
보컬 제거 알고리즘의 실시간 구현

김현태 · 도진규 · 박장식*

동의대학교

*동의과학대학

Real-Time Implementation for Vocal-Removal Algorithm

Hyuntae Kim · Jingyu Do · Jangsik Park*

Donggeui University

*Donggeui Institute of Technology

E-mail : htaekim@deu.ac.kr

요 약

최근 고품질 원음 노래방기에 대한 관심이 증가함에 따라, 기존 미디반주기 제조사들은 원음 레코딩 방법 대신 보다 저렴한 방법으로 원음 반주 음악을 만드는 방법을 시도하고 있다. 본 논문에서는 스테레오로 보컬 및 음악이 녹음된 AR(all recorded)에서 주파수 영역에서의 에너지 차 정보를 이용하여 MR(music recorded)을 만드는 방법과 DSP(TMS320C6713)를 이용하여 이를 실시간으로 구현하는 방법에 대해 연구하였다. DSP 보드의 출력단에 6채널 오디오출력 인터페이스를 설계하여 실시간으로 원음, 보컬이 제거된 MR, 그리고 분리된 보컬이 동시에 스테레오로 출력되도록 구현하였다. DSP를 이용한 청취 테스트를 통해 보컬의 분리 및 제거가 잘 이루어지는 것을 확인할 수 있다.

ABSTRACT

Recently, According to increasing interest to original sound Karaoke instrument, MIDI type karaoke manufacturer attempt to make more cheap method instead of original recoding method. In this paper, we developed how to create MR from AR, recorded in stereo, by using the energy difference in the frequency domain and how to implement in DSP(TMS320C6713) were developed. At the output of the DSP board, 6-channel audio output interface designed for real-time stereophonic generating original sound, vocals removed MR, and separated vocals simultaneously. Real-time listening test using DSP show vocal separating and removal task successfully.

키워드

실시간 구현, 보컬 제거, 원음반주기, 주파수 영역 처리

1. 서 론

음성의 분리에 관한 연구는 광범위하게 연구 되어왔으나, 음악 속에서 보컬(vocal)을 분리하는 연구는 많지 않다. 보컬은 음성과 많은 유사성을 지닌다. 예를 들어 둘 다 유성음과 무성음으로

구성되어있다. 그러나 보컬과 음성은 중대한 차이점이 있다. 현재의 잘 알려진 차이점은 특정 노래 장르에서 보컬 포먼트가 더해진다는 것이며, 주파수 범위가 2000 ~3000Hz 사이에 위치한다[1].

노래에서 가수는 보통 반주 음악과 맞추기 위

해 고의적으로 유성음을 늘리고 무성음을 줄이는 경향이 있다. 이에 따른 직접적인 결과로서 노래를 하는 동안 유성음과 무성음의 비율이 변하며 유성음이 전체의 90%를 차지한다[2]. 음성은 이 보다 많은 비율로 무성음이 차지한다[3]. 사운드 분리 측면에서 보컬과 음성의 가장 큰 차이점은 음성은 일반적으로 하모닉성분, 비하모닉성분, 협대역 또는 광대역 성분들이 간섭하는 환경을 가지나 보컬은 항상 반주 음악과 함께 존재하며 반주 음악은 주로 보컬과 관련된 하모닉성분이며, 광대역성분을 가진다. 이러한 사실은 보컬 분리 문제가 상대적으로 보다 어렵다는 것을 말한다[4-5].

본 논문에서 스테레오(stereo)로 녹음된 AR로부터 주파수 영역에서 개별 스테레오 신호와 스테레오 채널간 차신호간의 주파수 빈별 에너지 비교를 통해 보컬의 유무를 판별하여 분리하고 DSP 보드를 통해 실시간으로 구현한다.

