

# 커널 백피팅 알고리즘 기반의 가중 $\beta$ -지수승 최소평균제곱오차 추정방식을 적용한 보컬음 분리 기법

## Vocal separation method using weighted $\beta$ -order minimum mean square error estimation based on kernel back-fitting

조혜승, 김형국<sup>†</sup>

(Hye-Seung Cho and Hyoung-Gook Kim<sup>†</sup>)

광운대학교 전자공학과

(Received August 13, 2015; accepted September 17, 2015)

**초 록:** 본 논문에서는 커널 백피팅 알고리즘에 가중  $\beta$ -지수승 최소평균제곱오차 추정방식(weighted  $\beta$ -order minimum mean square error: WbE)을 적용한 보컬음 분리 방식에 대해 제안한다. 음성 향상 방식에서, WbE는 진폭 성분 기반 MMSE(Minimum Mean Square Error) 추정방식, 로그 스펙트럼 진폭 기반 MMSE 추정방식 등과 같은 기존의 베이지안(Bayesian) 기반의 추정방식들 보다 객관적 및 주관적 측면에서 모두 보다 높은 성능을 나타내는 방식으로 잘 알려져 있다. 이에 본 논문에서는 기본적인 반복적 커널 백피팅 알고리즘에 WbE를 적용하여 음악 신호에서의 보컬음 분리 성능을 향상시키고자 하였다. 실험결과는 본 논문에서 제안한 방식이 기존의 분리 방식보다 분리 성능이 더 뛰어나다는 것을 보인다.

**핵심용어:** 보컬음 분리, 커널 백피팅, 가중  $\beta$ -지수승 최소평균제곱오차 추정 방식

**ABSTRACT:** In this paper, we propose a vocal separation method using weighted  $\beta$ -order minimum mean square error estimation (WbE) based on kernel back-fitting algorithm. In spoken speech enhancement, it is well-known that the WbE outperforms the existing Bayesian estimators such as the minimum mean square error (MMSE) of the short-time spectral amplitude (STSA) and the MMSE of the logarithm of the STSA (LSA), in terms of both objective and subjective measures. In the proposed method, WbE is applied to a basic iterative kernel back-fitting algorithm for improving the vocal separation performance from monaural music signal. The experimental results show that the proposed method achieves better separation performance than other existing methods.

**Keywords:** Vocal separation, Kernel back-fitting, Weighted  $\beta$ -order MMSE estimation

**PACS numbers:** 43.60.-c, 43.60.Jn

### 1. 서 론

모노 혹은 스테레오 음악 신호에서의 배경음과 보컬음의 분리는 악기 또는 가수 인식 및 분류, 배경음 또는 보컬음의 악보 전사, 음악 리믹싱, 오디오 복원 등 실생활에 쓰이는 다양한 어플리케이션에서 공통

으로 사용되는 기본적인 전처리 단계이다.

기본적으로 음악 신호의 배경음과 보컬음의 분리는 음원의 추정과 추정된 음원에 대한 분리 기술의 적용으로 수행된다. 이에 따라 NMF(Non-negative Matrix Factorization),<sup>[1]</sup> REPET(Repeating Pattern Extraction Technique)<sup>[2]</sup> 등 다양한 접근 방식을 통해 배경음과 보컬음을 추정한 후 이를 기반으로 각 음원을 분리하는 연구가 진행되었으며 최근에는 KAM(Kernel Additive Modeling)<sup>[3]</sup>을 적용한 음원 분리 방식이 제

<sup>†</sup>Corresponding author: Hyoung-Gook Kim (hkim@kw.ac.kr)  
Department of Electronics Convergence Engineering, Kwang-woon University, 20 Gwangun-ro, Nowon-gu, Seoul 01897, Republic of Korea  
(Tel: 82-2-940-5574, Fax: 82-2-913-5006)

안되었다.

KAM은 음악 신호가 일정한 주기를 가진 다수의 퍼커시브 음원, 하모닉 음원, 보컬 음원의 합으로 구성된다는 가정 하에 각각의 음원을 개별적으로 추정하는 방식으로, 반복적 백피팅(back-fitting) 알고리즘을 적용하여 각 음원의 스펙트럼을 재추정하여 갱신한다. 또한 음원 분리 단계에서 위너 필터(Wiener filter) 방식을 적용하여 추정한 음원을 분리한다.

