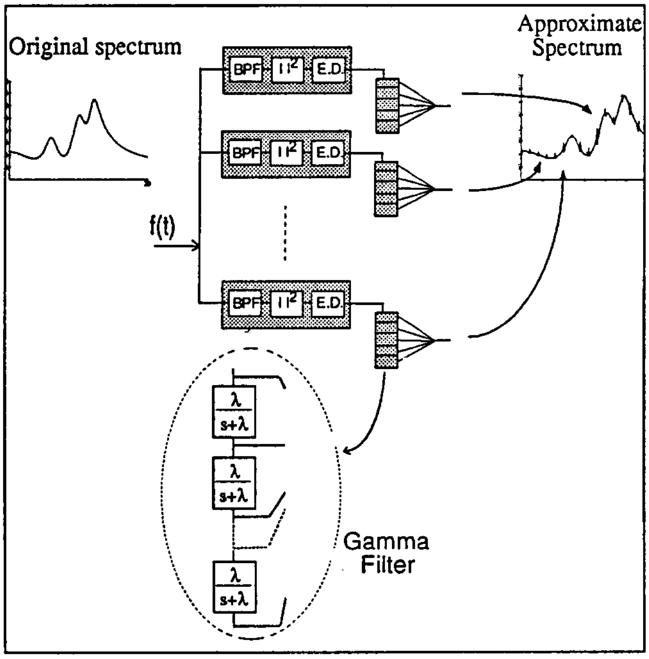

Figure 2 Spectral Approximation via Poisson moments

INTERPOLATION VIA POISSON MOMENTS

In this section we will give the form of an interpolating polynomial that approximates the magnitude spectrum .

Let's partition the frequency axis into N bands whose lowest cutoff frequencies are $\underline{\Omega} = \{0, \Omega_1, \Omega_2, \ldots, \Omega_{N-1}\}$ such that the width of the j^{th} band will be $W_{j+2}-W_{j+1}$. Eventually, we will have a moment vector

$$M_j = [f_{oj}(t_o), f_{1j}(t_o), \ldots f_{K-1,j}(t_o)]^T \qquad j = 0, 1, \ldots N-1$$

Although the approximation of $F_W(\Omega)$ can be carried out along all the frequency bands yielding a single, large order polynomial, due to its simplicity we will prefer to do the interpolation band by band yielding several small order polynomials. Again, these piecewise approximations can be put together to yield a global approximation.

Given two Poisson moment vectors M_j, M_{j+1} that belong to adjacent bands a polynomial $F_{app}(\Omega)$

$$F_{app_j}(\Omega) = \sum_{r=0}^{2K-1} a_{rj}\Omega^r \qquad (6)$$

$$A_j = [a_{oj}, a_{1j}, \ldots, a_{2K-1,j}]^T$$

can be constructed to approximate the j^{th} band. The coefficients of the polynomial are related to the Poisson moments as follows

$$A_j = C_j^{-1}\begin{bmatrix} M_j \\ ----- \\ M_{j+1} \end{bmatrix} \qquad C_j = [\Delta_j | \Delta_{j+1}]^T_{2Kx2K}$$

where the $(n,m)^{th}$ element of the matrix Δ_j is $\Omega_j(n+m)!/m!$. Coefficient vector A_j is a linear weighted sum of the vectors M_j and M_{j+1}. If an ANN is used to process the coefficients A_j, the neural network can be made to learn the weight matrix C_j^{-1}, too. With this convenience, the moments can be directly fed into the ANN.

EXAMPLE

Figure 3 illustrates the magnitude spectrum of a 20 msec segment of the vowel /e/ and its approximation obtained using Poisson moments. The frequency axis was divided into 15 bands of 160 Hz each. Each band was shifted to the origin and filtered by a Gamma filter of order 4, thereby creating Poisson moment vectors of size 4. Poisson moments were further used to approximate the original magnitude spectrum. The Poisson derived spectrum shows a better definition of the pitch frequency, while showing the first two formants equally well. For comparison purposes, we also included in this figure the special case of the approximation where one moment per band is used. This corresponds to the sampling of the frequency axis. The improvement in the approximation gained by the utilization of additional Poisson moments is obvious.

CONCLUSIONS

In this study we have shown how the Gamma filter can be used to form a time-frequency representation of its input. The representation is readily available at the taps of the Gamma filter in the form of Poisson moments. Compared to conventional spectral representation schemes like Fourier series or cepstral coefficients, this is a computationally inexpensive method. The discrete time processor that operates on the moments is not constrained by the Nyquist rate of the input signal, but by the rate moments vary. The highest frequency of the input signal affects the number of bands that need to be implemented to cover the required bandwidth with a given precision. In a sense, this method trades speed for parallelism, since each frequency band operates totally independent of the others. For spectral analysis of very high frequency signals that can not be digitized with the present technology, this method is very appealing. Analog VLSI chips can be fabricated to implement the analog bandpass filters and the Gamma structure, where the Poisson moments will be measured.

