

FDBM의 음원분리 성능평가

이채봉*

Evaluation of a signal segregation by FDBM

Chai-Bong Lee *

요약

음원분리 방법으로는 여러 가지가 제안되고 있으나 그 중에서도 주파수영역 두 귀 모델(Frequency Domain Binaural Model : FDBM)은 저 연산량과 울림 제거에 효과적이다. FDBM에 의한 두 귀 보청 시스템은 SNR이나 기여도 함수(Coherence function)에 의한 평가로 하기 때문에 인간의 청취특성을 고려하지 않고 있다. 본 논문에서는 음질의 문제 해결을 위하여 FDBM의 음원분리 성능평가를 하였다. SNR, 기여도 함수, PESQ의 세 가지 수법을 이용하여 기본 특성에 대하여 시뮬레이션을 통하여 확인하였다. 모든 결과가 FDBM에 의해 좌우 채널간 평가치의 차가 작게 되었고, 좌우 채널이 거의 같은 레벨까지 개선되었음을 확인하였다. 그리고 음원방향을 바꾸고 음원 수를 증가시킨 경우에도 일정한 개선이 보였다. SNR과 기여도 함수, PESQ의 결과를 비교하면 PESQ의 평가에서는 입력 SNR를 변동시킨 경우에도 거의 모든 조건에서 분리에 의한 평가가 개선되었다.

ABSTRACT

Various approaches for sound source segregation have been proposed. Among these approaches, frequency domain binaural model(FDBM) has the advantages of low computational load and effective howling cancellation. A binaural hearing assistance system based on FDBM has been proposed. This system can enhance desired signal based on the directivity information. Although FDBM has been evaluated in terms of signal-to-noise ratio (SNR) and coherence function, the evaluation results do not always agree with the human impressions. These evaluation methods provide physical measures, and do not take account of perceptual aspect of human being. Considering a binaural hearing assistance system as a one of major applications, the quality of segregated sound should keep level enough.

In the paper, signal segregation performance by means of FDBM is evaluated by three objective methods, i.e., SNR, coherence and Perceptual Evaluation of Speech Quality(PESQ), to discuss the characteristic of FDBM on the sound source segregation performance. The simulation's evaluation results show that FDBM improves the quality of the left and right channel signals to an equivalent level. And the results suggest the possibility that PESQ provides a more useful measure than SNR and coherence in terms of the segregation performance of FDBM. The evaluation results by PESQ show the effects from segregation parameters and indicate appropriate parameters under the conditions. In the paper, signal segregation performance by means of FDBM is evaluated by three objective methods, i.e., SNR, coherence and PESQ, to discuss the characteristic of FDBM on the sound source segregation performance. The simulation's evaluation results show that FDBM improves the quality of the left and right channel signals to an equivalent level. And the results suggest the possibility that PESQ provides a more useful measure than SNR and coherence in terms of the segregation performance of FDBM. The evaluation results by PESQ show the effects from segregation parameters and indicate appropriate parameters under the conditions.

키워드

FDBM, SNR, Coherence Function, PESQ, Sound Source Segregation
주파수영역 두 귀 모델, 신호대 잡음비, 기여도 함수, 음성 품질의 지각평가, 음원분리

* 교신저자(corresponding author) : 동서대학교 정보시스템공학부(lcb@gdsu.dongseo.c.kr)

접수일자 : 2013. 10. 28

심사(수정)일자 : 2013. 11. 25

게재 확정일자 : 2013. 12. 16

1. 서 론

대표적인 음원분리 방법으로는 마이크로폰 배열을 이용한 빔 포밍[1]이나 독립성분 분석(Independent Component Analysis : ICA)[2, 3] 등이 있다. 빔 포밍은 마이크의 위치나 음원 위치 등을 고정 조건으로 주어야 할 필요가 있고, 사전 정보의 제약이 있다. 일반적으로 n 개의 마이크로폰의 배열에서 제거되는 잡음수의 상한은 $n-1$ 개이므로 복수 음원에서는 마이크의 규모가 크게 되며, 연산량도 증가한다. ICA에서는 사전 정보가 필요하지 않으나 선형변환이므로 분리되는 음원 수는 배열을 구성하는 마이크로폰 수 이하가 된다. 따라서 빔 포밍과 마찬가지로 복수 음원에서는 시스템의 규모가 커지며, 연산량도 증가하는 문제점이 생긴다. 이러한 수법 중에서도 주파수영역 두 귀 모델(Frequency Domain Binaural Model : FDBM)은 2 소자라는 소수 소자에 의한 복수 음원분리나 효과적인 울림 제거를 가지는 점에서 뛰어나다[4, 5]. 그리고 이 모델은 시간영역 두 귀 모델에 대하여 주파수영역에서 연산을 하기 때문에 저 연산량으로 실현 가능성이 있으며 실제 장착에 있어서도 실시간 처리가 가능하다[6, 7].

