논문 2004-41SP-6-33

청각기강의 모델을 이용한 음성강조 시스템

(Speech Enhancement System Using a Model of Auditory Mechanism)

최 재 승*

(Jae-Seung Choi)

요 약

음성 신호처리의 분야에서 잡음처리의 문제는 지금도 중요한 연구 과제이다. 특히 배경잡음이 음성의 인식율을 현저히 저하시키는 것은 오래 전부터 주목 받고 있다. 배경잡음으로는 실제 환경에 존재하는 비정상적인 다양한 잡음, 예를 들면 도로에서의 자동차의 주행잡음, 프린터의 구동잡음 등이 있다. 이런 종류에 대한 잡음 대책은 단순하지 않고, 종래의 위너 필터(Wiener filter) 등에 의한 선형적인 잡음제거 법보다도, 보다 고도한 잡음억제 기술이 필요하다. 본 논문에서는, 이러한 방법의 한 가지 시도로써 백색잡음 및 위에 기술한 비정상적인 배경잡음에 의해 열화된 음성을 상호억제로 불리는 인간의 청각기관에서의 잡음억제 기능 모델을 사용하여 음성강화 법의 알고리즘을 소개한다. 제안된 알고리즘은 스펙트럴 왜곡(SD)의 평가방법을 통하여 백색잡음 및 유색잡음에 대해서 효과적인 것을 보여준다.

Abstract

On the field of speech processing the treatment of noise is still important problems for speech research. Especially, it has been noticed that the background noise causes remarkable reduction of speech recognition ratio. As the examples of the background noise, there are such various non-stationary noises existing in the real environment as driving noise of automobiles on the road or typing noise of printer. The treatment for these kinds of noises is not so simple as could be eliminated by the former Wiener filter, but needs more skillful techniques. In this paper as one of these trials, we show an algorithm which is a speech enhancement method using a model of mutual inhibition for noise reduction in speech which is contaminated by white noise or background noise mentioned above. It is confirmed that the proposed algorithm is effective for the speech degraded not only by white noise but also by colored noise, judging from the spectral distortion measurement.

Keywords: speech enhancement, noise reduction, background noise

I. 서 론

근년, 음성인식 장치의 성능이 향상하고 실용화가 광범위하게 진행되고 있지만, 여러 종류의 잡음에 의한 장치의 신뢰성이 크게 저하하는 것이 문제되고 있다. 특히, 정상적인 배경잡음 이외에, 도로에서의 자동차의 주행잡음이나 프린터의 구동잡음 등의 비정상적인 다양한 잡음이 문제되고 있다. 이러한 비정상적인 잡음 중에는 스펙트럴의 형상이 음성이나 음성의 피치(pitch) 성분에 유사한 것도 많기 때문에 음성의 인식율을 현저하게 저하시키는 원인이 되고 있다[1].

잡음의 제거와 경감법에 있어서 선형적인 위너 필터(Wiener filter)[2], 적응 필터법[3], 신경회로망[4] 등에 의한 것 이외에, 적응적 필터로 작용하는 청각의 상호억제 기강을 음성강조 및 잡음제거에 응용하려는 여러 종류의 논문들이 발표되었다. 최근의 이런 종류의 연구로서는 Cheng 등[5], Dang 등[6] 및 Shamma 등[7]의 연구가 있다. 특히 Cheng 등은 헬리콥터 잡음에 대해서 상호억제가 유효하다는 것을 기술하고 있고, SNR(Signal-to-Noise Ratio)에 대해서도 이론적인 해석을 하여 실험적으로도 그 유효성을 증명하고 있다. 그들은 음성을 샘플링 한 후, 이산 캡스트럼(cepstrum) 변환을 실시하여 이 출력에 대해서 상호억제를 하고 있다. 본 연구는 청각적 기강을 생리학적이 아닌 공학적으로 상호억제를

* 정회원, 일본 오사카시립대학교 정보통신공학과
(Department of Information and Communication Engineering, Osaka City University)
접수일자: 2004년2월12일, 수정완료일: 2004년11월4일

응용하려는 입장에서, Cheng 등의 음성강화 시스템에 진폭 조정 계수를 도입해, (1) 잡음의 종류, (2) 상호억제의 주파수 대역폭등의 변화에 대한 SD(Spectral Distortion)를 구하여, 음성의 특성 개선에 필요한 기초자료를 얻고 있다. Cheng 등은 이러한 평가를 SNR을 사용하고 있는 것에 비해, 본 연구에서는 음성의 명료도에 관계가 깊은 SD를 사용하였다.

