www.ietdl.org

Published in IET Circuits, Devices & Systems
Received on 14th August 2012
Revised on 14th April 2014
Accepted on 6th May 2014
doi: 10.1049/iet-cds.2013.0326

ISSN 1751-858X

Analysis and implementation of low-power perceptual multiband noise reduction for the hearing aids application

Cheng-Wen Wei, Cheng-Chun Tsai, Yi FanJiang, Tian-Sheuan Chang, Shyh-Jye Jou

Department of Electrical Engineering, National Chiao Tung University, Hsinchu 300, Taiwan
E-mail: jackiewei.ee95g@nctu.edu.tw

Abstract: Traditional noise reduction designs provide good performance but suffer from high complexity and long latency, which limits their application to hearing aids. Targeted for strict low-power and low-latency requirement of completely-in-the-canal type hearing aids, this study analyses and implements a previously proposed sample-based perceptual multiband spectral subtraction with a multiplication-based entropy voice activity detection. Simulation results reveal that the authors design can provide similar speech quality as others, but with lower computational complexity and simple control effort. The corresponding core-based architecture design further exploits processing characteristics of the proposed approach to reduce power consumption with a sign-magnitude and a preprocessed input data reuse scheme. Chip measurement shows that the design only consumes 83.7 µW at 0.6 V operation with 90 nm high threshold voltage (HVT) (high V_T) standard cell library.

1 Introduction

Background noise not only degrades speech quality and intelligibility, but also increases the discomfort of hearing aid users. Thus, noise reduction (NR) is necessary for the hearing aid designs.

For single microphone hearing aids, such as completely-in-the-canal (CIC) type ones, NR cancels noise based on the statistic information of noise. For such system, many NR methods have been proposed [1, 2]. These methods include those based on low-level expansion [3], envelope tracking [4], adaptive low-pass [5] and high-pass filtering [6], envelope modulation filtering [7], spectral sharpening [8], Winner filtering [9, 10] and spectral subtraction [11–18]. Among all these methods, spectral subtraction can provide a good tradeoff between quality and computational complexity, which has been a good choice for hearing aid applications.

The spectral subtraction, firstly proposed by Boll [11], cancels noise in speech based on the assumption that speech and noise are uncorrelated, which was improved by Berouti *et al.* [12] to reduce the artefacts (also known as musical noise [1]) caused by NR process. Those methods can be further generalised to improve quality by proper tradeoff the parameters [13]. Based on this idea, Sim *et al.* [14] proposed a method for optimally selecting the parameters in the sense of minimum mean squared error. In addition, Hu and Yu [15] proposed an adaptive noise estimation method for quality improvement.

Another well-known spectral subtraction method is the multiband spectral subtraction (*mband*) [16] which exploited the fact that the noise is not uniformly distributed in all frequency bands. Thus, noise attenuation can be adaptive to the noise of each band, and thus results in better speech quality [1]. Based on the *mband* method, perceptual *mband* [17, 18] was also proposed by further exploiting perceptual frequency decomposition for efficient noise suppression. However, conventional *mband* methods cannot be directly applied to hearing aids because of their strict low-power and low-latency requirement. The low-power problem is mainly caused by the high computational complexity of *mband* and its voice activity detection (VAD) according to the programme from [1]. Our power target is set to be <100 µW since the averaged power consumption for commercial CIC hearing aids is about 1 mW (from 0.8 to 1.3 mA at 1.2 V operation) [19, 20]. In addition, both of them are frame-based methods which have high latency owing to the inherent lag in buffering the initial frame which is usually larger than the 15 ms limitation of hearing aids [21].

Targeting for above hearing aids requirement, this paper presents an NR design by exploiting the idea of perceptual *mband* [17, 18] and entropy-based VAD [22, 23] as the basis method for further development. This paper is based on our previous publication [24], but provides an in-depth analysis of algorithm, complexity, speech quality and power consumption as well as detailed architecture and chip implementation. The entropy-based VAD methods detect speech based on signal randomness and are robust to the background noise when compared with other low-complexity methods, such as energy and zero crossing ones. For low latency, this paper adopts the sample-based decomposition by the 1/3 octave ANSI S1.11 filter bank of auditory compensation [25, 26]. For low computational complexity, this paper proposes four low-power techniques beyond conventional low-power design methods. The first one is to eliminate overlap-add computations with the proposed sample-based approach. The second one is to eliminate complex divisions and logarithms in the multiplication-based entropy VAD. The third one is a sign-magnitude input data format [27] and preprocessed input data reuse between different steps with a sliding window integration scheme. The fourth one is a functional gating for high signal-to-noise ratio (SNR) input. The final implementation adopts a core-based architecture to provide good programmability for future extension.

The rest of this paper is organised as follows. Section 2 presents the details of our design and complexity comparison. Section 3 shows the simulation results and comparison for speech quality. Section 4 presents the results of chip implementation. Conclusions are finally drawn in Section 5.

2 NR design

Fig. 1 illustrates the block diagram of our digital hearing aid chip [28]. It consists of analysis and synthesis filter bank (AFB and SFB), NR, insertion gain (IG) and wide dynamic range compensation (WDRC). The input speech is first decomposed into subband signals by ANSI S1.11 AFB designed based on 1/3 octave distribution for human auditory system and multirate processing for low-power signal decomposition [29]. The background noise in each subband is reduced by the NR to improve speech quality. The IG and WDRC are designed for auditory compensation to improve speech intelligibility [25, 28–32]. The B2C and SFB serve as frequency reconstruction tools, where the B2C merges 18 subbands (from IG) to three compression channels (to WDRC) and the SFB synthesises the three WDRC channels for final output generation [32]. The B2C and SFB include constant delays to compensate the different delays between subbands.

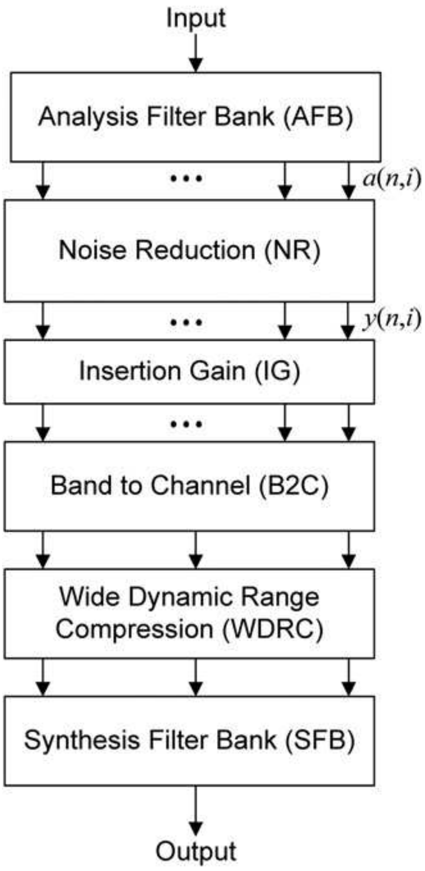

Input
Analysis Filter Bank (AFB)
a(n,i)
Noise Reduction (NR)
y(n,i)
Insertion Gain (IG)
Band to Channel (B2C)
Wide Dynamic Range Compression (WDRC)
Synthesis Filter Bank (SFB)
Output

Fig. 1 *Block diagram of digital hearing aid [28]*

Fig. 2 depicts the block diagram of our NR design. This method is a fully sample-based one and utilises the perceptual decomposed signal to detect the voice activity by multiplication-based entropy VAD and adaptive threshold for different noise conditions. The noise in speech is reduced by the sample-based *mband* for efficient noise cancellation and low-power consumption. The sample-based *mband* is turned off by a functional gating mechanism to reduce power consumption in the high SNR condition. The details of those blocks are described below.