FDBM에 의한 음원분리의 응용으로서는 두 귀 보청시스템이 제안되고 있다[8]. 음원에서 나온 음은 인간이 좌우 귀에 도달할 때까지 안면에 의한 회절, 감쇠 등의 영향을 받아 시간차나 강도차가 생긴다. 인간은 이 두 귀 정보를 수단으로 음상정위를 하며 위험감지나 선택적 청취를 한다. 일반적으로 보청기를 장착하는 경우에 있어서도 마찬가지로 한 귀용 보청기를 두 귀에 장착하는 경우, 두 귀로 도달하는 음에서 적절한 두 귀 정보가 얻어지기 때문에 음상정위 보조에 효과가 기대되나, 동시에 주위의 잡음까지 증폭되기 때문에 소음환경하에서 청취는 곤란하다.

두 귀 보청시스템은 장착자가 청취를 쉽게 하기 위해 고려한 것이므로 실용화에서는 청취음의 음질이 고려된다. 즉, FDBM에 의한 음원분리 후의 음질저하 문제이다. 음원분리에 의한 잡음 억제에 어느 정도 효과가 얻어지고 있으나, 사용하는 환경에 따라 잡음 추정오차에 의한 음원분리가 적절히 이루어지지 않으며 잡음신호의 잔류나 신호성분의 손실에 의한 왜곡이 생긴다. FDBM 음원분리에 의한 음원저하는 해결해

야 할 문제이다. FDBM의 음원분리 성능을 음질의 면에서 평가하고 문제점을 명확하게 하기 위한 적절한 평가방법에 대한 검토를 할 필요가 있다.

분리음의 평가수법은 주관적 평가와 객관적 평가가 있으나 청취특성을 고려한 평가를 위해서는 청취실험에 의한 주관적 평가가 필요하다[9]. 그러나 주관적 평가는 조건에 따라 결과가 생기며 시간이나 노력이 필요하므로 많은 시스템에 대한 평가는 쉽지 않고 다른 시스템과의 비교도 어렵다. 따라서 객관적이며 인간의 청취특성을 고려한 평가방법이 필요하다. 본 논문에서는 인간의 청취특성을 고려한 전화대역 음성의 객관음성 품질평가 방법인 음성 품질의 지각평가(Perceptual Evaluation of Speech Quality : PESQ)[10]를 이용한 평가를 하였다. 종래의 연구나 다른 시스템과의 비교를 위해 널리 이용되고 있는 신호 대 잡음 비(Signal to Noise Ratio : SNR)와 기여도 함수를 이용한 평가도 병행하였다. 이러한 세 가지의 방법을 이용하여 FDBM의 음원분리에 대한 기본 특성에 대하여 검토를 하였다.

II. FDBM과 음원분리 방법

2.1 주파수영역 두 귀 모델

FDBM은 인간의 청각기능을 모델화 한 것으로 두 귀간 정보를 이용한 2소자만으로 복수 음원의 방위각 및 양각(귀와 지평선상이 0° 로 기준)방향을 추정하여 얻어진 방향 추정정보를 바탕으로 하여 복수 음원분리를 가능하게 하는 2입력 2출력 모델이다. FDBM의 개요는 그림 1과 같다. FDBM은 두 귀 입력신호를 고속 푸리에 변환(Fast Fourier Transform : FFT)하여 두 귀간 위상차(Interaural Phase Difference : IPD), 두 귀간 레벨차(Interaural Level Difference : ILD)[11]를 계산한 후, 미리 구하여 둔 IPD, ILD 데이터베이스와 비교함으로써 음원방향 추정정보를 얻어 이 추정정보를 근거로 음원분리를 한다.

m 개의 목표신호를 $S_m(n)$ ($m=1,2,3,\dots$)이라고 하면 좌우 귀의 입력신호 $l(n)$, $r(n)$ 는 식(1)과 같다.

$$l(n) = l_1(n) + l_2(n) + l_3(n) + \dots = \sum_m s_m(n) * h_{l,m}(n), \quad (1)$$

$$r(n) = r_1(n) + r_2(n) + r_3(n) + \dots = \sum_m s_m(n) * h_{r,m}(n)$$

여기서, $h_{l,m}(n), h_{r,m}(n)$ 은 m 번째 음원의 방위각-양각 조합에 대응하는 좌우 귀의 머리전달함수(Head Related Transfer Functions : HRTFs)이다[12]. 두 귀 입력신호 $l(n), r(n)$ 을 FFT에 의해 스펙트럼 $L(k), R(k)$ 로 변환함으로써 대역분할을 한다. 두 귀 입력신호의 스펙트럼 $L(k), R(k)$ 에 의해 크로스 스펙트럼 $C_{lr}(k)$ 은 식(2)와 같이 나타난다.

$$C_{lr}(k) = L^*(k)R(k) \quad (2)$$

입력신호 $l(n), r(n)$ 의 IPD $\theta_{lr}(k)$ 는 식(3)에 의해 구하여진다.

$$\theta_{lr}(k) = \tan^{-1} \left[\frac{\text{Im}\{C_{lr}(k)\}}{\text{Re}\{C_{lr}(k)\}} \right] \quad (3)$$

그리고 입력신호 $l(n), r(n)$ 의 ILD $\varepsilon_{lr}(k)$ 는 식(2)에서 얻어진 크로스 스펙트럼 $C_{lr}(k)$ 와 파워 스펙트럼 $C_{ll}(k)$ 의 비에 의해 식(4)와 같이 구하여진다.

$$\varepsilon_{lr}(k) = 20 \log \left| \frac{C_{lr}(k)}{C_{ll}(k)} \right| \quad (4)$$

입력신호로부터 얻어진 IPD, ILD를 HRTF에 의해 작성된 IPD, ILD의 데이터베이스[13]와 비교하는 것에 의해 음원방향 추정정보를 얻을 수 있다.

