

## 스마트TV향 VoIP 컨퍼런스 기능을 위한 잡음제거 알고리즘의 성능비교

\*서광덕 \*\*최홍재 \*\*\*김형국

광운대학교 전파공학과

\*[kwangduk.s@gmail.com](mailto:kwangduk.s@gmail.com) \*\*[cozierman@kw.ac.kr](mailto:cozierman@kw.ac.kr) \*\*\*[hkim@kw.ac.kr](mailto:hkim@kw.ac.kr)

### Comparison of Noise Reduction Algorithm for Smart TV in VoIP Conference Facility

\*Seo, Kwang-duk \*\*Choi, Hong-Jae \*\*\*Kim, Hyoung-Gook

Dept. of Radio Science and Engineering, Kwangwoon University

#### 요약

본 논문에서는 스마트TV향 VoIP(Voice over Internet Protocol) 컨퍼런스 기능을 위한 잡음제거 알고리즘의 성능비교 하였다. 기존에 연구 되어져 있는 Improved Minima Controlled Recursive Averaging(IMCRA)방식과 Gaussian분포 기반의 잡음 제거 알고리즘, IMCRA방식과 Gamma분포 기반의 잡음제거 알고리즘, IMCRA방식과 Mel-filter를 적용한 잡음제거 알고리즘, R&L 알고리즘들의 방식을 비교하였으며, 성능 비교를 위해 각 알고리즘을 통해 나온 다양한 잡음 환경에서의 잡음이 제거된 신호의 PESQ와 연산속도를 비교한다.

#### 1. 서론

최근 스마트TV의 보급으로 VoIP를 통한 영상회의를 이용하여 시간과 장소의 제약 없이 업무를 수행하고 회의를 하는 형태의 근무가 가능해졌다. 하지만 각 VoIP를 통한 회의시 사용자가 위치한 환경에서의 교통소음, 인접한 방음소음, 각종 기기들로 부터의 소음은 화상회의의 음성 품질을 저하시킨다. 저하된 음성 품질은 청취자에 있어서 많은 불편을 느끼게 하고 회의의 질을 떨어뜨리게 된다. 따라서 본 논문에서는 스마트 TV의 VoIP를 통한 화상회의의 잡음을 제거하여 쾌적한 회의 환경을 만들 수 있는 잡음제거 알고리즘들의 성능을 비교한다.

#### 2. 잡음제거 알고리즘 비교

본 논문에서 비교하는 잡음제거 알고리즘은 그림1 과 같이 크게 잡음추정과 음성추정으로 구분된다.

##### 2.1 IMCRA(Improved Minima Controlled Recursive Averaging) 방식과 Gaussian분포 기반의 잡음제거 알고리즘(알고리즘1)

평균화된 잡음음성신호를 이용한 잡음추정 방식인 IMCRA 방식 [5]을 이용하여 잡음을 추정한다. 다음 음성 추정에 있어서는 Gaussian 분포기반의 log-spectral의 mean-square error를 최소화 할 수 있는 음성추정 이득함수를 사용하여 잡음이 제거된 음성을 획득한다.[1]

##### 2.2 IMCRA방식과 Gamma분포 기반의 잡음제거 알고리즘(알고리즘2)

알고리즘1은 Gaussian분포기반의 음성추정이득을 사용하지만, 알고리즘2는 minimum Mean-square error(MMSE) 음성추정의 잡음제거 이득을 위해 Generalized Gamma 분포 기반의 음성추정 이득함수를 사용한다.[2]

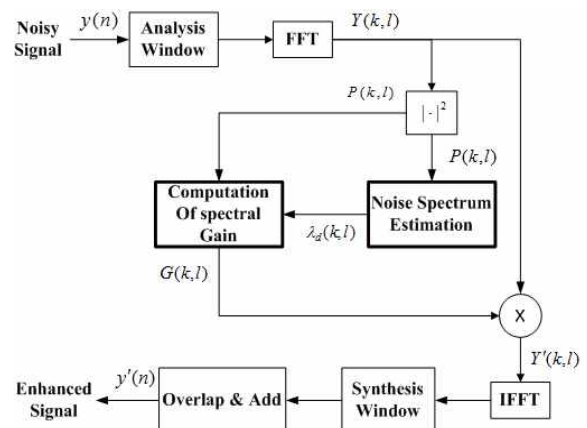


그림 1. 잡음제거 알고리즘의 전체 구성도

음성추정 이득  $G$ 는 식(1)과 같이 유도되고, generalized Gamma 분포를 적용시킨 잡음제거 이득  $G(k,l)$ 은 식(2)과 같이 구할 수 있다.

