HAL
open science

Psychoacoustical improvement of Wiener filtering : some recent approaches and a new method

Asmaa Amehraye, Dominique Pastor, Ahmed Tamtaoui

To cite this version:

Asmaa Amehraye, Dominique Pastor, Ahmed Tamtaoui. Psychoacoustical improvement of Wiener filtering : some recent approaches and a new method. [Research Report] Traitement Algorithmique et Matériel de la Communication, de l'Information et de la Connaissance (Institut Mines-Télécom-Télécom Bretagne-UEB); Institut national des postes et télécommunications de Rabat (Institut national des postes et télécommunications de Rabat). 2007, pp.13. hal-02316497

HAL Id: hal-02316497

https://hal.science/hal-02316497v1

Submitted on 15 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Collection
des Rapports
de Recherche
de l'ENST Bretagne

RR-2007001-SC

Psychoacoustical improvement of Wiener filtering : some recent approaches and a new method

Amélioration psychoacoustique du filtrage de Wiener : quelques approches récentes et une nouvelle méthode

Rapport Interne GET / ENST Bretagne

2007

Asmaa AMEHRAYE
Dominique PASTOR
Ahmed TAMTAOUI

Psychoacoustical improvement of Wiener filtering: some recent approaches and a new method

Amélioration psychoacoustique du filtrage de Wiener: quelques approches récentes et une nouvelle méthode

Rapport Interne GET / ENST Bretagne

Asmaa Amehraye*, **Dominique Pastor**
Ecole Nationale Supérieure des Télécommunications de Bretagne,
CNRS TAMCIC (UMR 2872),
Technopôle de Brest Iroise, CS 83818,
29238 BREST Cedex, FRANCE
e-mail: `asmaa.amehraye, dominique.pastor@enst-bretagne.fr`

Ahmed Tamtaoui
Institut National des Postes et Télécommunications,
2, av ALLal EL Fasse, Madinat AL Irfane, Rabat, Morocco
e-mail: `tamtaoui@inpt.ac.ma`

*Asmaa Amehraye is also with Faculté des Sciences de Rabat, GSCM-LRIT, B.P. 1014, Morocco

Abstract

This paper deals with musical noise resulting from subtractive type algorithms and especially Wiener filtering. We compare several methods that introduce perceptually motivated modifications of the standard Wiener filtering and we propound a new speech enhancement technique. This one aims to improve the quality of the enhanced speech signal provided by the standard Wiener filtering by controlling the latter via a second filter regarded as a psychoacoustically motivated weighting factor. According to objective measures and the observation of some spectrograms, the described process results in significant reduction of musical noise.

Keywords

Musical noise, Wiener filtering, psychoacoustics, speech distortion.

Résumé

Dans ce papier, on s'intéresse à la réduction du bruit de type musical qu'engendrent des méthodes basées sur la soustraction de bruit et en particulier, le filtrage de Wiener. On compare plusieurs méthodes qui introduisent des modifications du filtre de Wiener, ces modifications étant basées sur les propriétés du système auditif humain. Nous proposons aussi une nouvelle méthode. Celle-ci améliore la qualité de la parole débruitée en sortie du filtre de Wiener usuel. Cette amélioration résulte d'un contrôle du filtre de Wiener par un second filtre qui peut être considéré comme un facteur de pondération perceptuelle. En se basant sur l'observation des spectrogrammes et des mesures objectives de qualité de la parole débruitée, la méthode que nous proposons apporte une amélioration significative du bruit musical en comparaison avec les autres méthodes traitées dans ce papier.

Keywords

Bruit musical, distorsion, filtre de Wiener, psychoacoustique, signal de parole.

