청각 시스템의 특징을 이용한 음성 명료도 향상

*이상훈, 정 홍
포항공과대학교 전자전기공학과
e-mail : god1710@postech.ac.kr, hjeong@postech.ac.kr

Speech Enhancement based on human auditory system characteristics

*Sang-Hoon Lee, Hong Jeong
Department of Electronic and Electrical Engineering
Postech

Abstract

본 논문에서는 인간 청각 시스템의 특징을 이용한 음성명료도 향상 알고리즘을 제안한다. 기존의 연구들은 음성과 잡음이 같이 섞여 있는 Single-Channel에서의 명료도 향상의 대해 주로 다루었다. 하지만 잡음에 섞이기 전의 깨끗한 음성과 주변 잡음이 분리된 Dual-Channel에서의 명료도 향상에 관한 연구는 거의 다루어지지 않았다. 본 논문에서 음성을 잡음이 섞이기 전에 미리 강화시켜 나중에 잡음에 섞였을 때 명료도가 강화되도록 하는 방법을 제안한다. 인간 청각 시스템의 마스킹 효과를 적절히 이용하여 음성을 강화시키는 방법을 사용하였다. 실험 결과 이 방법은 단순히 볼륨만을 높이는 방법에 비해 명료도가 더 향상되는 것으로 나타났다.

I. Introduction

음성 명료도 향상에 관련된 지금까지의 연구들은 대부분 Single-Channel의 경우만 다루어져 왔다. 하지만 일상적인 휴대전화 통화 환경이나 군사 목적의 무선통신 등 실제 상황에서는 Dual-Channel 음성 명료도 향상이 매우 필요하다. 휴대전화 통화 환경에서 주변의 잡음이 통화에 방해되는 경우, 즉 주변 잡음으로 인하여 상대방의 음성을 알아 들을 수 없는 경우 대부분의 사용자들은 볼륨을 높임으로써 이 상황을 해결하려 하지만 주변 잡음이 많이 시끄러울 경우 이 방법은 금방 한계에 부딪히게 된다. 이 논문에서는 주변 환경 잡음에 따라 음성을 강화시켜 명료도를 향상시키는 방법을 제시하는데 이 방법은 단순히 볼륨 또는 파워를 올리는 것보다 명료도를 더 향상시킨다. 우리가 제안한 알고리즘은 dual-channel 환경에서 음성을 잡음에 알맞게 강화시켜 명료도를 향상시키는 방식이다. 이 때 인간 청각시스템의 마스킹 효과를 감안하여 잡음에 취약한 음성 부분을 선택적으로 강화하는 방식을 사용하였다.

II. Algorithm

본 논문에서 제안된 음성명료도 향상 알고리즘은 다음과 같다. 먼저 깨끗한 음성과 주변 잡음을 입력으로 받아 FFT 한 후 주파수 도메인에서 잡음의 Masking Threshold를 형성한다. 크리티컬 밴드별로 음성을 분석하여 음성이 주변 잡음의 영향을 받는 상황, 즉 잡음이 음성을 Masking하는 밴드에서는 음성을 강화하고 그 외에의 밴드는 그대로 나두는데 이 때 주변 잡음 상황만을 기준으로 하여 음성을 강화하면 음성이 잡음화 되는 현상이 일어난다. 이 문제를 해결하기 위하여 음성 신호 중에서 의미 있는 부분, 즉 유성음의 경우 포먼트 부분을 더욱 강조하는 방법을 사용하여 음성의 잡음화 문제를 해결하였다. 노이즈의 Masking Threshold는 [1]의 과정을 따라서 생성되었다. 우리는

음성 신호 중에서 의미 있는 밴드를 강조하기 위하여 다음과 같은 추정값을 정의하였다.

$$\gamma(i) = \Lambda\left(\frac{\sum_{b=1}^{B_i} s^2(b)}{\sum_{w=1}^{W} s^2(w)}\right) \quad \text{Eq. (1)}$$

여기서 s는 음성 신호를 FFT한 결과이며 i는 크리티컬 밴드의 index이며 B_i는 크리티컬 밴드의 크기이다. 그리고 $\Lambda(k)$는 soft decision을 위한 함수이다.

