Journal of Korea Multimedia Society Vol. 16, No. 10, October 2013(pp. 1171-1179)
http://dx.doi.org/10.9717/kmms.2013.16.10.1171

하모닉 정보를 이용한 SAOC의 보컬 신호 제거 방법에 관한 연구

박지훈[†], 장대근[††], 한민수[††]

요 약

IAS는 대게 사용자가 자신의 취향에 맞는 음악을 직접 제작 및 편집 가능한 기능을 제공하는 서비스이다. SAOC는 낮은 전송률로 IAS가 가능한 다객체 오디오 코딩 기술이다. 하지만 SAOC 기법은 특정 객체를 제거하는 경우, 특히 보컬 객체를 제거하는 경우 배경음악에 보컬 객체의 하모닉이 남아있는 문제점이 있다. 그래서 본 논문은 하모닉 추출과 제거를 사용한 보컬 객체 제거 기법을 제안한다. 제안 하는 기법은 부호화기에서 추출한 하모닉 정보를 이용하여 복호화기에서 보컬 객체 신호를 다운믹스 신호에서 제거하는 기법이다. 하모닉 정보로써, 기본 주파수, MVF, 하모닉 크기를 사용한다. 성능평가로 객관적, 주관적 실험을 수행하였으며 모든 실험 결과를 통해 SAOC 기법보다 제안하는 기법이 우수함을 확인한다.

A Study on Vocal Removal Scheme of SAOC Using Harmonic Information

Ji-Hoon Park[†], Dae-Geun Jang[††], Min-Soo Hahn[†††]

ABSTRACT

Interactive audio service provide with audio generating and editing functionality according to user's preference. A spatial audio object coding (SAOC) scheme is audio coding technology that can support the interactive audio service with relatively low bit-rate. However, when the SAOC scheme remove the specific one object such as vocal object signal for Karaoke mode, the scheme support poor quality because the removed vocal object remain in the SAOC-decoded background music. Thus, we propose a new SAOC vocal harmonic extranction and elimination technique to improve the background music quality in the Karaoke service. Namely, utilizing the harmonic information of the vocal object, we removed the harmonics of the vocal object remaining in the background music. As harmonic parameters, we utilize the pitch, MVF(maximum voiced frequency), and harmonic amplitude. To evaluate the performance of the proposed scheme, we perform the objective and subjective evaluation. As our experimental results, we can confirm that the background music quality is improved by the proposed scheme comparing with the SAOC scheme.

Key words: SAOC, Vocal Removal(보컬 제거), Harmonic Information(하모닉 정보)

※ 교신저자(Corresponding Author) : 박지훈, 주소 : 대전광역시 유성구 구성동 한국과학기술원 N1동 313호(305-701), 전화 : 042) 350-8778, FAX : 042) 350-8700, E-mail : batho2n@kaista.ac.kr
접수일 : 2013년 08월 30일, 수정일 : 2013년 09월 12일
완료일 : 2013년 09월 22일
[†] 스마트 IT 융합 시스템 연구단
[††] 한국과학기술원 전기및전자공학과
(E-mail: jangdg85@kaist.ac.kr)
[†††] 한국과학기술원 전기및전자공학과
(E-mail: mshahn2@kaist.ac.kr)
※ 이 논문은 2012년 정부(교육과학기술부)의 재원으로 (재)스마트 IT 융합 시스템 연구단(글로벌프론티어사업)의 지원을 받아 수행된 연구임 ((재)스마트 IT 융합시스템 연구단-2012054202)

1. 서 론

기존의 음악 서비스 제공자는 사용자에게 여러 가지 오디오 객체들이 믹싱(mixing) 되어있는 음악만을 제공한다. 그리고 서비스를 제공받는 사용자는 제공된 음악의 볼륨만 조절 가능한 제한적인 서비스를 제공받았다. 하지만 근래에 들어 음악 신호로부터 특정 객체의 소리만을 제거하거나 각 객체의 편집이 가능한 새로운 오디오 서비스의 요구가 사용자들로부터 발생하였고, 이에 따라서 각 오디오 객체 신호를 모두 제공하는 IAS(interactive audio service)라 불리는 MUSIC 2.0 서비스가 제공되었다[1,2]. 믹싱된 음악만을 제공하던 기존의 오디오 서비스와 달리 IAS는 음악을 구성하는 각각의 객체신호들과 프리셋(preset) 정보가 사용자에게 제공된다. IAS는 크게 프리셋 모드와 인터랙티브(interactive) 모드를 제공 가능해야 한다. 프리셋 모드는 프로듀서가 미리 만들어놓은 음악을 재생 가능하게 각 객체의 믹싱 게인(gain)에 관한 정보인 프리셋 정보를 통해 사용자가 음악을 청취하는 모드이다. 반면에 인터랙티브 모드는 사용자가 제공된 프리셋 정보를 사용하지 않고 자신의 취향에 따라 제공된 오디오 객체 신호를 조절하여 자신만의 음악을 청취하는 모드이다. 현재의 유일한 IAS인 MUSIC 2.0 서비스는 앞의 두 가지 모드를 모두 제공하지만, 이 서비스는 다음 그림 1에 나타나듯이 각 객체를 따로 부호화하여 사용자에게 제공하기 때문에 비트율(bit rate)이 오디오 객체의 숫자에 비례하여 증가하여 모바일(mobile) 환경에는 적합지 않다. 즉, 현재 제공되고 있는 IAS는 광대역 네트워크가 보장되거나 많은 용량을 보유한 장치에 적합한 서비스이다. 이러한 IAS가 모바일 환경에서도 제공하기 위해서는 비트율을 감소시켜야 할 필요가 있다. 그래서 SAOC(spatial audio object coding) 기법이 해결책으로써 사용되었다.

