

Spectral Subtraction과 Two Channel Beamformer를 이용한 음성 강조 기법

Speech Enhancement using Spectral Subtraction and Two Channel Beamformer

김 학 윤*
(Hack Yoon Kim*)

요약

본 연구에서는 단일 채널 단구간 진폭 스펙트럼 추정 기법의 하나인 Spectral Subtraction 방법과 2 채널 Griffiths-Jim Beamformer를 결합한 음성 강조 기법을 제안한다. 기존의 단구간 진폭 스펙트럼 추정 기법에서는 관측된 신호의 스펙트럼에서 잡음의 평균 스펙트럼을 감산하여 잡음을 제거하고 있지만, 이 방법을 이용하여 잡음을 제거 할 경우에는 잡음 변동 시 잡음 억제 능력이 미약하고, 목적 신호의 단구간 진폭 스펙트럼 추정 성능이 낮아진다는 단점을 갖고 있다. 그 이유는 실제 잡음의 스펙트럼은 평균값 주위에 분산되어 있기 때문이다. 그러므로, 2 채널 Beamformer의 사각(Blocking Matrix)를 이용하여 분석 구간에서의 잡음의 단구간 진폭 스펙트럼을 추정하고, 이 추정된 값을 이용하여 목적 신호의 스펙트럼을 추정하는 기법을 제안하고, 컴퓨터 시뮬레이션을 통하여 그 유효성을 입증한다.

ABSTRACT

In this paper, a new spectral subtraction technique with two microphone inputs is proposed. In conventional spectral subtraction using a single microphone, the averaged noise spectrum is subtracted from the observed short-time input spectrum. This results in reduction of mean value of noise spectrum only, the component varying around the mean value remaining intact. In the method proposed in this paper, the short-time noise spectrum excluding the speech component is estimated by introducing the blocking matrix used in Griffiths-Jim-type adaptive beamformer with two microphone inputs, combined with the spectral compensation technique. A simulation was conducted to verify the effectiveness of the method.

I. 서론

보청기 사용자가 환경에 잡음이 존재 할 경우에 음성을 청취하게 되면, 음성의 품질이 떨어지고, 명료도가 저하되기 때문에 사용자가 피로를 호소하는 경우가 자주 있다. 이와 같은 문제점을 해결하기 위하여, 지금까지 여러 가지의 음성강조(Speech Enhancement) 기법이 연구되어 왔다[1-8].

음성강조기법은 시스템 입력 수에 따라 단일 채널 신호를 이용하는 방법[1-5]과 멀티 채널 신호를 이용하는 방법[6-7,9]으로 구별할 수 있다.

단일 채널의 신호를 이용하여 음성을 강조하는 기법은, 소규모의 하드웨어로 구성이 가능한 것이 많다. 그중

단일 채널의 신호를 이용하고 있는 대표적 예인 S.S.(Spectral Subtraction)방법[1]은 실시간 시스템에 종종 응용되고 있다. 그러나, 단일 채널의 신호를 이용하여 음성을 강조하는 기법은 입력신호에서 얻어지는 정보가 극히 한정되어 있으므로, 예를 들면, 잡음은 정상 혹은 준 정상일 것이라는 등의 가정이 필요하게 된다[4].

그 반면, 멀티 채널의 신호를 이용하여 음성을 강조 할 경우에는, 위에서 언급한 제약 조건이 훨씬 적어져, 보다 다양한 잡음에 대처 가능하게 되며, 잡음 억제 능력도 일 반적인 단일 채널 방법보다 훨씬 좋다고 알려져 있다[3]. 그러나, 그 단점으로서 는 보다 대규모의 하드웨어가 필요 하게 되고, 연산량도 많이 요구되어 진다. 특히, 디지털 보 청기에의 응용을 생각할 경우, 마이크로폰의 개수는 많아 도 두개 정도가 제일 적당하고, 또한 실시간 처리, 시스템 의 휴대성등의 관점으로부터 연산량도 적은 것이 요구되 어진다.

*충주대학교 이공대학 전자·정보통신·반도체 공학부
접수일자:1998년 9월 1일

그러므로, 본 연구에서는 단일채널에 의한 음성 강조 기법인 S.S.법과 멀티 채널의 Beamformer의 장점을 조합하여, 디지털 보청기에의 응용을 고려하여 효과적으로 잡음을 억제할 수 있는 방법을 제안하고자 한다.

II. 기존의 음성강조법

제안하는 방법을 논하기 전에, 본 연구의 기저가 되는 단일 채널의 음성 강조 기법인 S.S.법 및 멀티 채널 음성 강조 기법중의 하나인 2 채널 Griffiths-Jim Beamformer에 관하여 간단히 서술한다.