본 논문에서는 백색잡음뿐만 아니라 실제 환경에서 존재하는 도로에서의 자동차의 주행잡음이나 프린터의 구동잡음과 같은 유색의 배경잡음에 의해서 열화된 음성을 대상으로 하여, 청각생리학을 기초로 한 음성강조 시스템을 구현하였다. 본 연구에서는 음성품질을 개선시켜 보다 명료도가 높은 음성의 재생이 가능하도록 실험적으로 명백히 하였다. 특히 본 연구에서의 시스템은 다양한 잡음 환경 하에서 잡음 레벨(level)의 감소에 효과적이다.

본 논문에서는 Ⅱ장에서 음성강조 시스템에 대해서 설명하며, Ⅲ장에서는 음성의 특성 개선법에 대해 소개한다. Ⅳ장에서는 실험조건과 평가법을, Ⅴ장에서는 실험결과 및 고찰 대해서 기술하고, 마지막으로 Ⅵ장에서는 결론을 맺는다.

Ⅱ. 음성강조 시스템

본 연구에 사용한 음성강조 시스템을 그림 1에 나타낸다. 먼저, 8kHz로 샘플링(sampling)된 잡음이 중첩된 음성신호는 128샘플(sample)의 hamming window $W_1(t)$를 통과한 후 cepstral 변환(cepstral trans-form)을 한다(위의 경로). 구해진 캡스트럼(cepstrum) 성분은 방형창 $W_2(t)$에 의해서 단시간 영역 성분만이 출력되어진 후 FFT(Fast Fourier Transform)에 의해 음성신호의 스펙트럴 포락(spectral envelope)을 얻는다. 프레임(flame) 단위로 지연(D_i)을 한 후, 프레임 단위로 가중치(W_i)를 부가하여 스펙트럴 평균(spectral ave-rage)를 취한다. 이 때, 3프레임 분의 지연이 일어난다. 다음에 이 스펙트럴 성분을 주파수 공간에서 상호억제(Function of Spatial Lateral Inhibition, FSLI)를 한다. 상호억제에 의해서 나타난 부(negative)의 성분은 정류기(rectifier)에 의해서 제거되며, 이 것을 진폭성분으로 한다. 한편, 다른 경로(하단의 경로)에서 FFT되어 3프레임 분이 지연되어 추출된 신호를 위상성분과 진폭성분으로 분리한다. 여기에서 진폭성분은 (입력의 진폭

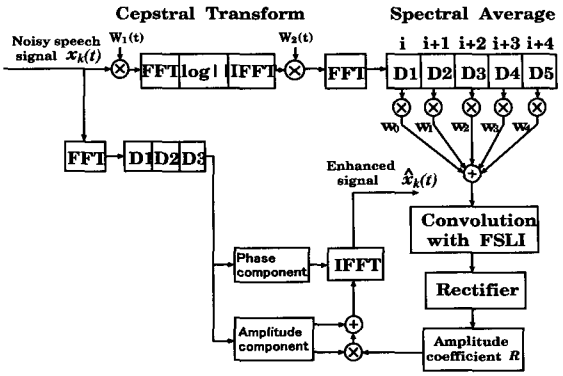


그림 1. 음성강조 시스템
Fig. 1. Speech enhancement system.

성분×(1+R×정류기 출력))으로 변환되어 최종 출력되며, 이 성분과 위상성분을 합성하여 역 푸리에 변환(IFFT)을 함으로써 강조된 음성신호를 재생한다.