근래에 MPEG(moving picture experts group) 표준화가 완료된 SAOC는 다객체(multi-object) 오디오 신호를 한 개의 다운믹스(down-mix) 신호와 공간파라미터로 표현하는 기법이다[3-5]. 하지만 이 SAOC 기법은 특정 객체를 완전히 제거하는 경우, 재생된 음악의 음질이 저하되는 문제점이 있어 곧바로 IAS로써 사용할 수 없다. 특히 보컬(vocal) 객체 신호를 제거하는 노래방 모드에서 제거한 보컬 객체의 하모닉(harmonic) 성분이 배경음악에 남아 있어 음질이 저하된다. 그래서 본 논문에서는 특정 객체가 제거된 음악을 많이 듣는 시나리오를 가정하여 사용자가 가장 많이 사용하는 노래방 모드 즉, 보컬 객체를 제거하는 경우 음질을 향시키는 방법을 제안한다.

SAOC 기법으로 보컬 신호가 제거된 배경음악의 음질을 향상시키기 위해 본 논문에서는 하모닉 성분을 추출하는 방법과 하모닉 성분 제거 방법을 포함한 S-VHC (SAOC vocal harmonic coding) 기법을 제안한다. 보컬 객체의 성공적인 제거를 위하여 부호화 단에서는 하모닉 정보로써 하모닉의 간격, 하모닉의 크기, 하모닉의 범위를 추출하고 전송한다. 복호화 단에서는 사용자가 보컬객체 제거를 원할 때, SAOC 기법으로 보컬이 제거된 배경음악에 남아있는 보컬

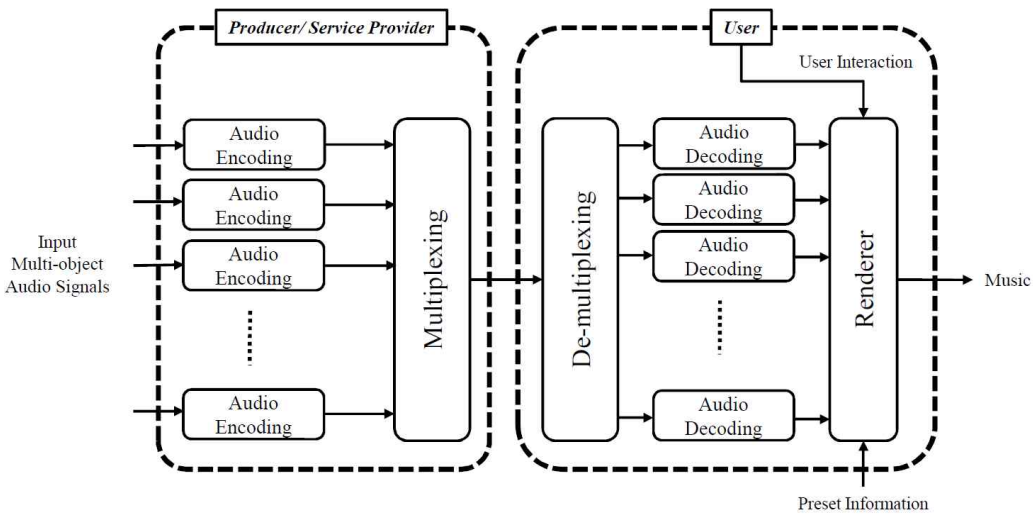


그림 1. MUSIC 2.0 서비스 개념도

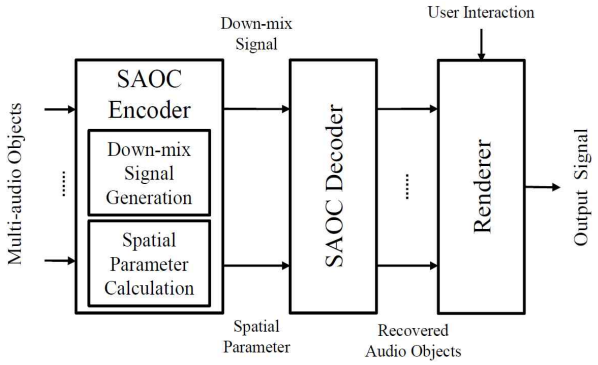


그림 2. 기본 SAOC 구조도

객체의 하모닉 성분을 전송된 하모닉 정보를 이용하여 제거한다.

본 논문의 구성은 다음과 같다. 2장에서는 기존 SAOC기법과 특정 객체신호(보컬 객체)를 제거할 때의 문제점에 대해 다루고, 3장에서는 제안하는 S-VHC 기법에 대하여 자세하게 설명한다. 4장에서 실험결과를 통해 제안하는 방법의 성능을 확인하고, 5장에서 결론을 맺는다.

2. Spatial Audio Object Coding (SAOC) 기법

2.1 SAOC의 기본적인 개념

SAOC는 그림 2에 나타나듯이 크게 부호화기(encoder), 복호화기(decoder), 렌더러(renderer) 세 부분으로 구성되어 있다[3-5].

