A hybrid descent method for optimal sigmoid filter design

Kit Yan Chan[a], Sven Nordholm[a], Siow Yong Low[b], Pei Chee Yong[a], Ka Fai Cedric Yiu[c]

***Abstract* - In this letter, a hybrid descent method is used to determine a set of filter parameters for a sigmoid filter which attempts to work under various SNR conditions. It overcomes the limitations of the current sigmoid filters that performs effectively only at a single SNR. Results show that significant improvement in terms of better speech qualities can be achieved by the proposed sigmoid filter when working under various SNR conditions.**

Index Terms - Noise suppression, sigmoid function, speech enhancement, hybrid optimization method, various SNR conditions

I. Introduction

Single-channel speech enhancement remains a viable choice for enhancing noisy speech due to its simplicity and convenience of implementation. Two main tasks associated with single-channel techniques are noise suppression and speech preservation. It remains a challenge to optimally achieve both tasks. It is made difficult when the noise characteristic is highly non-stationary [1]. In noisy environments, the noise estimator is prone to erroneous estimation in the noise statistics, thereby causing a mismatch in the suppression. The mismatch in the noise statistics also produces annoying musical artifacts, which reduce the intelligibility of the enhanced speech.

Recently, a speech enhancement scheme in the modulation domain has been proposed in order to reduce the musical noise [2]. As a result, a good trade-off between temporal slurring and musical noise can be tuned [2]. Also, progress has been made in the algorithmic development for SNR based cost functions and/or different prior statistical models to improve speech quality [3,5,7]. However, given their complexity, these aforementioned algorithms demand heavy computational requirements. This is critical for computationally-sensitive applications like cochlear implants and hearing aids.

Motivated by the desire to achieve low computational cost, Hu et al. [8] proposed the sigmoid (SIG) filter, which requires that various parameters be estimated. Hu et al. [8] showed that the SIG can be readily used to enhance the noisy signal as the SIG follows a pattern similar to the human listener's psychometric function of intelligibility versus SNR. To further improve the effectiveness of the original SIG, Yong et al. [9] have developed a modified SIG (MSIG) with an a priori SNR estimate to cater real-time estimation of the SNR. It overcomes the one-frame delay when estimating the *a priori SNR* by matching the estimated clean speech spectrum and the *a priori SNR* as opposed to the previous frame. Results show that the modified SIG significantly outperforms the original *a priori SNR* based method and reduces audible noise and increases noise reduction for the commonly used gain functions, when the MSIG is developed under a particular SNR level.

Nevertheless, it is impractical to develop a single MSIG which can work effectively only at a single SNR level. It remains critical room for further improvement in determining the appropriate filter parameters. It is necessary to estimate appropriate filter parameters without having access to the statistics of SNR conditions. This paper proposes a hybrid descent algorithm (HAD) in order to determine appropriate filter parameters for various SNR conditions.

II. Sigmoid filter

Given a noisy signal $y(t)=x(t)+v(t)$, with the clean speech, $x(t)$, and the uncorrelated additive noise, $v(t)$, the enhanced speech spectrum, $\hat{X}(\omega,\ell)$, with respect to $x(t)$ can be generated by the MSIG [9] namely $G_{\text{MSIG}}(\omega,\ell,\kappa(\sigma))$, as

$$\hat{X}(\omega,\ell)=G_{\text{MSIG}}(\omega,\ell,\kappa(\sigma))\cdot Y(\omega,\ell). \qquad (1)$$

In (1), $Y(\omega,\ell)$ is the *M*-point short time Fourier transformation of $y(t)$; $\omega\in\left[\omega_{0,}\omega_{1,}....,\omega_{K-1}\right]$ is a real angular center frequency with *K* bands; $\ell\in[0,1,....,L-1]$ is the time frame index given by with *L* frames; and $\kappa(\sigma)$ is the filter parameters used to control the characteristic of G_{MSIG}. It works under the estimated SNR, namely σ . G_{MSIG} is formulated as shown in (2):