II-1. S.S.법

이 방법은 Wiener법과 마찬가지로 목적신호의 단구간 진폭 스펙트럼 추정법의 한 분야로서, 잡음이 부가된 입력 신호의 스펙트럼으로부터 잡음신호의 스펙트럼을 감산함에 의하여 목적신호의 스펙트럼을 추정하는 기법이다.

음성신호를 $s(n)$, 잡음을 $d(n)$ 이라 하면, 관측된 신호 $y(n)$ 은 다음 식과 같이 표현된다.

$$y(n) = s(n) + d(n) \quad (1)$$

이 신호를 시간축상에서 짧은 구간으로 분할하여, 시간창을 걸고 단구간 푸리에 변환을 하면, 입력 신호의 단구간 파워 스펙트럼 $|Y(k)|^2$ 을 구할 수 있다. 여기서, 목적신호(음성신호) $s(n)$ 과 잡음 신호 $d(n)$ 이 서로 무상관이고, 또한 잡음원이 정상이라고 가정하면, 종래의 S.S.법에서는 다음 식과 같이 목적신호의 단구간 파워 스펙트럼의 추정값 $|\hat{S}(k)|^2$ 을 관측신호의 파워 스펙트럼으로부터 잡음의 평균 파워 스펙트럼 $E\{D(k)|^2\}$ 을 감산하여 구한다.

$$|\hat{S}(k)|^2 = |Y(k)|^2 - E\{D(k)|^2\} \quad (2)$$

여기서 $E\{\cdot\}$ 는 기대치를 나타낸다.

위 식으로부터 목적신호의 단구간 진폭 스펙트럼의 추

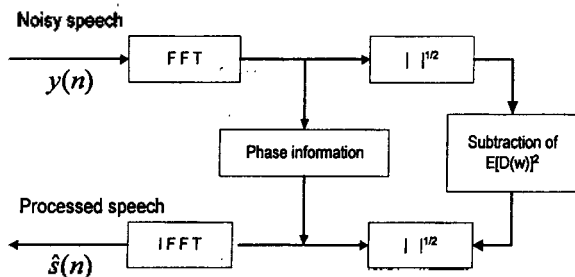


그림 1. S.S.법의 시스템 Block Diagram
Fig. 1. A system block diagram of spectral subtraction.

정값 $|S(k)|$ 를 구하고, 그 값에 입력신호의 위상을 넣은 다음, 역 푸리에 변환을 행하게 되면, 원하는 목적신호를 추정하게 된다. 그림 1에 S.S.법의 시스템 Block Diagram을 나타낸다.

II-2. 2 채널 Griffiths-Jim Beamformer

Beamformer는 적응형과 비적응형으로 구분할 수 있고, Griffiths-Jim beamformer는 적응형 Beamformer의 대표적인 예이다. 그림 2에 2 채널일 경우의 Griffiths-Jim Beamformer의 Block Diagram을 보인다.

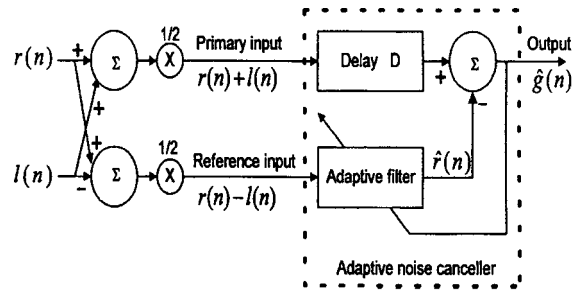


그림 2. 2 channel Griffiths-Jim Beamformer의 Block Diagram
Fig. 2. A block diagram of two-channel Griffiths-Jim beamformer.

그림 2에서와 같이 목적신호원이 두개의 마이크로폰을 잇는 선분에 대해 수직 방향에 있을 경우(디지털 보청기 착용자가 소리나는 방향으로 향하고 있을 경우를 상정함), 목적신호 $s(n)$ 은 두 개의 마이크로폰에 동시에 도달한다. 한편, 한개의 잡음원 $d(n)$ 이 그 이외의 방향에 위치하고 있다고 가정하면, 잡음은 약간의 시간차를 갖고 도달하게 된다. 그러면, 두개의 마이크로폰에서 얻어지는 신호는 다음 식과 같이 표현된다.

$$\begin{aligned} r(n) &= s(n) + d'(n) \\ l(n) &= s(n) + d''(n) \end{aligned} \quad (3)$$

여기서 $d'(n)$ 및 $d''(n)$ 은 잡음원 $d(n)$ 의 지연을 나타내고, 기호 $[\cdot], [\cdot]$ 는 지연시간차를 나타낸다.