부호화기는 입력 다객체 오디오 신호로부터 다운믹스 신호와 공간파라미터를 계산하여 부호화한다. 먼저 다운믹스 신호 $x_d(n)$은 입력 객체 신호들의 가중치 합으로 쉽게 생성된다. 공간파라미터 계산을 위해 각 i번째 객체 신호 $x_i(n)$을 주파수도메인 신호 $X_i(k)$로 변환한다. 변환된 객체신호는 사람 청각특성을 반영한 대역폭이 ERB(equivalent rectangular band-width)가 되는 파라미터 서브밴드(sub-band)로 나누어져 각 서브밴드마다 공간파라미터를 계산하게 된다. 공간파라미터로는 OLD(object level difference)를 사용한다. b번째 서브밴드의 i번째 객체의 OLD는 서브밴드에서 가장 큰 파워를 갖는 객체의 파워로 정규화(normalization)로 정의되며 다음 식과 같이 계산할 수 있다.

$$OLD_i(b)=\frac{P_i(b)}{P_{\max}(b)},\quad \begin{matrix} i=1,\cdots,N \\ b=1,\cdots,B \end{matrix}, \tag{1}$$

여기서 N과 B는 각각 입력 객체신호와 파라미터 서브밴드의 개수이고 파라미터 서브밴드 파워 $P_i(b)$는 다음 식 2와 같이 정의된다.

$$P_i(b)=\sum_{k=A_{b-1}}^{A_b-1}|X_i(k)|^2\ , \tag{2}$$

여기서 A_b는 b번째 서브밴드의 구분 경계 인덱스(index)이다. 본래 표준의 SAOC에서는 객체신호의 주파수 변환을 위하여 QMF(quadrature mirror filter)를 사용하지만, 본 논문에서는 제안하는 하모닉 파라미터 추출을 위하여 DFT(discrete Fourier transform)을 사용하여 주파수 변환을 수행하였다. 따라서 파라미터 QMF와 동일한 서브밴드의 대역폭을 계산하여 다음 표 1과 같이 파라미터 서브밴드의 경계를 정리하였다.

복호화기에서는 전송된 다운믹스 신호와 공간파라미터를 이용하여 각 오디오 객체 신호를 복원한다. 각 객체 신호를 복원하기 위한 각 객체 신호의 서브밴드 게인은 공간파라미터로 계산가능하고 다운믹스 신호에 곱함으로써 각 객체 신호 $\hat{X}_i(k)$를 복원한다. 각 객체 신호를 구하는 과정은 다음 식과 같다.

표 1. 파라미터 서브밴드 경계 인덱스 (DFT 크기: 2048, 표본화율: 44.1 kHz)

A_0	A_1	A_2	A_3	A_4	A_5	A_6	A_7
0	3	7	11	15	19	23	27
A_8	A_9	A_{10}	A_{11}	A_{12}	A_{13}	A_{14}	A_{15}
31	39	47	55	63	79	95	111
A_{16}	A_{17}	A_{18}	A_{19}	A_{20}	A_{21}	A_{22}	A_{23}
127	159	191	223	255	287	318	367
A_{24}	A_{25}	A_{26}	A_{27}	A_{28}			
415	479	559	655	1025			

$$\hat{X}_i(k) = X_d(k)\sqrt{OLD_i(b)/\sum_{j=1}^{N} OLD_j(b)}\,,\ k \in b \qquad (3)$$

렌더러에서는 복원된 객체 신호들을 사용자의 취향에 따라 리믹스(remix)하고 렌더링하고 최종적으로 IDFT(inverse DFT)를 통해 출력 신호를 생성한다. 본 논문에서는 복호화기와 렌더러를 전체 복호화기 한 과정으로 가정하고 SAOC 구조를 크게 부호화기와 복호화기로 분류하여 연구를 진행한다.

2.2 SAOC기법의 문제점

SAOC는 특정 객체를 완전히 제거할 때, 특히 노래방 서비스처럼 보컬 객체를 제거하고자할 때 성능 저하가 크게 발생한다. 그 이유는 SAOC 기법이 낮은 주파수 해상도를 갖는 파라미터 서브밴드기반의 처리 기법이고, 전송된 다운믹스 하나의 신호에서 각 객체의 파워 비율로써 객체 신호를 복원하기 때문에 보컬 객체 제거시 배경음악에 보컬 객체가 남아있는 문제가 발생한다. 보컬 객체가 남아 있는 이유로는 보컬의 유성음 구간의 하모닉이 가장 큰 원인이다. SAOC 기법에서 객체 신호를 제거할 때 OLD 파라미터만을 사용하는데, OLD파라미터는 파라미터 서브밴드의 파워 비율을 계산하므로, OLD 파라미터는 주파수 신호의 세세한 신호 값이 아닌 전반적인 스펙트럼 포락선(envelope) 정보를 나타낸다. 그러므로 OLD 파라미터만을 가지고 보컬 객체를 제거하고자 하면, 하모닉 성분에 의해 주파수 신호의 편차가 심한 유성음 구간에서 신호가 잘 제거되지 못하는 문제점이 발생한다. 다음 그림 3은 다운믹스 신호에서 OLD 파라미터만을 사용하여 보컬 객체를 제거했을 때 보컬 객체의 하모닉 신호가 배경음악에 남아있는 스펙트로그램(spectrogram)들이다. 마지막 그림 3-(c)에서 확인 가능하듯이 그림 3-(a)에는 없는 하모닉 성분이 배경음악에 남아 있는 것을 확인 가능하다.

3. 제안하는 보컬제거 기법

기존의 OLD 파라미터만을 사용하여 객체를 제거했던 SAOC 기법은 보컬객체 제거시 음질 열화가 발생한다. 그래서 본 논문에서는 다음 그림 4와 같은 보컬 객체의 하모닉 정보를 이용한 보컬객체 제거 기법인 SAOC 보컬 하모닉 코딩 (SAOC vocal harmonic coding: S-VHC) 기법을 제안한다. 제안하는 기법은 상대적으로 낮은 비트율만 증가하면서도 노래방모드로 사용가능할 정도의 배경음악을 제공한다. 제안하는 기법을 위해서 시간-주파수 변환은 QMF 대신에 DFT를 사용하며, 파라미터 서브밴드 경계는 위의 표 1의 값들을 사용한다.

