SPEECH ENHANCEMENT USING A WIENER FILTERING UNDER SIGNAL PRESENCE UNCERTAINTY

A. AKBARI AZIRANI - R. LE BOUQUIN JEANNÈS - G. FAUCON

Laboratoire du Traitement du Signal et de l'Image - Université de Rennes 1
Bât. 22 - Campus de Beaulieu - 35042 RENNES CEDEX - FRANCE
Regine.Lebouquin@univ-rennes1.fr

Abstract

Noise reduction is a key-point of speech enhancement systems in hands-free communications. A number of techniques have been already developed in the frequency domain such as an optimal short-time spectral amplitude estimator proposed by Ephraim and Malah in [1] including the estimation of the *a priori* signal-to-noise ratio. This approach reduces significantly the disturbing noise and provides enhanced speech with colorless residual noise. In this paper, we propose a technique based on a Wiener filtering under uncertainty of signal presence in the noisy observation. Two different estimators of the *a priori* signal-to-noise ratio are tested and compared. The main interest of this approach comes from its low complexity.

1. INTRODUCTION

In this paper a technique for enhancing a speech signal degraded by uncorrelated stationary additive noise is investigated in the scope of hands-free telecommunication systems. The problem of speech enhancement and mainly noise reduction in speech remains a key-point of such systems. Generally, in a moving car, speech is degraded by ambient noises due to the engine, traffic and wind and the Signal to Noise Ratio (SNR) is low. A great number of techniques have been already studied [2,3]. Some of them are based on the well-known spectral subtraction approach that is suitable for enhancing speech embedded in stationary noise. In these methods there remains usually a level of residual, unnatural background noise, called musical noise.

The main interest of the technique we propose comes from its colorless residual noise and its low complexity. It uses the same mechanism that eliminates the residual musical noise in the method proposed by Y. Ephraim and D. Malah (called Ephraim and Malah estimator) [1]. We propose a Wiener filtering which takes into account the uncertainty of signal presence and is function of the *a priori* SNR and the *a posteriori* SNR. The *a priori* SNR is estimated recursively using the former estimates of the speech spectrum.

2. AMPLITUDE ESTIMATOR

2.1. Ephraim and Malah estimator

Let $s(t)$ and $n(t)$ denote the speech and noise processes respectively. The observed signal $x(t)$ may be written:

$$x(t) = s(t) + n(t), \tag{1}$$

which is equivalent in the frequency domain to:

$$X(f,k)e^{j\varphi_X(f,k)} = S(f,k)e^{j\varphi_S(f,k)} + N(f,k)e^{j\varphi_N(f,k)} \tag{2}$$

where $X(f,k)$, $S(f,k)$ and $N(f,k)$ are the amplitudes of the Short Time Fourier Transforms (STFT) of the signals $x(t)$, $s(t)$ and $n(t)$ on each frame k of 256 samples respectively and $\varphi_X(f,k)$, $\varphi_S(f,k)$, $\varphi_N(f,k)$ determine their phases. A well-known amplitude estimator of the signal $\hat{S}(f,k)$ that minimizes the mean-squared spectral error is the conditional mean:

$$\hat{S}(f,k) = E[S(f,k)|X(f,k)] \tag{3}$$

where the expectation operator $E[.]$ is used to indicate averaging over the ensemble of noise sample functions, speech envelopes and phases and the ensemble of speech events. The above estimator is developed for the two states "silence and nonsilence". We obtain an estimator composed of a weighted sum of individual estimators relative to the speech signal in the two states. The weights are the posterior probabilities of the two states given the noisy signal. Since the optimal estimator of the clean signal given that this signal is absent in the noisy observation equals zero, the resulting composite estimator is the product of the estimator of the clean signal given that this signal is present in the noisy observation and the posterior probability of signal presence given the noisy signal. This estimator depends on the *a priori* SNR $R_{prio}(f,k)$ and on the *a posteriori* SNR $R_{post}(f,k)$ [4] defined as:

$$R_{prio}(f,k) = \frac{E\left[S^2(f,k)\right]}{E\left[X^2(f,k)\right]} \tag{4}$$

