International Conference on Control, Engineering & Information Technology (CEIT'14)

Proceedings - Copyright IPCO-2014

ISSN 2356-5608

Perceptual Speech Enhancement Using Multi_band Spectral Attenuation Filter

Sana Alaya, Novlène Zoghlami and Zied Lachiri
Signal, Image and Information Technology Laboratory
National Engineering School of Tunis
Tunis, Tunisia

Zied Lachiri
Instrumentation and Measures Department
National Institute of Applied Science and Technology
Tunis, Tunisia

***Abstract*—This paper addresses the enhancement of the speech signal using perceptual properties. We propose to implement classic spectral attenuation on a gammachirp perceptual filterbank with nonlinear frequency distributions in ERB scale in association with a Johnston model which will provide us a masking frequency threshold used to improve perceptual appearance of speech signal. Objective and subjective assessment tests are applied to prove the performance of our method especially on perceptual appearance.**

***Keywords*—Gammachirp filterbank, nonlinear frequency distribution, spectral attenuation, Johnston model, perceptual properties**

I. INTRODUCTION

The speech signals are transmitted and processed through channels that can affect it. Anything that modifies the clean signal can be assumed to be noise. Noise reduction in continuous speech [1] represents one of the major problems in the field of signal processing. The essential objective is to improve the signal intelligibility and its perceptual appearance. Several techniques have been developed to improve the voice quality as described in [1][2][3][4][5][6][7][8]. Spectral attenuation technique [9][10][11] have the same goal especially finding a compromise between minimizing distortion introduced into the signal and maximizing noise reduction through the spectral estimation signal. The major problem of these techniques is the occurrence of audible artifacts introduced during the denoising process thereby deteriorating the quality and intelligibility of the signal found. The goal of this paper is to create a new method based on the classical spectral attenuation associated with a gammachirp filterbank, which mimics the functioning of the human ear, following the ERB scale with a thresholding filterbank following the Johnston model [12].

This paper is organized as follows. Section 2 describes the developed perceptual speech enhancement method. Section 3 outlines the objective evaluation results as well as the subjective results.

II. THE PROPOSED SPEECH ENHANCEMENT METHOD

Spectral attenuation methods are used for noise reduction in continuous speech. Minimum Mean Square Error-Short-Term Spectral Amplitude (MMSE) [11] is one of the spectral attenuation techniques that aim to estimate the level of noise present in a signal by performing a uniform spectral decomposition of the noisy signal by windowing followed by a Fourier transform. A uniform linear frequency analysis is then applied to the signal. However, the frequency analysis should be performed with a finer scale to obtain a more robust denoising method. By deepening research on the laws of psychoacoustics, it was found that the human ear could perceive signals in a noisy environment by performing an accurate frequency analysis. This precision is the result of the non-uniform frequency decomposition of the signal. This leads us to realize a denoising method based on a non-uniform frequency-analysis filterbank. Then it is interesting to analyze a signal using a non-uniform filterbank in combination with a spectral attenuation.

In the present work we will use a non-uniform decomposition of the noisy speech signal $y(t) = s(t) + n(t)$ with $s(t)$ is the clean signal, $n(t)$ is the additive noise and $t = 0,1, \dots, M-1$ is the time index. For this reason, it was chosen to use the gammachirp filterbank which imitates the functioning of the human ear [13]. The impulse response of this filter is defined as follows:

$$g_c(t) = At^{n-1}\exp(-2\pi b\,\mathrm{ERB}(f)t)\cos(2\pi ft + c\,\ln t + \varphi) \quad (1)$$

ISSN 2356-5608

Where $(t > 0)$, A is the amplitude, n and b are parameters defining distribution, f_c is the asymptotic frequency modulation, and φ is the initial phase; $\ln t$ is a natural logarithm of time, ERB(f) is the equivalent rectangular bandwidth of the filter at f [14], and at moderate levels, $\mathrm{ERB}(f) = 21.4\ln(0.00437f + 1)$ in Hz, when c=0, this equation represents a complex impulse response of the gammatone[15]. It has been demonstrated [13] that the gammachirp filter fits human psychoacoustic masking data [16] when the parameter c is associated with the sound pressure level typically as $c = 3.38 - 0.107Ps$ where Ps is the threshold level of a probe sinusoid in noticed noise [13]. In our method the gammatone filter decomposes the noisy signal into sub-bands as $y_i(t) = y(t) * g_i(t)$ where $g_i(t)$ is the impulse response of the i^{th} band.

