은닉 마르코프 모델을 이용한 스테레오에서 서라운드 오디오 신호로의 변환

정석희, 전찬준, 김홍국
광주과학기술원
{jeongsh, cjchun, hongkook}@gist.ac.kr

Conversion of Stereo to Surround Audio Signal Using Hidden Markov Model

Seok Hee Jeong, Chan Jun Chun, Hong Kook Kim
Gwangju Institute of Science and Technology (GIST)

요 약

본 논문에서는 hidden Markov model (HMM) 기반의 스테레오 신호로부터 서라운드 오디오 신호를 생성하는 기법을 제안한다. 먼저 5.1 채널 오디오 훈련 데이터베이스로부터 MDCT 영역에서 전방/서라운드 채널의 서브밴드 에너지를 프레임 단위로 계산하고, 이를 특징 벡터로 하여 좌측과 우측 채널 두 개의 HMM 이 구성된다. 다음으로, 입력된 스테레오 신호에 대해 HMM decoding 을 통해 서라운드 채널의 MDCT 영역의 서브밴드 에너지가 예측된다. 이 예측된 서브밴드 에너지로부터 역 MDCT 를 통해 서라운드 오디오 신호가 생성된다. 제안된 방법의 성능평가를 위해 MUSHRA 청취 실험을 수행한 결과, 제안된 HMM 기반의 방식으로 생성된 서라운드 오디오 신호가 기존의 패시브 서라운드 디코딩 기반으로 생성된 서라운드 신호에 비해 높은 선호도를 보였다.

1. 서론

최근 들어 실감나는 오디오 효과를 사용자에게 제공할 수 있는 5.1 채널 오디오 시스템에 대한 수요가 급증하고 있다. 하지만, 5.1 채널 오디오 콘텐츠를 제작하기 위해서는 다수의 마이크로폰을 활용해야 하는 등 비용 및 공간적 제약이 존재하게 된다. 이에 따른 대안으로 스테레오에서 5.1 채널로의 업믹싱 기법에 대한 연구가 활발히 진행되고 있다. 이러한 5.1 채널 업믹싱 기법에서 중요한 부분 중에 하나는 서라운드 채널을 생성하는 것이다. 대부분의 업믹싱 기법들은 스테레오 오디오 신호를 신호상관도가 높은 부분과 낮은 부분으로 분리하여 서라운드 채널을 생성하는 신호상관기반의 업믹싱 기법을 사용한다[1]. 하지만 5.1 채널 원음의 신호상관관계는 시간에 따라 변하기 때문에 신호상관기반 업믹싱 기법은 5.1 채널 원음과 유사한 서라운드 채널 오디오 신호를 생성하기 어렵다.

따라서 본 논문에서는 이러한 단점을 극복하기 위하여 채널간의 에너지 특성과 오디오 프레임간의 에너지 변화 추이를 고려한 모델 기반의 서라운드 채널 오디오 생성 기법을 제안한다. 우선, modified discrete cosine transform (MDCT) 영역에서 5.1 채널 오디오 신호의 서브밴드 에너지를 특징 벡터로 하여 HMM 을 학습하고, 이를 기반으로 주어진 스테레오 신호에 대한 서라운드 채널의 MDCT 영역에서의 서브밴드 에너지를 예측한다.

본 논문의 구성은 다음과 같다. 2 절에서는 HMM 기반 서라운드 오디오 신호 생성 기법에 대해서 기술한다. 3 절에서는 제안된 서라운드 오디오 신호 생성 기법의 성능을 평가한다. 마지막으로 4 절에서 결론을 맺는다.

