RESEARCH　　　　　　　　　　　　　　　　　　　　　　　　**Open Access**

An improved adaptive gain equalizer for noise reduction with low speech distortion

Markus Borgh[1*], Magnus Berggren[2], Christian Schüldt[2], Fredric Lindström[1] and Ingvar Claesson[2]

Abstract

In high-quality conferencing systems, it is desired to perform noise reduction with as limited speech distortion as possible. Previous work, based on time varying amplification controlled by signal-to-noise ratio estimation in different frequency subbands, has shown promising results in this regard but can suffer from problems in situations with intense continuous speech. Further, the amount of noise reduction cannot exceed a certain level in order to avoid artifacts. This paper establishes the problems and proposes several improvements. The improved algorithm is evaluated with several different noise characteristics, and the results show that the algorithm provides even less speech distortion, better performance in a multi-speaker environment and improved noise suppression when speech is absent compared with previous work.

Keywords: speech enhancement, noise reduction, noise-level estimation

1 Introduction

When communicating using hands free devices such as speakerphones, the speech signal is typically corrupted by background noise such as ventilation noise or computer fan noise. One commonly used method for reducing this type of noise is spectral subtraction [1,2]. Although typically achieving well in terms of noise reduction, the basic spectral subtraction algorithm has often the effect that musical noise appears due to spectral flooring [3]. Ways of reducing the musical noise has been proposed by e.g. Ephraim and Malah [4], although this method still tends to give audible artifacts which could in some cases even result in reduced listening comfort compared to the original unprocessed signal [5]. Further improvements have been made by Plapous et al. [6] in which they introduce a two-step noise reduction technique that reduces the noise without adding artifacts to the speech signal. However, this algorithm aims at reducing speech harmonics distortion and does nothing for the unvoiced speech.

A time domain speech enhancement ("booster") algorithm, in this paper denoted the *speech booster algorithm* (SBA), has been proposed by Westerlund et al. [7] in which the audio signal is amplified according to a signal-to-noise ratio (SNR) estimate in subbands. The gain is calculated for a subband divided signal, and the gains in each subband are independent of each other. Advantages of SBA are the low computational complexity compared to other algorithms with similar amount of speech enhancement [8] as well as the ease of implementation and the absence of musical noise if the gains are controlled with care [7].

However, SBA suffers from a massive drawback which manifests itself in situations with intense continuous speech. In this type of situations, the subband SNR estimates will gradually become inaccurate, resulting in undesired damping and ultimately reduced speech signal quality.

This paper demonstrates the drawback and proposes a modification to avoid this drawback. Further, the paper presents additional improvements in the form of a gain modified to produce less speech distortion and to provide more noise damping in speech pauses.

The outline of the paper is as follows. In Section 2, the original SBA presented in [7] is described, and in Section 3, the proposed improvements are presented. Section 4 describes the simulation setup used for comparing the original SBA to the proposed method and Section 5 presents the results. Section 6 compares the SBA and the proposed method using objective speech distortion and SNR increase measures during speech. A short comment on

* Correspondence: markus.borgh@limesaudio.com
[1]Limes Audio AB, Box 7961, 90719 Umeå, Sweden
Full list of author information is available at the end of the article

subjective evaluation is presented in Section 7, followed by the conclusions in Section 8.

2 The speech booster algorithm

The noisy speech is denoted $x(n)$, where n is the sample index, and is assumed to consist of the desired speech signal $s(n)$ and additive noise $v(n)$

$$x(n) = s(n) + v(n). \qquad (1)$$

A filterbank consisting of K bandpass filters is used to divide the input signal $x(n)$ into K subband signals, each denoted $x_k(n)$ where $k \in [0, K - 1]$. The output signal is then formed by weighting and summation of the subband signals according to

$$y(n) = \sum_{k=0}^{K-1} g_{1,k}(n)x_k(n), \qquad (2)$$

where $g_{1,k}(n)$ is the subband gain based on estimation of the SNR in subband k. Calculation of the subband gain is performed as

$$g_{1,k}(n) = min\left\{ \left(\frac{A_k(n)}{B_k(n)} \right)^{p_k}, L_k \right\}, \qquad (3)$$

where $A_k(n)$ is an estimate of the noisy speech signal level, $B_k(n)$ is an estimate of the noise level, L_k is a threshold determining the maximum allowed gain in subband k and $p_k \geq 0$ is a constant denoted the gain rise exponent [7].

The noisy speech level is estimated by taking a short-time average of the input signal according to

$$A_k(n) = \alpha_k A_k(n - 1) + (1 - \alpha_k) |x_k(n)|, \qquad (4)$$

where $0 \leq \alpha_k \leq 1$ is a forgetting factor constant.

Estimation of the noise level is based on the short-time average $A_k(n)$ as

$$B_k(n) = \begin{cases} A_k(n) & \text{if } A_k(n) \leq B_k(n - 1), \\ (1 + \beta_k)B_k(n - 1) & \text{otherwise,} \end{cases} \qquad (5)$$

where β_k is a positive constant defining the increase rate of the noise level.

3 The proposed method

One problem with the SBA as described in the previous section is the noise-level estimation in (5). During intense continuous speech, the noise-level estimate $B_k(n)$ will increase and cause reduction of the speech boosting gain, see (3).

To overcome this problem, an alternative noise estimation method is proposed. The proposed noise estimator utilizes a modified update scheme according to

$$B_k(n) = \begin{cases} A_k(n) & \text{if } A_k(n) \leq B_k(n - 1), \\ B_k(n - 1) & \text{if } A_k(n) > B_k(n - 1), \\ & \text{and } \phi(n) = 1 \\ (1 + \beta_k)B_k(n - 1) & \text{otherwise,} \end{cases} \qquad (6)$$

where $\varphi(n)$ is an update controller, which can take on the values 1 (no update) or 0 (update). Use of the noise estimation update controller $\varphi(n)$ prevents noise estimation during speech and thus eliminates the problem of speech boosting gain reduction during intense continuous speech.

