Journal of The Institute of Electronics and Information Engineers Vol.53, NO.11, November 2016 http://dx.doi.org/10.5573/ieie.2016.53.11.082
ISSN 2287-5026(Print) / ISSN 2288-159X(Online)

논문 2016-53-11-11

스펙트럼 성형기법을 이용한 멀티미디어 콘텐츠의 명료도 향상

(Intelligibility Enhancement of Multimedia Contents Using Spectral Shaping)

지 유 나*, 박 영 철**, 황 영 수***

(Youna Ji, Young-cheol Park, and Young-su Hwang©)

요 약

본 논문에서는 스펙트럼 성형기법을 이용한 멀티미디어 콘텐츠 명료도 향상 알고리즘을 제안한다. 영화, 동영상과 같은 오디오-비주얼 미디어 콘텐츠에서 다이얼로그는 영상의 내용을 이해하기 위한 중요한 요소이다. 하지만 종종 영상내의 효과음, 배경음악 등과 같이 함께 믹싱 된 오디오 성분에 의해 중요한 정보를 지닌 다이얼로그의 명료도가 떨어지는 문제점이 제기되어왔다. 뿐만 아니라 멀티미디어 콘텐츠의 이용 환경이 다양해지면서 청자의 주변 환경 또한 오디오 볼륨에 영향을 미치는 요소가 된다. 본 논문에서는 이러한 문제점을 해결하기 위해 영상의 중요 단서를 담고 있는 사운드트랙의 음성 성분 명료도를 높이고자 한다. 제안된 알고리즘은 먼저 영상의 스테레오 오디오 신호에서 음성 존재 확률(Speech Presence Probability)을 이용한 소프트 마스커를 통해 다이얼로그 성분을 검출한다. 추출된 다이얼로그 성분은 스펙트럼 성형 기법을 적용하여 명료도에 중요한 영향을 미치는 고주파대역의 성분을 증폭시키는 등 음성 신호 스펙트럼의 에너지를 재분배하여 신호의 명료도를 향상시켰다. 마지막으로 크기 정규화 과정을 통해 프로세스 전과 후의 전체 오디오의 파워를 동일하게 유지함으로써 증폭으로 인한 스피커의 오디오 포화(saturation)를 방지하였다. 실험을 통해 본 알고리즘이 동일한 오디오 볼륨에서 영상의 명료도를 향상시킴을 확인 할 수 있었다.

Abstract

In this paper, we propose an intelligibility enhancement algorithm for multimedia contents using spectral shaping. The dialogue signals is essential to understand the plot of audio-visual media contents such as movie and TV. However, the non-dialogue components as like sound effects and background music often degrade the dialogue clarity. To overcome this problem, this paper tries to improves the dialogue clarity of audio soundtracks which contain important cues for the visual scenes. In the proposed method, the dialogue components are first detected by soft masker based on speech presence probability (SPP) which is widely used in speech enhancement field. Then, extracted dialogue signals are applied to the spectral shaping method. It reallocate the spectral-temporal energy of speech to enhanced the intelligibility. The total energy is maintained as unchanged via a loudness normalization process to prevent saturation. The algorithm was evaluated using the modeled and real movie soundtracks and it was shown that the proposed algorithm enhances the dialogue clarity while preserving the total audio power.

Keywords : 음성 명료도 향상, 스펙트럼 성형, 소프트 마스커

Ⅰ. 서 론

최근 태블릿, 노트북과 같은 모바일 기기의 활용도가 점점 넓어짐에 따라 멀티미디어 기기를 사용하는 환경 또한 다양해지고 있다. 이러한 다양한 사용 환경은 멀티미디어 콘텐츠의 오디오 신호의 크기에도 영향을 미치게 된다. 특히 잡음이 존재하는 외부 환경에서 영화와 같은 멀티미디어 콘텐츠를 감상하게 될 경우 잡음은 영화 줄거리의 이해를 돕는 다이얼로그의 명료도를 떨어뜨리게 된다[1]. 뿐만 아니라 조용한 실내에서도 오디오 볼륨은 상황에 따라 제약을 받게 된다. 예를 들어 심야 시간에 TV 또는 영화 시청의 경우 큰 오디오 볼륨은 주변 이웃에 피해를 줄 수 있으며 너무 작은 볼륨은 영상에서 전달하고자하는 대사를 명확히 인지하기 어려

*학생회원 ** 정회원, 연세대학교 컴퓨터정보통신공학부
(Yonsei University)
*** 평생회원, 가톨릭 관동대학교 전자공학과
(Catholic kwandong University)
© Corresponding Author (E-mail : hysoo@cku.ac.kr)
Received ; August 25, 2016 Revised ; October 5, 2016
Accepted ; October 26, 2016