and

$$R_{post}(f,k) = \frac{X^2(f,k)}{E\left[N^2(f)\right]} - 1 \tag{5}$$

The estimate of the signal amplitude is given by:

$$\hat{S}(f,k) = G_1(f,k)G_2(f,k)X(f,k) \tag{6}$$

where $G_1(f,k)$ is the optimal amplitude estimator:

$$G_1(f,k)=\frac{\sqrt{\pi}}{2}\frac{\sqrt{V(f,k)}}{1+R_{post}(f,k)}\exp\left(-\frac{V(f,k)}{2}\right)\cdot \left[(1+V(f,k))I_0\left(\frac{V(f,k)}{2}\right)+V(f,k)I_1\left(\frac{V(f,k)}{2}\right)\right] \quad (7)$$

and $G_2(f,k)$ is a term which takes into account the uncertainty of signal presence:

$$G_2(f,k)=\frac{\Lambda(f,k)}{1+\Lambda(f,k)} \quad (8)$$

where $\Lambda(f,k)$ is the generalized likelihood ratio:

$$\Lambda(f,k)=\frac{1-q(f,k)}{q(f,k)}\frac{\exp(V(f,k))}{1+R_{prio}(f,k)} \quad (9)$$

and

$$V(f,k)=\frac{R_{prio}(f,k)}{1+R_{prio}(f,k)}\left(1+R_{post}(f,k)\right) \quad (10)$$

$I_0(.)$ and $I_1(.)$ denote the modified Bessel functions of zero and first order respectively and $q(f,k)$ is the probability of signal absence in the spectral component *f*.

The important idea to reduce the musical noise effect in the Ephraim and Malah estimator lies in the estimation of the *a priori* SNR [1,5] given by the non-linear recursive relation:

$$R_{prio}(f,k)=\lambda\frac{S^2(f,k-1)}{E\left[N^2(f)\right]}+(1-\lambda)P\left[R_{post}(f,k)\right] \quad (11)$$

$S(f,k-1)$ is the amplitude estimator of the signal spectral component in the $(k-1)$th analysis frame. $E\left[N^2(f)\right]$ is the estimate of the noise power spectral density (psd) learned in the intervals where speech is absent. λ is a weighting factor close to 1 and $P[.]$ is an operator defined by:

$$P[u]=\begin{cases}u & \text{if } u\geq 0\\ 0 & \text{otherwise}\end{cases} \quad (12)$$

to ensure the positiveness of $R_{post}(f,k)$.

The *a posteriori* SNR is computed on each frame *k* using (5).

2.2. Proposed method

In the new method, the optimal amplitude estimator $G_1(f,k)$ is replaced by the Wiener filtering $W(f,k)$:

$$W(f,k)=\frac{R_{prio}(f,k)}{R_{prio}(f,k)+1} \quad (13)$$

and as previously, the posterior probability $G_2(f,k)$ is expressed in terms of the *a priori* SNR and the *a posteriori* SNR and is computed as in (8).

Two estimators of the *a priori* SNR are studied. The first one is given by (11). We propose a second estimator of this SNR:

$$R_{prio}(f,k)=\lambda\frac{S^2(f,k-1)}{E\left[N^2(f)\right]}+(1-\lambda)P\left[\frac{E\left[X^2(f,k)\right]}{E\left[N^2(f)\right]}-1\right] \quad (14)$$

where $E\left[X^2(f,k)\right]$ is the psd of the noisy observation evaluated in each short-time frame as follows:

$$E\left[X^2(f,k)\right]=\lambda_2 E\left[X^2(f,k-1)\right]+(1-\lambda_2)X^2(f,k) \quad (15)$$

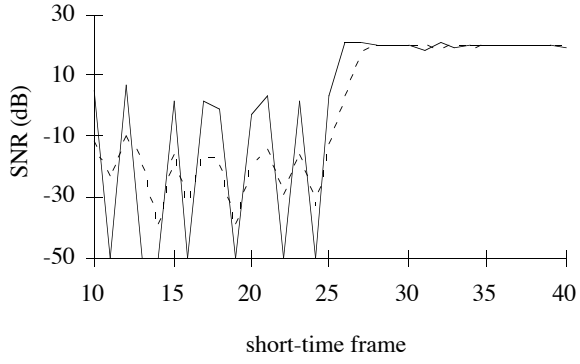
and λ_2 is a forgetting factor between 0.5 and 1.

