

심리음향모델을 적용한 난청 보정 알고리즘의 연구

노형철, 김헌중, 한헌수, 차형태
송실대학교 전자공학과

The Study of the Sensorineural Hearing Loss Compensation Algorithm using Psychoacoustics Model

Hyungchul Roh, Hunjoong Kim, Hernsoo Hahn, Hyungtai Cha
Dept. of Electronics Engineering, Soongsil University

E-mail : hcroh@mmslab.soongsil.ac.kr

요 약

본 논문에서는 청각 장애인의 보다 향상된 보청 환경을 조성하고자 청각손실을 심리음향 모델을 적용하여 감음 신경성 난청을 보정하는 알고리즘을 제안한다.

제안한 알고리즘에서는 난청의 유형은 나이에서부터 중추 뇌에 걸친 감음계와 신경계의 장애에서 비롯되는 감음신경성 난청(sensorineural hearing loss)으로 주파수 영역상에서 MTH(minimum hearing threshold)가 균일하지 않게 상승하게되어 가청영역이 좁아지는 문제점을 해결하기 위한 방법으로 각각의 주파수 밴드마다 멀티밴드 압축 알고리즘을 적용하였다. 그러나 이 경우 각각의 주파수 밴드에 따른 서로 다른 가청 영역의 영향에 의한 변형된 스펙트럼 모양으로 인해 spectral contrast reduction과 변형된 마스킹 특성으로 인해 음성 변별력에 제한을 가하게 된다. 이것은 주변 주파수 성분들에 의한 마스킹 효과에 의한 것으로, 신호에 대한 난청인이 느끼는 지각 영역(perceptual domain)에서의 해석과 심리음향 모델 파라미터를 통한 보청기의 개발이 이루어져야 하며, 본 논문에서 그 알고리즘을 적용하였다.

I. 서 론

인간에게 있어 청각의 손실은 의사 전달과 정보 교환의 수단이 부분적으로 단절됨을 의미하며 이것은 일상 생활에서의 많은 불편을 감수하게 할 뿐 아니라 정서

적, 심리적으로도 적지 않은 영향을 미친다[1]. 따라서 청각 손실로 어려움을 겪고 있는 난청인 들은 보청기가 필요하며 그 역할 또한 중요하다. 현대사회의 복잡한 생활환경은 난청인의 수를 더욱 증가시켜 앞으로는 보청기의 수요가 보다 증가할 것으로 예상되며 향상된 보청기의 개발이 요구되어질 것이다[2].

사람의 청각 특성은 정상인의 경우 음성신호는 가청 한계보다 높기 때문에 모든 음성을 받아들인데 문제가 없지만, 청각 손실이 일어나면 손실 정도에 따라서 가청한계가 변하게 된다.

난청의 유형은 크게 전음성 난청(傳音性 難聽: conductive hearing loss)과 감음신경성 난청(感音信經性 難聽: sensorineural hearing loss)으로 나누어 볼 수 있다[1][2]. 본 논문에서는 선형 증폭형 보청기를 쉽게 이용할 수 있는 전음성 난청을 배제하고 보다 복잡한 비선형적이고 시변적인 신호처리를 필요로 하는 감음 신경성 난청에 대해 새로운 알고리즘을 제시하였다. 이 알고리즘으로 본 논문에서는 주파수 대역 특성을 고려하고 심리음향모델을 적용하여 고주파 영역에서 두드러지게 나타나는 난청인의 가청한계가 마스킹 곡선보다 낮은 특성을 소프트웨어를 통하여 실험적으로 입력되는 음성 신호 레벨에 따라 난청인의 보청기 이득을 적절히 변화시켜서 성능을 개선하는 알고리즘은 연구하였다.

II. 동적 범위 보상

청각 손실 외에 난청인의 가장 큰 문제점 중의 하나

는 순음 가청한계(pure-tone threshold)와 불편한 수준(discomfort level) 사이의 범위가 줄어드는 동적 범위 보상(Dynamic Range Compensation) 문제이다. 감음 신경성 난청의 경우 주파수 영역 상에서 MHT(Minimum Hearing Threshold)가 균일하지 않게 상승하게 되어 가청 영역이 좁아지게 된다. 그 결과 이러한 가청영역감소와 변형된 스펙트럼의 대조로 인해 음성 인식이 저하되고 청자가 느끼는 소리의 주관적 크기가 정상인의 경우에 비해 크게 변화하는 라우드니스누가(loudness recruitment) 현상이 일어나게 되어, 신호음 레벨의 작은 상승이 청자에게는 크게 증가한 것으로 느껴지게 된다[3].