워 몰입을 방해 할 수 있다. 또한 일반적으로 멀티미디어 콘텐츠의 오디오 신호는 내용 이해에 중요한 영향을 미치는 음성 신호 뿐만 아니라 음악, 효과음 그리고 배경음 등 다양한 오디오 신호가 혼합되어 나타나며 이러한 주변 효과음은 오디오 볼륨이 작은 환경에서는 명료도를 떨어뜨린다는 단점을 가지고 있다. 이러한 이유로 최근 멀티미디어 사운드트랙의 음성 명료도를 높이고자 하는 연구들이 진행되어 왔다. 대표적인 방법 중 하나는 패턴 인식 기법을 이용한 기술이다[2]. 이 논문은 입력 사운드 트랙에서 음성을 검출하여 증폭시킴으로써 명료도를 향상시키는 방식이다[2]. 하지만 이 기법은 트레이닝에 필요한 사전 데이터가 있어야 하며 연산량이 많아 모바일 기기에서의 사용은 부적합하다.

반면 최근에는 스테레오 또는 5.1 채널 사운드 트랙으로부터 음성 존재 확률(Speech presence probability, SPP)을 기반으로 계산된 소프트 마스커를 적용하여 음성 성분만을 선택 증폭시키는 알고리즘들이 소개되었다[1, 3]. 이러한 기술들은 먼저 멀티 채널 오디오 신호를 방향성이 있는 패닝 음원들이 포함된 주 성분과 그 외 잔향, 배경음 등이 포함된 주변 신호로 분리한다. 이때 중요 정보인 음성 신호는 주 신호에 포함되어 있으며 주 신호에서 음성을 추출한 후 증폭함으로써 다이얼로그의 명료도를 향상 시킨다.

반면 음성 통신 시스템 분야에서는 청자의 주변 환경을 고려하여 통신 채널을 통해 수신되는 음성 신호의 명료도를 향상시키는 알고리즘이 제안되었다[4~6]. 이 중 대표적인 접근 방법 중 하나는 수신되는 음성 스펙트럼의 에너지를 재분배하여 명료도에 영향을 미치는 대역을 강화 시키는 방식이다. 이는 신호의 전체 라우드니스를 유지하여 과도한 증폭으로 인해 발생할 수 있는 신호의 포화, 고막의 손상을 방지하면서 신호의 명료도를 강화시킬 수 있다는 장점이 있다.

본 논문에서는 음성 명료도 강화 분야에서 대표적으로 사용되는 스펙트럼 성형 기법을 활용한 멀티미디어 콘텐츠 다이얼로그 향상 알고리즘을 제안한다. 제안 된 알고리즘은 입력 신호로부터 소프트 마스커를 이용해 다이얼로그 성분을 분리한 후 스펙트럼 성형 기법을 통해 명료도에 중요한 영향을 미치는 대역으로 에너지를 재분배함으로써 음성 명료도를 향상 시킨다. 이후 객관적, 주관적 실험을 통하여 본 논문의 성능을 확인한다. 본 논문의 구성은 다음과 같다. 먼저 2장에서 본 논문에서 제안하는 다이얼로그 명료도 향상 알고리즘을 소개하고 3장에서 제안된 알고리즘을 검증하기 위한 시뮬레이션 결과를 보이며 마지막 4장에서 결론으로 마무리한다.

Ⅱ. 다이얼로그 명료도 향상 알고리즘

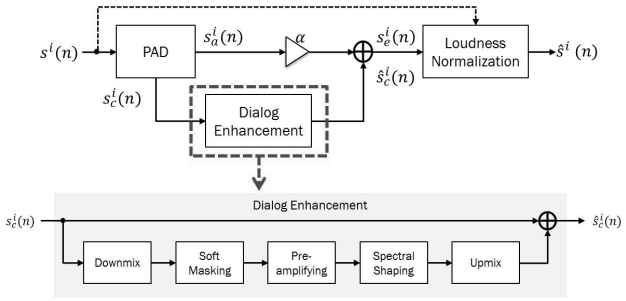

그림 1. 다이얼로그 명료도 향상 알고리즘 블록 다이어그램
Fig. 1. Block diagram of the proposed dialogue intelligibility enhancement algorithm.