3. ESTIMATION OF THE *A PRIORI* SNR

In the following, for reasons of clarity, the indices *f* and *k* are dropped. Simulations have been performed to analyze the variations of R_{prio} and R_{post}. The analyzed signal contains noise only in the first 10 frames; for the next 15 frames the input SNR at 1 kHz is -15 dB; for the last 15 frames it is equal to 20 dB. λ is set to 0.98 as in [1]. The probability q is set to 0.5. Fig. 1 represents the first estimator R_{prio} according to eq. (11) and R_{post} at 1 kHz versus the short-time frame number on the last two sequences. For low SNR, where R_{post} has large variations, R_{prio} is a smoothed version of R_{post} and its variance is much smaller than that of R_{post}. For high input SNR, R_{prio} follows R_{post} with a delay of one frame. Since the Wiener filter in (13) is only a function of R_{prio}, the musical noise due to the fluctuations of the disturbances is strongly reduced for low input SNR.

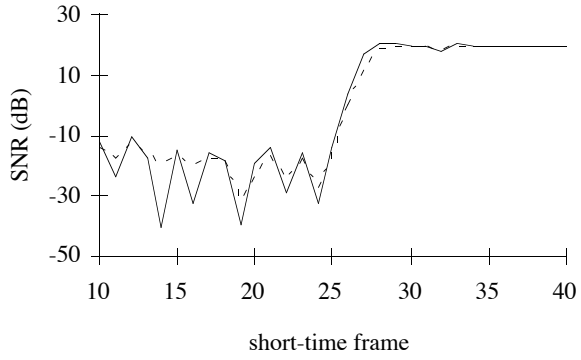
The smoothness of the *a priori* SNR helps reducing the musical noise and decreasing the fluctuations of R_{prio} in the low input SNR is desirable. We can not use a parameter λ too close to one in (11) because the averaged value of R_{prio} decreases strongly in the low input SNR [5]. Moreover, it introduces an important delay between the appearence of the transient component and the moment when R_{prio} reaches 20 dB.

In Fig. 2, the variations of the new estimator of R_{prio} given by eq. (14) is compared to that of (11) for the same input signal and the same parameter λ. The forgetting factor λ_2 is equal to 0.7. It is obvious that in the low input SNR the fluctuations of R_{prio} are reduced using (14) and R_{prio} reaches 20 dB with a delay slightly larger than that of (11) in the transient period.



——— *a posteriori* SNR at 1 kHz
- - - - - *a priori* SNR at 1 kHz estimated by (11)

Fig. 1. Estimated *a priori* SNR and *a posteriori* SNR versus short-time frame



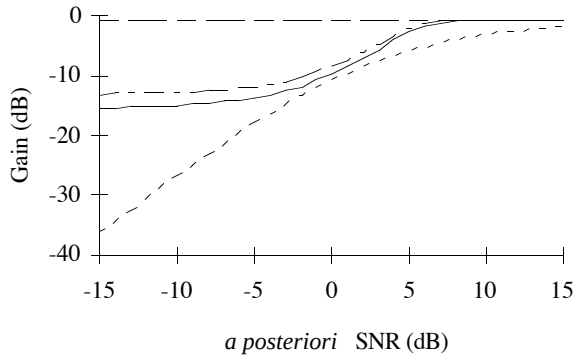
——— *a priori* SNR at 1 kHz estimated by (11)
- - - - - *a priori* SNR at 1 kHz estimated by (14)

Fig. 2. Estimated *a priori* SNR versus short-time frame

4. COMPARISON OF THE GAIN CURVES

We compare the gain of the proposed estimator $W(f,k)G_2(f,k)$ to 3 other estimators that are used in speech enhancement: the magnitude spectral subtraction, the Wiener filter in (13) that does not take into account the uncertainty of speech presence in the noisy signal and the Ephraim and Malah estimator.

