고품질 인터넷 전화 서비스를 위한 스마트 디바이스에서 홈 유비쿼터스 환경적 노이즈 제거 방안 연구

장병옥*

요약

본 논문은 유비쿼터스 홈 환경에서 고품질 인터넷전화 서비스를 제공하기 위한 노이즈 신호 제거 방안을 제안하였다. 원음 신호의 손실 없이 유비쿼터스 환경에서 환경적 노이즈 신호를 제거하기 위해서 새로운 위너필터 기반의 전처리 노이즈 제거 방안을 적용하였다. 제안된 방안은 다양한 스마트 디바이스에서 팬 잡음, 에어컨디션 잡음, 샤워실 잡음, 등 다양한 노이즈와 함께 실험되었으며 PESQ 평가 방안으로 제안된 방안이 검증되었다.

Ubiquitous Environmental Noise Removal Method on Smart Devices for High-Quality Internet Telephony Service

Byeong-Ok Jang*

ABSTRACT

This paper presents noisy signal removal method on smart devices for internet telephony service at the ubiquitous environment. In order to reduce the environmental noise signal without speech distortion efficiently in a pervasive home environment, we provide a new Wiener filtering based reduction technique with pre-processing on various smart devices at a pervasive home. For various noisy conditions like White Gaussian, fan, air-conditioner, babble, shower-bath noises, etc., at home, the perceptual evaluation of speech quality (PESQ) evaluation is performed to evaluate the performance of the proposed method.

Key Words : Noise, Internet Telephony, Quality, Smart Devices, Ubiquitous Home Environment

* 나사렛대학교 디지털콘텐츠학과(✉bojang@kornu.ac.kr)
· 제1저자(First Author) : 장병옥 · 교신저자(Correspondent Author) : 장병옥
· 접수일(2011년 1월 14일), 수정일(1차 : 2011년 2월 14일), 게재확정일(2011년 2월 17일)

www.kci.go.kr

Ⅰ. Introduction

Internet telephony service has been normally expected with providing high speech quality in our ordinary residential environment today.

Research on noise reduction/speech enhancement can be traced back to about 40 years ago [1]-[4]. Nowadays, the noise reduction techniques are used in a wide range of applications such as communication, automatic speech recognition, and sound source localization systems [5]-[8]. Also, many approaches including spectral subtraction schemes and adaptive filtering techniques, including Kalman and Wiener filtering ,have been proposed to suppress the background noise [9]-[13]. Nevertheless, the challenging problem of noise reduction for speech enhancement with various noise reduction techniques has still remained as one of the main research issues.

We are designing an IP-based pervasive home environment to provide an Internet telephony service using smart devices which are connected by the Internet (see Fig. 1). For example, smart mirrors in the bathroom or bedroom can function by using a call connection over the IP-based network. A Smart table in the living room and a smart window in kitchen are also connected by the IP-based network to provide a call service. However, call conversation is interrupted by various environmental background noises such as the sound of dripping tap water (shower-bath), the sound of a fan (air-conditioner), the sound of a TV, etc. This paper chooses a Wiener filter-based noise reduction scheme because it satisfies the above constraints and shows a basically high performance as one of the popular approaches.

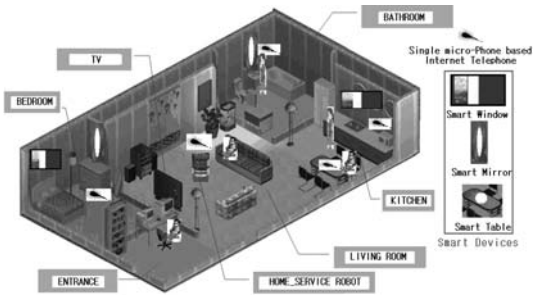

그림 1. 유비쿼터스 홈 환경 투시도
Fig 1.Perspective Plan for Ubiquitous Home Environment

The ITU-T P.862 PESQ is used as an objective performance measure [14]. Various noise conditions such as White Gaussian, fan, water, babble, and air-conditioner noises are considered. The performance of the proposed method is compared with that of the noise reduction methods in the IS-127 EVRC [15] because it includes the noise reduction block while speech codecs have no noise reduction. And, the performance is compared with that of the noise reduction method in the ETSI standard for the distributed speech recognition front-end because it adopts the standardized noise reduction method based on the Wiener filter. The proposed noise reduction scheme is applied as pre-processing of the Internet telephony speech codec such as G.723.1.

