Time-Varying Filtering of Speech Signals Using Linear Prediction

Srdjan Stanković, Jan Tilp

Abstract— **Time-varying filtering of noisy speech signals based on linear prediction (LP) is presented. This approach is tested on noisy speech signals apparent in hands-free telephone systems.**

I. Introduction

Basic difficulties in filtering of noisy speech signals arise from their highly nonstationary and multicomponent nature [6]. When dealing with short-time speech spectra, the dynamic range of those components is usually as high as 40 $[dB]$. This means that the energies of speech and noise have vary significantly at different frequencies. By using linear prediction (LP) [3], we can simultaneously consider the excitation part of speech signals (with approximately constant spectral amplitudes) and a remaining part obtained by subtraction of the excitation signal from the speech signal itself. This remaining part of the speech signal will be referred to as the pseudo envelope. The aim of this paper is to separatey apply time-varying filtering [2], [5] to the excitation signal and to the pseudo envelope of speech signals. An efficient method for time-varying filtering of speech signals is thus obtained. The basic concept is illustrated by an experimental example of filtering a noisy speech signal in a hands-free telephone system.

II. Linear Predictive Analysis

A linear predictive analysis of a speech signal is based on an all-pole model [3], where the speech signal is defined by:

$$s(n)=\sum_{i=1}^{p}a_i(n)s\left(n-i\right)+e\left(n\right). \qquad (1)$$

The error or LP residual (i.e., the excitation) is denoted by $e(n)$, and $a_i(n)$ are the weights

Electronics Letters, Vol.36, No.8, Apr.2000.

applied to the previous speech samples, for any time n. The weights correspond to the direct-form coefficients of a non-recursive filter, the transfer function of which is given as:

$$A(n,z)=1-\sum_{i=1}^{p}a_i(n)z^{-i}. \qquad (2)$$

Passing the speech signal through the filter $A(n,z)$ results in the removal of the near-sample correlations and produces the excitation $e(n)$. The LP analysis has both time- and frequency-domain interpretations. The key aspect for applying LP in our approach is that it is possible to extract the excitation signal and, on the basis of the excitation signal, the pseudo envelope, as well. To keep the signal delay short, it is possible to determine the coefficients $a_i(n)$ sequentially with an adaptive LMS-type algorithm.

III. Time-varying filtering

Time-varying filtering of the noisy signal $x(t)=s(t)+n(t)$ can be performed as follows [2], [5]:

$$(Hx)(t)=\int_{-\infty}^{\infty}h(t+\frac{\tau}{2},t-\frac{\tau}{2})w(\tau)x(t+\tau)d\tau. \qquad (3)$$

Using Parseval's theorem, equation (3) can be written in the form:

$$(Hx)(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}L_H(t,\omega)STFT_x(t,\omega)d\omega, \qquad (4)$$

where the impulse response $h(t+\frac{\tau}{2},t-\frac{\tau}{2})$ is defined using Weyl symbol mapping $L_H(t,\omega)$ into the time frequency plane, i.e., $h(t+\frac{\tau}{2},t-\frac{\tau}{2})=\int_{-\infty}^{\infty}L_H(t,\omega)e^{-j\omega\tau}d\omega$, and $STFT_x(t,\omega)=FT\{w(\tau)x(t+\tau)\}$ is the short-time Fourier transform. The lag window is denoted by $w(\tau)$.

For optimal time-varying filtering, when the signal and noise are not correlated, $L_H(t,\omega)$ can be derived (in analogy with the Wiener filter) in the form [2], [5]:

$$L_H(t,\omega) = 1 - \frac{\overline{WD}_{nn}(t,\omega)}{\overline{WD}_{xx}(t,\omega)}, \qquad (5)$$

where $\overline{WD}_{xx}(t,\omega) = \int_{-\infty}^{\infty} E\{x(t+\tau/2)x^*(t-\tau/2)\}e^{-j\omega\tau}d\tau$ is the mean value of the Wigner distribution (WD) of the noisy signal, i.e. the Wigner spectrum (WS), and $\overline{WD}_{nn}(t,\omega)$ is the WS of the noise. In general, $\overline{WD}_{ss}(t,\omega)$ will eliminate uncorrelated cross-terms in the WD, because $E\{s_i(t+\tau/2)s_j^*(t-\tau/2)\} = 0$, for uncorrelated components $s_i(t)$ and $s_j(t)$, [1]. Thus, the problem of cross-terms does not exist in the WS if we are able to use a large number of random signals belonging to the same random process. In most practical applications filtering must be based on a single realization of the noisy signal. Therefore, it is not possible to use averaging of the WD, i.e., the WS. It is for the reason why we substitute the WD in equation (5) with the SM, [4], as a cross-terms free version of the WD:

$$L_H(t,\omega) = 1 - \frac{SM_{nn}(t,\omega)}{SM_{xx}(t,\omega)}, \qquad (6)$$

where $SM_{xx}(t,\omega)$ and $SM_{nn}(t,\omega)$ represent the SM of the noisy signal $x(t)$ and noise $n(t)$, respectively. The discrete version of the SM is given by [4]:

$$SM_{xx}(n,k) = |STFT_x(n,k)|^2 + 2Re\left\{\sum_{l=1}^{L} STFT_x(n,k+l)STFT_x^*(n,k-l)\right\}, \qquad (7)$$

where $|STFT_x(n,k)|^2$ is the spectrogram, while $2L+1$ is the frequency-domain window length; n and k are the discrete time and frequency, respectively.