일반적으로 음성향상 알고리즘에서 사용되는 베이지안(Bayesian) 최소평균제곱오차(Minimum Mean Square Error: MMSE) 추정방식<sup>[4]</sup>은 음성 추정에 있어서 기존의 위너 필터 방식의 성능보다 우수하고, 가중  $\beta$ -지수승 MMSE 추정방식(Weighted  $\beta$ -order MMSE Estimation: WbE)<sup>[4]</sup>은 일반적인 베이지안 MMSE 추정방식에 지수 파라미터를 적용하여 사람의 청각적 지각 특성을 반영한 것으로 기존의 베이지안 MMSE 추정방식 기반의 다양한 음성 추정 방식 중 가장 우수한 성능을 제공하는 것으로 널리 알려져 있다.

이에 본 논문에서는 WbE와 KAM의 백피팅 알고리즘을 결합하여 배경음과 보컬음의 분리 성능을 개선시킨 음원 분리 방식에 대해 제안한다.

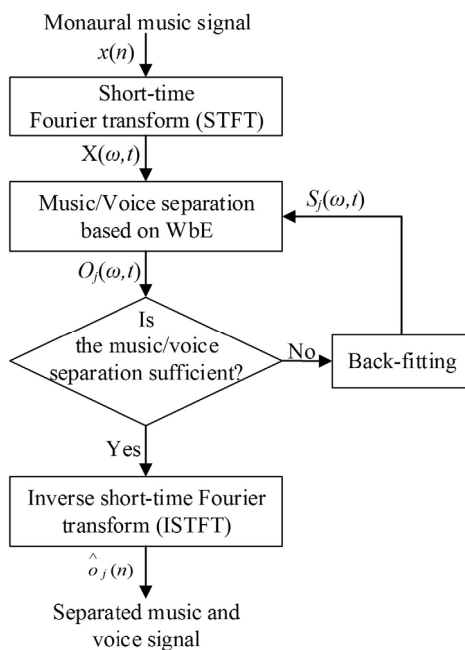


Fig. 1. Overall flow chart of proposed music/voice separation algorithm.

## II. 배경음/보컬음 분리 알고리즘

Fig. 1은 본 논문에서 제안하는 배경음과 보컬음의 분리 알고리즘의 전체 블록도이다. 제안하는 알고리즘은 STFT(Short-Time Fourier Transform), WbE 기반의 배경음/보컬음 분리부, 백피팅 결정부, 백피팅부, ISTFT(Inverse Short-Time Fourier Transform)의 5개의 모듈로 구성된다.

입력되는 모노의 음악 신호  $x(n)$ 은 일정한 주기를 가지는 다수의 퍼커시브 음원, 하나의 하모닉 음원, 하나의 보컬 음원으로 구성된  $J$ 개의 음원  $o_j(n)$ 의 합으로 구성된다는 가정 하에 다음과 같이 나타낼 수 있다.

$$x(n) = \sum_{j=1}^J o_j(n), \quad (1)$$

여기서  $j(1 < j < J)$ 는 각 음원의 인덱스,  $n$ 은 시간 축 샘플의 인덱스를 나타내며  $o_j(n)$ 은 개별 음원의 오디오 신호를 나타낸다.

제안하는 배경음과 보컬음의 분리 알고리즘은 다음과 과정을 통해 수행된다.

먼저, 입력된 모노의 음악 신호  $x(n)$ 은 다음과 같이 STFT 과정을 거쳐 시간 축 신호가 주파수 축의 복소수 스펙트럼으로 변환된다.

$$X(\omega, t) = \sum_{n=0}^{N-1} x(Rt+n)w(n)\exp\left(\frac{-i2\pi\omega n}{N}\right), \quad (2)$$

여기서  $R$ 은 프레임의 크기,  $t$ 는 프레임 인덱스,  $w(n)$ 은 윈도우 함수를 나타내며  $N$ 은 윈도우의 크기,  $\omega$ 는 주파수 빈 인덱스를 나타낸다.

변환된 복소수 스펙트럼  $X(\omega, t)$ 는 WbE 기반의 배경음 및 보컬음 분리부에 입력되며 WbE를 통해 개별 음원의 복소수 스펙트럼  $O_j(\omega, t)$ 로 분리된다.