Ⅲ. 음성의 특성 개선법

1. 잡음 모델

음성신호를 $s(t)$로 하고, 잡음이 중첩된 음성신호를 식 (1)과 같이 나타낸다.

$$x_k(t) = s(t) + k \times n(t) \tag{1}$$

단, $n(t)$는 잡음신호, k는 잡음강도를 나타내는 계수이다. 이 식을 푸리에 변환(Fourier Transform)을 하면 식 (2)와 같이 된다.

$$X_k(e^{jw}) = S(e^{jw}) + k \times N(e^{jw}) \tag{2}$$

2. 캡스트럼 분석

음성파형 $x(t)$는, 음원파형(pitch 파형) $g(t)$와 성도의 인펄스 응답$v(t)$와의 컨벌루션(convolution)으로 표현된다(방사의 영향은 생각하지 않는 것으로 한다)[8]. 즉, 식 (3)과 같다.

$$x(t) = g(t) * v(t) \tag{3}$$

이산 푸리에 변환의 절대치의 대수 연산조작을 D라고 하면,

$$D\{x(t)\} = \log | X(e^{j\omega}) | = D\{g(t) * v(t)\}$$
$$= D\{g(t) + D\{v(t)\} \tag{4}$$

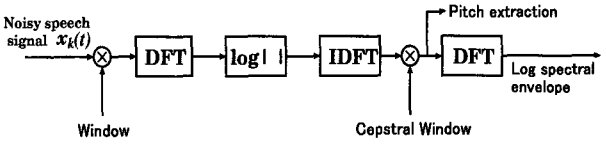


그림 2. 캡스트럼 분석의 순서
Fig. 2. Sequence of cepstrum analysis.

$D\{x(t)\}$의 역 푸리에 변환(Inverse Fourier Trans-form)을 캡스트럼(cepstrum) c(t)라고 부른다. 즉, 식 (5)와 같다.

$$c(t) = \frac{1}{2\pi}\int_0^{2\pi} \log|X(\omega)|\, e^{j\omega t}$$
$$X(\omega) = X(z)|_{z=e^{j\omega t}} \tag{5}$$

따라서, $x(t)$의 캡스트럼은 $g(t)$의 캡스트럼과 $v(t)$의 캡스트럼의 합이 된다. 캡스트럼의 독립변수는 시간의 차원(quefrency)를 가진다.

유성음의 경우에, $D\{g(t)\}$의 캡스트럼은 시간 축 상에서 $1/F_0$ (F_0: pitch 주파수)의 근방에 있는 성분으로 나타나고, $D\{v(t)\}$의 캡스트럼은 단시간 영역성분으로 나타난다. 그러므로, 캡스트럼에 창(window)를 씌워서 단시간 영역성분 만을 추출한 후($g(t)$의 제거), 이것을 이산적 푸리에 변환을 하면 스펙트럴의 포락이 얻어진다. 캡스트럼 분석의 순서를 그림 2에 나타낸다. 즉, 그림 1의 음성강화 시스템의 컨벌루션 입력으로써 평균화된 음성신호의 스펙트럴 포락이 얻어진다.

3. 스펙트럴의 평균

음성신호의 스펙트럴은 프레임 사이에서 급격하게 변동되지는 않지만, 잡음은 프레임 사이에서 불규칙한 스펙트럴 변동을 발생시키는 것으로 생각된다. 그러므로, 프레임 사이에서의 잡음에 의한 불규칙적인 피크(peak)를 감소시키는 하나의 방법으로 스펙트럴의 평균을 취한다. 가중치가 부가된 스펙트럴의 평균을 식 (6)과 같이 나타낸다.

$$\overline{P}_x^{(i)}(\omega) = \frac{1}{2N+1}\sum_{j=-N}^{N} W_j P_x^{(i-j)}(\omega) \tag{6}$$

본 실험에서는 $N=2$로 하고, 가중치를 $[W_0, W_1, W_2, W_3, W_4]$=[0.7, 1.1, 1.4, 1.1, 0.7]로 하였다.

