



HAL
open science

On the application of recent results in statistical decision and estimation theory to perceptual filtering of noisy speech signals

Asmaa Amehraye, Dominique Pastor, Sofia Ben Jebara

► To cite this version:

Asmaa Amehraye, Dominique Pastor, Sofia Ben Jebara. On the application of recent results in statistical decision and estimation theory to perceptual filtering of noisy speech signals. ISCCSP 2006: 2d international symposium on Control, Communications and Signal Processing, Mar 2006, Marrakech, Morocco. hal-02137078

HAL Id: hal-02137078

<https://hal.science/hal-02137078>

Submitted on 22 May 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.

On the Application of Recent Results in Statistical Decision and Estimation Theory to Perceptual Filtering of Noisy Speech Signals

Asmaa Amehraye, Dominique Pastor and SoËa Ben Jebara

Abstract—This paper combines perceptual filtering to recent results in statistical decision and estimation theory in order to denoise speech signals corrupted by additive and independent white Gaussian noise. The resulting technique requires no Voice Activity Detector and its performance is significantly close to that obtained when the noise standard deviation is known and the masking threshold computed on the basis of the clean speech signals.

I. INTRODUCTION

Speech enhancement has greatly progressed over the past decades. Traditional methods basically improve the Signal to Noise Ratio (SNR). However, they introduce unpleasant residual noise known as musical noise. Various algorithms have been proposed for reducing the effects of musical noise ([1], [2], [3]). They admittedly succeed under some conditions but there is still a need for more performant methods that reduce the amount of residual noise at very low input SNRs without introducing speech distortion.

The current trend is to exploit the auditory masking properties, widely used in perceptual audio coding. By using these properties, it is expected to make residual noise inaudible ([4], [10], [11]). In this respect, the masking threshold becomes a constraint to take into account because the human auditory system is not sensitive to any residual noise masked by coexistent speech signal. The most usual method to calculate the masking threshold is presented in [5] and relies on a critical band analysis modelling the behaviour of the inner ear.

In this paper, to denoise speech signals corrupted by independent and additive white Gaussian noise (AWGN), we combine a perceptually motivated method with a non parametric estimator of the noise standard deviation. This estimator avoids the use of any Voice Activity Detector (VAD). Because of the novelty of this estimator, we do not address the case of coloured noise yet. The estimate performed by this estimator is used twice. First, to adjust a recursive Wiener filtering whose outcome serves to estimate the masking threshold of the speech signals to denoise; second, to tune the perceptual filter proposed by [9]. We evaluate the quality of the filtered speech signals by means

of two objective criteria, namely the standard Segmental Signal to Noise Ratio (SSNR) (see [8]) and the Modified Bark Spectral Distortion (MBSD) (see [13]). The latter is a perceptually motivated metric that employs the concept of masking threshold in order to introduce no bias due to possible inaudible distortions. It proves to be highly correlated with subjective speech quality assessment [13].

Section II presents the perceptual filtering we consider. In section III, we describe the theoretical tools, the noise standard deviation estimator derived from these results and the application of this estimator to speech processing. The performance of the speech enhancement method that combines this estimator with the perceptual filtering of section II is addressed in section IV. Concluding remarks and perspectives are given in section V.

II. PERCEPTUAL SPEECH ENHANCEMENT

Let $s(n)$ be some speech signal corrupted by additive and independent stationary noise $x(n)$. The observed signal is

$$y(n) = s(n) + x(n). \quad (1)$$

Given a frame of N samples $y(n)$, $n = 1, \dots, N$, let Y_k , S_k and X_k , $k = 0, \dots, N - 1$, denote the Discrete Fourier Transform (DFT) coefficients of $y(n)$, $s(n)$ and $x(n)$ respectively. Generally, speech enhancement techniques consist in estimating the frequency components S_k by

$$\hat{S}_k = H_k Y_k, k = 0, \dots, N - 1, \quad (2)$$

where H_k is a linear estimator chosen according to a suitable criterion. The error signal generated by this estimator is

$$e_k = \hat{S}_k - S_k = (H_k - 1)S_k + H_k X_k. \quad (3)$$

The values $(H_k - 1)S_k$, $k = 0, \dots, N - 1$, are the DFT coefficients of the speech distortion due to the filtering and the frequency components $H_k X_k$, $k = 0, \dots, N - 1$, are the DFT coefficients of the residual noise. The quantity $|H_k|^2 E[|X_k|^2]$ is then the residual noise spectral power for the k th frequency component.