The paper is organized as follows. Various smart devices at pervasive home network environment are introduced in Section II and our Wiener filter-based noise reduction is proposed in Section III. In Section IV, our experiments are described and the performance evaluation results are shown. Finally, in Section V, we draw the conclusions including further studies.

www.kci.go.kr

II. Internet Telephony Service on Ubiquitous Home Environment

The system is designed to support a range of activities, from home-based settings to collaboration between distant sites. The system incorporates with the Internet telephone for a voice communication service at-home network environment. However, there are various noisy sounds that exist in a home environment during the call-connection. Thus, the proposed noise signal reduction scheme using an optimized Wiener filtering technique based on input-SNR estimation is applied on the designed smart devices such as the smart mirror (see Fig. 2) and smart table (see Fig. 3) in the pervasive home environment.

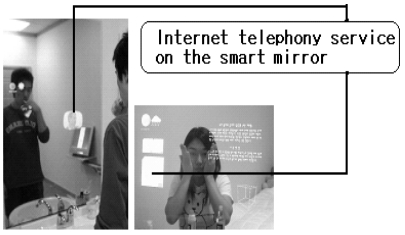

그림 2. 스마트 거울
Fig 2.Smart Mirror

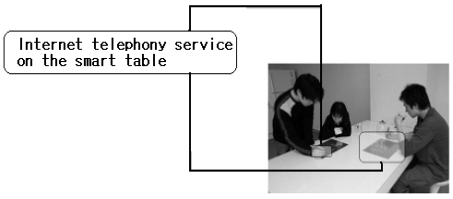

그림 3. 스마트 테이블
Fig 3.Smart Table

III. Noise Reduction Scheme on IP-based Ubiquitous Home Environment

For The proposed scheme has two advantages over the existing approach of techniques:

1.More effective noise reduction to remove clearly by using the input-SNR estimation technique on smart devices at a pervasive home.

2. In addition, an efficient reduction method for environmental background noise.

For a non-causal infinite impulse response (IIR) Wiener filter, n noisy, a clean speech signal $d(n)$, a background additive noise $v(n)$, and an observed signal $x(n)$ can be expressed as

$$x(n) = d(n) + v(n). \quad (1)$$

It is assumed that $d(n)$ and $v(n)$ are jointly wide-sense stationary. A Wiener filter is designed to minimize the mean square error.

$$r_{dx} = \sum_{l=-\infty}^{\infty} w(l) r_x(k-l). \quad (2)$$

where $r_x(k)$ is the autocorrelation of $x(n)$ and $r_{dx}(k)$ is the cross-correlation between $x(n)$ and $d(n)$. Because it is assumed that $d(n)$ and $v(n)$ are uncorrelated, the autocorrelation and the cross-correlation can be respectively given as

$$r_x(k) = r_d(k) + r_v(k), \quad (3)$$

$$r_{dx}(k) = r_d(k). \quad (4)$$

Thus, (4) can be expressed in the frequency domain as

www.kci.go.kr

$$P_d(e^{jw}) = W(e^{jw})\left[P_d(e^{jw}) + P_v(e^{jw})\right]. \quad (5)$$

and the frequency response of the Wiener filter becomes

$$W(e^{j\omega}) = \frac{P_d(e^{j\omega})}{P_v(e^{j\omega}) + P_d(e^{j\omega})} = \frac{\zeta(e^{j\omega})}{1+\zeta(e^{j\omega})}. \quad (6)$$

in which $\zeta(e^{j\omega})$ is considered as the SNR (Signal-to-Noise Ratio) defined by

$$\zeta(e^{j\omega}) = \frac{P_d(e^{j\omega})}{P_v(e^{j\omega})}. \quad (7)$$

Since a noncausal IIR filter is unrealizable in practice, we propose a causal finite impulse response (FIR) Wiener filter. The proposed noise reduction procedure is shown in Fig. 4. Noise reduction is processed frame-by-frame. The length of the processing frame is 80 samples or 10 ms. A total 100 samples, i.e., the current 80 and the past 20 samples, are used for the power spectrum calculation of the processing frame.

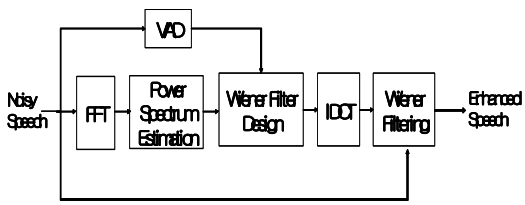

그림 4. 잡음 제거 절차
Fig 4.Proposed Noise Removal Procedure

The past samples are initialized to zero in the first frame. There is no buffering delay to analyze the power spectrum. For the power spectrum analysis, the signal is windowed by the 100 sample-length asymmetric window $h(n)$ whose center is located at the 70th sample. The length and the center of the asymmetric window are empirically chosen to make the algorithm produce the best performance. The asymmetric window is expressed as (8).

$$h(n) = \begin{cases} 0.5 - 0.5\cos\left(\dfrac{2\pi n}{139}\right) & ;0 \le n < 70 \\ \cos\left(\dfrac{2\pi(n-70)}{119}\right) & ;70 \le n < 100 \end{cases} \quad (8)$$