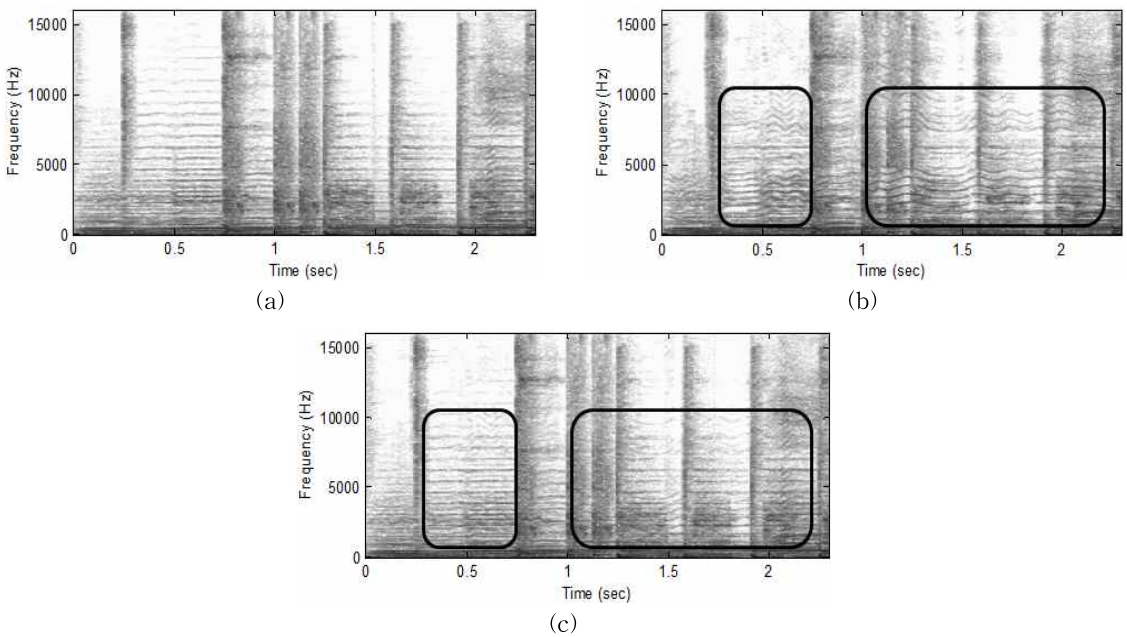


그림 3. SAOC기반 객체 저거의 문제점 (a: 보컬객체가 없는 다운믹스 신호 스펙트로그램, b: 보컬객체가 포함된 다운믹스 신호 스펙트로그램, c: b 신호로부터 OLD 파라미터를 사용하여 보컬객체를 제거한 신호 스펙트로그램)

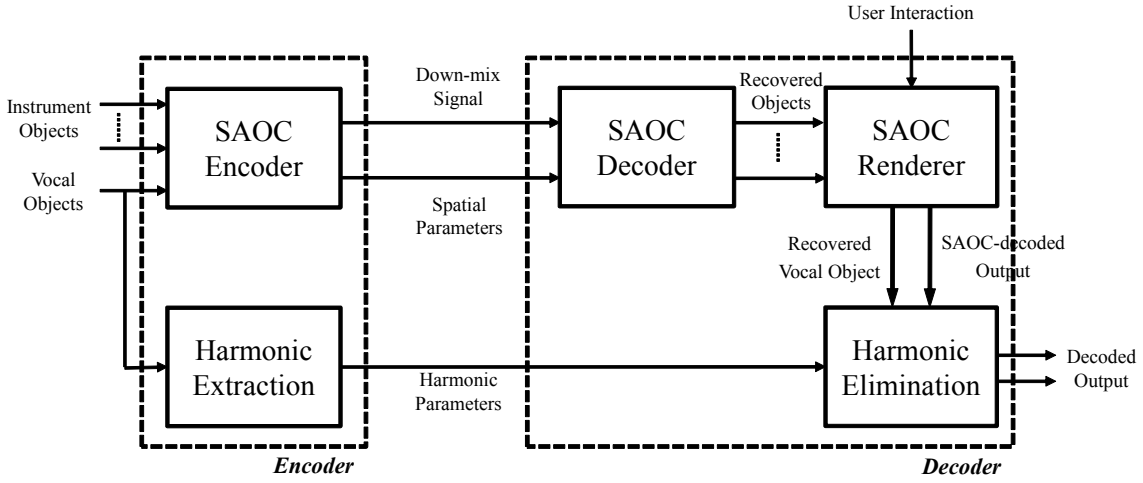


그림 4. 제안하는 보컬 제거 기법 구조도

제안하는 보컬 하모닉 코딩 기법 부호화기는 먼저 기존 SAOC 부호화기와 동일하게 다객체 입력 신호로부터 다운믹스 신호와 공간 파라미터인 OLD 파라미터를 계산한다. 다음으로 하모닉 추출(harmonic extraction) 블록에서 보컬 객체로부터 하모닉 파라미터들을 추출하여 복호화기단으로 전송한다. SAOC 복호화기에서는 전송된 다운믹스 신호와 공간 파라미터를 이용하여 각 객체 신호를 복원하고 렌더러에서 사용자의 취향에 따라 각 객체 신호를 조절한다. 마지막으로 하모닉 제거(harmonic elimination) 블록에서는 사용자가 노래방 모드, 즉 보컬객체 제거를 원하지 않으면 렌더러에서 출력된 신호를 그대로 출력한다. 반면에 사용자가 노래방 모드를 원한다면 전송된 하모닉 정보를 이용하여 OLD 파라미터로 보컬 객체가 제거된 배경음악에 남아있는 보컬 하모닉 성분을 마저 제거하여 배경음악을 출력한다.

