The Effect of the Speech Enhancement Algorithm for Sensorineural Hearing Impaired Listeners

Dong Wook Kim, Young Woo Lee[1], Jong Shill Lee, Young Joon Chee, Sang Min Lee[2], In Young Kim, Sun I. Kim

Department of Biomedical Engineering, Hanyang University, Seoul, KOREA
[1]Digital Media R&D Center, SAMSUNG ELECTRONICS, CO., LTD., KOREA
[2]School of Electrical Engineering, Inha University, Incheon, KOREA
(Received July 5, 2007. Accepted November 5, 2007)

Abstract

Background noise is one of the major complaints of not only hearing impaired persons but also normal listeners. This paper describes the results of two experiments in which speech recognition performance was determined for listeners with normal hearing and sensorineural hearing loss in noise environment. First, we compared speech enhancement algorithms by evaluation speech recognition ability in various speech-to-noise ratios and types of noise. Next, speech enhancement algorithms by reducing background noise were presented and evaluated to improve speech intelligibility for sensorineural hearing impairment listeners. We tested three noise reduction methods using single-microphone, such as spectrum subtraction and companding, Wiener filter method, and maximum likelihood envelop estimation. Their responses in background noise were investigated and compared with those by the speech enhancement algorithm that presented in this paper. The methods improved speech recognition test score for the sensorineural hearing impaired listeners, but not for normal listeners. The results suggest the speech enhancement algorithm with the loudness compression can improve speech intelligibility for listeners with sensorineural hearing loss.

Key words : speech enhancement, noise reduction, digital hearing aid, spectral subtraction, companding

I. INTRODUCTION

Hearing aids are currently the primary means of treating hearing impairment, a disorder that affects up to 10% of the world's population. However, the availability of these newer, high technology digital hearing aids required that we justify their value related. The hope is that the new possibilities would open up a market that has very little penetration, 10% to 20% of the hearing impaired use hearing aids, and satisfaction, about 45% of hearing aid user that have stopped using them give reasons revolving around the standard hearing aids ineffectiveness in increasing intelligibility.

The objectives of this study are to find out what is the most compliant from hearing impaired listeners and how we can compensate theses issues. Digital Signal Processing (DSP) hearing aid devices employ sophisticated miniature digital electronic circuits to amplify and process sound using signal processing algorithms that are adjusted for each individual patient. Recent development of digital technology has offered new possibilities for noticeable advances of hearing aids. However, background and feedback noise are still one of the major complaints of the hearing impaired persons who use digital hearing aids [1]. Also, it is clear that sensorineural hearing impairment listeners may have a substantially reduced ability to understand speech in background noise [2]. The first explanation is audibility deficit to understand speech in noise by masking effect [3]. The other explanations of the problems of understanding in noise for persons with sensorineural hearing impairment are the loss of binaural "squelch" effect [4], the phenomenon of upward spread of masking [5], the effects of temporal smearing, and so on.

In mid 30's, loudness recruitment, the progressive alleviation of hearing impairment as the input level of sound increases, is discovered by audiologists. A normal cochlea is an active system with about 50 to 60dB amplification measured at threshold. When the outer hair cells are damaged, cochlear amplification decreases and the regulation process is

This study was supported by a Nano-bio technology development project, Ministry of Science & Technology, Republic of Korea. (2005-01249)
Corresponding Author : In Young Kim, M.D., Ph.D.
Dept. of Biomedical Engineering, Hanyang University Sungdong P.O. Box 55, Seoul, KOREA 133-605
Tel : 02-2291-1713 / Fax : 02-2296-5943
E-mail : iykim@hanyang.ac.kr

increasingly lost. Transmission lowers from a compressive to a linear character. With a loss of hearing over 60dB, the cochlear amplification is already inactive and the loss of the inner hair cells brings about a further linear shift of the hearing threshold. The amplification is regulated nonlinearly by the outer hair cells. Nonlinear compression digital hearing aids can compensate the loss of outer hair cells, the loss of sensitivity, by compressing the input sound. But only this kind of compression algorithm is not enough, especially, in the noisy environment. Since, this sensitivity loss and the speech-to-noise ratio (SNR) loss can be relatively independent quantity [6,7,8]. The hearing aid can not solve the SNR problem caused by loss of inner hair cells, and/or actual nerve deafness. Furthermore, noise has a different effect on speech intelligibility even among persons with similar hearing loss. Without increase the SNR, the amplification function of hearing aids has limited improvement of speech discrimination for persons with sensorineural hearing loss [9]. They need to not only hear sound better, but also understand speech better. We need to consider not only the SNR loss, but also the reduced dynamic range associated with loudness recruitment for listeners with sensorineural hearing loss simultaneously.

However, the physiological correlates of loudness and recruitment are still not well understood. Recent psychophysical results challenge some of the classical concepts of the recruitment, and suggest that the dynamic range of loudness is also reduced. These results have potentially significant implications for hearing aids that have renewed an interest in how sensorineural hearing loss affects level encoding in the auditory nerve. In this paper, we compared speech enhancement algorithms, i.e. such as spectrum subtraction and companding, Wiener filter method, and maximum likelihood envelop estimation, to improve the speech intelligibility in various noise environments and present the spectrum subtraction-based algorithm using the recent electrophysiology research on cochlear, especially for sensorineural hearing impairment listeners.

II. SPEECH ENHANCEMENT IN HEARING AID

Any speech enhancement algorithm, including those used in digital hearing aids, is based on a set of explicit or implicit assumptions about the acoustic characteristics of the background noise. These assumptions provide the criteria for determining when noise is present [10, 11]. The success of the speech enhancement algorithm depends primarily on the validity and robustness of these assumptions, and secondarily on the computational resources available to implement the method of noise reduction. There are two fundamentally different sets of assumptions that have been used for hearing aid speech enhancement or noise reduction.

One approach is based on the directional properties of sound. This approach assumes that noise sources have different spatial locations than desired signals. According to this approach, signals originate in front of the listener, and noises originate from other directions. If these assumptions are valid, directional microphones, spatial filtering algorithms, or spatially defined source separation algorithms may be effective noise reduction algorithm. These assumptions are likely to be valid in free space, in large enclosures, or in small enclosures with good sound absorbing surfaces. However, in reverberant spaces or in spaces of any type where there are noise sources surrounding the individual, the assumptions are not met, and, consequently, speech enhancement algorithm based on the directional properties of sound will have limited effectiveness. In addition, when desired signals originate from directions other than in front of the listener, these algorithms will make it more difficult to detect and recognize these signals.

