한국음향학회지 제36권 제4호 pp. 261~266 (2017)
The Journal of the Acoustical Society of Korea Vol.36, No.4 (2017)
https://doi.org/10.7776/ASK.2017.36.4.261

pISSN : 1225-4428
eISSN : 2287-3775

커널 모델과 장단기 기억 신경망을 결합한 보컬 및 비보컬 분리

Vocal and nonvocal separation using combination of kernel model and long-short term memory networks

조혜승,[1] 김형국[1†]

(Hye-Seung Cho[1] and Hyoung-Gook Kim[1†])

[1]광운대학교 전파공학과

(Received February 20, 2017; revised March 24, 2017; accepted July 31, 2017)

초 록: 본 논문에서는 커널 모델과 장단기 기억(Long-Short Term Memory, LSTM) 신경망을 결합한 보컬 및 비보컬 분리 방식을 제안한다. 기존의 음원 분리 방식은 비보컬 음원만 있는 구간에서 음원을 오추정하여 불필요한 비보컬 음원을 출력하는 한계가 있다. 따라서 본 논문에서는 커널 모델 기반의 보컬음 분리 방식에 LSTM 신경망 기반의 보컬 구간 분류 방식을 결합하여 보컬 음원의 오추정 문제를 개선하고 분리 성능을 향상시키고자 하였다. 또한 본 논문에서는 방식간의 결합 구조에 따라 병렬 결합형 분리 알고리즘과 직렬 결합형 분리 알고리즘을 제안하였으며, 실험을 통해 제안하는 방식들이 기존의 방식에 비해 더욱 향상된 분리 성능을 보이는 것을 확인할 수 있었다.
핵심용어: 보컬 및 비보컬 분리, 커널 모델, 장단기 기억 신경망, 심층 신경망

ABSTRACT: In this paper, we propose a vocal and nonvocal separation method which uses a combination of kernel model and LSTM (Long-Short Term Memory) networks. Conventional vocal and nonvocal separation methods estimate the vocal component even in sections where only non-vocal components exist. This causes a problem of the source estimation error. Therefore we combine the existing kernel based separation method with the vocal/nonvocal classification based on LSTM networks in order to overcome the limitation of the existing separation methods. We propose a parallel combined separation algorithm and series combined separation algorithm as combination structures. The experimental results verify that the proposed method achieves better separation performance than the conventional approaches.
Keywords: Vocal and nonvocal separation, Kernel model, LSTM (Long-Short Term Memory), DNN (Deep Neural Newtorks)

PACS numbers: 43.60.Uv, 43.75.Rs

I. 서 론

음악 신호에 담긴 보컬 신호는 가수의 정보, 가사 및 음악의 감성 등 해당하는 음악에 대한 유용한 정보를 포함한다. 따라서 혼합 음악 신호에서의 보컬 및 비보컬의 분리는 음원 기반의 다양한 응용 분야에서 중요하게 다뤄지는 기술 중 하나이다.[1]

보컬 음원을 동반하는 일반적인 음악 신호에서는 비보컬 음원으로만 이루어진 간주 구간이나 보컬이 묵음으로 진행되는 구간 등이 포함되어있다. 따라서 음악 신호에서 보컬 음원을 정확히 추정하기 위해서는 보컬과 비보컬이 혼합된 구간과 비보컬만 존재하는 구간에서 각각 발생하는 문제를 해결해야 한다. 즉, 보컬과 비보컬이 혼합된 구간에서는 여러 음원이 겹쳐있는 가운데 보컬 음원만 정확히 추정해야

†**Corresponding author:** Hyoung-Gook Kim (hkim@kw.ac.kr)
Department of Radio Sciences and Engineering, Kwangwoon University, 20 Gwangun-Ro, Nowon-Gu, Seoul 01897, Republic of Korea
(Tel: 82-2-940-5574, Fax: 82-2-913-5006)

하며, 비보컬만 존재하는 구간에서는 되도록 오추정 성분을 줄여야 한다. 그러나 대부분의 음원 분리 방식은 이러한 구별 없이 음원 추정을 수행하므로 비보컬만 존재하는 구간에서 불필요한 비보컬 음원이 보컬 음원으로 오추정되며, 이로 인해 분리 성능이 저하된다는 문제가 있다. 이러한 문제를 개선하기 위해 본 논문에서는 음원 분리 방식과 보컬 및 비보컬 분류 방식을 결합한 분리 방식에 대해 제안한다.