(3)식에서 $r(n)$ 과 $l(n)$ 을 역상으로 합하면, 목적 음원은 동상이므로 서로 상쇄되고, 잡음신호만을 관측할 수 있으므로, 이것을 보조 입력으로 하고, $r(n)$ 과 $l(n)$ 은 동 위상으로 합한 신호를 주입력으로 하여, 적응 필터를 구성한다. 적응 필터의 출력에서는 주 입력원으로부터 보조 입력과 상관 관계가 있는 성분만이 제거되어 원하는 목적신호를 출력으로 얻게 된다.

II-3. 기존의 방법의 문제점

단일 채널을 이용하여 단구간 진폭 스펙트럼을 추정하는 방법(S.S.법)에서는, 잡음의 진폭 스펙트럼의 기대값

$E[|D(k)|^2]$ 을 음성이 존재하지 않는 구간에서 구한 스펙트럼의 평균값을 기대값으로 대응하고 있지만[3], 이와 같은 방법은 비 음성 구간의 검출이 필요하게 되고, 또한, 잡음이 변동할 때 목적신호의 추정 성능이 크게 떨어진다는 단점을 갖고 있다. 그러므로 잡음의 평균 파워스펙트럼이 아닌, 현재 처리를 행하고 있는 구간에서의 잡음의 파워스펙트럼을 추정하여, 그 추정 값을 이용하게 된다면, 위에서 언급한 문제점이 해결될 뿐만 아니라 잡음의 억제 능력도 향상되리라 기대된다.

반면, 2 채널 Griffiths-Jim Beamformer 방법에서는, 잡음의 단구간 진폭 스펙트럼 값뿐만 아니라, 위상까지 포함된 단구간 복소 스펙트럼을 추정하여, 이것을 마이크로폰 입력의 복소 스펙트럼에서 감산하는 것과 동가이며, 마이크로폰의 수가 잡음원보다 많을 경우에는 상당히 큰 잡음억제 능력을 갖고 있지만, 잡음원이 마이크로폰의 수보다 많이 존재할 경우에는, 잡음 억제능력이 크게 떨어진다고 알려져 있다[6]. 즉, 그림 2에서 적용 필터는, 잡음의 위상과 진폭을 추정하고 있는 것과 같은 역할을 하고 있으며, 원리적으로는 한 개의 적용 필터는 한 개의 잡음원의 위상과 진폭만이 조정 가능하므로, 잡음이 복수일 경우에는 마이크로폰과 적용 필터를 동시에 증가시키지 않으면 안 된다. 이 때문에 잡음원이 마이크로폰 수보다 많을 경우에는 각각의 잡음원에 대한 그 진폭과 위상을 엄밀히 추정할 필요가 있다는 점이 잡음원이 다수 존재할 경우의 Griffiths-Jim Beamformer의 문제점이라 할 수 있다.

III. 2 채널 신호를 이용한 잡음의 단구간 진폭 스펙트럼 추정

여기에서는 앞 절에서의 검토를 토대로, 그림 3과 같이 공간에 한 개의 목적신호원과 잡음원이 한 개 혹은 복수 개(본 연구에서는 두 개) 존재할 경우, 2 채널의 입력신호에서 잡음의 단구간 진폭 스펙트럼을 추정하는 방법에 관하여 논하고, 또한 추정되어진 값을 이용하여 목적신호의 단구간 진폭 스펙트럼을 추정하는 기법을 제안한다.

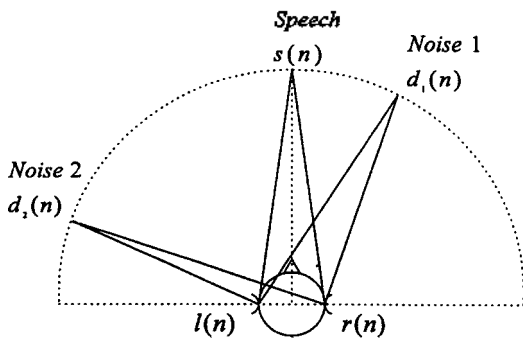


그림 3. 목적신호원과 잡음원 및 마이크로폰 배치도
Fig. 3. Broadside array orientation.

III-1. 잡음원이 하나일 경우

우선, 잡음원이 하나 존재할 경우에 대하여 검토한다.

목적신호가 두 개의 마이크로폰을 잇는 수직면상에 있고, 잡음원이 그 외의 방향에 위치하여 있을 때(그림 3에서 Noise1 만이 존재한다고 가정), 마이크로폰에 입력되는 신호는 다음 식과 같이 나타낼 수 있다.

$$\begin{aligned} r(n) &= s(n) + d_1'(n) \\ l(n) &= s(n) + d_1''(n) \end{aligned} \tag{4}$$

위 식에서, $d_1'(n)$, $d_1''(n)$ 은 잡음 $d_1(n)$ 의 지연을 나타내고, ' ' , ' ' '는 앞 절에서와 같이 지연 시간의 차를 의미한다.