우선, IPD에 의한 음원방향 추정을 보면 식(3)에서 얻은 주파수 성분마다의 IPD $\theta_{lr}(k)$ 와 HRTF에 의해 얻어진 주파수의 두 귀 위상차 정보 데이터베이스(IPD-DOA map) $\theta_{map}(k, \phi, \psi)$ 의 차인 $\Delta\theta(k, \phi, \psi)$ 를 구한다. 여기서 ϕ 는 방위각, ψ 는 양각을 나타낸다. $\Delta\theta(k, \phi, \psi)$ 에 의해 IPD에 근거한 음원방향 추정정보는 식(5)와 같이 얻어진다.

$$D_\theta(k, \phi, \psi) = 1 - \frac{1}{1 + e^{-g(\alpha_1(k)\Delta\theta(k, \phi, \psi) - \varepsilon/2)}} \quad (5)$$

여기서, $\alpha_1(k)$ 는 주파수 의존의 가중함수이며 g, ε

은 분리 필터의 지향성에 대응하는 변수이다.

ILD에 의한 음원방향 추정은 식(4)에서 얻은 ILD를 IPD와 마찬가지로 HRTF를 이용하여 얻어진 주파수마다의 레벨차 정보인 데이터베이스(ILD-DOA map) $\varepsilon_{map}(k, \phi, \psi)$ 와 차분 $\Delta\varepsilon(k, \phi, \psi)$ 를 구하면 식(6)과 같이 ILD에 근거한 음원방향 정보를 얻을 수 있다.

$$D_\varepsilon(k, \phi, \psi) = 1 - \frac{1}{1 + e^{-g(\alpha_2(k)\Delta\varepsilon(k, \phi, \psi) - \varepsilon/2)}} \quad (6)$$

식(5), 식(6)에 의해 구한 주파수마다의 음원방향 추정정보로부터 음원추정을 한다.

음원분리 처리에 대해서는 신호분리 필터 $H_m(k)$ 는 m 번째 음원위치 ϕ'_m, ψ'_m 에서 식(7)과 같이 얻어진다.

$$H_m(k) = D_m(k, \phi'_m, \psi'_m) \quad (7)$$

여기서, $D_m(k, \phi'_m, \psi'_m)$ 은 m 번째 음원에 대한 방향 정보 추정정보이며, m 가 두 개 이상인 경우는 식(8)과 같이 정의한다.

$$D_m(k, \phi, \psi) = \begin{cases} 0, & \text{if } D_{m-1}(k, \phi'_{m-1}, \psi'_{m-1}) \neq 0 \\ D_{m-1}(k, \phi, \psi), & \text{otherwise} \end{cases} \quad (8)$$

분리신호 $\hat{l}_m(t)$ 및 $\hat{r}_m(t)$ 는 입력신호 스펙트럼 $L(k), R(k)$, 신호분리 필터 $H_m(k)$ 에서 역 푸리에 변환을 하면 식(9)와 같이 얻어진다.

$$\begin{aligned} \hat{l}_m(t) &= \text{IFFT}H_m(k) \cdot L(k) \\ \hat{r}_m(t) &= \text{IFFT}H_m(k) \cdot R(k) \end{aligned} \quad (9)$$

그리고 관측신호 스펙트럼 $L(k)$ 및 $R(k)$ 에 대하여 같은 필터를 이용하여 분리처리를 하기 때문에 분리신호의 IPD 및 ILD는 관측신호와 같다. 따라서 분리신호도 공간정보를 보유하고 있다.

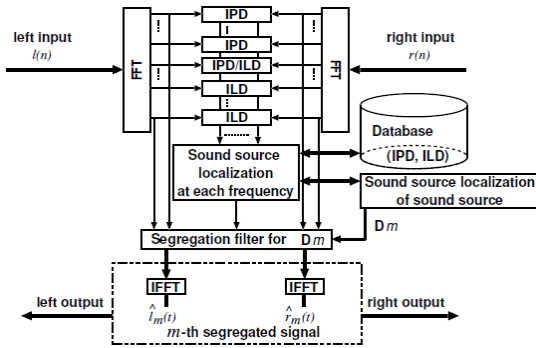


그림 1. FDBM 개요
Fig. 1 FDBM diagram

2.2 신호 대 잡음비

신호 대 잡음비(Signal to Noise Ratio : SNR)는 음원분리 성능의 평가에 널리 사용되는 일반적인 수법이다. 음원분리 후의 SNR을 비교하여 개선량을 구하는 것으로 평가한다. 원하는 신호 $s_t(n)$, 잡음을 포함한 관측신호를 $s_o(n)$, 분리 후의 신호를 $s_p(n)$ 이라고 하면 분리 전후의 SNR은 식(10), 식(11)과 같이 된다.