$$G = \frac{1}{r} \frac{\int_0^{\infty} \int_{-\pi}^{+\pi} a e^{j(\phi-\theta)} f_{X|S}(x|a, \Phi) f_A(a) d\Phi da}{\int_0^{\infty} \int_{-\pi}^{+\pi} f_{X|S}(x|a, \Phi) f_A(a) d\Phi da} \quad (1)$$

$$G(k,l) = \frac{v \zeta(k,l)}{v + \zeta(k,l)} \frac{M\left(v + 1; 2; \frac{\gamma(k,l) \zeta(k,l)}{v + \zeta(k,l)}\right)}{M\left(v; 1; \frac{\gamma(k,l) \zeta(k,l)}{v + \zeta(k,l)}\right)} \quad (2)$$

여기서  $k$ 는  $k$ 번째 스펙트럼 성분,  $l$ 은 시간 프레임 지수,  $v$ 는 generalized Gamma 분포의 Shape 파라미터,  $M(a;b;z)$ 는 CHF(confluent hypergeometric function),  $a, b$ 는 CHF coefficient,  $z$ 는 랜덤변수이다.

**2.3 IMCRA 방식과 Mel-filter를 적용한 잡음제거 알고리즘(알고리즘 3)**

알고리즘3은 알고리즘1, 2와 같은 최소잡음추정, 잡음성분추정의 단계를 거쳐 추정잡음성분  $\lambda_d$ 를 사용하여 음성추정단계를 통해 잡음제거 이득을 구한다. 하지만 음성추정시 Mel-filter[7]를 사용하여  $N$ 개의 차원의 성분을 Mel-filter bank를 통과시켜  $m$ 개의 차원으로 감소시켜 음성추정을 함으로써 연산량을 감소시킨다.

Gaussian 분포기반의  $m$ 개의 차원으로 감소된 잡음제거 이득은 다시 Mel-filter[7]를 통해  $N$ 개의 차원으로 복원하여 사용함으로써, 잡음을 제거한 음성을 획득한다. [3]

**2.4 R&L 알고리즘 (알고리즘 4)**

R&L 알고리즘은 최소잡음성분  $P_{\min}(k,l)$ 을 식(3)와 같이 구한다.

$$P_{\min}(k,l) = \begin{cases} \beta P_{\min}(k,l-1) + \frac{1-\beta}{1-\eta} (P(k,l) - \eta P(k,l-1)) & \text{if } P_{\min}(k,l-1) < P(k,l) \\ P(k,l) & \text{otherwise} \end{cases} \quad (3)$$

여기서  $\eta$ 는  $P_{\min}(k,l)$ 의 가중치를 제어하는 상수이며,  $\beta$ 는 스무딩 파라미터이다. 기존의 최소잡음성분 추정은 정해진 윈도우크기만큼의 스펙트럼을 비교하였지만, R&L 알고리즘에서는 식(3)과 같이 프레임 단위로 비교하여 연산량이 줄어드는 효과를 기대할 수 있다.