In order to take into account the properties of the auditory system, the filter H_k can be designed so as to yield inaudible residual noise. This can be achieved by forcing the residual noise spectral power to be below the masking threshold in each frequency bin; indeed, the human auditory system does not perceive any noise with spectral power less than the masking threshold. This is the approach followed in [9]. The filter is then constrained by the inequality $|H_k| \leq \sqrt{\frac{T_k}{\gamma_k}}$, $k = 0, \dots, N - 1$, where T_k and γ_k are the value of the masking

A. Amehraye is with the GSCM-Facult  des Science de Rabat, Rabat, Morocco and also with Ecole Nationale Sup rieure des T l communications de Bretagne asmaa.amehraye@enst-bretagne.fr

D. Pastor is with Ecole Nationale Sup rieure des T l communications de Bretagne, Technop le de Brest-Iroise, CS 83818, 28238 Brest, France dominique.pastor@enst-bretagne.fr

S. Ben Jebara is with Research unit TECHTRA , Ecole Sup rieure des Communications de Tunis, Route de Raoued 3.5 Km, Cit  El Ghazala, 2083, Ariana, Tunisia, sofia.benjebara@supcom.rnu.tn

threshold and that of the noise power spectral density in the k th bin. We choose real values for H_k such that $0 \leq H_k \leq 1$. This inequality is the same as that satisfied by the standard Wiener filter and guarantees that the spectral power of the residual noise is less than or equal to that of the input noise. The expression of the perceptual filter is then

$$H_k = \min\left(\sqrt{\frac{T_k}{\gamma_k}}, 1\right) \quad (4)$$

This solution can also be derived by considering the filtering problem addressed in this section as a constrained optimization problem [4]. The filter defined by (4) reduces only the noise frequency components that are above the masking threshold because the others are not audible and can even mask residual noise. In practice, the two terms of the ratio in the right hand side of (4) must be estimated and the performance of the resulting filtering strongly depends on the accuracy of these estimates.

The masking threshold can be estimated on the basis of the outcome of a spectral subtraction ([4], [11]). This solution reduces the amount of additive noise but introduces musical noise. The tone-like nature of musical noise increases the energy per critical band. This can induce an overestimation of the masking threshold.

To avoid such a drawback, we estimate the speech signals by Wiener filtering. This method introduces less musical noise in comparison with spectral subtraction methods [3]. The Wiener filtering version that we use is based on the *a priori* SNR $\xi_k(m)$ estimated by [3]

$$\tilde{\xi}_k(m) = (1-\alpha)h(\chi_k(m) - 1) + \alpha \frac{|\tilde{S}_k(m-1)|^2}{|\tilde{X}_k(m-1)|^2}, 0 \leq \alpha < 1.$$

In this equation, $|\tilde{S}_k(m-1)|$ and $|\tilde{X}_k(m-1)|$ are the amplitude estimates of the k th spectral components of the clean speech signal and noise in the $(m-1)$ th analysis frame; $h(x) = x$ if $x \geq 0$ and $h(x) = 0$ otherwise; α is some weighting factor, we chose $\alpha = 0.98$ for our experiments; $\chi_k(m) = |Y_k(m)|^2 / |\tilde{X}_k(m)|^2$ is the *a posteriori* SNR where $|Y_k(m)|$ is the amplitude of the k th spectral component of the noisy speech signal in the m th analysis frame. The Wiener filter is then

$$W = \frac{\tilde{\xi}_k(m)}{1 + \tilde{\xi}_k(m)}, k = 0, \dots, N-1. \quad (5)$$

The estimate performed by the Wiener filtering thus defined is then employed to calculate the masking threshold.

In order to estimate the input noise spectral density, standard solutions are based on the use of a VAD. Frames detected by this VAD as noise alone serve to estimate the values γ_k , $k = 0, \dots, N-1$. The accuracy of the estimates depends on the performance of the VAD. When speech signals are corrupted by independent and AWGN with noise standard deviation σ_0 , we have $\gamma_k = \sigma_0^2$. In this case, what follows explains how to get an estimate of σ_0 without using any VAD.