It is important to note that when the directional sound

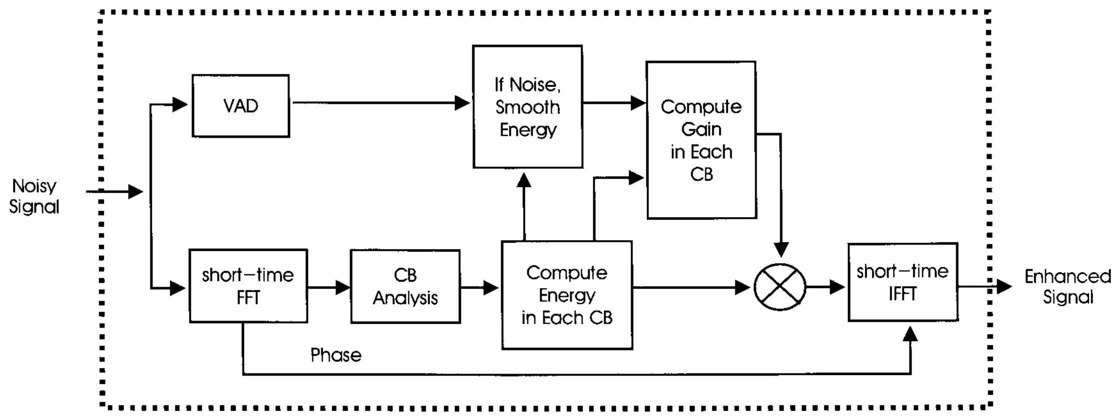

Fig. 1. A Block Diagram of the Speech Enhancement Algorithm in Frequency Domain.

assumptions are met, this class of noise reduction algorithms can actually improve the signal/noise ratio, even when the signal and noise with overlapping spectra are present at the same time, e.g., in a group of people talking. Thus, these algorithms can potentially improve speech intelligibility, as well as reducing the level of the noise.

The second approach is based on the spectral and/or temporal statistics of the noise and the desired signal. This approach assumes that the noise and desired speech signal have different spectral and temporal properties that can be detected and distinguished from the output patterns seen across the filter bands in a multiband hearing aid. When properties associated with noise are identified, the gain of one or more of the bands assumed to contain primarily noise is reduced. Likewise, when properties associated with speech are identified, the gain of bands assumed to contain important speech information is increased [12, 13]. There are several important strengths and limitations of speech enhancement algorithms based on these assumptions, just as there were for the directionality assumptions. These methods can work well for reducing the audibility and annoyance of stationary background noise, including hearing aid circuit noise, during intervals when no speech is present. However, these methods are not effective when the interfering noise has the same spectral and temporal properties as the desired speech signal, e.g., in a group of people talking. Nor will these methods be effective in improving intelligibility when speech and noise are present at the same time. In these situations, one or more frequency bands assumed to contain noise will be attenuated, but the level of the speech in those bands will also be attenuated without changing the signal/noise ratio. Also by computational limitation of digital hearing aid, only few of more complex speech enhancement algorithms are implant able in current hearing aid.

We evaluated these spectral enhancement algorithms i.e. the spectral enhancement, Wiener filtering, and maximum likelihood envelop estimation, and proposed a speech enhancement algorithm using compressing-and-expanding, i.e., companding, strategy for spectral enhancement algorithm based on formant enhancement methodology.

III. SPECTRAL ENHANCEMENT ALGORITHM

A speech enhancement algorithm based on spectral enhancement using compressing-and-expanding, i.e., companding, strategy for spectral enhancement algorithm based on formant enhancement methodology is shown on Figure 2.

Let $x(n)$ and $d(n)$ denote speech and uncorrelated additive noise signals, respectively, where n is a discrete-time index. The noise signal $y(n)$, given by $y(n)=x(n)+d(n)$, is divided into overlapping frames by the application of a window function and analyzed using the Fast Fourier Transform (FFT). This is defined by

$$Y(k,\ell)=\sum_{n=0}^{N-1} y(n+\ell R)h(n)e^{-j(2\pi/N)nk} \tag{1}$$

where k is the frequency bin index, ℓ is the time frame index, h is an analysis window of size N, and R is the framing step.

Spectral weighting $Q(k,\ell)$ is determined using the proposed subtraction rule and the estimated noise spectrum. Noise reduction is performed by applying a spectral weighting to spectral componentsof the noisy speech [14].

$$X(k,\ell)=Q(k,\ell)\,Y(k,\ell) \tag{2}$$

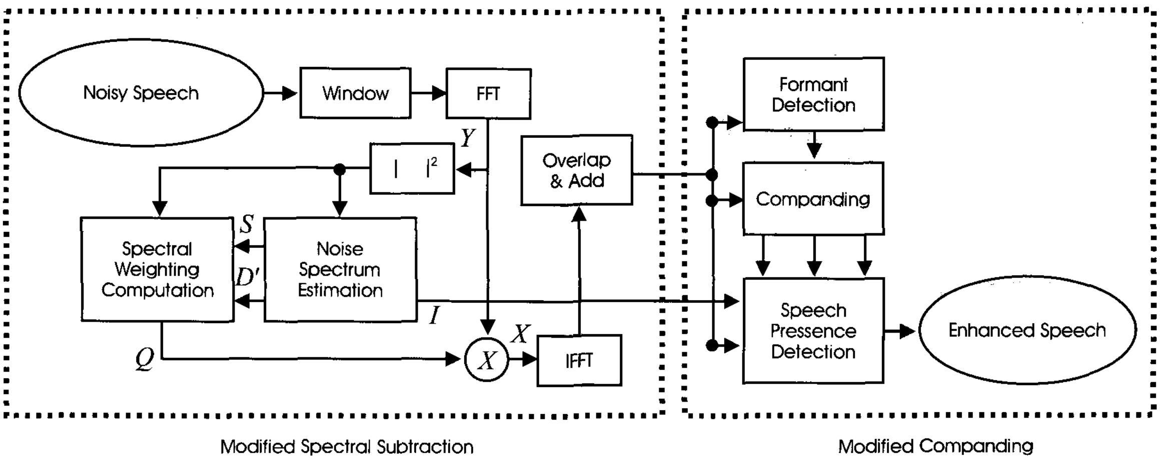

Fig. 2. Block diagram of the proposed speech enhancement method.