2.1 Sample-based multiband spectral subtraction and its low-power implementation

The sample-based *mband* works in a similar way as traditional *mband*, but with sample-based processing, sign-magnitude scheme and simplified operations. For the input signal treated as voice, clean speech $\hat{y}(n, i)$ is estimated by noise subtraction as below

$$\left|\hat{y}(n, i)\right| = \left|a(n, i)\right| - \mu_i\left|Pn(n, i)\right| \tag{1}$$

$$\mu_i = \begin{cases} B_1, & \text{if } i = F22 - F30 \\ B_2, & \text{if } i = F31 - F39 \end{cases} \tag{2}$$

where $a(n, i)$ and $Pn(n, i)$ are the AFB output and estimated noise power of subband i at time n, respectively. μ_i is oversubtraction factor which is a set of constants designed to enhance the cancellation of non-speech subbands. B_1 and B_2 are set as 1.0 and 2.5 in our case, respectively, where the multiplication by 2.5 can be realised by simple 1 bit shift and addition. This setting results in low attenuation below $F30$ (1 kHz) to reduce speech distortion. To avoid negative values resulted from (1), $\left|\hat{y}(n, i)\right|$ is floored as follows

$$\left|\hat{y}(n, i)\right| = \begin{cases} \left|\hat{y}(n, i)\right|, & \text{if} \left|\hat{y}(n, i)\right| > \beta\left|a(n, i)\right| \\ \beta\left|a(n, i)\right|, & \text{else} \end{cases} \tag{3}$$

where β is 0.0625 and can be implemented by right shift 4 bit.

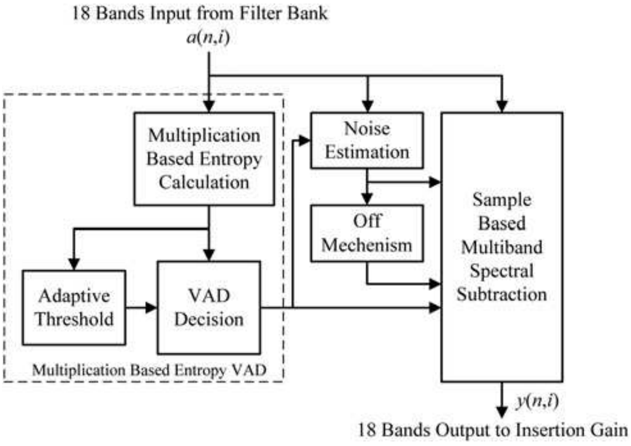

18 Bands Input from Filter Bank
a(n,i)
Multiplication Based Entropy Calculation
Noise Estimation
Off Mechenism
Sample Based Multiband Spectral Subtraction
Adaptive Threshold
VAD Decision
Multiplication Based Entropy VAD
y(n,i)
18 Bands Output to Insertion Gain

Fig. 2 *Block diagram of our NR design*

Moreover, to reduce musical noise, a small input signal is added after noise subtraction as (4)

$$\left|\tilde{y}(n,\ i)\right| = \left|\hat{y}(n,\ i)\right| + \gamma\left|a(n,\ i)\right| \tag{4}$$

where γ is a small constant and can also be implemented by bit shifting. The final denoised output of subband i at time n is expressed as sign-magnitude scheme

$$y(n,\ i) = \operatorname{sign}\big(a(n,\ i)\big) \times \left|\tilde{y}(n,\ i)\right| \tag{5}$$

For input signal treated as noise, the input is processed by spectral attenuation as sign-magnitude scheme, shown in (6)

$$y(n,\ i) = \operatorname{sign}\big(a(n,\ i)\big) \times \left|a(n,\ i)\right| \times \mathrm{Att} \tag{6}$$

where *Att* can also be implemented by bit shifting.

The sample-based *mband* leads to one sample latency. However, the latency of our design depends on system architecture. In our hearing aid, since the AFB adopts multirate structure and delivers output $a(n,\ i)$ after the new sequence (length N_i) of associated subband i is collected in buffer, where N_i is as below

$$N_i = \begin{cases} 32, & i = F39,\ F38,\ F37 \\ 16, & i = F36,\ F35,\ F34 \\ 8, & i = F33,\ F32,\ F31 \\ 4, & i = F30,\ F29,\ F28 \\ 2, & i = F27,\ F26,\ F25 \\ 1, & i = F24,\ F23,\ F22 \end{cases} \tag{7}$$

Therefore the latency of our design is 1.3 ms (32 samples for 24 kHz sampling rate). The VAD which will be described later does not cause latency because it does not locate at the forward path of output generation.

The end-to-end formulation of our design can be written as below

$$y(n,\ i) = \begin{cases} \operatorname{sign}(a(n,\ i)) \times \left[\left(1+\gamma\right)\left|a(n,\ i)\right| - \mu_i\left|Pn(n,\ i)\right|\right] \\ \operatorname{sign}(a(n,\ i)) \times \left[\left(\beta+\gamma\right)\left|a(n,\ i)\right|\right] \end{cases} \tag{8}$$

Equation (8) reveals that the main computation of our design can be represented by the power efficient sign-magnitude format. In which, the magnitude can be divided into signal part $|a(n,\ i)|$ and noise part $|Pn(n,\ i)|$, whereas the sign of input $a(n,\ i)$ can be stored for final output.

2.2 Noise power estimation

The noise estimation estimates noise power based on sign-magnitude scheme as below

$$Pn(n,\ i) = \frac{1}{4N_i}\sum_{m=n-4N_i+1}^{n}\left|a(m,\ i)\right|, \quad n = 0,\ N,\ 2N,\ \ldots \tag{9}$$

where $Pn(n,\ i)$ is the noise power at time n in each subband i. $a(m,i)$ is the AFB output of the subband i at time m.

2.3 Multiplication-based entropy VAD and its low-power implementation

2.3.1 Entropy calculation:

Original entropy calculation [22, 23] is based on frequency domain and uniform bandwidth decomposition. To estimate the entropy in time domain using the non-uniform bandwidth decomposition of ANSI S1.11 filter bank, a modified entropy $H'(n)$ as shown in (10) is designed

$$H'(n) = -\sum_{i=F22}^{F39} P'(n,\ i)\log_{10} P'(n,\ i), \quad n = 0,\ 4N,\ 8N,\ \ldots \tag{10}$$

where

$$P'(n,\ i) = \left(\frac{Pa(n,\ i)G_i + K}{\sum_{i=F22}^{F39}\left(Pa(n,\ i)G_i + K\right)}\right) \tag{11}$$

$$Pa(n,\ i) = \frac{1}{8N_i}\sum_{m=n-8N_i+1}^{n}\left|a(m,\ i)\right| \tag{12}$$

N is the maximum of N_i and is 32 in our case. $Pa(n,\ i)$ is also based on sign-magnitude scheme. The VAD operates once every $4N$ samples. K is a preset constant proposed by [22] to reduce the impact by spectrum variation and maintain the signal uncertainty for entropy computation. This enhances the difference between noise and speech, and thus improves VAD discrimination. G_i is a set of constants inversely proportional to subband bandwidth to improve discrimination clarity by enhancing the entropy difference between speech and noise which is degraded by non-uniform bandwidth, as shown in Fig. 3.