그림 1은 본 논문에서 제안하는 다이얼로그 명료도 향상 알고리즘의 블록 다이어그램을 보이고 있다. 먼저 스테레오 오디오 소스 신호는 주성분분석(Principal Component Analysis, PCA)기반의 Primary Ambient Decomposition (PAD)을 통해 주변 신호와 주 신호로 각각 분리되게 된다[7]. 그 후 소프트 마스커를 통해 주 신호에서 음성 성분을 검출 및 사전 증폭시킨 후 스펙트럼 성형 과정을 통해 밴드 간 에너지를 재분배한다. 향상된 다이얼로그 신호는 프로세스 전의 주 신호와 더해져 향상된 주 신호를 만들어 내고 이는 마지막으로 분리되었던 주변 신호와 더해진다. 최종적으로 오디오 신호는 정규화 과정을 통해 프로세스 전과 전체적으로 동일한 크기를 가지도록 출력된다.

1. PAD

스테레오 시스템에서 i-채널의 오디오 신호는 다음과 같이 모델링 될 수 있다.

$$s^i(n) = s_a^i(n) + s_c^i(n), i = L, R \quad (1)$$

이때 $s_a^i(n)$와 $s_c^i(n)$는 각각 i-채널의 주변 신호와 주 신호를 나타낸다. 본 논문에서는 먼저 PCA 기반의 PAD 알고리즘을 이용해 스테레오 신호로부터 두 성분을 분리한다[7]. PCA 기법은 주 성분 신호가 채널 간의 상관도가 높은 반면 주변 성분은 서로 다른 채널에 독립적으로 존재하여 매우 낮은 상관도를 갖는 특성을 이용하여 두 성분을 분리하는 방식이며 최근 채널 간의 고유치값의 비를 이용한 향상된 PCA 기반의 PAD 알고리즘이 제안되었다.

따라서 채널 간 상관도가 높은 패닝 음원인 다이얼로그 성분은 주 신호로 분리 되어 나온다. 분리된 주 신호는 먼저 다운믹스 과정을 통해 모노 신호로 변환된다.

$$s_c'(n) = \left(s_c^L(n) + s_c^R(n)\right)/\sqrt{2} \quad (2)$$

2. 다이얼로그 검출 및 사전 증폭

다운믹스 과정까지 거친 모노 주 신호 $s_c'(n)$에는 다이얼로그뿐만 아니라 특정한 방향성을 가지는 신호와 같이 채널 간 상관도가 높은 성분들이 모두 포함되어있다. 따라서 추출된 주 신호에서 다이얼로그 성분을 검출해 선택적으로 증폭시키고자 한다. 이를 위해 음질 개선 분야에서 대표적으로 이용되는 음성 존재 확률 기반의 소프트 마스커를 계산한다[8]. SPP는 0과 1사이의 정규화 값을 가지며 1에 가까울수록 음성이 존재 할 확률이 높음을 뜻한다. 본 논문에서는 소프트 마스커를 얻기 위해 고정된 사전 정보들을 이용해 통계적으로 음성 존재 확률을 계산하는 단 채널 알고리즘을 이용하였다[8]. 다운 믹스 후 얻어진 단 채널 신호의 STFT (Short-Time Fourier Transform)에 의한 주파수 도메인 신호는 다음과 같이 표현 할 수 있다.

$$S_c'(k,l) = D(k,l) + N(k,l) \quad (3)$$

이때 $D(k,l)$와 $N(k,l)$는 각각 다이얼로그 성분과 주 신호에서 다이얼로그를 제외한 그 외 오디오 신호를 뜻한다. k,l은 각각 주파수, 프레임 인덱스이다. 다이얼로그 성분 외의 신호를 잡음으로 간주할 때 주어진 주 신호에 음성이 존재할 확률은 다음과 같이 계산 될 수 있다[8].

$$p_s(k,l) = \left(1 + \frac{P(H_0)}{1-P(H_0)}(1+\xi_{H_1})e^{-\frac{|S_c'(k,l)|^2}{\sigma_N^2(k,l)}\frac{\xi_{H_1}}{1+\xi_{H_1}}}\right)^{-1} \quad (4)$$

위 수식에서 $P(H_0)$ 그리고 ξ_{H_1}는 고정된 값으로 사전 음성 부재(a priori speech absence probability) 그리고 사전 신호대음성비(a priori SNR)를 나타낸다. $\sigma_N^2(k,l) = E[N(k,l)\cdot N^*(k,l)]$는 추정된 잡음의 PSD로 음성이 존재하지 않는 구간에서 업데이트 된다[8]. 구해진 음성 존재 확률 값은 연속된 프레임에서의 음성의 강한 상관성을 고려하여 시간 축에서 1차 재귀 평균을 적용해 최종 소프트 마스커를 얻는다.

