A decision-directed adaptive gain equalizer for assistive hearing instruments

Kit Yan Chan, *Member, IEEE,* Siow Yong Low, Sven Nordholm, *Senior Member, IEEE,* and Cedric Ka Fai Yiu,

***Abstract*—Assistive hearing instruments have a significant impact on speech enhancement when the signal to noise ratio is low. These instruments are usually developed by using the conventional adaptive gain equalizer (AGE) which has low computational complexity and low distortion in real-time speech enhancement. The conventional adaptive gain equalizers are intended to boost the speech segments of speech signals but they are incapable of suppressing noise segments. The overall speech quality of the assistive hearing instruments may be reduced, as the noise segments still cannot be filtered out. In this paper, a decision-directed adaptive gain equalizer (AGE) is proposed for assistive hearing instruments. It aims to overcome the limitation of the conventional AGE, which is capable only of boosting speech segments in noisy speech but incapable of suppressing noise segments. The proposed approach simultaneously boosts the speech segments and suppresses noise segments in noisy speech. Experimental results with different types of real-world noise indicate that the proposed method achieves better speech quality than does the conventional AGE. The resulting method provides improved functionality for assistive hearing instruments.**

Keywords - Assistive hearing instruments, single channel filter, speech booster, adaptive gain equalizer, Perceptual Evaluation of Speech Quality (PESQ), particle swarm optimization, heuristic algorithms, speech enhancement

I. Introduction

Assistive hearing instruments are in high practical demand in scenarios where there is loud environmental noise such as those found in factories or workshops which affects the speech intelligibility and effectiveness of telecommunication media [1]. Noisy speech often results in lower intelligibility and listener fatigue. Hence, it is necessary to use assistive hearing instruments to reduce additive noise from noisy speech while ensuring that speech components remain as undistorted as possible [2].

The adaptive gain equalizer (AGE) [3] can efficiently and economically be used in assistive hearing instruments, since it is a single channel approach and thus it offers low complexity, low processing delay and low distortion in real-time speech enhancement [4]–[6]. It overcomes the multi-channel microphone approaches [7] which are more expensive as they require more microphones.

The conventional AGE can be implemented in assistive hearing instruments by modifying the magnitude spectrum of a speech signal [8], [9]. It adjusts the weighting on the magnitude spectrum, in order to impose a high gain if speech is presented. Alternatively, a unity gain is applied if no speech or only ambient noise is presented. By doing so, the conventional AGE acts like a speech booster. It amplifies the magnitude spectrum when speech is active, and remains idle when speech is inactive. However, one of the main limitations of the conventional AGE is that during the "idling period", the background noise is not suppressed and this results in a reduction of noise suppression capability. Although the speech signal is boosted by the assistive hearing instrument implemented with the conventional AGE, the background noise can still be heard.

In this paper, a novel AGE, namely a decision-direct AGE, is proposed by merging the mechanisms of the conventional AGE [3] and the decision-directed approach [10]. Unlike the conventional AGE, the proposed technique not only boosts the magnitude spectrums containing the speech components, but it also suppresses ambient noise. Hence, background noise can be suppressed when the direct-decision AGE is implemented in the assistive hearing instrument. Also, a tradeoff weight is introduced in the direct-decision AGE in order to control the amount of acceptable speech distortion and noise suppression. It allows for engineering judgement to obtain the desired tradeoff with respect to a perceived quality of the speech signal, while developing the assistive hearing instrument.

To determine the optimal tradeoff weight, we could ask numerous people to listen to the enhanced speech signals, rank the audio quality with a number between one and five, and then average these for the final result. However, this is a time consuming and expensive procedure. Here, the tradeoff weight is optimized with respect to a commonly used objective measure for speech quality namely the Perceptual Evaluation of Speech Quality (PESQ) [11], which has been standardized by the International Telecommunication Union. The PESQ scores yield a high correlation with hearing intelligibility scores [12]. It is commonly used by speech algorithm to reflect an increase in speech intelligibility. Because the assistive hearing instrument is developed based on the optimal direct-decision AGE with respect to the PESQ, the audibility and annoyance of complex distortions can be optimized.

However, the relationship between the PESQ score and the enhanced speech signal is not linear as an improvement in signal to noise ratio (SNR) does not translate to an improvement in speech intelligibility. [13]. The problem on determining an optimal tradeoff weight with respect to the PESQ is multi-optimum. Conventional gradient-based optimization methods might not be able to determine this tradeoff weight properly, as there is no guarantee that the global optimum will be obtained.

Kit Yan Chan and Sven Nordholm are with the Department of Electrical and Computer Engineering, Curtin University; Siow Yong Low is with the School of Electronics and Computer Science University of Southampton, Malaysia Campus; Cedric Ka Fai Yiu is with the Department of Applied Mathematics, The Hong Kong Polytechnic University; The corresponding author of this paper is Kit Yan Chan. Tel: +618 9266 2945; Fax: +618 9266 2819; e-mail: kit.chan@curtin.edu.au

Manuscript received xxxx xx, xxxx.

Therefore, to solve this optimization problem, a hybrid optimization algorithm is proposed that integrates the mechanisms of both local and global search methods. It first uses a global optimizer, namely the particle swarm optimization algorithm (PSO) [14], [15], to generate an initial solution with good PESQ scores, since the PSO is effective in solving many difficult optimization problems [16] particularly relating to noise suppression [17]–[20]. Then, it uses a gradient-based optimization method [21] to locate the optima with respect to the PESQ.

The effectiveness of the proposed direct-decision AGE is evaluated by performing speech enhancement under four noisy environments namely white noise, factory noise, babble noise and volvo noise. Experimental results indicate that the proposed direct-decision AGE obtains better PESQ compared to the conventional AGE. Furthermore, analytical results show that the improved performance also contributes to better noise suppression, and negligible expense on both target signal distortion and musical tones which are produced by the proposed direct-decision AGE. Hence, higher hearing intelligibility scores can be obtained when the direct-decision AGE is used in the development of assistive hearing instruments.

II. Assistive hearing Instruments with AGE

The conventional AGE can be used to boost or suppress the signal with the specified frequency range by controlling the gain in a particular frequency band [8], [9]. It acts as a speech booster when speech is present, and it remains idle when speech does not exist. It first decomposes the input signal into M number of sub-bands, which is illustrated in Figure 1. After that, each sub-band signal is individually adapted by a particular gain function, based on the estimated signal to noise ratio (SNR) in each sub-band at every time instant. The weighting of each sub-band is adaptively increased by the magnitude of its corresponding gain function when the speech is dominant. Then, all the sub-band signals is added in order to reconstruct a full-band signal. Hence, this speech enhancement approached is named as adaptive gain equalizer (AGE) as aforementioned.

Since the conventional AGE is implemented in the time domain, both the sub-band decomposition and reconstruction processes can be performed in a straightforward manner. It can be implemented either on digital or analog circuits with small computational complexity in terms of few million instructions per second [4]. Also, no voice activity detection is required for the conventional AGE since speech enhancement is based on a continuous estimate of the SNR in each sub-band. It eliminates the step of voice activity detection which are difficult to tune under low SNR. Therefore, it can be implemented effectively and economically in real-time speech enhancement for assistive hearing instruments [5].