최근 제안된 커널 모델 기반의 음원 분리 방식은[2] 음원의 특성에 따른 커널을 이용해 음원을 추정하여 분리하는 방식으로, 보컬과 비보컬이 혼합된 구간에서 기존의 방식들보다 높은 추정 성능을 나타내었다. 또한 순환 신경망 기반의 방식 중 하나인 LSTM (Long-Short Term Memory) 신경망[3]은 기존의 순환 신경망에서 발생하는 사라지는 경사 문제를 해결하기 위해 등장한 구조로, 시계열 데이터를 다룸에 있어 현재 뛰어난 성능을 보이고 있다. 이에 본 논문에서는 기존의 커널 모델 기반의 음원 분리에 LSTM 신경망 기반의 보컬 및 비보컬 구간 분류를 결합한 보컬 및 비보컬 분리 방식을 제안한다. 이때 결합 구조에 따라 병렬 결합형 분리 알고리즘과 직렬 결합형 분리 알고리즘을 각각 제안한다.

II. 보컬 및 비보컬 분리 방식

본 논문에서 제안하는 커널 모델과 LSTM 신경망을 결합한 보컬 및 비보컬 분리 방식은 결합 구조에 따라 병렬 결합형 분리 알고리즘과 직렬 결합형 분리 알고리즘으로 나뉜다.

제안하는 병렬 결합형 분리 알고리즘은 커널 모델 기반의 음원 분리 방식과 LSTM 신경망 기반의 보컬 및 비보컬 구간 분류 방식을 각각 독립적으로 연결한 알고리즘을 말한다. Fig. 1의 좌측 그림은 병렬 결합형 분리 알고리즘의 구조도를 나타낸다.

본 방식에서, 모노의 혼합 음악 신호가 입력되면 STFT(Short-Time Fourier Transform) 적용부를 거쳐 커널 기반의 음원 분리부, 구간 분류를 위한 특징값 추출부로 각각 입력된다.

커널 기반의 음원 분리부에서는 입력된 신호에 대해 음원의 특성에 따른 커널이 적용되며 이로부터

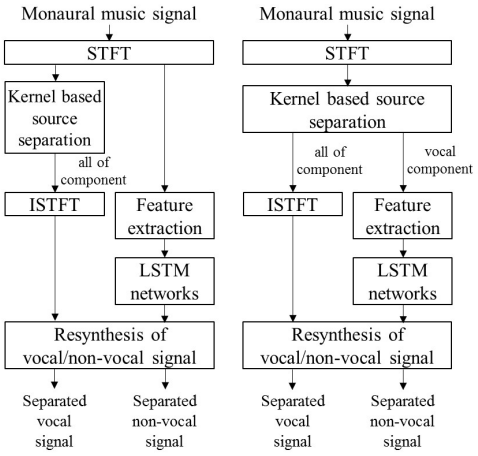

Fig. 1. The proposed vocal and non-vocal separation method. (Left) Parallel combined separation algorithm. (Right) Series combined separation algorithm.

보컬 성분과 비보컬 성분(퍼커시브, 하모닉, 주기 성분)이 분리된다. 분리된 신호들은 ISTFT(Inverse Short-Time Fourier Transform) 적용을 거쳐 보컬 및 비보컬 신호 재합성부로 입력된다.

이와 병렬적으로 입력단에서 STFT 적용부를 거친 신호는 구간 분류를 위해 특징값 추출부로 입력되며, 출력된 특징값은 다시 LSTM 신경망 적용부로 입력된다. 미리 학습된 LSTM 신경망을 통해 구간 별 보컬 및 비보컬 분류 결과가 출력되어 보컬 및 비보컬 신호 재합성부로 입력된다.