여기서, 2 채널 Griffiths-Jim Beamformer와 같은 방법으로 두 개의 마이크로폰에서 얻어진 신호의 차를 구하면, 목적신호인 음성 신호는 서로 상쇄되어 없어지게 되고 다음 식과 같이 잡음 성분만 남게 된다.

$$r(n) - l(n) = d_1'(n) - d_1''(n) \tag{5}$$

위 식에서, $d_1'(n)$ 에 대하여, $d_1''(n)$ 은 지연 성분을 갖고 있으므로, 차신호의 진폭스펙트럼에는 지연 간격에 대응한 스펙트럼에 Dip이 발생하게 된다. 그러므로 차신호의 단구간 푸리에 변환 값을 그대로 잡음의 단구간 진폭 스펙트럼 추정 값으로 이용하게 되면 지연 성분에 의해 생긴 스펙트럼의 Dip으로 인하여 추정 성능이 저하되게 된다.

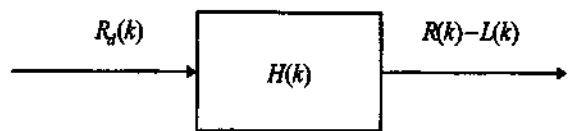


그림 4. 가상적인 전달함수
Fig. 4. Transfer function defined in this paper.

이와 같이, 지연에 의해 생긴 Dip을 보상하기 위해 그림 4와 같은 가상적인 전달함수 $H(k)$ 를 고려하고, 이를 식으로 표시하여 보면, 다음 식과 같이 나타낼 수 있다.

$$R(k) - L(k) = H(k)R_q(k) \tag{6}$$

여기서, $R(k)$ 및 $L(k)$ 는, 각각 $r(n)$ 및 $l(n)$ 의 단구간 푸리에 변환을 나타내며, $R(k)-L(k)$ 는 다음 식과 같이 표시되는 차신호의 단구간 푸리에 변환을 나타낸다.

$$R(k) - L(k) = D_1'(k) - D_1''(k) \tag{7}$$

식(7)에서 $D_1'(k)$ 및 $D_2'(k)$ 는 각각 $d_1'(n)$ 과 $d_2'(n)$ 의 단구간 푸리에 변환 값을 나타내고, 또한 $R_s(k)$ 는 한쪽 마이크로폰에 입력되는 잡음만의 스펙트럼을 나타내고 있으며, 이는 미지의 값이다. 잡음원이 하나일 경우, 한쪽 마이크로폰에서의 잡음만의 진폭스펙트럼 $R_s(k)$ 는

$$R_s(k) = D_1'(k) \quad (8)$$

이므로,

$$D_1'(k) = D_1(k)e^{-j2\pi k/N} \quad (9)$$

이다. 식(9)에서, τ 는 이산 시간차를 나타내며, N 은 이산 푸리에 변환의 Point를 나타낸다. 따라서, 잡음원이 하나일 경우의 전달함수 $H(k)$ 는 다음 식과 같이 된다.

$$H(k) = 1 - e^{-j2\pi k/N} \quad (10)$$

위 식과 같이 잡음이 하나일 경우의 전달함수 $H(k)$ 의 추정은, 지연시간을 추정하는 문제로 귀착하게 된다. 따라서, 지연 시간의 추정은 자기 상관 함수등에 의하여 추정할 수 있으므로 구한 전달함수를 이용하여, 잡음 성분만의 단구간 진폭 스펙트럼은 다음 식과 같이 구하여진다.

$$|\hat{R}_s(k)| = \sqrt{|R(k) - L(k)|^2 / |H(k)|^2} \quad (11)$$

III-2. 잡음원이 복수일 경우

그림 3에서와 같이, 목적음원은 두개의 마이크로폰을 잇는 선분의 수직면상에 있고 잡음원이 그 이외의 방향에 복수 개(본 연구에서는 두개) 존재 할 경우, 두개의 마이크로폰에서 얻어지는 신호는 다음 식과 같이 된다.

$$\begin{aligned} r(n) &= s(n) + d_1'(n) + d_2'(n) \\ l(n) &= s(n) + d_1''(n) + d_2''(n) \end{aligned} \quad (12)$$

앞 절과 같은 방법으로 두개의 마이크로폰의 차신호는

$$r(n) - l(n) = d_1'(n) - d_1''(n) + d_2'(n) - d_2''(n) \quad (13)$$

으로 되고, 목적신호인 음성신호는 제거되고, 잡음 신호 성분만이 남게된다. 앞 절에서 논한바와 같이, 식(13)의 차신호의 진폭 스펙트럼은 복수의 잡음과 그 각각의 지연 성분이 포함되어 있으므로, 잡음원이 하나일 경우와 마찬가지로 잡음원이 두개일 경우도 또한, 스펙트럼에 Dip이 생기게 된다. 여기서, 잡음이 하나일 경우와 마찬가지로