$$G_{\text{MSIG}}(\omega,\ell,\kappa(\sigma))=\frac{1}{1+e^{-k_1(\sigma)\cdot\left(\hat{\xi}(\omega,\ell)-k_2(\sigma)\right)}}\cdot\frac{1-e^{-k_3(\sigma)\cdot\hat{\xi}(\omega,\ell)}}{1+e^{-k_3(\sigma)\cdot\hat{\xi}(\omega,\ell)}}, \qquad (2)$$

where $\kappa(\sigma)$ is given as $\kappa(\sigma)=\left[k_1(\sigma),k_2(\sigma),k_3(\sigma)\right]$; $\hat{\xi}(\omega,\ell)$ is the *a priori* SNR estimated by the recently developed approach [9]. G_{MSIG} is effectively in real-time implementation when compared with the conventional decision-directed approach [5]; σ is estimated by the average of the *a posteriori* SNR which is given by:

$$\sigma=E\left[\frac{\left|Y(\omega,\ell)\right|^2}{E\left(\left|V(\omega,\ell)\right|^2\right)}\right]=E\left(\frac{\left|Y(\omega,\ell)\right|^2}{\lambda_v(\omega,\ell)}\right), \qquad (3)$$

with the *M*-point short-time Fourier transformation of the noise $V(\omega,\ell)$ and the noise power spectral density $\lambda_v(\omega,\ell)$. The first term of the G_{MSIG} represents a sigmoid function [8]. The slope and the mean of which are controlled by $k_1(\sigma)$ and $k_2(\sigma)$ respectively. They control the amount of musical noise, speech distortion and noise reduction of the enhanced speech. To balance the three quality measures, the sigmoid slope has to be sensitive towards speech and insensitive towards the variation of noise. The mean of the magnitude of the sigmoid function cannot be zero although $\hat{\xi}(\omega,\ell)$ is very small. Hence, poor noise reduction can be produced.

[a]K.Y. Chan, S. Nordholm and P.C. Yong are with the Department of Electrical and Computer Engineering, Curtin University, Australia. [b]S.Y. Low is with School of Electronics and Computer Science, University of Southampton, Malaysia campus, Johor, Malaysia. [c]K.F.C. Yiu is with the Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong. Sven Nordholm and Ka Fai Cedric Yiu are supported by RGC Grant PolyU. (5301/12E).

To improve the original SIG, the hyperbolic tangent function is introduced as the second term of G_{MSIG}. Figure 1 shows the magnitude of G_{MSIG} with respect to $\hat{\xi}(\omega,\ell)$, where different magnitudes are obtained with different filter parameters. Hence, the magnitude is decreased when $\hat{\xi}(\omega,\ell)$ is low. Also, the magnitudes of G_{MSIG} are smaller than those of the original SIG, when $\hat{\xi}(\omega,\ell)$ is low. Hence, it provides more noise reduction at low $\hat{\xi}(\omega,\ell)$ than those of the original SIG.

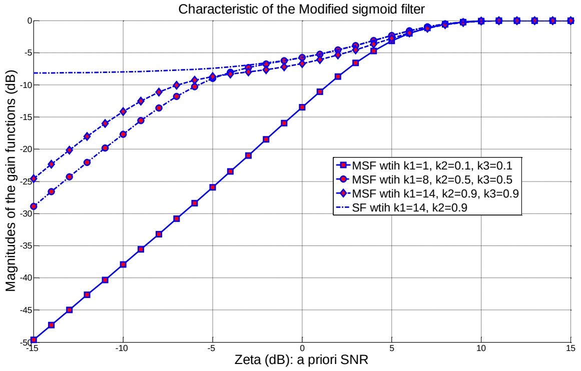

Figure 1 Characteristic of G_{MSIG} with different $\hat{\xi}(\omega,\ell)$

The speech quality of the enhanced speech can be optimized with appropriate $\kappa(\sigma)=\left[k_1(\sigma),k_2(\sigma),k_3(\sigma)\right]$. Therefore, $\kappa(\sigma)$, is necessary to be optimized with respect to an specified speech quality measure, $\Theta(..)$, by solving (4):