3.1 하모닉 정보 추출 방법

하모닉 정보를 추출하기 위해 먼저 하모닉 정보를 구성하는 파라미터들을 정의할 필요가 있다. 본 논문에서는 하모닉의 간격을 나타내는 신호의 주기(pitch), 하모닉의 크기(harmonic amplitude), 하모닉의 범위(maximum voiced frequency)를 하모닉을 구성하는 파라미터로써 정의한다. 제안하는 하모닉 정보 추출 블록의 구성도는 다음 그림 5와 같다.

신호의 주기(pitch)는 하모닉 성분의 간격을 나타내는 파라미터로써, MVF와 하모닉 크기를 계산하기 위해 꼭 필요한 파라미터이다. 신호의 주기 성분

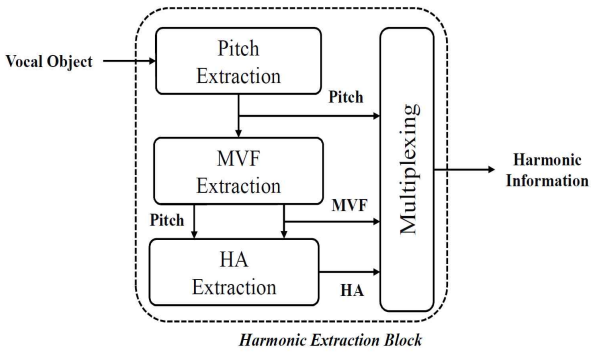


그림 5. 하모닉 정보 추출 블록 구성도

을 추출하는 연구는 이미 오래전부터 다양하게 연구되어 왔다[6-12], [19]. 본 논문에서는 다양한 신호 주기 추출 알고리즘 중에서 돌출 함수(salience function) 알고리즘을 사용하였다[11]. 이 알고리즘은 일정 간격의 하모닉 크기들의 합으로 신호주기를 계산하는 함수이다. 돌출 함수를 이용하기 위하여 전처리 단계로써 신호의 백색화(whitening)을 수행한다. 백색화가 수행된 신호 $Y(k)$에 돌출함수를 신호주기 후보 τ에 따라 계산하면

$$s(\tau)=\sum_{m=1}^{M}\max|Y(k)| \quad , \quad k\in\kappa_{\tau,m} \tag{4}$$

여기서, m은 정수이고, M 하모닉의 개수이다. 그리고 $\kappa_{\tau,m}$는 다음과 같은 범위로 정의된다.

$$\kappa_{\tau,m}=\left[\left\langle\frac{mK}{\tau-\triangle\tau/2}\right\rangle,\ldots,\left\langle\frac{mK}{\tau+\triangle\tau/2}\right\rangle\right] \tag{5}$$

여기서 K는 DFT크기이고, $\langle\bullet\rangle$는 반올림 연산자이다. 이때 돌출함수의 가장 큰 값을 갖는 τ가 신호의

주기가 된다.

사람이 발성한 유성음은 대게 하모닉 성분을 갖고 있지만, 발성기관을 거치면서 에너지의 마찰 및 손실로 고주파 대역까지 하모닉 성분이 남아 있지 않다. 그래서 주파수상에서 상대적으로 주기적인 하모닉 성분이 존재하는 끝점을 MVF(maximum voiced frequency)라 한다[13-16]. 이 MVF 파라미터는 HMM(hidden Markov model) 기반 음성 합성기에서 소개된 파라미터로써, 본 논문에서는 하모닉 성분이 어디까지 존재하는지에 대한 파라미터로써 이용된다. f_c의 차단 주파수(cut-off frequency)를 갖는 고주파 통과필터 $H_{f_c}(n)$을 통과한 보컬 객체 $x_{v,f_c}(n)$를 이용하여 신호주기만큼의 딜레이(delay)를 갖는 신호와의 정규화 자기상관도를 차단주파수를 증가시키며 계산한다. 정규화 자기상관도 식은 다음과 같이 정의된다.

$$R_{f_c}(\tau)=\frac{\sum_{n=0}^{L-1} x_{v,f_c}(n)x_{v,f_c}(n+\tau)}{\sqrt{\sum_{n=0}^{L-1}\left[x_{v,f_c}(n)\right]^2\sum_{n=0}^{L-1}\left[x_{v,f_c}(n+\tau)\right]^2}}, \quad (6)$$

여기서, L은 프레임(frame) 크기이다. 차단주파수를 500 Hz씩 증가시키며 식 (6)의 값을 계산할 때, 처음으로 0.5보다 작아지는 고주파 대역 통과 필터의 차단 주파수를 MVF라 정의한다.

하모닉의 크기 정보로는 하모닉의 피크점 크기를 추출한다. 하모닉의 피크(peak)는 신호의 주기의 역수인 기본주파수(F0: fundamental frequency)의 정수배의 자리에 위치한다. 하모닉 크기 $H(m)$은 다음 식 (7)과 같이 계산한다.

$$H(m)=|X_v(mF_0)|^2 \ , \ m=1,\ldots,M \quad (6)$$

여기서 m은 정수이고, M은 하모닉의 개수로, $M=\langle f_{mvf}/F_0\rangle$과 같이 구해진다.