We consider $s(k)$ to be the original speech where the index k represents the sampled time index for time instant, $t = kT_s$, in which T_s is the sampling period. Also, let $v(k)$ be the environmental noise such that the noisy speech, $x(k)$ is given by

$$x(k) = s(k) + v(k). \tag{1}$$

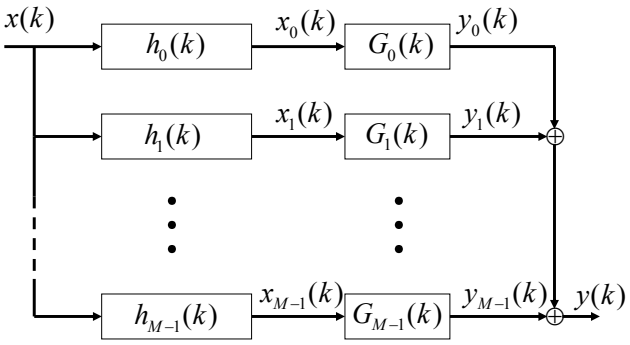


Figure 1. The structure of the adaptive gain equalizer (AGE) with weighting $G_m(k)$.

The m-th sub-band representation of the noisy speech illustrated in Figure 1 is given by:

$$x_m(k) = x(k) * h_m(k) = s_m(k) + v_m(k) \tag{2}$$

where $*$ denotes the convolution operator; $h_m(k)$ is the bandpass filter in the m-th band; and $s_m(k)$ and $v_m(k)$ are the m-th sub-band representation of the original speech and noise respectively.

The enhanced speech processed by the conventional AGE namely $G_m^{SB}(k)$ is given as

$$y(k) = \sum_{m=0}^{M-1} G_m^{SB}(k) x_m(k), \tag{3}$$

where $G_m^{SB}(k)$ denotes the gain function of the m-th sub-band and M is the number of subbands. $G_m^{SB}(k)$ acts as the speech booster [3]. It boosts the signal power during the active speech period and remains idle during the speech silence period.

$G_m^{SB}(k)$ is represented by the ratio of the estimate of the noisy signal level, $A_m(k)$, to the estimate of the noise floor, $N_m(k)$, and is illustrated as:

$$G_m^{SB}(k) = \left[\frac{A_m(k)}{N_m(k)}\right]^{\frac{1}{p}}. \tag{4}$$

where $A_m(k)$ and $N_m(k)$ are given as (5) and (6) respectively when the noise is slowly varying:

$$A_m(k) = (1-\alpha)A_m(k-1) + \alpha|x_m(k)|^p; \tag{5}$$

$$N_m(k) = \begin{cases} (1+\beta)N_m(k-1), & A_m(k) > N_m(k-1) \\ A_m(k), & A_m(k) \le N_m(k-1); \end{cases} \tag{6}$$

$|\cdot|$ denotes the absolute value; $|x_m(k)|$ is the magnitude spectrum of $x_m(k)$; and p is the gain rise which represents the subtraction type. In the conventional spectral substraction technique, $G_m^{SB}(k)$ is used for magnitude subtraction, when $p = 1$. When $p = 2$, $G_m^{SB}(k)$ is used for power substraction; α is a smoothing constant which is used to control the sensitivity of $G_m^{SB}(k)$ with respect to the envelope of the sub-band signal; and β is used to control how fast $N_m(k)$ can be adapted to the noise changes.

The current estimate of the noise floor in (6), $N_m(k)$, is incrementally increased when the current estimate of the noisy signal level, $A_m(k)$, is greater than the previous estimate of the noise floor, $N_m(k-1)$. Conversely, when $A_m(k)$ is smaller than $N_m(k-1)$, $N_m(k)$ is updated based on $A_m(k)$. Therefore, this update acts as an "attack and decay" change, and the estimate of the noise floor is imposed as $N_m(k) > 0$ in order to ensure the stability of $G_m^{SB}(k)$. Also, the following limiter is imposed to avoid excessive amplification of $G_m^{SB}(k)$:

$$G_m^{SB}(k) = \begin{cases} G_m^{SB}(k), & G_m^{SB}(k) \leq C \\ C, & G_m^{SB}(k) > C \end{cases} \tag{7}$$

where C is the upper limit.

Based on (4), if $A_m(k) \approx N_m(k)$, the estimate of the noisy signal level, $A_m(k)$, during the speech silent period, is approximately the same as the estimate of the noise floor. Hence, $G_m^{SB}(k) \approx 1$, and $G_m^{SB}(k)$ approaches the function of a by-pass filter. This is the basic limitation of the conventional AGE in that the noise is fully by-passed and still exists, when there is no speech signal generated by the user. Therefore, background noise can still be heard, when the assistive hearing instrument is implemented using the conventional AGE. In order to filter the noise during the speech silent period, an improved AGE namely decision-directed AGE is proposed in the following section.

III. Decision-Directed AGE

A. Mechanism of Decision-Directed AGE

In this paper, a decision-directed AGE, namely $G_m^{DD}(k)$, is proposed to operate in a twofold manner by incorporating the mechanisms of the noise suppression filter [22] and the conventional AGE [3]. When the speech is active, it amplifies the speech component and boosts the speech signal as similar operation as the conventional AGE. When the speech is silent, it imposes a lower gain on the noise envelope and suppresses noise contribution in an operation similar to the noise suppression filter. Hence, the noise can still be processed during the speech silent period, when the assistive hearing instrument is implemented with the conventional AGE.

Prior to enhancing noisy speech, the decision-directed AGE first estimates the *a priori* SNR for the m-th sub-band, $R_m^{prio}(k)$, based on the following decision-directed approach [10],

$$\begin{aligned} R_m^{prio}(k) &= (1-\gamma)\max\{R_m^{post}(k-1), 0\} \\ &\quad +\gamma\frac{|G_m^{NS}(k-1)x_m(k-1)|^p}{N_m(k)^p} \end{aligned} \tag{8}$$

where the *a posteriori* SNR, $R_m^{post}(k-1)$ is defined by

$$R_m^{post}(k-1) = \frac{|x_m(k-1)|^p}{N_m(k-1)^p} - 1; \tag{9}$$

the parameter, γ, is a weighting constant and the noise envelope; $N_m(k)$ is readily defined in equation (6); and $R_m^{post}(k-1)$ in equation (9) is a local estimate of the SNR computed from the current data frame and $R_m^{prio}(k)$ in equation (8) represents the SNR of the estimated unknown spectrum from the previous frame.

By coupling the estimations of $R_m^{prio}(k)$ in (8) and $R_m^{post}(k-1)$ in (9), the noise suppression filter, $G_m^{NS}(k)$, can be produced, where $R_m^{prio}(k)$ can be referred to the information on the unknown envelope signals. The resulting $R_m^{prio}(k)$ can be used to minimize any annoying musical phenomenon that is prevalent in conventional methods to an acceptable level [23]–[27]. Here the noise suppression filter, $G_m^{NS}(k)$, is represented by a commonly used Wiener filter [22] which is given as

$$G_m^{NS}(k) = \frac{R_m^{prio}(k)}{R_m^{prio}(k)+1}; \tag{10}$$

Based on (10), during the speech silent period (i.e., $s_m(k) \approx 0$), the *a priori* SNR becomes $R_m^{prio}(k) \approx 0$, and the gain function reduces consequently to $G_m^{NS}(k) \approx 0$. This in turn results in an artificial sounding output whereas during the speech silent period, there is a complete silence. Hence, this may result in an unnatural sounding background.