Contents

1 Introduction **4**

2 The standard filtering and its limitations **5**

3 Perceptually motivated speech denoising **6**

4 Experimental Results **9**

5 Conclusion **11**

1 Introduction

The objective of a speech enhancement process is to improve the quality and intelligibility of speech in noisy environments. The problem has been largely discussed over the years. Many approaches have been proposed. Basic methods are subtractive type algorithms such as those described in [1], [2]. Such methods return residual noise known as musical noise. This type of noise turns out to be quite annoying. In order to reduce the effect of musical noise, several solutions have been proposed. Some involve adjusting parameters of the spectral substraction so as to offer more flexibility as in [3] and [4]. Others, such as that proposed in [5], are based on signal subspace approaches. Despite the effectiveness of those techniques to improve the Signal to Noise Ratio (SNR), the problem of eliminating or reducing musical noise is still a challenge to many researchers.

In the last few decades, the introduction of psychoacoustic models has attracted a great deal of interest. The objective is to improve the perceptual quality of the enhanced speech signal. In [4], a psychoacoustic model is used to control the parameters of the spectral subtraction in order to find the best trade-off between noise reduction and speech distortion. To make musical noise inaudible, the linear estimator proposed in [6] incorporates the masking properties of the human auditory system. In [7], the masking threshold of tones and an intermediate signal, which is slightly denoised and free of musical noise, are used to detect musical tones generated by spectral subtraction methods. This detection can be used by a post-processing aimed at reducing the detected tones.

Even though the psychoacoustic models are usually developped in the frequency domain, signal subspace approaches can also involve perceptual models by resorting to some suitable frequency to eigendomain transform as described in [8], [9], [10].

In this work, we are particularly interested by methods related to the standard Wiener filter for two reasons. First, the Wiener filter is easy to implement. Second, it can reasonably be expected that if we succeed in reducing the perception of residual noise resulting from Wiener filtering, the quality of the denoised speech will be improved and yield a rather satisfactory listening comfort.

On the basis of such remarks, the authors in [11] propose to apply the Wiener filter only when noise is audible and, thus, to not process frequency components where noise is masked. Similarly, in [12], a perceptually motivated modification is applied to the Wiener filtering of the noisy speech signal sub-band components, these components being calculated via a filterbank.

In the present paper, we propose to control the standard Wiener filtering

by a psychoacoustically motivated filter that can be regarded as a weighting factor. The purpose is to minimise the perception of musical noise without degrading the clarity of the enhanced speech. We compare the proposed method to those introduced in [11], [12] and [13] when the noisy speech signals are analysed in the Fourier domain. This is the reason why we adapt the method proposed in [12] to the frequency domain.

The organization of this paper is as follows. Section 2 reminds the reader with the basics concerning the standard Wiener filtering of noisy speech signals. With the same notations and assumptions of section 2, section 3 introduces several enhancement processes, amongst which the new method we propose. Section 4 presents the performance evaluation by means of objective measures and the observation of spectrograms. Section 5 concludes this paper.

2 The standard filtering and its limitations

The observed noisy speech signal is assumed to be some speech signal additively corrupted by independent noise. The processing is performed frame by frame in the frequency domain. Each frame involves the same number M of samples. For the kth frame, let $s_k(t), n_k(t)$ and $y_k(t)$, $t = 0, 1, \ldots, M-1$, stand for the M samples of the speech signal, noise and the observed noisy speech signal, respectively. We thus have $y_k(t) = s_k(t) + n_k(t)$. Now, let $Y_k(\nu)$, $S_k(\nu)$ and $N_k(\nu)$, $\nu = 0, \ldots, M-1$, denote the Discrete Fourier Transform (DFT) coefficients of $y_k(t)$, $s_k(t)$ and $n_k(t)$, $t = 0, 1, \ldots, M-1$, respectively. For every $\nu = 0, 1, \ldots, M-1$, we have $Y_k(\nu) = S_k(\nu) + N_k(\nu)$.