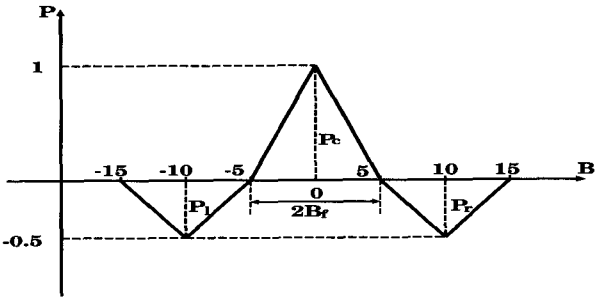


그림 3. 상호억제 함수의 인펄스 모델
Fig. 3. Impulse model of a lateral inhibition function.

$\overline{P}_x^{i}(\omega)$는 평균화된 ($i$)번째의 프레임의 단시간 전력 스펙트럴이다.

4. 상호 억제

FSLI(Function of Spatial Lateral Inhibition)는 내이(internal ear)의 기저막에 있어서 신경상호간의 상호억제 기강을 모의한 것이며, 음성의 스펙트럴의 높은 부분(산과 같은 부분)을 날카롭게 하며, 낮은 부분(계곡과 같은 부분)의 잡음을 경감하는 것으로부터 음성강조에 유효하다고 생각되어진다. 상호억제의 일반적인 기능은 공간적인 입력 패턴 혹은 시간적인 입력 변동을 날카롭게 하는 것이다[5, 6, 7, 9].

그림 3은 본 실험에서 사용한 FSLI의 특성을 나타낸 것이며($B_f=5$의 경우), 가로 축은 주파수 표본점을 나타내고, 세로 축은 주파수 $B_f=0$의 위치에 입력 1이 부가된 경우에 그 근방의 표본점에서 얻어진 출력(인펄스 응답)을 나타내고 있다. 그리고 B_f는 FSLI의 넓이를 결정하는 요소이다.

FSLI의 요소(parameter)를 식 (7)과 같이 설정한다.

$$P_l + P_c + P_r = 0 \tag{7}$$

식(7)의 제한은 상호억제에 의해 잡음의 합의 평균치가 영(zero)으로 되어서 잡음이 경감되기 때문이다. 본 실험에서는 식 (8)과 같은 값을 사용하였다.

$$P_c = 1 \text{ and } P_l = P_r = -0.5 \tag{8}$$

상호억제된 출력은 $\overline{P}_x^{i}(\omega)$와 그림 3에서 나타내는 상호억제 함수와의 컨벌루션에 의해서 구해진다. 그림 4는 스펙트럴 평균이 끝난 후의 입력(…표시)과 상호억

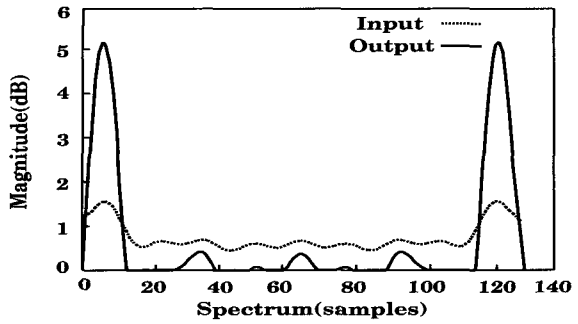


그림 4. 스펙트럴의 평균(점선)과 상호억제에 의한 포르만트의 강조(실선)

Fig. 4. Spectrum average(doted line) and formant enhancement by lateral inhibition(solid line).

제된 출력(-표시)을 나타낸다. 음성의 입력신호는 일본어 "aioi"를 8kHz로 샘플링 했을 때의 음성신호의 /a/의 부분에 백색잡음이 부가된 것을 사용했으며, 다음 그림은 이 신호에 대한 스펙트럴의 평균을 나타낸다. 그림에서 스펙트럴의 높은 부분(산과 같은 부분)이 포르만트(formant)이다. 상호억제 효과에 의해서, 이 포르만트 부분이 강조되고 있는 모양을 알 수 있다.