$$SNR_{input} = 10 \log \left| \frac{E[s_t(n)^2]}{E[s_t(n) - s_o(n)]^2} \right| \tag{10}$$

$$SNR_{output} = 10 \log \left| \frac{E[s_t(n)^2]}{E[s_t(n) - s_p(n)]^2} \right| \tag{11}$$

음원분리 전후의 SNR를 비교하기 위해서는 관측 신호와 분리신호에 포함된 신호성분이 일치할 필요가 있으나 많은 음원분리에 있어서 완전히 일치하는 것은 어렵기 때문에 원하는 신호가 분리신호와 거의 손실이 없다는 가정하에 평가를 한다. 식(11)에서의 잡음 항에서도 알 수 있듯이 원하는 신호성분의 손실과 잡음성분의 잔존을 구별할 수 없어 같은 것으로 평가하고 있다. 그리고 스펙트럼 구조를 고려하지 않는 경우에는 원하는 신호와 분리신호의 시간차나 강도차를 잡음으로 평가하고 있다. 인간이 지각하지 못하는 성분의 왜곡에 대해서도 지각 가능한 왜곡과 같은 잡음으로서 취급하고 있다. 이와 같이 SNR은 신호의 물리적 특성을 측정하는 척도이나 인간의 청각특성 등

을 고려하지 않고 있기 때문에 SNR가 동일 레벨에 있어도 청각상의 인상이 크게 다른 경우가 있으며 평가치도 인간의 인상과는 반드시 일치하지 않는다.

2.3 기여도 함수

기여도 함수는 여러 가지 영역에서 전달계의 동정이나 선형성의 평가 등에 이용되고 있다. 신호 $x(n)$, $y(n)$ 의 기여도 함수 $\gamma_{x,y}^2(k)$ 는 식(12)와 같이 나타낸다.

$$\gamma_{x,y}^2(k) = \frac{|C_{xy}(k)|^2}{C_{xx}(k)C_{yy}(k)} \tag{12}$$

$C_{xy}(k)$ 는 $x(n)$, $y(n)$ 의 크로스 스펙트럼이며, $C_{xx}(k)$ 는 $x(n)$ 의 파워 스펙트럼이다. 기여도 함수는 주파수의 함수로서 널리 사용되는 신호의 유사 척도로 $0 \leq \gamma^2 \leq 1$ 의 범위로 표시되며 $\gamma_{x,y}^2(k) = 1$ 이면 신호 $x(n)$, $y(n)$ 가 일치하고, $\gamma_{x,y}^2(k) = 0.8$ 이면 신호 $y(n)$ 의 성분 80%가 신호 $x(n)$ 에 근거한 성분이며 나머지 20%는 외부잡음 성분이나 비선형 왜곡 성분을 나타낸다. SNR와 비교하면 신호간의 정상적인 시간차나 강도차에 의한 영향은 받기 어려우나 특정의 주파수에 있어서는 SNR에 영향을 받기 쉽다. 그리고 기여도 함수도 청취시의 심리 음향효과를 고려하지 않고 인간이 지각하지 못하는 잡음까지 평가를 하고 있다. 그러므로 기여도 함수도 인간의 청각적인 인상과는 반드시 일치하지 않는다.

2.4 음성품질의 지각평가

SNR, 기여도 함수의 향으로 청각특성을 고려하지 않은 SNR, 기여도 함수의 물리척도에서는 실제 청취자의 인상과 평가가 일치하지 않는 경우가 있다. 인간의 청각에 근거한 심리 음향모델을 고려한 전화대역 객관 음성품질 평가방법인 음성품질의 지각평가(Perceptual Evaluation of Speech Quality : PESQ)를 FDBM의 평가수법의 하나로 채용하고 있다.

지각모델은 IRS 필터를 적용시킨 신호에 대하여 지연에 따라 플레이를 움직이면서 FFT를 한다. 각각의 신호는 인간의 청각특성을 고려한 Bark 척도[14]를 근거로 주파수로 변환시켜 보정처리를 한다. 그 다

음은 같은 인간의 청각특성을 고려한 Sone 척도[15]에 근거하여 라우드네스 스케일로 변환시켜 인간의 지각에 따른 표현으로 나타낸다. 이 신호의 차에서 인간이 청취시 느끼는 왜곡이 구하여진다. 왜곡에 대한 신호손실보다도 가산적인 잡음을 지각하기 쉬운 인간의 비대칭적인 지각특성을 반영하는 처리를 함으로서 비대칭 왜곡을 구할 수 있다. 청취시 느끼는 왜곡과 비대칭 왜곡 중 지각되기 쉬운 참조신호에 있어 무음 구간에 생기는 왜곡을 강조하고 주파수 축으로 통합 함으로서 프레임 왜곡 및 비대칭 프레임 왜곡을 구할 수 있다. 최종적으로 시간축으로 통합한 평균왜곡과 평균 비대칭왜곡을 최대 평가치 4.5에서 감산함으로서 PESQ값을 구할 수 있다[10]. PESQ값은 -0.5~4.5의 범위로 표시되며 참조신호와 평가신호가 같은 음질이라면 4.5를 나타낸다.