보컬 및 비보컬 신호 재합성부에서는 먼저 분리된 퍼커시브, 하모닉, 주기 신호를 모두 합하여 비보컬 신호를 생성한다. 그 후 보컬 및 비보컬 분류 결과에 따라 분리된 보컬 신호에서 비보컬 구간이라고 판단되는 구간의 신호를 비보컬 신호로 전달하여 각 신호를 재합성한다. 이를 통해 최종적으로 분리된 보컬 신호와 비보컬 신호가 출력된다.

Fig. 1의 우측 그림은 직렬 결합형 분리 알고리즘을 나타낸다. 본 알고리즘은 커널 모델 기반의 음원 분리 방식과 보컬 및 비보컬 구간 분류 방식이 순차적으로 연결된 알고리즘을 말한다. 본 방식에서는 모노의 혼합 음악 신호가 STFT 적용부를 거쳐 커널 기반의 음원 분리부, 특징값 추출부, LSTM 신경망 적용부를 순차적으로 통과한다.

커널 기반의 음원 분리부로부터 출력된 보컬 성분

은 구간 분류를 위해 특징값 추출부로 입력된다. 특징값 추출부로부터 출력되는 분리된 보컬 성분의 특징값은 LSTM 신경망 적용부로 입력되고, 이로부터 보컬 및 비보컬 구간 분류가 수행된다. 분류 결과는 보컬 및 비보컬 신호 재합성부로 입력된다. 또한 STFT 적용부를 거쳐 커널 기반의 음원 분리부에서 출력되는 전체 성분에 대해 각각 ISTFT가 적용되어 보컬 및 비보컬 신호 재합성부로 입력된다. 보컬 및 비보컬 신호 재합성부에서는 병렬 결합형 분리 알고리즘과 마찬가지의 과정을 통해 보컬과 비보컬 신호가 재합성된다. 이로부터 최종적으로 분리된 보컬 및 비보컬 신호가 출력된다.

2.1 커널 기반의 음원 분리

커널 기반의 음원 분리 방식은 음악 신호가 보컬 음원, 하모닉 음원, 그리고 주기를 가진 다수의 퍼커시브 음원으로 구성된다는 가정 하에 개별적인 커널을 이용하여 반복적 백피팅 알고리즘을 통해 각각의 음원을 추정하고 이를 기반으로 각 음원에 따른 이득값을 계산하여 음원을 분리한다.

먼저 모노의 음악 신호 $x(n)$와 그에 포함된 개별 음원의 신호 $o_j(n)$에 대해 STFT 적용부를 거친 신호를 다음과 같이 나타낼 수 있다.

$$X(\omega,t)=\sum_{j=1}^{J} O_j(\omega,t), \tag{1}$$

여기서 ω와 t은 주파수 빈과 프레임 인덱스를 나타내며, $j\in(1,...,J)$는 각 음원의 인덱스를 나타낸다.

변환된 $X(\omega,t)$에 대해 다음과 같이 분리 이득값을 적용하여 독립적인 초기 $\widehat{O}_j(\omega,t)$를 얻을 수 있다. $\widehat{O}_j(\omega,t)$는 j번째 음원의 추정된 복소수 스펙트럼을 나타낸다.

$$\widehat{O}_j(\omega,t)=G_j(\omega,t)\ \bullet\ X(\omega,t), \tag{2}$$

이때의 이득값 $G_j(\omega,t)$은 분리 이득값이며, 그 초기값은 모두 1로 설정된다.

초기 $\widehat{O}_j(\omega,t)$에 대해 j번째 음원의 형태에 따른 커널을 적용하여 각 음원의 스펙트럼을 추정한다. 이때, 커널은 개별 음원이 가지고 있는 시간-주파수 스펙트럼 상의 고유한 모양에 따라 이진의 주기 커널($j=1,2,...,J-3$), 퍼커시브 커널($j=J-2$), 하모닉 커널($j=J-1$), 보컬 커널($j=J$)로 구성된다. 개별 음원의 스펙트럼 상의 고유한 모양은 다음과 같다. 주기 성분은 일정한 주기를 따라 배열되어있는 형태를 가진다. 퍼커시브 성분은 주파수축 영역에 걸쳐 있는 형태를 가진다. 하모닉 성분은 스펙트럼 상에서 시간축으로 걸쳐있는 모양을 보인다. 시간-주파수 스펙트럼 상에서 보컬 성분은 주파수 축, 시간 축 영역 모두 존재하는 십자가 형태를 보인다. 이러한 개별 음원의 특성에 따라 각 음원에 따른 커널 모양은 Fig. 2와 같이 나타난다.[2]