이러한 문제점을 보상하기 위해 난청인의 순음 audio gram에 근거한 압축 알고리즘이 적용되는데, 이러한 동적 범위 보상을 통한 보정기는 라우드니스 스케일링 함수의 기울기를 변화시킬 수 있으므로 모든 입력 레벨에 대해서 인지 측면에서 적용할 수 있게 하는 라우드니스를 복원할 수 있다.

본 논문에서는 음성신호와 배경음의 비선형(non-linear)적이고 시변(time-varying)적인 특성을 고려하여 각각의 주파수 밴드에 대해서 각기 다른 입력 레벨에 대해서 이득 값을 조정하는 보상 방법을 적용하였다. 임의의 채널에 대한 동적 범위 보상은 다음 그림 1.에서와 같이 입력 신호의 에너지를 난청인의 축소된 동적 범위로 매핑 시키게 된다.

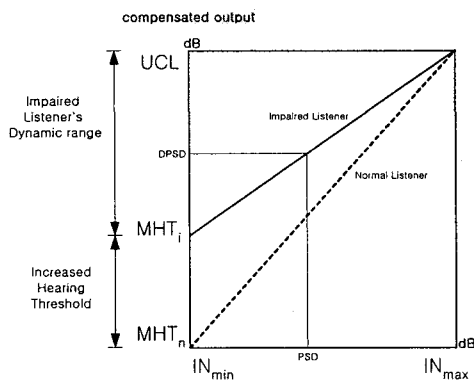


그림 1. Dynamic range scaling function

여기에서 LSP(Loudness Scaling Procedure)를 위한 파라미터인 입력 신호 에너지 범위($IN_{min} \sim IN_{max}$)는 동적 범위 보상을 위한 프로세서 신호 표현을 위한 Bit 해상도를 통해 결정되어 질 수 있으며, MHT_n 은 정상인들을 통한 통계적인 측정을 통해 결정되어진 값 또는 ATH(Absolute threshold)로 표현될 수 있으며, UCL과 MHT_i 는 난청인의 각각의 주파수 채널에 대한 순음 audio gram 측정을 통한 audio profile을 통해서 구해지게 된다.

III. 마스크 특성에 기초한 음성 개선 (Speech Enhancement Based on Masking Properties)

난청인은 이와 같이 최소 가청 한계의 상승과 동적 범위가 줄어든 것 외에 감소된 주파수 변별력에 의해 줄어든 음성 인식을 겪게되는데, 이것은 주변 주파수 성분들에 의한 마스크 효과에 의한 것으로, 이러한 현상은 배경 잡음이 존재할 때 더욱 심하게 나타난다[4].

이러한 난청인의 지각 특성을 해석하기 위한 시뮬레이션 환경은 다음과 같다. 스펙트럼 분석은 입력 음성 데이터를 8 kHz로 샘플링하고 16bit의 해상도를 갖는 32ms분량의 256 PCM 데이터를 hanning window와 FFT를 통해 이루어진다.

Bark Index	Number of FFT bins	Frequency Range	Band Index	Center Frequency	MHT _i
1	4	0 - 93.75	
2	3	93.75 - 187.50			
3	3	187.50 - 281.25	1	281.25	9.2
4	4	281.25 - 406.25			
5	3	406.25 - 500.00	2	500	10
6	4	500.00 - 625.00			
7	4	625.00 - 750.00	3	750	15
8	5	750.00 - 906.25			
9	5	906.25 - 1062.50	4	1062.50	17
10	6	1062.50 - 1250.00			
11	6	1250.00 - 1437.50	5	1437.50	25
12	8	1437.50 - 1687.50			
13	8	1687.50 - 1937.50	6	1937.50	32
14	11	1937.50 - 2281.25			
15	13	2281.25 - 2687.50	7	2687.50	40
16	16	2687.50 - 3187.50			
17	20	3187.50 - 3812.50	8	3812.50	35
18		3812.50 - 4000.00			

표 1. The Critical Band and 8 Critical Band Channels at the 8 kHz Sampling Rate and Frame Size N= 256

다음 그림 2.는 표 1.에서와 같은 임계 대역에 근거한 중심 주파수를 통한 8개의 밴드를 통해 근사화된 난청인의 audio profile을 나타낸다. 이 때 난청인의 MHT_i (minimum hearing threshold)는 정상인의 MHT_n 에 비해 그 차이만큼 상승한 것으로 해석할 수 있다. 그러므로 입력 신호에 대한 난청인의 지각된 신호 에너지는 정상인의 MHT_n 에 비해 상승된 난청인의 임계치의 증가량만큼 감소시킴으로서 정상인의 마스크 모델을 통해서 해석할 수 있다.