To prevent filtering the noises which are initially inaudible and may become audible if masks are filtered, the Johnston masking threshold $T_{i,k}(f)$ is used in the output of the spectral attenuation block in each sub-band. Figure 1 illustrates the principle of the perceptual denoising method.

The perceptual filter is defined by the following equation:

$$\begin{cases} SS_{i,k}(f): \text{Spectral Attenuation Filter} \\ G_{i,k}(f) = \dfrac{\left|\tilde{S}_{i,k}(f)\right|^2}{\left(\left|\tilde{S}_{i,k}(f)\right|^2 + \max\left((\gamma_{i,k}(f) - T_{i,k}(f)\right),0\right)}\,[6] \end{cases} \quad (2)$$

$$\text{Where} SS_{i,k}(f) = \left|\frac{\sqrt{\pi}}{2}\right| \sqrt{\frac{1}{1+R_{post_{i,k}}}} \left(\frac{R_{priori_{i,k}}}{1+R_{priori_{i,k}}}\right) M\left(\left(1 +\right.\right.$$

Rposti,kRpriorii,k1+Rpriorii,k represent the gain of the spectral attenuation filter in the sense of Ephraim and Malah [11] where $M(\theta)$ is given by:

$$M(\theta) = e^{-\frac{\theta}{2}}\left[(1+\theta)I_0\frac{\theta}{2} + \theta.I_1\left(\frac{\theta}{2}\right)\right] \quad (3)$$

I_0 and I_1 are respectively the Bessel modified function of order 0 and 1. The posteriori level defines the measured value to the m^{th} window defined by:

$$R_{post_{i,m}}(f) = \frac{|y_{i,m}(f)|^2}{\hat{B}_{i,m}(f)}. \quad (4)$$

Priori signal to noise ratio is defined by:

$$R_{priori_{i,m}}(f) = (1-\alpha)\left[R_{post_{i,m}}(f) - 1\right] + \alpha\frac{\left|y_{i,(m-1)}(f)\right|^2}{\hat{B}_{i,m}(f)} \quad (5)$$

With α parameter between 0 and 1, $\left|y_{i,m-1}(f)\right|^2$ is the power spectrum of the previous window and $R_{post_{i,m}}(f)$ is the post local relative level. The next step consists in applying the perceptual filter $G_{i,k}(f)$ to the output of spectral attenuation filter. $T_{i,k}(f)$represents the auditory masking threshold and $\gamma_{i,k}(f)$ represents the power noise estimation $E\left[\left|(N_{i,k}(f)\right|^2\right]$. The gain perceptual filter $G_{i,k}(f)$ depends on the value of the Johnston threshold$T_{i,k}(f)$. When the Johnston threshold is greater than the noise value $\gamma_{i,k}(f) < T_{i,k}(f)$ it means that the noise is inaudible then it is unnecessary to go through the second stage of filtering because we risk to filter noise mask and noise which initially inaudible risk of becoming. In this situation $G_{i,k}(f) \approx 1$. Otherwise we found $\gamma_{i,k}(f) > T_{i,k}(f)$, noise is then audible by the human ear. Therefore it is necessary to go through the second filter at the output of the spectral attenuation step. The principal role of the second filter is to improve the perceptual appearance of the enhanced signal. We find the signal $\hat{y}_i(t)$ using the inverse Fourier transform. Synthesis step allows us to find the enhanced signal $\hat{y}$ (t) by summing all treated sub-band $\hat{y}_i(t)$.