2. HMM 기반 서라운드 오디오 신호 생성 기법

먼저, HMM 훈련을 위해 전방/서라운드 채널 신호로 구성된 오디오 신호에 대해 프레임 단위로 나눈 뒤, MDCT 를 적용하여 MDCT 영역에서의 15 개의 서브밴드 에너지를 계산한다. 이때 전방/서라운드 채널 신호의 서브밴드 에너지 시퀀스는 아래와 같다.

$$E_{F,m,t} = [E_{F,m,t}(0), E_{F,m,t}(1), \cdots, E_{F,m,t}(14)] \quad (1)$$

$$E_{R,m,t} = [E_{R,m,t}(0), E_{R,m,t}(1), \cdots, E_{R,m,t}(14)] \quad (2)$$

여기서 $E_{F,m,t}$ 와 $E_{R,m,t}$ 는 각각 전방과 서라운드 채널 신호의 서브밴드 에너지를 의미하고, t 와 m 은 각각 프레임 인덱스와 채널(좌측 (m=L) 또는 우측 (m=R))을 의미한다. (1)와 (2)로 정의된 서브밴드 에너지 벡터 시퀀스를 특징 벡터로 활용하여 상태 전이 확률, 초기 확률과 같은 HMM 파라메타를 기대치최대화(Expectation-Maximization, EM) 알고리즘을 통해 추정한다[2]. 본 논문에서는 3 분 길이의 48 kHz 표본화율을 갖는 오디오 파일 50 개를 훈련 데이터베이스로 사용하였으며, 1280-point MDCT 를 사용하였다.

그림 1 은 제안된 HMM 기반 서라운드 오디오 신호 생성 기법에 대한 블록도를 보여준다. 그림에서 보는 바와 같이 입력된 신호는 MDCT 를 통해 주파수 영역으로 변환되고, 서브밴드 에너지가 계산된다. 또한, 서라운드 신호로의 합성을 위해 전방 채널 신호는 해당 서브밴드 에너지에 의해 $\overline{X}_{F,m,t}(k)$ 로 정규화된다. 다음으로, 훈련 과정을 통해 구성된 HMM 으로부터 아래 식과 같이 서라운드 오디오 신호의 서브밴드 에너지를 예측할 수 있다[2].

$$\hat{E}_{R,m,t} = \sum_{j=1}^{N_s} P(q_t = s_j \mid E_{F,m,1:t}, \lambda) \sum_{m=1}^{N_m} P(m \mid E_{F,m,t}, \Theta_{jm}, q_t = s_j) \mu_{jm}^{E_R} \quad (3)$$

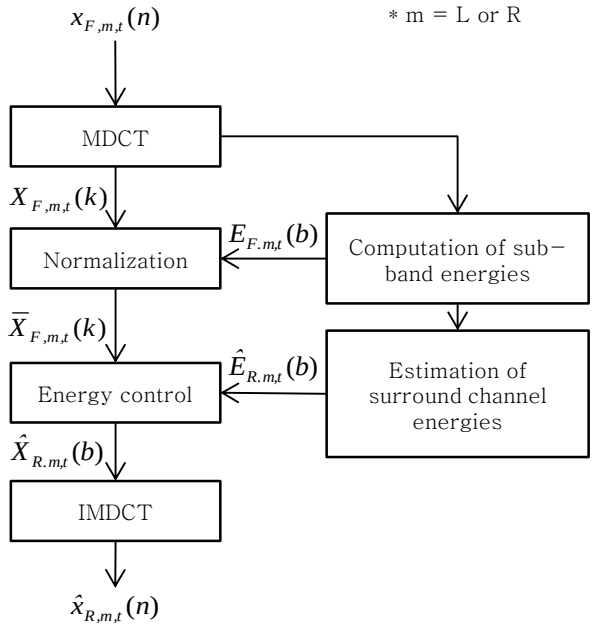

그림 1. 제안된 HMM 기반의 서라운드 오디오 신호 생성 기법의 블록도

여기서 N_s 와 N_m 은 각각 HMM 의 상태수와 각 상태에서의 혼합 가우시안 수를 의미하며, $E_{F,m,1:t}$ 은 $E_{F,m,t}$ 의 첫 번째 프레임에서 t 번째 프레임까지 구성되는 시퀀스이다. 또한, μ_{jm}^{Es} 은 j 번째 상태에서 m 번째 혼합 가우시안 모델에서의 서라운드 채널 신호의 평균 벡터를 나타낸다.