With the modified entropy $H'(n)$, the voice activity $\mathrm{VA}(n)$ is decided once every $4N$ samples based on a threshold $\mathrm{VA\_thr}(n)$ as below

$$\mathrm{VA}(n) = \begin{cases} 1, & \mathrm{if}(-H'(n)) > \mathrm{VA\_thr}(n) \\ 0, & \mathrm{if}(-H'(n)) \le \mathrm{VA\_thr}(n) \end{cases} \tag{13}$$

where $\mathrm{VA\_thr}(n)$ is adaptive to the environment variation.

The noise estimation (9) and entropy computation (12) use $4N_i$ (5.2 ms) and $8N_i$ (10.4 ms) data and modulus value average as power estimation to reduce the computational

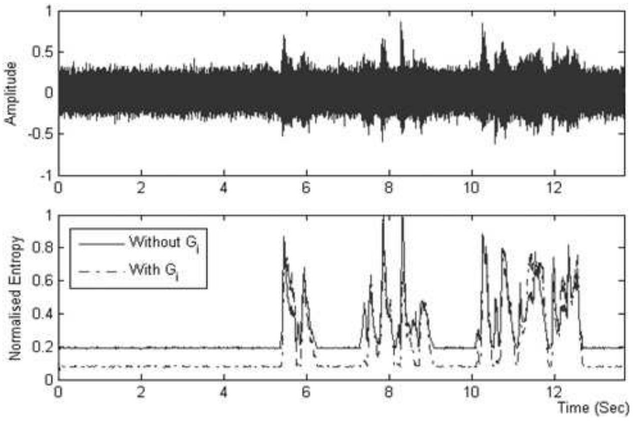


Fig. 3 *White noisy speech and resulted entropy with and without G_i at 0 dB SNR*

complexity for low-power consideration. According to our experience, the simplification only degrades 10% noise estimation and 3% VAD error which are not notable in the noise suppressed speech with proper parameter tuning.

The low-power design of the entropy-based VAD is through two simplifications as below. The first simplification, denoted as the multiplication-based entropy VAD, replaces the complex logarithms and divisions in (10) and (11) by multiplications with the following three steps. First, $\log_{10} P'(n,i)$ can be expressed as

$$\log_{10} P'(n,\ i) = \log_2 P'(n,\ i) \times \log_{10} 2 \tag{14}$$

where $\log_{10} 2$ is a constant multiplication and can be ignored because each subband has the same factor. Second, $\log_2 P'(n,i)$ computation can be approximated by the Mitchell's algorithm [33] as

$$\log_2 P'(n,\ i) = \log_2\big(2^q(1+r)\big) \simeq q + r \tag{15}$$

where $P'(n,\ i) = 2^q\ (1+r)$ is implemented by a lookup table. Finally, let $P'(n,i)$ in (11) as

$$\begin{aligned} P'(n,\ i) &= \frac{\big(Pa(n,\ i)G_i + K\big)}{\sum_{i=F22}^{F39} (Pa(n,\ i)G_i + K)} \\ &= \frac{\text{num\_}P'(n,\ i)}{\text{den\_}P'(n)} \end{aligned} \tag{16}$$

Substitute (16) into (11) and rewrite it as

$$\begin{aligned} &\text{den\_}P'(n) \times H'(n) \\ &= -\sum_{i=F22}^{F39} \big(\text{num\_}P'(n,\ i) \\ &\quad \times \big[\log_2\big(\text{num\_}P'(n,\ i) - \log_2\big(\text{den\_}P'(n)\big)\big)\big]\big) \end{aligned} \tag{17}$$

Thus, the conditional expression in (13) can be derived as

$$\text{VA}(n) = \begin{cases} 1, & \text{if den\_}P'(n) \times (-H'(n)) \\ & > \text{den\_}P'(n) \times \text{VA\_thr}(n) \\ 0, & \text{if den\_}P'(n) \times (-H'(n)) \\ & \le \text{den\_}P'(n) \times \text{VA\_thr}(n) \end{cases} \tag{18}$$

With (17) and (18), the VA(n) can be calculated without divisions and logarithms.

The second simplification, denoted as the preprocessed input data reuse scheme, exploits the common absolute value of input data at the noise power estimation (9), signal power calculation (12), spectral subtraction (5) and attenuation (6) use the absolute value of input data for processing. In addition, integration in (9) and (12) use almost the same data except different integration lengths. Therefore reusing the absolute value and partial integration results can significantly save computation and storage access.

This scheme is based on sign-magnitude, sliding window and partial integration. First, the incoming two's complement subband data are converted to the sign-magnitude format, where the magnitude is stored into a register as (19)

$$\boldsymbol{a}(n,\ i) = \big|a(m,\ i)\big| \quad m = n,\ n-1,\ \ldots,\ n - 8N_i + 1 \tag{19}$$

In contrast, the sign bit is stored into a buffer to generate the final denoised output.

Second, intermediate results can be reused by using a sliding window with the partial integration. The integration in noise estimation (9) calculates once every N samples using $4N$ samples data, whereas the integration in entropy (12) operates once every $4N$ samples using $8N$ samples data. To share those computations, the integration result of each N samples data are stored into registers as (19) for cross-N integration

$$\begin{aligned} &\text{sum\_a}(n,\ i) \\ &= \Bigg\{ \sum_{m=n-N_i+1}^{n} \big|a(m,\ i)\big|,\ \sum_{m=n-2N_i+1}^{n-N_i} \big|a(m,\ i)\big|,\ \ldots, \\ &\quad \times \sum_{m=n-8N_i+1}^{n-7N_i} \big|a(m,\ i)\big| \Bigg\} \end{aligned} \tag{20}$$

By using (20), the integration of signal (12) can be simplified by two partial integrations as (21). The old term in (21) is replaced by new partial integration at each $4N$ samples with a sliding window approach

$$\begin{aligned} &\text{sum\_a}_{\text{VAD}}(n,\ i) \\ &= \Bigg\{ \sum_{m=n-4N_i+1}^{n} \big|a(m,\ i)\big|,\ \sum_{m=n-8N_i+1}^{n-4N_i} \big|a(m,\ i)\big| \Bigg\} \end{aligned} \tag{21}$$

The integration in noise estimation (9) can also be simplified to N samples partial integrations as (22)

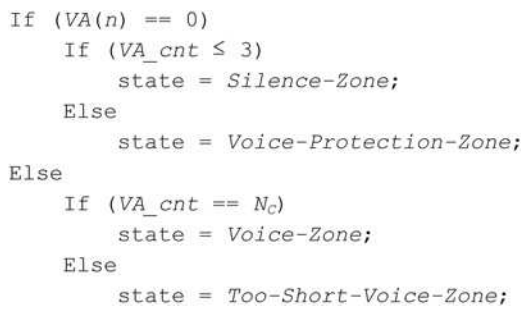
$$\begin{aligned} &\text{sum\_a}_{\text{NE}}(n,\ i) \\ &= \Bigg\{ \sum_{m=n-N_i+1}^{n} \big|a(m,\ i)\big|,\ \sum_{m=n-2N_i+1}^{n-N_i} \big|a(m,\ i)\big|,\ \ldots, \\ &\quad \times \sum_{m=n-4N_i+1}^{n-3N_i} \big|a(m,\ i)\big| \Bigg\} \end{aligned} \tag{22}$$

2.3.2 VAD decision:

Since speech will last for a period of time, we use a counter VA_cnt to work as a confidence counter of voice activity to decide signal state for NR: If VA(n) = 1, VA_cnt will be increased by one until it reaches to N_{C}. Otherwise VA_cnt will be decreased by one until it reaches to zero. N_{C} is an empirical constant and is eight in this paper.