III. ESTIMATION OF THE NOISE STANDARD DEVIATION

A. Theoretical results

The random variables encountered below are assumed to be defined on the same probability space and we write (a-s) for almost surely. Given a positive real value σ_0 , a sequence $X = (X_k)_{k \in \mathbf{N}}$ of random complex variables is said to be a *complex white Gaussian noise* (CWGN) with standard deviation σ_0 if the random variables X_k , $k = 1, 2, \dots$, are complex, mutually independent and identically Gaussian distributed with null mean and variance σ_0^2 ($X_k \sim \mathcal{N}_c(0, \sigma_0^2)$).

The *minimum amplitude* $\mathfrak{a}(S)$ of a sequence $S = (S_k)_{k \in \mathbf{N}}$ of random complex variables is defined by

$$\mathfrak{a}(S) = \sup \{ \alpha \in [0, \infty) : \forall k \in \mathbf{N}, |S_k| \geq \alpha \text{ (a-s)} \}. \quad (6)$$

If f is some map of the set of all the sequences of complex random variables into \mathbf{R} , we say that the limit of f is $\ell \in \mathbf{R}$ when $\mathfrak{a}(S)$ tends to ∞ and write that $\lim_{\mathfrak{a}(S) \rightarrow \infty} f(S) = \ell$ if, for any positive real value η , there exists some $\alpha_0 \in (0, \infty)$ such that, for every $\alpha \geq \alpha_0$ and every S such that $\mathfrak{a}(S) \geq \alpha$, $|f(S) - \ell| \leq \eta$.

Let $L_{\mathbf{C}}^2(\Omega)$ stand for the set of those complex random variables Y such that $E[|Y|^2] < \infty$. We denote by $\ell^\infty(\mathbf{N}, L_{\mathbf{C}}^2(\Omega))$ the set of those sequences $S = (S_k)_{k \in \mathbf{N}}$ of complex random variables such that $S_k \in L_{\mathbf{C}}^2(\Omega)$ for every $k \in \mathbf{N}$ and $\sup_{k \in \mathbf{N}} E[|S_k|^2]$ is finite.

The following result is a corollary of the limit theorem established in [6] for n -dimensional real random vectors. Given any random vector Y and any real number τ , $\mathcal{I}(|Y| \leq \tau)$ stands for the indicator function of the event $\{|Y| \leq \tau\}$.

Proposition 3.1: Let $Y = (Y_k)_{k \in \mathbf{N}}$ be some sequence of complex random variables such that, for every $k \in \mathbf{N}$, $Y_k = \varepsilon_k S_k + X_k$ where $S \in \ell^\infty(\mathbf{N}, L_{\mathbf{C}}^2(\Omega))$, $X = (X_k)_{k \in \mathbf{N}}$ is some CWGN with standard deviation σ_0 and $\varepsilon = (\varepsilon_k)_{k \in \mathbf{N}}$ is a sequence of random variables valued in $\{0, 1\}$ respectively.

Assume that

- (A1) for every $k \in \mathbf{N}$, S_k , X_k and ε_k are mutually independent;
- (A2) the random variables Y_k , $k \in \mathbf{N}$, are mutually independent;
- (A3) the random variables ε_k , $k \in \mathbf{N}$, are mutually independent;
- (A4) the priors $P(\{\varepsilon_k = 1\})$, $k \in \mathbf{N}$, are less than or equal to one half.

Given any natural number m and any pair (σ, T) of positive real numbers, define the random variable $\Delta_m(\sigma, T)$ by

$$\Delta_m(\sigma, T) = \frac{\sum_{k=1}^m |Y_k| \mathcal{I}(|Y_k| \leq \sigma T)}{\sum_{k=1}^m \mathcal{I}(|Y_k| \leq \sigma T)} - 2\sigma \frac{\int_0^T u^2 e^{-u^2} du}{1 - e^{-T^2}}.$$

Then, σ_0 is the unique positive real number σ such that, for every $\beta_0 \in (0, 1]$,

$$\lim_{\mathfrak{a}(\Lambda) \rightarrow \infty} \left\| \overline{\lim}_m \Delta_m(\sigma, \beta_0 g(\mathfrak{a}(\Lambda)/\sigma)) \right\|_{\infty} = 0 \quad (7)$$

uniformly in $\beta \in [\beta_0, 1]$ where, for every $x \in \mathbf{R}$, $g(x) = I_0^{-1}(e^{x^2})/2x$ with $g(0) = 1$ and I_0 is the standard zeroth-order modified Bessel function of the first kind.

