PAPER

An Improved Perceptual MBSS Noise Reduction with an SNR-Based VAD for a Fully Operational Digital Hearing Aid

Zhaoyang GUO^{†,††}, Xin'an WANG^{†,††a)}, *Nonmembers*, Bo WANG^{††}, *Member*,
and Shanshan YONG^{††}, *Nonmember*

SUMMARY This paper first reviews the state-of-the-art noise reduction methods and points out their vulnerability in noise reduction performance and speech quality, especially under the low signal-noise ratios (SNR) environments. Then this paper presents an improved perceptual multiband spectral subtraction (MBSS) noise reduction algorithm (NRA) and a novel robust voice activity detection (VAD) based on the amended sub-band SNR. The proposed SNR-based VAD can considerably increase the accuracy of discrimination between noise and speech frame. The simulation results show that the proposed NRA has better segmental SNR (segSNR) and perceptual evaluation of speech quality (PESQ) performance than other noise reduction algorithms especially under low SNR environments. In addition, a fully operational digital hearing aid chip is designed and fabricated in the 0.13 μm CMOS process based on the proposed NRA. The final chip implementation shows that the whole chip dissipates 1.3 mA at the 1.2 V operation. The acoustic test result shows that the maximum output sound pressure level (OSPL) is 114.6 dB SPL, the equivalent input noise is 5.9 dB SPL, and the total harmonic distortion is 2.5%. So the proposed digital hearing aid chip is a promising candidate for high performance hearing-aid systems.

key words: digital hearing aid, noise reduction, multiband spectral subtraction, voice activity detection

1. Introduction

Every year, the need for hearing aid (HA) rises exponentially due to the dramatic number of individuals who suffer from some degree of hearing loss [1]. And the speech processing of hearing aid often operates under noisy environments. The background noise not only degrades speech quality and intelligibility, but also increases the discomfort of hearing aid users. Thus, the noise reduction is necessary for hearing aid designs.

Many noise reduction methods have been proposed [2], [3]. These methods include those based on low-level expansion [4], envelope tracking [5], envelope modulation filtering [6], adaptive low-pass [7] and high-pass filtering [8], Wiener filtering [9]–[11], statistical model based minimum mean square error (MMSE) estimators [12], subspace algorithm [13] and spectral subtraction [14]–[23]. In addition,

designing modern HA is a constant compromise between power, size and performance. So as one of the most important speech processing algorithms in HA, the noise reduction algorithm needs better performance and lower computational complexity. Among all these methods, Wiener-filtering-based or MMSE-based approaches are generally providing a good enhancement quality from the viewpoint of both noise reduction performance and the amount of nonlinear artifacts generated. But these methods require more complex operation than spectral-subtraction-based approaches. And the noise signal does not affect the speech signal uniformly over the whole spectrum, some frequencies are affected more adversely than others, so the noise reduction should be processed in the multiband. Exactly the multiband spectral subtraction (MBSS) can provide a good tradeoff between quality and computational complexity. So it has been a good choice for hearing aid applications.

The first version of spectral subtraction was proposed by Boll [15]. This method is capable of eliminating the stationary noises. Then, this method was developed by Berouti in order to improve the spectral subtraction to avoid musical noises [16]. In 1994, Martin [17] developed an algorithm which eliminates the need for explicit speech pause detection without substantial increase in computational complexity based on spectral subtraction and minimum statistics. After that, one of the popular improvements was made by Kamath in 2002 with his MBSS method [18]. There are several researches to MBSS with different approaches in order to make better speech enhancement performance. In 2008, a non-linear multi-band Bark scale frequency spacing approach was proposed to reduce coloured noises [22]. Then, in 2009, the perceptual frequency weighting function was introduced for improving the MBSS. This method can efficiently maximize additive noise reduction while minimizing the speech distortion [24]. The MBSS method based on phase modification and magnitude compensation was introduced in 2011. This method outperforms the original multi-band spectral subtraction method resulting in superior speech quality and well reduced musical noises [25]. The iterative processing of MBSS was introduced in 2012 [26]. In the same year, the spectral subtraction method combined with auditory masking properties was investigated [27]. The test result showed that it can considerably enhance the noise speech corrupted by white noise and reduce the musical noise. The noise reduction method based on MBSS and entropy-based VAD was introduced

Manuscript received December 7, 2016.

Manuscript revised January 18, 2017.

Manuscript publicized February 17, 2017.

[†]The authors are with Department of Institute of Microelectronics, Peking University, Yiheyuan Road 5, 100871 Beijing, China.

^{††}The authors are with Department of Key Laboratory of Microsystems, Peking University Shenzhen Graduate school, Lishui Road 2199, 518055 Shenzhen, China.

a) E-mail: anxinwang@pku.edu.cn

DOI: 10.1587/transinf.2016EDP7484

in 2014. This method can be simplified and still achieves good performance [28], [29]. In 2016, a speech enhancement approach based on analysis-synthesis framework was proposed [30]. The proposed approach takes advantage of the analysis-synthesis framework and an improved multi-band summary correlogram (MBSC) pitch detection algorithm to effectively eliminate musical noise.

In addition, in the recent years, a new generation of binaural hearing aids is available. So different binaural noise reduction techniques have been proposed for enhancement of the target signal arriving from any arbitrary direction (i.e., sound localization). These binaural noise-reduction strategies are based on spectral subtraction [31], statistical methods [32], and multichannel Wiener filtering (MWF) [33], [34]. The majority of the reported binaural noise-reduction techniques is fully exploiting this binaural unmasking effect to further improve speech intelligibility. But the binaural techniques are just compensatory for the basic noise reduction algorithms (i.e., spectral subtraction), so it is also necessary to improve the noise reduction algorithm for the binaural or monaural hearing aids.

