TIME-DOMAIN NOISE EXTRACTION IN MOBILE COMMUNICATION SPEECH SIGNALS

Voicu FILIMON[+1], **Maurice BELLANGER**[++]

[+] Ohio State University, Electro-Science Lab., 1320 Kinnear Road,
43212 Columbus, OH, USA
[++]Conservatoire National des Arts et Metiers, 292 Rue Saint Martin,
75141 Paris, CEDEX 03, France

RÉSUMÉ

Le but de cet article est de présenter une méthode de débruitage temporel pour la parole en radiocommunication mobiles, en supposant que le bruit du moteur est essentiellement une somme des sinusoïdes de fréquence variable. Un algorithme adaptatif rapide appliqué à un prédicteur en cellules du seconde ordre est utilisé pendant les pauses de parole pour identifier les sinusoïdes en temps réel et les retirer. Pendant les périodes parlées, le bruit est retiré en utilisant des filtres à bande étroite suivant les fréquences préalablement déterminées. Les résultats d'écoute sont satisfaisants est ils sont présentés sous forme graphique.

ABSTRACT

This paper describes a method of time-domain noise cancellation in a mobile communication setting, based upon the assumption of noise being a periodic waveform related to engine activity and using pauses in speech to linearly predict and thus remove the noise. During speech intervals, noise is further removed using narrow band filters tracing the frequencies previously detected. Frequency identification in speech pauses is done in real time using a gradient algorithm to decompose the predictor into its second order components. Auditive experiments are successful and results are visually presented.

1. INTRODUCTION

Elimination of noise affecting speech signals in car phone communication is desirable since engine related noise often proves highly damaging to speech. Two approaches are generally used, which differ in the way apriori knowledge about noise is collected in order to perform the extraction. One approach uses another way to prelevate noise, independent of the vocal path, usually a second microphone or noise simulation after measurement of the rotation speed of the engine. The other method relies on pauses in speech, which often occur in conversations, to identify noise. Again there exist two ways of further proceeding. One consists of a frequency approach to noise reduction by attenuating speechless frequency bands or by subtracting the noise spectrum out of the signal. The second way attempts an extraction in the time domain, thus avoiding the inherent errors due to the Short Fourier Transform which can affect the quality of speech. It is this latter method that we will employ.

Noise will be adaptively identified in speechless periods using a prediction algorithm and an additional assumption on noise content based on the car environment: namely that it consists of a sum of sinusoids possibly of variable frequency. In these periods the predictor will also delete the noise. When speech is present, that is when the adaptive algorithm can no longer latch onto the noise, a method of following the changes in frequency of the identified disturbing sinusoids and of further extracting them is developed, thus fulfilling the task of noise elimination. The two phases of the method are further described.

2. IDENTIFICATION OF NOISE

Prediction of noise in speechless intervals will be done using the linear predictor in Fig.1, which is already decomposed into its second order terms for easy 'on-line' recognition of the sinusoidal components. Consider the predictor H to be of the form:

$$H(z) = \prod_{m=1}^{p} H_m(z) = \prod_{m=1}^{p} (1 + a_{1m} z^{-1} + a_{2m} z^{-2}) \qquad (1)$$

If the signal to be predicted consists of a sum of sinusoids, the coefficients a_{1m} and a_{2m} provide information about:

- the presence of the sinusoids if :

$$a_{2m} = 1$$

[1] supported by a 'Tempus' grant through a research stage at CNAM, Paris

and the number of unitary a_{2m} coefficients indicates the number of sinusoids.

- frequency of the sinusoids , because :

$$a_{1m} = -2 \cdot \cos(\frac{2\pi}{8192} k_m) \qquad (2)$$

when $a_{2m} = 1$, with k_m being the discrete frequency and 8192 the sampling rate.

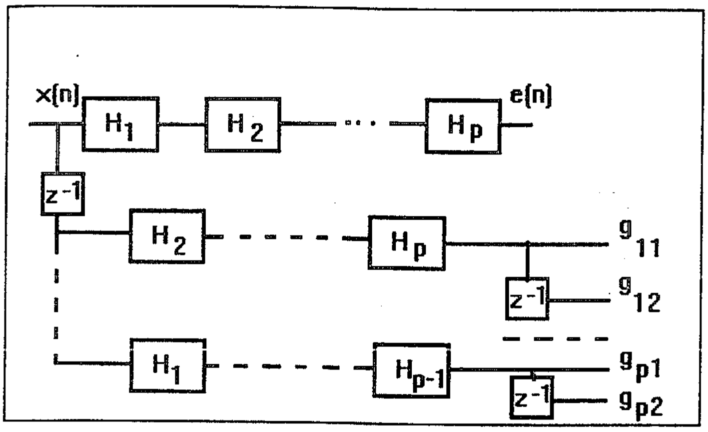

Fig.1 Linear predictor of order 2p as a cascade of second order terms.