The signal power spectrum is computed for this windowed signal using the 256-FFT. Based on the voice activity detection (VAD) decision, the noise power spectrum is updated only for non-speech intervals in the Wiener filter design. For speech intervals, the last noise power spectrum is reused. Plus, the speech power spectrum is estimated by the difference between the noise power spectrum and speech power spectrum. In our proposed Wiener filter, the frequency response is expressed as

$$W(k) = \frac{\zeta^{\alpha}(k)}{1+\zeta^{\alpha}(k)}, \qquad 0 < \alpha \quad 1. \quad (9)$$

and $\zeta(k)$ is defined as

$$\zeta(k) = \frac{P_d(k)}{P_v(k)}. \quad (10)$$

where k is the frequency bin, $\zeta(k)$, $P_d(k)$, and $P_v(k)$ are the SNR, the speech power spectrum, and

www.kci.go.kr

the noise power spectrum, respectively. Therefore, filtering can be controlled by the parameter α. As α increases, $\zeta^*(k)$ also increases for $\zeta(k)$ greater than one, while $\zeta^*(k)$ decreases for $\zeta(k)$ less than one. Therefore, the signal is more strongly filtered out to reduce the noise for small $\zeta^*(k)$. On the other hand, the signal is more weakly filtered with little attenuation for large $\zeta^*(k)$.

To analyze the effect of α, we evaluate the performances for α values from 0.1 to 1. The performance is evaluated not for the coded speech but for the original speech in White Gaussian conditions. As α is increased up to 0.7, the performance is improved. The performance becomes worse after that.

Thus, we can adaptively select the optimal α according to the estimated SNR by a logistic function. The logistic function is trained to decide the optimal α for the estimated SNR at each frequency bin. The logistic function used in this paper can be expressed as

$$p(SNR) = Min + \frac{2(Max - Min)}{1 + e^{(|n-1|/\beta)}} . \quad (11)$$

Because the shape of the logistic function changes with the variation of β, if the maximum and the minimum values of the logistic function are fixed, we should find the appropriate β.

The appropriate β value is decided by the simple gradient search algorithm. At the first iteration for the initial β value, the corresponding α (as the output of the logistic function) is calculated with the estimated SNR as the input of logistic function at each frequency bin for each frame.

The designed Wiener filter coefficients in the frequency domain are transformed into the time-domain ones by the inverse discrete cosine transform (IDCT). Finally, the noise is suppressed by the convolution sum between the impulse response of the proposed Wiener filter and the noisy speech. Because the proposed Wiener filter is a causal filter, the algorithmic delay is unavoidable. However, the delay is almost the half of the filter length and almost negligible compared to the total voice over IP network-delay.

IV. Experimental Results

To test the proposed algorithm in pervasive home network environment, the designed smart devices such as smart mirror, smart table, and smart window are applied for Internet telephony service with various noisy signals.

To evaluate the performance, the PESQ is measured as the objective speech quality assessment. After comparing the original signal with its degraded signal, the PESQ gives us the reliable estimation of the subjective measurement as an MOS-like value from -0.5 to 4.5.

In our experiments, one hundred mono spoken sentences sampled at 8 kHz with 16 bit resolution are used as the clean speech. The duration of each utterance is almost 10 seconds. The utterances are spoken by 2 males and 2 females. As the open test, 40 spoken sentences are used to train the logistic

www.kci.go.kr

function and the others to evaluate the performance. In the training process, the maximum and the minimum values of the logistic function are fixed as 0.65 and 0.6, respectively. After the training process, β is determined to be 1.65. Thus, in the Wiener filter design, α is decided by the output of the logistic function characterized by these parameter values when the estimated SNR is the input of the logistic function at each frequency bin for each frame. The number of the proposed Wiener filter coefficients is 65. Therefore, the algorithmic delay becomes 4 ms.

표 1. G.723.1 (6.3 kbps)를 활용한 PESQ 결과
Table 1. Pesq Results in G.723.1 (6.3kbps)

Noise type	SNR(dB)	None	EVRC	ETSI	Proposed
White Gaussian noise	0	1.44	1.40	1.78	1.86
	5	1.70	1.90	2.20	2.27
	10	2.03	2.36	2.55	2.58
	15	2.34	2.70	2.85	2.84
	20	2.64	2.95	3.07	3.09
Fan noise	0	1.54	1.68	1.72	1.77
	5	1.91	2.07	2.12	2.14
	10	2.28	2.46	2.51	2.51
	15	2.58	2.77	2.82	2.83
	20	2.87	3.05	3.09	3.11
Babble noise	0	1.59	1.72	1.79	1.76
	5	1.91	2.11	2.19	2.14
	10	2.27	2.48	2.52	2.46
	15	2.55	2.77	2.79	2.79
	20	2.84	3.04	3.07	3.06

Table I shows the PESQ results in G.723.1 codec, respectively. In most noisy conditions, the performance of the proposed method is higher than those of the EVRC noise suppression method and those of ETSI noise reduction method.