식(6)과 같은 전달함수를 생각해 보자. 잡음원이 두개일 경우, $R(k) - L(k)$ 와 $Rd(k)$ 는 다음 식과 같이 된다.

$$R(k) - L(k) = D_1'(k) - D_1''(k) + D_2'(k) - D_2''(k) \quad (14)$$

$$R_d(k) = D_1'(k) + D_2'(k) \quad (15)$$

위 식에서, $D_1'(k)$, $D_1''(k)$, $D_2'(k)$, $D_2''(k)$ 는, 각각 $d_1'(n)$, $d_1''(n)$, $d_2'(n)$, $d_2''(n)$ 의 단구간 푸리에 변환을 나타낸다. 식(15)에서 한쪽 마이크로폰에서의 잡음만의 푸리에 변환 값인 $R_d(k)$ 가 미지수이기 때문에, 잡음원이 하나일 경우와 같이, 전달함수 $H(k)$ 를 직접 추정하는 것은 불가능하다. 또한, 잡음원이 복수이므로 잡음원이 하나일 경우와 같이 단순한 지연 추정에 의하여 전달함수를 동정하는 것은 불가능하다. 그러나, 후술하는 바와 같이 본 연구에서 제안하는 음성 강조 방법에 필요한 것은 식(11)에서 나타낸 것과 같이, $|H(k)|^2$ 인 전달함수의 파워스펙트럼이다. 그러므로, 여기서 전달함수 $H(k)$ 를 선형 시불변이라고 가정하고, 식(6)의 양변을 제곱하고 기대값을 취하면, 다음 식과 같이 나타낼 수 있다.

$$E[|R(k) - L(k)|^2] = |H(k)|^2 E[|R_d(k)|^2] \quad (16)$$

또한, 잡음 $d_1(n)$ 과 $d_2(n)$ 이 서로 무상관이라 가정하면,

$$E[|R_d(k)|^2] = |D_1'(k)|^2 + |D_2'(k)|^2 \quad (17)$$

으로 된다.

더욱이, 잡음원 $d_1(n)$ 과 $d_2(n)$ 을 백색잡음으로 하면, 좌변은 정수가 되고, 이 정수를 σ^2 이라고 놓으면,

$$|H(k)|^2 = (1/\sigma^2) E[|R(k) - L(k)|^2] \quad (18)$$

으로 된다. 그러므로, 잡음원이 하나일 경우와 마찬가지로 관측신호의 차신호 스펙트럼에서 직접 추정할 수 있게 된다.

III-3. 제안하는 음성강조법

앞 절에서는 잡음의 단구간 진폭 스펙트럼을 추정하는 기법에 관하여 논하였다. 여기에서는 추정된 잡음의 단구간 진폭 스펙트럼을 이용하여 목적신호(음성신호)를 추정하는 기법에 대하여 논한다. S.S.법에서는 식(2)와 같이 입력신호의 단구간 파워 스펙트럼으로부터 잡음의 평균 파워 스펙트럼을 감산함으로써 목적신호의 파워 스펙트럼을 추정하고 있지만, 본 연구에서 제안한 방법은 다음 식과 같이 입력신호의 파워 스펙트럼에서 앞 절에서 구한

잡음의 단구간 파워 스펙트럼을 감산함으로써 목적신호의 단구간 파워 스펙트럼을 추정하는 것이다.

$$|\hat{S}(k)|^2 = R(k) + L(k)^2 - |\hat{R}_d(k)|^2 \quad (19)$$

본 연구에서 제안한 방법의 Block Diagram을 그림 5에 보인다.

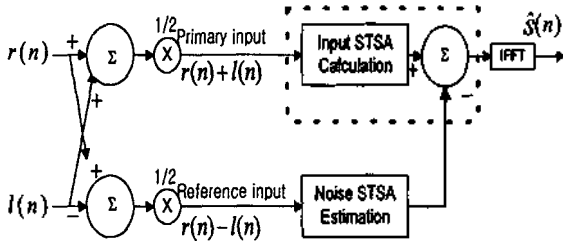


그림 5. 제안한 음성강조법의 Block Diagram (STSA : Short-Time Spectral Amplitude)
 Fig. 5. A block diagram of the proposed speech enhancement method(STSA : Short-Time Spectral Amplitude).

그림 5에 제시한 바와 같이, 본 연구에서 제안한 방법은 앞단은 2 채널 Griffiths-Jim Beamformer의 방법을 이용하고 뒷단은 S.S.법을 연결시켜 구성된다. 즉, 앞단의 Griffiths-Jim Beamformer에서 두개의 마이크로폰의 차신호를 이용하여 잡음의 단구간 진폭 스펙트럼을 구한 후, 얻어진 전달함수를 이용하여 차신호의 스펙트럼에 생긴 Dip을 보상하고, S.S.법을 사용하여 주입력의 파워 스펙트럼에서 감산함으로써 목적신호의 파워 스펙트럼을 추정한 다음, 입력신호의 위상을 부가함으로써 추정된 음성신호를 구하는 기법이다.