However, these algorithms might introduce artificial noise problem mainly reflected in weakening the smoothness of the enhanced signals. This noise is mainly caused by the switches among non-linear functions of these algorithms. And the reason of generating non-linear functions is that the noise reduction is processed based on independent signal frame. So the proposed improved perceptual multiband spectral subtraction NRA takes advantage of the neighbourhood information of the adjacent signal frames and smooths the signal power spectrum of each sub-band in the different signal frames. Furthermore, in order to improve the smoothness of the enhanced signals, the proposed NRA is implemented using frame-based fast Fourier transform (FFT) processing with a frame size of 320 samples (zero-padded to 512 samples for the FFT) corresponding to 20 ms, and a frame advance of 160 samples (a half signal frame, corresponding to 10 ms) instead of 320 samples (a whole signal frame).

Besides, for high de-noise efficiency, the noise reduction always embeds VAD in order to distinguish between speech dominated duration and noise dominated duration. The traditional VAD usually detects voice based on zero crossing rate [35] or entropy [28]. The accuracy is quite high in high SNR stationary noise environment, but, under the low SNR or non-stationary noise environments, the accuracy of VAD is quite low due to the inaccurate background noise estimation. In order to improve the accuracy of VAD especially under the low SNR noise environments, the proposed SNR-based VAD adopts a decision rule based on two critical parameters, the normalized SNR of each sub-band and the deviation value of power spectrum in different sub-bands, which can be adaptively amended according to the SNR level and the power spectrum of each sub-band.

In addition, a fully operational digital hearing aid chip is designed and fabricated in the $0.13\mu\text{m}$ CMOS process based on the proposed NRA. Moreover, the internal status

controller and parameter interface are proposed in the hearing aid chip to adjust the degree of noise reduction, so that the noise reduction method is more appropriate for different background noise environments.

The remainder of this paper is organized as follows. Section 2 presents the architecture of the fully operational digital hearing aid chip. Section 3 discusses the improved perceptual multiband spectral subtraction NRA and the novel robust VAD based on amended sub-band SNR. Section 4 describes the performance analysis and implementation results. Conclusions are finally drawn in Sect. 5.

2. Chip Architecture

Figure 1 shows the system block diagram of the proposed fully operational digital hearing aid chip together with its Sigma-Delta Codec, dedicated digital signal processor (DSP), parameter interface and internal status controller. The Sigma-Delta Codec is a low power intellectual property (IP) core which consists of the preamp, the 16 bit Sigma-Delta analog to digital converter (ADC) and digital to analog converter (DAC), and the driver of the receiver. The dedicated DSP includes FFT, inverse fast Fourier transform (IFFT), four speech processing algorithms, and the shared arithmetic unit. The four speech processing algorithms consist of noise reduction (NR), wide dynamic range compression (WDRC), frequency shifting (FS), and feedback canceller (FC). WDRC compresses the dynamic range of the amplified signal to match the residual dynamic range of patients and also protects the residual hearing ability of patients. FS is designed for compressing and shifting the high frequencies above a pre-calculated cut-off frequency to a lower frequency range to improve high-frequency intelligibility. FC is designed for reducing the feedback of the hearing aid system. We did sufficient analysis of the four speech processing algorithms and extracted the shared arithmetic unit to reduce the computational complexity for low-power consideration. The internal status controller is designed to achieve the fully operational speech processing algorithms system. Because of the internal status controller we can

Fig. 1 System block diagram of the proposed digital hearing aid chip

selectively enable or disable these speech processing algorithms and change the processing degree of these algorithms according to the background noise condition. We can also program the parameters of the status controller and the dedicated DSP through the parameter interface. In this paper we mainly discuss the noise reduction algorithm (NRA) of the dedicated DSP, which is described as follows.

3. Proposed Improved Perceptual MBSS NRA and SNR-Based VAD

The block diagram of the proposed NRA is shown in Fig. 2. The proposed NRA has similar procedure with the conventional MBSS, but improves the latter in the aspects as follows. 1) In order to improve the accuracy of VAD especially under the low SNR noise environments, the proposed SNR-based VAD adopts a decision rule based on two critical parameters, the normalized SNR of each sub-band and the deviation value of power spectrum in different sub-bands. 2) To solve the artificial noise problem and achieve the smoothness of enhanced signals, the proposed NRA employs smoothing the signal power spectrum of each sub-band in the different signal frames when the sub-band SNR is calculated and the sub-band noise power is estimated. 3) Furthermore, the proposed NRA is implemented using frame-based FFT processing with a frame size of 320 samples (zero-padded to 512 samples for the FFT) corresponding to 20 ms, and a frame advance of 160 samples (a half signal frame, corresponding to 10 ms) instead of 320 samples (a whole signal frame).

In this system, each current noisy speech signal frame consists a new set of 160 samples and the front half of last signal frame. Then the current signal frame is windowed and transformed into frequency domain through FFT. The noisy speech signal is decomposed into K number of sub-bands based on critical frequencies. Then the SNR of each sub-band is calculated before the VAD and amended after the VAD. The gain function of each sub-band can be calculated based on the amended SNR and the estimated noise power. Finally, the enhanced signal is reconstructed from IFFT and Overlap-Add.

3.1 Sub-Band SNR Calculation

To suppress residual noise, the noise power spectrum must be estimated accurately. So the input signal is decomposed into K number of sub-bands to calculate noise power spectrum respectively.

As shown in Table 1, the different frequency ranges have different contributions to speech intelligibility [36]. To avoid spectral distortions in the sounds [37], according to the relation, the frequency range is divided into six bands shown in Table 1, so K is assigned 6.