The noise-reduced speech $X(k, \ell)$ is converted back to the time domain using the inverse FFT(IFFT). The formant-enhanced signal $x''(n)$ is obtained by applying modified companding to the synthesized signal $x'(n)$ based on the speech presence function I.

$$\begin{aligned} x''(n) &= x'(n) & & if \sum(I) < \delta \quad (speech\ absent) \\ &= FE(x'(n)) & & else \quad (speech\ present), \end{aligned} \tag{3}$$

where δ is the speech presence threshold value. $I(k, \ell)$ denotes an indicator function for speech presence in each frequency bin, i.e., $I(k, \ell) = 1$ if speech is present and $I(k, \ell) = 0$ otherwise. FE is the algorithm of formant enhancement that is composed of modified companding, a formant detection algorithm and speech presence detection algorithms.

The sampling frequency of the proposed algorithm is 8kHz and the input segments with the length of 256 samples were weighted by Hamming window and overlapped of 128 samples. In order to reduce phase distortion due to IIR filter, we used zero-phase IIR filter method. Therefore, the total delay of the proposed algorithm is 16ms.

We applied noise estimation by minima controlled recursive averaging (MCRA) to track the noise spectrum [11]. The resultant noise spectrum D' is obtained by averaging past noise spectra using a smoothing parameter α_d' that is adjusted by the signal presence probability P defined by

$$D'(k, \ell) = \alpha_d' \cdot D(k, \ell - 1) + (1 - \alpha_d') \cdot S(k, \ell), \tag{4}$$

where $\alpha_d'(k, \ell) = \alpha_d + (1 - \alpha_d) p(k, \ell)$ is a time-varying smoothing parameter.

The local energy of the noisy speech S is obtained by a first order recursive averaging, given by

$$S(k, \ell) = \alpha_s \cdot S(k, \ell - 1) + (1 - \alpha_s) \cdot |Y(k, \ell)|^2 \tag{5}$$

The conditional speech presence probability is defined as

$$p(k, \ell) = \alpha_p \cdot p(k, \ell - 1) + (1 - \alpha_p) \cdot I(k, \ell), \tag{6}$$

where α_p is a smoothing parameter and $I(k, \ell)$ is defined as an indicator function for speech presence. The speech presence in a given frame is obtained by the ratio of the local energy of the noisy signal S and its minimum $S_{\min}$ within a specified time window. The ratio S_r is compared with a frequency-dependent threshold value λ determined experimentally by Cohen [11] if the ratio is larger than λ, the value of $I(k, \ell)$ is 1, if not then the value of $I(k, \ell)$ is 0.

We compute the spectral weighting using two parameters (S, D') obtained by MCRA. The subtraction rule is important in noise reduction algorithm. Therefore, we modified the conventional subtraction rule by adding two algorithms to decrease the residual noise. The noise spectrum D' is compensated using the biases of each frequency bin computed by the subtraction factor f:

$$P_n(k, \ell) = b(k, \ell) \cdot D'(k, \ell) \tag{7}$$

The estimated noise spectrum $D'(k, \ell)$ obtained by MCRA is based on minimum tracking with recursive averaging. Therefore, the estimated noise spectrum is lower than the real noise spectrum and this bias needs to be compensated. The

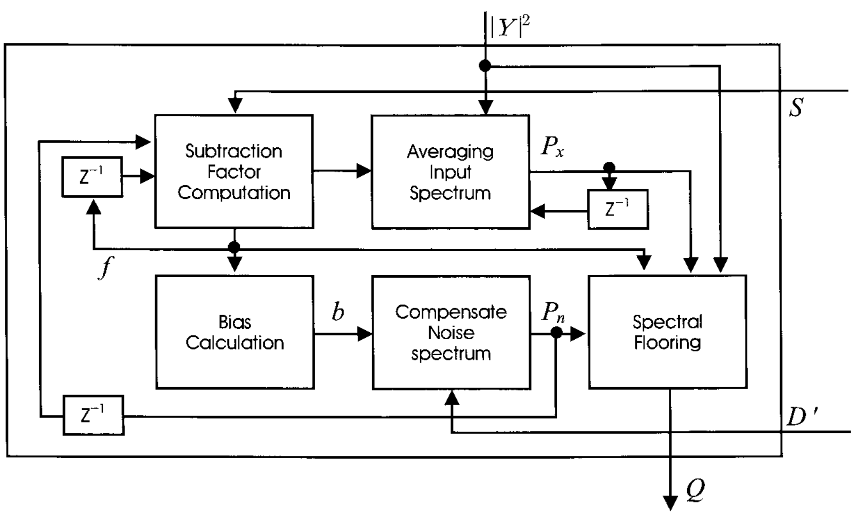

Fig. 3. Block diagram of the spectral computation.

subtraction factor $f(k, \ell)$, which indicates the SNR, is used to adjust the bias of each frequency bin.

The biases of every frequency bin $b(k, \ell)$ are calculated using the bias weight b_w, the bias shift b_s and the subtraction factor $f(k, \ell)$:

$$b(k, \ell) = b_w \cdot f(k, \ell) + b_s \tag{8}$$

The subtraction factors for each frequency bin are smoothed using

$$f(k, \ell) = \alpha_q \cdot f(k, \ell - 1) + (1 - \alpha_q) \cdot \left(1 + \frac{P_n(k, \ell - 1)}{P_n(k, \ell - 1 + S(k, \ell))}\right) \tag{9}$$

where $f(k, \ell)$ is the subtraction factor, $P_n(k, \ell)$ is the compensated noise spectrum, and $\alpha_q$is a smoothing constant. A large change in the subtraction factor due to a sudden change of SNR generates musical noise. Therefore, we smooth the subtraction factor to decrease its variance. The subtraction factor is computed as a function of the SNR using the local energy of noisy speech S and the compensated noise spectrum P_n.