Popular Internet telephony speech codec such as G.723.1 is tested. The results are compared with those of the noise suppression in the IS-127 EVRC and the noise reduction in the ETSI standard as mentioned. The ETSI noise reduction scheme generates 40 ms buffering delay for the power spectrum analysis while there is no buffering delay in the EVRC noise suppression scheme.

V. Conclusions

We have shown that the speech quality based on the simulation results can be significantly improved by using the proposed scheme. The proposed noise reduction scheme is applied before encoding as a pre-processing of the VoIP speech codecs. For all noisy conditions, the average PESQ gains of 0.30 is achieved for G.723.1 , respectively, by the proposed method. The PESQ results show that the performance of the proposed method outperforms both the EVRC noise suppression and the ETSI noise reduction methods. Thus, our proposed noise reduction method can be effectively used to reduce additive background noises in a pervasive home environment.

Acknowledgement

본 논문은 2010년도 나사렛대학교 학술연구비 지원에 의해 연구되었음.

www.kci.go.kr

References

[1] R. J. Mcaulay and M. L. Malpass, "Speech enhancement using a soft decision noise suppression filter," *IEEE Trans. Acoust., Speech, Signal Process.,* vol. 28, no. 2, pp. 137-145, Apr. 1980.

[2] 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, Apr. 1979.

[3] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE trans. Acoust., Speech, Signal Process.,* vol. 32, no. 6, pp. 1109-1121, Dec. 1984.

[4] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE trans. on Acoustics, Speech, and Signal Process.,* vol. 33, no. 2, pp. 443-445, April 1985.

[5] M. Park, H. R. Kim, and S. H. Yang, "Frequency-Temporal Filtering for a Robust Audio Fingerprinting Scheme in Real-Noise Environments," *ETRI Journal,* vol.28, no.4, pp.509-51, Aug. 2008.

[6] H. C. Lee and D. R. Halverson, "Design of Robust Detector with Noise Variance Estimation Censoring Input Signals over AWGN," *ETRI Journal,* vol.29, no. 1, pp.110-112, Deb. 2009.

[7] Nidal S. Kamel, and Varun Jeoti, "A Linear Prediction Based Estimation of Signal-to-Noise Ratio in AWGN Channel," ETRI Journal, vol.29, no.5, pp.607-613, Oct. 2010.

[8] Kyoung Ho Bang, Young Cheol Park, and Jeongil Seo, "Audio Transcoding for Audio Streams from a T-DTV Broadcasting Station to a T-DMB Receiver," *ETRI Journal*, vol.28, no.5, pp.664-668,Oct. 2006.

[9] K. K. Paliwal and A. Basu, "A speech enhancementmethod based on Kalman filtering," *Proc. ICASSP'87*, pp. 177-180, 1987.

[10] S. Gannot, D. Burshtein, and E. Weinstin, "Iterative and sequential Kalman filter-based speech enhancement algorithms," *IEEE Trans. Speech Audio Process.*, vol. 6, no. 4, pp. 373-385, Jul. 1998.

[11] R. Martin, "Noise power spectral density estimation based on optimal smoothing and minimum statistics," *IEEE Trans. Speech Audio Processing,* vol. 9, no. 5, pp. 504-512, Jul. 2001.

[12] M. H. Hayes, Statistical Digital Signal Processing and Modeling, John Wiley & Sons, 1996.

[13] B. Widrow and S. D. Stearns, Adaptive Signal Processing, Englewood Cliffs, NJ 07632: Prentice Hall, 1985.

[14] Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codec, ITU-T Recommend. P.862, Feb. 2001.

[15] Enhanced variable rate codec, Speech service option 3 for wide-band spectrum digital systems, 1996.

저자소개

장병옥(Byeongok Jang)

1990년 : 서울산업대학교 전자계산학과 학사

1995년 : 동국대학교 정보관리학과 (경영학석사)

1999년 : 경기대학교 전자계산학과 (이학박사)

1981년 ~ 1990년: (주)크라운그룹 전산팀장

1990년 ~ 1998년: (주)건영그룹 전산실장

1998년 ~ 2000년: 한빛정보 CEO

1998년 ~ 1999년: 경기대학교 겸임교수

1999년 ~ 2000년: 광운대학교 대우교수

2001년 ~ 현재: 나사렛대학교 디지털콘텐츠학과 교수

※ 관심분야 : 디지털콘텐츠, 웹 디자인, 게임공학 등

www.kci.go.kr