The signal power spectrum $E_y(m, k)$ of each sub-band is calculated by the average of the square of the original signal spectrum $Y(m, k, i)$ according to

Fig. 2 Block diagram of the proposed NRA

Table 1 Relation between frequency range and intelligibility

Frequency range (Hz)	Intelligibility (%)
0-250	2
250-500	3
500-1000	35
1000-2000	35
2000-4000	13
4000-8000	12

$$E_y(m, k) = \frac{1}{N_k} \sum_{i=1}^{N_k} |Y(m, k, i)|^2, \quad (1)$$

where $m = 1, 2, \dots, M$ is the frame index, $k = 1, 2, \dots, K$ is the frequency bin index in each sub-band of one signal frame. M is the total number of frames and K is the total number of sub-bands in one signal frame. N_k is the number of data for sub-band k .

Because there is a relationship between the adjacent signal frames, we can smooth the signal power spectrum of each sub-band in different signal frame, as shown in

$$E_y(m, k) = \max\{E_{min}, \alpha(m)E_y(m-1, k) + (1 - \alpha(m))E_y(m, k)\}, \quad (2)$$

where $E_{min} = 0.0625$, which is the allowed minimum signal power spectrum of each sub-band. And, E_{min} is determined by the average value of minimum signal power spectrum of each sub-band, and through a large number of experiments, it is found that 0.0625 is the best choice for E_{min} in terms of the final output SNR of the algorithm, especially when the signal power spectrum of the sub-band is small. And 0.0625 can be implemented by right shift 4 bit, so it can reduce the computational complexity. $E_y(m-1, k)$ is the signal power spectrum of last signal frame. $\alpha(m)$ is the smooth factor and is assigned to the value 0.45, which is an empirical value for simple numerical smooth [17], [38], as shown in

$$\alpha(m) = \begin{cases} 0 & \text{if } m \leq 1; \\ 0.45 & \text{if } m > 1. \end{cases} \quad (3)$$

The SNR of each sub-band can be calculated using the Eq. (4), as shown in

$$SNR(m, k) = 10 \log \frac{E_y(m, k)}{E_n(m, k)}, \quad (4)$$

where $E_n(m, k)$ is the estimated noise power spectrum, which is calculated in Sect. 3.2.

3.2 Proposed SNR-Based VAD and Noise Power Estimation

In order to improve the accuracy of VAD and the performance of noise reduction especially under the low SNR noise environments, the proposed SNR-based VAD adopts a decision rule based on two critical parameters, the normalized SNR of each sub-band and the deviation value of power spectrum in different sub-bands, which can be adaptively amended according to the SNR level and power spectrum of each sub-band.

The normalized SNR (NSNR) is defined as the Eq. (5) and applied to achieve accurate logical judgment of VAD with low computational complexity.

$$NSNR(m, k) = \max \left\{ 0, \min \left\{ 89, \text{round} \left\{ \frac{SNR(m, k)}{0.375} \right\} \right\} \right\}, \quad (5)$$

where 89 and 0.375 are determined according to the range of $SNR(m, k)$. 0.375 is the quantization step of SNR, and the maximum value of the quantized sub-band SNR is around 89. The sum NSNR of all sub-bands, $NSNR_{sum}(m)$ is obtained by accumulating $NSNR(m, k)$ in one signal frame, as given in

$$NSNR_{sum}(m) = \sum_{k=0}^{K-1} NSNR(m, k). \quad (6)$$

Under the low SNR noise environments, the excessive low NSNR will lead to low gain function, and then the low gain function will weaken the smoothness of the enhanced signal. In order to improve this situation, the NSNR of each sub-band should be amended according to the $NSNR_{sum}(m)$ and $NSNR(m, k)$, as shown in

$$\begin{aligned} & \text{if}((NSNR_{sum}(m) \leq NSNR_{sum-TH}) \\ & \text{or} (NSNR(m, k) \leq NSNR-TH)), \\ & NSNR(m, k) = NSNR_{cons}, \end{aligned} \quad (7)$$

where $NSNR_{sum-TH}$, $NSNR-TH$, and $NSNR_{cons}$ are pre-defined thresholds and can be adaptively amended according to the background noise condition and programed through the parameter interface. In addition, $NSNR_{sum-TH}$ and $NSNR-TH$ are determined experimentally according to the minimum value of $NSNR_{sum}(m)$ and $NSNR(m, k)$, respectively. $NSNR_{cons}$ is assigned higher than $NSNR-TH$ to avoid that $NSNR(m, k)$ is too small to result in a small gain function, which will lead to obvious discontinuity in the enhanced signal.

The $NSNR_{sum}(m)$ is one of the parameters which are used for conditioning the noise signal frame. In order to improve the accuracy of VAD, another parameter is introduced, which is the deviation value between the signal power spectrum of the current signal frame and the average long time

signal power spectrum (the average signal power spectrum of the last signal frames).