$$\mathrm{Opt}: J_{\mathrm{s}}\left(\kappa(\sigma)\right)=\Theta\left(\kappa(\sigma),X(\omega,\ell),Y(\omega,\ell)\right), \qquad (4)$$

where $\Theta(..)$ evaluates the speech quality of the enhanced speech $G_{\mathrm{MSIG}}\left(\omega,\ell,\kappa(\sigma)\right).Y(\omega,\ell)$ with respect to the original speech $X(\omega,\ell)$. The speech quality can be evaluated by the objective measurement algorithms which are used to evaluate the performance of speech enhancement algorithms [12]. As perceptual evaluation of speech quality (PESQ), SNRseg, log-likelihood can be the commonly used measures, the optimization problem (4) can be either a minimization or maximization problem, depending on the speech quality measure $\Theta(..)$.

However, the magnitude of G_{MSIG} varies with respect to $\hat{\xi}(\omega,\ell)$. As $\hat{\xi}(\omega,\ell)$ is related to the *a posteriori* SNR, a particular filter parameter $\kappa(\sigma)$ may work satisfactorily only for a single σ. In this paper, the HDA is proposed in order to generate a set of filter parameters namely $\mathrm{K}=\left[\kappa(\sigma_1),\kappa(\sigma_2),...,\kappa(\sigma_N)\right]$. Each $\kappa(\sigma_i)$ is determined with respect to a particular σ_i with i=1,2,…N. $\sigma_i\in\left[\sigma_1,\sigma_2,...,\sigma_N\right]=\Omega$, where Ω is the set for the *a posteriori* SNR and each σ_i is calculated based on (3). The noisy speech, $Y(\omega,\ell)$, is contaminated artificially by adding $V(\omega,\ell)$ with $X(\omega,\ell)$ under an artificial SNR namely ϕ_i, where $\phi_i<\phi_j\in\left[\phi_1,\phi_2,...,\phi_N\right]=\Phi$ and $i<j$ with i ,j=1,2,…N, and all ϕ_i is within the predefined upper bound, ϕ_1, and low bound, ϕ_N.

III. Parameter Estimation for Sigmoid Filters

As the dimension of $J_{\mathrm{s}}\left(\kappa(\sigma)\right)$ is not high, the simplex search algorithm could be used to locate the optima [14]. However, the global optimum may not be located due to the nonlinearity of $J_{\mathrm{s}}\left(\kappa(\sigma)\right)$. To seek the global optimum, simulated annealing [15] could be used. As the convergence rate of simulated annealing is relatively slow, a HDA comprising the simulated annealing and the simplex search technique is proposed to solve $J_{\mathrm{s}}\left(\kappa(\sigma)\right)$. It is used to generate a set of appropriate $\kappa^{opt}(\sigma_i)$ with respect to σ_i. Each σ_i is aligned with the artificial SNR ϕ_i, where $X(\omega,\ell)$, is artificially corrupted by $V(\omega,\ell)$ with the artificial SNR, ϕ_i.

In *Step* 1 of the HDA, i is initialized by setting i=1. *Step* 2 creates $Y(\omega,\ell)$, artificially with the signal to noise ratio, ϕ_i, and *Step* 3 estimates the *a posteriori* SNR, σ_1. The HDA searches the filter parameter with respect to σ_1. It generates an initial guess, $\kappa_k(\sigma_i)$, randomly with k=0 and i=1. Then, *Step* 5 uses the simplex search method to determine the optimum for $\kappa_k^*(\sigma_i)$, and *Step* 6 uses the simulated annealing [10] to explore a solution, $\kappa_k(\sigma_i)$ which is better than the optimum $\kappa_k^*(\sigma_1)$. *Steps* 5 and 6 keep iterating mutually until no better solution can be found. The searching process with respect to σ_i stops in *Step* 7, after converged. The converged solution $\kappa_k(\sigma_i)$ is returned as the outcome for σ_i with $\kappa^{opt}(\sigma_i)=\kappa_k(\sigma_i)$.