In this statement, Y models a sequence of observations; for every given $k \in \mathbf{N}$, S_k stands for some possible random signal, ε_k is the possible occurrence of S_k and the complex noise is modelled by X . The assumption that $S \in \ell^\infty(\mathbf{N}, L^2_{\mathcal{C}}(\Omega))$ corresponds to the practical case of interest where the energies of the signals are finite and bounded.

B. The algorithm

With the same notations as above, suppose that we have m observations Y_1, \dots, Y_m . Let $L \in \mathbf{N}$ and set $\beta_\ell = \ell/L$ for every $\ell \in \{1, \dots, L\}$. The result stated above suggests estimating σ_0 by a possibly local minimum of

$$\sup_{\ell \in \{1, \dots, L\}} \Delta_m(\sigma, \beta_\ell g(a(S)/\sigma)).$$

According to proposition 3.1, the larger the minimum amplitude, the better the estimate. However, good results can be obtained even when the minimum amplitude of the signals is not large [7]. Therefore, we propose to perform an estimate of the noise standard deviation by taking $a(S) = 0$, a trivial bound for any signal norm. By so proceeding, we discard any assumption about the probability distributions of the signals and simply assume that these signals are less present than absent. Since $g(0) = 1$, we compute a minimum $\hat{\sigma}_0$ of

$$\sup_{\ell=1, L} \left\{ \frac{\sum_{k=1}^m |Y_k| \mathcal{I}(|Y_k| \leq \beta_\ell \sigma)}{\sum_{k=1}^m \mathcal{I}(|Y_k| \leq \beta_\ell \sigma)} - 2\sigma \frac{\int_0^{\beta_\ell} u^2 e^{-u^2} du}{1 - e^{-\beta_\ell^2}} \right\} \quad (8)$$

by means of a minimization routine for scalar bounded non-linear functions. For instance, the experimental results presented in the next section were obtained with the MATLAB routine `fminbnd.m`.

The search interval $[\sigma_{\min}, \sigma_{\max}]$ is constructed as follows. Sort the complex values Y_1, \dots, Y_m , $k = 1, \dots, m$, by increasing modulus. Let $Y_{[k]}$, $k = 1, \dots, m$, be the resulting sequence. The right endpoint of the search interval is $\sigma_{\max} = |Y_{[m]}|/\sqrt{2}$. Now, choose a real number Q close to 1 but less than or equal to $1 - \frac{m}{4(m/2-1)^2}$. A typical choice is $Q = 0.95$, provided that $m \geq 24$. Set $h = 1/\sqrt{4m(1-Q)}$ and $k_{\min} = m/2 - hm$. The left endpoint is $\sigma_{\min} = |Y_{[k_{\min}]}|/\sqrt{2}$. The reader is asked to refer to [6] and [7] for justifications of this construction.

Preliminary tests of the same type as those described in [7] show that $\hat{\sigma}_0$ tends to overestimate the value of σ_0 and suggest to estimate this noise standard deviation by

$$\tilde{\sigma}_0 = \sqrt{\frac{\sum_{k=1}^m |Y_k|^2 \mathcal{I}(|Y_k| \leq \hat{\sigma}_0)}{\sum_{k=1}^m \mathcal{I}(|Y_k| \leq \hat{\sigma}_0)}}. \quad (9)$$

We work on theoretical justifications of this final estimate.

According to [6] and [7] and as a good trade-off between computational load and accuracy, it is recommended to use the algorithm proposed above on a few hundred observations and to choose $L = m$. For instance, the experimental results of section IV were obtained on the basis of observation sets with 200 hundred samples each.

C. Application to speech processing

Consider K samples of speech signals corrupted by independent AWGN with standard deviation σ_0 . Let F_s stand for the sampling frequency. Split this set of observations into M disjoint frames of $N = 2^p$ samples each. We have $K = MN$. In practice, p will be chosen such that $NF_s \approx 20$ ms. Apply an N -DFT on each frame. We obtain a matrix $[U_{i,j}]_{i \in \{1, \dots, M\}, j \in \{0, \dots, N-1\}}$ of complex values where i is the frame index and j the DFT bin number. Because of the Hermitian symmetry of the DFT, we restrict our attention to the values $U_{i,j}$, $i \in \{1, \dots, M\}$, $j \in \{0, \dots, N/2 - 1\}$.