3.2 하모닉 제거 방법

하모닉 제거 블록은 그림 6에 보이듯이 SAOC로 복호화 된 신호가 입력 신호이다. 사용자가 노래방 모드를 원하지 않는다면, SAOC로 복호화된 신호가 그대로 출력된다. 반면에 사용자가 노래방 모드를 원한다면, 하모닉 제거 블록에서는 전송된 하모닉 정보와 하모닉 필터링과, 스무싱 필터링을 통해 SAOC로 복호화된 신호에 남아있는 보컬 하모닉 신호를 마저

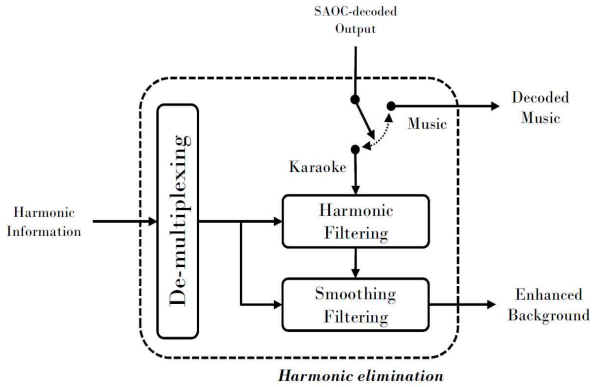


그림 6. 하모닉 제거 블록 구성도

제거한다.

사용자가 노래방 모드를 원할 때, SAOC 복호화기에서 복원된 보컬 신호 $\widehat{X}_v(k)$와 배경음악 $\widehat{X}_b(k)$는 다운믹스 신호 $X_d(k)$와 OLD 파라미터로 다음과 식 (7), (8)과 같이 계산할 수 있다.

$$\widehat{X}_v(k)=X_d(k)\sqrt{OLD_v(b)/\sum_{j=1}^{N}OLD_j(b)}\,, \quad (7)$$

$$\widehat{X}_v(k)=X_d(k)\sqrt{1-OLD_v(b)/\sum_{j=1}^{N}OLD_j(b)}\,. \quad (8)$$

하모닉 제거 필터를 설계하기 위해서 다운믹스 신호와 공간 파라미터를 사용하여 주파수 서브밴드의 파워를 계산한다. 그리고 식 (7)과 식(8)로 구해진 신호들의 파워 스펙트럼을 구하고, 하모닉 제거 필터 게인 $G_E(k)$를 다음 식과 같이 설계한다.

$$G_E(k)=\begin{cases}\sqrt{1-\dfrac{H^2(m)-|\widehat{X}_v(k)|^2}{|\widehat{X}_b(k)|^2}}, & k=m\times F_0\\ 1 & ,otherwise\end{cases}. \quad (9)$$

하모닉 제거필터를 통해 하모닉이 제거된 배경음악 $\widehat{X}_m(k)$은 기본주파수의 정수배 위치에 있는 하모닉 피크만을 제거한 필터이다. 그래서 하모닉 피크 주위의 불연속성을 제거하기 위한 스무싱 처리가 필요하다. 스무싱 필터 게인 $G_S(k)$는 다음과 같이 설계한다.

$$G_S(k)=\begin{cases}\dfrac{\sum_{q=-W/2}^{W/2}\left[\widehat{X}_m(k+q)\right]^2}{W\left[\widehat{X}_m(k)\right]^2}, & \lambda-\dfrac{W}{2}\le k\le\lambda+\dfrac{W}{2}\\ 1 & , \ otherwise\end{cases}, \quad (10)$$

여기서 W는 하모닉의 대역폭이고, λ는 기본주파수의 정수 곱이다. 하모닉이 제거된 배경 음악에 스

무싱 필터를 통과시켜 최종 출력 신호인 개선된 배경 음악 $\hat{X}_e(k)$을 출력한다.

4. 실험과 결과

4.1 실험 환경

성능 평가를 위하여 5곡의 한국 가요를 실험 콘텐츠로 사용하였다. 각각의 노래는 보컬, 기타, 피아노, 드럼 등의 4-6 개가량의 스테레오 오디오 객체로 구성되어있다. 모든 객체 신호들은 44.1 kHz의 표본화율(sampling rate)을 갖고 16 bit로 양자화 되었으며, 평균 음악 길이는 20초이다. 분석 윈도우 크기는 2048 샘플, 오버랩(overlap) 크기는 1024 샘플을 사용하였다. 신호의 주파수 변환을 위해 2048-FFT(fast Fourier transform)을 수행하였고, 파라미터 서브밴드의 경계는 표 1을 사용하였다. 성능평가는 객관적 성능 평가와 주관적 성능평가를 모두 수행하였고, 다운믹스 신호 이외의 추가정보에 대한 비트율도 같이 측정하였다. 다운믹스 신호는 128 kbps AAC(advanced audio coding) 기법으로 부호화하여 복호화기로 전송하였다. 비교 방법인 SAOC는 OLD 파라미터만을 사용하여 보컬 객체를 제거하고, S-VHC는 제안하는 방법으로 보컬 객체를 제거한다.

4.2 객관적 성능 평가

객관적 성능 평가는 세그먼트 SNR(segment signal-to-noise ratio)과 SKLD(symmetric Kullback-Leibler distance)를 측정한다[17]. 세그먼트 SNR과 SKLD를 측정하는 식은 다음과 같다.

$$SNR_{SEG} = 10\log\left(\frac{\sum_{n\in L} p(n)^2}{\sum_{n\in L}(p(n)-q(n))^2}\right), \quad (11)$$

$$D_{SKLD} = 10\log\left(\sum_{k\in K}(P(k)-Q(k))\log\frac{P(k)}{Q(k)}\right), \quad (12)$$

여기서 $p(n)$와 $P(k)$는 기준 신호이고, $q(n)$와 $Q(k)$는 비교 신호이다.