백피팅 결정부에서는 현 단계에서 분리된 음원의 추정 스펙트럼과 이전 단계의 추정 스펙트럼을 비교한다. 현 단계와 이전 단계의 스펙트럼의 차이가 백피팅 문턱값보다 작거나 같다면 현 단계의 추정 스펙트럼에 ISTFT를 적용하여 시간 축 신호로 변환한

후 분리된 음원을 출력한다. 반대로 현 단계와 이전 단계의 스펙트럼의 차이가 백피팅 문턱값보다 크면 차이가 문턱값에 수렴할 때까지 백피팅이 반복적으로 수행된다.

백피팅 과정에서는 앞서 추정된 음원의 스펙트럼이 커널과 함께 매디언 필터링을 통해 재추정된다. 음원의 특성에 따라 각기 다른 커널이 적용되며 커널의 적용 과정에 대해서는 다음 장에서 다룬다.

### 2.1 백피팅 알고리즘을 통한 음원 스펙트럼 재추정

백피팅 결정부에서 현재의 음원이 충분히 분리되지 않았다고 판단될 경우 백피팅 알고리즘을 통한 음원 스펙트럼의 재추정이 수행된다.

백피팅 알고리즘을 통한 음원 스펙트럼의 재추정 과정에서는 앞서 추정된 음원의 파워스펙트럼에 음원 별 특성에 따른 커널이 매디언 필터링을 통해 적용되며, 이러한 과정을 거쳐 분리하고자 하는 음원의 특성에 부합하는 스펙트럼이 재추정된다.

먼저, 추정된 각 음원의 복소수 스펙트럼  $O_j(\omega, t)$ 에 대한 파워 스펙트럼  $V_j(\omega, t)$ 을 구한다.

$$V_j(\omega, t) = |O_j(\omega, t)|^2. \quad (3)$$

구해진 개별 음원의 파워스펙트럼  $V_j(\omega, t)$ 에 음원 별 특성에 따른 커널을 기반으로 매디언 필터링이 다음과 같이 수행된다.

$$M_j(\omega, t) = \text{median} [ V_j(\omega, t) | K_j(\omega, t) ], \quad (4)$$

여기서  $M_j(\omega, t)$ 은 재추정된 음원의 파워스펙트럼,  $K_j(\omega, t)$ 는 커널을 나타내며 인덱스  $j$ 에 따라서 각각 퍼커시브 커널( $j=1, 2, \dots, J-2$ ), 하모닉 커널( $j=J-1$ ), 보컬 커널( $J$ )로 구성된다.

이 때 커널은 각 음원의 시간 축, 주파수 축의 분포 특성에 따라 퍼커시브 커널, 하모닉 커널, 보컬 커널<sup>[3]</sup> 등으로 다음과 같이 구분되어 적용된다: 1) 퍼커시브 음원에는 스펙트럼 상에서 시간축을 따라 일정한 주기를 가지며 주파수축에 존재하는 커널로서 퍼커시

브 커널이라 정의한다. 2) 하모닉 음원에는 스펙트럼 상에서 시간 축으로 일정하게 존재하는 형태의 하모닉 커널이 적용된다. 3) 보컬 음원에는 스펙트럼 상에서 주파수 축, 시간 축 영역에 모두 존재하는 십자가 형태의 보컬 커널이 적용된다.

반복적 백피팅 알고리즘은 스펙트럼 전체를 대상으로 수행되는 방식이므로 각 반복 단계마다 대용량의 변수를 처리해야 하며, 따라서 연산량이 증가하고 방대한 저장 공간을 필요로 한다는 문제가 발생한다. 본 논문에서는 이러한 문제를 해결하고 연산 효율을 높이면서 분리 성능을 개선시키기 위해 앞서 매디언 필터를 거쳐 재추정된 음원의 파워스펙트럼  $M_j(\omega, t)$ 에 SVD(Singular Value Decomposition)를 적용하여 스펙트럼을 분해한다.

$$S_j(\omega, t) = D_j \Sigma_j C_j = SVD [ M_j(\omega, t) ], \quad (5)$$

여기서 재추정된 음원의 파워스펙트럼  $M_j(\omega, t)$ 는  $M \times M$  형태의 열 기반의 행렬  $D_j$ ,  $M \times L$  형태의 대각선 행렬  $\Sigma_j$ ,  $L \times L$  형태의 행 기반의 행렬  $C_j$ 의 곱으로 분해된다.