The noise estimation update controller is defined as

$$\phi(n) = \begin{cases} 1 \text{ if } S_k(n) \geq T_{\phi,k} \text{ for any } k \\ 0 \text{ otherwise} \end{cases} \qquad (7)$$

where $T_{\varphi,k}$ is a threshold and $S_k(n)$ is the ratio between the maximum and minimum signal magnitudes in accumulated blocks defined as

$$S_k(n) = \frac{\max\limits_{q\in\{0,\ldots,N_b-1\}} F_k(l - q)}{\delta + \min\limits_{q\in\{0,\ldots,N_b-1\}} F_k(l - q)}. \qquad (8)$$

In (8), N_b is the number of blocks, used for the estimation of $S_k(n)$, $0 < \delta \ll 1$ is a constant included for avoiding division by zero, and $F_k(l)$ is the accumulated signal block

$$F_k(l) = \sum_{i=0}^{N_s-1} |x_k(lN_s - i)|, \qquad (9)$$

where N_s is the number of samples accumulated in every block. The block index $l \in \mathbb{Z}$ fulfills

$$lN_s \equiv n. \qquad (10)$$

The essence of (8) is to compare the largest accumulated signal block (numerator) with the smallest block (denominator), out of the N_b most recent (in time) blocks. A high ratio $S_k(n)$ indicates that the signal $x_k(n)$ currently could be regarded as non-stationary under the considered time-frame, meaning in this context that the current signal content is likely to be dominated by speech. A low ratio $S_k(n)$ on the other hand means that the signal $x_k(n)$ is likely to be dominated by stationary (still under the considered time-frame) noise. The noise estimation update controller (7) then allows noise estimation once $S_k(n)$ is below the threshold $T_{\varphi,k}$ for all k.

A second problem with the original SBA is that if L_k is set too high, there is a risk of fast pumping of the noise and distortion of the speech [7]. To avoid this while still providing significant reduction of the noise in speech pauses, a second gain factor is proposed. This gain factor denoted the fullband gain, $g_2(n)$, only provides damping,

i.e. reduces the noise, in longer speech pauses and is applied to the input signal as

$$\gamma(n) = g_2(n) \sum_{k=0}^{K-1} g_{1,k}(n) x_k(n). \tag{11}$$

The proposed fullband gain is based on a gain controller $\phi_0(n)$, which is defined as

$$\varphi_0(n) = \begin{cases} 1 \text{ if } \dfrac{1}{K} \sum_{k=1}^{K} g_{1,k}(n) \geq T_\varphi \\ 0 \text{ otherwise} \end{cases} \tag{12}$$

where T_ϕ is a threshold. Further, to avoid changes in $g_2(n)$ during short speech pauses a hold function of n_h samples is introduced for the gain controller $\phi(n)$ which then becomes

$$\varphi(n) = \max_{q \in \{0, \dots, n_h - 1\}} \varphi_0(n - q). \tag{13}$$

The fullband gain is expressed as

$$g_2(n) = \lambda(n) g_2(n-1) + (1 - \lambda(n)) L(n) \tag{14}$$

where $\lambda(n)$ is the forgetting factor and $L(n)$ is the target damping value. The speech pause-driven gain $g_2(n)$ is designed to quickly adapt to a certain value L_f with smoothing parameter λ_f and adapt slowly to the level L_s $<L_f$ with a smoothing parameter $\lambda_s > \lambda_f$. The shift between these regions is decided with

$$L(n) = \begin{cases} 1 & \text{if } \varphi(n) = 1 \\ L_f & \text{if } \varphi(n) = 0 \\ & \text{and } g_2(n-1) > L_f(1 + \Delta) \\ L_s & \text{otherwise} \end{cases} \tag{15}$$

and

$$\lambda(n) = \begin{cases} 0 & \text{if } \varphi(n) = 1 \\ \lambda_f & \text{if } \varphi(n) = 0 \\ & \text{and } g_2(n-1) > L_f(1 + \Delta) \\ \lambda_s & \text{otherwise} \end{cases} \tag{16}$$

where Δ is a small positive constant defining the limit of transition between the regions of fast and slow damping.

As can be seen in (12), the proposed fullband gain directly depends on the subband gains $g_{1,k}$; if sufficient gain is applied in the subbands (during speech), the gain controller $\phi(n)$ will be 1, indicating that the fullband gain should rise, see (15) and (14). On the other hand, if little subband gain is applied (when only stationary noise is present), the gain controller $\phi(n)$ will be 0, indicating that the fullband gain should fall, see (16) and (14).

The fullband gain $g_2(n)$ could be said to consist of three regions. The first region, $L(n) = 1$, is used when

speech is present. The second region, $L(n) = L_f$, is used directly after a speech segment in the audio signal. In this region, the gain is quickly reduced, which reduces the noise that is no longer masked by the speech. Since the adaption to the lowest gain in this region is relatively fast, the amount of noise suppression cannot be too large since that would give a non-comfortable sounding alteration of the noise level. Instead, the third region, $L(n) = L_s$, is used to adapt to the lowest desired gain. This adaption is fairly slow in order to make the transition between the noise levels less apparent.

Further, instead of the full-rate filterbank structure used in [7], it is proposed to use a polyphase filterbank with downsampling [9] to provide reduction in computational complexity. In this paper, a decimation rate of 32 was used. For detailed information about polyphase filterbanks, the reader is referred to [9,10] and the references therein.

4 Simulation setup

To compare the performance of the SBA and the proposed algorithm, several simulations were conducted. The audio signals used in the evaluation were speech signals consisting of recorded speech and a noise signal consisting of recorded ventilation noise. All signals were sampled with 16-kHz sampling frequency. Evaluation was performed with different SNRs, which was achieved by varying the noise level through multiplication with a noise gain factor η_v as