$$p_s'(k,l) = \alpha_p p_s'(k,l-1) + (1-\alpha_p)p_s(k,l) \quad (5)$$

이때 α_p는 스무딩 파라미터로 0과 1사이의 값을 갖는다. 본 논문에서는 $\alpha_p = 0.93$으로 적용되었다.
검출된 다이얼로그 성분은 먼저 사전 증폭 상수를 이용해 일차적인 증폭이 이루어졌다.

$$\tilde{S}_c(k,l) = \left(S_c'(k,l)\cdot p_s'(k,l)\right)\cdot g \quad (6)$$

위의 수식에서 g는 사전 증폭 상수로 실험적으로 결정될 수 있으며 본 논문에서는 3dB를 사전 증폭 값으로 설정하였다.

3. 다이얼로그 명료도 향상을 위한 스펙트럼 성형

사전 증폭된 다이얼로그 신호는 스펙트럼 성형 기법의 입력으로 이용된다. 스펙트럼 성형은 주파수 도메인에서 다이얼로그 스펙트럼 중 명료도에 중요한 영향을 미치는 대역으로 에너지를 재분배하여 음성의 명료도를 높이는 기법으로 통신 환경에서 수신단의 음성 강화를 위해 제안된 기술 중 하나이다[6]. 일반적으로 고주파대역이나 무성음 구간 등의 음성을 강화시켰을 때 명료도가 향상되었다는 연구 결과들이 있다[5~6]. 본 논문에서는 기존에 제안된 스펙트럼 성형 과정을 추출된 다이얼로그에 적용하여 주변 배경음과 효과음이 존재에서도 명료도를 높여 콘텐츠의 이해를 돕고자 하였다.

스펙트럼 성형은 총 세 단계로 이루어진다[6]. 먼저 입력으로 들어온 사전 증폭 된 주 신호에서 유성음이 존재하는 확률을 계산한다. 그 다음 적응 스펙트럼 성형 단계에서 앞서 계산한 유성음이 존재할 확률 값을 기반으로 음성의 포만트 성분을 강화시킨다. 마지막으로 적응 그리고 고정 프리엠퍼시스 필터를 이용하여 상대적으로 명료도에 중요한 영향을 미치는 고주파대역의 에너지부분을 증폭시킨다.

$$S_c^{ss}(k,l) = \tilde{S}_c(k,l)\cdot H_s(k,l)\cdot H_p(k,l)\cdot H_r(k,l) \quad (7)$$

위의 수식에서 $H_s(k,l), H_p(k,l)$ 그리고 $H_r(k,l)$는 각각 적응 스펙트럼 성형, 적응 그리고 고정 프리엠퍼시스 필터를 뜻한다. 먼저 $H_s(k,l)$는 현재 음성 신호가 유성음일 확률을 기반으로 음성의 포만트 성분을 강조시키는 역할을 한다. $H_p(k,l)$는 음성의 명료도에 중요한 영향을 미치는 1100 Hz 이상의 대역에 6dB/octave를 증가시키며 이 또한 유성음 구간에 대해서만 적용시킨다. 마지막으로 $H_r(k,l)$은 음성의 1000-4000 Hz 대역 신호의 에너지를 강화시키는 동시에 500 Hz 이하의 대역에 대해서는 6dB/octave 만큼을 에너지를 감소시키도록 한다.

필터 설계의 자세한 내용은 기존 논문에[6] 기술되어 있으며 본 논문에서는 기존 논문의 파라미터를 그대로 적용하였다.

4. 최종 스테레오 오디오 신호 크기 정규화

스펙트럼 성형을 통해 다이얼로그의 에너지가 재분배 된 주 신호는 ISTFT(Inverse STFT)를 이용해 시간 도메인으로 옮겨 온 뒤 업믹스 과정을 거쳐 프로세스 전의 주 신호와 더해진다.

$$\hat{s}_c^i(n) = s_c^{ss}(n) + s_c^i(n), i = L,R \tag{8}$$

위의 수식에서 $s_c^{ss}(n) = ISTFT\left[S_c^{ss}(k,l)\right]$는 스펙트럼 성형이 적용된 다이얼로그 신호의 시간 도메인 표현이다. 에너지가 재분배된 다이얼로그 신호는 프로세스를 거치기 전의 주 신호와 더해지게 된다. 이 과정을 통해 소프트 마스커의 에러로 인해 발생할 수 있는 신호의 왜곡을 완화시키는 동시에 증폭의 효과를 얻을 수 있다.
최종 출력 신호를 생성하기 위해 프로세스 된 주 신호는 앞서 분리해냈던 주변 신호와 더해진다. 이때 주변 신호에 0과 1사이의 상수 이득 α를 적용해 그 크기를 조절하였다.

