

## 음성신호 에너지 및 LP 분석 기반 적응적 배경음혼합 알고리즘

\*강진아 \*전찬준 \*김홍국 \*\*김명보 \*\*김지운

\*광주과학기술원 정보통신공학부

{jinari, cjchun, hongkook}@gist.ac.kr

\*\*삼성전자 디지털이미징사업부 개발팀

{kmbo.kim, jiwoon.kim}@samsung.com

## An Adaptive Background Sound Mixing Algorithm Based on Energy and LP Analysis of Speech Signal

\*Jin Ah Kang \*Chan Jun Chun \*Hong Kook Kim \*\*Myeong Bo Kim \*\*Ji Woon Kim

\*School of Information and Communications, GIST

\*\*Digital Imaging Business R&amp;D Team, Samsung Electronics

## 요약

본 논문에서는 제작된 콘텐츠에 배경음을 간편하고 효과적으로 혼합하기 위해서 녹음된 신호(전경음)를 분석하여 배경음 에너지를 적응적으로 조절하는 배경음혼합 알고리즘을 제안한다. 이를 위해, 제안된 알고리즘은 등청감 곡선 (equal-loudness curve) 및 linear prediction (LP) 분석에 기반하여 전경음신호의 청감 에너지 및 음성신호 존재여부를 결정한다. 이에 따라 전경음에 음성신호가 존재하는 경우에는 음성이 명확하게 들릴 수 있도록 혼합된 배경음의 에너지를 하향 조절하고, 반대로 전경음에 음성신호가 존재하지 않는 경우에는 배경음이 명확하게 들릴 수 있도록 혼합된 배경음의 에너지를 상향 조절한다. 제안된 알고리즘의 효율성을 검증하기 위해, 고정 가중치를 이용하여 배경음을 혼합하는 경우와의 음질 신호도 조사를 실시한 결과, 제안된 알고리즘에 대한 높은 신호도를 보였다.

## 1. 서론

최근 사용이 편리하고 간편한 휴대용 디지털 캠코더 및 카메라의 보급과 youtube[1]와 같은 동영상 콘텐츠 공유 플랫폼의 확산은 일반 사용자들에 의한 다양한 콘텐츠 제작 및 공유 문화를 형성하는 기반이 되었다. 이에 따라, 디지털 캠코더 및 카메라, 또는 콘텐츠 공유 플랫폼 제공업체들은 일반 사용자들이 보다 쉽고 편리하게 콘텐츠를 제작하고 감상할 수 있도록 다양한 부가기능을 제공하고 있다. 이러한 기능 중에 하나로 배경음혼합 기능은 제작된 콘텐츠를 보다 효과적으로 표현 또는 감상할 수 있도록 해 준다. 그러나 TV 등의 대중매체를 통해 보여지는 콘텐츠와 같이 배경음을 효과적으로 혼합하기 위해서는 별도의 장비 및 기술이 필요로 된다. 이에 따라 본 논문에서는 제작된 콘텐츠에 보다 쉽고 간편하게 배경음을 혼합하는 알고리즘을 제안한다.

본 논문에서 제안되는 배경음혼합 알고리즘은 녹음된 신호(전경음)에 음성신호가 존재하는 구간에 대해서는 배경음 에너지를 낮게 조절하고, 음성신호가 존재하지 않는 구간에 대해서는 배경음 에너지를 높게 조절하도록 설계된다. 이를 위해, 제안된 알고리즘은 등청감 곡선 (equal-loudness curve)[2]과 linear prediction (LP)[3] 분석을 이용하여 전경음신호의 에너지 및 상태를 결정한다. 이를 통해 결정된 전경음 신호 상태는 혼합되는 배경음의 에너지를 조절하는데 사용된다.