$$v(n) = \eta_v w(n) \tag{17}$$

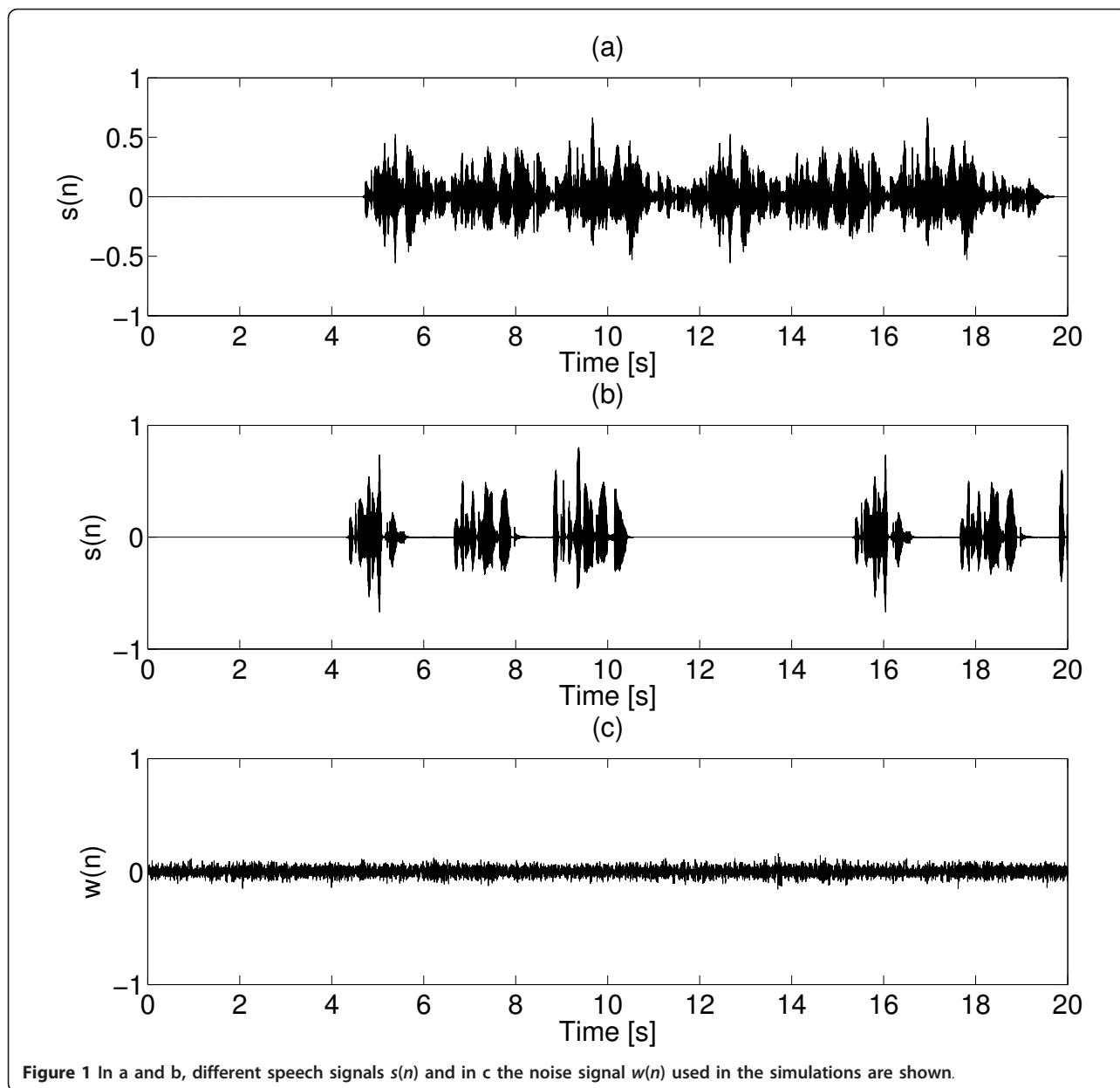where $w(n)$ is the ventilation noise signal. The signal $w(n)$ is shown in Figure 1 along with both versions of speech signal $s(n)$.

4.1 Common parameter setup

In this section, the setup of the parameters used by both the SBA and the proposed algorithm is discussed. It should be noted that the same parameter settings were used for both algorithms when possible in the simulations.

To avoid artifacts such as musical noise, the difference in gain between two separate subbands cannot be too large. On the other hand, the larger the allowed difference–the more noise reduction is achieved. A suitable choice of maximum subband gain is in the region $10 \leq | 20 \log_{10} L_k| \leq 25$ dB [7].

The forgetting factor α_k is chosen so that the gain $g_{1,k}(n)$ will be stable and less affected by impulsive noises compared to a lower setting of α_k. Westerlund et al. recommend a lower setting of α_k but also mention that tweaking this parameter could lead to improved performance depending on the noise environment.

Figure 1 In a and b, different speech signals *s*(*n*) and in c the noise signal *w*(*n*) used in the simulations are shown.

Further, the relationship between the SNR estimate $\frac{A_k(n)}{B_k(n)}$ and the subband gain $g_{1,k}(n)$ is decided by the gain rise exponent p_k, see (3). If a linear relationship is desired, then $p_k = 1$ and if $p_k > 1$, an alteration of the SNR estimate will have a larger effect on the gain than if $p_k < 1$. For the simulations, a setting of $p_k = 1$ was chosen.

4.2 Parameter setup for the proposed algorithm

The proposed algorithm contains a number of additional parameters that should be tuned. In this section, the setup of the additional parameters is discussed.

As described in Section 3, the proposed algorithm incorporates a fullband gain $g_2(n)$, which has the purpose of damping noise in longer speech pauses. The gain limitation L_f describes the first damping limit of g_2 (n). If this is too large, there is a risk of rapid noise pumping. The last gain limitation parameter L_s should be set according to the desired maximum total noise damping $|20 \log_{10} (L_k L_s)|$ dB.

The setup of the gain controller $\phi(n)$ was done by adjusting the parameters T_ϕ and n_h. The hold time parameter n_h is to be altered depending on how fast the additional noise damping $g_2(n)$ should start to affect the signal. A short hold time would imply noticeable additional noise reduction in short speech pauses but could on the other hand cause annoying pumping of the noise

level. A longer hold time lessens this noise level pumping effect, but would not cause any noticeable additional noise damping in short speech pauses. Further, the threshold T_ϕ should be set with the maximum allowed subband noise damping L_k in mind. The threshold should be $T_\phi > L_k$ for the controller to be able to deactivate. A recommended threshold setting is $T_\phi \approx 2L_k$. For the simulations in this paper, the setting $T_\phi = 0.5$ was used.

The setup of the noise estimation update controller $\varphi(n)$ was done by adjusting the parameters $T_{\varphi,k}$, N_b and N_s. The controller makes a decision based on the previous $N_b N_s$ samples, which implies that by adjusting these parameters, the behavior of $\varphi(n)$ is greatly affected. The threshold $T_{\varphi,k}$ marks the decision point for distinguishing between speech and noise. If $|20 \log_{10} T_{\varphi,k}| = 10$, the ratio between the largest and smallest signal block has to be at least 10 dB for the noise estimation to halt. This is the setting used in the simulations. Moreover, the smoothing parameter β_k was adjusted so that the adaption to an increased noise level would be approximately 2 dB/s for both the SBA and the proposed algorithm. This corresponds to $\beta_k = 2.8 \times 10^{-4}$ for the SBA and $\beta_k = 6.3 \times 10^{-4}$ for the proposed algorithm.

5 Behavior of the algorithm
In this section, the two main advantages of the proposed algorithm over the original SBA are demonstrated. The parameter values used are listed in Table 1.