The procedures for the spectral weighting computation are shown in Figure 3. The smoothed input spectrum P_x is determined by first order recursive averaging with a smoothing parameter α:

$$P_x(k, \ell) = \alpha \cdot P_x(k, \ell - 1) + (1 - \alpha) \cdot |Y(k, \ell)|^2 \tag{10}$$

Following the proposal of Martin [12], we subtract from the spectral magnitude a subtraction factor f with a limit on the maximum subtraction of a spectral floor constant $subf$. The spectral weighting Q is calculated using the noise spectrum P_n, obtained by the proposed subtraction rule with a limited maximum subtraction by a spectral floor constant $subf$:

$$\begin{aligned} &if \quad 1 - \sqrt{f(k, \ell) \cdot \frac{P_n(k, \ell)}{P_x(k, \ell)}} \le surf \cdot \sqrt{\frac{P_n(k, \ell)}{|Y(k, \ell)|^2}} \\ &\qquad Q(k, \ell) = surf \cdot \sqrt{\frac{P_n(k, \ell)}{|Y(k, \ell)|^2}} \\ &else \\ &\qquad Q(k, \ell) = 1 - \sqrt{f(k, \ell) \cdot \frac{P_n(k, \ell)}{P_x(k, \ell)}} \end{aligned} \tag{11}$$

Companding has been used to enhance spectral contrast [15]. It naturally emphasizes spectral peaks and attenuates spectral valleys. However, companding requires configuration of a multi-channel filter bank to decrease distortion. Consequently, we searched formants using a formant detection algorithm and applied companding only to channels of the formant frequencies. In addition, we applied companding only in the speech frame using speech presence detection.

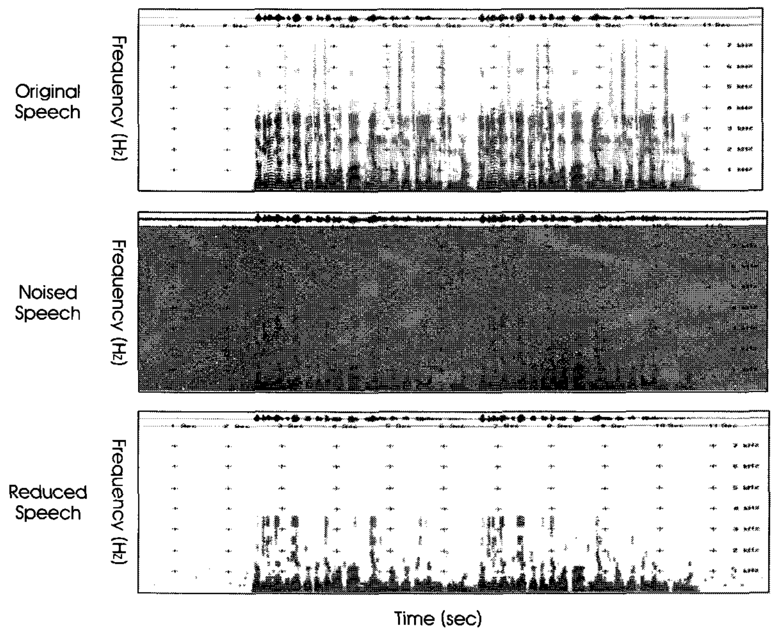

Fig. 4. Spectrum of speech: original, noise induced and speech enhanced signals.

Single channel companding consists of a prefilter, compression block, postfilter, and expansion block. The prefilter is a band pass filter with a broad bandwidth and the postfilter is band pass filter with a narrow bandwidth. The prefilter and postfilter have the same formant frequency. The spectral energies of frequencies near the formant are suppressed by the prefilter and compression block, and the spectral energy of the formant frequencies is enhanced by the postfilter and expansion block. This is defined by

$$x_t' = F_i(x')\cdot(ED(F_i(x')))^{ci-1} \tag{12}$$

and

$$x_i = G_i(x_i')\cdot(ED(G_i(x_i')))^{(ei-ci)/ci} \tag{13}$$

where F is the prefilter, G is the postfilter, ED is the envelope detector, ci is the compression index, ei is the expansion index and i is the order of formant that has a value between 1 and 3. The envelope is obtained by taking a filtered signal $f(n)$ and extracting its Hilbert envelope $abs(H(f(n)))$ where H represents the Hilbert transform. This is defined by

$$x'_{env}(n) = abs(H(f(n))) \tag{14}$$

We chose band pass filters for F and G with transfer functions described by $F_i(s) = F_i'^2(s)$ and $G_i(s) = G_i'^2(s)$, where $F_i'(s)$ and $G_i'(s)$ are given by equations (15) and (16):

$$F_i'(s) = \left(\frac{2\left(\frac{\tau_i}{q_1}\right)s}{\tau_i^2 s^2 + 2\left(\frac{\tau_i}{q_1}\right)s + 1} \right)^2 \tag{15}$$

$$G_i'(s) = \left(\frac{2\left(\frac{\tau_i}{q_2}\right)s}{\tau_i^2 s^2 + 2\left(\frac{\tau_i}{q_2}\right)s + 1} \right)^2 \tag{16}$$

Because we require zero-phase versions of $F_i(s)$ and $G_i(s)$, we apply $F_i'(s)$ or $G_i'(s)$ once in the forward time direction and once in the reverse time direction, namely $F_i'^2(s)$ or $G_i'^2(s)$ mean that $F_i'(s)$ or $G_i'(s)$ are applied twice. The coefficients are computed using the bilinear (Tustin) approximation. The filters are designed using formant frequencies that are extracted using linear prediction coding (LPC) and an algorithm for peak detection [11]. The speech presence detection compares the sum of I with the speech presence threshold value δ. If the sum of I is larger than δ, $x''(n) = x''(n) = x'(n) + x_1(n) + x_2(n) + x_3(n)$, or if not, $x''(n) = x'(n)$. $x_i(n)$ is the speech with formant enhancement by companding with the resonant frequency of ith formant.

Ⅳ. PERFORMANCE EVALUATION

Three speech enhancement algorithms, such as the spectral enhancement, Wiener filtering, and maximum likelihood envelop (MLE) estimation, are being evaluated. We simulated and implemented the speech enhancement algorithm using Matlab and a DSP board. The processed signals are presented to both normal and sensorineural hearing impairment listeners. Their responses in background noise were investigated and compared with those by the speech enhancement algorithms that presented in this paper. We implemented the spectral enhancement algorithm using both bandpass filters according to critical band rate in time domain and short-time Fourier transform in frequency domain to compare the results in real time [16]. An estimation of the noise spectrum is updated during periods of non-speech duration. A voicing and pitch detector using both zero-crossing rate method and modified Roberts algorithm was used for VAD (Voice Activity Detector). To avoid side effect from VAD methods, we used optimal method to confirm voice duration for the clinical test. This VAD method is also compared with MCRA algorithm. A critical band processing is employed for reflecting human psychoacoustics. The energy rather than magnitude of input signal at each critical bandis calculated and the gain is computed according to the VAD. Finally, phase term is added from the frequency spectrum, and the signal is converted in time domain using a short-time IFFT routine.