Basic speech enhancement approaches consists in estimating every frequency component $S_k(\nu)$ by $\widetilde{S}_k(\nu) = H_k(\nu)Y_k(\nu)$ where $H_k(\nu)$ is an estimator chosen according to a suitable criterion. The error signal generated by this estimator is

$$\begin{aligned} e_k(\nu) &= \widetilde{S}_k(\nu) - S_k(\nu) \\ &= (H_k(\nu) - 1)S_k(\nu) + H_k(\nu)N_k(\nu). \end{aligned} \tag{1}$$

The values $(H_k(\nu)-1)S_k(\nu)$ are the DFT coefficients of the speech distortion due to the filtering and the frequency components $H_k(\nu)N_k(\nu)$ are the residual noise DFT coefficients. Musical noise then results from pure tones present in residual noise. The Wiener filtering based on Malah's decision-directed approach (see [2]) is one of the most famous method aimed at reducing musical noise. In this case, the estimator is $H_k(\nu) = W_k(\nu)$ and $\widetilde{S}_k(\nu) = W_k(\nu)Y_k(\nu)$ is the Wiener estimate of $S_k(\nu)$ where $W_k(\nu)$ is here-

after called the Wiener gain function given by

$$W_k(\nu) = \xi_k(\nu)/(1 + \xi_k(\nu)) \tag{2}$$

where

$$\xi_k(\nu) = (1-\alpha)h\left(\chi_k(\nu) - 1\right) + \alpha \frac{|\widetilde{S}_{k-1}(\nu)|^2}{\gamma_k(\nu)} \tag{3}$$

is the so-called decision-directed estimate of the *a priori* SNR

$$\mathbb{E}\left[|S_k(\nu)|^2\right]/\mathbb{E}\left[|N_k(\nu)|^2\right]. \tag{4}$$

In Eq. (3), $\widetilde{S}_{k-1}(\nu) = W_{k-1}(\nu)Y_{k-1}(\nu)$ is the νth spectral component of the Wiener denoised speech signal in frame $k-1$; $\gamma_k(\nu)$ is the estimate of $\mathbb{E}\left[|N_k(\nu)|^2\right]$; $h(x) = x$ if $x \geq 0$ and $h(x) = 0$ otherwise; $\chi_k(\nu) = |Y_k(\nu)|^2/\gamma_k(\nu)$ is the estimate of the *a posteriori* SNR $|Y_k(\nu)|^2/\mathbb{E}\left[|N_k(\nu)|^2\right]$; the weighting factor α is set to 0.98 for a good compromise between musical noise and speech distortion [2] .

The estimate $\xi_k(\nu)$ takes into account the current frame, with weight $(1-\alpha)$, and the result of the processing of the previous frame, with weight α. The smoothing character of the decision-directed approach reduces the level of the musical noise, which, however, remains present and perceptually annoying.

3 Perceptually motivated speech denoising

The block diagram of figure 1 summarizes the different speech enhancement processes discussed in this section and compared in the next one. It allows for some improvement of the Wiener filtering of frame k by choosing $F_k(\nu)$ equal to $W_k(\nu)$ and introducing perceptual criteria through the filter $G_k(\nu)$. The purpose is to achieve a good compromise, in a perceptual sense, between residual noise and speech distortion. For frame k, the resulting estimate $\widehat{S_k}(\nu)$ of $S_k(\nu)$ is $\widehat{S_k}(\nu) = H_k(\nu)Y_k(\nu)$, where $H_k(\nu) = G_k(\nu)F_k(\nu) = F_k(\nu)G_k(\nu)$.

Figure 1 also points out that the computation of the masking threshold $T_k(\nu)$ will be based on the Wiener estimate of the clean speech signal for all the methods described below. The masking threshold could also be estimated on the basis of the outcome of a spectral substraction as in [4]. However, the tone-like nature of musical noise increases the energy per critical band and the presence of too much musical noise can therefore induce an overestimation of the masking threshold. The Wiener estimate is thus preferable because the