## 2. 제안된 배경음혼합 알고리즘

그림 1은 제안된 음성신호 에너지 및 LP 분석에 기반한 적응적 배경음혼합 알고리즘의 전체 흐름도를 보여준다. 그림에서 보는 바와 같이, 제안된 알고리즘은 전경음과 배경음 두 신호를 입력받아 전경음 원음에 배경음을 가중합산한 혼합신호를 출력한다. 이를 위해, 먼저 두 입력신호를 주파수 영역으로 변환한 뒤 등청감 곡선을 적용하여 에너지를 구한다. 등청감 곡선은 사람의 귀가 느끼는 주파수 민감도 특성에 따라 청감적으로 동일하게 느껴지는 음압레벨을 측정된 것으로, 제안 알고리즘에서는 청취자에게 들리는 청감 에너지를 예측하기 위해 A-weighting curve[2]를 적용한다. 다음으로, 전경음신호에 대한 LP 분석을 수행하여 LP 잔차신호 에너지를 구한다. 이는 다음의 signal activity classification (SAC) 단계에서 전경음에 음성신호의 존재여부를 판별하는데 사용된다.

그림 2는 SAC 흐름도를 나타낸 것으로, 먼저 전경음 원신호 에너지에 대한 LP 잔차신호 에너지 비율,  $R_1$ 을 구한다. 이때,  $R_1$ 이  $thres_1$ 보다 크면 전경음 상태는 비음성구간(inactive)으로 판별된다. 반대로,  $R_1$ 이  $thres_1$  이하이면 전경음 음성구간(active) 평균 에너지에 대한 입력신호 에너지 비율인  $R_2$ 를 구한다. 이때,  $R_2$ 가  $thres_2$ 보다 크면 전경음 상태를 음성구간으로 판별하고, 그렇지 않으면 중간구간

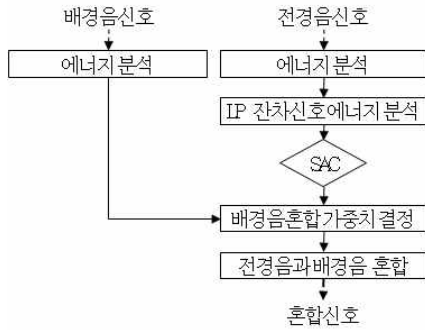


그림 1. 제안된 배경음혼합 알고리즘 흐름도.

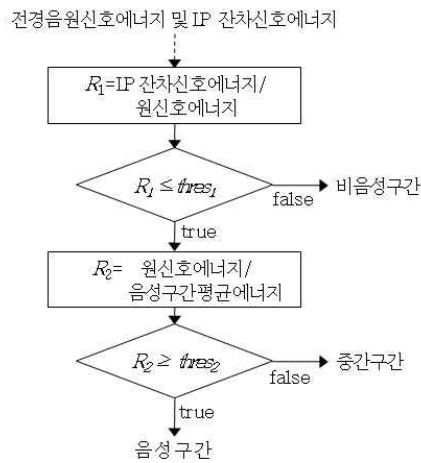


그림 2. 전경음 SAC 흐름도.

(mid-active)으로 판별한다. 즉, 음성신호와 같이 인접 신호간 상관도가 높아서 LP 잔차신호 에너지가 원신호 에너지 대비 어느 정도 작고, 동시에 입력신호 에너지가 평균 음성구간 에너지 대비 어느 정도 이상이 되면 전경음 SAC 상태를 음성구간으로 판별하고, 그 외의 경우에 대해서는 각각 중간구간과 비음성구간으로 판별한다.

