World Academy of Science, Engineering and Technology
International Journal of Electronics and Communication Engineering
Vol:3, No:11, 2009

From Maskee to Audible Noise in Perceptual Speech Enhancement

Asmaa Amehraye, Dominique Pastor, Ahmed Tamtaoui, and Driss Aboutajdine

Abstract—A new analysis of perceptual speech enhancement is presented. It focuses on the fact that if only noise above the masking threshold is filtered, then noise below the masking threshold, but above the absolute threshold of hearing, can become audible after the masker filtering. This particular drawback of some perceptual filters, hereafter called the maskee-to-audible-noise (MAN) phenomenon, favours the emergence of isolated tonals that increase musical noise. Two filtering techniques that avoid or correct the MAN phenomenon are proposed to effectively suppress background noise without introducing much distortion. Experimental results, including objective and subjective measurements, show that these techniques improve the enhanced speech quality and the gain they bring emphasizes the importance of the MAN phenomenon.

Keywords—Perceptual speech filtering, maskee to audible noise, distorsion, musical noise.

I. Introduction

MANY perceptual speech enhancement algorithms have been proposed in the last few decades. They are based on psychoacoustic models to take advantage of the masking phenomenon inherent to the human auditory system. For instance, in [1], the 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 [2] incorporates the masking properties of the human auditory system. In general, the objective of perceptual speech enhancement is to improve the perceptual quality of the enhanced speech signal. Since human ears cannot perceive noise with level below the noise masking threshold, perceptual methods basically aim at reducing audible noise only. By so proceeding, these methods reduce speech distortion.

Although perceptual methods perform well in comparison with classical subtractive type algorithms, most of them still return some audible and annoying musical residual noise. The reasons are manifold. To begin with, biases are introduced by estimating the noise spectrum and the masking threshold. However, some experiments show that even when the noise spectrum and the masking threshold are known, musical noise is still present after denoising. The cornerstone of the letter is the following claim: the attenuation of speech components after perceptual speech enhancement lowers the masking threshold level and, therefore, may reveal noise components initially masked and not processed. This is hereafter called the maskee-to-audible-noise (MAN) phenomenon.

The first contribution of this letter is to highlight the relevance of the MAN phenomenon. The second contribution is the presentation of an elementary approach that takes into account this phenomenon to perform perceptual speech enhancement. This elementary approach involves weighting a standard perceptual filter.

Two weighting functions are considered for application to the same perceptual filter, chosen for its efficiency to process audible noise [3], [5]. This leads to two weighted perceptual filters (WPF). The first one, WPF1, has been introduced in [5]. It is recalled here for comparison to the second one, WPF2, proposed below. In contrast to the former, the latter corrects the MAN phenomenon only at the specific frequencies that are candidates to the MAN phenomenon. By so proceeding, we aim to avoid introducing undesirable distortion.

It is worth noticing that WPF1 and WPF2 basically derive from the standard Wiener filter for two reasons. First, the Wiener filter is easy to implement and second, it can reasonably be expected that if we succeed in reducing the perception of residual noise resulting from some Wiener-like filtering, the quality of the denoised speech will be improved and yield a fairly satisfactory comfort of listening.

The organisation of the letter is as follows. Section II discusses the MAN phenomenon in perceptual speech enhancement. Section III presents the perceptual filter to which the weighting functions are applied and completes the presentation of WPF1 by the results of subjective tests. These experimental results motivate the design of WPF2, introduced in section IV. The two filters WPF1 and WPF2 are then compared by means of subjective and objective tests. Section V concludes this letter.

II. The MAN phenomenon in perceptual speech enhancement

A. Frequency masking

The masking phenomenon derives from the limited frequency selectivity of the human auditory system. In this letter, we consider only the so-called frequency masking. This masking occurs when some powerful signal distorts the absolute threshold of hearing and, thus, makes inaudible weak signals that would be perceptible otherwise. How effective the masker is at increasing the masking threshold of hearing

A. Amehraye is with Lab-STICC (CNRS FRE 3167), Institut Telecom, Telecom Bretagne, Technopole Brest Iroise, 29238 Brest, France and with Laboratoire GSCM-LRIT, Faculté des Sciences, Université Mohammed V, B.P. 1014, Rabat-Agdal, Morocco. asmaa.amehraye@telecom-bretagne.eu.

D. Pastor is with Lab-STICC (CNRS FRE 3167), Institut Telecom, Telecom Bretagne, Technopole Brest Iroise, 29238 Brest, France. dominique.pastor@telecom-bretagne.eu.