For each frame i and each bin j , we assume the random presence of a speech frequency component $S_{i,j}$ and that the probability of presence of $S_{i,j}$ is less than or equal to one half. This probability of presence may be larger than one half for low frequency components; however, for high frequency components, this probability of presence becomes less than or equal to one half and even relatively small. We thus assume that $U_{i,j} = \varepsilon_{i,j} S_{i,j} + X_{i,j}$. As above, $\varepsilon_{i,j} \in \{0, 1\}$ indicates whether the speech frequency component $S_{i,j}$ is present or absent in the j th bin of the i th frame. Since noise is white and Gaussian with standard deviation σ_0 , the complex random variables $X_{i,j}$ are mutually independent and identically distributed with $X_{i,j} \sim \mathcal{N}_{\mathcal{C}}(0, NC^2\sigma_0^2)$ when the DFT coefficients U_k , $k = 0, \dots, N-1$, of a sequence of N samples u_0, \dots, u_{N-1} are $U_k = C \sum_{n=0}^{N-1} u_n e^{-i2\pi nk/N}$.

Instead of performing an estimate of $\sqrt{NC^2\sigma_0^2}$ on the basis of the $MN/2 = K/2$ values we have, we follow the recommendation of the previous section and split our set of observations into subsets of $m = 200$ observations each. Each subset is used to perform an estimate of $\sqrt{NC^2\sigma_0^2}$. We then compute the average value of the $MN/2m$ estimates thus obtained. It then suffices to divide this average by $C\sqrt{N}$ to get an estimate of σ_0 . In order to deal with m observations that can reasonably be considered as mutually independent, these observations can be chosen randomly amongst the $MN/2$ values we have. However, this randomization does not affect significantly the results presented below.

IV. PERFORMANCE EVALUATION

The perceptual filtering associated with the estimator proposed in section III was tested as follows. We considered speech signals from the TIMIT database. These speech signals are recorded at 16 kHz and we downsampled them to 8 kHz before adding white Gaussian noise.

To estimate the noise standard deviation, we proceeded as described above with frames of $N = 256$ samples each. This estimate serves to estimate the masking threshold and tune the perceptual filtering. Both the estimation of the masking threshold and the perceptual filtering are achieved on the

basis of 32ms-duration frames (256 samples per frame) with a 50% overlap. The computation of the masking threshold is based on 18 critical bands.

We computed the SSNR and the MBSD of the proposed method. The SSNR and the MBSD are objective criteria for measuring the performance of a speech enhancement system. The SSNR is the average of the SNR values on short segments [8]. However, the SSNR is not relevant enough to measure the distortion of denoised speech signals and to assess the quality of perceptually motivated speech enhancement approaches. Such approaches purposely keep noise components that are inaudible because suppressing them could introduce unpleasant speech distortion. Many other objective measures have been developed. They correlate well with subjective measures of speech quality. Amongst them, the MBSD [13] is an improved version of the Bark Spectral Distortion (BSD) [12]. It extends the BSD by incorporating the masking threshold so as to differentiate audible distortions from inaudible ones. The MBSD proves to be even more correlated with speech quality than the BSD [13].

We computed the SSNR and the MBSD when the exact value of the noise standard deviation is used and the masking threshold is computed on the basis of the clean speech signals (before adding noise). The SSNR and the MBSD thus obtained are hereafter called the “Theoretical limits”.

The SSNR and the MBSD were also measured for the perceptual filter of section II when the noise standard deviation is estimated on the basis of signal-free time frames provided by an ideal DAV (“perceptual filter + DAV”).

Consider the Wiener filtering based, as in section II, on the recursive computation of the *a priori* SNR. We computed the SSNR and the MBSD when the estimate of the noise standard deviation is achieved via the estimator of section III (“Wiener + Noise Estimator”). We also calculated the SSNR and the MBSD when this noise standard deviation is estimated on the basis of signal-free periods of time (“Wiener + DAV”).

Fig. 1 presents the performance measurements obtained. The proposed method yields performance significantly close to the theoretical limits at various input SNRs. It outperforms the Wiener filtering. Surprisingly enough, it also performs better than the perceptual filtering tuned by the noise standard deviation estimate derived from an ideal DAV. We think that this relates to the fact that the number of observations used by our estimator is significantly larger than that available on the basis of signal-free time frames. Note also that the SNR improvement achieved by the proposed method is particularly significant for low input SNRs.