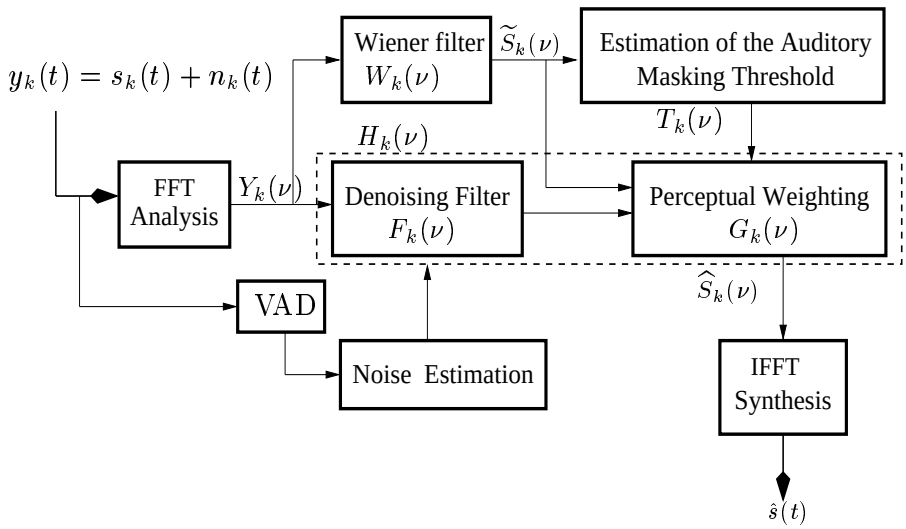

Figure 1: Block diagram of the proposed enhancement process

Wiener filter introduces less musical noise than spectral subtraction methods [2]. In this paper, the power spectrum of the noisy speech is estimated on the basis of signal-free time frames, which amounts using an ideal Voice Activity Detector (VAD).

Before introducing the speech enhancement method we propose, we describe two recent techniques that can be regarded as perceptually motivated modifications of the Wiener filter. The first one, described in [12], can be regarded as a Wiener filtering of only the amount of noise that exceeds the masking threshold. In [12], this approach is applied to the sub-band components obtained by using an auditory filterbank. In fact, this method can easily be adapted to the usual case where the time-frequency analysis is performed by the standard DFT. With respect to the block diagram of figure 1, it involves choosing

$$\begin{cases} F_k(\nu) = 1, \\ G_k(\nu) = |\widetilde{S}_k(\nu)|^2/(|\widetilde{S}_k(\nu)|^2 + \max\left(\gamma_k(\nu) - T_k(\nu), 0\right)) \end{cases} \tag{5}$$

where $\widetilde{S}_k(\nu)$ is the Wiener estimate defined before.

In the second method, introduced in [11], the Wiener filtering is controlled by the result of the comparison between noise and the masking threshold. This comparison makes it possible to perform the denoising only for the noise frequency components that are audible in the sense that their amplitudes exceed the masking threshold. Comparing to the block diagram of figure 1 and with the same notations as those used so far, the perceptually motivated

modification of the Wiener filter proposed in [11] consists in setting

$$\begin{cases} F_k(\nu) = 1, \\ G_k(\nu) = \begin{cases} W_k(\nu) & \text{, if } \gamma_k(\nu) > T_k(\nu) \\ 1 & \text{, otherwise.} \end{cases} \end{cases} \tag{6}$$

Remark 3.1 At this stage, it is crucial to note the following. Perceptual methods basically aim at reducing noise without introducing much distortion. The gain in SNR is not the main objective. It follows that noise is not fully eliminated since only its audible frequency components are reduced. Therefore, noise components that are not audible, thanks to some maskers in the original noisy signal, can be still present after denoising and even become audible if the maskers are filtered.

The method we introduce now is an attempt to overcome this type of drawback by using a filter G that acts as a perceptual weighting factor controlling the Wiener gain function. Among the several perceptual filters $G_k(\nu)$ described in this section (see also Eq. (8) below), we have chosen the filter given in Eq. (5) because, according to the experimental results of the next section, it performs better than those specified by Eqs. (6) and (8). Therefore, comparing to the block diagram of figure 1, we set

$$\begin{cases} F_k(\nu) = W_k(\nu), \\ G_k(\nu) = |\widetilde{S}_k(\nu)|^2/(|\widetilde{S}_k(\nu)|^2 + \max\left(\gamma_k(\nu) - T_k(\nu), 0\right)) \end{cases} \tag{7}$$