IV. 컴퓨터 시뮬레이션

본 장에서는 잡음원이 하나 및 두개인 경우에 대하여 본 연구에서 제안한 방법의 유효성을 다른 기존의 방법들과 비교하면서 검토한다.

IV-1. 시뮬레이션 조건

목적 음원으로 이용한 음성 데이터는 표본화 주파수가 20kHz인 단 문장을 이용하였으며, 잡음은 서로 무상관한 백색 잡음을 사용하였다. 그림 3에 본 연구에서 이용한 목적음원 및 잡음원이 위치한 배치도를 나타내었다. 단구간 푸리에 변환 및 음성강조 처리에 이용한 시간창은 256 Point의 Hamming 창을 사용하였고, 인접하는 프레임과 프레임간의 처리는 128 Point씩 중복하여 처리하고 있다. 또한, SNR은 0dB로 하였다.

IV-2. 잡음원이 하나일 경우의 시뮬레이션 결과

우선, 잡음원이 하나일 경우에 관하여 검토한다. 잡음원의 놓인 위치는 그림 3의 d(n)만이 존재하였을 경우를 가정하였다. 본 연구에서 이용되는 식(6)의 전달함수 H(k)는 두개의 마이크로폰의 차신호 r(n)-l(n)의 자기상관 함수를 구하여 추정하였다.

그림 6은 기존의 S.S.법과 제안한 방법에 의해 음성강조를 행한 결과를 2048 point 구간에 의한 평균 파워 스펙트럼을 보인 것이다. 그림 6에서 그림의 맨 위로부터 음성에 잡음이 부가된 경우의 스펙트럼을, 기존의 S.S.법에 의하여 강조된 스펙트럼을, 본 연구에서 제안한 방법에 의하여 강조된 신호의 스펙트럼을, 맨 아래는 원음성의 스펙트럼을 나타내고 있다. 이 결과를 보면 기존의 S.S.법과 제안한 방법에 의한 음성강조 효과를 비교해 보면, 제안한 방법의 잡음억제 능력이 기존의 방법보다 약 9dB 정도 향상되었음을 알 수 있다. 그 이유는, 잡음의 단구간 평균 진폭 스펙트럼 대신에 분석 구간에 의한 잡음의 단구간 진폭 스펙트럼을 추정하여 그 값을 이용하여 잡음을 억제하였기 때문에 목적음원의 단구간 진폭 스펙트럼의 추정 성능이 향상되었다고 본다.

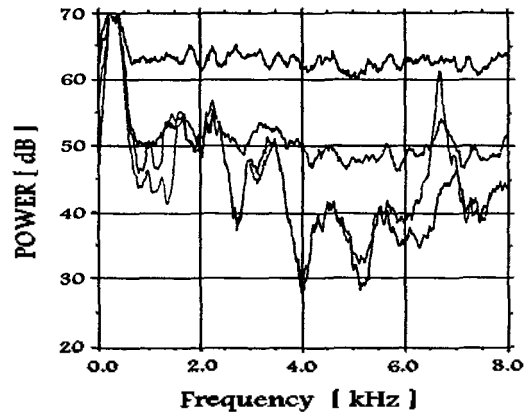


그림 6. 잡음원이 한 개일 경우의 시뮬레이션 결과(SNR 0 dB) : 위로부터 잡음이 부가된 신호, 기존의 S.S.법의 경우, 제안한 방법을 이용하였을 경우 및 원음성의 파워 스펙트럼

Fig. 6. Results of simulation for a single noise source for SNR 0 dB: from the top curve, power spectrum of noisy speech, enhanced power spectrum by conventional spectral subtraction, enhanced power spectrum by the proposed method, and power spectrum of the original speech.

그림 7은 2 채널 Griffiths-Jim Beamformer에 의한 방법과 제안한 방법에 의하여 음성 강조가 수행된 결과를 비교한 것이다. 이 그림을 보면 Beamformer에 의한 방법이 제안한 방법보다 잡음 억제 능력이 약 10dB 정도 높고, 원음성의 스펙트럼의 포락선과 Beamformer에 의하여 강조된 신호의 포락선을 비교하여 보면, 사카이 형성된 6.7kHz 이외의 주파수 지역은 거의 원음성의 스펙트럼의 포락선과 거의

의 유사하게 추정되었음을 알 수 있다. 이것은 앞에서 언급한 바와 같이, Beamformer에서는 잡음의 단구간 진폭 스펙트럼만이 아닌 위상까지도 추정하여 그 값을 이용하기 때문에, 잡음원의 수가 마이크로폰 수보다 작다는 조건하에서는 진폭 스펙트럼 값만을 사용하는 본 기법보다 좋은 결과가 얻어졌다고 본다.