To avoid this unnatural sound, the decision-directed AGE, $G_m^{DD}(k)$, is proposed by injecting the comfort noise onto the overall output. It is intended to produce two benefits namely: i) it avoids unnatural complete silence, and ii) it masks the presence of any potential noise artifacts. This comfort noise is injected based on the tradeoff between $G_m^{SB}(k)$ and $G_m^{NS}(k)$ which is given as:

$$G_m^{DD}(k) = \lambda G_m^{NS}(k) + (1-\lambda)G_m^{SB}(k) \tag{11}$$

where the tradeoff weight, λ, controls the amount of comfort noise that is injected in the update. Based on (11), the amount of comfort noise is injected from the $G_m^{SB}(k)$ to $G_m^{DD}(k)$ according to the value of λ, since $G_m^{SB}(k)$ is a speech booster which passes the noise to $G_m^{DD}(k)$. When $\lambda = 0$, $G_m^{DD}(k)$ in (11) reverts to $G_m^{SB}(k)$. Hence, it provides little alteration of the noise during speech silent periods. Better speech quality can be produced, when the assistive hearing instrument is implemented with the conventional AGE.

B. Optimization of Decision-Directed AGE

As an illustration, the m-th enhanced speech, $y_m(k)$, obtained by $G_m^{DD}(k)$ is given by:

$$y_m(k) = G_m^{DD}(k)x_m(k) = G_m^{DD}(k)[s_m(k) + v_m(k)]. \tag{12}$$

By substituting (11) with (12), we obtain

$$\begin{aligned} y_m(k) &= [\lambda G_m^{NS}(k) + (1-\lambda)G_m^{SB}(k)][s_m(k) + v_m(k)] \\ &= (1-\lambda)G_m^{SB}(k)v_m(k) + (1-\lambda)G_m^{SB}(k)s_m(k) \\ &\quad + \lambda G_m^{NS}(k)v_m(k) + \lambda G_m^{NS}(k)s_m(k), \end{aligned} \tag{13}$$

which can be simplified as:

$$y_m(k) = f_{\text{AGE}}(x_m(k), k, \lambda), \tag{14}$$

where $x_m(k)$ is the noisy speech corrupted by the noise $v_m(k)$, and f_{AGE} represents the mapping of the decision-directed AGE from $x_m(k)$ to $y_m(k)$ with the tradeoff parameter λ.

When the speech is not active (i.e. $s_m(k) \approx 0$), $G_m^{NS}(k) \approx 0$ according to (10) and the output of the decision-directed AGE is simply the noise floor which is given by:

$$y_m(k) \approx (1-\lambda) G_m^{SB}(k) v_m(k). \tag{15}$$

Based on equation (4), $G_m^{SB}(k) \approx 1$ during the speech inactive periods. Hence, the output is a scaled version of the actual background noise. After injecting the comfort noise as shown in equation (15), the overall output avoids the unnatural total silence periods during the speech inactive periods. Therefore, the tradeoff weight λ provides a variable adjustment with respect to noise suppression, speech comfort and distortion levels. When λ approaches unity, a higher speech distortion level is being generated for more noise suppression. When λ approaches zero, the enhanced speech has less distortion but less noise is filtered from the noisy speech. Therefore, the speech quality in terms of speech reduction and noise distortion for $G_m^{DD}(k)$ can be traded-off by using an appropriate λ.

However, perceptional speech quality may not be improved, although we optimize either the speech reduction and noise distortion given by the $G_m^{DD}(k)$, since the perceptional speech quality is not closely correlated to these two fundamental audio criteria. For example, perceptually masked coding noise, at a typical SNR of 13dB, can be completely inaudible, whereas random noise at the same value of SNR would be extremely disturbing [28]. Hence, optimization with respect to either speech reduction or noise distortion might not generate the optimal speech quality under low SNR.

In order to obtain a more reliable speech quality for the enhanced speech, we use the popular objective quality measure namely PESQ to determine the tradeoff parameter λ, as the PESQ can yield a modestly high correlation with intelligibility scores [12]. Also it can be used to correctly distinguish between audible and inaudible distortions and this has proven to be the best way of accurately predicting the audibility and annoyance level of complex distortions [29], [30]. Hence, speech quality measure with respect to more audio criteria can be indicated for speech enhancement mechanisms [13], [31]. Better speech quality can be produced, when the assistive hearing instrument is implemented using the directed-decision AGE optimized with respect to PESQ.

The mean opinion score (namely r_{MOS}) for the PESQ can be given by the formulation which creates an intrusive test between the enhanced speech, $y_m(k)$, and the original speech, $s_m(k)$:

$$r_{\text{MOS}} = F_{\text{PESQ}}[s_m(k), y_m(k)]. \tag{16}$$

To determine r_{MOS} given by F_{PESQ}, the detailed computation is given in [28]. First, $s_m(k)$ and $y_m(k)$ are aligned to the same constant power level which is corresponded to the listening level used in subjective tests. The aligned signals, $s_m(k)$ and $y_m(k)$, undergo an auditory transformation in order to mimic the key properties of human hearing. Hence, the speech components, which are inaudible to listeners, can be filtered. Then, the two disturbance parameters, namely the absolute (symmetric) disturbance and the additive (asymmetric) disturbance are calculated using non-linear averages over specific areas of the error surface, where the absolute (symmetric) disturbance is a measure of absolute audible error and the additive (asymmetric) disturbance is a measure of audible errors that are much louder than the original speech signal.

By substituting (14) into (16), the optimal tradeoff weight, λ^{opt}, with respect to the PESQ can be determined by solving the following optimization problem:

$$J = \min_{\lambda \in [0..1]} \quad F_{\text{PESQ}}[s_m(k), f_{\text{AGE}}(x_m(k), k, \lambda)] \tag{17}$$

Solving the optimization problem (17) could be difficult since it consists of two nonlinear functions, F_{PESQ} and f_{AGE}, although the dimension of optimization problem (17) is not high. In the following section, a Hybrid Optimization Algorithm (HOA) that integrates the mechanisms of both local and global searches is introduced in order to solve this optimization problem.

IV. Hybrid Optimization Algorithm

The gradient-based method could be used to find the optimum, λ^{opt}, of the optimization problem (17) by systematically moving the solution space [21]. However, there is no guarantee that the global optimum can be obtained due to the nonlinearity of (16). Therefore, a global optimizer, namely PSO [14], [15], is used to seek the global optimum of (17), since the effectiveness of the PSO has been demonstrated in solving many difficult optimization problems [16], particularly in the development of noise suppression approaches [17]–[20], [32]–[34]. Also, the PSO operation is mostly based on two formulations on determining particle velocities and particle locations. It is much simpler than the evolutionary algorithms which involves a few evolutionary operations such as crossover, mutation, reproduction, chromosome selection and gradient determination (when the differential evolutionary algorithm is used). Hence, computational time can be saved, as the simpler PSO operation is used. The PSO is selected to be used in this research.