V. CONCLUSION AND FUTURE WORK

By combining a perceptually motivated approach for speech enhancement with a new estimator of the noise standard deviation, we denoise speech signals corrupted by independent AWGN without using any VAD. The proposed method reduces residual noise, limits speech distortions and outperforms Wiener filtering. Above all, its performance

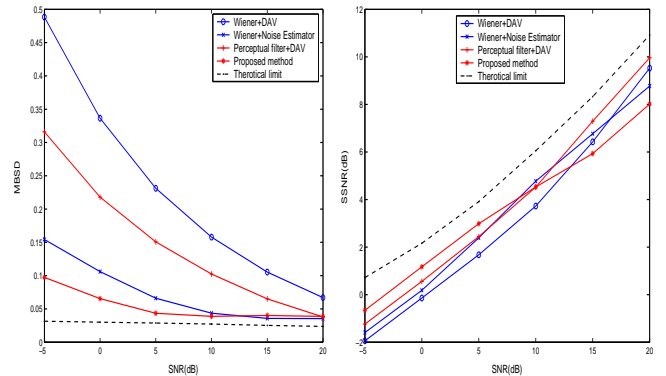


Fig. 1. MBSD and SSNR improvement for speech signals in independent AWGN with various SNRs.

measurements are very promising because they significantly approach those obtained by using the exact value of the noise standard deviation and computing the masking threshold on the basis of clean speech signals. Forthcoming work will involve further comparison to other techniques such as Ephraim and Malah’s [3] as well extension of the approach to non-white Gaussian noise.

REFERENCES

- [1] M. Berouti, R. Schwartz, and J. Makhoul, “Enhancement of speech corrupted by acoustic noise”, in *Proc. IEEE Int. Conf. on Acoustics, Speech, Signal Processing*, Washington DC, 1979, pp. 208-211.
- [2] S. Boll, “Suppression of acoustic noise in speech using spectral subtraction”, *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 27, no. 2, 1979.
- [3] Y. Ephraim and D. Malah, “Speech enhancement using a minimum mean square error short-time spectral amplitude estimator”, *IEEE Trans. on Acoust., Speech, Signal Processing*, vol. ASSP-32, 1984, pp. 1109-1121.
- [4] Y. Hu and P. Loizou, “Incorporating a psychoacoustic model in frequency domain speech enhancement”, *IEEE Signal Processing Letters*, 11(2), 2004, pp. 270-273.
- [5] Johnston, J. D, “Transforming of audio signals using perceptual noise criteria”, *IEEE Jour. Selected Areas Commun.*, vol. 6, no. 2, 1988, pp. 314-323.
- [6] D. Pastor, R. Gay, “A limit theorem for sequences of independent random vectors with unknown distributions and its applications to non-parametric detection”, *Collection des Rapports de Recherche de l’ENST Bretagne*, RR-2004001-SC, 2004.
- [7] D. Pastor, “Un théorème limite et un test pour la détection non paramétrique de signaux dans un bruit blanc gaussien de variance inconnue”, *20e colloque GRETSI sur le traitement du signal et des images, GRETSI’05*, Louvain-La-Neuve, 2005.
- [8] S. Quackenbush, T. Barnwell, and M. Clements, “Objective Measures of Speech Quality”. *Englewood Cliffs, NJ: Prentice-Hall*, 1988.
- [9] Te-Won Lee and Kaisheng Yao, “Speech enhancement by perceptual filter with sequential noise parameter estimation”, *International Conference on Acoustics, Speech, and Signal Processing, ICASSP ’04*, Volume 1, 2004, pp. 693-696.
- [10] D. Tsoukalas, M. Paraskevas, and J. Mourjopoulos, “Speech enhancement using psycho-acoustic criteria”, in *Proc. IEEE ICASSP 1993*, Minneapolis, MN, 1993, pp. 359-361.
- [11] N. Virag, “Singlechannel speech enhancement based on masking properties of the human auditory system”, *IEEE Trans. Speech and Audio Processing*, vol. 7, 1999, pp. 126-137.
- [12] S. Wang, A. Sekey and A. Gersho, “An objective measure for predicting subjective quality of speech coders”, *IEEE Journal on Selected Areas in Communications*, vol. SAC-10, 1992, pp. 819-829.
- [13] W. Yang, M. Dixon and R. Yantorno, “Modified bark spectral distortion measure which uses noise masking threshold”, *IEEE Speechcoding Workshop*, Pocono Manor, 1997, pp. 55-56.