A. Tamtaoui is with Institut National des Postes et Télécommunication, Madinat Al Irfane, Rabat, Morocco. tamtaoui@inpt.ac.ma.

D. Aboutajdine Laboratoire GSCM-LRIT, Faculté des Sciences, Université Mohammed V, B.P. 1014, Rabat-Agdal, Morocco. aboutaj@fsr.ac.ma.

depends on the frequency of the maskee and on the frequency of the masker. The maximum masking effect occurs when the masker and the maskee are at the same frequency; the masking effect diminishes when the frequency of the maskee moves away from that of the masker.

B. MAN phenomenon

The main idea of perceptual speech enhancement is to incorporate the masking properties of the human auditory system to reduce audible noise only and, thus, avoid much distortion. Noise components that are not audible because of some maskers in the original noisy signal are still present after denoising. They can become audible if they are initially above the absolute threshold of hearing and their maskers are filtered. This is what we call the MAN phenomenon. It can affect the performance of perceptual filtering that processes audible noise only.

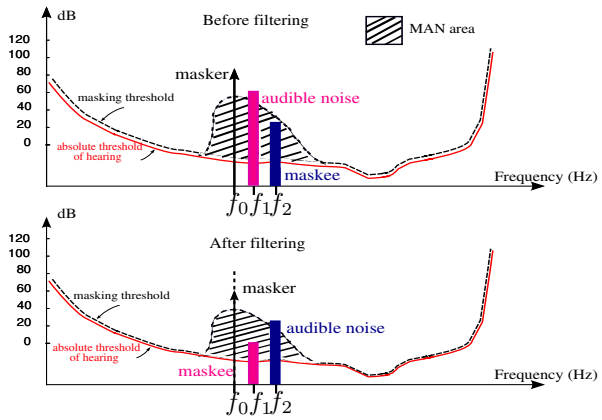

Fig. 1. Description of the MAN phenomenon

Figure 1 is an illustration of the MAN phenomenon located in the dashed area. At frequency f_0 a masker is present. The resulting masking threshold (dashed black curve) is below a first noise component at frequency f_1 and above another noise component, the maskee, located at frequency f_2. We do not perceive the maskee at frequency f_2. After filtering the additive noise around the audible frequency f_1, the masker signal is attenuated, which normally lowers the masking threshold, so that noise at frequency f_2 becomes audible. This phenomenon may occur at each frequency where the energy of the noise maskee lies between the initial masking threshold and the absolute threshold of hearing ATH. These conditions can be easily satisfied physically and carrying out an experiment to illustrate the matter is simple. Consequently, the MAN phenomenon favours the emergence of isolated audible tones that contribute to musical noise. In fact, musical noise consists of rapidly changing random tones that are noticeable in the background of speech. They sound metallic and the denoised speech might be even more unpleasant than the original noisy signal.

III. Perceptual weighting by Wiener filtering

A. Principle

We begin by presenting the perceptual filter on which WPF1 and WPF2 are based. Then, we recall the expression of WPF1. The notations introduced henceforth are kept throughout the rest of the paper with always the same meaning.

Let us assume that each frame of noisy signal contains the same number M of samples. Given frame k, $Y_k(\nu) = S_k(\nu) + N_k(\nu)$ denotes the Discrete Fourier Transform (DFT) Y_k of the noisy speech signal at frequency $\nu = 0, 1, \ldots, M-1$ where S_k (resp. N_k) stands for the DFT of the speech signal (resp. additive and independent noise). The weighting functions proposed in this letter are applied to

$$G_k(\nu) = \frac{|\widehat{S}_k(\nu)|^2}{|\widehat{S}_k(\nu)|^2 + \max\left(\gamma_k(\nu) - T_k(\nu), 0\right)}. \qquad (1)$$

where $\widehat{S}_k(\nu) = Y_k(\nu)W_k(\nu) = Y_k(\nu)(\xi_k(\nu)/(1+\xi_k(\nu)))$, W_k is the Wiener filter, $\xi_k(\nu)$ is the so-called decision-directed estimate of the *a priori* SNR [4], $T_k(\nu)$ is the masking threshold and $\gamma_k(\nu)$ is the noise power spectrum estimate. Filter G_k performs a Wiener filtering of only the amount of noise that exceeds the masking threshold. In [3], this method applies to the sub-band components returned by an auditory filter bank. Here, the filter G_k results from the adaptation of this method to the usual case where the time-frequency analysis is performed by the standard DFT. The choice of this filter is motivated by the objective test results given in [5]. According to these results, G_k outperforms the perceptual filters proposed in [6], [2] and the standard Wiener filter. Filter G_k, as well as those in [1], [6], [2], are typical examples of perceptual filters which do not take into account the MAN phenomenon. In fact, they process audible noise only.