ISSN 2356-5608

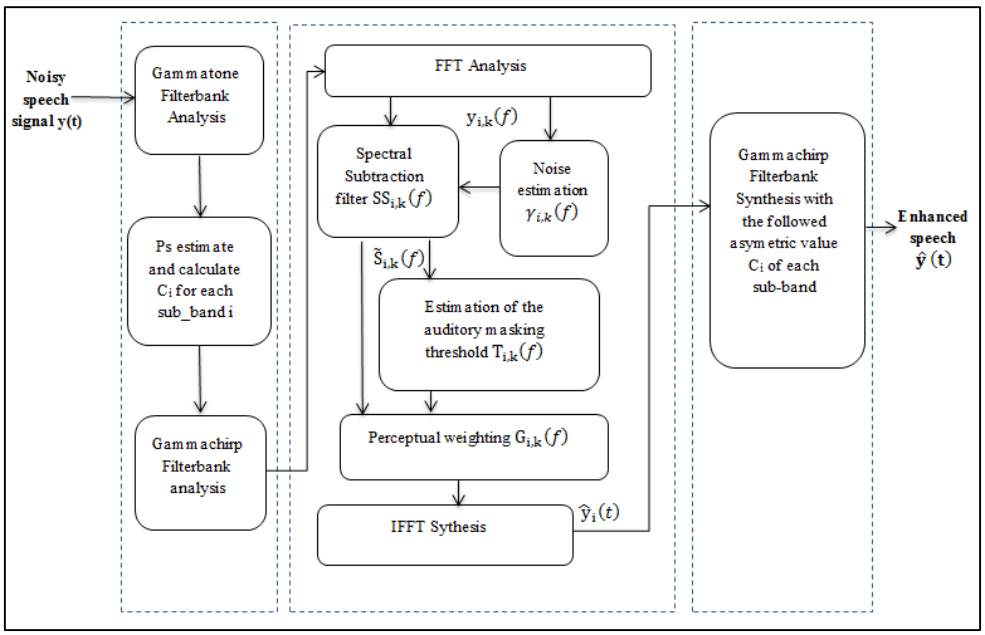

Figure 1: Schematic of the proposed perceptual method

III. EXPERIMENTAL RESULTS

The method was assessed using TIMIT database sampled at 16khz and corrupted by two types of noise: the car noise and the babble noise at different SNR level: 0dB, 5dB, 10dB and 15dB. We use the Hamming window of 256 samples with an overlap of 50%. The decomposition is performed with the gammachirp filter using 32 bands according to the ERB scale with variable value of asymmetry parameter C. In order to assess the method we have chosen PESQ measures [17] as an objective method. We have strengthened these measures by subjective listening tests. A total of 5 listeners participated in the listening tests. Listeners are invited to give three ratings (SIG, BAK, OVRL) for each processed signal [18]. Standing successively for the speech quality, the level of degradation of the background noise and the overall quality.

Table 1 list the mean results of the PESQ, SIG, BAK and OVRL score from the proposed perceptual spectral attenuation (PSA) in comparison with the Ephraim and Malah method (MMSE).

From the PESQ results of the proposed method, we can see that it improves significantly the speech quality at different SNR levels with different noises. At 5 dB we have 2.71 for the car noise by using the proposed PSA method against 2.55 with the MMSE method. For the babble noise we have 3.28 at 15 dB for the proposed PSA method against 3.19 for the MMSE compared method.

From the SIG value, which informs us about the signal distortion level, we can see that the proposed method provides less distortion to the enhanced signal compared with the MMSE method. We obtain a value of 3.48 with car noise and 3.29 with babble noise at 0 dB for the proposed PSA method against 3.32 with car noise and 3.15 with babble noise for the MMSE compared method. This means that the enhanced speech with our method (PSA) did not contain notable speech distortion compared with the Ephraim and Malah spectral attenuation method (MMSE).

From the background intrusiveness (BAK) results, we note that the proposed method brings about less distortion for the enhanced signal with the different noise types and SNR level. In fact at 15 dB we obtain for the car noise the value of 3.59 for the proposed PSA method against 3.21 with the MMSE method. For the babble noise at 5dB we obtain 1.79 for the proposed method against 1.36 with compared method.

From the OVRL value we note a significant improvement in overall quality by comparing it with the other method. For the babble noise at 10dB we obtain 3.59 for the proposed method against 3.31 with the compared method. For the car noise at 0dB we obtain 3.20 for the proposed method against 3.09 with the compared method.

TABLE 1. THE MEAN SCORE OF PESQ, SIG, BAK AND OVRL FOR PROPOSED PERCEPTUAL SPECTRAL ATTENUATION METHOD (PSA) AND THE EPHRAIM AND MALAH METHOD (MMSE)

		Objective tests		**Subjective tests**					
		PESQ		***SIG***		***BAK***		***OVRL***	
		METHODS		***METHODS***		***METHODS***		***METHODS***	
		MMSE	***PSA***	***MMSE***	***PSA***	***MMSE***	***PSA***	***MMSE***	***PSA***
SNR of Babble Noise	***0 dB***	2.15	2.19	3.15	3.29	1.09	1.23	2.77	3.15
	5dB	2.51	2.56	3.55	3.63	1.36	1.79	2.66	3.22
	10dB	2.86	2.93	4.36	4.47	2.32	2.59	3.31	3.59
	15dB	3.19	3.29	4.64	4.78	3.09	3.21	3.56	4.02
SNR of Car Noise	***0dB***	2.21	2.36	3.32	3.48	1	1.68	3.09	3.20
	5dB	2.55	2.71	3.74	3.81	1.55	2.15	3.18	3.37
	10dB	2.89	3.03	4.59	4.67	2.33	3.05	3.47	3.82
	15dB	3.21	3.35	4.82	4.86	3.21	3.59	4.09	4.35