However, the PSO takes a longer convergence time to locate the optimum compared with the gradient-based method. Consequently, the Hybrid Optimization Algorithm (HOA) comprising the mechanisms of the PSO algorithm and the gradient-based method is proposed to solve (16). In the HOA, the solution obtained by the PSO is used as the initial solution of the gradient-based method. Based on the initial solution, the optimum can be located more effectively by the gradient-based method than by solely using the PSO. The pseudo code of the HOA given in Algorithm IV.1 is used to determine an optimum, λ^{opt}, with respect to the PESQ formulated in (16), when the noisy speech, $x_m(k)$, and the original speech, $s_m(k)$, are given.

In the HOA, $\lambda_i(t)$ denotes the i^{th} particle at the t^{th} iteration where i = 1, 2,... , N_s. Each particle $\lambda_i(t)$ represents the tradeoff weight of the directed-decision AEG. At the 1^{st} iteration (i.e. $t = 1$), N_s particles are generated randomly within the range from 0 to 1, and then each particle is evaluated

based on the objective function (16) with respect to $x_m(k)$ and $s_m(k)$. In (19), the current particle position, $\lambda_i(t)$, at the t^{th} iteration is governed by the current particle velocity, $v_i(t)$, and the previous particle position, $\lambda_i(t-1)$, at the $(t-1)^{th}$ iteration, where the velocity of the i^{th} particle at the t^{th} iteration is given by:

$$v^i(t)=\begin{cases} v_{max} \text{ if } v_{max} < v^i(t)' \\ v^i(t) \text{ if } v_{max} > v^i(t)' \end{cases} \text{ with}$$

$$v^i(t)' = k \cdot \{\omega(t) \cdot v^i(t-1)' + \varphi_1 \cdot r_1 \cdot (p^i - \lambda_i(t-1)) \varphi_2 \cdot r_2 \cdot (g - \lambda_i(t-1))\}; \quad (18)$$

the i^{th} particle position at the t^{th} iteration is:

$$\lambda_i(t) = \lambda_i(t-1) + v^i(t); \quad (19)$$

p^i is the position of the best particle at the i^{th} iteration; g is position of the best particle among all iterations; and r_1 and r_2 are the random numbers in the range of [0,1]. $\omega(t)$ is the inertia weight factor. φ_1 and φ_2 are acceleration constants. k is the constriction factor derived from the stability analysis of (18) for assuring the convergence [35]. By setting appropriate $\omega(t)$, k, φ_1 and φ_2 based on [36], the particle can be converged effectively toward p^i and g.

Algorithm IV.1: PSEUDO CODE(HOA)

Input $x_m(k), s_m(k)$
- $t \leftarrow 0$
- Generate $\lambda_i(t)$ randomly with $i = 1, 2, ..., N_s$.
- Evaluate $\lambda_i(t)$ based on equation (16).
- **while** (t is less than the predefined iteration)
  - **do**
    - Set $t \leftarrow t + 1$
    - Update all velocities $v^i(t)$ based on (18).
    - Generate all new particles $\lambda_i(t)$ based on (19).
    - Evaluate $\lambda_i(t)$ based on (16) with respect to $y_m(k)$ and $s_m(k)$.
- Set $\lambda^0 \leftarrow g$
- *// g is the best particle among all particles*
- Use simplex search method to locate the optimal solution λ^{opt} by using λ^0 as the initial solution.

return (λ^{opt})

After performing several iterations, the particles start moving within a small region on the search domain and the searching progress slows down. To speed up the process of locating the optimum, the swarm movement is terminated when it cannot locate a solution with good PESQ. Then the searching process is conducted by using the gradient-based method which is effective in locating the local optimum. The following best particle is used as the initial solution, λ^0, for the gradient-based method:

$$\lambda^0 = \max_{i \in [1..N_s]} (\lambda_i(T)) \quad (20)$$

After running the gradient-based method for some iterations, the optimal solution, λ^{opt}, can be obtained. The gradient-based method terminates when the objective function has a certain degree of decline such that:

$$\mid F_{\text{PESQ}}\{s_m(k), f_{\text{AGE}}(x_m(k), \lambda^{\text{opt}})\} - F_{\text{PESQ}}\{s_m(k), f_{\text{AGE}}(x_m(k), \lambda^0)\} \mid \leq \varepsilon_g \quad (21)$$

V. PERFORMANCE EVALUATIONS

A. Experimental Settings

The performance of the proposed decision-directed AGE was evaluated with five speech sequences voiced by ten speakers including seven males and three females. The five speech sequences consisted of the five Christmas carol titles: 'Jingle Bells', 'Santa Claus is Coming to Town', 'Sleigh Ride', 'Let It Snow', and 'Winter Wonderland'. Hence, 50 recorded speech sequences were included. These recorded speech sequences were assumed to be noise free, and were contaminated artificially with four noisy environments from four noise data files, namely white noise, factory noise, babble noise and volvo noise (or car noise) which were all collected from the NOISEX-92 database. They were contaminated by the four noise sources with different settings of SNRs ranging from -16 dB to 16 dB.

For the decision-directed AGE, the following parameters were used: $M = 16$ (given in (3)); $\alpha = 0.004$ (given in (5)); $\beta = 10^{-4}$ (given in (6)); $\gamma = 0.9$ (given in (8)); and a Kaiser window with a bandwidth of $1/16$ was used for the design of a bank of finite impulse response (FIR) filter. These parameters have been empirically determined to give the best performance possible for a wide range of input SNRs. For a fair comparison purpose, all the experiments have been conducted using these parameters.

For the hybrid optimization algorithm, the following parameters were used: N_s=50; φ_1=φ_2=2.05. The effectiveness of those parameters have been evaluated by solving a set of parametrical problems with different landscapes including multi-optimum, non-convex, discontinuous and undifferentiate functions [36]. Also, these parameters have been used on PSO for solving many real-world problems [36], [37] including optimal power flow, nonlinear electronic packaging, and neural network design. Also, satisfactory results have been obtained on solving those problems.

Based on those parameters, the tradeoff weight in the decision-directed AGE can be optimized with respect to the PESQ using the hybrid optimization algorithm. As the operational range for the decision-directed AGE is pre-defined to be between -16 dB and 16 dB, the optimal tradeoff weights were determined with respect to the SNRs of -16dB, -12dB, -8dB, -4dB, 0dB, 4dB, 8dB, 12dB and 16dB respectively.

The following two speech performance indices, namely segmental SNR measure (SNRseg) [28] and normalized signal distortion, were used to quantify the noise suppression, and the distortion to the speech source respectively.

The SNRseg measure, $\mathcal{S}$, is given as:

$$\mathcal{S} = \frac{10}{\mathcal{M}} \sum_{m=0}^{\mathcal{M}-1} 10 \log_{10} \frac{\|\mathbf{y}_m\|^2}{\|\mathbf{y}_m - \hat{\mathbf{y}}_m\|^2} \quad (22)$$

where $\mathbf{y}_m$ represents a clean speech frame (in time domain); $\hat{\mathbf{y}}_m$ is the enhanced speech frame; and $\mathcal{M}$ is the number of frames of the signal. In order to discard the non-speech frames, each frame had threshold of a -10 dB lower bound and 35 dB upper bound. Here the frame length of both $\mathbf{y}_m$ and $\hat{\mathbf{y}}_m$ is chosen to be 20 msecs.