$$s_c^i(n) = \alpha s_a^i(n) + \hat{s}_c^i(n), i = L,R \tag{9}$$

마지막으로 최종 출력 신호와 입력 신호의 크기를 맞추기 위해 정규화 과정을 거친다. 이는 신호처리과정에서 오디오 신호의 볼륨이 과하게 증폭되는 것을 방지하는 것으로 이를 통해 입력 신호와 최종 출력 신호의 전체 오디오 볼륨은 같지만 내부적으로 다이얼로그만 향상된 효과를 얻을 수 있다.

$$\hat{s}^i(n) = s_c^i(n) \cdot \sqrt{\sum_n |s^i(n)|^2 / \sum_n |s_c^i(n)|^2} \tag{10}$$

Ⅲ. 실 험

본 논문에서 제안한 알고리즘의 성능을 검증하기 위해 다음과 같은 컴퓨터 시뮬레이션이 진행되었다. 실험에는 서로 다른 상황의 총 7개의 사운드 트랙이 사용되었다. 그 중 3개의 신호(Mov1M-Mov3M)는 실험의 성능 평가를 위해 모델링된 스테레오 신호로 기존의 영화 사운드트랙에 TIMIT 데이터베이스의 음성 신호를 결합하여 생성되었다. 또한 실제 영화 사운드 트랙에서의 성능을 알아보기 위해 총 4편의 영화에서(Mov4S - Mov7S) 서로 다른 상황의 입력 신호를 추출하였다. 각 입력 신호의 목록은 아래 표 1과 같으며 샘플링 주파수는 44.1 kHz 이다.

표 1. 실험에 사용된 영화 클립
Table1. Input signal for simulation.

		언 어	상 황
1	Mov1M	영어	배경 음악
2	Mov2M	영어	헬리콥터 소리
3	Mov3M	영어	총격 소리
4	Mov4S	영어	배경음악
5	Mov5S	영어	빗속에서 남녀의 대화
6	Mov6S	한국어	전쟁 신, 남성 음성
7	Mov7S	한국어	빗속에서 다수의 남녀 대화

먼저 다이얼로그 검출 블록의 성능을 확인하기 위한 실험이 진행되었으며 그 결과는 그림 2, 3과 같다. 그림 2는 모델링 신호 1에 대하여 (a) 분리된 주 신호와 (b) 정답

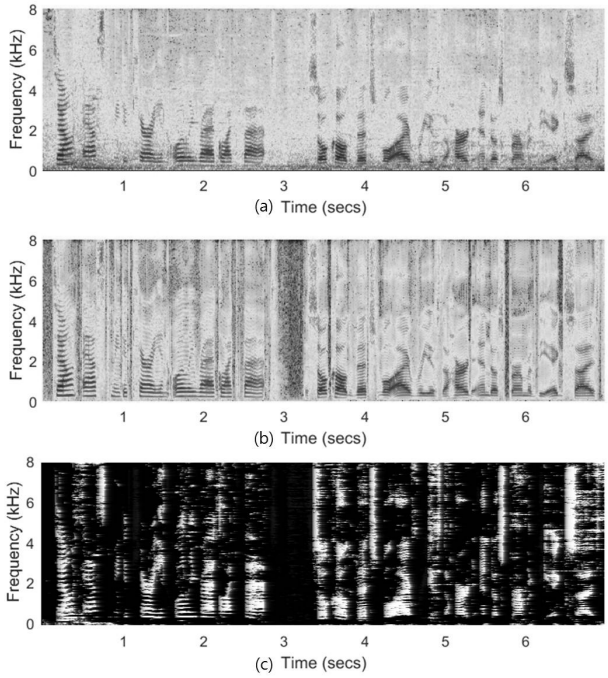

그림 2. 모델링 신호에 대한 다이얼로그 검출결과. (a) Mov1M 신호의 주 신호, (b) 음성 신호의 스펙트로그램 그리고 (c) 계산된 소프트 마스커
Fig. 2. The dialogue detection results for modeling signal(Mov1M). The spectrogram of (a) primary, (b) speech signal and (c) generated soft masker.

음성 신호 스펙트로그램 그리고 이를 기반으로 계산된 (c) 소프트 마스커를 보이고 있다. 소프트 마스커의 하얀색 부분이 음성이 존재하는 영역으로 판별된 지역이다.

그림을 통해 계산된 소프트 마스커가 분리된 주 신호로부터 음성 신호 성분을 잘 검출하고 있음을 확인 할 수 있다. 실제 영화 클립 신호를 이용한 실험인 그림 3을 통해서도 동일한 결과를 확인 할 수 있다.