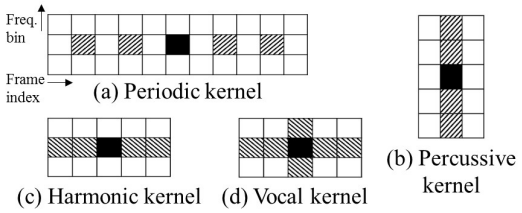

Fig. 2. The shape of kernel components.

다음과 같이 추정한 j번째 음원 $\widehat{O}_j(\omega,t)$의 파워스펙트럼에 대해 해당하는 j번째 커널이 중간값 필터링을 통해 적용되고, 이를 통해 개별 음원의 스펙트럼을 재추정할 수 있다.

$$M_j(\omega,t)=median\left[\left|\widehat{O}_j(\omega,t)\right|^2\middle|K_j(\omega,t)\right], \tag{3}$$

여기서 $K_j(\omega,t)$는 음원 j에 해당하는 커널을 나타내며 $M_j(\omega,t)$은 중간값 필터링을 통해 추정된 j번째 음원의 파워스펙트럼을 나타낸다.

추정된 음원의 파워스펙트럼은 음원 별 분리 이득값을 갱신하는데 사용되며, 갱신된 분리 이득값을 통해 다시 Eq. (2) 과정을 거쳐 갱신된 $\widehat{O}_j(\omega,t)$를 얻을 수 있다. 갱신된 $\widehat{O}_j(\omega,t)$은 다시 Eq. (3)로 입력되어 커널과 중간값 필터를 이용한 음원 스펙트럼 추정 과정에 사용되며, 음원이 충분히 분리될 때 까지 이러한 음원 스펙트럼 추정 및 분리 이득값 갱신 과정이 반복적으로 수행된다.

이와 같은 반복적 백피팅 알고리즘은 전체 스펙트럼을 대상으로 진행되므로 각 반복 단계마다 대용량의 연산을 필요로 한다. 이러한 문제를 해결하기 위해 다음과 같이 SVD(Singular Value Decomposition)을 적용하여 스펙트럼을 분해한다.

$$D_j \Sigma_j C_j = SVD[M_j(\omega,t)], \tag{4}$$

$$S_j(\omega,t) = D_j \times C_j, \tag{5}$$

여기서 D_j는 M×M형태의 열 기반의 행렬, Σ_j는 M×L 형태의 대각선 행렬, C_j는 , L×L 형태의 행 기반의 행렬을 나타낸다. 따라서 분리 이득값 $G_j(\omega,t)$의 계산 과정은 다음과 같이 나타낼 수 있다.

$$G_j(\omega,t) = S_j(\omega,t)\left[\sum_{j'=1}^{J} S_{j'}(\omega,t)\right]^{-1}. \tag{6}$$

최종적으로 계산되는 각 음원의 분리 이득값이 Eq. (2)를 통해 입력 신호의 복소수 스펙트럼에 적용되어 개별 음원의 분리된 성분을 얻을 수 있다.

2.2 특징값 추출

본 논문에서는 현재 오디오 분류에 광범위하게 사용되고 있으며 우수한 결과를 도출하는 것으로 알려진 로그 스케일의 멜 밴드 에너지 값[4]을 추출하여 사용하였다. 본 논문에서 사용한 멜 밴드 에너지 특징값의 추출 과정은 다음과 같다. 먼저 입력되는 복소수 스펙트럼에 대해 파워 스펙트럼을 계산한 후 각 프레임에 대해 로그 스케일의 멜 필터 밴드를 적용하여 특징값을 추출한다. 본 논문에서는 60개의 멜 밴드를 사용하였으며 보컬 및 비보컬 구간 분류의 효율을 높이면서 충분한 특징값을 얻기 위해 프레임 단위로 추출한 특징값에 26 프레임의 관측 윈도우와 50 % 오버래핑을 적용하여 세그먼트 단위를 설정하여 사용하였다.