음성추정 구간에서의 R&L 알고리즘은 다른 알고리즘과 달리 사전 사후확률을 구하지 않고 입력된 잡음음성과 위스펙트럼에서 추정된 잡음  $\lambda_d$ 를 차감시킴으로써, 음성 스펙트럼  $E(k,l)$ 을 추정하고, 가중치  $\mu_k$ 를 잡음스펙트럼과 음성스펙트럼을 이용하여 잡음제거이득을 구하게 된다. 이는 알고리즘 1, 2, 3과 달리 사전확률, 사후확률의 연산을 하지 않음으로서 연산량을 줄일 수 있다.[4]

$$E(k,l) = \max\{|Y(k,l)|^2 - E(k,l), \nu\lambda_d(k,l)\} \quad (4)$$

$$G(k,l) = \frac{E(k,l)}{E(k,l) + \mu_k \lambda_d(k,l)} \quad (5)$$

**3. 실험 결과 및 고찰**

각 알고리즘의 성능을 측정하기 위해 기존의 실제 다양한 SNR과 잡음환경(White, Bubble, Car, Train)에서 녹화된 20분길이의 화상회의 영상 20개를 사용하여 PESQ 측정결과와 연산속도를 비교하여 그림2, 표1에 각각 나타내었다.

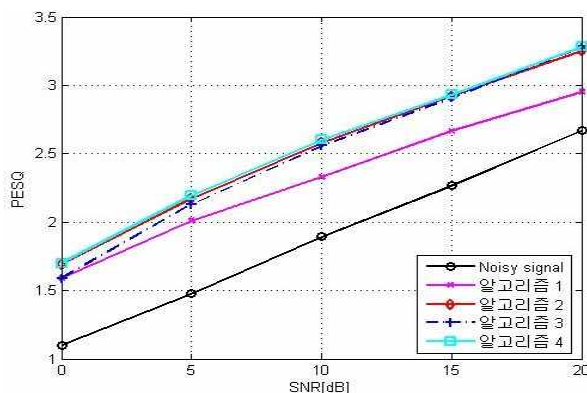


그림 2. SNR에 따른 각 알고리즘의 PESQ 비교

알고리즘	연산속도(ms)
알고리즘 1	251.210
알고리즘 2	248.164
알고리즘 3	31.509
알고리즘 4	30.418

표 1. 각 알고리즘의 연산속도 비교

실험결과 알고리즘1, 2, 3이 PESQ가 높게나와 좋은 잡음제거 성능을 보였고, 연산속도에서는 알고리즘3, 4가 다른 알고리즘에 비해 7배 이상 높게 나왔다.

**4. 결론**

본 논문에서는 스마트TV용 VoIP 회의 기능을 위한 잡음제거 알고리즘을 비교하였다. VoIP 시스템은 적은 연산량이 요구되며, 적은 연산량과 다른 알고리즘과 동등한 PESQ를 가지는 알고리즘4가 VoIP를 통한 영상회의에 적용하기 적합한 잡음 제거 알고리즘이라고 판단 된다.

**감사의 글**

이 논문은 2011년도 정부(교육과학기술부)의 재원으로 한국연구재단의 기초연구사업 지원을 받아 수행된 것임(2011-0004311).

**5. 참고문헌**

- [1] Cohen Israel, "Noise estimation by minima controlled recursive averaging for robust speech enhancement." *IEEE Signal Process*, 9 (1), 12-15, 2002.
- [2] R.C. Hendriks, J.S. Erkelens, J. Jensen, and R.Heusdens, "Minimum mean-square error amplitude estimators for speech enhancement under the generalized Gamma distribution," *Proc. International Workshop on Acoustic Echo and Noise Control(IWAENC)*, vol. 10, pp. 1-4, Sept. 2006.
- [3] The xiph open source community. <http://www.speex.org> (July 14, 2011)
- [4] Loizou P. , Sundarajan R. (2006) "A Noise estimation Algorithm for highly non - stationary Environments", *speech Communication* 48 (2006) Science direct pp. 220-23.
- [5] Y. Ephraim, D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Trans. Acoust. Speech Signal Process.* ASSP-33 (2) (April 1985) 443 - 445.
- [6] Cohen, "On speech enhancement under signal presence uncertainty," *Proceedings of the 26th IEEE International Conference on Acoustics Speech, and Signal Processing*, ICASSP-01, Salt Lake City, Utah, 7 - 1 May 2001.
- [7] Tokyo Institute of Technology, Yokohama, Japan, "Cepstral Analysis Synthesis on th Mel Frequency Scale," *IEEE International Conference on ICASSP '83*, pp. 93-96, Apr 1983.