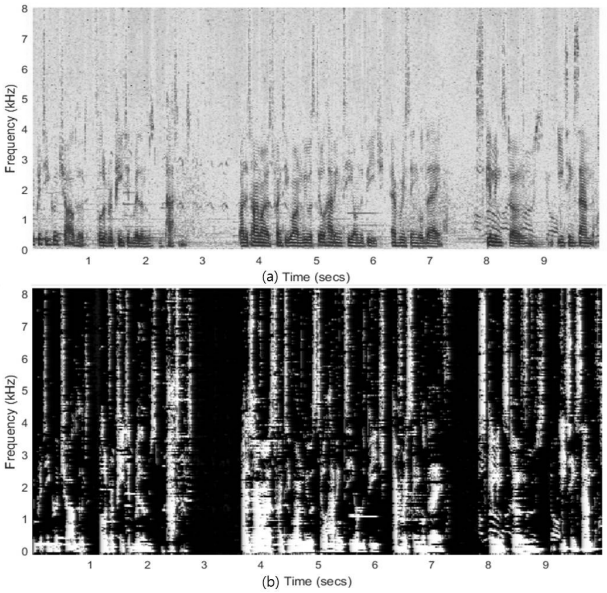

그림 3. 실제 영화 신호에 대한 다이얼로그 검출결과. (a) Mov4S 신호의 주 신호의 스펙트로그램 그리고 (b) 계산된 소프트 마스커

Fig. 3. The dialogue detection results for Movie soundtracks (Mov4S). The spectrogram of (a) primary signal and (b) generated soft masker.

다음 그림 4는 입력 스테레오 신호와 제안 알고리즘을 거친 최종 출력 신호의 PSD(Power spectral density) 결과를 보이고 있다. 비교를 위해 제안된 알고리즘에서 스펙트럼 성형 과정을 제외하고 다이얼로그 성분만 증폭한 결과가 함께 게재되었다. 그림 4의 (a)는 각각 주 성분 신호 (수식 (3)), 다이얼로그가 선택 증폭 된 신호 (수식 (4)) 그리고 스펙트럼 성형 기법이 적용된 신호 (수식 (7))의 PSD 결과를 보이고 있다. 그림을 통해 스펙트럼 성형 기법을 도입한 경우 음성 명료도에 결정적인 영향을 미치는 고주파대역에 에너지가 증폭된 것을 확인 할 수 있다. 그림 4. (b)는 (a)의 출력 결과를 기반으로 만들어진 최종 출력 신호의 PSD 결과이다. 다이얼로그 성분을 상수 증폭 시킨 경우 크기 정규화까지 적용한 마지막 출력 신호에서 명료도에 상대적으로 영향이 적은 저주파대역의 하모닉 성분에 에너지가 집중되는데 비해 스펙트럼 성형을 적용한 최종 출력 신호는 고주파영역에서 충분한 강화를 보이는 것을 확인 할 수 있다.

알고리즘의 정량적 평가를 위해 객관적 그리고 주관적 음질평가가 시행되었다. 먼저 객관적 음질평가로는 음성의 명료도를 나타내는 SII(Speech intelligibility index)가 측정되었다[9]. 성능 비교를 위해 프로세스 전의 스테레오 입력 신호와 검출된 다이얼로그 신호를 상수 증폭시킨 경우를 함께 측정하였다. SII 측정을 위해서는 정답 음성 신호 정보가 필요하기 때문에 모델링 된 3개의 신호 Mov1M – Mov3M에 대해서만 실험이 진행되었으며 그 결과는 그림 5에서 보이고 있다. 스펙트럼 성형까지 적용한 출력 신호가 3개의 입력 신호에서 모두 프로세스 전, 그리고 다이얼로그만 증폭 시킨 오디오 신호에 비해 더 높은 SII 값을 가짐을 확인 할 수 있다.

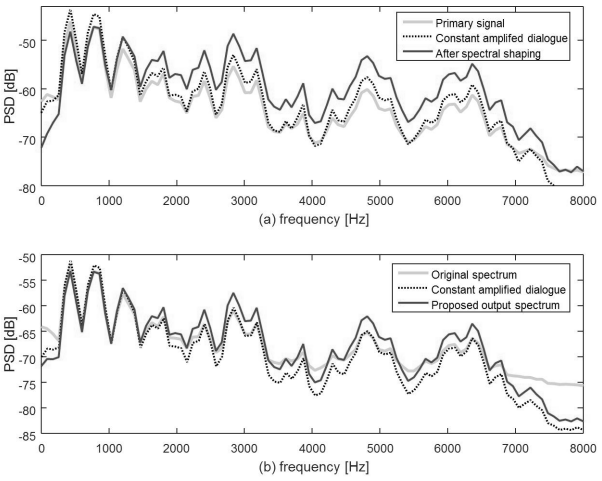

그림 4. 프로세스 전, 후의 PSD 결과 비교. (a) 분리 된 주 신호의 상수 증폭, 스펙트럼 성형 결과 PSD, (b) 최종 출력 신호 PSD

Fig. 4. The PSD results of (a) primary and (b) final output signals.