다음으로, 판별된 전경음 SAC 상태에 따라 배경음신호 가중치,  $w(n)$ ,를 다음 식과 같이 구한다.

$$w(n) = \begin{cases} k_a \frac{E_f(N, n)}{E_b(n)}, & N: \text{음성구간} \\ k_m \frac{E_f(N, n)}{E_b(n)}, & N: \text{중간구간} \\ k_i \frac{\overline{E_f}(N-1)}{E_b(n)}, & N: \text{비음성구간} \end{cases} \quad (1)$$

식 (1)에서, 음성구간과 중간구간에 대해서는 배경음신호 에너지,  $E_b(n)$ ,가 전경음신호 에너지,  $E_f(N, n)$ ,에 대해 각각  $k_a$ 와  $k_m$ 의 비율이 되도록 구하고, 비음성구간에 대해서는  $E_b(n)$ 이 이전에 발생한 전경음 음성구간 평균 에너지인  $\overline{E_f}(N-1)$ 에 대해  $k_i$ 의 비율이 되도록 구한다. 따라서, 현재 전경음신호 상태가 음성구간 또는 중간구간인 경우에는 배경음신호 에너지를 현재 전경음신호 에너지의 일정 비율이 되도록 하향 조절하고, 반대로 현재 전경음신호 상태가 비음성구간인 경우에는 이전 전경음 음성구간의 평균 에너지를 반영하여 배경음신호 에너지를 상향 조절한다.

표 1. 청취 선호도 조사 결과.

전경음	배경음	선호도 (%)	
		고정가중치 기반 혼합 알고리즘	제안된 알고리즘
대화1	연주곡1	10	90
대화2	연주곡2	30	70
대화3	연주곡3	0	100
대화4	연주곡4	10	90
대화5	연주곡5	10	90
대화6	가요1	10	90
대화7	가요2	0	100
시낭송	연주곡6	10	90
내레이션1	새소리	10	90
내레이션2	연주곡7	20	80
평균		11	89

### 3. 성능평가

제안된 배경음혼합 알고리즘의 효용성을 검증하기 위해서는 고정 가중치로 배경음을 혼합하는 방법과의 제안된 방법과의 청취 선호도를 조사하였다. 먼저 제안된 알고리즘에서는  $thres_1$ ,  $thres_2$ 는 각각 0.25, 0.5로 설정하였고,  $k_a$ ,  $k_m$ ,  $k_i$ 는 각각 0.25, 0.4, 0.9로 설정하였다. 또한 LP 차수는 16차로 설정하였다. 다음으로, 청취 선호도 조사는 32 kHz 샘플링 주파수와 스테레오 채널로 녹음된 전경음과 배경음 10쌍을 표 1과 같이 장르별로 구성하여 고정 가중치를 이용한 혼합음원과 제안 알고리즘을 이용한 혼합음원을 남녀 10명에게 비교청취하도록 하였다. 이에 대한 결과는 표 1에 나타내었으며, 제안된 알고리즘에 의한 혼합음원을 더욱 선호하는 것으로 확인되었다.

### 4. 결론

본 논문에서는 전경음 음성신호의 에너지 및 LP 분석을 수행하여 그에 따라 적응적으로 배경음 에너지를 조절하는 배경음혼합 알고리즘을 제안하였다. 제안된 알고리즘은 전경음 에너지 및 LP 분석결과에 기반하여 전경음 상태를 음성구간과 중간구간, 비음성구간으로 분류하였고, 이에 따라 음성구간과 중간구간일 때에는 입력된 전경음 에너지를 기준으로 배경음 에너지를 하향 조절하고, 비음성구간일 때에는 이전 전경음 음성구간의 평균 에너지를 반영하여 상향 조절하도록 설계되었다. 제안된 알고리즘의 효용성을 검증하기 위해, 고정 가중치를 이용한 배경음 혼합음원과 청취 선호도 조사를 수행한 결과, 제안된 알고리즘에 의한 배경음을 혼합할 경우에 대한 선호도가 높았다.

### 감사의 글

이 논문은 2010년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2010-0000135).

### 참고문헌

[1] <http://www.youtube.com/>.  
 [2] H. Fletcher and W. A. Munson, "Loudness, its definition, measurement and calculation," *Journal of the Acoustical Society of America*, vol. 5, pp. 82-108, Oct. 1933.  
 [3] T. P. Barnwell, III., "Recursive windowing for generating autocorrelation coefficients for LPC analysis," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 29, no. 5, pp. 1062-1066, Oct. 1981.