The VAD decision rule shown in Fig. 4 classifies input into four zones to improve speech quality and noise estimation.

For current $4N$ samples with strong entropy, that is, VA(n) = 1, the state of this period will be decided as the *Voice-Zone*



```
If (VA(n) == 0)
    If (VA_cnt ≤ 3)
        state = Silence-Zone;
    Else
        state = Voice-Protection-Zone;
Else
    If (VA_cnt == N_C)
        state = Voice-Zone;
    Else
        state = Too-Short-Voice-Zone;
```

Fig. 4 *The VAD decision rule*

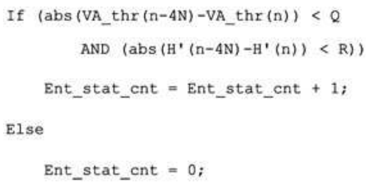
if strong entropy has persisted for a long period (VA_cnt = N_C). Otherwise, the short period ($0 <$ VA_cnt $< N_C$) having strong entropy is treated as the *Too-Short-Voice-Zone*. In contrast, the current $4N$ samples that have small entropy, that is, VA(n) = 0, will be viewed as noise. However, VAD will treat those noise periods (at most $N_C \times 4N$ samples) following the *Voice-Zone* as the *Voice-Protection-Zone* to protect the small volume vowel or consonant samples from over-cancellation. In other words, N_C defines a guarded duration between true speech and true noise, which is set as eight (42.7 ms or 1024 samples for 24 kHz sampling rate) based on our observation that signal having such long and strong entropy is probably speech. We use the same setting for the *Voice-Protection-Zone*. The noise periods except the *Voice-Protection-Zone* are regarded as the *Silence-Zone*. In this state, signal is regarded as noise only and noise estimator (9) will update its noise power estimation.

For *Voice-Zone* and *Voice-Protection-Zone*, input signal is processed by the spectral subtraction as (1) to improve speech quality. For *Too-Short-Voice-Zone* and *Silence-Zone*, input is attenuated using (6), where the gain *Att* are implemented by right shift 4 and 5 bit for *Too-Short-Voice-Zone* and *Silence-Zone*, respectively.

2.4 Adaptive threshold

The VAD threshold VA_thr(n) is adaptive to environment conditions with two steps as below. First, we use a counter *Ent_sta_cnt* as a confidence counter to monitor the stationary condition, distance between entropy *H'(n)* and threshold VA_thr(n), as the procedure shown in Fig. 5:

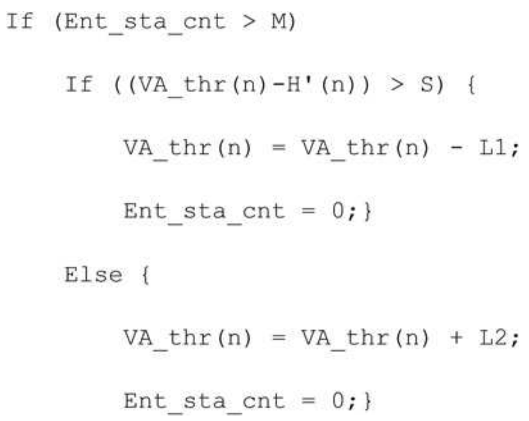
In which, VA_thr($n-4N$) and $H'(n-4N)$ are the threshold and entropy estimation of previous computation, respectively. Q and R are pre-defined stationary regions for threshold and entropy and are set to 0.03 empirically.



```
If (abs(VA_thr(n-4N)-VA_thr(n)) < Q
      AND (abs(H'(n-4N)-H'(n)) < R))
   Ent_stat_cnt = Ent_stat_cnt + 1;
Else
   Ent_stat_cnt = 0;
```

Fig. 5 *The first step of VAD adaptive threshold*

Second, if the stationary condition is kept for M computations (M is 24 in our case), VA_thr(n) will be adaptively adjusted according to whether it is larger than $H'(n)$ as shown in Fig. 6:



```
If (Ent_sta_cnt > M)
   If ((VA_thr(n)-H'(n)) > S) {
      VA_thr(n) = VA_thr(n) - L1;
      Ent_sta_cnt = 0;}
   Else {
      VA_thr(n) = VA_thr(n) + L2;
      Ent_sta_cnt = 0;}
```

Fig. 6 *The second step of VAD adaptive threshold*

where M and S are selected according to the environment. $L1$ and $L2$ are step sizes for attack and recovery based on the requirement of adjusting speed. In our case, S is set to 0.02, whereas $L1$ and $L2$ are 0.002 and 0.001, respectively. Above procedure is to gradually adjust VA_thr(n) until it is larger than $H'(n)$ by S. When VA_thr(n) updates, *Ent_sta_cnt* will be reset to zero for next stationary detection.

M and R are designed based on our observation that the entropy of speech tends to have large variation in a short period. By using this feature, the signal probably belongs to stationary noise if its entropy almost remains constant for a long period. Proper settings for M and R make the adaptive threshold tend to update threshold in the noise region according to our experience.

2.5 Off mechanism

Spectral subtraction and attenuation for the high SNR condition are unnecessary because of small noise amount. Under such condition, the spectral subtraction and attenuation will generate speech distortion. In addition, those operations also consume unnecessary power which cannot be neglected. To solve this problem, an off mechanism is designed to turn off the spectral subtraction and attenuation if the noise estimation in each subband is lower than a pre-defined threshold for a long period (controlled by a 4 bit counter). For our design, the pre-defined threshold is simply based on the noise estimation of each subband and its corresponding factor G_i as below

$$\begin{aligned} P_{\text{off}}(n,\ i) &= Pn(n,\ i) \times G_i \\ &= \left(\frac{1}{4N_i} \sum_{m=n-4N_i+1}^{n} \left| a(m,\ i) \right| \right) \times G_i, \qquad (23) \\ n &= 0,\ N,\ 2N,\ \ldots \end{aligned}$$

If the $P_{\text{off}}(n,\ i)$ of each subband is less than L_{off}, the 4 bit counter will be added by one. L_{off} is a constant based on the maximum acceptable noise level. In our case, L_{off} is four according to the noise level of 15 dB SNR. When the counter is accumulated to 15 (about 20 ms), the gating function is activated to turn off the spectral subtraction and attenuation.