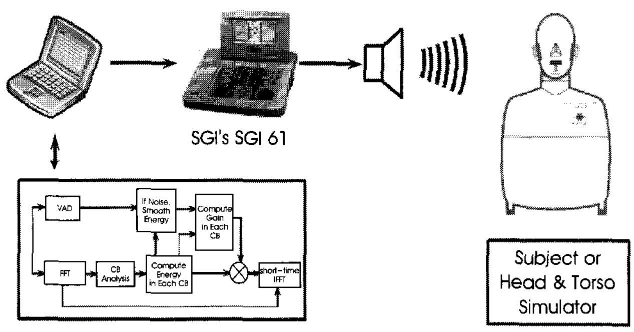

Fig. 5. Test set-up for comparing speech enhancement algorithms

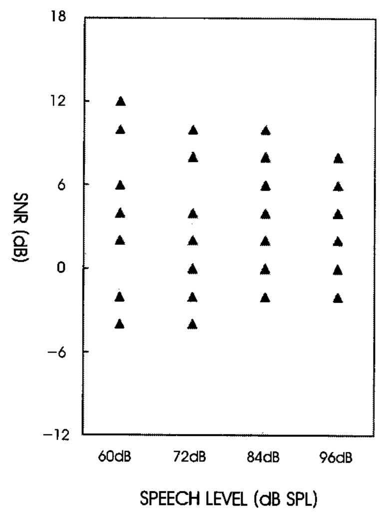

Fig. 6. Speech-to-Noise Ratios need to achieve 50% performance using SP words at four speech levels

For real time evaluation of comparing three speech enhancement algorithms, a floating point DSP processor and 16-bit audio CODEC with additional analog circuits are comprised in the DSP system. The system could complete the entire processing in allowed processing time. A Grason-Stadler GSI 61 audiometer with a TDH-50p earphone were used to evaluate the algorithm for subject tests.

A. Subjects

Sensorineural hearing impairment listeners were selected from a hearing research volunteer group at Samsung Medical Center. The hearing-impaired group was composed of 18 subjects (age: 28 to 79, mean: 58.4 years) and 22 ears were tested for the comparison. The audiological criteria for the selection were: sensorineural hearing losses, Speech Recognition Threshold (SRT) scores are 47.1 dB HL of mean value from range 24 to 84, Speech Discrimination Test (SDT) scores are 75.4% of mean from range 22 to 100%. 15 cases of the subjects have descending-type audiograms, one has flat form, and the others have bell-type audiograms. Additionally, on the basis of prior medical examination, the subjects were restricted to those with no active pathology. Summary descriptions of relevant characteristics of the hearing impaired subjects are given in Table 1.

B. Test Materials and Instrumentation

The speech materials are two-syllabic spondee word lists and monosyllabic Phonically Balanced Word (PB word) lists.

Table 1. Subject characteristics.
a)Pure-tone-average (PTA) is the average hearing loss (ANSI, 1969b) at 500, 1000, and 2000 Hz.
b) We used MCL instead of the speech presentation level (SDT 60) for 5 of 22 subject tests

Group	Subject	Ear	Age	PTAa)	SDT	SDT60	Note
Falling	JKT	R	43	45	82%	80%	
	HIW	R	71	43	74%	42%	
	KDS	R	69	43	98%	64%	
	CHM1	R	50	40	100%	88%	
	PYS1	L	51	40	98%	90%	
	LSB1	R	28	42	92%	74%	
	LSB2	L	28	42	98%	80%	
	JKS	R	49	38	68%	48%	
	CHM2	L	50	30	100%	76%	
	LHT	R	79	47	82%	46%	
	LSB3	L	67	37	96%	80%	
	HBC	L	75	38	90%	44%	
	LGG	L	70	50	88%	44%	
	LKW	R	73	87	36%		MCLb)
	KYK	R	60	78	38%		MCLb)
Bell	KBJ	R	57	53	56%	30%	
	KYS	R	71	35	84%	66%	
	KGC	L	67	17	100%	96%	
	PYS2	R	51	67	48%		MCLb)
	KCH	R	43	47	76%	48%	
Flat	KHY	L	59	58	54%		MCLb)
Ungroup	NJS	R	66	55	32%		MCLb)
	Mean		58	46.9	76.8%	65.8%	

These word lists, which are clinically used at Audiology and Vestibular Laboratory of Samsung Medical Center, were taken from K-SPIN (Korean Speech In Noise) test and K-HINT (Korean-Hearing In Noise Test).

The speech noise was generated by a GSI61 audiometer. Speech-shaped, multi-talker babble and environmental sounds were used as background noise. The signal and speech noise were mixed by the audiometer and presented monaurally to the subjects.

C. Procedure

On the basis of prior medical examination, the subjects were restricted to those with no active pathology. Initially, in order to assess performance at speech levels encountered under condition of linear amplification and investigate the change of speech reception ability of our selected subject group in the noisy environment, speech recognition performance was measured over a relatively wide intensity range, from 60 to 96 dB HL. The SNR required for 50% correct pointusing spondee word lists was estimated at four speech levels, 60, 72, 84, and 96 dB HL. Table 1presents the individual performance data for 22 ears using spondee word lists. In Figure 6, as higher speech level, the data is less scattered and this shows similar result of Dirks's study [2].

To compare the performance in noise, the 50% performance for words from the spondaic words was estimated at three different types of noise environment, SNRs of +6, 0, and -6 dB. We measured this performance with the speech through the speech enhancement algorithms. We confirmed duration of speech in the VAD processing to prevent side effect of a voice detection algorithm itself. The word recognition tests were performed using PB word lists at the presentation level, 60 dB HL of speech. For five of subjects, the presentation level of 60 dB HL was not enough to the speech test, so subjects' MCLs (Most Comfortable Level)were used as presentation levels. Noisy environment was same as the 50% performance test, +6, 0, and 6 dB of SNRs.