다음으로, 예측된 서라운드 신호의 서브밴드 에너지 시퀀스, $\hat{E}_{R,m,t}$ 와 정규화된 전방 채널 신호, $\overline{X}_{F,m,t}(k)$ 의 합성 과정을 통해 예측된 서라운드 신호의 MDCT 계수, $\hat{X}_{R,m,t}(k)$ 를 생성할 수 있다[2]. 마지막으로 역 MDCT 를 통해 시간 영역의 서라운드 오디오 신호를 생성할 수 있다. 본 논문에서는 혼합 가우시안 수와 상태수를 각각 128 와 3 으로 설정하였다.

3. 성능 평가

제안된 서라운드 오디오 생성 기법에 대한 성능평가를 위해서 MUSHRA 테스트를 실시하였다[3]. 이를 위해 클래식, 오케스트라, 팝 뮤직, 락 뮤직, 발라드 등의 5 가지 장르에 해당하는 오디오 파일을 사용하였다. 총 7 명의 실험자가 참여하였으며 5.1 채널 스피커의 배치는 ITU-R BS.775-1 에 정의된 배치를 준수하였다. 실험에 사용된 5.1 채널 오디오의 서라운드 신호는 제안된 기법을 통해 생성된 서라운드 신호로 대체하여 실험을 진행하였다. 성능 비교를 위해 패시브 서라운드 디코딩(Passive Surround Decoding, PSD) 기법[1] 또한 같은 방식으로 대체되었다. MUSHRA 청취 실험에 사용된 비교 음원은 1) hidden reference, 2) 14 kHz 의 차폐주파수를 갖는 저역필터로 처리된 앵커 신호, 3) 7 kHz 의 차폐주파수를 갖는 저역필터로 처리된 앵커신호, 4) 패시브 서라운드 디코딩 기법으로 처리된 음원, 그리고 5) 제안된 기법으로 처리된 음원이다.

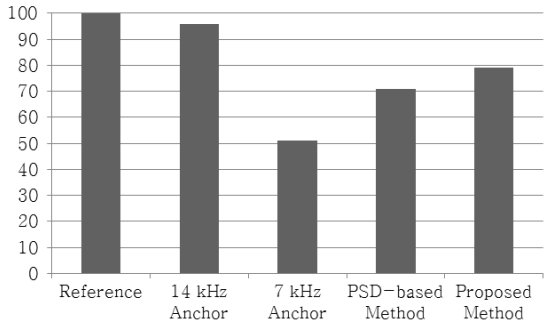

그림 2. MUSHRA 평가 결과 비교.

그림 2 는 MUSHRA 테스트 결과를 나타낸 것이다. 그림에서 보는 바와 같이 제안한 방식은 평균 79 점으로 PSD 기반의 업믹싱 기법에 비해 높은 점수를 얻은 것을 확인할 수 있었다.

4. 결론

본 논문에서는 스테레오 오디오 신호를 활용한 HMM 기반의 서라운드 오디오 신호 생성 기법을 제안하였다. 제안된 서라운드 오디오 신호 생성 기법의 성능평가를 위해 주관적 음질 평가를 실시한 결과, 기존의 신호상관기반 업믹싱 알고리즘인 패시브 서라운드 디코딩 기법보다 제안된 서라운드 오디오 생성 기법이 높은 선호도를 보임을 확인할 수 있었다.

감사의 글

본 연구는 미래창조과학부 및 정보통신산업진흥원의 대학 IT 연구센터육성 지원사업(NIPA-2014-H0301-14-1019) 및 2014 년도 미래창조과학부의 재원으로 한국연구재단의 지원(No. 2012-010636)을 받아 수행된 연구임.

참고문헌

[1] C. J. Chun, Y. H. Lee, Y. G. Kim, H. K. Kim, and C. S. Cho, "A real-time audio upmixing method from stereo to 7.1-channel audio," *Communications in Computer and Information Science*, vol. 120, pp. 162-171, Dec. 2010.

[2] N. I. Park, K. M. Jeon, S. H. Choi, and H. K. Kim, "Artificial stereo extension based on hidden Markov model for the incorporation of non-stationary energy trajectory," *Audio Engineering Society 135th Convention*, New York, NY, Preprint 8980, Oct. 2013.

[3] ITU-R BS 1534, *Method for Subjective Assessment of Intermediate Quality Level of Coding Systems*, June 2001.