The following analysis describes some properties of this "double filtering". If $\gamma_k(\nu) < T_k(\nu)$, which means that noise is inaudible in frame k, we have $G_k(\nu) = 1$. The Wiener filter is however applied for two reasons. First, it favours the gain in SNR. Second, it reduces the risk that non audible noise components might become audible after the filtering of audible maskers present in the original noisy signal (see remark 3.1). Note that if $\gamma_k(\nu) \ll T_k(\nu)$, that is, when the SNR is very good, the Wiener filter and the perceptual weighting factor both equal 1 so that no distortion is introduced.

When $\gamma_k(\nu) > T_k(\nu)$, we couple the high noise suppression capability of the Wiener filtering with the effect of the weighting factor so as to enhance the speech quality and reduce musical noise. In the limit case where $\gamma_k(\nu) \gg T_k(\nu)$, we have $\xi_k(\nu) \ll 1$ and $W_k(\nu)G_k(\nu)$ tends more quickly to 0 than $W_k(\nu)$. We can say that the proposed method accentuates the denoising when noise is perceptually significant.

The last perceptual filter considered in this section is that proposed in [13]. Comparing to the block diagram of figure 1, it obeys the following equation

$$\begin{cases} F_k(\nu) = 1, \\ G_k(\nu) = \min\left(\sqrt{T_k(\nu)/\gamma_k(\nu)}, 1\right) \end{cases} \tag{8}$$

This filter is designed so as to yield inaudible residual noise by forcing the residual noise spectral power to be below the masking threshold.

Summarizing, the common feature of the three perceptual filters defined by (5), (6) and (8) is to not process the noisy speech signal when noise is perceptually insignificant. In contrast, our approach (see Eq. 7) involves activating the Wiener filtering even when noise is not audible. By so proceeding, it is expected to reduce the amount of backround noise that could result in audible musical noise after the filtering of adjacent maskers.

4 Experimental Results

We have compared the five methods presented in the foregoing, namely the adaptation to the frequency domain of the "Lin" filter introduced in [12] (see Eq. (5)), the "Tee Won Lee" filter proposed in [13] and summarized by Eq. (8), the "Beaugeant" filter propounded in [11] and specified by Eq. (6), the standard Wiener filter (see Eq. (2)) employing the decision-directed approach of Eq. (3) and, finally, the "double filtering" of Eq. (7). This comparison has been performed on speech signals from the TIMIT database downsampled to 8 KHz before adding white Gaussian noise or babble noise from the NOISEX database at specific SNR's.

The experimental results of this section have been obtained through the following protocol. Short-time windows (32ms) of noisy speech, with 50% overlap, are transformed into the frequency domain using the short-time Fast Fourier transform. As mentioned above, the auditory masking threshold is computed on the basis of the Wiener estimate. The different calculation steps of the masking threshold are those described in [16]. The enhanced speech signal $\widehat{S_k}(t)$ in the time domain is obtained using the overlap-and-add approach after transformation back into the time domain via the Short-Time Inverse Fast Fourier Transform. If we consider the spectrograms of the enhanced speech signals returned by the several tested methods, the "Lin" filter and the "double filter" yield the best results. In fact, the spectrograms provided by these two methods slightly differ from each other (see figure 4).

The five methods addressed in this paper have also been assessed by means of objective measures, namely the standard Segmental Signal to Noise Ratio (SSNR) and the Modified Bark Spectral Distortion (MBSD). The SSNR is the average of the SNR values on short segments. The MBSD proves to be highly correlated with subjective speech quality assessment [18]. Figure 2 (resp. figure 3) presents the average MBSD and the average SSNR for 5 TIMIT sentences corrupted by additive white gaussian noise (resp. babble noise) with SNR from -5dB to 20dB .