실험 결과는 다음 표 2에 정리되어 있다. 표 2에 보이듯이 기존 SAOC 기법의 성능이 SEGSNR은 20.91 dB, SKLD는 33.81 dB일 때, 제안하는 방법은 6 kbps의 정보를 더 사용하면서 SEGSNR이 23.22 dB, SKLD는 26.63 dB로 모두 성능이 향상되었다.

표 2. 객관적 성능평가 결과

비교 방법	SEGSNR (dB)	SKLD (dB)	Bit-rate (kbps)
SAOC	20.91	33.81	18.84
S-VHC	23.22	26.63	24.03

4.3 주관적 성능 평가

주관적 성능 평가로는 10명의 청취자가 5개의 콘텐츠에 관하여 MUSHRA(multiple stiuli with hidden reference and anchor) 실험을 수행했다[18]. MUSHRA 실험은 오디오 신호의 주관적 성능평가 방법으로 가장 널리 알려진 방법으로, 평가 신호로, 히든(hidden) 기준 신호와, 3.5 kHz 저대역 통과 필터를 통과한 앵커(anchor) 신호를 같이 평가하는 방법이다. 실험 결과에서 히든 기준 신호의 MUSHRA 값이 95점 이상이어야 실험의 신뢰성이 있다고 판단된다.

실험의 결과는 다음 그림 7에 정리되어있다. 각 마크는 '+'는 히든 기준 신호, '×'는 앵커 신호, '○'은 기존 SOAC 기법으로 보컬 신호를 제거한 신호, '□'는 제안하는 방법으로 보컬 신호를 제거한 신호들에 대한 MUSHRA 결과 평균값이다. 그리고 각 마크들을 관통하는 선들은 MUSHRA 값의 95% 신뢰도 구간을 나타낸다. MUSHRA 결과에서 보이듯이 모든 콘텐츠에서 제안하는 S-VHC 방법으로 보컬 객체를 제거한 배경음악이 기존의 SAOC 기법으로 보컬 객체를 제거한 배경음악보다 우수한 음질을 들려주는 것을 확인가능하다.

5. 결 론

SAOC는 낮은 전송률로 IAS가 제공 가능하한 기술이다. 하지만 노래방 서비스를 위한 보컬객체를 제거하는 것과 같은 특정 객체를 제거할 때의 성능은 아직 충분하지 않다. 그래서 본 논문에서는 SAOC 코더의 노래방 모드에서 보컬 객체 신호를 제거하는 방법에 대해서 다루고, 보컬 신호의 하모닉 정보를 함께 이용하여 보컬 객체 제거하는 기법을 제안했다. 제안하는 S-VHC 기법의 부호화기에서는 하모닉 정보를 추출하는 블록을 제안하였다. 제안한 블록은 보컬 객체 신호로부터 하모닉 파라미터로써 신호의 주기, MVF, 하모닉 크기를 추출하여 복화기로 전송

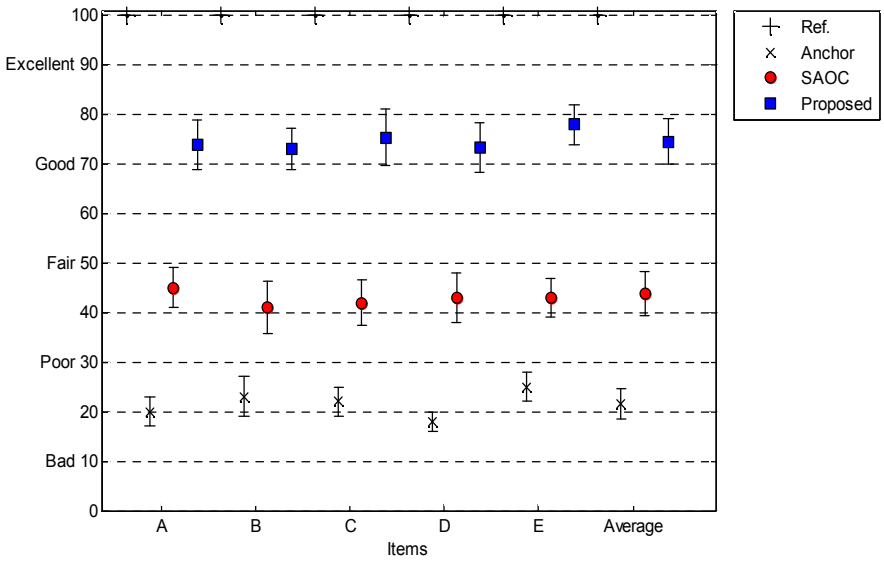


그림 7. MUSHRA 성능평가 결과

한다. 그리고 S-VHC의 복호화기에서는 SAOC기법으로 보컬 객체가 제거된 배경음악에 남아있는 보컬 하모닉을 제거하기 위하여 하모닉 추출단에서 전송된 하모닉 정보를 이용해서 하모닉 필터링, 스무싱 필터링을 수행한다. 제안하는 기법은 기존의 SAOC 기법의 전송률에 비해 6 kbps의 추가 전송률만 사용하면서도, SEGSNR과 SKLD에서 각각 2.4 dB, 7.2 dB의 이득을 얻었으며, 주관적 성능평가인 MUSHRA 실험에서도 30점 이상이 향상된 우수한 성능을 확인하였다. 향후 연구로써, 보컬 객체가 아닌 다른 악기 객체를 제거할 때 성능을 향상 시킬 수 있는 연구가 필요하다.

참 고 문 헌

[1] D. Jang, T. Lee, Y. Lee, and J. Yoo, "A Personalized Preset-based Audio System for Interactive Service," *121st AES Convention*, 2006.

[2] Consideration of Interactive Music Service, ISO/IEC JTC1/SC29/WG11 (MPEG), Archamps, Document M15390, 2008.