III. 두 귀 모델의 음원분리 성능평가

3.1 평가방법에 의한 시뮬레이션

SNR, 기여도 함수, PESQ의 세 가지 객관평가 수법을 이용하여 시뮬레이션을 하였다. 시뮬레이션 조건으로서는 샘플링 주파수 16kHz, 양자화 비트수는 16비트로 하였다. FDBM용 데이터베이스의 도출 및 관측신호의 FFT 처리의 프레임 길이는 512 샘플로 하였다. 프레임마다 해닝창(Hanning window)을 적용하였고 프레임 시프트는 1/4 프레임으로 하였다. 기준이 되는 입력 SNR는 목표음원과 잡음의 파워 비로 정의하고 -30~30dB의 범위에서 10dB마다 변화시켰다. 음성자료는 일본 음향학회 연구용 연속음성 데이터 베이스[16]에 수록된 것을 사용하였다. 1개의 목표음원과 1개 혹은 2개의 잡음원을 사용하였다. 그림 2는 더미헤드와 목표음원, 잡음원의 배치를 나타내었다. 방위각 ϕ 는 정면을 0° , 오른쪽을 +방향으로 하였고 양각 ψ 는 수평면을 0° , 위의 방향을 +로 하였다. 목표음원은 남성 음성으로 제시하고, 방위각과 양각을 0° 로 배치하였다. 잡음원의 종류와 배치는 표 1과 같다.

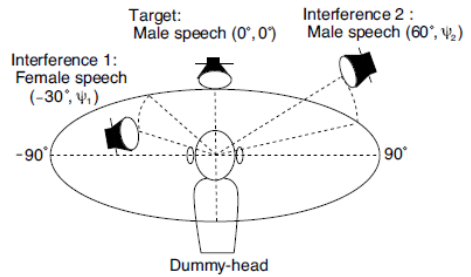


그림 2. 목표음원 및 잡음원의 배치
Fig. 2 Arrangement of a target and noise source

표 1. 잡음원의 종류 및 위치
Table 1. A class and position of noise source

elevation angle	term	noise source	bearing angle, elevation angle
0°	A	female	$-30^\circ, 0^\circ$
	B	male	$60^\circ, 0^\circ$
	C	female male	$-30^\circ, 0^\circ$ $60^\circ, 0^\circ$
$-20^\circ, 20^\circ$	A'	female	$-30^\circ, -20^\circ$
	B'	male	$60^\circ, 20^\circ$
	C'	female male	$-30^\circ, -20^\circ$ $60^\circ, 20^\circ$

3.2 평가 결과 및 고찰

음원을 양각 0° 에 배치한 경우, 그림 3은 조건 A, 그림 4는 조건 B, 그림 5는 조건 C의 결과를 나타내었다. 모든 그림에서 (a)는 입력신호의 SNR인 SNR_{in} 이 좌우로 다르며 입력 SNR도 다르게 나타나고 있다. 이것은 SNR_{in} 이 HRTF 데이터베이스에 등록된 머리 임펄스 응답을 시간영역으로 계산하였을 때 좌우 채널에 SNR로서 정의되고 있기 때문이다. 그림 3의 (a)에서 입력 SNR 5dB부근에서는 $SNR_{in,r}$ 과 $SNR_{out,r}$ 이 교차하고 있다. 이 교차점에 의해 입력 SNR가 낮은 범위에서는 $SNR_{in,r}$ 에 대하여 $SNR_{out,r}$ 의 값이 높고 FDBM 처리 후의 SNR 개선이 보인다. 그러나 교차점 보다 입력 SNR가 높은 범위에서는 $SNR_{in,r}$ 에 대하여 $SNR_{out,r}$ 가 낮고, FDBM 처리에 의해 SNR가 저하한 것을 알 수 있다. 이것은 FDBM 입력신호의 SNR가 충분히 높기 때문에 FDBM 처리에 의해 잡음성분이 생겼다고 생각되어진다. 그리고 왼쪽 채널 신호에서도 같은 경향이 보이거나 오른쪽 채널에 비하면 교차점이 약간 오른쪽으로 이동되고 있