The gain curves versus *a posteriori* SNR are depicted in Fig. 3 and 4 for *a priori* SNR of 10 and -10 dB respectively. As it can be seen the magnitude spectral subtraction is a function of the *a posteriori* SNR only and has the largest variations compared with the 3 other techniques. Fig. 3 corresponds more particularly to the predominant speech components (high *a priori* SNR), whereas Fig. 4 shows the gain curves for high noise components (low *a priori* SNR).

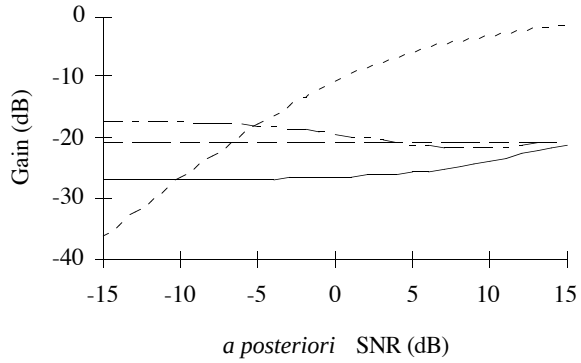


- - - - - Magnitude spectral subtraction
— - — Ephraim and Malah estimator
— — Wiener filter
——— Proposed method

Fig. 3. Gain curves for the *a priori* SNR of 10 dB

For the 3 techniques depending on the *a priori* SNR, the attenuation is lower than that of the magnitude spectral subtraction for high *a priori* SNR (Fig. 3). In consequence, this last approach leads to some undesirable distortion on the useful signal. As for $R_{prio} = -10$ dB, the attenuation of the same 3 techniques is greater when $R_{post} > -5$ dB. In this way, noise components are well suppressed compared to the magnitude spectral subtraction.

Concerning the gains obtained by the Ephraim and Malah estimator and the proposed technique, they are comparable, but globally, the gain of the new method is slightly lower. It is important to note that for low *a priori* SNR (Fig. 4) this technique seems better because its gain increases along with the *a posteriori* SNR contrary to the previous one. Consequently, for low *a priori* and *a posteriori* SNR, the noise components are well attenuated using the new technique.

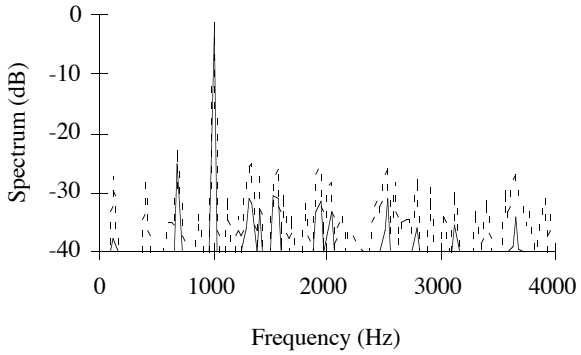


- - - - - Magnitude spectral subtraction
— - — EME
— — Wiener filter
——— Proposed method

Fig. 4. Gain curves for the *a priori* SNR of -10 dB

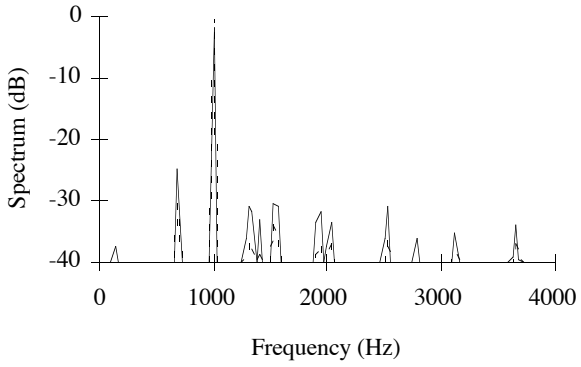
5. SIMULATIONS

The above results are confirmed by the following simulation: a sinusoidal signal at 1 kHz is added to a white noise to obtain a SNR of 6 dB at this frequency. The SNR is defined as the ratio of the sinusoidal signal power to the variance of the noise. The short-time spectrum of the signal estimated by Ephraim is compared to that of the proposed technique where the *a priori* SNR is estimated by (11) (Fig. 5). The STFT is realized by a 256 point Fourier transform using a Hamming window and the sampling rate is 8 kHz. The overlap between successive frames is equal to 75%. As can be seen from this figure, the 1 kHz signal component estimated by these two techniques has the same spectral amplitude but the level of noise components is lower using the new estimator. Fig. 6 corresponds to the estimation of the speech signal by the proposed method using either (11) or (14). This figure shows that the noise components are more attenuated using (14) but the level of the sinusoidal signal is the same.