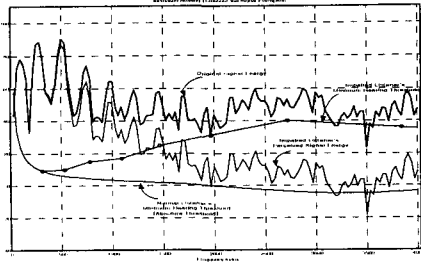


그림 2. Impaired listener의 audio profile과 perceived signal energy

$$THD = MHT_i - MHT_n$$

$$IPSD = PSD - THD \quad (\text{in dB})$$

단, 여기서 THD 는 난청인과 정상인의 최소 청각 임계치의 차를, PSD 는 입력 신호의 전력 스펙트럼, $IPSD$ 는 난청인의 입력 신호에 대한 지각적인 전력 스펙트럼을 나타낸다.

지각적인 영역에서의 마스킹 특성을 해석하기 위한 절차는 마스킹 성분이 순음 또는 잡음인가에 따라 마스킹 곡선이 달라지므로 스펙트럼 정보로부터 순음 성분과 잡음 성분을 찾아내야 한다. 다음으로 주요 마스크를 찾고, 개별적인 마스킹 임계치를 계산하고, 개별 마스킹 임계치와 가청 임계치의 합으로 구해지는 전체 마스킹 임계치를 계산한다[5].

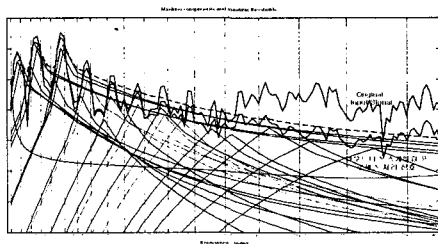


그림 3. The Processed signals with dynamic range compensation

그림 3은 청자에게 정상 라우드니스 지각을 복원 시켜주기 위해 난청인의 audio profile에 근거해서 각기 다른 주파수 밴드에 대해서 입력신호의 에너지 레벨에 따라, 지각측면에서 adaptive하게 라우드니스를 복원한 결과와 마스킹 임계치이다. 그림 3은 원래 신호의 에너지 레벨과 라우드니스 보상된 지각된 신호의 에너지 레벨과의 차이는 각 주파수 밴드에서의 정상인과 난청

인의 동적 범위의 차이만큼 전이된 것을 나타내고 있다. 그림에서 살펴볼 수 있듯이 각 주파수 밴드에 따른 서로 다른 동적 범위의 영향으로 청자에게 인식되어지는 신호의 에너지가 원래 신호에너지에 비해 변형된 스펙트럼 대비와 마스킹 특성을 갖게되어 많은 부분의 주파수 구성요소들이 마스킹되어 음성 변별력에 제한을 가하고 있다.

이와 같은 문제점은 주파수 domain에서는 어떻게 보상할지 모호해지지만, 이것을 사람의 청각인지 특성을 이용한 지각 domain에서 각각의 주파수 밴드에 대한 전력 레벨과 마스킹 임계치 사이의 거리로 정의되는 SMR(signal-to-masking ratio)를 통해 개선시킬 수 있다.

다음 그림 4.은 이러한 SMR을 이용하여 원래 신호의 마스킹 특성과 스펙트럼 대비를 갖도록 라우드니스 스케일링된 신호를 보상시켜준 것의 특성을 나타낸다.

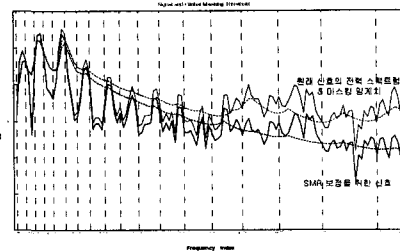


그림 4.a. Masking threshold comparison with SMR compensation

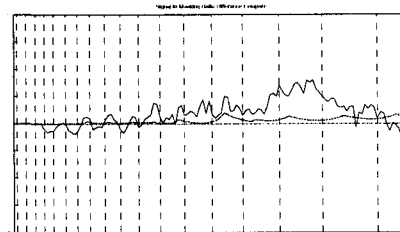


그림 4.b. SMR compensated signal's SMR difference

그림 4.a.은 이러한 SMR 보정된 신호에 대한 마스킹 임계치의 비교이다. 그림에서 볼 수 있듯이 전체적으로 라우드니스 스케일링된 신호의 전력 스펙트럼이 원 신호의 SMR에 맞게 reshaping이 이루어져 원신호와 같은 마스킹 특성을 복원하여 원래 신호같이 인식할 수 있게 되었다. 또한 그림 4.b.에서의 원래신호와 SMR difference 비교에서 볼 수 있듯이 전체적인 변화가 낮