IV. 실험 조건과 평가법

1. 실험 조건

다음에서 기술하는 것과 같은 조건으로 실험을 실시하였다.

1. 음성 데이터는 일본인 남성화자에 의한 단어와 여성화자에 의한 문장의 두 종류를 사용하였으며, 모두 샘플링 주파수는 8 kHz이다. 남성화자에 의한 음성 데이터로서는 "aioi", "hachioji"를 사용하였다. 여성화자에 의한 음성 데이터로서는"kondonoonseikenkyuukaio"(금번의 음성 연구회를)를 사용하였다. 이 여성화자에 의한 음성데이터는 일본 정보처리 개발협회에서 배부한 연구용 연속 음성 데이터 베이스 중의 성인 남자 화자에 의한 문장이다.

2. 배경잡음으로 사용한 백색잡음은 컴퓨터에 의해서 생성한 가우스(gauss) 백색잡음이다. 그리고 자동차의 주행잡음은 교통량이 많은 도로에서 녹음한 배경잡음이며, 프린터의 구동잡음은 프린터 동작 시에 녹음한 것(구형 프린터 사용)을 배경잡음으로 사용하였다. 이러한 배경잡음으로 사용한 데이터는 모두 샘플링 주파수 8 kHz로 A/D 변환한 것을 사용한다.

3. 음성 및 잡음의 주파수 대역은 0~3 kHz이다.

4. 음성 처리는 1 프레임을 256 샘플로 하고, ham-ming window에 의해서 중첩되지 않게 잘라낸다.

2. 실험 결과의 평가법

재생 신호의 평가는 음성 명료도와 관계가 깊은 식 (9)의 SD[10]를 사용했다.

$$SD = \sqrt{\frac{1}{N_F}\sum_{i=1}^{N_F}\frac{1}{W}\int_0^W \{S_s^{(i)}(f) - S_y^{(i)}(f)\}^2\, df} \quad \text{(dB)} \quad (9)$$

여기에서, N_F는 측정구간의 프레임 수, W는 신호의 대역폭, $S_s^{(i)}(f)$ 및 $S_y^{(i)}(f)$는 입력신호 $s(t)$와 출력신호 $y(t)$의 대수 스펙트럴(dB)이며, 식 (10)과 같이 정의한다.

$$S_s^{(i)}(f) = 10\log_{10}|S(f)|^2$$
$$S_y^{(i)}(f) = 10\log_{10}|Y(f)|^2 \quad (10)$$

여기에서, $S(f)$, $Y(f)$는 각각 주파수 f에서의 입력 신호와 출력 신호의 스펙트럴이다.

V. 실험 결과 및 고찰

지금까지 기술한 것과 같은 기본적인 구성조건 아래에서, 백색잡음, 자동차의 주행잡음, 프린터의 구동잡음 등이 부가된 음성 입력에 대해, 진폭성분 조정계수 R을 바꾸었을 경우의 FSLI의 효과를 SD를 평가 기준으로 하여 구했다.

1. 스펙트럴의 진폭성분 조정계수 R에 관한 효과

표 1과 표2는 각각 백색잡음의 경우, 모음이 많은 단어와 자음이 많은 단어에 대해서, 잡음계수 k와 진폭성분 조정계수 R을 요소(parameter)로 하였을 경우의 시간 신호 $x_k(t)$의 입력 SNR과 출력 SD를 구한 것이다. 표 1과 표 2의 SD의 평가를 보면, 스펙트럴 조정계수 R을 조정함으로써, 각각의 잡음계수 k에 대해서 최적인 R의 값(dB)이 존재하는 것을 알 수 있다(고딕체로 강조). 예를 들면, 표 2의 $k=10$의 경우, 최적인 R의 값에 대해서 SD가 13.0dB이다. 이것은 $R=0$의 잡음을 포함한 원음에 대한 SD 값인 20.4dB보다 7.4dB의 개선되었다. SD의 절대치에 의한 평가부터 SNR의 평

표 1. R의 효과("aioi", $B_f=8$)

Table 1. The effect of R (in the case of "aioi" and $B_f=8$).