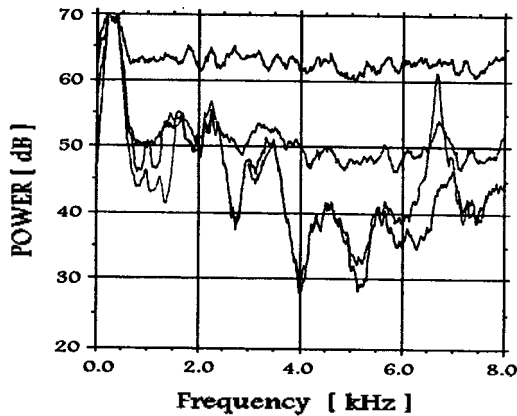


그림 7. 잡음원이 한 개일 경우의 시뮬레이션 결과(SNR 0 dB) : 위로부터, 잡음이 부가된 음성, 제안한 방법을 이용하였을 경우, 2 채널 Beamformer의 경우 및 원음성의 파워 스펙트럼

Fig. 7. Results of simulation for a single noise source for the SNR 0 dB : from the top curve, power spectrum of noisy speech, enhanced power spectrum by the proposed method, enhanced power spectrum by two-channel Griffiths-Jim beamformer, and power spectrum of the original speech.

IV-3. 잡음원이 두개일 경우의 시뮬레이션 결과

그림 8은 잡음원이 두개일 경우(그림 3에서 잡음원 $d_1(n)$ 과 $d_2(n)$ 가 모두 존재하였을 경우)에 관하여 제안한 방법과 2 채널 Beamformer에 의한 결과를 비교한 것이다. 또한, 식(6)의 전달함수 $H(k)$ 는 식(18)을 이용하여 구한 함수를 이용하였다. 그 결과를 보면 2 채널 Beamformer에 의하여 강조된 스펙트럼은, 잡음원의 수가 마이크로폰 수보다 적어야 한다는 조건을 만족시키고 있지 않기 때문에, 잡음원이 하나일 경우(그림 7 참조)와 비교하여 보면 잡음 억제 능력이 크게 떨어지고 있음을 알 수 있다. 그렇지만, 본 연구에서 제안한 방법도 잡음원이 복수일 경우의 전달함수가 잡음원이 하나일 경우의 전달함수와 달리 최소 자승적인 의미로만 존재하기 때문에 전달함수를 정확히 추정하는 것은 불가능하므로, 잡음원이 하나일 경우의 잡음 억제 능력과 비교하여 보면 잡음 억제 능력이 저하되었음을 알 수 있다.

그 반면, 본 연구에서 제안한 방법의 잡음 억제 능력과 2 채널 Beamformer의 잡음 억제 능력을 비교하여 보면, 본 연구에서 제안한 방법의 잡음 억제 능력이 2 채널 Beamformer의 방법보다 약 5 dB 정도 향상되어 있음을 볼

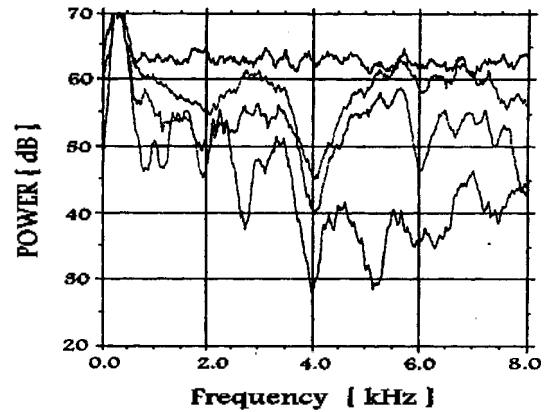


그림 8. 잡음원이 두 개일 경우의 시뮬레이션 결과(SNR 0 dB) : 위로부터 잡음원이 부가된 음성, 2 채널 Beamformer의 경우, 제안한 방법을 이용하였을 경우 및 원음성의 파워 스펙트럼

Fig. 8. Results of simulation for a two-noise sources for the SNR 0 dB: from the top curve, power spectrum of noisy speech, enhanced power spectrum by two-channel Griffiths-Jim beamformer, enhanced power spectrum by the proposed method, and power spectrum of the original

speech. 이것은 본 방법에 있어서 단구간 스펙트럼의 위상까지 엄밀히 추정하는 것이 아니고, 진폭 스펙트럼만을 고려한 결과라고 생각한다.

V. 결론

본 연구에서는 S.S.법과 2 채널 Griffiths-Jim Beamformer를 융합시켜 잡음의 단구간 진폭 스펙트럼을 추정하고 이것을 입력의 단구간 진폭 스펙트럼으로부터 감산하여 음성을 강조하는 기법을 제안하였다.