So in order to identify the sinusoids one would have to decompose filter H. In order to avoid a time consuming search for roots, this can be done on-line by updating the coefficient vector A= $[a_{11}\ a_{21}\ ....a_{1p}\ a_{2p}]$ using a gradient algorithm operating in the tracking mode.

The function to be minimized according to Fig. 1 is:

$$J = E\{e^2_{(n)}\} = E\{(H \otimes x)^2_{(n)}\}$$

with the gradient of e with respect to each coefficient a_{im} [2] :

$$g_{im} = \frac{\partial\,(e_{n+1})}{\partial\,(a_{im})} = Z^{-1}\{z^{n-i}\frac{H(z)}{H_m(z)}X(z)\} \quad i = 1,2 \quad m = 1,p$$

which can also be seen to comply with the implementation in Fig. 1. Let $G=[g_{11}\ g_{12}\ g_{21}\ g_{22}\ .....\ g_{p1}\ g_{p2}]^T$ be the gradient vector, then the update relation for the coefficient vector A turns out to be:

$$A_{(n+1)} = A_{(n)} - R^{-1}{}_{GG} \cdot G \cdot e \qquad (3)$$

with R_{GG} the autocorrelation matrix of the gradient sequence.

We implemented the recursion in (3) using the Fast Least Square (FLS) algorithm [1], chosen for its speed and stability and for providing the possibility of tracking variable frequencies. This is a viable alternative because the correlation matrix of the gradient sequence g_{ij} tends to be block diagonal since its coefficients are highly correlated in each block : $g_{2m}(n)= g_{1m}(n-1)$, but little correlated between different blocks: $R_{GG}(0,i,m_1,j,m_2)\sim 0$ if $m_1 \neq m_2$. The FLS algorithm described in equations (4) was applied to each of the second order sections in filter H in Fig.1 with partial gradient sequences for input, thus in effect implementing the recursion in (3). Here e_1 and e_2 stand for forward and backward prediction error, FP and BP for forward and backward predictor and K for gain. $W < 1$ is the forgetting factor, E the starting error, while g is the input signal with G the vector containing the input sequence, which actually consists of the gradient values g_{im} in Fig.1.

$$\begin{aligned}
&e_1(n+1) = g(n+1) - FP(n)\cdot G(n)^T \\
&FP(n+1) = FP(n) + K(n)\cdot e_1(n+1) \\
&\varepsilon_1(n+1) = g(n+1) - FP(n+1)\cdot G(n)^T \\
&E(n+1) = W\cdot E(n) + e_1(n+1)\cdot\varepsilon_1(n+1) \\
&\begin{bmatrix} M(n+1) \\ m(n+1) \end{bmatrix} = \begin{bmatrix} 0 \\ K(n) \end{bmatrix} + \frac{\varepsilon_1(n+1)}{E(n+1)}\cdot\begin{bmatrix} 1 \\ -FP(n+1) \end{bmatrix} \qquad (4) \\
&e_2(n+1) = g(n+1-N) - BP(n)\cdot G(n+1)^T \\
&K(n+1) = \frac{M(n+1) + m(n+1)\cdot BP(n)}{1 - m(n+1)\cdot e_2(n+1)} \\
&BP(n+1) = BP(n) + K(n+1)\cdot e_2(n+1)
\end{aligned}$$

Dimension N in our case is 2. After obtaining the forward and backward predictor FP and BP the gain K is used to further compute:

$$e = H^T \cdot X$$

$$\begin{bmatrix} a_{1m}(n+1) \\ a_{2m}(n+1) \end{bmatrix} = \begin{bmatrix} a_{1m}(n) \\ a_{2m}(n) \end{bmatrix} - K \cdot e \qquad (5)$$

for each m in (1..p), with vector X containing the signal sequence in Fig.1. Differentiation with respect to the roots of predictor H is achieved by updating only p-1 of the p second order blocks at one iteration, each starting with the same initial value.