잡음 강도	SNR (dB)	진폭성분 조정계수				
		R=0	R=1	R=2	R=3	R=4
k=0	∞	**0.0**	12.3	16.1	18.5	20.1
k=2	4.2	13.5	**9.3**	10.3	11.5	12.6
k=4	-1.9	15.8	10.8	**10.2**	10.5	11.1
k=6	-5.4	16.7	11.9	10.7	**10.6**	10.9
k=8	-7.9	17.1	12.6	11.2	**10.9**	11.2
k=10	-9.8	17.4	13.1	11.5	**11.1**	11.4

표 2. R의 효과("hachioji", $B_f=8$)

Table 2. The effect of R (in the case of "hachioji" and $B_f=8$).

잡음 강도	SNR (dB)	진폭성분 조정계수				
		R=0	R=1	R=2	R=3	R=4
k=0	∞	**0.0**	13.4	17.8	20.1	21.7
k=2	0.9	16.2	**12.5**	13.1	14.1	15.1
k=4	-5.1	18.4	13.1	**12.6**	13.1	13.5
k=6	-8.7	19.5	13.8	12.8	**12.7**	13.1
k=8	-11.2	20.1	14.4	13.1	**12.8**	13.3
k=10	-13.1	20.4	14.8	13.4	**13.0**	13.5

과 방법과 다르기 때문에 단순하게 SNR과 비교하기는 어렵지만, 전화의 PCM 방식의 평가 데이터에 의하면, SNR에서의 20dB의 개선량은 SD에서는 약 3.5dB에 대응하므로(참고 문헌[10]의 Fig. 3), SD에서의 7.4dB의 개선량은 상당히 효과적이라고 말할 수 있다. 여기에서, 표에 나타나있는 SNR은 $x_k(t)$의 신호 대 잡음 비이다. 실제로, 출력된 음성을 들어본 결과, SD가 적을수록 양호한 재생 음성임을 확인할 수 있었다. 표 2의 자음이 많은 단어에 대한 결과로부터, 자음이 많으면 본 시스템에서는 SD의 값이 크게 되는 것을 알 수 있다. 표 3은 FSLI의 효과를 나타낸 것이며, "No Processing"은 잡음을 포함한 원음에 대한 SD를 표시하고(즉, $R=0$), "Without FSLI"는 그림 1에서 "Convolution with FSLI"와 "Rectifier"를 제외한 R만을 최적의 값으로 조정한 경우의 SD를 나타내고 있다. 그리고 "With FSLI"는 II장에서 제안한 본 방식에 의해 구한 SD이다.

이상의 결과로부터 알 수 있듯이, FSLI의 처리를 한 경우가 모음이 많은 단어와 자음이 많은 단어의 모두에 대해서 SD의 값(dB)이 적게 되므로 FSLI가 효과적이라는 것을 알 수 있다.

표 3. FSLI의 효과를 나타내는 SD("hachioji", $B_f=8$)

Table 3. SD effect of FSLI (in the case of "hachioji" and $B_f=8$).

잡음강도	No processing	Without FSLI	With FSLI
k=0	0.0	0.0	0.0
k=2	16.2	15.2	12.5
k=4	18.4	17.7	12.6
k=6	19.5	19.1	12.7
k=8	20.1	19.8	12.8
k=10	20.4	20.2	13.0

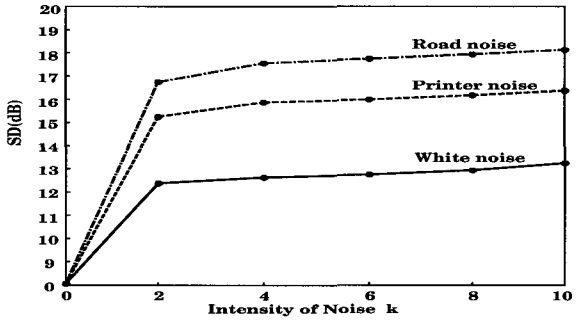


그림 5. FSLI를 사용한 경우의 효과의 비교("hachioji", $B_f=8$)

Fig. 5. Comparison between the effects of FSLI (in the case of "hachioji" and $B_f=8$).