The logarithm $E_{db}(m, k)$ of the current signal frame power spectrum $E_y(m, k)$ can be calculated using the Eq. (8), as shown in

$$E_{db}(m, k) = 10 \log E_y(m, k). \quad (8)$$

The deviation value $\Delta E(m)$ is calculated as

$$\Delta E(m) = \sum_{k=0}^{K-1} |E_{db}(m, k) - \bar{E}_{db}(m, k)|, \quad (9)$$

where $\bar{E}_{db}(m, k)$ is the average long time signal power spectrum and is calculated as

$$\bar{E}_{db}(m, k) = \begin{cases} E_{db}(1, k) & \text{if } m = 1; \\ \beta(m)\bar{E}_{db}(m-1, k) + (1-\beta(m))E_{db}(m, k) & \text{if } m > 1, \end{cases} \quad (10)$$

where $\beta(m)$ is the self-adaptive averaging factor [38] and interrelated to $E_{sum}(m)$ which is the sum power spectrum of all sub-bands in one signal frame. $\beta(m)$ and $E_{sum}(m)$ are given by the Eq. (11) and Eq. (12) respectively.

$$\beta(m) = \max \left\{ \beta_L, \min \left[\beta_H, \beta_H - \left| \frac{\beta_H - \beta_L}{E_H - E_L} \right| (E_H - E_{sum}(m)) \right] \right\}, \quad (11)$$

$$E_{sum}(m) = 10 \log \sum_{k=0}^{K-1} E_y(m, k), \quad (12)$$

where $E_H = 50$, $E_L = 30$, $\beta_H = 0.99$, $\beta_L = 0.5$. Because $\beta(m)$ is the self-adaptive averaging factor, it should be more than 0.5; therefore, we assign $\beta_L = 0.5$. The $\beta(m)$ should be less than 1, so we assign $\beta_H = 0.99$. The experimental results indicate that $E_{sum}(m)$ swings around 50, and the swing range is about 20. So we assign $E_H = 50$, and $E_L = 30$.

The VAD decision rule is based on the $NSNR_{sum}(m)$ and $\Delta E(m)$, as shown in Fig. 3. The $NSNR_{sum-THR}$ and $\Delta E-THR$ are pre-defined thresholds, which can be adjusted adaptively. Assuming that the first 32 signal frames are noise frames, so the $NSNR_{sum}(m)$ and $\Delta E(m)$ of the 32 signal frames can be calculated according to the Eq. (6) and Eq. (9). Their averages are regarded as the initial thresholds of $NSNR_{sum-THR}$ and $\Delta E-THR$. The thresholds will be updated when the current frame is judged as a noise frame, as shown in the Eq. (13) and Eq. (14).

$$NSNR_{sum-THR} = \partial 1 \cdot NSNR_{sum-THR} + (1 - \partial 1) \cdot NSNR_{sum}(m), \quad (13)$$

$$\Delta E-THR = \partial 2 \cdot \Delta E-THR + (1 - \partial 2) \cdot \Delta E(m) \quad (14)$$

where $\partial 1 = 0.98$ and $\partial 2 = 0.8$. They are the thresholds update coefficients which can be programed through the parameter interface. And in order to make $NSNR_{sum-THR}$

Fig. 3 Decision rule of the proposed VAD

and ΔE_THR update mainly based on their previous value, $\partial 1$ and $\partial 2$ are empirically assigned to around 1 [38].

As shown in Fig. 3, if $VAD_flag = 0$, the current signal frame is a noise frame; and when $VAD_flag = 1$, the current signal frame is a speech frame. If the current signal frame is the speech frame continuously and the deviation value $\Delta E(m)$ is continuously less than ΔE_THR , the VAD_cnt will increase. When the VAD_cnt is equal to 32, the $E_n(m, k)$ will also be updated so that the noise estimation will keep changed if the condition of background noise has a big change. The $E_n(m, k)$ is updated as the Eq. (15) when the current signal frame is a noise frame.

$$E_n(m+1, k) = \max \{E_{min}, \alpha_n E_n(m, k) + (1 - \alpha_n) E_y(m, k)\}, \quad (15)$$

where $E_{min} = 0.0625$ and it is the allowed minimum signal power spectrum of each sub-band, α_n is the smooth factor of sub-bands noise power spectrum. In order to make $E_n(m, k)$ contribute more to $E_n(m+1, k)$, α_n is assigned to around 0.9, which is an empirical value for simple numerical smooth [38].

3.3 Gain Function of Each Sub-Band Calculation

The gain function of each sub-band can be calculated based on the amended NSNR and the estimated noise power. First, the gain factor of the current signal frame is calculated according to the Eq. (16), as shown in

$$G_f(m) = \max \left\{ G_{min}, -10 \log \sum_{k=0}^{K-1} E_n(m, k) \right\}, \quad (16)$$

where $G_f(m)$ is the gain factor of the current signal frame, G_{min} is the allowed minimum gain which can be programmed through the parameter interface, and G_{min} can indirectly adjust the degree of noise reduction.

Then we can calculate the linear gain function of each

sub-band based on the gain factor $G_f(m)$ and the amended NSNR, as shown in

$$G(m, k) = \min \left\{ 1, 10^{\left(\mu_g (NSNR(m, k) - \sigma_{th}) + G_f(m) \right) / 20} \right\}, \quad (17)$$

where $\mu_g = 0.39$, $\sigma_{th} = 0.6$, and they are determined empirically. $G(m, k)$ is the linear gain function of each sub-band. We can get small $G(m, k)$ when the estimated noise power spectrum of the sub-band is large, on the contrary the large $G(m, k)$ can be obtained when the clean speech power spectrum of the sub-band is large.

The enhanced signal $Y_s(m, k)$ in frequency domain can be obtained by multiplying $G(m, k)$ and $Y(m, k)$, as shown in

$$Y_s(m, k) = G(m, k) \cdot Y(m, k). \quad (18)$$

And then the enhanced signal in time domain can be reconstructed from IFFT and Overlap-add.