If the input signal x is periodic, thus perfectly predictable, the output e will rapidly tend to zero when using the FLS. Thus at the end of every speech pause, the frequencies k_m and their number are already identified using relation (2) if the a_{2m} coefficients are equal to 1.

3. NOISE REMOVAL IN SPEECH INTERVALS

The cascade structure in Fig.1 also allows for easy removal of the noise sinusoids during the speech periods, when the FLS is no longer used. Adding to the second order FIR filters an IIR part one can obtain a frequency efficient Notch stop-band filter described by the following equation:

$$\begin{aligned}
y_{m(n)} &= x_{m(n)} + a_{1m}\cdot x_{m(n-1)} + a_{2m}\cdot x_{m(n-2)} - \\
&-(1-\varepsilon)\cdot a_{1m}\cdot y_{m(n-1)} - (1-\varepsilon)^2\cdot a_{2m}\cdot y_{m(n-1)} \qquad (6)
\end{aligned}$$

where x_m and y_m are the input and output of the m^{th} second order filter and $\varepsilon > 0$ is a small constant. This can be achieved since, as can be seen, the IIR coefficients are directly related to the FIR part. The number of these filters applied in this phase of the noise removal depends on the number of sinusoids identified in the previous speechless interval and the coefficients a_{im} are the ones determined by the adaptive algorithm. An advantage of this kind of extractor with respect to a predictor generating the sinusoids for subsequent extraction is that changes in the amplitude of the noise do not matter.

The filter just described also removes signal components of the same frequency with the noise sinusoids. It is an inconvenience but not a large one according to the experimental results. To insure that the frequency modifications of the noise sinusoids are traced even in the speech intervals, the output of the filter H in Fig.1 is controlled, independently of the general signal output, by pairs of two bandpass Notch filters, with central frequencies of $k_m \pm$ a small offset, according to the relations:

$$
\begin{aligned}
z_{m1(n)} &= \varepsilon \cdot y_{m(n)} + (1-\varepsilon)\cdot a'_{1m}\cdot z_{m1(n-1)} + (1-\varepsilon)^2 \cdot a_{2m}\cdot z_{m1(n)} \\
z_{m2(n)} &= \varepsilon \cdot y_{m(n)} + (1-\varepsilon)\cdot a''_{1m}\cdot z_{m2(n-1)} + (1-\varepsilon)^2 \cdot a_{2m}\cdot z_{m2(n)}
\end{aligned} \tag{7}
$$

The number of pairs corresponds to the number of sinusoids previously detected in the speechless interval. Coefficients a_{1m}' and a_{1m}'' are obtained from (2) by adding or subtracting a fixed offset to frequency k_m , whereas coefficients $a_{2\,m}$ are equal to 1 if a sinusoid was previously detected.

$$
\begin{aligned}
a'_{1m} &= -2 \cdot \cos\left(\frac{2\pi}{8192}(k_m + offset)\right) \\
a''_{1m} &= -2 \cdot \cos\left(\frac{2\pi}{8192}(k_m - offset)\right)
\end{aligned} \tag{8}
$$

By detecting the envelope of the z_{mi} filters, a simple algorithm for varying the central frequency of the H_m filters in Fig.1 along with the frequency shifts in the noise sinusoids can be developed by taking advantage that one is in the discrete domain:

$$\text{if} \quad anv(z_{m1}) \gtrless anv(z_{m2}) \Rightarrow k_{m(n+1)} = k_{m(n)} \pm \Delta \tag{9}$$

where Δ is a chosen constant and 'anv' stands for the envelope detector. It can happen as will be seen in the experimental results, that the output of the filters z_{mi} is affected at certain instants by signal components in the frequency band that these filters pass. In this case the frequency k_m is lost, but since speech pauses occur frequently thus restarting the adaptive prediction algorithm, the overall loss in signal is little.

In this way noise extraction is simply performed in both periods with or without speech, taking into account that all the algorithms described, amount to filtering in the discrete domain with few coefficients. In contrast to frequency-domain noise extraction, this algorithm is far less computational intensive.

4 EXPERIMENTAL RESULTS

Experiments where conducted on portions of speech signals perturbed by sinusoids of variable frequency. Fig.2 shows a 3 seconds long portion of a speech signal. Noise was simulated by adding two different chirp signals with frequencies increasing and decreasing in this interval. Fig. 3 shows the result of adding the noise to the initial speech signal. Auditive tests prove that the disturbance is fairly significant.