5.1 Estimation of the noise level
In Figure 2, the subband gain $g_{1,k}(n)$ in one subband ($k = 1$) (plot a) and the corresponding level estimates $A_k(n)$ and $B_k(n)$ (plot b) are shown for an input signal containing both noise ($\eta_v = 1$, SNR $\approx$ 3dB) and continuous speech. The speech signal consists of multiple speakers overlapping, a situation which frequently occurs in a normal discussion with a large number of participants. The noise

Table 1 Parameter values used in simulations

Parameter	Value
$L_k, \forall k$	0.25
L_f	0.5
L_s	0.125
Δ	0.05
$\alpha_k, \forall k$	0.984
λ_f	0.9687
λ_s	0.999
$p_k, \forall k$	1
N_b	64
N_s	8
n_h	100
δ	2.2×10^{-16}

estimation approach in the SBA and the proposed method are compared. For the SBA, the noise-level estimate gradually rises during the speech segments of the audio signals. This causes the subband gain $g_{1,k}(n)$, shown in Figure 2 plot a (dashed), to decrease during longer speech segments since the SNR estimate will be lower than the actual SNR. It is clear that the original SBA suffers from problems in this case, whereas the proposed solution does not. For the proposed solution, displayed in Figure 2 plot b (dotted), the update controller $\varphi(n)$ activates during the speech segment of the displayed signal. This produces a stable noise-level estimate during the speech segment and thus a more correct subband gain is applied. It should be noted that the difference in subband gain between the proposed solution and the SBA is sometimes as large as 10 dB, which is a highly audible difference.

In Figure 3, the subband gain $g_{1,k}(n)$ in one subband ($k = 1$) and the corresponding level estimates $A_k(n)$ and $B_k(n)$ are shown for an input signal containing only noise ($\eta_v = 1$), with a sudden noise level increase ($\eta_v = 3$) after 20 s. It can be seen that the performance of the proposed algorithm is similar to that of the original SBA.

Thus, by using an update controller, the noise-level estimation performance is improved. With a suitable choice of $T_{\varphi,k}$, the noise estimation update controller $\varphi(n)$ becomes active during speech segments while still being able to adapt to changing noise levels. Without the proposed update controller, i.e. the SBA, the noise-level estimation will over time rise to a higher level than the actual background noise level. The only way of reducing this effect would be to decrease the value of β_k, but this would in turn also result in slower adaption to an increased noise level.

Further, one important property of the update controller $\varphi(n)$ is that it should never fail to activate when speech is present. In this case, it is better to halt the update too often than too seldom. A faulty update causes the estimated noise level to increase during speech which in the long term could cause a noise-level estimation $B_k(n)$ as high as the actual speech level $A_k(n)$, as discussed previously and shown in Figure 3 for the SBA.
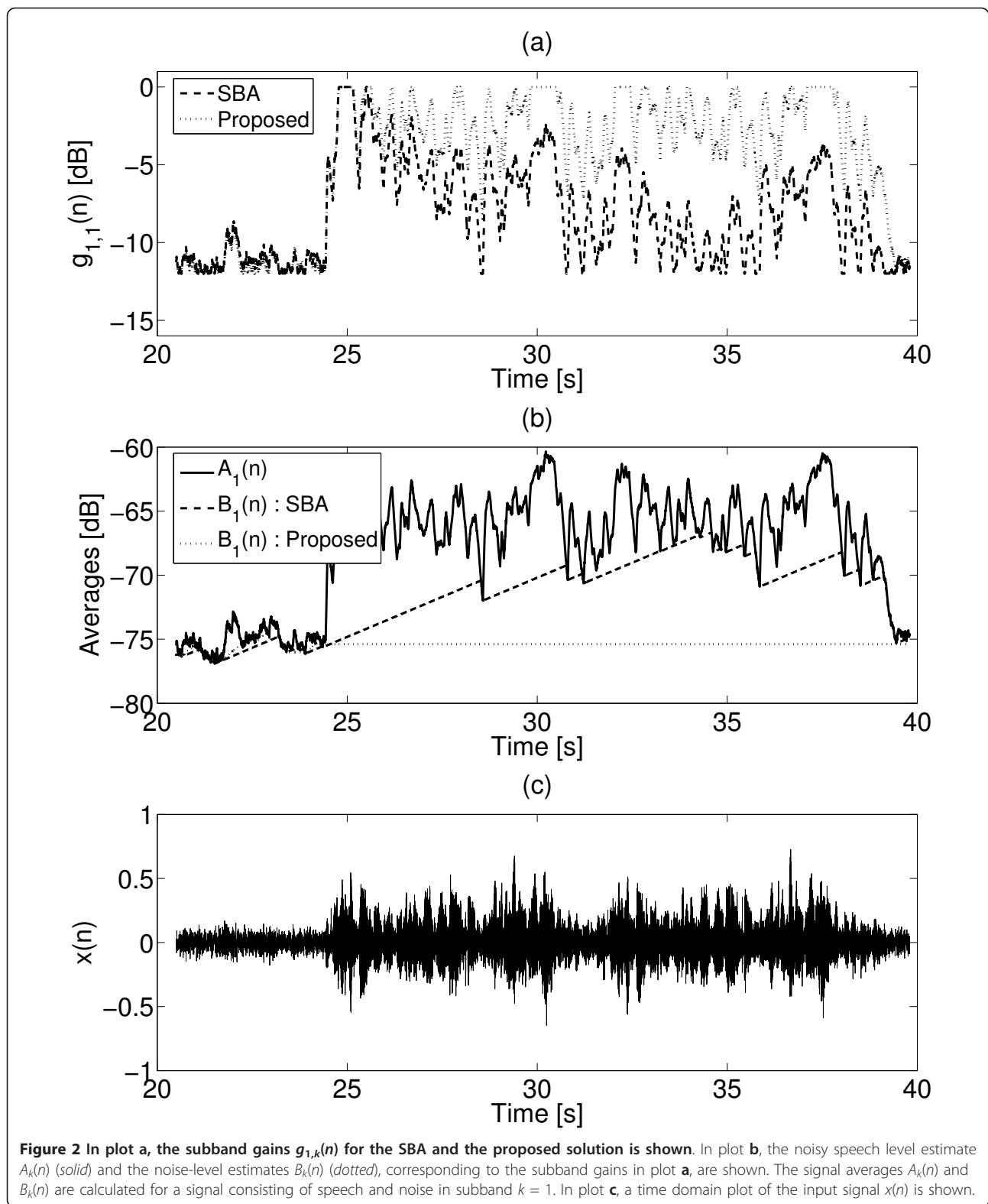
5.2 Noise damping in longer speech pauses
In Figure 4, the effect of the proposed algorithm on a noisy speech signal (Figure 1 plot b and $\eta_v = 1$, SNR $\approx$ 4 dB) is shown for the SBA and the proposed algorithm. The total subband gain $G_k(n)$, defined as $G_k(n) = g_{1,k}(n)$ in the SBA case and $G_k(n) = g_{1,k}(n) g_2(n)$ for the proposed algorithm, is plotted along with the resulting output signals in a specific subband ($k = 1$). From Figure 4 plot a, it can be seen that for the proposed algorithm, the noise is reduced with as much as 27 dB after 26 s.