3.4 Low Computational Complexity Consideration

In the proposed algorithm, multiplications are the bottleneck of the computational complexity. To reduce these costs, all multipliers are designed based on shifting and addition logic. For example, A (n bits) can be split as shown in Eq. (19), where $a_{(n-1)} \sim a_{(0)}$ is 0 or 1. So the multiplication of two operands, A and B, can be expressed as the sum-shift format, as shown in Eq. (20).

$$A[(n-1):0] = a_{(n-1)} \cdot 2^{(n-1)} + a_{(n-2)} \cdot 2^{(n-2)} + \dots + a_{(0)} \cdot 2^0 \quad (19)$$

$$A * B = a_{(n-1)} \cdot [B \ll (n-1)] + a_{(n-2)} \cdot [B \ll (n-2)] + \dots + a_{(0)} \cdot B \quad (20)$$

And the complicated logarithms and divisions are replaced by the multiplications according to the simplification method introduced in [29]. Moreover, the exponents are implemented in table lookup.

In addition, we did sufficient analysis of the four speech processing algorithms and extracted the shared arithmetic units to reduce the computational complexity for low-power consideration. Because the four speech processing algorithms are executed one by one in proper order, so they can invoke the shared arithmetic units independently.

4. Performance Analysis and Implementation Results

The simulations are executed to evaluate the performance of the proposed VAD and NRA. For VAD, the detection accuracy presented as percentage is calculated. The VAD accuracy is the percentage of the number of correctly detected frames to total number of frames. The number of correctly detected frames is total number of frames minus the number of missing frames (speech frame but detected as noise frame) and false alarmed frames (noise frame but

detected as speech frame). For noise reduction, segmental SNR (segSNR) and ITU-T P.862.2 Perceptual Evaluation of Speech Quality (PESQ), whose assessing programs are provided by [2], are calculated to evaluate the reduction performance and speech quality respectively. In the performance simulations, 32 Mandarin Chinese two-characters are concatenated into four sentences and corrupted by white, babble, factory and car noise provided by NOISEX-92 database [39]. The system sampling rate is 16 kHz. The proposed NRA is implemented using frame-based FFT processing with a frame size of 320 samples (zero-padded to 512 samples for the FFT) corresponding to 20 ms, and a frame advance of 160 samples (corresponding to 10 ms).

4.1 Proposed VAD Performance

Figure 4 displays the accuracy of the proposed SNR-based VAD compared with the sample-based entropy VAD [28], [29] and FFT-based entropy VAD [40] at five various input SNR levels and under four types of background noise environments. The sample-based VAD [28], [29] proposes a feature parameter, called spectral entropy, which exploits the use of the banded structure on speech spectrogram. And it adaptively selects useful bands not corrupted by noise to distinguish a speech from a nonspeech. The FFT-based entropy VAD [40] first calculates the probability density function of spectrum and then compares these values of features with preset thresholds to enhance the discriminability between speech signals and noise signals.

The simulation results indicate that the accuracy of the proposed VAD is ranged from 68% to 93% for different SNR levels and noise conditions. And the accuracy of the proposed SNR-based VAD is about 5% to 20% higher than the sample-based VAD and FFT-based entropy VAD especially for signals under 6 dB SNR. The reason is that the detection accuracy of the sample-based VAD and FFT-based entropy VAD is susceptible to large background noise and becomes lower for low SNR environment. On the contrary, the proposed SNR-based VAD adopts the decision rule based on two critical parameters, the normalized SNR of each sub-band and the deviation value of power spectrum in different sub-bands, which can be adaptively amended according to the SNR level and power spectrum of each sub-band. So it can be appropriate for different SNR environments especially under the low SNR noise environments.

4.2 Proposed NRA Performance

Figure 5 shows the simulation results of the proposed NRA with a noisy speech segment corrupted by 10 dB white noise from NOISEX-92 database [39]. It is obvious that the proposed method can achieve better noise reduction performance than the traditional MBSS. The proposed method reduces the noise in the speech period but maintains the formats of the speech, and the noise in the period of non-speech is also considerably reduced.

To evaluate the reduction performance (evaluated by

Fig. 4 Comparison of the proposed SNR-based VAD with the sample-based entropy VAD and FFT-based VAD under (a) White, (b) Babble, (c) Factory and (d) Car noise condition

Fig. 5 Simulation results of the MBSS and the proposed NRA with noisy speech corrupted by 10 dB white noise

segSNR) and speech quality (evaluated by PESQ) respectively, the proposed NRA is compared with other different noise reduction methods such as Wiener filter (wiener_wt) [11], subspace (klt) [13], the *mband* [18], the

Table 2 Comparison of the average PESQ under four types of background noise environments

Noise type	Method	0 dB	3 dB	6 dB	9 dB	12 dB
Average PESQ (white, babble, factory, car)	wiener_wt	1.4916	1.6423	1.9180	2.2364	2.3526
	<i>mband</i>	1.5226	1.7012	2.0012	2.2018	2.3415
	klt	2.1015	2.0634	2.3846	2.5612	2.8590
	sample-based	2.0430	2.2190	2.5648	2.7656	3.0175
	MBSC	2.4023	2.4942	2.6890	2.8256	3.0235
	Proposed	2.5512	2.6050	2.8432	2.9618	3.1015

Fig. 6 Comparison of the average PESQ under four types of noise environments

sample-based [28], [29], and MBSC [30] at five different input SNR levels (0, 3, 6, 9, 12 dB) and under four types of background noise environments (white, babble, factory and car). The programs of the wiener_wt, klt, and *mband* algorithms are provided by [2], and the programs of the sample-based and MBSC methods are provided by [28], [29] and [30], respectively.

The average PESQ scores with four types of background noises at the different input SNR levels are shown in Table 2 and Fig. 6. It is noted that an average of around 0.9 ~ 0.3 point improvement over the conventional methods are achieved in the various conditions. So in comparison with the different noise reduction methods, the proposed method achieves better objective speech quality.