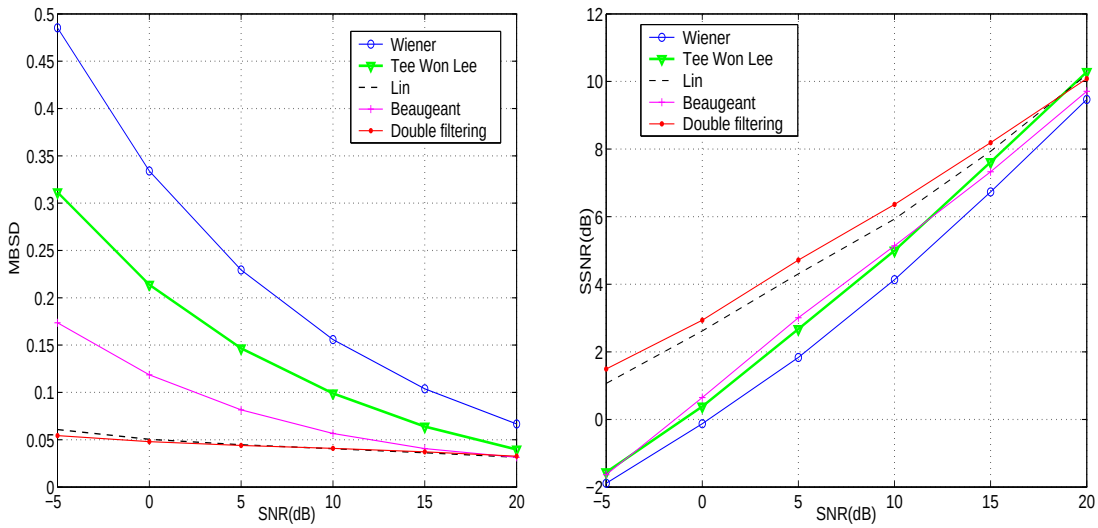

Figure 2: Mean MBSD and segmental SNR achieved by the several speech enhancement approaches studied in this paper for speech signals originally corrupted by white Gaussian noise

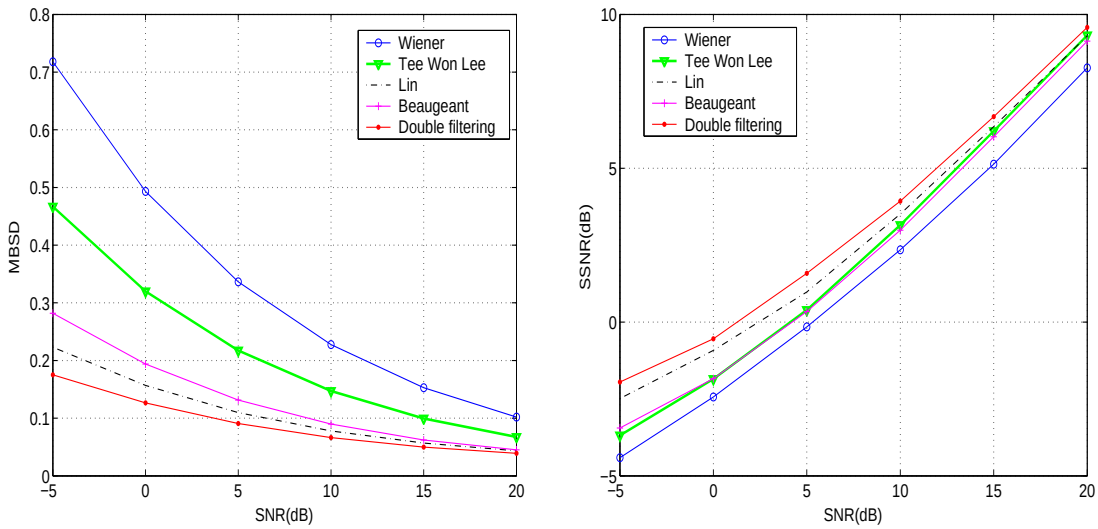

Figure 3: Mean MBSD and segmental SNR achieved by the several speech enhancement approaches studied in this paper for speech signals originally corrupted by babble noise