The weighting function considered in this section is the standard Wiener filter W_k. The resulting WPF, namely WPF1, is thus specified by

$$G_k^{\mathrm{WPF1}}(\nu) = W_k(\nu)G_k(\nu). \qquad (2)$$

The resulting filtering accentuates the denoising for the frequencies ν where noise is perceptually annoying, that is, when the noise power spectral density or spectrum $\gamma_k(\nu)$ is above the masking threshold. By attenuating every frequency with W_k, we avoid that the MAN phenomenon occurs.

In the next section, we experimentally assess the performance of WPF1 in comparison to the standard Wiener filter W_k, G_k (see Eq. (1)) and

$$H_k(\nu) = \begin{cases} W_k(\nu) & \text{if } \gamma_k(\nu) > T_k(\nu) \\ 1 & \text{otherwise,} \end{cases} \qquad (3)$$

which is the perceptually motivated filter of [6]. This filter performs the Wiener denoising only for audible noise frequency components. A smoothing correlogram is applied to each perceptual filter so as to avoid discontinuities of the filter frequency response.

TABLE I
THE MEAN SCORES ON THE SIG, BACK AND OVRL SCALES FOR 4 METHODS EVALUATED IN CAR AND BABBLE NOISE AT SNR LEVELS OF 5DB AND 10DB (ENGLISH NATIVE LISTENERS)

Car noise		Noisy	$\mathbf{W_k}$	$\mathbf{G_k}$	$\mathbf{H_k}$	WPF1
5 dB	SIG	4.53	4.5	4.53	4.39	4.61
	BACK	2.44	3.96	3.90	3.96	4.54
	OVRL	3.09	4.28	3.88	3.85	4.39
10 dB	SIG	4.66	4.64	4.61	4.64	4.64
	BACK	2.99	4.59	4.33	4.26	4.65
	OVRL	3.46	4.38	4.28	4.18	4.40
Babble noise		**Noisy**	$\mathbf{W_k}$	$\mathbf{G_k}$	$\mathbf{H_k}$	**WPF1**
5 dB	SIG	3.96	3.08	2.31	2.31	3.01
	BACK	1.41	3.33	2.88	2.79	3.88
	OVRL	2.09	2.70	1.99	1.81	2.71
10 dB	SIG	4.19	3.85	2.94	2.89	3.78
	BACK	1.73	3.68	3.20	3.38	4.26
	OVRL	2.35	3.26	2.46	2.44	3.45

TABLE II
THE MEAN SCORES ON THE SIG, BACK AND OVRL SCALES FOR 4 METHODS EVALUATED IN CAR AND BABBLE NOISE AT SNR LEVELS OF 5DB AND 10DB (NON-NATIVE ENGLISH LISTENERS)

Car noise		Noisy	$\mathbf{W_k}$	$\mathbf{G_k}$	$\mathbf{H_k}$	WPF1
5 dB	SIG	4.43	4.42	4.35	4.23	4.46
	BACK	2.52	4.49	3.83	3.84	4.63
	OVRL	3.32	4.29	3.88	3.82	4.41
10 dB	SIG	4.40	4.47	4.30	4.42	4.49
	BACK	2.78	4.50	4.18	4.20	4.54
	OVRL	3.41	4.30	4.08	4.22	4.37
Babble noise		**Noisy**	$\mathbf{W_k}$	$\mathbf{G_k}$	$\mathbf{H_k}$	**WPF1**
5 dB	SIG	4.13	2.98	2.26	2.26	2.94
	BACK	1.87	3.49	3.29	3.23	3.78
	OVRL	2.65	2.86	2.11	2.16	2.79
10 dB	SIG	4.39	3.54	2.91	2.86	3.51
	BACK	2.08	3.83	3.42	3.52	3.98
	OVRL	2.92	3.24	2.63	2.61	3.34

B. Subjective tests

Our purpose is now to assess the enhanced speech quality achieved by WPF1. If the MAN phenomenon is significant, we should notice some improvement by using WPF1 in comparison with filters that do not take into account the MAN phenomenon . Therefore, after the evaluation by means of objective tests presented in [5], we carried out subjective listening tests to compare WPF1, G_k, H_k and the Wiener filter.