Thus, the inclusion of the proposed additional gain $g_2(n)$ leads to a reduced noise level during speech pauses,

Figure 2 In plot a, the subband gains $g_{1,k}(n)$ for the SBA and the proposed solution is shown. In plot **b**, the noisy speech level estimate $A_k(n)$ (*solid*) and the noise-level estimates $B_k(n)$ (*dotted*), corresponding to the subband gains in plot **a**, are shown. The signal averages $A_k(n)$ and $B_k(n)$ are calculated for a signal consisting of speech and noise in subband $k = 1$. In plot **c**, a time domain plot of the input signal $x(n)$ is shown.

without affecting the quality of the speech. The additional gain will cause no speech distortion as the gain is constant (with value $g_2(n) = 1$) during speech. Further,

it does not change the spectral characteristics of the noise since all subbands are equally attenuated and the damping is changing slowly. The damping level L_s can
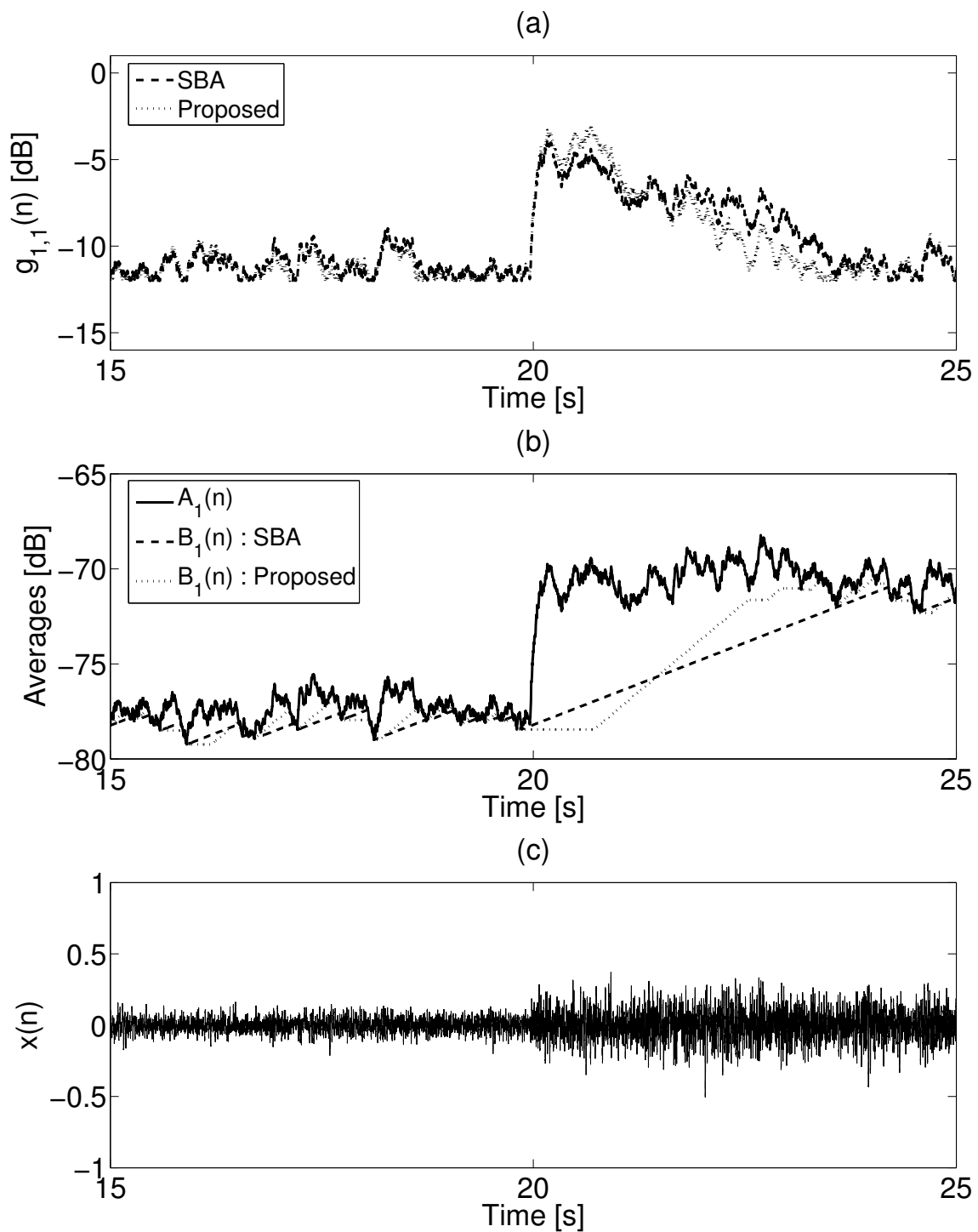
Figure 3 In plot a, the subband gains $g_{1,k}(n)$ for the SBA and the proposed solution is shown. In plot **b**, the noisy speech level estimate $A_k(n)$ (*solid*) and the noise-level estimates $B_k(n)$ (*dotted*), corresponding to the subband gains in plot **a**, are shown. The signal averages $A_k(n)$ and $B_k(n)$ are calculated for a signal consisting of only noise in subband $k = 1$. A sudden increase in the actual noise level takes place after 20 s. In plot **c**, a time domain plot of the input signal $x(n)$ is shown.
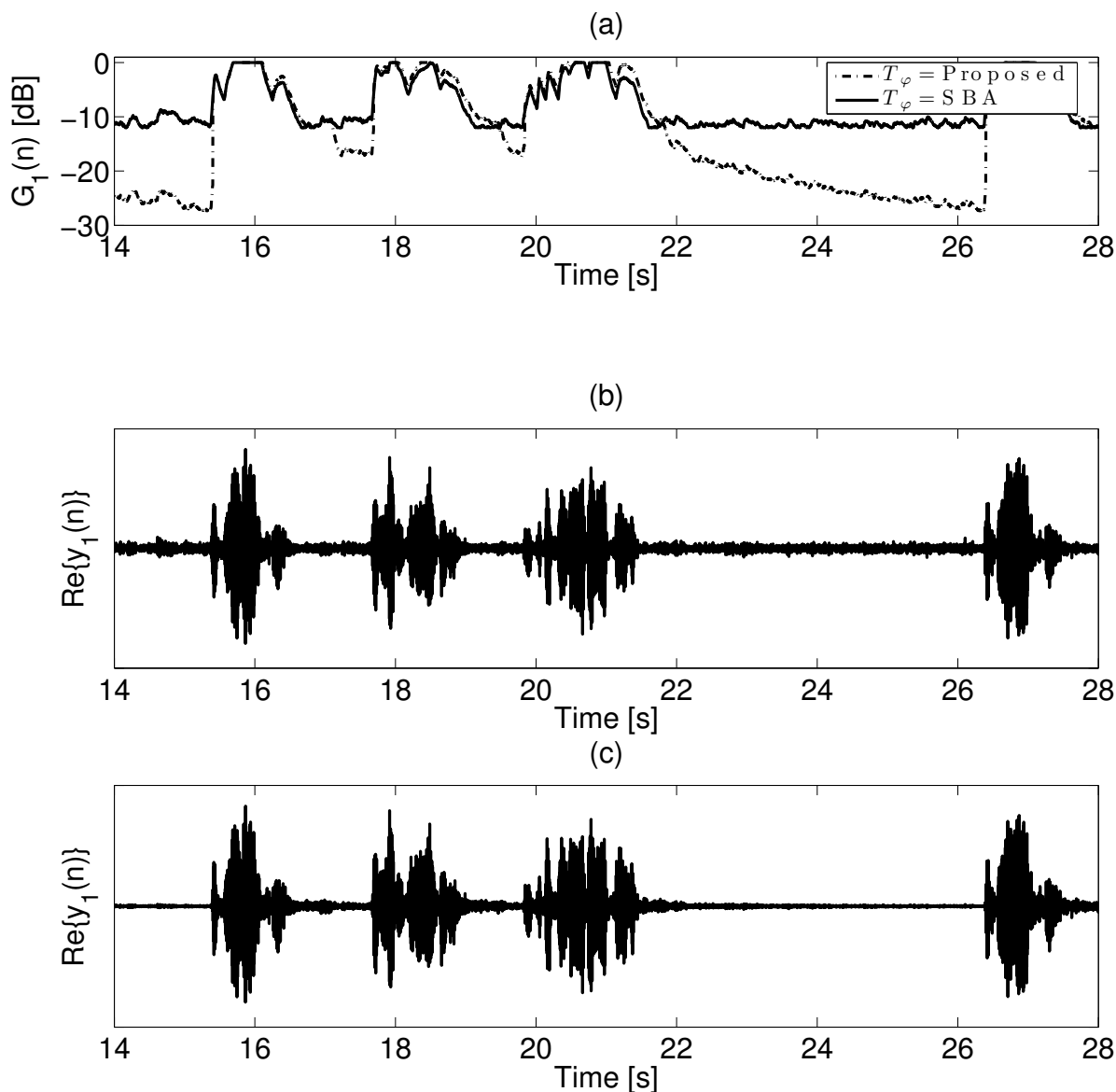
Figure 4 In plot a, the total gain $G_k(n)$ in subband $k = 1$ for the SBA and the proposed algorithm is shown. In plot **b**, the processed audio signal $y_k(n)$ in the same subband is shown for the SBA. In plot **c**, the processed audio signal $y_k(n)$ in the same subband is shown for the proposed algorithm.