II. 제안하는 방법

2-1. 보컬 제거 알고리즘

AR(all recorded)된 스테레오 음악에는 왼쪽채널과 오른쪽채널에 각각 보컬의 소리가 똑같이 들어가 있다는 가정과 주파수 영역에서 MR (music recorded)의 경우는 80 ~ 400Hz사이에 분포하게 되며 보컬의 음높이는 소프라노(soprano)와 같은 경우 1400Hz까지 높게 발생하므로 MR과 보컬은 음높이의 차이점을 이용하여 보컬을 제거한다.

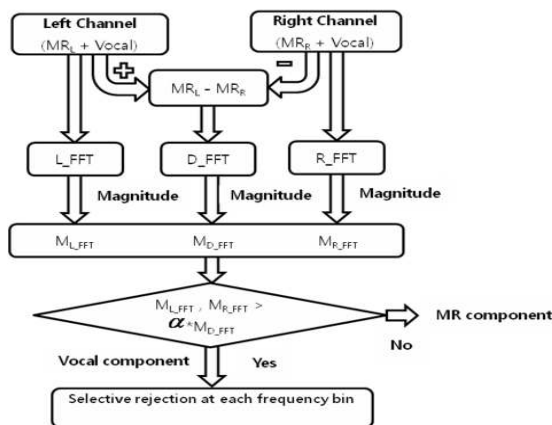


그림 1. 보컬 제거 블록도

Fig. 1. Block diagram of vocal remover

그림 1에서 AR을 스테레오로 가정한 상태에서 왼쪽채널(MR_L+Vocal)과 오른쪽 채널(MR_R+Vocal)에 각각 보컬의 목소리가 포함되어있다고 가정하여 왼쪽 채널에서 오른쪽 채널을 제거하여 목소리를 제거한 후 완전히 제거되지 않은 목소리를 제거하기 위해서 왼쪽채널과 오른쪽 채널을 각각 FFT(fast fourier transform)변환하여 음높이(pitch)가 반주음악만 포함되어있는 것의 임계치보다 크면 보컬부분이라고 판단하여 제거한다.

2-2. DSP 구현

제안하는 보컬 제거 알고리즘을 CCS(code composer studio) DSK v3.1을 이용하여 코드를 완성하였으며, 완성된 코드 중 주파수 빈별 보컬 여부를 판별하는 부분을 발췌하여 표 1에 나타내었다.

표 1. DSP 코드 일부분
Table 1. A part of DSP codes

```
// 일정한 값 이상이면 제거
for(i=0;i<FFT_N;i++) {
    if(diff_in_pw[i] < x_L_pw[i]) {
        if(diff_in_pw[i] < x_R_pw[i]) {

            // 왼쪽과 오른쪽 채널이 일정치 이상이면
            vocal_L[2*i] = x_L[2*i];
            vocal_L[2*i+1] = x_L[2*i+1];
            vocal_R[2*i] = x_R[2*i];
            vocal_R[2*i+1] = x_R[2*i+1];

            // 0으로 만들어 주어 제거한다.
            x_L[2*i] = 0;
            x_L[2*i+1] = 0;
            x_R[2*i] = 0;
            x_R[2*i+1] = 0;
        }
    }
}
```

완성된 코드를 DSP(digital signal processing) TMS320C6713 DSK 보드에 다운로드시켜 구현하였다. 보드의 오디오 출력단에 6개 채널 오디오 인터페이스를 설계하여 2개 채널에는 스테레오 AR, 또 다른 2개 채널에는 보컬 제거한 MR, 나머지 2개 채널에서는 분리된 보컬이 실시간으로 거의 동시에 출력되도록 설계하였다. 그림 2에 구현한 보컬 제거기를 나타내었다.