Noise signals from the Noisex database (babble and car noise) were added with two SNRs (5 dB, 10 dB) to 10 sentences randomly chosen from the TIdigits database downsampled to 8 KHz. The experimental protocol was the following one. Short-time windows (32 ms) of noisy speech, with 50% overlap, were transformed into the frequency domain using the short-time Fast Fourier Transform (FFT). The auditory masking threshold was estimated by means of the Johnston model [7] applied to the Wiener estimate $\widehat{S}_k$. The noise spectrum was assumed to be known in order to assess the filtering without taking the risk to introduce any bias due to noise spectrum estimation. The enhanced speech signal in the time domain was obtained using the overlap-and-add approach after transformation back into the time domain via the Short-Time Inverse FFT. We used the recently standardised methodology [8] for the subjective evaluation. In short, this methodology requires the listener to rate, using five-point scales, the distortion of speech alone (SIG), the background noise alone (BACK) and the overall quality (OVRL). The sentences were presented to 24 listeners amongst which 8 English natives.

The experimental results are those of tables I and II. The perceptual weighting WPF1 performs generally better than the other filters. An analysis of variance (ANOVA) on the data of tables I and II with level of significance $\alpha = 0.05$ shows the following. WPF1 achieves a statistically significant smaller noise distortion (higher BACK scores) than the other algorithms in every experimental condition (p-values< 0.05 for BACK). This is natural since WPF1 basically accentuates the denoising to reduce the background noise distortion (residual noise) and avoid the MAN phenomenon. As far as the signal quality is concerned, the results are not statistically different in car noise and the different methods are as good as each other (p-values > 0.05 for SIG); in contrast, in babble noise, the listeners statistically prefer the original noisy signal (p-values < 0.05 for SIG). The fact that WPF1 does not perform statistically better than the other filters with respect to signal quality can be explained as follows: the Wiener filtering alters speech quality and the perceptual filtering does not correct this distortion even though it reduces the residual noise intrusiveness. Now, regarding the overall quality, WPF1 performs significantly better in every situation except for babble noise at 5 dB and for non-native listeners.

According to these experimental results, the MAN phenomenon has a significant impact on the speech enhancement performance since such a simple method as WPF1 makes it possible to avoid it and to generally yield better performance than standard perceptual filters. In the next section, another perceptual weighting is proposed to overcome the limitation of WPF1 pointed out by the subjective ratings above.

IV. ATTENUATION AT FREQUENCY CANDIDATES TO THE MAN PHENOMENON

A. Principle

As noticed in the previous section, the main drawback of WPF1 is the following one: the Wiener filtering is applied to each frequency and this entails some signal distortion that cannot be compensated by the perceptual filtering. Instead of trying to avoid the MAN phenomenon, the new perceptual weighting WPF2 presented now intends to correct it. More specifically, the perceptual filtering of the noisy signal is now followed by a modified Wiener filter. This filter is a post-processing of the perceptual denoised signal only at frequencies ν where the value of the noise spectrum $\gamma_k(\nu)$ lies between the absolute threshold of hearing $\mathrm{ATH}_k(\nu)$and the masking threshold $T_k(\nu)$. Because noise at these frequencies is candidate to the MAN phenomenon. The expression of the resulting filter $G_k^{\mathrm{WPF2}}(\nu)$ is then

$$G_k^{\mathrm{WPF2}}(\nu) = G_k(\nu)W_k'(\nu) \quad (4)$$

TABLE III
COMPARATIVE PERFORMANCE MEASUREMENTS BETWEEN WPF1 AND WPF2, IN TERMS OF MEAN MBSD, PESQ AND SEGMENTAL SNR OVER 250 SENTENCES CORRUPTED BY BABBLE AND CAR NOISE AT DIFFERENT SNR LEVELS.

Babble noise		-5dB	0dB	5dB	10dB	15dB
MBSD	WPF1	0.054	0.045	0.037	0.031	0.027
	WPF2	0.038	0.033	0.030	0.027	0.024
PESQ	WPF1	1.75	2.18	2.63	3.00	3.30
	WPF2	2.10	2.46	2.80	3.12	3.41
SSNR	WPF1	1.15	2.44	4.29	6.45	8.80
	WPF2	1.93	3.54	5.50	7.60	9.89
Car noise		**-5dB**	**0dB**	**5dB**	**10dB**	**15dB**
MBSD	WPF1	0.025	0.022	0.020	0.016	0.013
	WPF2	0.023	0.020	0.017	0.013	0
PESQ	WPF1	3.42	3.65	3.86	4.05	4.21
	WPF2	3.45	3.69	3.91	4.10	4.26
SSNR	WPF1	6.50	8.47	10.41	12.25	14.13
	WPF2	7.96	10.13	12.29	14.5	16.61

where G_k is given by Eq. (1) and

$$W'_k(\nu) = \begin{cases} W_k(\nu) & , \text{ if } \mathrm{ATH}_k(\nu) < \gamma_k(\nu) \leq T_k(\nu) \\ 1 & , \text{ otherwise.} \end{cases} \quad (5)$$