The normalized signal distortion, $\mathcal{D}$ is given in decibels (dB) as

$$\mathcal{D} = 10 \log_{10} \left[\frac{1}{\mathcal{M}} \sum_{m=0}^{\mathcal{M}-1} |C\hat{P}_{\mathsf{out},s}(\omega_m) - \hat{P}_{\mathsf{in},s}(\omega_m)| \right] \quad (23)$$

where $\omega_m = 2\pi m/\mathcal{M}$, is the discretized and normalized frequency, and $\mathcal{M}$ is the number of FFT points. The normalizing constant, $\mathcal{C}$, is given as

$$\mathcal{C} = \frac{\sum_{m=0}^{\mathcal{M}-1} \hat{P}_{\mathsf{in},s}(\omega_m)}{\sum_{m=0}^{\mathcal{M}-1} \hat{P}_{\mathsf{out},s}(\omega_m)} \quad (24)$$

where $\hat{P}_{\mathsf{in},s}(\omega_m)$ is the spectral power estimate of the input signal and $\hat{P}_{\mathsf{out},s}(\omega_m)$ is the spectral power estimate of the output signal.

B. Experimental Results

The speech quality of the directed-decision AEG was indicated by the PESQ measure. Figure 2(a) shows the PESQ obtained by the original noisy speech signal corrupted with the white noise with the SNRs of -16, -12, -8, -4, 0, 4, 8, 12 and 16 respectively. It also shows the PESQ obtained by the three enhanced signals which were processed by the three approaches, conventional AGE [3], Wiener filter [22] and the proposed decision-directed AGE. It indicates that the PESQ obtained by the three enhanced signals is generally better than the PESQ of the original noisy speech signal. In general, the proposed decision-directed AGE can obtain better PESQ than those obtained by both the Wiener filter and the conventional AGE. Also, it can be observed that the proposed decision-directed AGE can obtain an improvement of 0.2 when the SNR is 16 dB. When the SNR is -16 dB, an improvement of 0.5 can be obtained by the proposed decision-directed AGE. For the factory noise, Figure 2(b) shows that more than 0.8 improvement in term of PESQ can be obtained by the proposed decision-directed AGE compared with the original noisy speech signal, when the SNR is 16 dB. When the SNR is -16dB, an improvement of 0.7 can be obtained.

When SNR=-16dB, the proposed decision-directed AGE obtains an improvement of 0.15 compared with the conventional AGE and Wiener filter. When the SNR is increasing, the proposed AGE can generally obtain better PESQ compared with the other two enhancement methods. At SNR=0dB, the proposed AGE can obtain an improvement of 0.25 compared the conventional AGE. Although the difference between the proposed AGE and the Wiener filter reduces when SNR increases, better PESQ can still be obtained by the proposed AGE compared with the conventional AGE. At SNR=16, the proposed AGE can obtain an improvement of 0.2 compared with the conventional AGE. In general, the proposed decision-directed AGE can obtain better PESQ than the other two enhancement approaches.

For both the babble noise and car noise, similar results can be found in Figure 2(c) and Figure 2(d) respectively which show that the proposed decision-directed AGE can produce improvement compared with both conventional AGE and Wiener filter. The proposed decision-directed AGE can generally obtain the best among the three enhancement approaches. Therefore, these results clearly indicate that better PESQ can be obtained by the proposed decision-directed AGE. This can be explained by the fact that the tradeoff weight of the proposed decision-directed AGE is optimized with respect to the PESQ in particular, while the conventional AGE is intended to boost just the speech signals and the Wiener filter is intended to suppress only the noise levels.

Table I shows the optimal tradeoff weights with respect to the tested noise and the considered SNRs. It indicates that there is no linear relationship between the tradeoff weight and the SNR. The PESQ of the directed-decision AEG cannot be improved by linearly adjusting the tradeoff weights, when the SNR reduces. Hence, the optimal tradeoff weights are necessary to be determined for different noisy conditions. Based on these optimal tradeoff weights, an expert system [38], [39] can be developed in order to perform a map between noisy conditions to tradeoff weight. When the expert system is incorporated with the proposed decision-directed AGE, the tradeoff weight can be adjusted automatically and better speech enhancement can be achieved under various SNR conditions.

Table I
OPTIMIZATION RESULT FOR λ

SNR (dB)	**-14**	**-10**	**-6**	**-2**	**0**	**2**	**6**	**10**	**14**
White noise	0.773	0.452	0.228	0.061	0.041	0.016	1.000	1.000	1.000
Factory noise	0.291	0.423	0.196	0.086	0.063	0.021	0.095	0.000	0.000
Babble noise	0.000	0.327	0.136	0.096	0.045	0.022	0.040	0.009	0.000
Car noise	0.000	0.482	0.205	0.126	0.091	0.063	0.029	0.007	0.000

C. Analytical Results

The speech signals processed by the decision-directed AGE were measured by the two speech quality criteria namely noise suppression level (segSNR) and normalized signal distortion. Figures 3(a)-3(d) show the noise suppression levels (segSNR) for the three approaches for the four types of noise with different SNRs. These results clearly shows that the proposed decision-directed AGE generally achieves a higher noise suppression compared to the conventional AGE across different SNRs. These figures indicate that the suppression capability of the conventional AGE is generally at the lower level of the proposed AGE.

The noise suppression improvement can be explained by Figure 4. It shows the evolution of the normalized gain function of the proposed decision-directed gain function at the 5-th sub-band (arbitrarily chosen) for the signal corrupted with car noise (SNR$=0$ dB) and varying values of λ. It shows clearly that the normalized gain function has a lower gain during speech silent periods. Hence, the gain function not only boosts the speech components but also imposes a lower gain on the noise components. As a result, more noise is being attenuated and consequently better noise suppression is obtained.

Figure 4 also indicates that the level of noise attenuation is controlled by the tradeoff parameter λ. For the case of $\lambda = 0.99$, the normalized gain function is close to zero during the detected non-speech periods, whereas for the case of $\lambda = 0.9$ and $\lambda = 0.6$, the noise floor is raised. When $\lambda = 0$, the gain function corresponds to the conventional AGE. The injection of noise to raise the noise floor can be mildly viewed as imposing a spectral floor [40]. However, this is contrary to the spectral floor method where the noise introduced is the estimated noise spectrum. Since the noise injected is the actual background noise with a reduced volume (see equation (15)), fewer unnatural sounding artifacts can intrude. The tradeoff parameter, λ, provides a user dependent tradeoff relationship for the amount of comfort noise injected against noise suppression. Therefore, better speech quality can be obtained by the proposed decision-directed AGE when the tradeoff parameter, λ, is optimized with respect to the PESQ.

Figures 5(a)-5(d) illustrate the normalized signal distortion for the three speech enhancement approaches. The results verify that the distortion levels for the proposed decision-directed AGE is in close proximity with the conventional AGE. These results suggest that the proposed decision-directed AGE achieves a better suppression with negligible expense of target signal integrity, and also verify that the parameter λ provides a variable tradeoff between distortion and suppression. This is because whilst more noise suppression can be achieved as λ is increased, the distortion level on the other hand increases. Therefore, the proposed decision-directed AGE can obtain the tradeoff between signal distortion and noise reduction but the conventional AGE can only boost the speech signal. It explains why the proposed decision-directed AGE can obtain better PESQ than the conventional AGE.