- - - - - STFT amplitude of the signal estimated by Ephraim
——— STFT amplitude of the signal estimated by the proposed method using (11)

Fig. 5. Comparison of the STFT amplitudes of the signals estimated by the Ephraim and Malah estimator and by the proposed method



——— STFT amplitude of the signal using (11)
- - - - - STFT amplitude of the signal using (14)

Fig. 6. Comparison of the STFT amplitudes of the signals estimated by the proposed method using (11) or (14)

6. RESULTS

The proposed estimator has been compared with the Ephraim and Malah estimator (Eph. & Mal.) and with the classical magnitude spectral subtraction (Spect. Sub.). The speech signals are recorded in a stopped car. Two english sentences are pronounced by a male and a female speaker. The car noise is recorded in the car moving at 130 km/h. Noise is added to speech signals to obtain a global SNR equal to 0 dB. Two objective measures - the gain G in the segmental SNR and the cepstral distance d_{cep} - are computed. Listening tests are conducted to appreciate the quality of the processed files. Thirty listeners have to judge the resulting files in terms of distortion, residual noise, and defaults brought by the processing. For each file, a note between 1 and 5 is given by each listener according to the following scale:
1: bad. 2: poor. 3: fair. 4: good. 5: excellent.

Table 1 presents for each method the objective measures and the mean opinion score obtained from the listening test.

Method	G (dB)	d_{cep}	score
Spect. Sub.	15.1	0.39	2.87
Eph. & Mal.	15.9	0.34	3.32
New method	15.9	0.50	3.35

Table 1: Objective and subjective results

The estimator developed by Ephraim and the proposed method have a gain in the SNR which is 0.8 dB greater than that of the magnitude spectral subtraction. The lowest cepstral distance is obtained by Ephraim. Concerning the listening test, the new method has a score slightly better than the two other ones. The listeners indicate that the disturbing noise and the musical noise are well reduced as using the Ephraim and Malah estimator.

7. CONCLUDING REMARKS

A Wiener filter which takes the uncertainty of signal presence into account in the noisy observation was presented. It was compared with the magnitude spectral subtraction, the standard Wiener filter and the estimator developed by Ephraim. Simulations have shown that the level of residual noise is lower. Informal listening tests confirm this remark and reveal that the proposed approach results in a significant noise reduction and provides enhanced speech with colorless residual noise. The noise is less annoying than that of the spectral subtraction. The complexity of the new approach is comparable with the complexity of the spectral subtraction and lower than that of the Ephraim and Malah estimator.

References

[1] Y. EPHRAIM, D. MALAH, "Speech Enhancement Using a Minimum Mean Square Error Short-Time Spectral Amplitude Estimator", IEEE Trans. on ASSP, vol. ASSP-32, n°6, pp. 1109-1121, December 1984.

[2] J.S. LIM, A.V. OPPENHEIM, "Enhancement and Bandwidth Compression of Noise Speech", Proceedings of the IEEE, vol. 67, n°12, pp. 1586-1604, December 1979.

[3] J.H.L. HANSEN, J.R. DELLER, "Speech Enhancement and Quality Assessment with Applications to Robust Recognition and Coding", Tutorial, ICASSP, May 1995.

[4] R.J. McAULAY, M.L. MALPASS, "Speech Enhancement Using a Soft-Decision Noise Suppression Filter", IEEE Trans. on ASSP, vol. ASSP-28, n°2, pp. 137-145, April 1980.

[5] O. CAPPÉ, "Elimination of the Musical Noise Phenomenon with the Ephraim and Malah Noise Suppressor", ICASSP, pp. 345-349, 1994.

Acknowledgment. The authors wish to thank Matra Communication (Paris) for the database.