The signal in Fig. 3 is further processed using the algorithms previously described to produce the output in Fig.4. During speechless portions of the signal the FLS algorithm applied to the gradient sequences of the second order filter coefficients easily finds the frequency of the sinusoids and latches onto it. Figs. 5 and 6 present the variable frequency detected during processing and Fig. 7 shows the two pairs of coefficients of the second order components of predictor filter H in (1), which adapt to the two disturbing sinusoids. As can be seen in this figures there exists an initial period of adaptation for the FLS algorithm, but this does not extend to further speechless intervals since initial conditions are set in the preceding speech periods.

During speech intervals the algorithm using narrow band Notch filters is used to extract the noisy sinusoids out of the signal and at the same time to keep track of the noise frequencies with the help of the offset filters in (6). Figures 5 and 6 show how the two chirp signals are tracked during these intervals . Parameter ε and the offset in (6) and (7) were equal to 0.1 and 40 respectively. There are moments in which the frequency of the noise sinusoids is lost due to the frequency content of the speech signal, but since speech pauses in which the adaptive algorithm relatches onto the sinusoids are frequent, this is not very damaging. Most of the time as can be seen the algorithm performs very well.

Fig.7 shows the separation of the roots of the predictor filter in Fig. 1. The FLS algorithm applied to the gradient sequences of the second order filter coefficients performs fast and well, and as can be seen the a_{2m} coefficients go to 1, signaling the presence of a sinusoid, whereas the a_{1m} coefficients indicate its frequency according to (2).

Comparing figures 2 and 4 one can ascertain the quality of noise removal. The output in Fig. 4 was audio tested and proved to be quite satisfactory.

BIBLIOGRAPHY

[1] M. Bellanger "Adaptive Digital Filters and Signal Analysis" Marcel Dekker Inc., New York, 1987

[2] L.B. Jackson, S.L Wood "Linear Prediction in Cascade Form", IEEE Trans. On ASSP-26, December 1978

[3] D.W. Tufts, R. Kumaresan "Estimation of Frequencies of Multiple Sinusoids: Making Linear Prediction Perform Like Maximum Likelihood" Proceedings of the IEEE, Vol 70, No 9, Sept 1982

[4] J.M. Romano et al. "Least Squares Adaptive Filters in Cascade Form for Line Pair Spectrum Modeling", Proceedings EUSIPCO-90 Conference, North Holland, Barcelona, Spain, Sept 1990

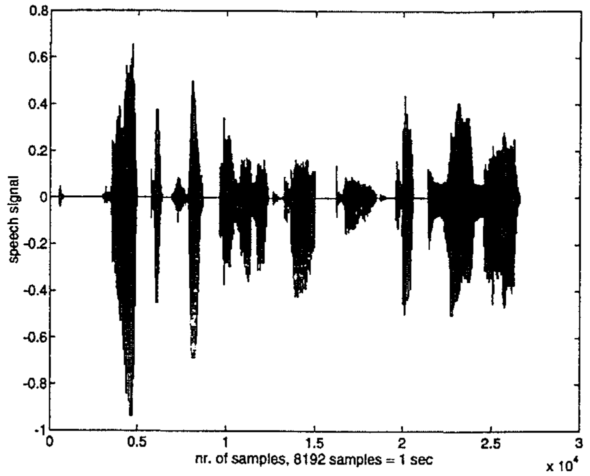

Fig.2 : a 3 seconds long speech signal

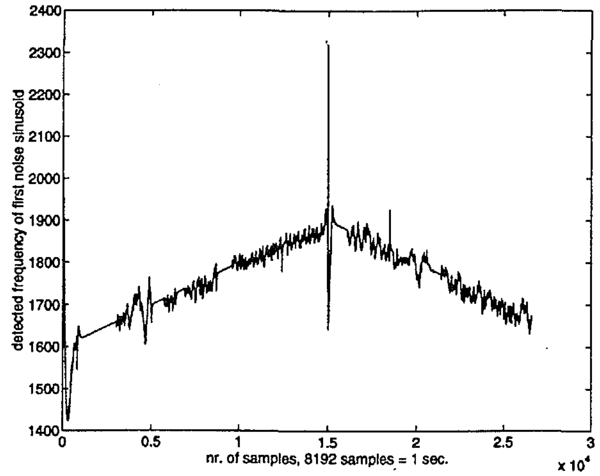

Fig.5 : frequency of the first chirp signal as detected by the FLS and the algorithm using narrow band filters