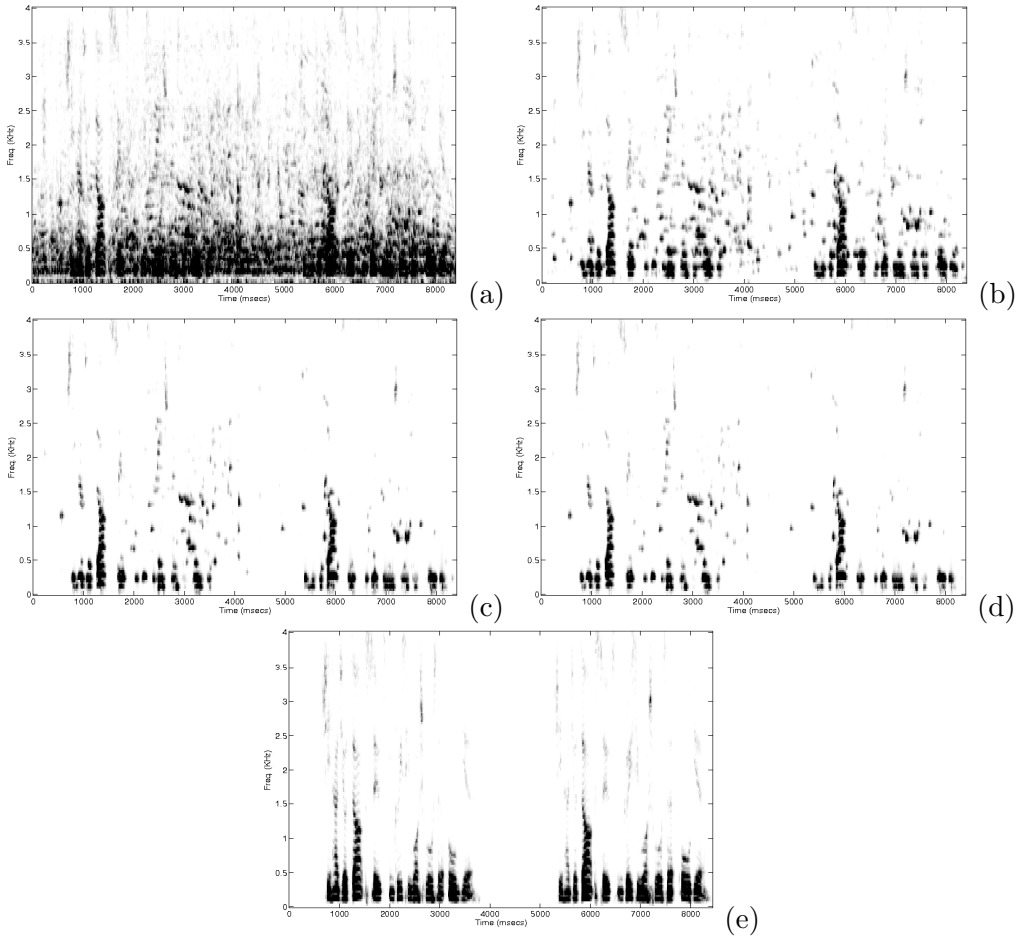

Figure 4: Spectrograms of: (a) Noisy speech signal at 0dB with babble noise, (b) Speech signal enhanced by Wiener filtering, (c) Speech signal enhanced by "Lin" filtering, (d) Speech signal enhanced by "double filtering" and (e) Clean speech. With the "Lin" filter and the "double filtering" approach, musical noise, which appears as isolated points randomly ditributed in time and frequence, is practically non-existent in (c) and (d)

.

5 Conclusion

In this paper, an effective approach for suppressing musical noise present after Wiener filtering has been introduced. Based on the perceptual properties of the human auditory system, a weighting factor accentuates the denoising process when noise is perceptually significant and prevents that residual noise components might become audible in the absence of adjacent maskers. When the speech signal is additively corrupted by white Gaussian noise or babble

noise, experimental results show the improvement brought by the proposed method in comparison to other filtering techniques of the same type.