[3] J. Herre and S. Disch, "New Concepts in Parametric Coding of Spatial Audio: From SAC to SAOC," *2007 International Conference on Multimedia and Expo*, pp. 1894-1897, 2007.

[4] J. Engdegard, B. Resch, C. Falch, O. Hellmuth, J. Hilpert, A. Hoelzer, L. Terentiev, J. Breebaart, J. Koppens, E. Schuijers, and W. Oomen, "Spatial Audio Object Coding (SAOC) -The Upcoming MPEG Standard on Parametric Object Based Audio Coding," *124th AES Convention*, 2008.

[5] O. Hellmuth, H. Purnhagen, J. Koppens, J. Herre, J. Engdegard, J. Hilpert, L. Villemoes, L. Terentiv, C. Falch, A. Holzer, M.L. Valero, B. Resch, H. Mundt, and H. Oh, "MPEG Spatial Audio Object Coding - the ISO/MPEG Standard for Efficient Coding of Interactive Audio Scenes," *129th AES Convention*, 2010.

[6] L.R. Rabiner, M.J. Cheng, A. Rosenberg, and C.A. McGonegal, "A Comparative Performance Study of Several Pitch Detection Algorithms," *IEEE Trans. on ASSP*, vol. ASSP-24, No. 5, pp. 399-418, 1976.

[7] M. Goto, "A Predominant-F0 Estimation Method for CD Recordings: MAP Estimation using an EM Algorithm for Adaptive Tone Models," *Proc. Int. Conf on Acoustics, Speech and Signal Processing*, Vol. 5, pp. 3365 -3368, 2001.

[8] A. de Cheveigne and H. Kawahara, "YIN, a Fundamental Frequency Estimator for Speech

and Music." *The Journal of the Acoust. Soc. Am.,* Vol. 111, No. 4, pp. 1917-1930, 2002.

[9] M. Wu, D. Wang, and G.J. Brown, "A Multipitch Tracking Algorithm for Noisy Speech," *Proc. IEEE Trans. Speech and Audio*, Vol. 11, No. 3, pp. 229-241, 2003.

[10] M. Goto, "A Real-Time Music-Scene-Description System: Predominant -F0 Estimation for Detecting Melody and Bass Lines in Real-World Audio Signals," *Speech Com.,* Vol. 43, No. 4, pp. 311-329, 2004.

[11] A. Klapuri, "Multiple Fundamental Frequency Estimation by Summing Harmonic Amplitudes," *Proc. International Conference on Music Information Retrieval*, pp. 216-212, 2006.

[12] H. Fujihara, M. Goto, J. Ogata, K. Komatani, T. Ogata, and H.G. Okuno, "Automatic Synchronization Between Lyrics and Music CD Recordings Based on Viterbi Alignment of Segregated Vocal Signals," *IEEE International Symposium on Multimedia*, pp. 257-264, 2006.

[13] S. Kim, J. Kim, and M. Hahn, "HMM-Based Korean Speech Synthesis System for Hand-Held Devices," *IEEE Trans. Consumer Electronics*, Vol. 52, No. 4, pp. 1384-1390, 2006.

[14] S. Kim, J. Kim, and M. Hahn, "Implementation and Evaluation of an HMM-based Korean Speech Synthesis System," *IEICE Transactions on Information and Systems*, Vol. E89-D, No. 3, pp. 1116-1119, 2006.

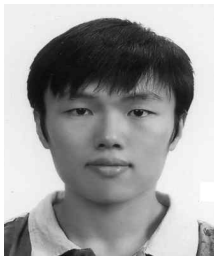
[15] S. Kim, J. Kim, and M. Hahn, "Two-band Excitation for HMM-based Speech Synthesis," *IEICE Trans. Information and Systems*, Vol. E90-D, No. 1, pp. 378-381, 2007.

[16] S. Han, S. Jeong, and M. Hahn, "Optimum MVF Estimation-Based Two-Band Excitation for HMM-Based Speech Synthesis," *ETRI Journal*, Vol. 31, No. 4, pp. 457-459, 2009.

[17] P.C. Loizou, *Speech Enhancement: Theory and Practice*, Talor & Francis, New York, 2009.

[18] ITU-R Recommendation, Method for the Subjective Assessment of Intermediate Sound Quality (MUSHRA), ITU, BS. 1543-1, 2001.

[19] T. Kim and J. Chang " A Study on Speech Period and Pitch Detection for Continuous Speech Recognition," Journal of Korea Multimedia Society, Vol. 8, no. 1, pp. 55-61, 2005.



박 지 훈

2013년 10월~현재 스마트 IT 융합 시스템 연구단, 연구교수
2013년 9월 KAIST 정보전자연구소, 연수연구원
2013년 2월 KAIST 전기및전자공학과(박사)
2007년 9월 KAIST 정보통신공학과(석사)
2005년 2월 KAIST 정보통신공학과(학사)
관심 분야 : 잡음제거, 음성 합성기, 오디오 코덱 등

장 대 근

2010년 3월~현재 KAIST 전기및전자공학과(박사과정)
2010년 2월 KAIST 정보통신공학과(석사)
2008년 2월 경희대학교 동서의료공학과(학사)
관심 분야 : 생체신호처리, 재택건강관리시스템, u-헬스케어 등

한 민 수

1998년 3월~현재 KAIST 전기및전자공학과 교수
1998년 2월 한국전자통신연구원, 책임연구원
1989년 2월 플로리다 주립대학 전기및전자공학과(박사)
1981년 9월 서울대학교 전기및전자공학과(석사)
1979년 2월 서울대학교 전기및전자공학과(학사)
관심 분야 : 음성신호처리, 음향신호처리, 생체신호처리 등