The average segSNR of the different noise reduction methods under four types of background noise at the different input SNRs are shown in Table 3 and Fig. 7. The simulation results reveal that the proposed NRA has obvious improvement in comparison with the conventional methods due to the good VAD results, especially at low input SNR. The results show that the average improvement of segSNR is about 5.7 ~ 1.2 dB over the conventional noise reduction methods, respectively.

4.3 Implementation Results

The fully operational digital hearing aid chip containing the

Table 3 Comparison of the average segSNR under four types of background noise environments

Noise type	Method	0 dB	3 dB	6 dB	9 dB	12 dB
Average segSNR (white, babble, factory, car)	wiener_wt	3.462	4.9264	6.7548	8.5654	9.1876
	<i>mband</i>	3.603	5.3014	6.8375	8.16	8.7468
	klt	5.5658	9.4876	10.965	12.986	13.576
	sample-based	5.2336	9.7267	12.246	13.782	14.267
	MBSC	5.853	10.268	12.584	13.965	14.523
	Proposed	6.585	11.532	13.323	14.252	15.335

Fig. 7 Comparison of the average segSNR under four types of noise environments **Fig. 8** Micro photograph of the digital hearing aid chip. (LDO: Low Dropout Regulator; PLL: Phase Locking Loop; POR: Power On Reset.)

proposed NRA and VAD is fabricated in the Semiconductor Manufacturing International Corporation (SMIC) 0.13 μm CMOS process. The micro photograph of the digital hearing aid chip is shown in Fig. 8. The parameter interface, status controller, and dedicated DSP was designed in digital integrated circuit and merged together.

Table 4 shows the implementation summary. The total power consumption is measured by the real chip (including the SRAM power), and the power consumption of NRA is estimated according to post-simulation of the design com-

Table 4 Implementation summary

Process	SMIC 0.13 μm
Maximum clock rate	50 MHz
Chip die area	3.1mm x 3.2mm
NR core area (post-simulation)	0.342888 mm ²
Power consumption of chip	1.3 mA@1.2V
Power consumption of NR core (post-simulation)	183.5 μW @1.2V

Fig. 9 Digital hearing aid based on the proposed fabricated chip

Fig. 10 Test scene of the proposed digital hearing aid with FONIX TYPE 7000 ANALYZER

piller. The power reduction is mainly because of the proposed low-power approach introduced in Sect. 3.4. In addition, the gated clock and operand isolation techniques are applied in order to reduce the power consumption. Comparison for area, power and latency to other implementations is difficult to perform, because of diverged sub-band decomposition method, sub-band number, the number of speech processing algorithms and the power consumption of the analog circuits.

A hearing aid is made based on our proposed fully operational digital hearing aid chip, as shown Fig. 9. We test the hearing aid on FONIX TYPE 7000 ANALYZER which is a professional instrument for hearing aid acoustic test and the test scene is shown in Fig. 10. The WDRC, FS and FC algorithms are programmed to be disabled through the parameter interface to show the performance of noise reduction. Figure 11 shows the performance of the proposed NRA on background noise reduction. It is obvious that the background noise is reduced when the NRA is enabled.

In addition, the sample-based NRA leads to at least one

Fig. 11 Performance of the proposed NRA on background noise reduction

Fig. 12 Time sequence diagram of the HA hardware

Fig. 13 Acoustic test result of the digital hearing aid with FONIX TYPE 7000 ANALYZER

sample latency of the algorithm. The frame advance of the algorithm is 160 samples (corresponding to 10 ms for 16 kHz sampling rate). So, the preparation time of a frame is 10 ms, and the time will be decreased if we decrease the samples of a frame. The timing sequence diagram of the hardware implementation is shown in Fig. 12. The processing latency of all the algorithms is about 0.79 ms, and the noise reduction takes about 0.13 ms. Because the proposed algorithms are implemented in application specific integrated circuit (ASIC) and the chip can work at 50 MHz

Table 5 Summary of acoustic performance

Maximum output (OSPL90)	114.6 dB SPL
Equivalent input noise (EQ Input noise)	5.9 dB SPL
Frequency range	0-8000 Hz
Total harmonic reduction	2.5%
Battery life (@A13 Zn-air battery)	120 h

maximum clock rate, the processing time of the algorithms is short. So the total latency of the proposed HA system is 10.79 ms, and it can meet the latency tolerance of HA (about 10 ms ~ 15 ms) [40].

The acoustic test result is shown in Fig. 13 and the summary of the acoustic test results is shown in Table 5. The maximum output sound pressure level (OSPL) is 114.6 dB SPL, the equivalent input noise is 5.9 dB SPL, and the total harmonic distortion is 2.5%. Acoustic test result shows that the proposed digital hearing aid chip is a promising candidate for high performance hearing-aid systems.

A limitation of this study is that the experiments involve software simulations, and the acoustic test result with the professional instrument. However, the subjective speech intelligibility improvement and discomfort feeling reduction should be evaluated in the clinical trials. These will be conducted in our future study.