2.3 LSTM 신경망

음악 신호에서의 보컬 및 비보컬 분류에 있어서 시간적 맥락은 매우 중요한 요소이다. 이는 음악 신호라는 특성에 기인하는데, 사람은 노래를 할 때에 목소리를 마치 악기처럼 사용하므로 짧은 구간에 대해서는 그 구간이 보컬신호를 포함하고 있는지 아닌지를 구분하기가 어렵다. 따라서 이를 구분하기 위해서는 보컬 구간의 시간적 맥락을 인지하여야 한다. 이에 본 논문에서는 앞서 설정한 각 세그먼트들을 보컬과 비보컬로 분류하는데 순환 신경망 구조를 기반으로 구성된 LSTM 신경망을 사용하였다.

시간에 따라 변화하는 맥락과 함께 시계열 데이터를 효과적으로 분석하기 위해서는 하나의 순간에 대해서만 해석할 것이 아니라 연속된 데이터의 관계를 해석해야 한다. 순환 신경망은 $t-1$시간의 은닉 노드가 t시간의 은닉 노드에 영향을 주는 구조로 구성된다. 따라서 순환 신경망을 이용하면 연속적인 시계열 데이터 속의 관계를 학습할 수 있다.

하지만 순환 신경망은 깊은 신경망 구조에서의 학습 시에 gradient가 사라지거나 발산하는 문제가 발생한다. 이러한 문제를 해결하고 효과적인 학습을 진행하기 위해 본 논문에서는 LSTM 유닛을 이용한 순환 신경망을 사용한다. Fig. 3는 LSTM 유닛을 나타낸다. LSTM 유닛은 기억 소자 C_t, 입력 게이트 i_t, 출력 게이트 o_t, 잊기 게이트 f_t로 이루어져 있으며 각 게이트와 기억 소자의 값은 다음과 같이 계산된다.

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f), \tag{7}$$

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i), \tag{8}$$

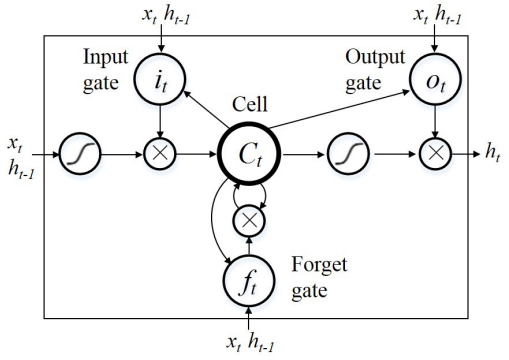

Fig. 3. The LSTM unit structure.

$$C_t = f_t C_{t-1} + i_t \phi(W_{xc} x_t + W_{hc} h_{t-1} + b_c), \quad (9)$$

$$o_t = \sigma(W_{xo} x_t + W_{ho} h_{t-1} + b_o), \quad (10)$$

$$h_t = o_t \odot \phi(C_t), \quad (11)$$

$$y_t = softmax(W_y h_t), \quad (12)$$

여기서 x_t는 시간 t에서의 입력데이터, y_t는 시간 t에서의 출력 데이터, h_t는 시간 t에서의 은닉 상태 값을 나타낸다. W, 는 각 상태에서의 가중치, b는 바이어스 값을 나타낸다. $\sigma(\bullet)$는 sigmoid 함수, $\phi(\bullet)$는 *tanh* 함수를 나타내며 $\odot$는 요소 곱을 나타낸다.