Speech recognition performance was determined to evaluate the proposed algorithm for listeners with normal hearing and sensorineural hearing loss, while listeners in the noise. First, we evaluated speech recognition ability in various speech-to-noise ratios and types of noise. Next, speech enhancement algorithms by reducing background noise were evaluated to improve speech intelligibility for sensorineural hearing impairment listeners. We tested three speech enhancement methods using single-microphone, such as the spectrum subtraction, Wiener filter method, and maximum likelihood envelop estimation. Their responses in background noise were investigated and compared with those by the speech enhancement algorithm that presented in this paper.

Most of subjects are used to wear their own hearing aid for years and they are very familiar with the sound of own hearing aids. To avoid any possible misconception of the processed sound, we evaluate the algorithm for the subject with wearing their own hearing aid on. The purpose of this test is to compare VAD and MCRA methods and evaluate sound acquaintancefor the subjects. The task for the hearing aid users was to repeat two syllables, either noisy or processed. Before the test the processed sound level is set to their MCL. An adaptive procedure was then used to determine the 50% correct point, in terms of SNR, for the speech recognition threshold [9]. The test words were initially presented at a 20 dB SNR. If the hearing aid users repeated the test words correctly, the SNR was decreased by 2 dB, and if the hearing aid users repeated the test words incorrectly, the SNR was increased by 2 dB. This procedure continued until 14 reversals had occurred and the SRT was taken as the average of the final 10 reversals. Each subject completed all 12 tests that comprised the three noise types and four processed test words. In this test we could compare VAD and MCRA methods to the subjects as well.

D. Analysis and Test Result

In the first experiment, the hearing thresholds of the 50% correct point for the spondee words were estimated at S/N ratio of +6, 0, and −6dB. A comparison was made of the two results with and without use of the speech enhancement method. In Figure 7, the left-hand side of the graph shows the data obtained without the speech enhancement algorithm. We could estimate the performance with the speech enhancement algorithm at 6dB and 0 dB of SNR, but only two of 16 subjects could make response at the −6dB of SNR.

The mean value of speech discrimination test in quiet is about 59%. Figure 8 represents the mean discrimination score at each S/N ratios. As can be seen, the means closely approximate a linear relationship between discrimination scores and SNRs. For the SNRs +6, 0, and −6dB, the mean scores are improved by 21, 28, and 29%, respectively. The improvement is highest at −6dB. We could compare the discrimination score (31%) at +6dB of SNR without the speech enhancement and the score (28%) at −6dB of SRT with the speech enhancement method. With the results from the comparison, about 12dB of SNR is improved in the speech discrimination test.

The subjects were divided into three groups according to the speech discrimination score, 0 to 49%, 50 to 79%, and 80 to 100%. When grouping the subjects, we considered audiological meaning such as a prognosis of a neural tumor, a marginal range of normal hearing, and so on. Figure 9

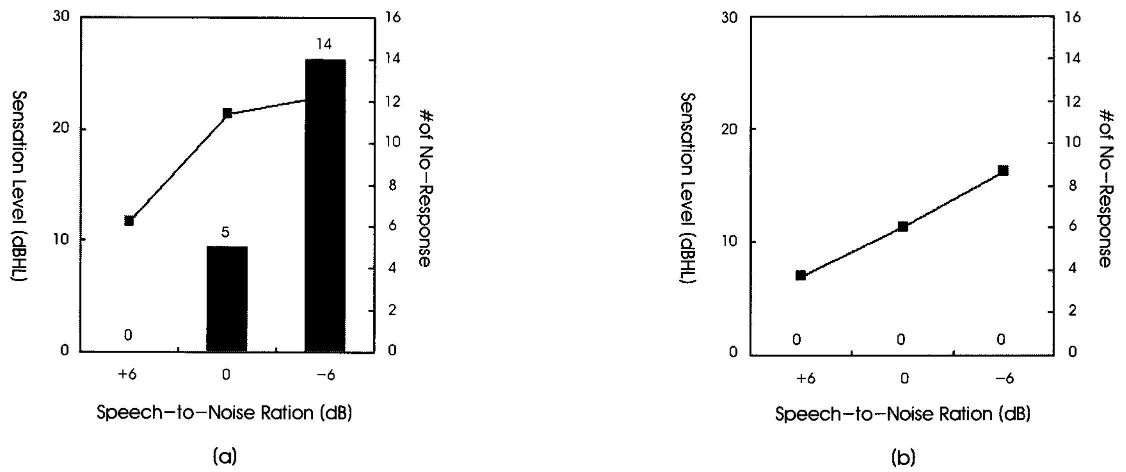

Fig. 7. Mean sensation levels of 50% performance using the spondee words at S/N ratio of +6, 0, and -6 dB (a) without and (b) with the speech enhancement algorithm.

represents the discrimination scores of three groups. The mean percentage values of each group were 41, 67, and 86%. From the results in Figure 9 and 10, the discrimination was increased with decreasing SNR, and in higher discrimination score obtained group. The mean discrimination score at −6dB of SNR is only 2% and the highest score is 16%. This result reflects we could estimate speech thresholds at −6dB of SNR for 2 subjects in 16 sensorineural hearing-impaired persons (Figure 7 (a)). For the subject with the descending type of audiogram, the improvement is about 5% at +6 and 0dB of SNRs, but there is no significant difference with other groups at −6dB of SNR. In addition, in the case of the standard speech enhancement or noise reduction system, the process of dynamic reduction of spectral energy introduces an audible artifact, 'a musical noise'. But 5 subjects could notify the musical noise that was the major factor for preference test by the normal hearing person.

As shown in Figure 11, WRT scores of −6dB of SNR shows different results in speech-like noise, e. g., speech noise or multi-talker babble noise or in environment noise like traffic sound. In this case the level of background noise is louder than the targeted speech sound. This shows the correlation of the speech and background noise is important factor to the subject listeners.