2.6 Complexity comparison

For hearing aids, power consumption is the main objective to be minimised. To fairly compare our method and other approaches without loss of generality, we use the computational complexity of different operations such as multiplications (MUL), divisions (DIV), logarithms (LOG), square roots (SQRT) and additions (ADD), with its corresponding power metrics for evaluation. According to our analysis, the power consumption of a 16-by-16 bit multiplier is about 18 times the power consumption of a 16 bit adder. The power consumption of a constant multiplication is equivalent to four ADDs and shifts in average, and one shift operation (using barrel shifter) consumes similar power as the power dissipation of one ADD. Regarding to DIV, it has larger but the same order complexity as MUL does. Furthermore, LOG and SQRT have 11 and 4 times more complexity in comparison with MUL [34, 35], respectively. Moreover, the power consumption of the table lookup in multiplication-based entropy VAD is equivalent to four ADDs.

Table 1 Comparison of averaged complexity per input sample based on the number of ADD, MUL, DIV, LOG and SQRT

	Arithmetic operations					Normalised estimated power
	ADD	MUL	DIV	LOG	SQRT	
specsub [12]	4.0	13.0	0	0	1.0	4.76
mband [16]	13.0	22.0	2.0	1.0	1.0	12.94
our design	48.4	0.3	0	0	0	1

Table 1 shows the complexity comparison with other approaches. To fairly compare frame and sample-based methods, the complexity of all frame-based methods are estimated for each frame and then normalised by frame length to estimate the averaged complexity per input sample. As to our design, the complexity is estimated for each new input sequence (e.g. 128 samples for VAD and 32 samples for noise power estimation) and then normalised by its sequence length as well to estimate the averaged complexity per input sample. The complexities of *specsub* and *mband* are estimated according to the default setting (e.g. overlap ratio is 50%) of MATLAB programmes provided by [1]. Additionally, for fair comparison, complexity of the frequency decomposition and reconstruction (including magnitude and phase computation) is excluded, whereas VAD complexity is included. According to this table, our complexity is much lower than others. In terms of normalised power consumption, our design only consumes about 7.7 and 21% estimated power compared with those of *mband* and *specsub*, respectively. Therefore our design is more suitable for the hearing aid applications.

3 Performance analysis

In the following simulation, we adopt 27 Mandarin Chinese two-characters for clean speech sources. Those two-characters are originally chosen from the Academia Sinica Balanced Corpus of Modern Chinese database [36] based on phonetic balance, tonal balance and familiarity for the intelligibility test of hearing aids under Mandarin Chinese environment [37]. The clean sources are concatenated into four sentences and corrupted by white, babble, factory and car noise from NOISEX-92 database [38] for stimuli. The stimuli are corrupted with 0 to 12 dB SNR. System sampling rate is 24 kHz. Furthermore, the IG and WDRC are disabled to show the performance of noise suppression.

3.1 VAD performance

Fig. 7 displays the accuracy of our sample-based VAD and FFT-based entropy VAD [22] whose frame length is $8N$ (or 256) for speech at five different input SNRs and under four noisy environments. The accuracy of the sample-based VAD is ranged from 60 to 90% for different SNR and noise conditions. Both VAD methods have their best performance under the white noise condition. For the non-white noise environment, the sample-based VAD performs well for signals over 3 dB SNR compared with the FFT-based one because of adaptive threshold. Moreover, the ANSI filter bank attenuates noise outside its passband, thereby improving the SNR of input signal before voice detection, especially for car noise. However, since the filter bank only provides 18 subbands, which is much less than the resolution of FFT, the detection accuracy of the sample-based VAD is susceptible to large background noise and becomes lower for low SNR environment.

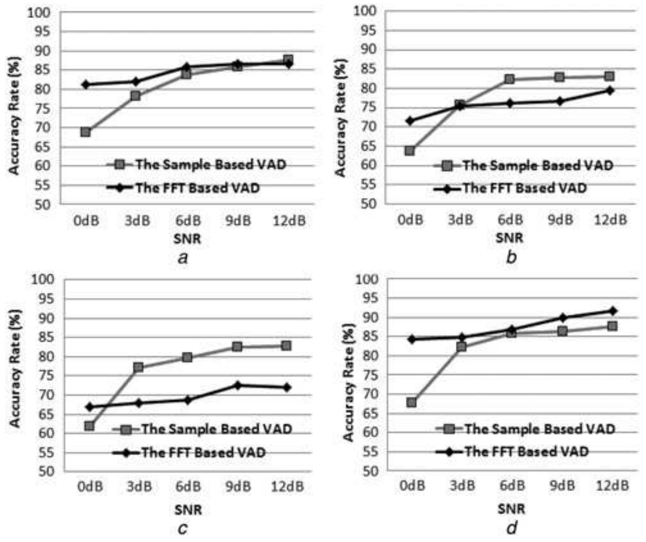


Fig. 7 *Comparison between our sample-based and the FFT-based entropy VAD [22] under*

a White
b Babble
c Factory
d Car noise condition

3.2 NR performance

Fig. 8 illustrates the performance evaluation of different NR methods listed in Table 2 with composite measure [1] under white, babble, factory and car condition. Except our design, the programmes of the six methods are from [1] without parameter tuning. According to these programmes, most of all adopt 20 ms for frame length and 10 ms for overlap size, whereas *pklt* uses 32 ms frame length and 16 ms overlap size. The adopted composite measure combines segmental SNR, formant information and perceptual assessment into one measurement. Therefore the composite measure has the best correlation to subjective evaluations among all of the objective evaluations presented in [1]. The index of composite measure is from one (worst) to five (best). The composite measure provides three scores, overall quality (OVL), signal distortion (SIG) and background intrusiveness (BAK).

Fig. 8*a* shows the OVL score under the white noise condition. Here only OVL is shown to illustrate the trend of performance. Under the white noise environment which has uniform and stationary nature being easy to predict, all methods can provide quality improvement, where *mmse* has the best performance because its high complexity method provides better attenuation curve to significantly reduce artefact. Among the three spectral subtraction methods (*specsub, mband* and our design), our design can perform more improvement for the low SNR (0 and 3 dB) input because of the voice protection of the VAD decision, smooth noise estimation and perceptual spectral subtraction. At the high SNR (12 dB) input, our design offers similar OVL compared with *mband*, *mt_mask* and *pklt*.

With respect to non-stationary environment, such as babble, factory and car noise, all methods have similar