SVD를 통해 분해된 스펙트럼  $S_j(\omega, t)$ 는 WbE 기반의 배경음 및 보컬음 분리부로 입력된다.

### 2.2 가중 $\beta$ -지수승 MMSE 추정방식 기반의 배경음 또는 보컬음 분리

본 논문에서 제안하는 배경음과 보컬음의 분리 과정에서는 더 향상된 분리 성능을 위해 SVD를 통해 분해된 스펙트럼에 위너 필터 대신 WbE가 적용된다.

#### 2.2.1 가중 $\beta$ -지수승 MMSE 추정방식

효과적인 WbE의 적용을 위해 지각 가중치 변수  $\alpha_j(\omega, t)$ 와 SVD 기반의 진폭 변수  $\beta_j(\omega, t)$ 가 적응적으로 적용된다. WbE 기반의 이득 값 적용을 통한 음원분리는 전체 스펙트럼에 대한 각 음원 스펙트럼의 이득 값을 계산한 후 이를 입력 스펙트럼에 곱하여 추정된 음원의 스펙트럼을 분리하는 방식으로 수행된다.

먼저 다음과 같이 입력되는 SVD 기반의 분해된 스

펙트럼  $S_j(\omega, t)$ 을 합하여 전체 스펙트럼을 구한다.

$$W(\omega, t) = S_1(\omega, t) + S_2(\omega, t) + \dots + S_f(\omega, t). \quad (6)$$

구해진 전체 스펙트럼  $W(\omega, t)$ 와 현재 음원의 스펙트럼  $S_j(\omega, t)$ 를 바탕으로 각 음원 별 사전 SNR  $\xi_j(\omega, t)$ 와 사후 SNR  $\gamma_j(\omega, t)$ 를 구한다.

$$\begin{aligned} \xi_j(\omega, t) &= \frac{S_j(\omega, t)}{W(\omega, t) - S_j(\omega, t)}, \\ \gamma_j(\omega, t) &= \frac{|X(\omega, t)|^2}{W(\omega, t) - S_j(\omega, t)}, \\ \chi_j(\omega, t) &= \frac{\xi_j(\omega, t)}{1 + \xi_j(\omega, t)} \gamma_j(\omega, t). \end{aligned} \quad (7)$$

추정된 개별 음원의 분리를 위해, WbE 기반의 이득 함수는 다음과 같이 주어진다.

$$\begin{aligned} G_j(\omega, t) &= \frac{\sqrt{\chi_j(\omega, t)}}{\gamma_j(\omega, t)} \\ &\cdot \left[ \frac{\Gamma\left(\frac{\beta_j(\omega, t) - 2\alpha_j(\omega, t)}{2} + 1\right)}{\Gamma(-\alpha_j(\omega, t) + 1)} \right]^{\frac{1}{\beta_j(\omega, t)}}, \quad (8) \\ &\cdot \left[ \frac{\Phi\left(-\frac{\beta_j(\omega, t) - 2\alpha_j(\omega, t)}{2}, 1; -\chi_j(\omega, t)\right)}{\Phi(\alpha_j(\omega, t), 1; \chi_j(\omega, t))} \right] \end{aligned}$$

여기서  $\alpha_j(\omega, t)$ 와  $\beta_j(\omega, t)$ 는 인간의 청각적 지각 특성에 기반한 변수를 나타내며  $\Gamma(\bullet)$ 는 감마함수,  $\Phi(\bullet)$ 는 confluent hypergeometric 함수를 나타낸다.

주어진 WbE의 이득 함수에 앞서 계산한 음원 별 사전 SNR, 사후 SNR을 적용하여 음원에 따른 이득 값을 도출한다. 이때  $\alpha_j(\omega, t)$ ,  $\beta_j(\omega, t)$  변수가 적응적으로 적용되며 이를 통해 더욱 효과적인 음원 분리가 수행된다.