Figure 12 shows the results of SRT in white noise environment for selected subject with wearing their own hearing aids. The result shows that the most of subjects could recognize speech in the lower SNR environment using the proposed method. For statistical analysis using Wilcoxon's signed rank test, p-values of the proposed speech enhancement

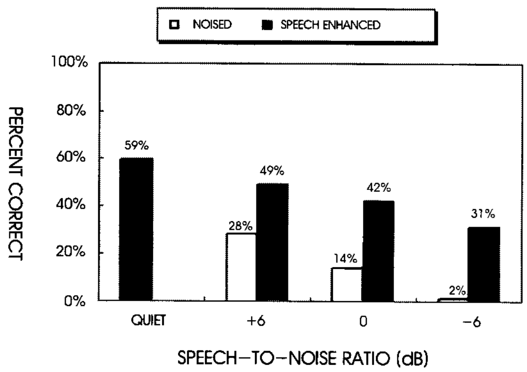

Fig. 8. Mean discrimination scores as a function of S/N ratios.

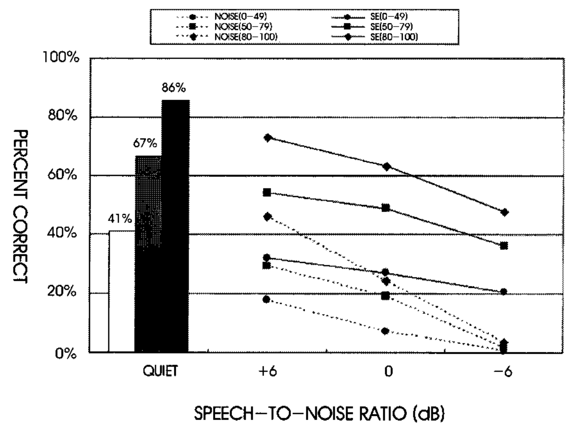

Fig. 9. Mean discrimination scores into three groups. The solid lines are scores without the enhancement and the dashed lines are with the enhancement method.

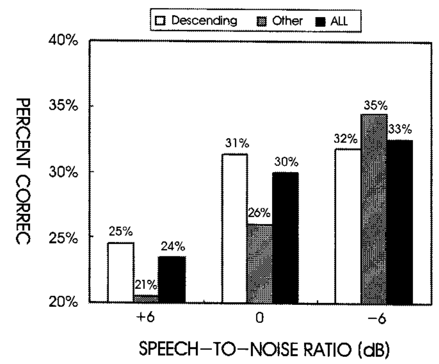

Fig. 10. Mean discrimination scores improvement of subjects by types of audiograms.

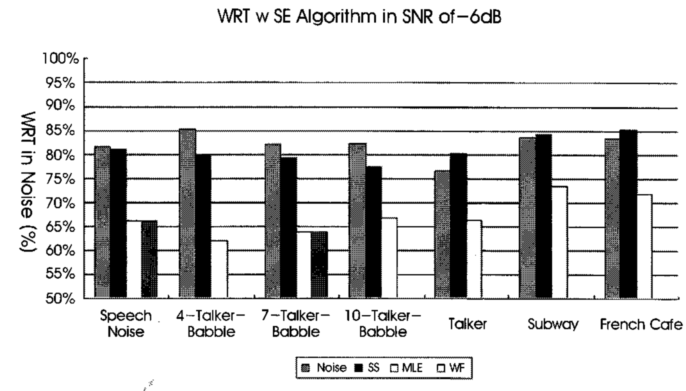

Fig. 11. Word Recognition Test scores with speech enhancement algorithms at various background noise for the normal persons.

algorithm using VAD are 0.008 in white Gaussian noise. The values are more significant when using the speech enhancement algorithm with MCRA in the white noise, 0.011.

The methods improved speech recognition test score for the sensorineural hearing impaired listeners, but not for normal listeners. The results suggest the speech enhancement algorithm may work in synergy with the gain function for the greater SNR improvement as the pre-processing algorithm of digital hearing aids.

V. DISCUSSION AND CONCLUSIONS

In this paper, we proposed the speech enhancement algorithm using modified spectral subtraction and companding for digital hearing aids and compared with other speech enhancement algorithms. The algorithm comprises the noise estimation by recursive averaging method, spectral weighting computation, and formant enhancement. The spectral weighting is obtained by modifying the subtraction rule for conventional spectral subtraction. The modification reduces the variance of the subtraction factor and adjusts the biases in the noise spectrum. Therefore, excellent noise suppression is obtained while avoiding musical residual noise and decreasing the distortion. Enhancement of formants is performed using companding. Formants of speech with noise suppression are enhanced based on the speech presence indicator. The proposed method has been tested and compared with other methods in various background noise and SNR levels. The superiority of the proposed method was demonstrated through subject evaluations. The result shows that the proposed method is beneficial for hearing aid users in diverse noise environments.

The auditory characteristics of each hearing impaired listeners are different from subject by subject. Also the fitting of the hearing aid is a process to make comfortable for each hearing impairment listeners and personal preference. However all of the hearing aid users could not be satisfied about their hearing aids. There are some possible explanations

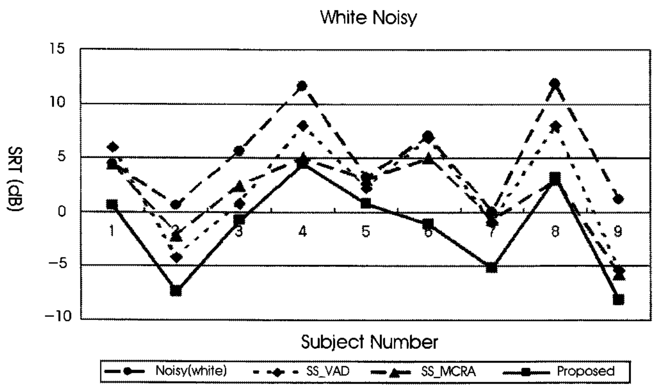

Fig. 12. SRT results of hearing aid users for white Gaussian noise by the speech enhancement algorithms: noisy, spectral subtraction using VAD, spectral subtraction using MCRA, and the proposed algorithm.

various noise environment, technical limitation of current hearing aid and/or lack of understanding about hearing itself [17, 18]. One example is the loudness recruitment which is used to be solved easily using wide dynamic loudness compression in hearing aid. There are three classical hypotheses about loudness recruitment. The common hypotheses are 1) loudness recruitment results from steeper auditory nerverate level functions after recruitment, 2) Loudness is based on the total auditory nerve discharge count, and recruitment results from an abnormally rapid spread of excitation after impairment, 3) loudness of narrowband stimulus is based on auditory nerve responses in a narrow Best Frequency (BF) region, and recruitment results from compression of the auditory nerve fiber threshold distribution after impairment. Auditory nerve activity growth rates are consistent with some, but not all, aspects of a new view of loudness recruitment. Similar auditory nerve growth rates near threshold are consistent with normal loudness growth near threshold in impaired listeners. Reduced occurrence of sloping saturation is consistent with loudness "catching up" at high levels. However, there is no obvious correlate of elevated loudness at threshold in the rate responses of impaired auditory nerve fibers. It is also possible that loudness is not directly related to total auditory nerve activity. Alternative hypotheses for neural correlates of loudness recruitment is based on temporal aspects of auditory nerve responses and a change in gain at more central levels.