Figure 6 presents the output power plots for a) the original noisy speech signals, b) the enhanced signals of the AGE, and c) the enhanced signals of the proposed decision-directed AGE for all the noise types considered. The plots show that a lower noise floor is obtained with the proposed decision-directed AGE. They show that the noise floor is just being lowered and not altered. By doing so, the proposed decision-directed AGE avoids the deleterious musical noise, which produces an unnatural sounding output. Hence, better PESQ can be obtained by the proposed decision-directed AGE compared with the conventional AGE.

VI. Conclusions

This paper presented an effective decision-directed AGE intended to be implemented in assistive hearing instruments. It overcomes the limitation of the conventional AGE which is able to boost speech segments only in noisy speech but is not able suppress noise segments. Unlike the conventional AGE, the proposed decision-directed AGE can simultaneously boost speech segments and suppress noise segments. Also, it injects comfort noise to avoid the deleterious musical phenomenon and unnatural total silence. A trade-off weight, which can balance noise suppression and the signal distortion levels through the injection of comfort noise, is incorporated in the proposed decision-directed AGE in order to further enhance the speech quality. Hence, the functionality of assistive hearing instruments can be improved.

The effectiveness of the proposed direct-decision AGE was evaluated based on a commonly-used speech quality measure, namely PESQ, under various SNR conditions with four noisy environments: white noise, factory noise, babble noise and car noise. Experimental results showed that the proposed direct-decision AGE obtained better PESQ compared to the conventional AGE. Also, analytical results suggested that the improved performance was a result of better noise suppression, and negligible effect of both target signal distortion and musical tones which were generated by the proposed direct-decision AGE. Therefore, better speech quality can be produced by the assistive hearing instrument which is implemented with the proposed direct-decision AGE.

As environmental noise can affect the accuracy of the instruments and measurement of speech recognition functions, the proposed method can be further extended by reformulating the decision-directed AGE in order to optimize speech recognition accuracy. The resulting method is expected to improve the instruments and measurement of speech recognition functions [41], [42].

VII. Acknowledgement

The fourth author is supported by the RGC Grant PolyU. (5301/12E).

References

[1] M. S. M. G. Vlaming, B. Kollmeier, W. A. Dreschler, R. Martin, J. Wouters, B. Grover, Y. Mohammadh, and T. Houtgast, "Hearcom hearing in the communication society," *Acta Acustica United With Acustica*, vol. 97, no. 2, pp. 172–192, 2011.

[2] A. H. K. Parsi and M. Bouchard, "Instantaneous binaural target psd estimiation for hearing aid noise reduction in complex acoustic environments," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 4, pp. 1141–1154, 2011.

[3] N. Westerlund, M. Dahl, and I. Claesson, "Speech enhancement for personal communication using an adaptive gain equalizer," *Signal Processing*, vol. 85, no. 6, pp. 1089–1101, June 2005.

[4] B. Sllberg, N. Grbic, and I. Claesson, *Implementation aspects of the adaptive gain equalizer, Research report No. 2006:04.* Boca Raton, FL: Blekinge Institute of Technology, 2006.

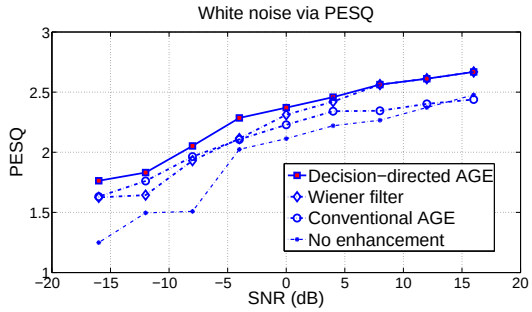
[5] N. Westerlund, M. Dahl, and I. Claesson, "Real-time implementation of an adaptive gain equalizer for speech enhancement purposes," *WSEAS Transactions on Computers*, vol. 3, no. 1, pp. 25–32, 2004.

[6] ——, "Speech enhancement using an adaptive gain equalizer with frequency dependent parameter settings," in *Proceedings of IEEE 60th Vehicular Technology Conference*, Los Angeles, 2004, pp. 3718–3722.

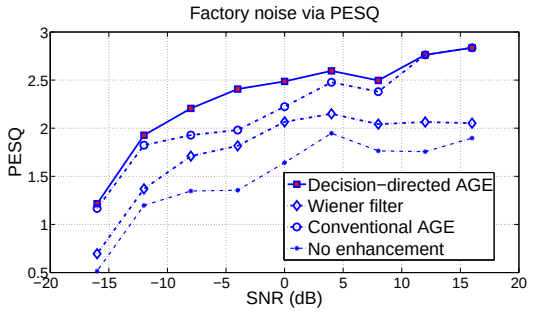
[7] J. P. Dmochowski and R. A. Goubran, "Decoupled beamforming and noise cancellation," *IEEE Transactions on Instrumentation and Measurement*, vol. 56, no. 1, pp. 80–88, 2007.

[8] U. Sauvagerd, "A ten-channel equalizer for digital audio-applications," *IEEE Transactions on Circuits and Systems*, vol. 36, no. 2, pp. 276–280, 1989.