Eq. (7)의 잊기 게이트는 h_{t-1}, x_t와 바이어스의 가중합에 sigmoid 함수를 취한 형태로, 0과 1사이의 값을 계산하며, 기존의 정보를 얼마나 잊어버릴지 결정한다. 여기서 1은 기존의 정보를 모두 기억하는 것을, 0은 모두 잊어버리는 것을 의미한다. 다음 단계의 Eq. (8)의 입력 게이트는 잊기 게이트와 유사하게 계산되며 어떤 새로운 정보를 읽어올 지를 결정한다. 다음으로, 잊기 게이트의 출력과 입력 게이트의 출력을 바탕으로 Eq. (9)와 같이 현재의 셀 상태를 갱신한다. Eq. (10)의 출력 게이트를 통해 무엇을 출력할지가 결정되고, Eq. (11)을 통해 최종 은닉 상태 값 h_t가 출력된다. 마지막으로, Eq. (12)를 통해 가중치 W_y와 곱해지고 *softmax* 함수를 거쳐 y_t가 출력된다.

III. 실험결과

본 논문에서 제안한 방식에 대한 성능 측정을 하기 위해 BSS Eval.(Blind Source Separation Evaluation)[5] 방식의 SDR(Source-to-Distortion Ratio), SIR(Source-to-Interference Ratio) 측정 실험을 진행하였다. SDR과 SIR 모두 값이 클수록 분리 성능이 높음을 의미한다. 실험을 위해 SiSEC 2016에서 제공하는 DSD100 dataset와 스튜디오에서 자가 녹음한 음원을 사용하였다. 음원은 스테레오, 44.1 kHz의 샘플링레이트로 구성되어 있다. 실험에서는 2채널의 신호의 평균을 내어 모노로 사용하였으며, 16 kHz로 down sampling하여 사용하였다. 또한 임의로 음원을 선정하여 60개를 학습에, 40개의 음원을 테스트에 사용하였다. 선정한 음원의 장르는 어쿠스틱, 컨트리, 락, 재즈, 오케스트라 팝, 인디 팝 등을 포함하며, 보컬에 전자음이 섞여있는 일렉트로닉, 디스코, 메탈 등의 장르는 포함하지 않았다. 또한 가사를 비교적 명확하게 발음하는 음원을 사용하였으며 허밍위주로 진행되거나 코러스가 과도한 음원은 배제하였다.

전체 실험을 위한 STFT 과정에서 80 ms의 프레임 크기와 80 % 오버래핑을 적용하였다. 또한 보컬/비보컬 구간 분류를 위해 개별 프레임에서 특징값을 추출 한 후 26개의 연속적인 프레임을 하나의 세그먼트로 설정하여 50 % 오버래핑과 함께 LSTM 신경망의 입력 특징 값으로 사용하였다. 신경망의 출력은 각 세그먼트에 대한 확률 값으로 출력되며, 앞서 50 %의 오버래핑을 적용하였으므로 출력 값을 overlap-add 한 후 문턱값을 적용하여 보컬 및 비보컬 구간 분류를 수행하였다. LSTM 신경망은 1개의 은닉 레이어, 256개의 노드를 사용하였으며 최대 15 epoch를 50의 batch 크기를 이용하여 학습하였다.

제안한 방식의 성능 측정을 위해 기존의 커널 기반의 분리 방식(Method1), 제안한 병렬 결합형 분리 알고리즘(Method2), 직렬 결합형 음원 분리 알고리즘(Method3)에 대한 실험을 각각 실시하여 그 결과를 비교하였다. Fig. 4는 원본 보컬 신호와 Method1, 2, 3에 대한 출력 보컬 신호를 차례로 나타내며 Table 1은 위의 3가지 방식의 SDR, SIR 측정 결과를 나타낸다.

Fig. 4의 그림을 비교해보면, Method1의 보컬 출력물에 남아있던 비보컬 잔여물이 Method2, 3의 출력물에는 제거됨을 알 수 있다. 또한 Table 1의 측정 실험 결과에서 제안한 Method2, 3의 결과가 보컬과 비보컬 신호 모두에서 Method1의 SDR, SIR 결과를 뛰어넘는 것을 확인할 수 있다. Table 1의 측정 결과에서 제안하는 병렬 결합형 분리 알고리즘과 직렬 결합형 분리 알고리즘이 거의 대등한 성능을 보이지만 미세하게 병렬 결합형 분리 알고리즘의 성능이 더 좋은 것을 확인할 수 있다. 이는 Fig. 4에서 보이다시피 병렬 결합형 분리 알고리즘의 보컬 및 비보컬 분류 성능이 직렬 결합형 분리 알고리즘보다 약간 더 우수하기 때문에 발생하는 결과이다.