even be set so that the noise becomes completely inaudible when maximum damping is applied.

6 Objective signal quality comparisons
To evaluate the performance of the SBA and the proposed algorithm in terms of speech quality and noise reduction, the SNR gain and speech distortion index [11,12] were used. The SNR gain, gSNR, is the difference between the input and output SNR, according to

$$gSNR = oSNR - iSNR. \tag{18}$$

In (18), iSNR and oSNR denote the input- and output SNR, respectively, defined as

$$iSNR = 10\,log_{10}\left(\frac{E\{s^2(n)\}}{E\{v^2(n)\}}\right), \tag{19}$$

and

$$oSNR = 10\,log_{10}\left(\frac{E\{\tilde{s}^2(n)\}}{E\{\tilde{v}^2(n)\}}\right), \tag{20}$$

where

$$\tilde{s}(n) = g_2(n) \sum_{k=0}^{K-1} g_{1,k}(n) s_k(n), \qquad (21)$$

and

$$\tilde{v}(n) = g_2(n) \sum_{k=0}^{K-1} g_{1,k}(n) v_k(n), \qquad (22)$$

where $sk(n)$ and $vk(n)$ are the subband versions of $s(n)$ and $v(n)$, respectively, and $\mathbb{E}\{\cdot\}$ denotes expected value.

The speech distortion index v_{sd} is a measure of how much the speech signal has been altered [11] and defined as

$$v_{sd} = 10 \log_{10} \left(\frac{E\left\{ (\tilde{s}(n) - s(n))^2 \right\}}{E\{s^2(n)\}} \right). \qquad (23)$$

Both the speech distortion index and the SNR gain are calculated globally. It should be noted that the SNR gain

and the speech distortion index are only evaluated when there is an active speech signal. Noise-only parts of the signal are not included in this part of the evaluation.

The objective comparison was performed with four different noise sources; noise recorded in a moving car traveling with a speed of 100 km/h, computer fan noise, ventilation noise and babble noise consisting of approximately 10 simultaneous speakers. Five different input SNR levels were used: 0, 6, 12, 18, and 24 dB. The increase rate of the noise-level estimation was set to 1 dB/s ($\beta_k = 2.3 \times 10^{-4}$), 3 dB/s ($\beta_k = 6.9 \times 10^{-4}$), 6 dB/s ($\beta_k = 1.4 \times 10^{-3}$), and 9 dB/s ($\beta_k = 2.1 \times 10^{-3}$) for both the SBA and the proposed method. The speech signals used in the evaluation were from the English speaking test samples of the ITU-T recommendation P.501 [13] and consisted of four speakers (2 male and 2 female) pronouncing one sentence each.