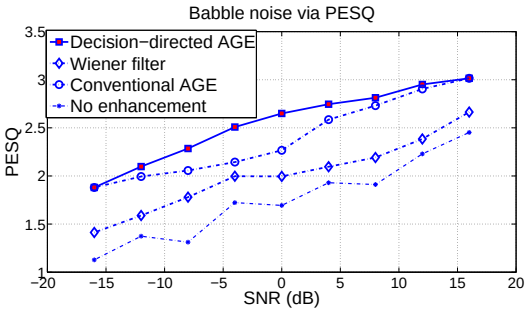
[9] S. M. Kuo, W. S. Gan, and F. L. Shau, "Design and synthesis of the audio equalizers," in *International Conference on Signal Processing*, vol. 1, 2004, pp. 579–582.
[10] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 32, no. 6, pp. 1109–1121, 1984.
[11] I.-T. Recommendation, Ed., *Perceptual evaluation of speech quality (PESQ): an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs.* Geneva, Switzerland, 2001.
[12] J. Ma, Y. Hu, and P. Loizou, "Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions," *Journal of the Acoustical Society of America*, vol. 125, no. 5, pp. 3387–3405, 2009.
[13] P. Paglierani and D. Petri, "Uncertainty evaluation of objective speech quality measurement in voip systems," *IEEE Transactions on Instrumentation and Measurement*, vol. 58, no. 1, pp. 46–51, 2009.
[14] Y. Shi, "Empirical study of particle swarm optimization," in *Proceedings of the 1999 Congress on Evolutionary Computation*, Washington, U.S., 1999, pp. 1945–1950.
[15] M. Clerc and J. Kennedy, "The particle swarm - explosion, stability, and convergence in a multidimensional complex space," *IEEE Transactions on Evolutionary Computation*, vol. 6, no. 1, pp. 58–73, 2002.
[16] K. Parsopoulos and M. Vrahatis, "On the computation of all global minimizers through particle swarm optimization," *IEEE Transactions on Evolutionary Computation*, vol. 8, no. 3, pp. 211–224, 2004.
[17] N. V. George and G. Panda, "A particle-swarm-optimization-based decentralized nonlinear active noise control system," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 12, pp. 3378–3386, 2012.
[18] S. T. Pan, "Evolutionary computation on programmable robust iir filter pole placement design," *IEEE Transactions on Instrumentation and Measurement*, vol. 60, no. 4, pp. 1469–1479, 2011.
[19] J. B. V. Reddy, P. K. Dash, R. Samantaray, and A. K. Moharana, "Fast tracking of power quality disturbance signals using an optimized unscented filter," *IEEE Transactions on Instrumentation and Measurement*, vol. 58, no. 12, pp. 3943–3952, 2009.
[20] N. K. Rout, D. P. Das, and G. Panda, "Particle swarm optimization based active noise control algorithm without secondary parth identification," *IEEE Transactions on Instrumentation and Measurement*, vol. 61, no. 2, pp. 554–563, 2012.
[21] R. Fletcher, Ed., *Practical Methods of Optimization*. Boca Raton, FL: Wiley, 2000.
[22] B. Widrow and S. D. Stearns, Eds., *Adaptive Signal Processing*. Englewood NJ: Prentice-Hall, 1985.
[23] O. Cappé, "Elimination of the musical noise phenomenon with the Ephraim and Malah noise supressor," *IEEE Transactions on Signal Processing*, vol. 2, no. 2, pp. 345–349, April 1994.
[24] S. Kanehara, H. Saruwatari, R. Miyazaki, K. Shikano, and K. Kondo, "Theoretical analysis of musical noise generation in noise reduction methods with decision-directed a priori snr estimator," in *Proceedings of IEEE International Workshop on Acoustic Signal Enhancement*, Aachen, Germany, 2012, pp. 1–4.
[25] R. Miyazaki, H. Saruwatari, T. Inoue, K. Shikano, and K. Kondo, "Musical-noise-free speech enhancement: Theory and evaluation," in *Proceedings of IEEE International Conference on Acoustic, Speech and Signal Processing*, 2012, pp. 4565–4568.
[26] C. Plapous, C. Marro, and P. Scalart, "Improved signal-to-noise ratio estimation for speech enhancement," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 6, pp. 2098–2108, 2006.
[27] P. Scalart and J. V. Filho, "Speech enhancement based on *a priori* signal to noise estimation," *IEEE International Conference on Acoustics, Speech and Singal Processing*, vol. 2, pp. 629–632, 1996.
[28] P. Loizou, *Speech Enhancement Theory and Practice*. Boca Raton, FL: CRC Press, 2007.
[29] Y. Hu and P. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Transactions on Audio, Speech, and Langange Processing*, vol. 16, no. 1, pp. 229–238, 2008.
[30] L. Ding, A. Radwan, M. S. Hennawey, and R. A. Goubran, "Measurement of the effects of temporal clipping on speech quality," *IEEE Transactions on Instrumentation and Measurement*, vol. 55, no. 4, pp. 1197–1203, 2006.
[31] A. Rix, J. Beerends, M. Hollier, and A. Hekstra, "Perceptual evaluation of speech quality (pesq) - a new method for speech quality assessment of telephone networks and codecs," in *Proceedings of IEEE International Conference on Acoustic, Speech, and Signal Processing*, vol. 2, Athens, Ohio, 2001, pp. 749–752.
[32] K. Chan, S. Nordholm, K. Yiu, and R. Togneri, "Speech enhancement strategy for speech recognition microcontroller under noisy environments," *Neurocomputing*, vol. 118, no. 22, p. 279288, 2013.
[33] K. Chan, S. Nordholm, and K. Yiu, "Multichannel filters for speech recognition using a particle swarm optimization," 2012, pp. 937–942.
[34] K. Chan, K. Yiu, T. Dillon, S. Nordholm, and S. Ling, "Enhancement of speech recognitions for control automation using an intelligent particle swarm optimization," *IEEE Transactions on Industrial Informatics*, vol. 8, no. 4, pp. 869–879, 2012.
[35] R. C. Eberhart and Y. Shi, "Comparing inertia weights and constriction factors in particle swarm optimization," in *Proceedings of IEEE Congress on Evolutionary Computation*, vol. 1, 2000, pp. 84–88.
[36] S. H. Ling, H. H. C. Iu, K. Y. Chan, H. K. Lam, C. W. Yeung, and F. H. F. Leung, "Hybrid particle swarm optimization with wavelet mutation and its industrial applications," *IEEE Transactions on Systems, Man and Cybernetic - Part B*, vol. 38, no. 3, pp. 743–763, 2008.
[37] N. Mo, Z. Zou, K. Chan, and T. Pong, "Transient stability constrained optimal power flow using particle swarm optimisation," *IET Proceedings on Generation, Transmission and Distribution*, vol. 1, no. 3, pp. 476–483, 2006.
[38] I. Tashev and M. Slaney, "Data driven suppression rule for speech enhancement," in *Proceedings of Information Theory and Applications Workshop*, 2013.
[39] S. Srinivasan, J. Samuelsson, and B. Kleijn, "Codebook driven short-term predictor parameter estimation for speech enhancement," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 1, pp. 163–176, 2006.
[40] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," in *IEEE International Conference on Acoustic, Speech, and Signal Processing*, Cambridge, MA, 1979, pp. 208–211.
[41] S. Pan and X. Li, "An fpga-based embedded robust speech recognition system designed by combining empirical mode decomposition and a genetic algorithm," *IEEE Transactions on Instrumentation and measurement*, vol. 61, no. 9, pp. 2560–2572, 2012.
[42] Y. Zhan, H. Leung, K. Kwak, and H. Yoon, "Automated speaker recognition for home service robots using genetic algorithm and dempstershafer fusion technique," *IEEE Transactions on Instrumentation and measurement*, vol. 58, no. 9, pp. 3058–3068, 2009.



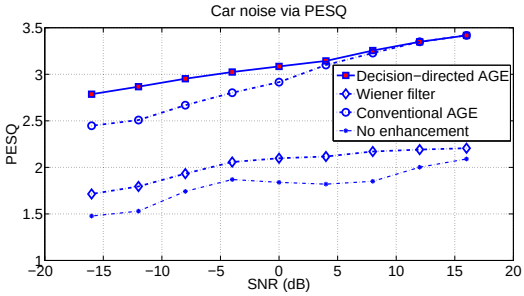
(a) White noise via PESQ



(b) Factory noise via PESQ

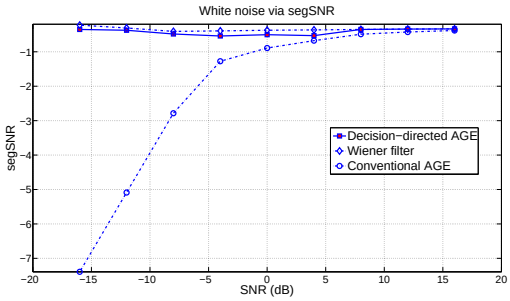


(c) Babble noise via PESQ

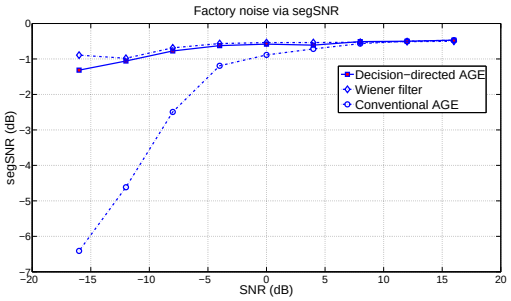


(d) Car noise via PESQ

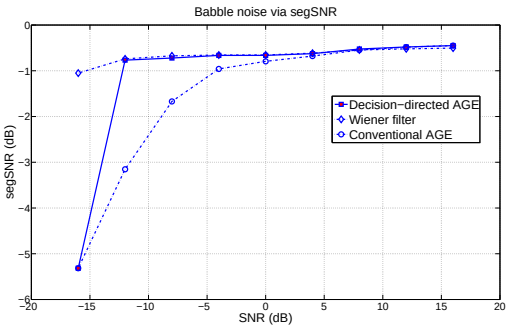
Figure 2. Results for the PESQ (a) White noise (b) Factory noise (c) Babble noise and (d) Car noise