Hybrid Descent Algorithm (HDA)

```
Input: $X(\omega,\ell)$, $V(\omega,\ell)$, $\Phi$ ;  Output: $\mathrm{K},\Omega$ ;
Step 1: Initialize by setting $i=1$ and $k=0$.
Step 2: Create $Y(\omega,\ell)$ by artificially adding $V(\omega,\ell)$ to $X(\omega,\ell)$
    with the artificial SNR, $\phi_i$.
Step 3: Calculate the a posteriori SNR, $\sigma_i$, based on (3).
Step 4: Generate $\kappa_k(\sigma_i)$ randomly within the operational range of
    the filter parameters and evaluate the enhanced speech quality
    based on $J_{\mathrm{s}}\left(\kappa_k(\sigma)\right)$.
Step 5: Search for the local optimum $\kappa_k^*(\sigma_i)$ for $J_{\mathrm{s}}\left(\kappa_k(\sigma_i)\right)$ using
    the simplex search method with the initial guess $\kappa_k(\sigma_i)$, where
    $J_{\mathrm{s}}\left(\kappa_k^*(\sigma_i)\right)-J_{\mathrm{s}}\left(\kappa_k(\sigma_i)\right)\le-\varepsilon_k$ with a positive $\varepsilon_k$.
Step 6: Starting from $\kappa_k^*(\sigma_i)$, execute the simulating annealing until
    a solution $\kappa_k(\sigma_i)$ is obtained, where
    $J_{\mathrm{s}}\left(\kappa_k(\sigma_i)\right)-J_{\mathrm{s}}\left(\kappa_k^*(\sigma_i)\right)\le-\delta_k$ with a positive parameter $\delta_k$.
Step 7: If $k$ is larger than a predefined value, set $\kappa^{opt}(\sigma_i)=\kappa_k(\sigma_i)$,
    and go to Step 8. Otherwise, set $k=k+1$ and go to Step 5.
```

Step 8: If $i < N$, then initialize with $k = 0$, put $\kappa_0(\sigma_{i+1}) = \kappa^{opt}(\sigma_i)$, set $i = i + 1$ and go to *Step* 2.

Step 9: Create and return the two sets Ω and K where

$$\Omega = [\sigma_1, \sigma_2, ..., \sigma_N] \text{ and } \mathrm{K} = [\kappa^{opt}(\sigma_1), \kappa^{opt}(\sigma_2), ..., \kappa^{opt}(\sigma_N)].$$

For σ_{i+1}, the filter parameter, $\kappa^{opt}(\sigma_{i+1})$ is determined by performing *Steps* 3 to 8, while $\kappa^{opt}(\sigma_i)$ is used as the initial guess. When *i* in *Step* 8 reaches *N*, the sets for the filter parameters, $\kappa^{opt}(\sigma_i) \in \mathrm{K}$, and the *a posteriori* SNR, $\sigma_i \in \Omega$, are created. $\kappa^{opt}(\sigma_i)$ is used in G_{MSIG} when the *a posteriori* SNR is closest to σ_i.

IV. Experimental Results

A. Experimental setup

The HDA was developed using the Matlab R2011b, whereby the routine ‘fmincon’ is used to determine the local optimum of the filter parameters and the inbuilt parameters of the simulated annealing were selected based on [15] where the iteration number was large enough to avoid premature convergence.

The HDA was evaluated based on a database of noisy speech sequences, namely NOIZEUS [13], which were corrupted by pink noise, babble noise and factory noise in the ranges of -15 to 15 dB. Three commonly used speech quality measures, namely PESQ [12], SNRseg [16] and the log-likelihood (LLR) [16], were used to evaluate the performance of the enhanced speech.

The operational range for the SNR is defined between -15 dB and 15 dB. A set of filter parameters is determined with respect to the noisy signal contaminated with the SNR of -15dB, -14dB, … and 15dB respectively, where the filter parameters are optimized using the HDA. Using the optimized filter parameters, the filter is evaluated by the test set, where the test set was corrupted with SNR -14.5dB in steps of 1dB till 14.5dB. Here, the filter parameters which were optimized with respect to the nearest *a priori* SNR were used in the filter working under an untrained noisy environment.