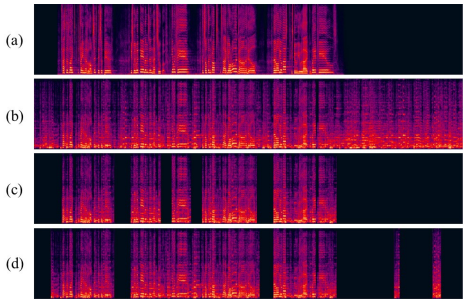

Fig. 4. Spectrogram of each signal. (a) original vocal signal, (b) separated vocal signal of method1, (c) separated vocal signal of method2, (d) separated vocal signal of method3.

Table 1. Performance for vocal and non-vocal separation.

Ratio	Separation performance for non-vocal		Separation performance for vocal	
	SDR	SIR	SDR	SIR
Method 1	6.42	15.43	0.35	8.45
Method 2	8.15	15.88	1.72	12.42
Method 3	8.13	15.82	1.68	12.37

위와 같은 결과를 통해 제안한 두 가지 결합 방식이 모두 기존의 커널 기반의 음원 분리 방식의 문제점을 개선하며 보컬 및 비보컬 분리의 성능을 향상시키는 것을 확인할 수 있다.

IV. 결 론

본 논문에서는 기존의 보컬 및 비보컬 분리 방식의 비보컬 구간에서 발생하는 음원 오추정 문제를 개선하기 위해 커널 기반의 음원 분리 방식과 LSTM 신경망 기반의 보컬 및 비보컬 분류 방식을 결합한 방식에 대해 제안하였으며 방식 간의 결합 구조에 따라 병렬형 결합 분리 알고리즘과 직렬형 결합 분리 알고리즘을 각각 제안하였다. 실험 결과를 통해 제안하는 방식들이 거의 대등한 성능을 보이며 모두 기존의 커널 모델 기반의 분리 방식에서 나타나는 문제점을 개선하고, 이를 통해 더욱 효과적인 보컬 및 비보컬 분리를 수행하는 것을 확인할 수 있었다.

향후 본 논문을 기반으로 음원 추정 과정에서 구간 분류를 실시간으로 적용하여 더욱 효과적으로 음원을 분리하는 방식에 대해 연구할 예정이다.

감사의 글

이 논문은 2015년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (NRF-2015R1D1A1A01059804).

References

1. E. Vincent, N. Bertin, R. Gribonval, and F. Bimbot, “From blind to guided audio source separation: How models and side information can improve the separation of sound,” IEEE Signal Processing Magazine **31**, 107-115 (2014).
2. A. Liutkus, D. Fitzgerald, and Z. Rafii, “Scalable audio separation with light kernel additive modeling,” IEEE ICASSP, 76-80 (2015).
3. S. Hochreiter and J. Schmidhuber, “Long short-term-memory,” Neural computation **9**, 1735-1780 (1997).
4. G. Parascandolo, H. Huttunen, and T. Virtanen, “Recurrent neural networks for polyphonic sound event detection in real life recordings,” IEEE ICASSP, 6440-6444 (2016).
5. E.Vincent, R. Griboncal and C. F´evotte, “Performance measurement in blind audio source separation,” IEEE Transactions on Audio, Speech and Language Processing, 1462-1469 (2006).

저자 약력

▸ 조 혜 승 (Hye-Seung Cho)

2015년 2월: 광운대학교 전자융합공학과 학사
2015년 3월~현재: 광운대학교 전파공학과 석박사 통합 과정

▸ 김 형 국 (Hyoung-Gook Kim)

1999년 ~ 2002년 7월: 독일 SIEMENS/ Cortologic AG 책임연구원
2002년 ~ 2005년 3월: 독일 베를린 공과대학교 Assistant Professor
2005년 ~ 2007년 2월: 삼성종합기술원 수석연구원
2007년 3월 ~ 현재: 광운대학교 전자융합공학과 교수