고 difference 값이 전 대역에 걸쳐 일정한 값을 유지해 지각된 신호는 원래신호의 스펙트럼 대비를 유지할 수 있음을 확인할 수 있다.

IV. 실험 및 결과

제한된 성능 평가를 위한 실험은 난청 시뮬레이션 (hearing impairment simulation)을 통해 1년 이상 음성 또는 오디오 신호처리 분야에서 종사한 정상 청력을 가진 전문인을 대상으로 어음 명료도 검사(speech discrimination test)를 통해 수행되었다.

정상인을 통한 난청 시뮬레이션 환경구성은 다음 그림 5와 같다.

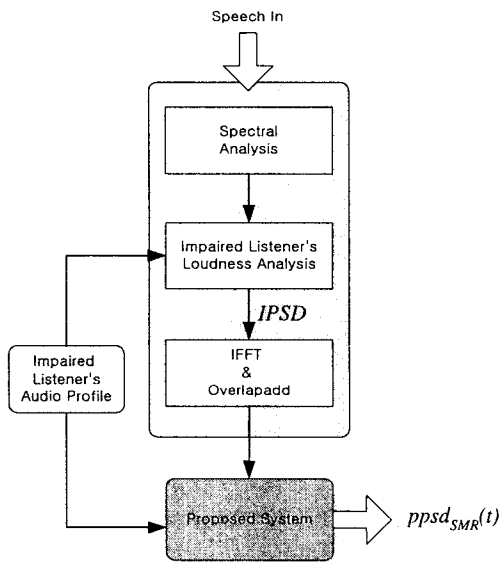


그림 5. 난청 시뮬레이션의 Block diagram

실험은 20대에서 40대의 정상인 남녀의 단음절 어음 (monosyllabic word)과 sentence를 통해 난청인이 $ppsd_{SMR}$ (SMR compensation된 신호에 대한 인지 신호)를 통해 실험 참가자의 어음 변별력을 측정하였다.

결과는 SDT 점수로 평균 97%로 나은 보청 환경을 제공할 수 있음을 알 수 있다.

V. 결론

본 논문에서는 디지털 보청 방식에 대한 보정 알고리즘으로 심리음향 모델을 적용하는 알고리즘을 연구하였다. 성능 평가에서 본 것처럼 난청인에게 큰 도움이 될 수 있는 가능성을 확인할 수 있다.

참고 문헌

- [1] Fred H. Bess et al, Audiology: The Fundamentals, *Williams & Wilkins*, 1995.
- [2] Maryanne Tate, Principles of Hearing Aid Audiology, *Chapman & Hall*, 1994.
- [3] Hanet C. Rutledge, "Compensation for Recruitment of Loudness in Sensorineural Hearing Impairments Using a Sinusoidal Model of Speech", *Proc. IEEE ICASSP*, pp. 3641-3644, 1991.
- [4] C. I. Berlin, Hair Cells and Hearing Aids, Singular Publishing Group, 1996.
- [5] ISO/IEC 13818-3, "Information Technology- Generic Coding of Moving Pictures and Associated Audio: Audio", pp 76-81, 1994.
- [6] E. Villchur, "Signal processing to improve speech intelligibility in perceptive deafness", *Journal of the Acoustical Society of America*, vol. 53, pp. 1646-1657, June 1973.
- [7] R. P. Lippmann, L. D. Braida, and N. I. Durlach, "Study of multichannel amplitude compression and linear amplification for persons with sensorineural hearing loss", *Journal of the Acoustical Society of America*, vol. 69, pp. 524-534, Feb. 1981.
- [8] Thomas Lunner and Johan Hellgren, "A Digital Filterbank Hearing Aid - Design, Implementation and Evaluation", *Proc. IEEE ICASSP*, pp. 3661-3664, 1991.
- [9] F. Asano. et al, "A digital hearing aid that compensates loudness for sensorineural impaired listeners", *Proc. IEEE ICASSP*, pp. 3625-3628, 1991.