최종적으로 다음과 같이 입력 복소수 스펙트럼  $X(\omega, t)$ 에 음원 별 이득 값  $G_j(\omega, t)$ 를 곱하는 방식으로 추정된 음원의 복소수 스펙트럼을 분리할 수 있다.

$$O_j(\omega, t) = G_j(\omega, t) \cdot X(\omega, t). \quad (9)$$

### 2.2.2 적응적 $\alpha_j(\omega, t)$ , $\beta_j(\omega, t)$ 변수 계산

지각 가중치 변수  $\alpha_j(\omega, t)$ 와 SVD 기반의 진폭 변수  $\beta_j(\omega, t)$ 는 마스킹 효과, 달팽이관의 비선형적 압축 특성, 음량 인지 특성(perceived loudness) 등의 인간의 청각적 지각 특성에 기인한다. 따라서 각 음원의 특성에 따라 적응적으로  $\alpha_j(\omega, t)$ 와  $\beta_j(\omega, t)$ 를 적용하여 음원 분리 성능의 향상을 불러올 수 있다.

$\alpha_j(\omega, t)$ 와  $\beta_j(\omega, t)$ 의 계산 과정은 다음과 같다. 음원에 따라 변화하는 시간 축, 주파수 축의 특성에 적응적으로  $\alpha_j(\omega, t)$ 와  $\beta_j(\omega, t)$ 를 도출하기 위해 먼저, 시간 축에서의 주파수 특성을 반영하는 Sub-band SNR  $Z_j(t)$ 를 다음과 같이 구한다.

$$Z_j(t) = 10 \log_{10} \frac{\sum_{\omega=0}^{\Omega-1} |W(\omega, t) - \sqrt{W(\omega, t) - S_j(\omega, t)}|^2}{\sum_{\omega=0}^{\Omega-1} [W(\omega, t) - S_j(\omega, t)]}. \quad (10)$$

구해진 Sub-band SNR  $Z_j(t)$ 와 마스킹 문턱값  $T^{[5]}$ 를 결합하여 변수  $\hat{\alpha}_j(t)$ 를 다음과 같이 2차 다항식 형태로 표현한다.

$$\begin{aligned} \hat{\alpha}_j(t) &= F[Z_j(t), T_j] \cong \sum_{p=0}^{\infty} \sum_{q=0}^{\infty} c_{pq} Z_j^p(t) T_j^q \\ &\cong e_0 + e_1 Z_j(t) + e_2 T_j + e_3 Z_j(t) T_j, \end{aligned} \quad (11)$$

여기서  $c_{pq}$ 는 다항 계수를 나타내며  $e_0 = 0.765$ ,  $e_1 = -0.123$ ,  $e_2 = -0.265$ ,  $e_3 = -0.07$ 은 실험을 통해 얻어진 경험적 값이다. 또한 마스킹 문턱값  $T$ 는 Deng *et al.*<sup>[5]</sup>이 도출한 값을 사용한다.

위와 같이 Sub-band SNR  $Z_j(t)$ 와 청각적 마스킹 효과를 반영한 변수  $\hat{\alpha}_j(t)$ 를 기반으로 다음과 같은 방식을 통해  $\alpha_j(\omega, t)$ 가 적응적으로 계산된다.

$$\begin{aligned} \alpha_j(\omega, t) &= \alpha_{low} + a \cdot \frac{(f_{\omega} - 2000)(\alpha_{high} - \alpha_{low})}{F_s/2 - 2000} \\ &\quad + (1 - a) \cdot \hat{\alpha}_j(t), \end{aligned} \quad (12)$$

여기서  $\alpha_{low}(=0.25)$ 와  $\alpha_{high}(=0.94)$ 는 현재 음원의 분리에 있어 나머지 음원들과의 상호보완적 역할을 수

행하며  $a$  ( $0 < a < 1$ )는 스무딩 파라미터,  $f_\omega$ 는 각 주파수 빈에 따른 주파수 값(Hz)을 의미한다. (즉,  $f_\omega = \omega F_s / N$ 이며  $F_s$ 는 샘플링레이트를 말한다.)