$$\Lambda(k) = \frac{1+\exp(k)}{2} \quad \text{Eq. (2)}$$

노이즈의 masking Threshold 정보와 음성 신호의 의미 추정값을 통하여 다음과 같은 band gain을 구하여 음성 신호에 적용시키고 energy normalization을 거치면 음성 신호는 명료도가 향상된다.

$$G(i) = \gamma(i) \times \left(\alpha \cdot R + \beta \frac{\sum_{b=1}^{B_i} n^2(b)}{\sum_{m=1}^{M} n^2(m)}\right) \quad \text{Eq.(3)}$$

여기서 $\gamma(i)$는 음성 신호 의미의 추정값을 반영하기 위해 곱해졌으며 α는 SNR reference factor이고 β는 noise distribution coefficient이다[2].

Ⅲ. Experiment

A. Test material

4명의 speaker가 실험용 음성 database를 만드는데 참여하였다. 각각의 speaker 는 77쌍의 1음절 CVC(consonant-vowel-consonant) 단어를 녹음하도록 하였다. 각각의 쌍은 CVC 중 하나만이 틀리게 구성되었다. 실험에 쓰일 노이즈는 Noisex-92 database에서 8가지 종류를 선택하였다.

B. Test method

실험 방법은 노이즈가 한쪽에서 플레이 되는 상황에서 다른 쪽에서는 하나의 1음절 단어를 들려준 후 해당되는 1쌍의 단어를 화면에 디스플레이 한 뒤 어떤 단어가 발화되었는지를 선택하도록 하였다. 이 때 노이즈는8가지가 랜덤으로 번갈아 나오도록 하였으며 음성은 다음과 같이 3종류로 처리된 음성이 랜덤하게 나오도록 하였다.

1) no processed speech
2) volume controlled speech
3) enhanced speech (our algorithm)

그리고 화자도 랜덤하게 선택되도록 하였다. 2)의 경우 3)과 같은 크기의 파워를 가지도록 볼륨을 조절한 것이다

C. Test result

그림 4는 전체적인 실험 결과이다. 아무런 처리를 하지 않은 음성의 평균적인 명료도는 74.3%였고, 우리 알고리즘에 의해 처리된 음성의 평균적인 명료도는 96.5% 로 차이는 22.2%만큼 났다. 그리고 우리 알고리즘에 의한 음성의 명료도 향상은 단순히 볼륨만 높인 음성보다 10.1% 향상되었다.

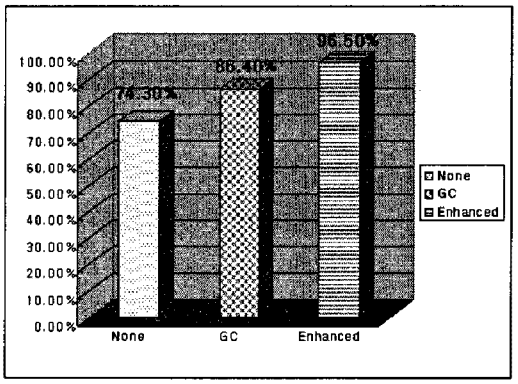

그림 1. 명료도 실험 결과

Ⅳ. Conclusion

본 논문에서는 dual-channel 환경에서 실시간으로 잡음을 분석하여 음성을 강화하는 알고리즘을 제시하였다. 제안된 알고리즘의 가장 큰 장점은 다음과 같다.

1) 계산량이 적다(가장 많은 연산을 요구하는 부분은 FFT 부분이다. 나머지 부분의 계산량은 사소하다)
2) 음성 신호중에서 의미 있는 부분이 선택적으로 강조되었다. 예를 들어 유성음의 경우 포먼트가 많이 강조되었다.
3) 잡음에 adaptive하다. 음성이 잡음보다 셀 경우 음성을 조금만 변화시키고 반대 경우 음성을 많이 변화시켜서 명료도를 향상시킨다.

참고문헌

[1] J. D. Johnston, "Transform coding of audio signal using perceptual noise criteria", IEEE J. Select. Areas Commun., Feb. (1988). Vol. 6, pp. 314-323.

[2] S,H Lee; H. Jeong, "Real-time speech intelligibility enhancement based on the background noise analysis", The IASTED International Conference on Signal Processing and Pattern Recognition(SPPRA 2007), Feb. 14-16, 2007