B. Performance evaluation

Figure 2 shows the PESQ obtained by four G_{MSIG} filters against the pink noise with respect to the test set: i) the adaptive G_{MSIG}, which uses the filter parameters optimized with respect to the training set; ii) the singly trained filter for low SNR namely G_{MSIG}-(-15) which was optimized with respect to -15dB; iii) another singly trained filter for high SNR namely G_{MSIG}-(15) which was optimized with respect to 15dB; and iv) the optimal G_{MSIG}, the filter parameters of which are optimized directly with respect to the test set. Hence, the performance of the adaptive G_{MSIG} can be compared with the optimal one.

The results indicate that the PESQ obtained by the adaptive G_{MSIG} are better than those obtained by the G_{MSIG}-(-15) and G_{MSIG}-(15), where G_{MSIG}-(-15) is close to the optimized PESQ with low SNRs and G_{MSIG}-(15) is close to those with high SNR. These results indicate that the singly trained filters, G_{MSIG}-(-15) and G_{MSIG}-(15), perform well only with respect to their own developed SNRs, while the adaptive G_{MSIG} performs well for the whole range of SNRs. Hence, the adaptive G_{MSIG} can align the optimized PESQ irrespective of the input SNR. Also, similar performance can be obtained by the adaptive G_{MSIG} compared with the optimal G_{MSIG}.

Figures 3 and 4 show the SNRseg and LLR with respect to the pink noise. Similar results are produced when the adaptive G_{MSIG} can obtain the best SNRseg and LLR which are closest to the optimized PESQ. Figures 5 to 7 show the speech quality measures for Factory noise, and Figures 8 to 10 show those for Babble noise. Similar results were obtained when the adaptive G_{MSIG} can obtain the best speech qualities. These results indicate that the adaptive G_{MSIG} performs well for the whole range of SNRs, while G_{MSIG}-(-15) and G_{MSIG}-(15) perform well only at their developed SNRs. Also, the performance of the adaptive G_{MSIG} aligns with the optimal G_{MSIG} under various SNR conditions.