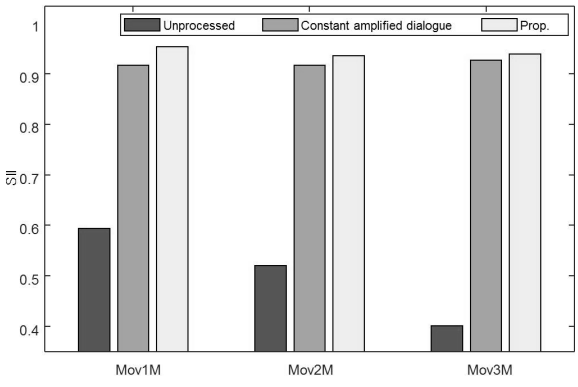

그림 5. SII 결과

Fig. 5. SII results.

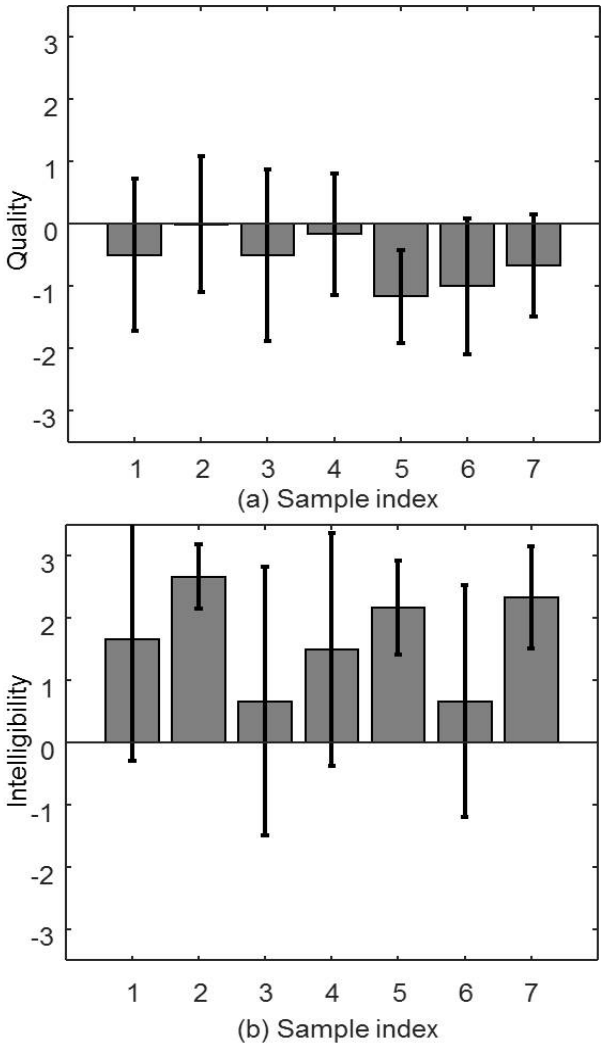

그림 6. 알고리즘 적용 후의 (a) 품질 그리고 (b) 명료도 주관적 평가 실험 결과
Fig. 6. The subjective test results of (a) quality and (b) intelligibility.

다음은 주관적 음질평가를 위해 ITU-T P.800의 주관적인 음질 측정 방법인 Comparison Category Rating (CCR) 실험이 진행되었다[10]. 이는 선호도를 평가하는 방법으로써, 비교 대상이 되는 두 신호 중 어느 신호가 더 음질이 좋은지, 더 명료하게 잘 들리는지를 판단하여 -3점부터 +3점까지 점수를 부여하는 방법이다. 입력 스테레오 신호와 제안 알고리즘을 적용한 향상된 출력 신호가 품질과 명료도 두 가지 측면에서 비교 측정되었다. 처리 후 신호의 품질과 명료도가 좋을수록 3에 가까운 값을 갖는다. 총 7명의 한국인 청자가 실험에 참여하였으며 7명 중 2명은 여성, 5명은 남성으로 구성되었다. 실험에는 청취 실험에서 일반적으로 이용되는 젠하이저 HD 280 pro 헤드폰이 사용되었다. 그림 6은 (a)품질과 (b) 명료도 측정 평균값과 표준편차를 보이고 있다. 그림 6의 샘플 인덱스는 표 1과 각각 대응되는 신호이다. 실험결과 제안된 알고리즘을 적용한 경우 전 음원에 대해 음성의 명료도가 향상된 것을 확인 할 수 있다. 다만 품질 측면에서 알고리즘 적용 전에 비하여 약간의 성능 열화를 느끼는 것을 보인다. 그렇지만 전반적으로 품질의 성능 열화 정도보다 명료도의 향상 정도가 더 큰 것을 확인 할 수 있다.