5. Conclusions

This paper presents an improved perceptual multiband spectral subtraction noise reduction algorithm and a novel robust voice activity detection based on amended sub-band SNR. The proposed NRA employs smoothing the signal power spectrum of each sub-band to deal with the artificial noise problem and achieve the smoothness of the enhanced signals. In order to improve the accuracy of VAD and the performance of noise reduction especially under the low SNR noise environments, the proposed SNR-based VAD adopts a decision rule based on two critical parameters, the normalized SNR of each sub-band and the deviation value of power spectrum in different sub-bands. The simulation results show that the accuracy of the proposed SNR-based VAD is about 5% to 20% higher than the sample-based VAD [28], [29] and FFT-based entropy VAD [40] especially for signals under 6 dB SNR noise conditions. In different noise environments, the average improvement of PESQ is 0.9 ~ 0.3 point and segSNR is 5.7 ~ 1.2 dB over the conventional noise reduction methods [11], [13], [28]–[30] respectively. In addition, a fully operational digital hearing aid chip is designed and fabricated in the 0.13 μm CMOS process based on the proposed NRA. The final chip implementation shows that the whole chip dissipates 1.3 mA at the 1.2 V operation. A hearing aid is made based on the proposed fully operational digital hearing aid chip. Acoustic test result shows that the maximum output sound pressure level (OSPL) is 114.6 dB SPL, the equivalent input noise is 5.9 dB SPL, and the total harmonic distortion is 2.5%. So the test results demonstrate that the proposed digital hearing aid chip is a promising candidate for high performance

hearing-aid systems.

Acknowledgements

This work is supported by R&D Project of Shenzhen Government (Project JSGG20130412145326444 and JCYJ20150331102721193).

References

- [1] S. Kim, N. Cho, S.-J. Song, and H.-J. Yoo, "A 0.9 V 96 μW Fully Operational Digital Hearing Aid Chip," *IEEE J. Solid-State Circuits*, vol.42, no.11, pp.2432–2440, 2007.
- [2] P.C. Loizou, ed., *Speech Enhancement: Theory and Practice*, CRC Press, Boca Raton, Florida, 2007.
- [3] J.M. Kates, *Digital hearing aids*, Plural Pub., 1st ed., 2008.
- [4] P.M. Clarkson and S.F. Bahgat, "Envelope expansion methods for speech enhancement," *J. Acoust. Soc. Am.*, vol.89, no.3, pp.1378–1382, 1991.
- [5] A. Schaub, *Digital hearing aids*, Thieme Medical Publishers, 1st edn, 2008.
- [6] H. Hermansky and N. Morgan, "RASTA processing speech," *IEEE Trans. Speech Audio Process.*, vol.2, no.4, pp.578–589, 1994.
- [7] J.M. Kates, "Signal processing for hearing aids," *Hear. Instrum.*, vol.37, pp.19–21, 1986.
- [8] A.L. Gingsjo, On transient noise and its reduction in hearing aids, PhD thesis, Department of Applied Electronics, Chalmers University of Technology, 1997.
- [9] J.S. Lim and A.V. Oppenheim, "All-pole modeling of degraded speech," *IEEE Trans. Acoust. Speech Signal Process.*, vol.26, no.3, pp.197–210, 1978.
- [10] H. Levitt, M. Bakke, J. Kates, A. Neuman, and T. Schwander, "Signal processing for hearing impairment," *Scand. Audiol.*, vol.38, pp.7–19, 1993.
- [11] Y. Hu and P.C. Loizou, "Speech enhancement based on wavelet thresholding the multitaper spectrum," *IEEE Trans. Speech Audio Process.*, vol.12, no.1, pp.59–67, 2004.
- [12] Y. Ephraim and D. Malah, "Speech enhancement using a minimum-mean square error short-time spectral amplitude estimator," *IEEE Transactions on Acoustics Speech & Signal Processing*, vol.32, no.6, pp.1109–1121, 1984.
- [13] Y. Hu and P.C. Loizou, "A generalized subspace approach for enhancing speech corrupted by colored noise," *IEEE Trans. Speech Audio Process.*, vol.11, no.4, pp.334–341, 2003.
- [14] A. Schaub and P. Straub, "Spectral sharpening for speech enhancement/noise reduction," *Proc. Int. Conf. Circuit Systems*, Toronto, Ontario, Canada, pp.993–996, May 1991.
- [15] S.F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust. Speech Signal Process.*, vol.27, no.2, pp.113–120, 1979.
- [16] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP*, pp.208–211, 1979.
- [17] R. Martin, "Spectral subtraction based on minimum statistics," *Proc. EUSIPCO*, pp.1182–1185, 1994.
- [18] S. Kamath and P. Loizou, "A multi-band spectral subtraction method for enhancing speech corrupted by colored noise," *Proc. IEEE Int. Conf. Acoustics Speech Signal Processing*, Orlando, FL, USA, p.4164, May 2002.
- [19] J.S. Lim and A.V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," *Proc. IEEE*, vol.67, no.12, pp.1586–1604, 1979.
- [20] B.L. Sim, Y.C. Tong, J.S. Chang, and C.T. Tan, "A parametric formulation of the generalized spectral subtraction method," *IEEE Trans. Speech Audio Process.*, vol.6, no.4, pp.328–337, 1998.