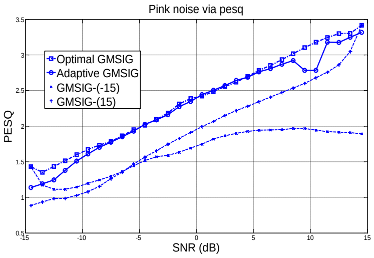

Figure 2 Pink noise via PESQ

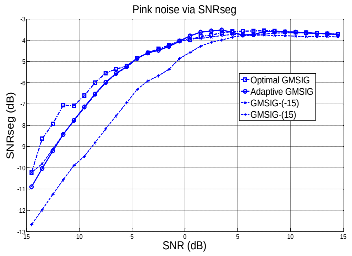

Figure 3 Pink noise via SNRseg

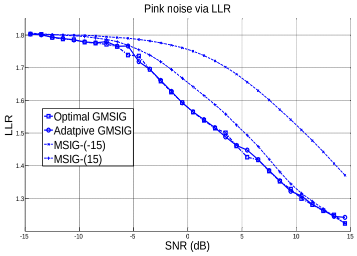

Figure 4 Pink noise via LLR

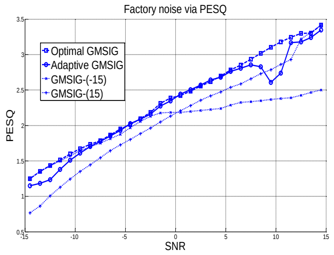

Figure 5 Factory noise via PESQ

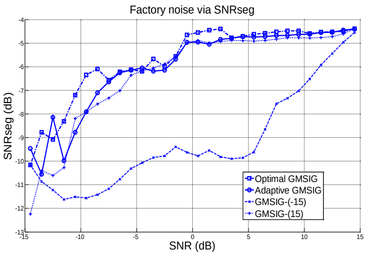

Figure 6 Factory noise via SNRseg

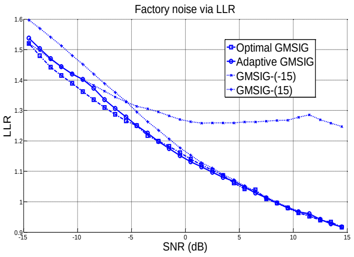

Figure 7 Factory noise via LLR

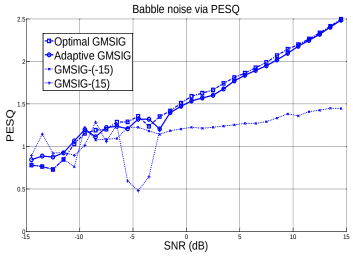

Figure 8 Babble noise via PESQ

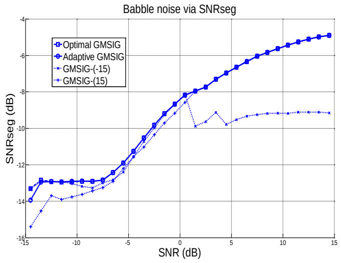

Figure 9 Babble noise via SNRseg

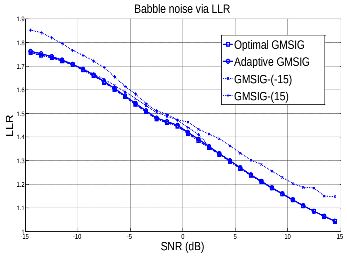

Figure 10 Babble noise via LLR

V. Conclusion

A HDA is proposed to optimize the filter parameters of the MSIG, in order to enhance the quality of speech under various SNR conditions. The HDA generates a training set of filter parameters for some SNRs, and those parameters are used as the closest neighboring SNRs for the MSIG working in an untrained noisy environment. Results show that significant improvement in terms of the three speech quality measures can be achieved.

References

[1] P. K. Ghosh, A. Tsiartas, and S. Narayanan. Robust voice activity detection using long-term signal variability. IEEE Trans. Speech Audio Process., vol. 19, no. 3, pp. 600-613, 2011.

[2] K. Paliwal, B. Schwerin and K. Wo, Speech enhancement using a minimum mean-square error short-time spectral modulation magnitude estimator, Speech Comm., vol. 54, no. 2, 282–305, 2012.

[3] I. Andrianakis and P. White, Speech spectral amplitude estimators using optimally shaped gamma and chi priors, Speech Comm., vol. 51, no. 1, pp. 1–14, 2009.

[5] C. Breithaupt and R. Martin, Analysis of the decision-directed SNR estimator for speech enhancement with respect to low-SNR and transient conditions, IEEE Trans. Audio Speech Lang. Process., vol. 19, no. 2, pp. 277–289, , 2011.

[7] E. Plourde and B. Champagne, Generalized bayesian estimators of the spectral amplitude for speech enhancement. IEEE Signal Process. Lett., vol. 16, no. 6, pp. 485–488, 2009.

[8] Y. Hu, P.C. Loizou, N. Li and K. Kasturi, Use of a sigmoidal-shaped function for noise attenuation in cochlear implants, JASA Express Letters, vol. 122, no. 4, pp., 128-134, 2007.

[9] P.C. Yong, S. Nordholm, H.H. Dam, Optimization and evaluation of sigmoid function with a priori SNR estimate for real-time speech enhancement, Speech Communication, vol. 55, no. 2, pp. 358-376, 2012.

[12] Y. Hu and P. Loizou, Evaluation of objective quality measures for speech enhancement. IEEE Trans. Audio Speech Lang. Process., vol. 16, no. 1, pp. 229–238, 2008.

[13] P. C. Loizou, Speech Enhancement Theory and Practice. CRC Press, 2007.

[14] R. Fletcher, Practical Methods of Optimization, Wiley, 1987.

[15] S. Kirkpatrick, C. D. G. Jr. and M. P. Vecchi, Optimization by simulated annealing, Science, vol. 220, pp. 671–680, 1983.

[16] S. Quackenbush, T. Barnwell, and M. Clements, Objective Measures of Speech Quality. Prentice Hall, Englewood Cliffs, NJ., 1988.