IV. Algorithm

In the LP analysis, a $p-th$ order adaptive LP-error filter (2) is applied to the noisy speech signal $x(n)$. We obtain two components: (i) the excitation signal $x_{exc}(n)$ and (ii) the pseudo envelope $x_{penv}(n)$ of the speech signal. Time-varying filtering will be applied to both of these components, i.e. using (6), we obtain $L_{excH}(n,k)$ and $L_{penvH}(n,k)$. The resultant Weyl symbol mapping is obtained by: $L_H(n,k) = L_{excH}(n,k) + L_{penvH}(n,k)$. Applying (4), while using the resultant $L_H(n,k)$, the filtered signal is obtained. A block diagram of this filtering approach is presented in Fig.1.

V. Experimental Illustration

In our example we take a noisy speech signal present in a hands-free telephone system. The signal was recorded in a car cruising travelling along a motorway. The STFT was calculated using a 256-samples rectangular window. Zero padding up to 1024 samples was carried out, and the SM with $L = 3$ was calculated. A modified version of (6) with a spectral floor β was applied [6]: $L_H(t,\omega) = \max\{1 - SM_{nn}(t,\omega)/SM_{xx}(t,\omega), \beta)$, where $\beta = 0.12$. Estimations of the SM of noise were performed during speech pauses. Prefiltering, using a high-pass filter with cut-off frequency $90\,Hz$, was applied prior to the time-varying filter procedure. The duration of the speech components was $2.6\,s$. The noisy speech signal (filtered by a high-pass filter) and the signal filtered using the previously described algorithm are presented in Figs.2a) and b). Figure 2a) shows that noise is present in the whole time-frequency plane. In particular, the noise has very high energy in the frequency range from $90\,Hz$ to $1kHz$. This frequency range is very important for the perception of speech signals. It is obvious that the energy of the noise is often as high as that of the speech signal itself. Figure 2b) shows that noise suppression and enhancement of the speech signal are rather high. Also, it can be noted that the noise is well suppressed between speech components (as can clearly be seen in the above stated frequency range). This is the result of our of time-varying filtering approach, where the filtering of the excitation plays the crucial role. As a measure of noise suppression, we calculated the ratio of noise energies during a speech pause (which was 1500 samples long) for the high-pass-filtered noisy signal and the

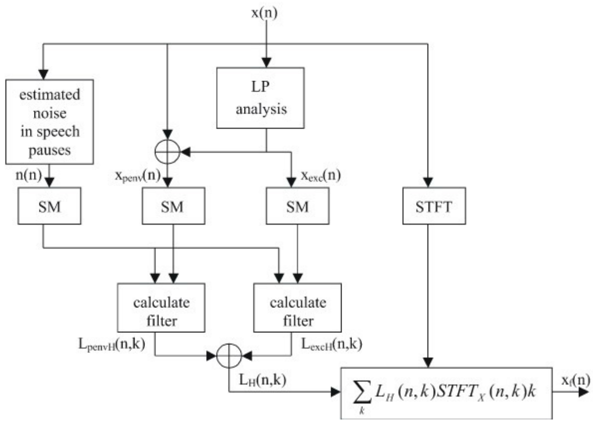

Fig. 1. Block scheme

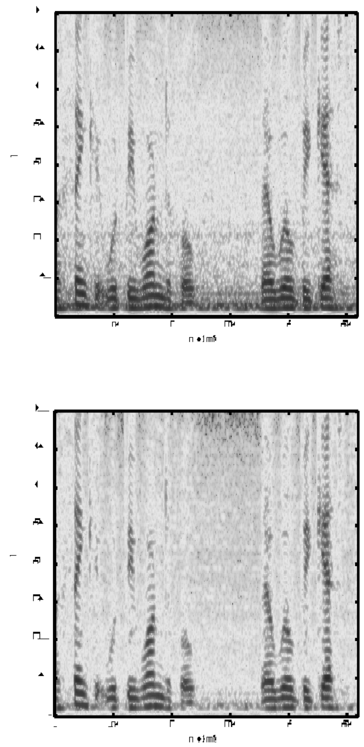

Fig. 2. Time-frequency representation (spectrogram): a) Noisy speech signal prefiltered by high-pass filter; b) Denoised speech signal filtered by the LP-based time-varying filtering.

denoised signal, filtered according to the LP-based time-varying filtering. This ratio is 9.57, i.e., approximately 10 $[dB]$.

VI. Conclusion

In this paper, LP has been used in the time-varying filtering of noisy speech signals. By using the described filtering approach it is possible to obtain an efficient filter which is able to significantly reduce noise in speech signals. Our theory has been illustrated using the example of filtering noisy speech signals in hands-free telephone systems.

References

[1] J. F. Boehme, D. Konig: "Statistical processing of car engine signals for combustion diagnostic", *Proc. IEEE 7th Workshop on statistical signal and array processing*, Quebec 1994, pp. 369-374.

[2] G. Matz, F. Hlawatsch, W. Kozek: "Generalized evolutionary spectral analysis and the Weyl spectrum of nonstationary random processes", *IEEE Transactions on Signal Processing*, vol.45, No.6, June 1997, pp.1520-1534.

[3] L. R. Rabiner, R. W. Schafer: "Digital Processing of Speech Signals", Englewood Cliffs, NJ, *Prentice Hall*, 1978.

[4] LJ. Stanković: "A method for time-frequency analysis", *IEEE Transactions on Signal Processing*, vol.42, No.1, Jan. 1994, pp.225-229.

[5] LJ. Stanković: "On the time-frequency analysis based filtering", *Annales des telecommunications*, vol.54, No. 5/6, May/June, 2000.

[6] N. Virag: "Single channel speech enhancement based on the masking properties of the human auditory system ", *IEEE Transactions on Speech and Audio Processing*, vol.7, no.2, March 1999, pp.126-137.