으며 입력 SNR 10dB 부근에서 교차하고 있다. 이것은 잡음원이 왼쪽 전방에 위치하고 있으므로 입력 신호에서는 왼쪽 채널 쪽이 낮은 평가치가 되고 있으나 방향정보에 근거한 음원분리를 한 결과, 좌우 모두 같은 평가치로 개선되었음을 나타내고 있다. 그림 3. (b)의 기여도 함수의 결과에 있어서도 SNR의 결과와 마찬가지로 C_{in_l} 과 C_{out_l} 의 교차점이 보인다. 입력 SNR가 낮은 범위에서의 평가저하와 입력 SNR가 높은 범위에서의 평가 개선도 같다. 그림 3의 (c)에서의 PESQ의 평가에서는 입력 SNR가 좌우 채널 모두 평가가 개선되고 있음을 알 수 있다. 따라서 입력 SNR가 높은 조건에서 SNR과 기여도 함수에 의한 평가에서는 분리처리에 의한 신호 왜곡이 생기는 결과가 나타났으나 같은 입력 SNR에서의 PESQ의 평가에서는 분리 처리에 의해 음질은 개선된 결과가 나타났다. 그림 4에서는 잡음원의 방향, 종류가 그림 3과 다르나 같은 결과가 나타났다. 그림 4의 (c)에서는 입력 SNR가 -30dB인 경우, 입력 평가치 -20dB보다 높게 나타나고 있으나 이것은 PESQ가 정상적인 SNR가 낮게 되는 환경을 가정하지 않았으므로 평가가 적절히 이루어지지 않았다고 생각되어진다. 그림 5에서는 음원 수가 3개이며 좌우에 음원이 배치되어 있기 때문에 입력신호에 있어서 좌우 평가치의 차가 적다. 그림 3, 4에 비하면 개선 정도가 저하되고 있으나 좌우의 신호를 같은 레벨까지 개선하고자 하는 것에 대해서는 같은 결과로 볼 수 있다.

그림 3, 4에 나타나고 있는 입력 SNR가 높게 되면 처리 후의 평가가 저하하는 점에 대해서는 그림 5에서는 현저하게 나타나지 않는다. 음원 수가 증가함에 따라 음원방향 추정의 조건이 까다롭기 때문에 음원 수가 2개의 경우와 비교하면 음원분리 시, 목표 음 이외의 성분을 추출하는 가능성이 높게 된다. 입력 SNR가 낮은 범위에서 잔류잡음이 많으나, 입력 SNR가 충분히 높은 범위에서는 과도한 잡음억제가 생기지 않기 때문에 신호의 왜곡이 적다고 생각되어진다. 표 1에서의 음원 위치를 양각방향으로 변화시킨 경우 (조건 : A', B', C')의 경우에서도 음원분리 성능에 큰 차이가 나타나지 않음을 확인하였다.

이상의 평가결과로부터 FDBM에서는 음원의 배치를 변화시킨 경우에도 현저한 저하는 없으며 같은 성능을 유지하고 있으며, 잡음원 방향의 의존성은 낮음

을 확인하였다. 음원 수가 증가한 경우에도 현저한 성능 저하가 보이지 않았으며 일정한 성능을 가지는 것을 확인하였다. 그리고 입력 SNR가 충분히 높은 조건하에서 SNR나 기여도 함수를 이용한 평가에서는 음질의 저하가 보이나, PESQ에 의한 평가에서는 같은 음질, 혹은 음질의 개선이 나타남에 따라 PESQ에 의한 평가가 인간의 청감에 가까운 것이며 분리음의 평가에 적용 가능함을 나타내었다.

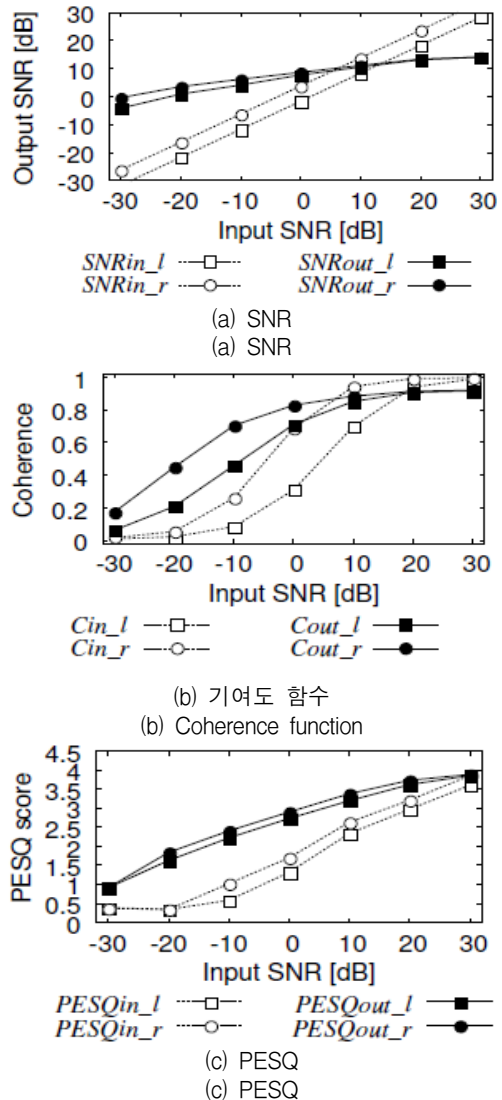


그림 3. SNR, 기여도 함수, PESQ의 평가결과(조건 A)
Fig. 3 Result of SNR, coherence, PESQ(term A)