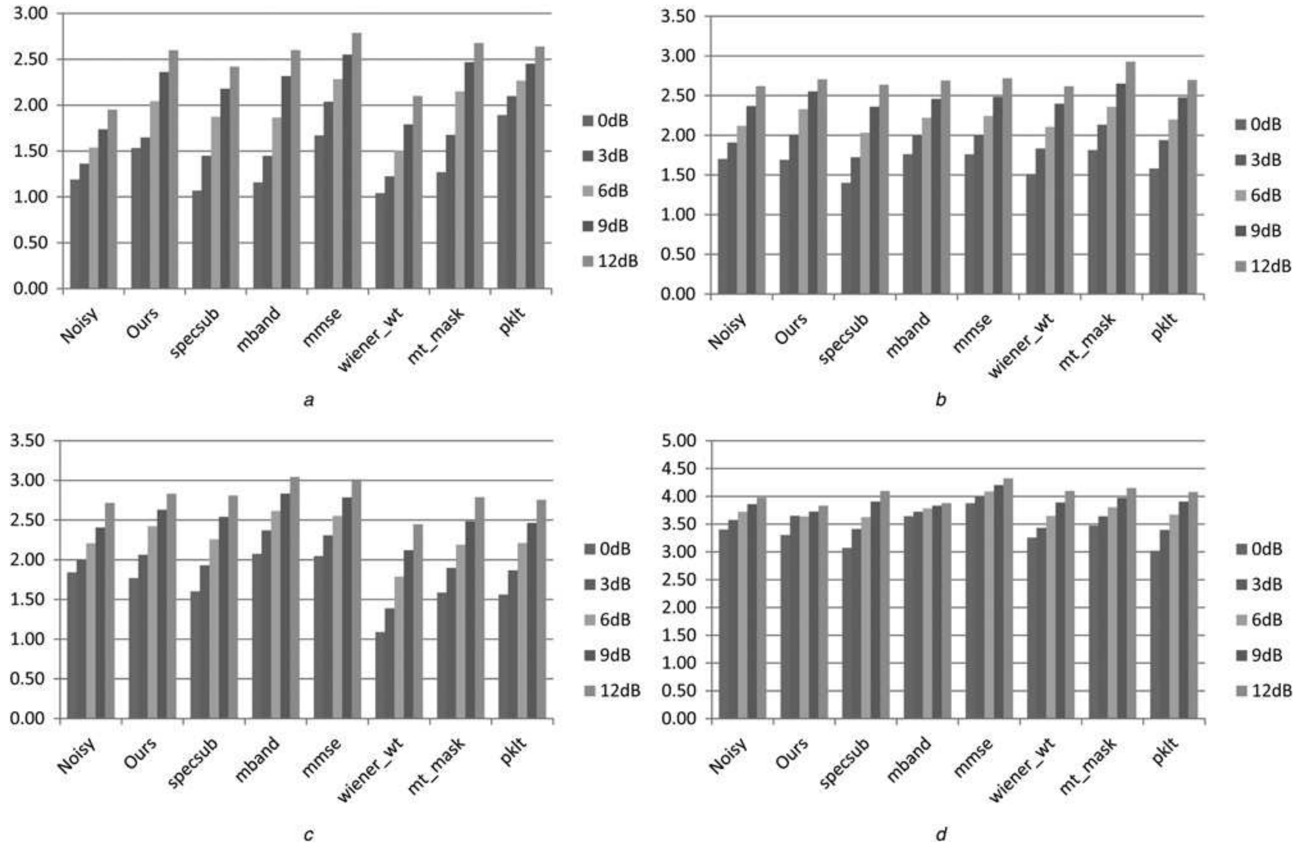


Fig. 8 *Comparison of OVL under*

a White
b Babble
c Factory
d Car noise condition

OVL score, but cannot provide significant improvement, because the non-stationary nature degrades VAD and noise estimation performance. In particular, under the babble condition having similar spectrum distribution as that of a speech signal, all tested methods obtain the worst performance, as shown in Fig. 8*b*. Among all methods, *mt_mask* has the best OVL at the babble noise, whereas our design can offer similar quality compared with *mband*, *mmse* and *pklt*. Regarding to the factory OVL shown in Fig. 8*c*, most methods can provide quality improvement, whereas *mband* and *mmse* have the best improvement. Our design performs better improvement than *specsub* except at the 0 dB SNR input because of poor VAD accuracy. For the car environment shown in Fig. 8*d*, most methods have good quality because the noise mainly distributes at the low frequency. Under such environment, our design is still better than *specsub* for the 0 and 3 dB SNR input and has similar performance as *mband* when above 3 dB SNR. In the car noise condition, *mmse* can still provide the best improvement.

Table 2 NR methods for speech quality comparison

Methods	Decompositions	Noise reduction approaches
our design	filter bank	spectral subtraction
specsub [12]	FFT	
mband [16]	FFT	
mmse [39]	FFT	statistical model based
wiener_wt [40]	DWT	
mt_mask [41]	FFT	
pklt [42]	eigen decomposition	subspace

Fig. 9 shows the spectrogram of input and output for our design, *specsub*, *mband* and *mmse* under white noise condition. Regarding to our design, as shown in Fig. 9*c*, noise in the silence region can be significantly reduced by spectral attenuation, whereas noise in the speech region are suppressed if it does not belong to a speech structure by multiband subtraction. In particular, the main parts of speech, such as harmonics, are preserved. Our design generates distortion at the onset of each word, since VAD cannot respond to it immediately. This problem could be reduced by tuning the duration of the *Too-Short-Voice-Zone* with the performance degradation of high variation noise.

The arrows in Fig. 9*c* indicate the musical noise [1] resulted from our design. The musical noise is the short term peaks in spectrogram that are produced by spectral subtraction. Since many peaks may occur at the same time, the total distortion is some kind of mixture of many tones. Our design usually has musical noise below 1 kHz, since oversubtraction is not applied to those subbands (μ_i is unity). This problem can be reduced by increasing oversubtraction.

The four methods shown in Fig. 9 suppress noise with different output spectrograms owing to their different strategies. Regarding to the noise region, *specsub* has the smallest noise energy while *mband* has the largest. However, the noise region of *specsub* is riddled with

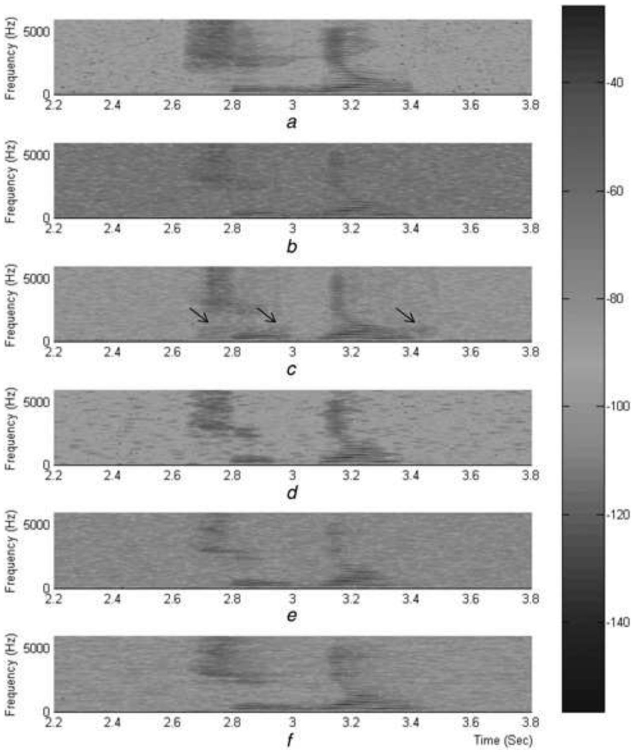


Fig. 9 *Spectrogram under the white noise condition*

a Clean speech
b 6 dB input noisy speech and the processed speech by
c Our design
d specsub
e mband
f mmse

musical noise. The high noise level of *mband* masks musical noise, but leads to more serious speech masking. Compared with *specsub* and *mband*, our design can provide a relatively clean noise region which is similar to that of *mmse*. Regarding to the speech region, *mmse* can provide the best performance. Among the three spectral subtraction methods, our design can preserve speech well compared with *specsub* and *mband* at the expense of more artefacts in the speech region.