In forthcoming work, our intention is to compare the several speech enhancement approaches proposed above by means of subjectives tests. We also plan to analyse the behaviour of these methods when the VAD is not ideal or when the noise estimate is computed via algorithms such as those proposed in [14] and [15]. Finally, we envisage using the synoptic of figure 1 with other denoising and perceptual filters, not necessarily to reduce musical noise but to improve the perceptual quality of the enhanced speech.

References

[1] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction", *IEEE Trans. Acoust., Speech and Signal Processing*, vol. 27, pp.113-120, 1979.

[2] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean square error short-time spectral amplitude estimator", *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, pp. 1109-1121, 1984.

[3] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise", *Proc. of ICASSP 1979*, Washington DC, 1979, pp. 208-211.

[4] N. Virag, "Single channel speech enhancement based on masking properties of the human auditory system", *IEEE Trans. Speech and Audio Processing*, vol. 7, pp. 126-137, 1999.

[5] Y. Ephraim and H.L. Van Trees, "A signal subspace approach for speech enhancement", *IEEE Trans. Speech, Audio Processing*, vol. 3, pp. 251-266, 1995.

[6] Y. Hu and P. Loizou, "Incorporating a psychoacoustic model in frequency domain speech enhancement", *IEEE Signal Processing Letters*, 11(2), pp. 270-273, 2004.

[7] A. Ben Aicha and S. Ben Jebara, "Utilisation de la courbe de masquage pour la détection des tonales musicales artificielles dans un signal de parole débruité par approche spectrales ", ISIVC'06, September 13-15, Hammamet, Tunisia, 2006.

[8] F. Jabloun and B. Champagne, "Incorporating the human hearing properties in the signal subspace approach for speech enhancement", *IEEE Trans. Speech, Audio Processing*, vol. 11, 2003, pp. 700-708.

[9] C. You, S. Koh, and S. Rahardja, "An invertible frequency eigendomain transformation for masking-based subspace speech enhancement", *IEEE Signal Processing Letters*, vol. 12, 2005.

[10] J. Kim, S. Kim and C. Yoo, "The incorporation of masking threshold to subspace speech enhancement", *Proc. ICASSP 2003*, vol. 1, pages 76-79, 2003.

[11] C. Beaugeant, V. Turbin, P. Scalart and A. Gilloire, "New optimal filtering approaches for hands-free telecommunication terminals", *Signal Processing* , Volume 64, Number 1, pp. 33-47(15), January 1998.

[12] L. Lin, W. H. Holmes and E. Ambikairajah, "Speech denoising using perceptual modification of Wiener filtering", *IEE Electronic Letters*, vol. 38, no. 23, pp. 1486-1487, November 2002.

[13] T. Lee and Kaisheng Yao, "Speech enhancement by perceptual filter with sequential noise parameter estimation", *Proc. of ICASSP 2004*, Montreal , Quebec, Canada, pp. 693-696, 2004.

[14] A. Amehraye, D. Pastor, S. Ben Jebara, "On the application of recent results in statistical decision and estimation theory to perceptual filtering of noisy speech signals", *ISCCSP 2006*, Marrakech, 2006.

[15] D. Pastor, A. Amehraye, "From non parametric statistics to speech denoising", 3rd international symposium on Image/Video Communications over fixed and mobile networks, ISIVC'06, September 13-15, Hammamet, Tunisia, 2006.

[16] J. D. Johnston, "Transform coding of audio signals using perceptual noise criteria", *IEEE Jour. Selected Areas Commun.*, vol. 6, no. 2, pp. 314-323, 1988.

[17] H. Lev-Ari and Y. Ephraim, "Extension of the signal subspace speech enhancement approach to colored noise", *IEEE Signal Processing Letters* , vol. 10, 2003, pp. 104-106.

[18] W. Yang, M. Dixon and R Yantorno, "Modified bark spectral distortion measure which uses noise masking threshold", *IEEE Speech coding Workshop*, Pocono Manor, pp. 55-56, 1997.