이와 병렬적으로, 진폭 변수  $\beta_j(\omega, t)$ 를 도출하기 위해서 다음과 같이  $\beta_{low}$ 와  $\beta_{high}$  사이의 선형 보간을 통해 중간 주파수에서의 압축률  $\hat{\beta}_j(\omega)$ 가 계산된다.

$$\hat{\beta}_j(\omega) = \beta_{high} - \frac{\frac{1}{\eta} \log_{10} \left( \frac{f_\omega}{A} + l \right)}{\frac{1}{\eta} \log_{10} \left( \frac{F_s}{2A} + l \right)} (\beta_{high} - \beta_{low})$$

for  $1 \leq j \leq J$ , (13)

여기서  $\beta_{high}$  ( $= 0.2$ )와  $\beta_{low}$  ( $= 1$ )는 각각 저 주파수대역, 고 주파수대역의 압축률을 나타내며  $\eta = 0.06$  mm,  $l = 1$ ,  $A = 165.4$  Hz를 사용한다.

현재 추정된 음원과 나머지 음원들과의 상호보완을 위해  $\tilde{\beta}_j(\omega)$ 는  $[\beta_{min}, \beta_{max}]$ 의 범위의 값을 사용하며 다음의 관계로 계산된다.

$$\tilde{\beta}_j(t) = \min [\max(\tau \cdot Z_j(t) + \lambda, \beta_{min}), \beta_{max}]. \quad (14)$$

이로써 Sub-band SNR과 달팽이관의 비선형적 압축특성, 음량 인지 특성 등을 고려하는 진폭 변수  $\beta_j(\omega, t)$ 는 다음과 같이 구해진다.

$$\beta_j(\omega, t) = b \cdot \hat{\beta}_j(\omega) + (1 - b) \cdot \tilde{\beta}_j(t), \quad (15)$$

여기서  $\tau = 0.45$ ,  $\lambda = 1.3$ ,  $\beta_{min} = 0.4$ ,  $\beta_{max} = 4.0$ 을 나타내며  $b$  ( $0 < b < 1$ )는 스무딩 파라미터를 의미한다.

### III. 실험결과

본 논문에서 제안한 음원 분리 방식에 대한 객관적인 성능 측정을 위해 BSS Eval(Blind Source Separation Evaluation)<sup>[6]</sup>의 NSIR(Normalized Source-to-Interference Ratio)과 NSDR(Normalized Source-to-Distortion Ratio)의 측정실험을 진행하였다.

NSIR, NSDR은 음악 신호와 분리된 보컬음 간의

Table 1. Comparative performance for music/voice separation.

Ratio	Separation Performance for Music		Separation Performance for Voice	
	NSDR	NSIR	NSDR	NSIR
SVD-GW-KAM	6.83	9.65	2.35	6.23
SVD-LSA-KAM	7.36	10.48	2.87	6.74
SVD-WbE-KAM	8.94	12.54	4.75	8.54

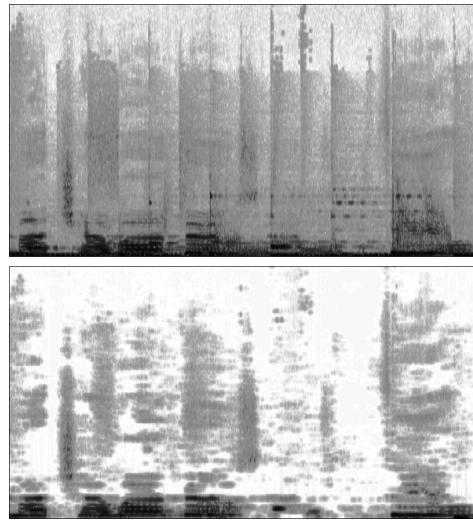


Fig. 2. Mixed music signal and separated voice.

SIR, SDR의 개선 정도를 추정하는 지표로 값이 클수록 분리 성능이 좋다는 것을 의미한다.

실험에 사용된 음악 신호는 44.1 kHz의 샘플링레이트, 16-bit의 깊이, PCM 포맷의 모노의 신호를 사용하였으며, ccMixer database에서 제공하는 다양한 장르의 음원 50곡, 스튜디오에서 자가 녹음한 음원 50곡, 총 100곡의 음원을 대상으로 실험을 진행하였다.

제안하는 음원 분리 방식의 성능 비교를 위해 다음과 같이 커널 백피팅 알고리즘 기반의 3가지 음원 분리 방식에 대한 배경음과 보컬음의 분리 실험을 실시하였다: 1) SVD-GW-KAM: SVD를 적용하여 분해된 파워 스펙트럼에 위너 필터 방식이 적용된 방식; 2) SVD-LSA-KAM: SVD를 적용하여 분해된 파워 스펙트럼에 로그 스펙트럼 진폭 기반의 MMSE 추정 방식이 적용된 방식; 3) SVD-WbE-KAM: SVD를 적용하여 분해된 파워 스펙트럼에 WbE가 적용된 방식.