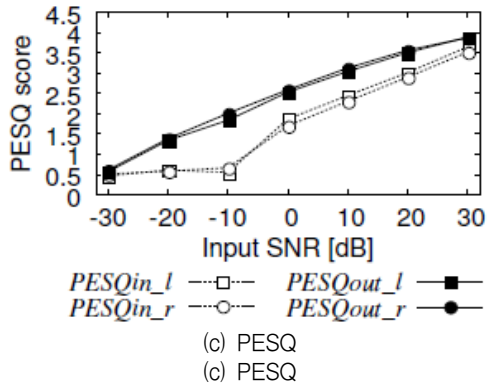
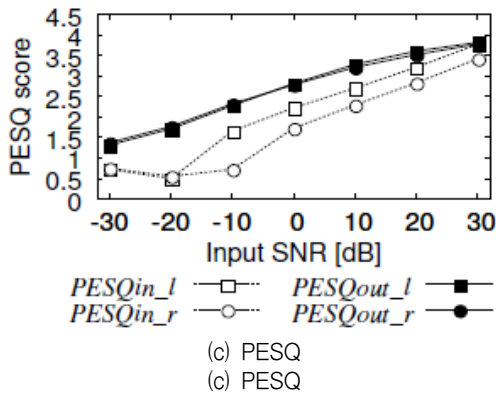
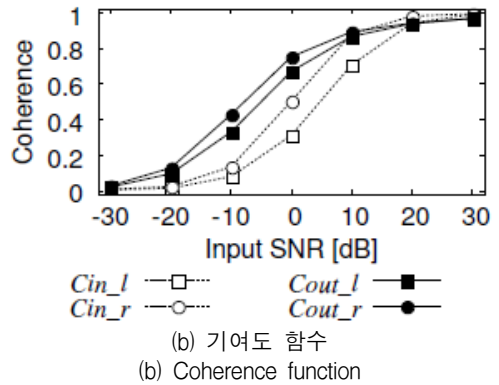
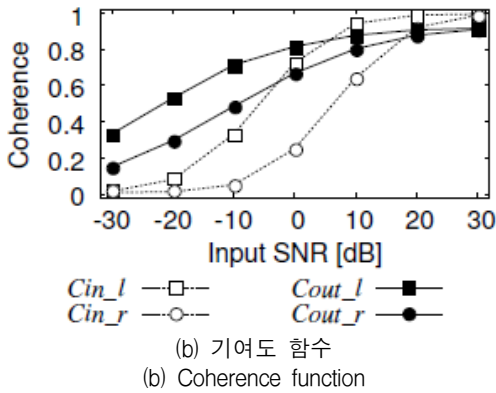
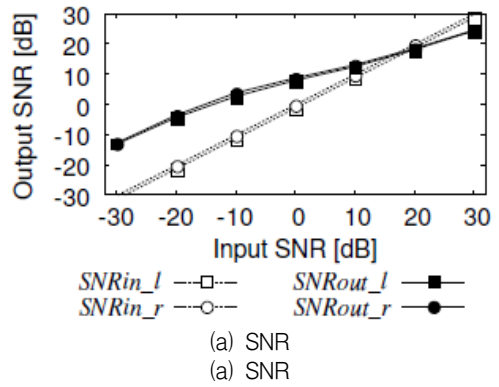
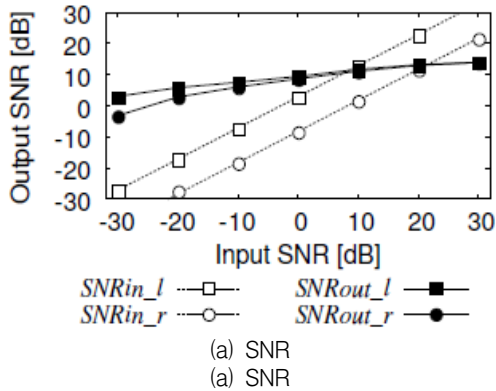


그림 4. SNR, 기여도 함수, PESQ의 평가결과(조건 B)
Fig. 4 Result of SNR, coherence, PESQ(term B)

그림 5. SNR, 기여도 함수, PESQ의 평가결과(조건 C)
Fig. 5 Result of SNR, coherence, PESQ(term C)

IX. 결론

본 논문에서는 주파수 영역 두 귀 모델(Frequency Domain Binaural Model : FDBM)에 의한 음원분리에

있어서 문제가 되고 있는 분리음의 음질 저하를 개선하기 위하여 FDBM의 성능평가를 하여 적절한 평가수법의 검토와 기본적인 특성에 대하여 검토를 하였다. 음질개선을 위한 적절한 성능평가에는 일반적으로 널리 사용되고 있는 방법인 SNR이나 기여도 함수는 물리적도이며 평가치가 인간이 청취한 경우의 인상과 일치하지 않는 문제가 있다. 본 연구에서는 보청 시스템 장착자를 고려한 음질개선을 목적으로 하고 있기 때문에 FDBM의 평가에는 인간의 청감특성이나 물리음향효과를 고려한 수법을 이용할 필요가 있다. 이 조건을 만족하는 전화 대역의 음성의 객관 음질 평가방법으로서 PESQ를 평가에 사용하였다. 음질의 문제 해결을 위하여 FDBM의 음원분리 성능평가를 하였으며 SNR, 기여도 함수, PESQ의 세 가지 방법을 이용하여 기본 특성에 대하여 확인하였다. 모든 결과가 FDBM에 의해 좌우 채널간 평가치의 차가 작게 되었고 좌우 채널이 거의 같은 레벨까지 개선되었음을 확인하였다. 그리고 음원방향을 바꾸고 음원 수를 증가시킨 경우에도 일정한 개선이 보였다. SNR과 기여도 함수, PESQ의 결과를 비교하면 PESQ의 평가에서는 입력 SNR를 변동시킨 경우에도 거의 모든 조건에서 분리에 의한 평가가 개선되었다. SNR와 기여도 함수는 입력 SNR가 높게 되면 분리음원에 의해 저하하는 결과가 나타났다. 두 귀 보청시스템의 음질개선 목적으로 FDBM의 음원분리 성능평가를 한 결과, PESQ를 이용한 평가가 SNR이나 기여도 함수를 이용한 평가보다 적절하다는 가능성을 시사하였다.