Figure 5 shows the speech distortion index for the SBA and the proposed algorithm. It can be seen that the speech distortion decreases with an increasing input SNR for both the SBA and the proposed method, which
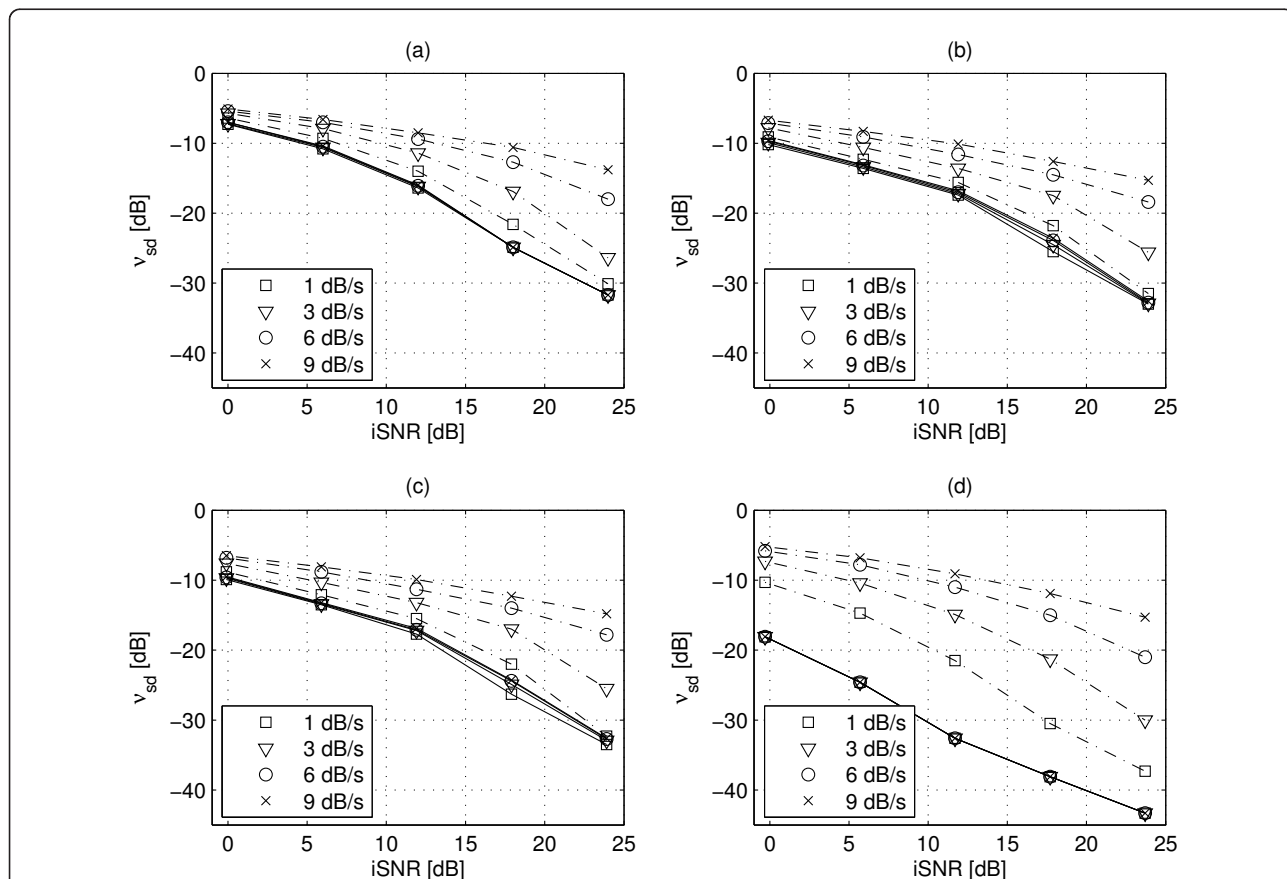
Figure 5 Speech distortion index for different noise characteristics and input SNR for both SBA (*dashdot*) and proposed (*solid*). Different increase rates, β_k, to a higher noise level (1, 3, 6, and 9 dB/s) were used. In **a**, the noise consists of noise recorded in a moving car, in **b** the noise comes from a computer fan, in **c** the noise comes from a ventilation system and in **d** the noise is babble noise from approximately 10 speakers.
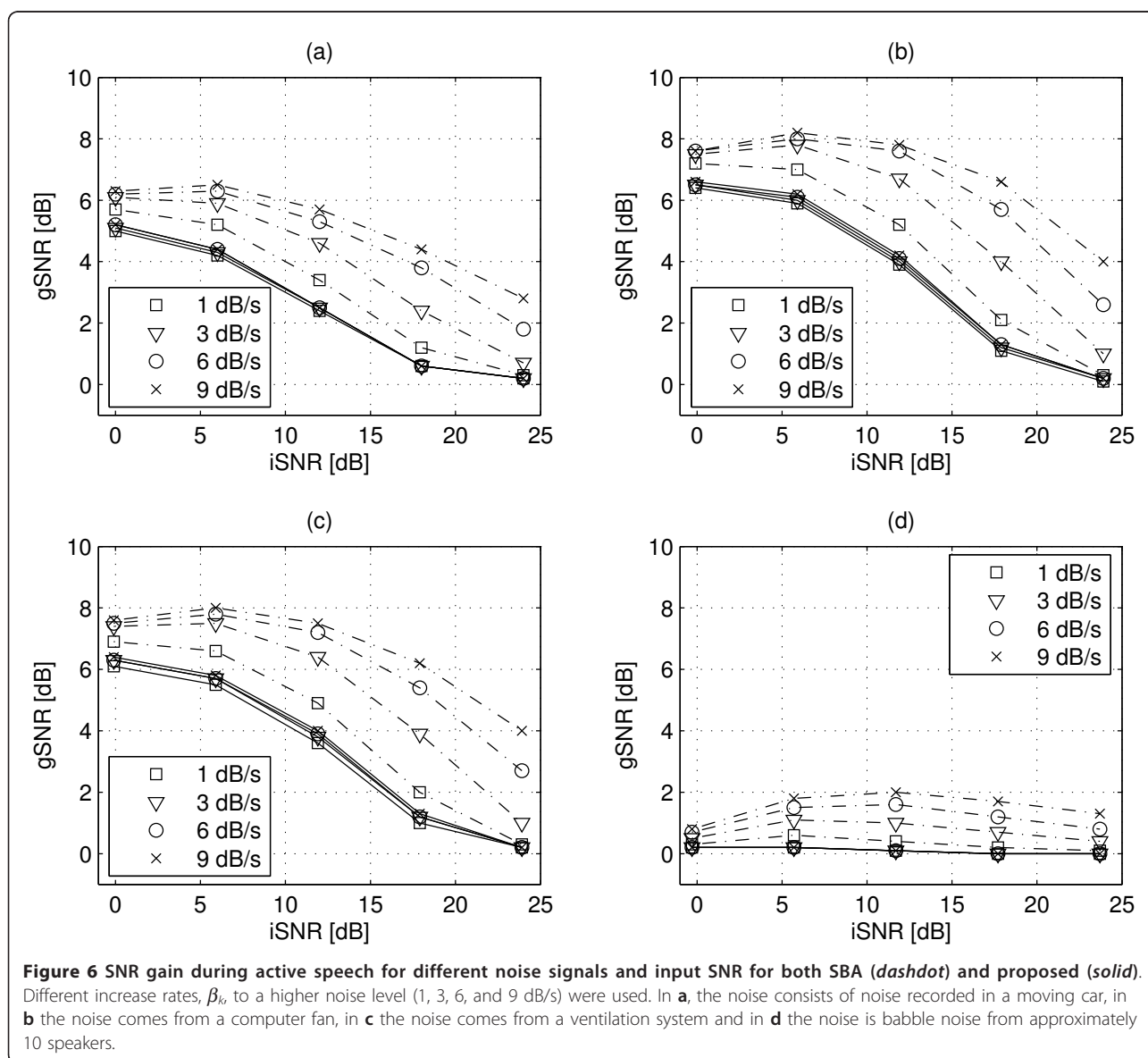
is expected since the fluctuations of the subband gains decrease as the input SNR increases. It can also be seen that the speech distortion of the proposed method is consistently lower than the SBA for all used noise sources and input SNRs.