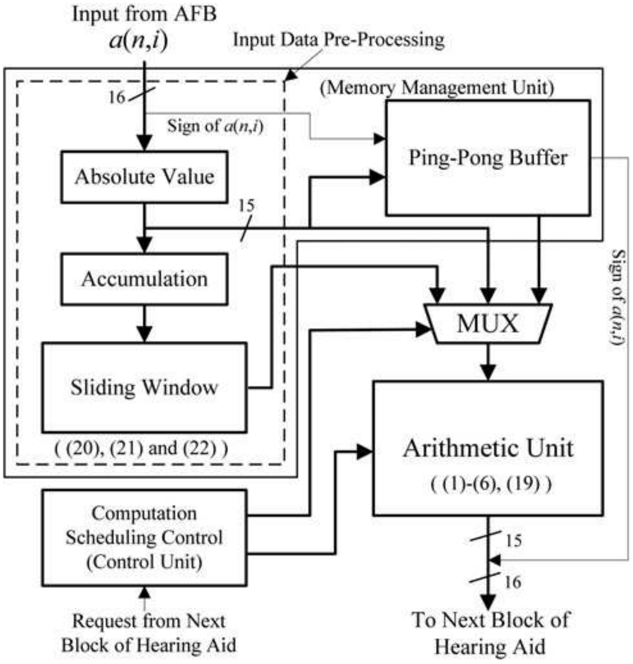


Fig. 10 *Proposed hardware architecture with their corresponding equations*

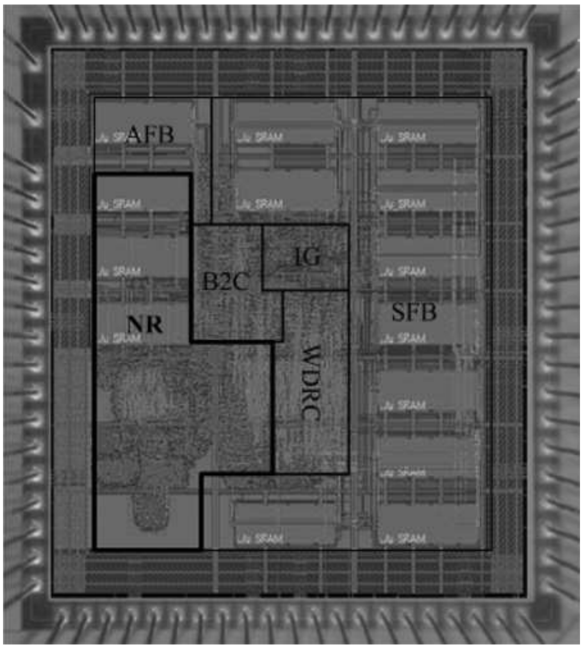


Fig. 11 *Die photo of our hearing aid chip*

The different delays between AFB subbands stretch signal energy and thus have impact on NR performance. However, in our NR algorithm, this effect is very little because the stretched signal makes VAD easier to enter *Voice-Zone* and *Voice-Protection-Zone* which preserve more speech for low distortion at expense of more residue noise.

4 Implementation result

The sample-based *mband* and VAD are implemented with a core-based architecture instead of a dedicated hardware, owing to the low data rate of speech signal and the consideration for software parameter tuning. Fig. 10 shows the block diagram of the proposed core-based architecture consisting of a memory management unit (including an input data preprocessing block and a ping-pong buffer) for data exchange, a computation scheduling control for control unit, and an arithmetic unit for arithmetic computation. The input data are first processed by the preprocessing block with the proposed data reuse scheme, and the results are

Table 3 Implementation summary

process	90 nm 1.0 V CMOS[*] HVT
gate count	101 697 (including SRAM[†]) 77 783 (excluding SRAM)
die area	1720 μm × 892 μm (exclude IO[‡])
clock rate	6 MHz
data rate	24 kHz
latency	1.3 ms
total power consumption	83.7 μW at 0.6 V
dynamic power consumption	39.1 μW at 0.6 V
leakage power consumption	44.6 μW at 0.6 V

[*] Complementary metal oxide semiconductor
[†] Static random-access memory
[‡] Input and output pads

sent into a ping-pong buffer for further computation. The following computations are implemented by the arithmetic unit controlled by the control unit.

Our NR design is integrated with other components as a completed hearing aid chip [28] by using cell-based flow and 90 nm high V_T complementary metal oxide semiconductor (CMOS) library with gated clock and low-power SRAM [43]. Fig. 11 depicts the die photo, where all abbreviations are shown in Fig. 1.

Table 3 shows the implementation summary. The power consumption is estimated according to total power consumption measured by the real chip (including the SRAM power), and the proportion of NR power to total chip power according to simulation. The power reduction is mainly because of the proposed low-power approach. Comparison for area, power and latency to other single microphone implementations is difficult to perform, because of diverged subband decomposition method, subband number, NR algorithm, implementation style and lack of speech quality measurement.

5 Conclusion

This paper presents a sample-based perceptual multiband spectral subtraction and a multiplication-based entropy VAD to meet the strict low-power and low-latency requirement of CIC hearing aids. With above approaches and further low-power simplification, our design has similar sound quality in terms of composite measure, but only requires 1.3 ms processing latency and about 7.7% estimated power owing to lower complexity compared with the conventional multiband methods. The final chip implementation shows that it dissipates 83.7 μW at the 0.6 V operation, and thus is suitable to be integrated to the hearing aid chips.

6 Acknowledgment

The authors would like to thank the UMC Inc., the National Chip Implementation Center and the National Science Council, R.O.C. under grant NSC98-2220-E-009-008. The authors would also like to thank Professor S. T. Young for providing speech database as well as Professor T. S. Chi for constructive discussion.

7 References

1 Loizou, P.C.: ‘Speech enhancement, theory and practice’ (CRC Press, 2007, 1st edn.)

2 Kates, J.M.: ‘Digital hearing aids’ (Plural Publishing, 2008, 1st edn.)

3 Clarkson, P.M., Bahgat, S.F.: ‘Envelope expansion methods for speech enhancement’, *J. Acoust. Soc. Am.*, 1991, **89**, (3), pp. 1378–1382

4 Schaub, A.: ‘Digital hearing aids’ (Thieme Medical Publishers, 2008, 1st edn.)

5 Kates, J.M.: ‘Signal processing for hearing aids’, *Hear. Instrum.*, 1986, **37**, pp. 19–21

6 Gingsjo, A.L.: ‘On transient noise and its reduction in hearing aids’, PhD thesis, Department of Applied Electronics, Chalmers University of Technology, 1997

7 Hermansky, H., Morgan, N.: ‘RASTA processing speech’, *IEEE Trans. Speech Audio Process.*, 1994, **2**, (4), pp. 578–589

8 Schaub, A., Straub, P.: ‘Spectral sharpening for speech enhancement/noise reduction’. Proc. Int. Conf. Circuit Systems, Toronto, Ontario, Canada, May 1991, pp. 993–996

9 Lim, J.S., Oppenheim, A.V.: ‘All-pole modeling of degraded speech’, *IEEE Trans. Acoust. Speech Signal Process.*, 1978, **ASSP-26**, (3), pp. 197–210