When $G_k(\nu)$ is inactive ($G_k(\nu) = 1$), the weighting factor $W'_k(\nu) = W_k(\nu)$ attenuates noise within the MAN area ($\mathrm{ATH}_k(\nu) < \gamma_k(\nu) \leq T_k(\nu)$).

B. Objective and subjective tests

This section aims to compare WPF1 and WPF2 on the basis of objective and subjective tests. We assessed WPF2 with respect to WPF1 only because the latter performs better than the other methods tested above. Also, this allowed us to reduce the test burden for listeners. Only non-native English listeners were involved in these tests. Indeed, according to the results of table I and II, there was no significant difference between scores of English natives and scores of non-native English listeners. This is due to the fact that TIdigits database involves English digits only.

The objective criteria were the standard Segmental Signal to Noise Ratio (SSNR), the Modified Bark Spectral Distortion (MBSD) and the Perceptual Evaluation of Speech Quality (PESQ). The experiments were carried out by using 250 sentences, randomly chosen from the TIDIGITS database. These sentences were corrupted by additive babble and car noise with SNR ranging from -5dB to 20dB. According to table III, WPF2 achieves significant performance improvement in comparison with WPF1, whatever the criterion. On the other hand, the subjective tests carried out according to the same protocol as in section III-B show that generally the listeners better rate WPF1 than WPF2 (see table IV). However, it follows from a t-test analysis that the performance measurements in table IV are not statistically significant (p-values > 0.05). These results confirm the impact of the MAN phenomenon and the relevance of correcting it.

V. CONCLUSION

The goal of this letter was to describe and emphasise the importance of the MAN phenomenon in perceptual speech enhancement. In fact, the weighted perceptual filters proposed in this paper and which take into account this phenomenon perform better than standard perceptual ones that process audible noise only. The approach proposed in this letter can be generalised as follows: first, by extending it to other perceptual filters; second, by looking for more sophisticated weighting factors.

TABLE IV
MEAN SCORES FOR THE SIG, BACK, AND OVRL SCALES FOR WPF1 AND WPF2 EVALUATED IN CAR AND BABBLE NOISE AT SNR LEVELS OF 5DB AND 10DB WITH NON-NATIVE ENGLISH LISTENERS

Car noise		WPF1	WPF2
5dB	SIG	4.35	4.24
	BACK	4.08	3.86
	OVRL	4.01	3.89
10dB	SIG	4.37	4.37
	BACK	4.21	4.03
	OVRL	4.12	4.02
Babble noise		**WPF1**	**WPF2**
5dB	SIG	2.81	2.654
	BACK	3.22	2.554
	OVRL	2.56	2.27
10dB	SIG	3.46	3.49
	BACK	3.51	3.03
	OVRL	3.23	2.99

REFERENCES

[1] 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.
[2] Y. Hu and P. Loizou, "Incorporating a psychoacoustic model in frequency domain speech enhancement," *IEEE Signal Processing Letters*, vol. 11(2), pp. 270–273, Feb 2004.
[3] L. Lin, W. H. Holmes, and E. Ambikairajah, "Speech denoising using perceptual modification of wiener filtering," *IEE Electronic Letters*, vol. 38, pp. 1486–1487, Nov 2002.
[4] 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, Dec 1984.
[5] A. Amehraye, D. Pastor, and A. Tamtaoui, "Perceptual improvement of wiener filtering." Proc. of ICASSP, 2008, pp. 2081–2084.
[6] C. Beaugeant, V. Turbin, P. Scalart, and A. Gilloire, "New optimal filtering approaches for hands-free telecommunication terminals," *Signal Processing*, vol. 64, pp. 33–47(15), Jan 1998.
[7] J. D. Johnston, "Transform coding of audio signals using perceptual noise criteria," *IEEE Jour. Selected Areas Commun*, vol. 6, pp. 314–323, 1988.
[8] ITU-T(2003), "Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm," *ITU-T Recommendation P.835*, 2003.