- [21] H.T. Hu and C. Yu, "Adaptive noise spectral estimation for spectral subtraction speech enhancement," *IET Signal Process.*, vol.1, no.3, pp.156–163, 2007.
- [22] R.M. Udrea, N. Vizireanu, S. Ciochina, and S. Halunga, "Nonlinear spectral subtraction method for colored noise reduction using multi-band Bark scale," *Signal Process.*, vol.88, no.5, pp.1299–1303, 2008.
- [23] C.V.R. Rao, M.B.R. Murthy, and K.S. Rao, "Noise reduction using Mel-SCALE spectral subtraction with perceptually defined subtraction parameters - a new scheme," *Signal Image Process.*, vol.2, no.1, pp.135–149, 2011.
- [24] F.A. Chowdhury, J. Alam, F. Alam, and D. Shaughnessy, "Perceptually weighted multi-band spectral subtraction speech enhancement technique," *International Conference on Electrical and Computer Engineering*, pp.395–399, 2009.
- [25] C. Li and W.-J. Liu, "A novel multi-band spectral subtraction method based on phase modification and magnitude compensation," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp.4760–4763, 2011.
- [26] N. Upadhyay and A. Karmakar, "Single channel speech enhancement utilizing iterative processing of multi-band spectral subtraction algorithm," *International Conference on Power, Control and Embedded Systems (ICPCES)*, pp.1–6, 2012.
- [27] L. Cao, T.-Q. Zhang, H.-X. Gao, and C. Yi, "Multi-band spectral subtraction method combined with auditory masking properties for speech enhancement," *International Congress on Image and Signal Processing (CISP)*, pp.72–76, 2012.
- [28] C.-W. Wei, C.-C. Tsai, T.-S. Chang, and S.-J. Jou, "Perceptual multi-band spectral subtraction for noise reduction in hearing aids," *2010 IEEE Asia Pacific Conference on Circuits and Systems*, pp.692–695, 2010.
- [29] C.-W. Wei, T.-S. Chang, S.-J. Jou, C.-C. Tsai, Y. FanJiang, "Analysis and implementation of low-power perceptual multiband noise reduction for the hearing aids application," *Iet Circuits Devices & Systems*, vol.8, no.6, pp.516–525, 2014.
- [30] B. Liu, J. Tao, Z. Wen, and F. Mo, "Speech Enhancement Based on Analysis-Synthesis Framework with Improved Parameter Domain Enhancement," *Journal of Signal Processing Systems*, vol.82, no.2, pp.141–150, 2016.
- [31] B.P.A. Rohman, C.B.A. Wael, and K. Paramayudha, "Adaptively compensated multiband spectral subtraction for robust noise reduction," *International Symposium on Intelligent Signal Processing and Communication Systems*, pp.139–142, 2015.
- [32] J. Li, S. Sakamoto, S. Hongo, M. Akagi, and Y. Suzuki, "Two-stage binaural speech enhancement with Wiener filter based on equalization-cancellation model," *Proc. IEEE Workshop Applcat. Signal Process. Audio Acoust. (WASPAA)*, pp.133–136, 2009.
- [33] S. Doclo, R. Dong, T.J. Klasen, J. Wouters, S. Haykin, and M. Moonen, "Extension of the multi-channel Wiener filter with ITD cues for noise reduction in binaural hearing aids," *Proc. IEEE Workshop Applcat. Signal Process. Audio Acoust. (WASPAA)*, pp.70–73, 2005.
- [34] T.V. Bogaert, S. Doclo, J. Wouters, and M. Moonen, "Speech enhancement with multichannel Wiener filter techniques in multimicrophone binaural hearing aids," *Journal of the Acoustical Society of America*, vol.125, no.1, pp.360–371, 2009.
- [35] J.C. Junqua, B. Reaves, and B. Mak, "A study of endpoint detection algorithms in adverse conditions: Incidence on a DTW and HMM recognize," *Proc. Conf. Eurospeech*, pp.1371–1374, 1991.
- [36] H. Dillon, *Hearing aids*, 2nd ed., Thieme, 2012.
- [37] U. Arioz, K. Arda, and U. Tuncel, "Preliminary results of a novel enhancement method for high-frequency hearing loss," *Computer Methods and Programs in Biomedicine*, vol.102, no.3, pp.277–287, 2011.
- [38] N. Virag, "Single channel speech enhancement based on masking properties of the human auditory system," *IEEE Trans. Speech Audio Process.*, vol.7, no.2, pp.126–137, 1999.
- [39] A. Varga, H.J.M. Steenneken, M. Tomlinson, and D. Jones, (1992), NOISEX-92, [Online], Available: http://spib.rice.edu/spib/select_noise.html, accessed July 2012.
- [40] C. Jia and B. Xu, "An improved entropy-based endpoint detection algorithm," *Int. Symp. Chinese Spoken Language Processing*, pp.96–99, Taipei, Taiwan, Aug. 2002.
- [41] M.A. Stone and B.C.J. Moore, "Tolerable hearing aid delays. II. Estimation of limits imposed during speech production," *Ear and Hearing*, vol.23, no.4, pp.325–338, 2002.

Zhaoyang Guo received the B.S. degree from Central South University, Changsha, China, in 2011. He is currently working toward the Ph.D. degree in Integrated Circuit and System at Peking University. His main research interests include RFID systems, Speech processing for hearing aid, SoC and ASIC design.

Xin'an Wang is a professor of Peking University, doctoral tutor, vice president of school of Electronic and Computer Engineering, Peking University, executive vice director of the Key Laboratory of Integrated Microsystems, the committee of China Computer Federation, the vice-chairman of Shenzhen Semiconductor Industry Association. His main research fields include microprocessor architecture, integrated circuit and system design, and life health monitoring.

Bo Wang received the B.S. degree from Beijing Jiaotong University, Beijing, China, in 1998, the M.S. degree in integrated electronic circuits from INSA de Lyon, Lyon, France, in 2005, and the Ph.D. degree in high-frequency electronics from University of Limoges, Limoges, France, in 2009. Since 2010, he has been an assistant professor in Peking university Shenzhen graduate school. His main research interests include RF front-end circuits, analog and mixed signal circuits.

Shanshan Yong received the B.S. degree from Dalian University of Technology, Dalian, China, in 2008, the Ph.D. degree in the Key Laboratory of Integrated Microsystems from Peking University, Beijing, China, in 2014. Since 2014, she has been a post-doctoral in Peking university Shenzhen graduate school. Her main research interests focuses on Digital Signal Processing and Digital IC design.