감사의 글

이 논문은 2012년도 동서대학교 "Dongseo Frontier Project" 지원에 이루어진 것이며, 데이터 정리를 도와 준 Minako Tomita 씨에게 감사를 표한다.

참고 문헌

- [1] Tsuyoshi Usagawa, Hirokazu Shimada, Yoshiaki Sawada, Yoshifumi Chisaki and Masanao Ebata, "A microphone array system using iterative echo suppression method as inverse filtering", *Acoustical Science & Technology*, Vol. 22, No. 4, pp. 315 - 317, 2001.
- [2] Shoji Makino, Shoko Araki, Ryo Mukai, Hiroshi Sawada and Hiroshi Saruwatari, "ICA-based blind source separation of sounds", *Proceedings of the Japan-China Joint Conference on Acoustics 2002*, pp. 83 - 86, 2002.
- [3] Tomoya Takatani, Tsuyoki Nishikawa, Hiroshi Saruwatari, "Blind Source Separation based on Binaural ICA", *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Vol. 5, pp. 321 - 324, 200
- [4] Hidetoshi Nakashima, Yoshifumi Chisaki, Tsuyoshi Usagawa and Masanao Ebata, "Frequency domain binaural model based on interaural phase and level difference", *Acoustical Science & Technology*, Vol. 24, No. 4, pp. 172 - 178, 2003.
- [5] Yoshifumi Chisaki, Kotaro Matsuo and Tsuyoshi Usagawa, "Howling canceler using interaural level difference for binaural hearing assistant system", *Acoustical Science & Technology*, Vol. 28, No. 2, pp. 90 - 97, 2007.
- [6] Markus Bodden, "Modeling human sound-source localization and the cocktailparty-effect", *Acta Acoustica*, Vol. 1, pp. 43 - 45, 1993.
- [7] Takashi Nakanishi, Norifumi Sato, Hidetoshi Nakashima, Yoshifumi Chisaki, Tsuyoshi Usagawa and Masanao Ebata, "Sound Source Segregation under reverberant condition using Frequency Domain Binaural Model", *Proceedings of Kyushu-Youngnam Joint Conference on Acoustics 2003*, pp. 129 - 132, 2003.
- [8] Tsuyoshi Usagawa, Rika Matsuo, Takashi Nakanishi, Hidetoshi Nakashima and Yoshifumi Chisaki, "Concurrent Speech Segregation based on DOA Information using Frequency Domain Binaural Model -An application for hearing aid-", *Proceedings of International Congress on Acoustics 2004*, Vol. 5, pp. 3655 - 3658, 2004.
- [9] Chai-bong Lee, "The effect of leading tone and following tone with single frequency on sound lateralization", *The Journal of the Korea Institute of Electronic Communication Sciences*, Vol. 5, No. 3, pp. 251-255, 2010.
- [10] ITU-T Recommendation, "Perceptual evalua-

- tion of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs", p. 862, 2001.
- [11] Chai-bong Lee, "The effect of a temporal masking on the sound laterlization", The Journal of the Korea Institute of Electronic Communication Sciences, Vol. 5, No. 4, pp. 352-356, 2010.
- [12] Chai-bong Lee, "A study on the simplification of HRTF within low frequency region," The Journal of the Korea Institute of Electronic Communication Sciences, Vol. 5, No. 6, pp. 581-587, 2010.
- [13] Bill Gardner and Keith martin, "HRTF measurements to a KEMAR dummy head microphone," MIT Media lab Perceptual Computing Technical Report#280, 1994.
- [14] Eberhard Zwicker, "Subdivision od the audible frequendy rang into critical bands", Journal of the Acoustical Society of America, Vol. 33, No. 2, pp. 248, 1961.
- [15] Eberhard Zwicker, Hugo Fastl, Psychoacoustics : Facts and Models, Spring-Verlag, Berlin, 1990.
- [16] The Acoustical Society of Japan, "A serial speech data base for research purpose", The Journal of the acoustical society of Japan, Vol. 48, No. 12, pp. 888-893, 1992.

저자 소개



이채봉(Chai-Bong Lee)

1985년 2월 동아대학교 전자공학과 졸업 (공학사)

1988년 3월 동북대학교 대학원 전기통신공학과 졸업(공학석사)

1992년 3월 동북대학교 대학원 전기통신공학과 졸업(공학박사)

1993~현재 동서대학교 전자공학과 교수

※ 관심분야 : 신호처리, 음향공학