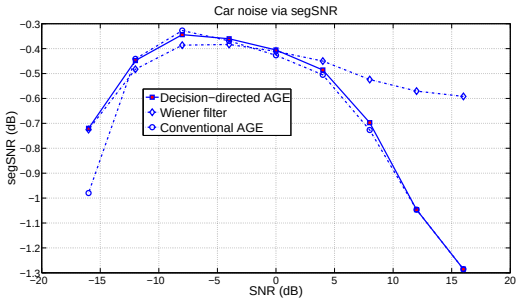
(a) White noise via segSNR



(b) Factory noise via segSNR



(c) Babble noise via segSNR



(d) Car noise via segSNR

Figure 3. Measurements for the segSNR (a) White noise (b) Factory noise (c) Babble noise and (d) Car noise

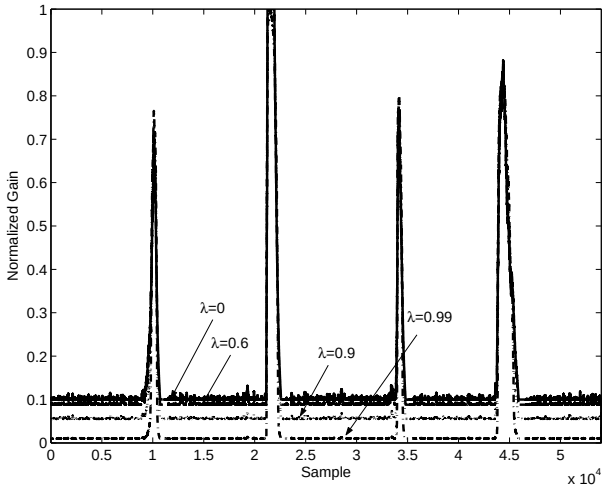
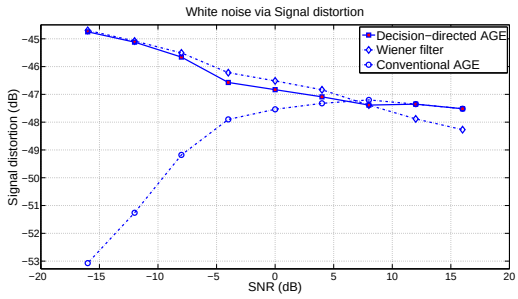
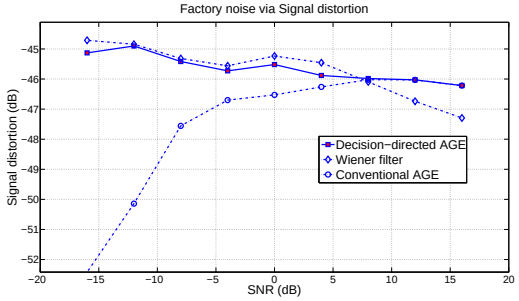


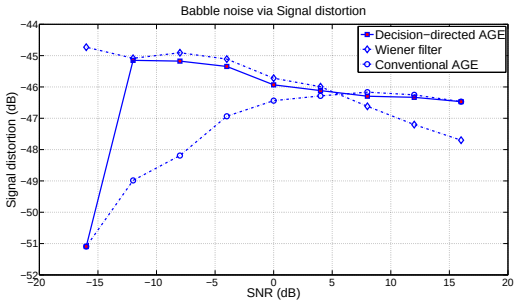
Figure 4. The normalized gain function for the conventional AGE, i.e., $\lambda = 0$ and the proposed decision-directed AGE for $\lambda = 0.6$, $\lambda = 0.9$ and $\lambda = 0.99$.



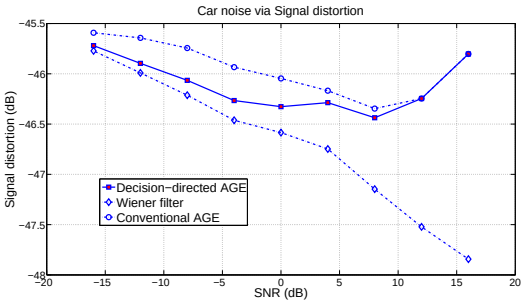
(a) White noise via normalized signal distortion



(b) Factory noise via normalized signal distortion



(c) Babble noise via normalized signal distortion



(d) Car noise via normalized signal distortion

Figure 5. Measurements for the normalized signal distortion (a) White noise (b) Factory noise (c) Babble noise and (d) Car noise

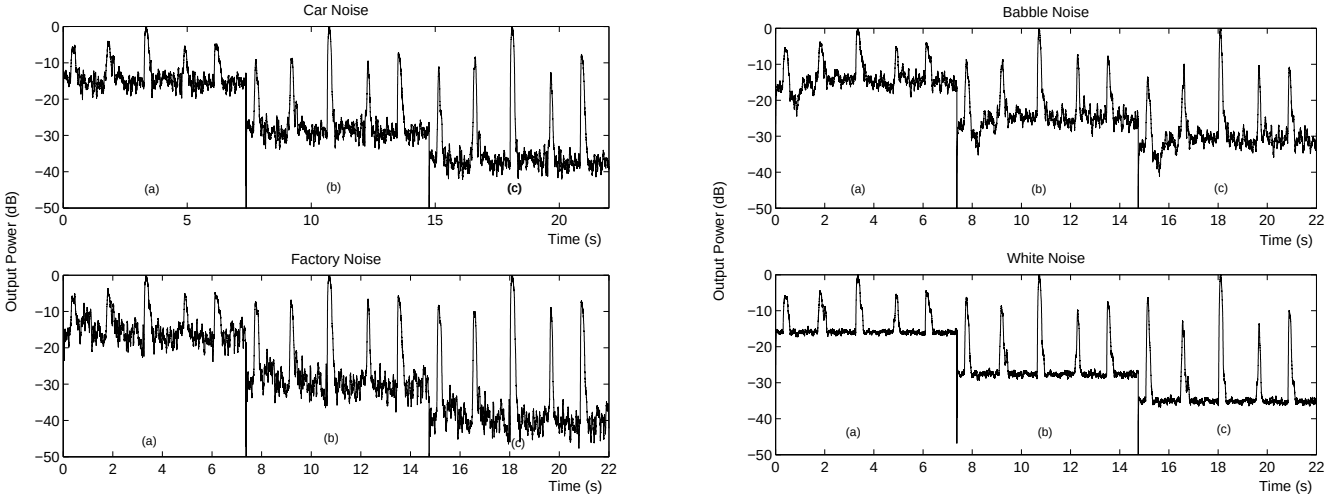


Figure 6. The output power plots for (a) corrupted observation; (b) the output from the conventional AGE; and (c) the output from the proposed directed-decision AGE for the four types of noise with SNR=0 dB.