IV. CONCLUSION

A new method was proposed for noise suppression without creating distortions in the speech signal and suppressing musical noise in order to improve the perceptual appearance of the enhanced signal. The method is based on the incorporation of three different filters. The first is the gammachirp perceptual filter which divides the signal in non-uniform bands imitating the functioning of the human ear. This latter is followed by a second filter based on spectral attenuation in Ephraim and Malah sense using a continuous noise estimate. The third filter is a perceptual filter using Johnston model to further improve perceptual appearance of the enhanced signal. The method was evaluated according to objective and subjective criteria. Depending on the objectives and subjective values found it can be concluded that the association of gammachirp with the Ephraim and Malah filter method and the Johnston perceptual filtering, lead us to say that the proposed enhancement method improves the quality of the signal as well as its perceptual aspect.

REFERENCES

[1] P. Loizou, Speech Enhancement: Theory and Practice, CRC Press, FL: Boca Raton, 2013.

[2] C.H. Taal,R.C.Hendriksand R. Heusdens,"A speech preprocessing strategy for intelligibility improvement in noise based on a perceptual distortion measure". ICASSP, Kyoto, pp. 4061 – 4064, 2012.

[3] C.V.R.Rao,M.B.R.Murthy, and K.S.Rao,"Speech enhancement using perceptual Wiener filter combined with unvoiced speech — A new Scheme". RAICS, Trivandrum, pp. 688 – 691, 2011.

[4] S.G.Sardaroudi and M.Geravanchizadeh, "A perceptual subspace approach for speech enhancement". IST, Tehran, pp. 878 – 881, 2010.

[5] 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.

[6] A.Amehraye,D. Pastor and A.Tamtaoui, "Perceptual improvement of Wiener filtering". ICASSP'08. Las Vegas, USA, pp. 2081-2084, 2008.

[7] N.Zoghlami, Z.Lachiri, and N.Ellouze, "Noise reduction based on perceptual speech analysis". 8th EURONOISE. Edinburgh, Scotland, pp. 26-28, 2009.

[8] C.V.R.Rao, M.B.R.Murthy, and K.S Rao,"Speech Enhancement Using Perceptual Wiener Filter Combined with Unvoiced Speech- A new Scheme". RAICS IEEE. Trivandrum, pp. 688-691, 2011.

[9] S.F.Boll, "Suppression of acoustic noise in speech using spectral subtraction". IEEE Trans. Acoust., Speech, Signal Process. Vol. 27, pp. 113-120, 1979.

[10] M.Berouti, R.Schwartz, and J.Makhoul, "Enhancement of speech corrupted by acoustic noise". in Proc. Int. Conf. on Acoustics, Speech, Signal Processing, pp. 208-211, 1979.

[11] Y.Ephraim and D.Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator". IEEE Trans. Acoust., Speech, Signal Process. Vol. 32(6), pp.1109-1121, 1984.

[12] J. D.Johnston, "Transform coding of audio signals using perceptual noise criteria". IEEE Jour. Selected Areas Commun. Vol. 6 ,pp. 314–323, 1988.

[13] T.Irino and R.D.Patterson, "A Dynamic Compressive Gammachirp Auditory Filterbank". IEEE Trans. Audio, Speech, and Language Process. Vol.14, pp. 2222-2232, 2006.

[14] B. C. J.Moore and B. R.Glasberg, "A revision of Zwicker's loudness model". ActaAcustica. Vol. 82, pp. 335-345,1996.

[15] R. D.Patterson, M.Allerhand, and C.Giguere, "Timedomain modelling of peripheral auditory processing: a modular architecture and a software platform". J. Acoust. Soc. Am., 98, pp. 1890-1894, 1995.

[16] S.Rosen and R.J.Baker, "Characterising auditory filternonlinearity", Hear. Res., 73, pp. 231-243, 1994.

[17] ITU-T recommendation P.862,"Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs", International Telecommunication Union,2000.

[18] ITU-T recommendation P.835,"Subjective test methodology for evaluating speech communication systems that include noise

ISSN 2356-5608

suppression algorithm", International Telecommunication Union, 2003.