One problem that needs to be addressed is the selection of the time scale λ. There are two conflicting factors in the choice of the time scale λ. [7] showed that in order to maximize the region of convergence of the approximation along the frequency axis, λ has to be as close as possible to zero. With a small λ, however, it is not possible to obtain time resolution. λ can be made a positive number at the cost of using more bands. In that respect λ should be selected such that the decaying exponential window will support the stationarity duration of the signal.

ACKNOWLEDGMENTS

This work has been supported by ONR grant N00014-94-1-0858.

REFERENCES

[1] B. de Vries, J. C. Principe, "The Gamma Model- A New Neural Model for Temporal Processing," *Neural Networks*, vol. 5, pp. 565-576, 1992.

[2] Tracey J.W., Principe J.C., "Isolated-Word Speech Recognition Using the Focused Gamma Neural Network," *World Congress on Neural Networks*, Portland, Oregon, vol III, pp. 87- 90, July 1993.

[3] Fairman F.W., Shen D.W.C.,"Parameter Identification for Linear Time-Varying Dynamic Processes," *Proc. IEE*, vol. 117, no. 10, Oct. 1970.

[4] Saha D.C., Rao G.P., *Identification of Continuous Dynamical Systems- The Poisson Moment Functional Approach.* Springer-Verlag, New York, 1983.

[5] Pickles J.O., *An Introduction to the Physiology of Hearing*, 2nd edition, London:Academic Press, 1988.

[6] de Boor, C., *A Practical Guide to Splines.* Springer-Verlag, New York, 1978.

[7] Celebi S., Principe J.C., 'Spectral Feature Extraction Using Poisson Moments,' *1993 Workshop on Neural Networks for Signal Processing*, Ermioni, Greece, 1993.

[8] Celebi S., Principe J.C., "Analysis of Spectral Feature Extraction Using the Gamma Filter," Proc. of IEEE Int. Conf. on Neural Networks, Orlando FL, vol VII, pp. 4494-4501, June 1994.

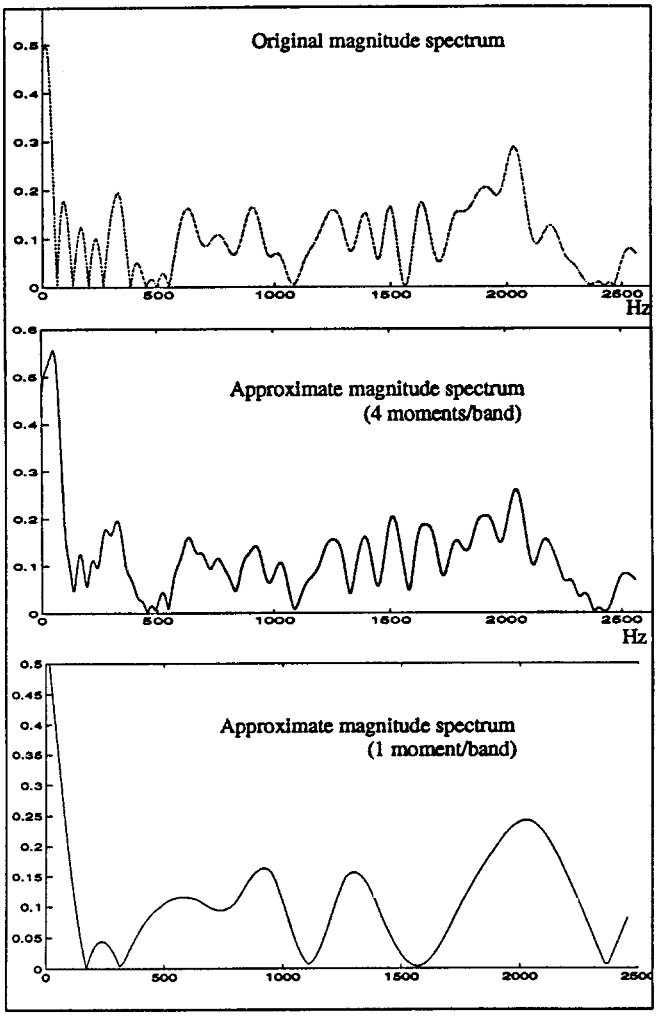

Figure 3 Original and the approximate spectra of vowel /ε/ obtained by Poisson moments