S.S.법에서는 목적음의 단구간 진폭 스펙트럼을 추정할 때, 잡음을 정상 혹은 준정상이라 가정하고, 잡음의 평균 진폭 스펙트럼을 이용하고 있다. 그러나, 잡음이 정상이라 하여도 잡음의 단구간 진폭 스펙트럼은 평균값 주변에 분산되어 있으므로, 그 분산값 때문에 목적음의 단구간 진폭 스펙트럼의 추정 성능이 저하되게 된다. 이와 같은 문제점을 해결하기 위한 수단으로서 본 연구에서는 2 채널 Beamformer를 이용하여 잡음의 단구간 진폭 스펙트럼을 추정하고, 이것을 S.S.법에 이용하는 기법을 제안하였다.

2 채널 Beamformer는 마이크로폰 수보다 한 개 적은 임의의 방향에 사각을 형성할 수 있다. 본 연구에서와 같이 디지털 보청기에 음성 강조 기법을 적용시키고자 할 때, 마이크로폰 수는 많아야 두개가 가장 적당하다고 본다. 이와 같이 마이크로폰 수가 한정되어 있을 경우 잡음 방향에 사각을 형성하고자 하면 잡음원의 수가 복수일 경우에는 추정 성능이 크게 떨어진다. 그러나, 보청기 사용자

와 같이 목적음원은 보통 한 개 존재 할 것이라는 전제로 부터 목적음원의 방향에 사각을 형성하여 잡음만을 관측 할 수 있다.

그러나, 이 잡음에는 자기자신의 지연이 포함되어 있으므로 이 잡음의 스펙트럼을 이용할 경우, 지연성분 때문에 진폭 스펙트럼에 Dip이 발생하여 추정성능이 저하 되게 된다. 이것을 보상하기 위하여 지연성분을 포함한 신호와 지연성분이 없는 신호간에 가상적인 전달함수를 설정하고, 그것을 최소 자승법으로 구함으로서 스펙트럼의 Dip을 보상하였다. 컴퓨터 시뮬레이션 결과 잡음원이 복수일 경우, 본 연구에서 제안한 기법이 기존의 기법보다 잡음 억제 능력이 약 5dB 정도 우수함을 보였다.

앞으로는, 본 방법을 이용한 가상적 전달함수를 이용 하였을 경우의 잡음의 단구간 진폭 스펙트럼의 추정성능의 검토 및 세부적인 검토가 필요하고, 또한 청감 특성을 위하여 청취자의 주관적 평가 방법인 단음절 명료도 실험을 할 예정이다.

참 고 문 헌

1. S. F. Boll, "Suppression of Acoustic Noise in Speech Using Spectral Subtraction," VOL. ASSP-27, APR., 1979
2. Yariv Epharim and David Malah, "Speech Enhancement using a Minimum Mean-square Error Short-Time spectral Amplitude Estimator," VOL. ASSP-32, NO. 6, DEC., 1984
3. J. S. Lim and A. v. Oppenheim, "Enhancement and Bandwidth Compression of Noisy Speech," Pro., IEEE, VOL 67, DEC 1979
4. R. J. McAulay, "Speech Enhancement Using a Soft-Decision Noise Suppression Filter," VOL. ASSP-28, NO. 2, APR., 1980
5. D. C. Farden and L. L. Scharf, "Statistical Decision of Nonrecursive Digital Filter," VOL. ASSP-22, NO. 3, JUN., 1974
6. Griffiths, L. J., and Jim C. W., "An alternative approach to linearly constrained adaptive beamforming," IEEE Trans. Antennas Propag. AP-30, 27-34, 1982
7. Widrow, B., Glover, J. R., McCool, J. M., Kaunitz, J., Williams, C. S., Hearn, R. H., Zeidler, J. R., Dong, E., Goodin, R. C., "Adaptive noise cancelling: Principles and Applications," Proc. IEEE 63, 1692-1716, 1975
8. 신원호, 양태영, 김원구, 윤대회, "잡음에 강한 특징 벡터 및 스펙트럼 차감법을 이용한 음성인식," 한국음향학회지, 15권 5호, pp.38-43, 1996년 10월
9. 강홍구, 윤대회, 차일환, "선형 위상 특성을 갖는 적

응 마이크로폰 어레이", 한국음향학회지, 11권 3호, 1992년 6월

김 학 윤(Kim, Hack-Yoon)



1978년~1982년 : 청주대학교 전자공학과(공학사)

1984년~1986년 : 연세대학교 대학원 전자공학과(공학석사)

1993년~1996년 : (일본)동북대학교 정보과학연구과 시스템정보과학전공(공학박사)

1997년~현재 : 청주대학교 전자, 정보통신, 반도체공학부 전임강사

*주관심분야: 전기음향, 신호처리