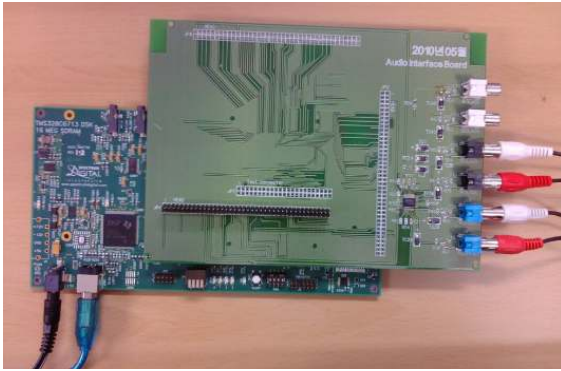


그림 2. DSP-TMS320C6713을 이용하여 구현한 보컬 제거기

Fig. 2 Implemented Vocal remover using DSP-TMS320C6713

표 2. 청취테스트 결과
Table 2. Result of listening test

파일명	곡명	가수	장르	평가		
				bad	normal	good
1.wav	cry	비바소울	힙합		o	
2.wav	그대만 보면	이정	발라드			o
3.wav	하늘을 봐	강찬	발라드			o
4.wav	이름만 들어도	강현주	발라드			o
5.wav	사랑한다는말	김동률	발라드			o
6.wav	사랑하면 버려야 할 아까운 것들	다이나믹 듀오	힙합		o	
7.wav	고집쟁이	드림콘타이거	힙합		o	
8.wav	술이야	바이브	발라드			o
9.wav	slowdownfeat	부가킹즈	힙합			o
10.wav	사랑	부활	락			o
11.wav	거리에서	성시경	발라드			o
12.wav	사랑 그대 곁에	양파	발라드			o
13.wav	사랑이 그래요	오종혁	발라드			o
14.wav	사랑 two	윤도현 밴드	락			o
15.wav	y	프리스타일	힙/발라드		o	
16.wav	Street life	DJ DOC	힙		o	
17.wav	이차선 다리	자태현	트로트			o
18.wav	어머나	장윤정	트로트			o
19.wav	편짜라	장윤정	트로트			o
20.wav	쓰러집니다	서주정	트로트			o
21.wav	루미	남진	트로트			o
22.wav	회장을 지우는 여자	강진	트로트			o
23.wav	백만송이 장미	심수봉	트로트			o
24.wav	정흔을 울러디오	나훈아	트로트			o
25.wav	Love Song	Sara Bareilles	팝송			o
26.wav	Paralyzer	Finger Eleven	팝송			o
27.wav	See You Again	Miley Cyrus	팝송		o	
28.wav	Tattoo	Jordin Sparks	팝송			o
29.wav	하루하루	빅뱅	댄스			o
30.wav	가자다가자다	브라운 아이즈	발라드			o
총계				0	5	25
%				0	0.17	0.83

III. 실험 결과 및 결론

AR로부터 보컬을 제거하는 알고리즘을 DSP 보드 상에 구현하고 6개 채널의 오디오 출력 인터페이스를 연결하여 AR, MR 및 분리된 보컬을 각각 스테레오로 동시에 실시간으로 출력하도록

구성하였다. 또한 DSP의 결과를 이용하여 다양한 장르 및 곡을 대상으로 연구에 참여한 연구원 3명을 대상으로 청취테스트를 실시하고 그 결과를 표 2에 나타내었다. 청취테스트 결과 보컬 제거한 MR의 음질에 대한 평가에서 약 83% 정도 우수한 평가를 얻었다.

참고문헌

- [1] J.Sundberg, "The acoustics of the singing voice, " Scidntific American, pp82-91, Mar.1977
- [2] Y.E.Kim,"Singing voice analysis/synthesis, "Ph.D.dissertation, MIT,Media Lab, 2003.
- [3] D.L.Wang,"Feature-based speech segregation" in Computational Auditory Scene Analysis: Principles, Algorithms, and Applications, D. L. Wang and G.J.Brown, Eds, New York:IEEE Press(dual imprint with Wiley), 2006, to appear.
- [4] Yipeng Li and DeLiang Wang, "Separation of Singing Voice from Music Accompaniment for Monaural Recordings" Technical Report: OSU-CISRC-9/05-TR61, 2005.
- [5] Jounghoon B., Hanseok K. "Spectral Subtraction Using Spectral Harmonics for Robust Speech Recognition in Car Environments," ICCS 2003, LNCS 2660, pp.1109-1116, 2003.