Table 1은 위의 3가지 방식의 배경음과 보컬음의 분리 실험 결과를 나타내며 Fig. 2는 본 논문에서 제

안하는 방식(SVD-WbE-KAM)을 통해 음악 신호에서 분리된 보컬음의 시간-주파수 축에서의 출력 결과를 나타낸다.

실험결과 본 논문에서 제안한 SVD-WbE-KAM 방식의 NSDR과 NSIR이 다른 방식들과 비교하여 가장 우수한 결과를 제공하며, 기본적인 위너 필터를 적용한 방식인 SVD-GW-KAM 방식은 실험 방식들 중 가장 낮은 성능을 나타낸다. 또한 Fig. 2로부터 제안하는 방식을 통해 음악신호에서 보컬음이 분리된 출력결과를 확인할 수 있다. 이로써 본 논문에서 제안하는 방식이 음악 신호에서의 배경음과 보컬음의 분리에 있어서 가장 우수한 성능을 보이는 것을 확인할 수 있다.

## IV. 결 론

본 논문에서는 백피팅 알고리즘을 기반으로 WbE를 적용하여 음악 신호에서 배경음과 보컬음을 분리하는 방식에 대해 제안하였다. 제안한 알고리즘은 인간의 청각적 지각 특성을 고려하는 음성 추정 및 분리 방식인 WbE를 반복적 커널 백피팅 알고리즘에 적용한 것으로, 이를 통해 배경음과 보컬음의 분리 성능을 기존의 방식보다 보다 더 향상시킬 수 있었다. 향후 본 논문에서 제안하는 방식을 입체음향 재현 모바일 어플리케이션에 적용하는 방안에 대해 연구할 예정이다.

## 감사의 글

본 연구는 미래창조과학부 및 정보통신기술진흥센터의 대학 ICT 연구센터육성 지원사업의 연구결과로 수행되었음(IITP-2015-H8501-15-1016). 본 논문은 2015년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(NRF-2015R1D1A1A01059804).

## References

1. S. Vembu and S. Baumann "Separation of vocals from polyphonic audio recordings," in Proc. International Society

- for Music Information Retrieval Conference, 337-344 (2005).
2. Z. Rafii and B. Pardo, "Repeating pattern extraction technique (REPET): a simple method for music/voice separation," IEEE Trans. Audio, Speech, Language Process. **21**, 71-82 (2013).
3. A. Liutkus, D. Fitzgerald, Z. Rafii, B. Pardo, and L. Daudet, "Kernel additive models for source separation," IEEE Trans. Signal Process. **62**, 4298-4310 (2014).
4. E. Plourde and B. Champagne, "Auditory-based spectral amplitude estimators for speech enhancement," IEEE Trans. Audio, Speech, Language Process. **16**, 1614-1623 (2008).
5. F. Deng, F. Bao, and C.-C. Bao, "Speech enhancement using generalized  $\beta$ -order spectral amplitude estimator," Speech Commun. **59**, 55-68 (2014).
6. E. Vincent, R. Griboncal, and C. F'evotte, "Performance measurement in blind audio source separation," IEEE Trans. Audio, Speech, Language Process. **14**, 1462-1469 (2006).
7. A. Liutkus, D. Fitzgerald, and Z. Rafii, "Scalable audio separation with light kernel additive modeling," in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, 76-80 (2015).

## 저자 약력

### ▶ 조 혜 승 (Hye-Seung Cho)



2015년: 광운대학교 전자융합공학과  
학사  
2015년 ~ 현재: 광운대학교 전파공학과  
석사과정

### ▶ 김 형 국 (Hyoung-Gook Kim)



1999년 ~ 2002년: 독일 SIEMENS/ Cortologic  
AG 책임연구원  
2002년 ~ 2005년: 독일 베를린 공과대학교  
Assistant Professor  
2005년 ~ 2007년: 삼성종합기술원 수석  
연구원  
2007년 ~ 현재: 광운대학교 전자융합공학과  
교수