However according to this recent electrophysiological research about cochlear and auditory system, we need to modify current loudness compression algorithm for hearing aids. The compression algorithms designed to overcome a reduced dynamic range may need to account for more than reduced BM compression, especially in cases of mixed OHC/IHC damage such as noise-induced hearing loss. Another problem that may have caused confusion is component 2 (C2) responses at high sound levels. Component 1 (C1) response dominates AN fibers at low to moderate sound levels and a transition to C2 is usually observed at high sound level. This is typically defined by a dip in the rate-level function and/or an abrupt change in the phase of response relative to the stimulus waveform. Noise-induced hearing loss can reduce or eliminate C1 with little evident effect on C2. Because the remaining C2 response shows steeper response growth, this effect may be significant for hearing aids that operate at high levels [19, 20]. This recent research in hearing science and speech processing method in background noise may elevate speech perception for the listeners with sensorineural hearing impairment.

We are currently working on implementing real time system for hearing aid using the speech enhancement algorithm and loudness compensation method based on the recent result in electrophysiological research about cochlear and auditory system. Also prospective ratio of the formant enhancement algorithm for types of hearing impairment is under development on the hearing aid fitting software.

REFERENCES

[1] S.M. Lee, S.W. Kim, S.Y. Kwon, Y.C. Park, I.Y. Kim, S.I. Kim, "Development of adaptive feedback cancellation algorithm for multi-channel digital hearing aids," *J. Biomedical Engineering Research*, vol. 25, no. 4, pp. 315-321, 2004.

[2] Dirks, D., Morgan, D., Dubno, J., "A procedure for quantifying for effects of noise on speech recognition," *J. Speech Hearing Disorder*, vol.47, pp.114-123, 1982.

[3] Hume, L. E., "Understanding the speech-understanding problems of the hearing Impaired," *Am. Acad. Audiol.*, vol.2, pp.59-69, 1991.

[4] Brronkhost, A. Q., Plomp, R., "A clinical test for the assessment of binaural speech perception in noise," *J. Audiology*, vol.29, pp275-285, 1990.

[5] Klein, A. J., Mills, J. H., Adkins, W. Y., "Upward spread of masking, hearing, and speech recognition in young and elderly listeners," *J. Acoustic. Soc. Am.*, vol.87, pp.1266-1271, 1990.

[6] Killion, M. C. and Fikret-Pasa, S., "The three types of sensorineural hearing loss: Loudness and intelligibility considerations," *Hear. J.*, vol. 46, no. 11, pp. 31-36, 1993.

[7] Moore, B. C. J., "Testing the concept of softness imperception: loudness near threshold for hearing-impaired ears," J. *Acoust. Soc. Am.*, vol.115, pp.3103-3111, 2004.

[8] Moore, B. C. J., Glasberg, B. R., Hess, R. F., Birchall, J. P., "Effects of flanking noise bands on the rate of growth of loudness of tones in normal and recruiting ears," *J. Acoust. Soc. Am.*, vol.77, pp.1505-1513, 1985.

[9] Hohmann, V., andKollmeier, B., "The effect of multichannel dynamic compression on speech intelligibility," *J. Acoust. Soc. Am.*, vol. 97, pp. 1191-1195, 1995.

[10] Stahl, V., Fischer, A., Bippus, R., "Quantile based noise estimation for spectral subtraction and Wiener filtering," in *Proc. 25th IEEE Internat. Conf. Acoust. Speech Signal Process.*, ICASSP2000, Istanbul, Turkey, 2000, pp. 18751878.

[11] Cohen, I., Berdugo, B., "Noise estimation by Minima Controlled Recursive Averaging for Robust Speech Enhancement," *IEEE Signal Processing Letter*, vol. 9, no. 1, pp. 1215, 2002.

[12] Martin, R., "Spectral subtraction based on minimum statistics," in *Proc. 7th European Signal Processing Conf., EUSIPCO94*, Edinburgh, Scotland, 1994, pp. 11821185.

[13] Doblinger, G., "Computationally efficient speech enhancement by spectral minima tracking in subbands," in *Proc. 4th European Conf. Speech, Communication and Technology*, EUROSP-EECH'95, Madrid, Spain, 1995, pp. 15131516.

[14] Cohen, I., Berdugo, B., "Speech enhancement for non-stationary noise environment," *Signal Processing: the official publication*

of the European Association for Signal Processing (EURASIP), vol. 81, no. 11, pp. 24032418, 2001.

[15] Turricchia, L., Sarpeshkar, R., "A Bio-Inspired Companding Strategy for Spectral Enhancement," IEEE Transactions on Speech and Audio Processing, vol. 13, no. 2, pp. 243253, 2005.

[16] Zwicker, E., Flottorp, G., Stevens, S. S., "*Critical bandwidth in loudness summation," J. Acoust. Soc. Am.*, vol. 29, pp.548-557, 1957.

[17] Dillon, Harvey, "Tutorial; Compression? Yes, But for low or high frequencies, for low or high intensities, and with what response times?," *Ear & Hearing*, vol. 17, no. 4, pp. 287-307, 1996.

[18] Zwicker, E. and Fastl, H., *Psychoacoustics*, Berlin, Springer-Verlag, 1990, pp. 133-147.

[19] Heinz, M. G., Kim, D. W., Sachs, M. B., Young, E. D., "A baseline Characterization of Activity Growth in Auditory-Nerve Fibers of Normal-Hearing Cats," the 25th Annual Midwinter Research Meeting, *Association For Research in Otolaryngology*, Florida, USA, vol. 25, January 27-31, 2002.

[20] Heinz, M. G., Young, E. D., "Response growth with sound level in auditory-nerve fibers after noise-induced hearing loss," *J. Neurophysiol.*, vol. 91, pp.784-795, 2004.