10 Levitt, H., Bakke, M., Kates, J., Neuman, A., Schwander, T., Weiss, M.: ‘Signal processing for hearing impairment’, *Scand. Audiol.*, 1993, **38**, pp. 7–19

11 Boll, S.F.: ‘Suppression of acoustic noise in speech using spectral subtraction’, *IEEE Trans. Acoust. Speech Signal Process.*, 1979, **27**, (2), pp. 113–120

12 Berouti, M., Schwartz, M., Makhoul, J.: ‘Enhancement of speech corrupted by acoustic noise’. Proc. Int. Conf. Acoustics, Speech, Signal Processing, Washington, DC, USA, April 1979, pp. 208–211

13 Lim, J.S., Oppenheim, A.V.: ‘Enhancement and bandwidth compression of noisy speech’, *Proc. IEEE*, 1979, **67**, pp. 113–120

14 Sim, B.L., Tong, Y.C., Chang, J.S., Tan, C.T.: ‘A parametric formulation of the generalized spectral subtraction’, *IEEE Trans. Speech Audio Process.*, 1998, **6**, (4), pp. 328–337

15 Hu, H.T., Yu, C.: ‘Adaptive noise spectral estimation for spectral subtraction speech enhancement’, *IET Signal Process.*, 2007, **1**, (3), pp. 156–163

16 Kamath, S., Loizou, P.: ‘A multi-band spectral subtraction method for enhancing speech corrupted by colored noise’. Proc. IEEE Int. Conf. Acoustics Speech Signal Processing, Orlando, FL, USA, May 2002, p. 4164

17 Udrea, R.M., Vizireanu, N., Ciochina, S., Halunga, S.: ‘Nonlinear spectral subtraction method for colored noise reduction using multi-band Bark scale’, *Signal Process.*, 2008, **88**, (5), pp. 1299–1303

18 Rama Rao, C.V., Rama Murthy, M.B., Srinivasa Rao, K.: ‘Noise reduction using Mel-SCALE spectral subtraction with perceptually defined subtraction parameters – a new scheme’, *Signal Image Process.*, 2011, **2**, (1), pp. 135–149

19 DiGiovanni, J.J.: ‘Hearing aid handbook 2011’ (Delmar, Cengage Learning, 2011, 1st edn.)

20 Dillon, H.: ‘Hearing aids’ (Boomerang Press, Sydney, Thieme, 2000)

21 Stone, M.A., Moore, B.C.J.: ‘Tolerable hearing aid delays. II. Estimation of limits imposed during speech production’, *Ear Hear.*, 2002, **23**, (4), pp. 325–338

22 Jia, C., Xu, B.: ‘An improved entropy-based endpoint detection algorithm’, Int. Symp. Chinese Spoken Language Processing, Taipei, Taiwan, ROC, August 2002, pp. 96–99

23 Wu, B.F., Wang, K.C.: ‘Robust endpoint detection based on the adaptive band partitioning spectral entropy in adverse environments’, *IEEE Trans. Speech Audio Process.*, 2005, **13**, (5), pp. 762–775

24 Wei, C.W., Tsai, C.C., Chang, T.S., Jou, S.J.: ‘Perceptual multiband spectral subtraction for noise reduction in hearing aids’. Proc. IEEE Asia Pacific Conf. Circuits Systems, Kuala Lumpur, Malaysia, December 2010, pp. 692–695

25 Chang, J.H., Tsai, K.S., Li, P.C., Young, S.T.: ‘Computer-aided simulation of multi-channel WDRC hearing aids’. 17th Annual Convention & Expo of the American Academy of Audiology, Reston, Virginia, USA, March 2005

26 ANSI S1.11–2004: ‘Specification for octave-band and fractional-octave-band analog and digital filters’, 2004

27 Wassner, J., Kaeslin, H., Felber, N., Fichtner, W.: ‘Waveform coding for low-power digital filtering of speech data’, *IEEE Trans. Signal Process.*, 2003, **51**, (6), pp. 1656–1661

28 Wei, C.W., Kuo, Y.T., Chang, K.C., Tsai, C.C., Lin, J.Y., FanJiang, Y., Tu, M.H., Liu, C.W., Chang, T.S., Jou, S.J.: ‘A low-power Mandarin-specific hearing aid chip’. Proc. IEEE Asian Solid-State Circuits Conf., Beijing, China, November 2010, pp. 333–336

29 Kuo, Y.T., Lin, T.J., Li, Y.T., Liu, C.W.: ‘Design and implementation of low-power ANSI S1.11 filter bank for digital hearing aids’, *IEEE Trans. Circuits Syst. I*, 2010, **57**, (7), pp. 1684–1696

30 Kuo, Y.T., Lin, T.J., Chang, W.H., Liu, Y.T., Liu, C.W.: ‘Complexity-effective auditory compensation for digital hearing aids’. Proc. Int. Symp. Circuits and Systems, Seattle, Washington, USA, May 2008, pp. 1472–1475

31 Chang, K.C., Kuo, Y.T., Lin, T.J., Liu, C.W.: ‘Complexity-effective dynamic range compression for digital hearing aids’. Proc. IEEE Int. Symp. Circuits and Systems, Paris, France, June 2010, pp. 2378–2381

32 Kuo, Y.T.: ‘Low-power auditory compensation for digital hearing aids’, PhD dissertation, Institute Electronics Engineering, National Chiao-Tung University, 2011

33 Mitchell, J.N.: ‘Computer multiplication and division using binary logarithms’, *IRE Trans. Electron. Comput.*, 1962, **EC-11**, (4), pp. 512–517

34 Venkat, K.: ‘Efficient multiplication and division using MSP430’ (Texas Instruments, 2006)

35 Jiang, J.: ‘General guide to implement logarithmic and exponential operations on a fixed-point DSP’ (Texas Instruments, 1999)

36 Available at http://www.rocling.iis.sinica.edu.tw/CKIP/engversion/20corpus.htm, accessed June 2013

37 Chang, J.H.: ‘Effect comparison of hearing aids prescriptions on Mandarin speech perception’, Master thesis, Institute Biomedical Engineering, National Yang-Ming University, 2005

38 Available at http://www.spib.rice.edu/spib/select_noise.html, accessed July 2012

39 Ephraim, Y., Malah, D.: 'Speech enhancement using a minimum mean square error short time spectral amplitude estimator', *IEEE Trans. Acoust. Speech Signal Process.*, 1984, **32**, (6), pp. 1109–1121

40 Hu, Y., Loizou, P.: 'Speech enhancement based on wavelet thresholding the multitaper spectrum', *IEEE Trans. Speech Audio Process.*, 2004, **12**, (1), pp. 59–67

41 Hu, Y., Loizou, P.: 'Incorporating a psychoacoustical model in frequency domain speech enhancement', *IEEE Signal Process. Lett.*, 2004, **11**, (2), pp. 270–273

42 Jabloun, F., Champagne, B.: 'Incorporating the human hearing properties in the signal subspace approach for speech enhancement', *IEEE Trans. Speech Audio Process.*, 2003, **11**, (6), pp. 700–708

43 Tu, M.H., Lin, J.Y., Tsai, M.C., Jou, S.J., Chuang, C.T.: 'Single-ended subthreshold SRAM with asymmetrical write/read-assist', *IEEE Trans. Circuits Syst. I*, 2010, **57**, (12), pp. 3039–3047