Ⅳ. 결 론

본 논문에서는 멀티미디어 콘텐츠를 위한 다이얼로그 명료도 향상 알고리즘을 제안하였다. 제안 알고리즘은 오디오-비주얼 콘텐츠의 스테레오 신호로부터 소프트 마스크를 이용해 다이얼로그 성분을 검출한다. 검출된 다이얼로그는 스펙트럼 성형 기법을 통해 고주파대역을 증폭 시키는 에너지 재분배 방식을 통해 명료도가 향상된 신호를 얻게 된다. 최종 출력 신호는 입력 스테레오 신호 기반의 크기 정규화 과정을 거쳐 얻게 된다. 다양한 실험 결과를 통해 제안 알고리즘 출력 신호가 프로세스 전에 비해 음성의 명료도를 강화시키는 것을 확인 할 수 있었다.

REFERENCES

[1] K. Lopatka, K. Bartosz, and C. Andrzej, “Novel 5.1 downmix algorithm with improved dialogue” Audio Engineering Society Convention 134. Audio Engineering Society, 2013.
[2] C. Uhle, H. Oliver, and W. Jan, “Speech enhancement of movie sound,” Audio Engineering Society Convention 125. Audio Engineering Society, 2008.
[3] K. Lopatka, C. Andrzej, and K. Bozena. “Improving listeners′ experience for movie playback through enhancing dialogue clarity in soundtracks.” Digital Signal Processing Vol. 48, pp. 40-49, 2016.
[4] J. H. Choi, and J. H. Chang, “Robust speech reinforcement based on gain-modification incorporating speech absence probability.” Journal of the Institute of Electronics Engineers of Korea SP, Vol. 47, no.1, pp. 175-182, 2010.
[5] B. Sauert, and P. Vary, “Recursive closed-form optimization of spectral audio power allocation for near end listening enhancement.” ITG-Fachbericht-Sprachkommunikation 2010 (2010).
[6] T.C. Zorila, K. Varvara, and S. Yannis. “Speech-in-noise intelligibility improvement based on

spectral shaping and dynamic range compression." Thirteenth Annual Conference of the International Speech Communication Association. 2012.
[7] Y. H. Baek, et al. "Efficient primary-ambient decomposition algorithm for audio upmix." Audio Engineering Society Convention 133. Audio Engineering Society, 2012.
[8] T. Gerkmann, B. Colin, and M. Rainer. "Improved a posteriori speech presence probability estimation based on a likelihood ratio with fixed priors." Audio, Speech, and Language Processing, IEEE Trans. on Vol. 16 no.5, pp. 910-919, 2008.
[9] ANSI, "Methods for calculation of the speech intelligibility index," S3.5-1997, (American National Standards Institute, NewYork), 1997.
[10] ITU-T P.800, Methods for Subjective Determination of Transmission Quality, Aug. 1996.

저 자 소 개

지 유 나(학생회원)
2011년 연세대학교 컴퓨터정보통신공학부 학사 졸업.
2011년～현재 연세대학교 전산학과 (석박통합 과정)
<주관심분야: 디지털 신호처리, 음질 개선, 음성 신호처리>

황 영 수(정회원)
1982년 연세대학교 전자공학과(학사)
1984년 연세대학교 전자공학과(석사)
1990년 연세대학교 전자공학과(박사)
1989년～현재 가톨릭 관동대학교 전자공학과 교수.
<주관심분야: 통신 신호처리, 음성 신호처리, 음향학>

박 영 철(평생회원)
1986년 연세대학교 전기전자공학과 (학사)
1988년 연세대학교 전기전자공학과 (석사)
1993년 연세대학교 전기전자공학과 (박사)
2002년～현재 연세대학교 컴퓨터정보통신공학부 교수
<주관심분야: 디지털 신호처리, 오디오 신호처리, 음성 신호처리, 적응 신호처리>