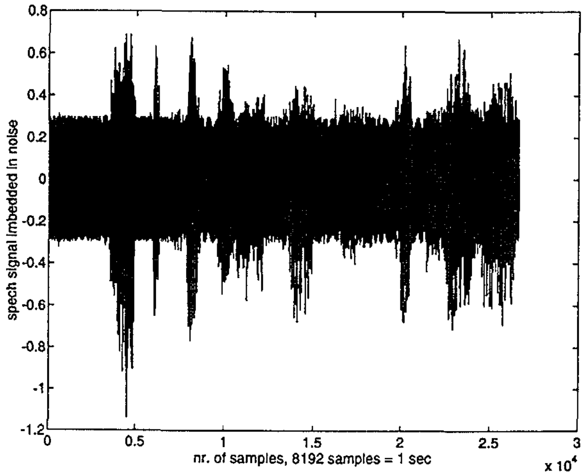

Fig.3 : two chirp signals added to the initial speech signal to result into a noisy signal

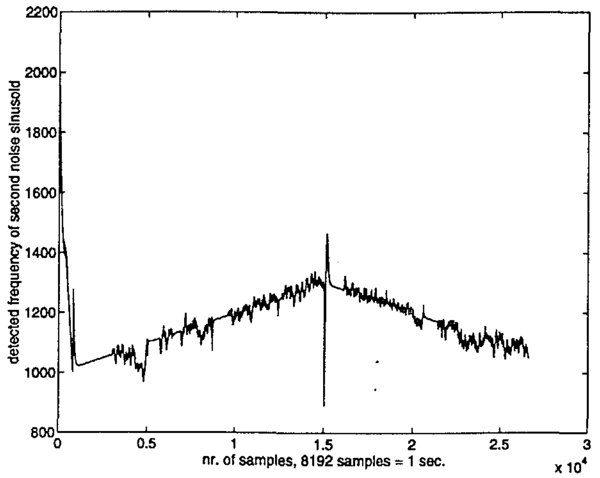

Fig.6 : frequency of the second chirp signal as detected by the FLS and the narrow band filter algorithm

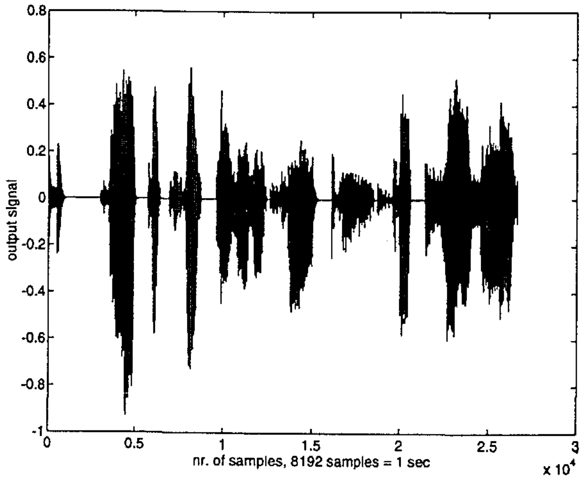

Fig.4 : output signal after the FLS algorithm and the narrow band filters were applied on-line

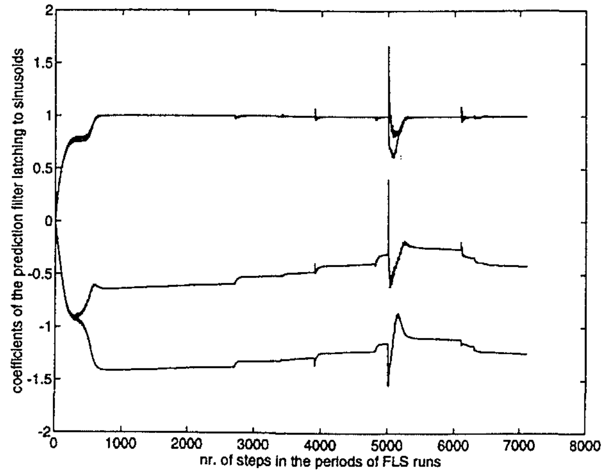

Fig. 7 : pairs of coefficients of the prediction filters latching onto the noise sinusoids during speechless intervals in which the FLS algorithm is applied