For rapid increase rates of the noise-level estimation (i.e. large β_k), the SBA distorts the speech more than for a slower increase rate. This is due to the adaption of the noise-level estimation during speech, as demonstrated in Section 5.1. The proposed method does not have this increase in speech distortion for higher noise-level estimation increase rates. Thus, the proposed method allows much more rapid noise-level adaptation without any significant increase in speech distortion, compared to the original SBA. This behavior is consistent for all used noise sources, even for the non-stationary babble noise.

Figure 6 shows the SNR gain during active speech for both methods. From this figure, it can be seen that the SBA shows slightly higher SNR gain than the proposed method. This demonstrates the well-known trade-off between speech distortion and SNR improvement [11].

Of particular interest are the results of the babble noise, see Figures 5d and 6d. In this case, neither the SBA nor the proposed algorithm achieve any significant SNR improvement (less than 2 dB), due to the highly non-stationary nature of the noise. However, the speech distortion is significantly less for the proposed algorithm owing to the improved noise estimation.

Figure 6 SNR gain during active speech for different noise signals and input SNR for both SBA (*dashdot*) and proposed (*solid*). Different increase rates, β_k, to a higher noise level (1, 3, 6, and 9 dB/s) were used. In **a**, the noise consists of noise recorded in a moving car, in **b** the noise comes from a computer fan, in **c** the noise comes from a ventilation system and in **d** the noise is babble noise from approximately 10 speakers.

7 Comments on subjective evaluation

The algorithm behavior presented in Section 6 only describes the performance in active speech regions. The contribution of the additional noise reduction, $g_2(n)$, applied during speech pauses cannot be discerned from these results. However, this additional gain will reduce the noise level even further resulting in a much lower noise level compared to the SBA. In a conference phone application, a typical scenario is that parts on one side "listen in" to an ongoing presentation conducted by talkers on the opposite side. The extra noise reduction by $g_2(n)$ in speech pauses reduces annoyance from continuous noise in these situations. The modifications in this paper were motivated by artifacts from the SBA algorithm, subjectively perceived by an evaluation panel of product managers and development engineers, in total 6 persons. The improvements proposed in this paper were considered as necessary improvements to the SBA, and the proposed algorithm was implemented in a commercially available product. Especially, the inclusion of the additional gain $g_2(n)$ in (11) was perceived as desirable.

8 Conclusions

The noise reduction algorithm presented in this paper is an improvement in the SBA approach presented in [7], which incorporates subband division of the audio signal with a noise damping in each subband. The subband damping is proportional to the current SNR estimate in the corresponding subband, yielding noise reduction with low levels of speech distortion. The proposed algorithm introduces an additional noise reduction functionality, which is applied in speech pauses, allowing the noise level to be further reduced without adding any speech distortion.

Moreover, the proposed algorithm introduces a noise estimation update controller and a gain controller is used to determine whether the audio signal contains speech or only background noise. Owing to this fact, it is possible to obtain a more reliable noise level estimation and thus the gain in each subband will correspond to the actual SNR, resulting in less speech distortion compared to the original SBA.

Comparisons between the SBA and the proposed algorithm in four different noise conditions, including nonstationary babble noise, show that the proposed method introduces less (in some cases up to 25 dB less) speech distortion for all evaluated input SNRs.

9 Competing interests

The authors declare that they have no competing interests.

Author details
[1]Limes Audio AB, Box 7961, 90719 Umeå, Sweden [2]Department of Electrical Engineering, Blekinge Institute of Technology, 37179 Karlskrona, Sweden

References
1. SF Boll, Suppression of acoustic noise in speech using spectral subtraction. IEEE Trans Acoust Speech Signal Process. **27**, 113–120 (1979). doi:10.1109/TASSP.1979.1163209
2. PC Loizou, *Speech Enhancement: Theory and Practice* (CRC Press, Taylor & Francis Group, 2007)
3. Z Goh, K-C Tan, BTG Tan, Postprocessing method for suppressing musical noise generated by spectral subtraction. IEEE Trans Speech Audio Process. **6**, 287–292 (1998). doi:10.1109/89.668822
4. Y Ephraim, D Malah, Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. IEEE Trans Acoust Speech Signal Process. **32**, 1109–1121 (1984). doi:10.1109/TASSP.1984.1164453
5. Y Uemura, Y Takahashi, H Saruwatari, K Shikano, K Kondo, Musical noise generation analysis for noise reduction methods based on spectral subtraction and MMSE STSA estimation, in *ICASSP '09: Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, 4433–4436 (2009)
6. C Plapous, C Marro, P Scalart, Improved signal-to-noise ratio estimation for speech enhancement. IEEE Trans Acoust Speech Signal Process. **14**, 2098–2108 (2006)
7. N Westerlund, M Dahl, I Claesson, Speech enhancement for personal communication using an adaptive gain equalizer. Signal process. **85**, 1089–1101 (2005). doi:10.1016/j.sigpro.2005.01.004
8. R Flynn, E Jones, Combined speech enhancement and auditory modelling for robust distributed speech recognition. Speech Commun. **50**, 797–809 (2008). doi:10.1016/j.specom.2008.05.004
9. E Hänsler, G Schmidt, *Acoustic Echo and Noise Control: A Practical Approach* (Wiley, 2004)
10. C Schüldt, F Lindström, I Claesson, A low-complexity delayless selective subband adaptive filtering algorithm. IEEE Trans Signal Process. **56**, 5840–5850 (2008)
11. J Benesty, J Chen, Y Huang, I Cohen, in *Noise Reduction in Speech Processing*, vol. 2. (Springer, 2009)
12. J Chen, J Benesty, Y Huang, S Doclo, New insights into the noise reduction wiener filter. IEEE Trans Audio Speech Language Process. **14**, 1218–1234 (2006)
13. ITU-T, *Test Signals for Use in Telephonometry*. Recommendation ITU-T P.501 (International Telecommunication Union, Geneva, 2009)