2. 각종 잡음과 상호억제 계수 B_f에 대한 FSLI의 효과

그림 5는 자음이 많은 단어에 대해서, FSLI를 사용한 경우의 백색잡음, 프린터의 구동잡음, 자동차의 주행잡음의 크기(진폭)에 대해서 최적인 계수 R을 선택한 경우의 SD의 변화를 표시한 것이며, 기술된 순서로 SD가 개선되었다.

그림 6은 자음이 많은 단어에 대해서, FSLI를 사용한 경우와 FSLI를 사용하지 않는 경우의 프린터의 구동잡음에 대한 SD를 비교한 것이다. FSLI를 사용한 경우가 FSLI를 사용하지 않는 경우 보다 SD가 약 3dB 개선되었다. 여기에서, $k=10$인 경우, 프린터의 구동잡음의 입력 SNR은 -33.3(dB)이다. 지금까지의 결과로부터, FSLI를 사용한 경우가 효과적이라는 것을 알 수 있으며, 특히 입력 SNR이 -33(dB) 이하의 극히 열악한 조건에서도 잡음 제거 효과가 높은 것을 확인할 수 있다.

그림 7은 자음이 많은 단어에 대해서, FSLI의 계수 B_f를 조정한 경우의 프린터의 구동잡음에 대해서 SD를 비교한 것이다. FSLI의 계수의 폭 B_f가 $B_f=8$, $B_f=11$, $B_f=5$의 순서로 SD가 개선되어 있다. 실험결과로부터, $B_f=8$인 경우가 가장 효과적인 것을 알

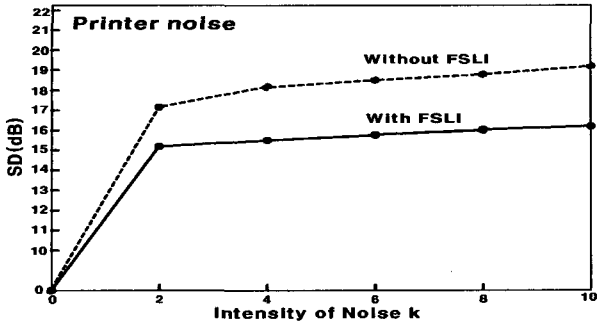


그림 6. FSLI를 사용한 경우와 FSLI를 사용하지 않은 경우의 비교("hachioji", $B_f = 8$)

Fig. 6. Comparison between the cases with FSLI and without FSLI (in the case of "hachioji" and $B_f = 8$).

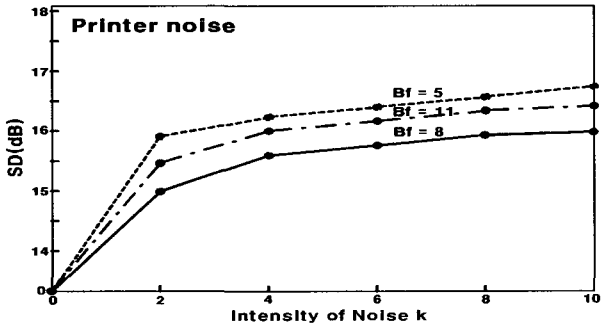


그림 7. FSLI의 계수B_f를 조정한 경우의 비교("hachioji", $B_f = 8$)

Fig. 7. Comparison between the cases when coefficient B_f of FSLI is adjusted (in the case of "hachioji" and $B_f = 8$).

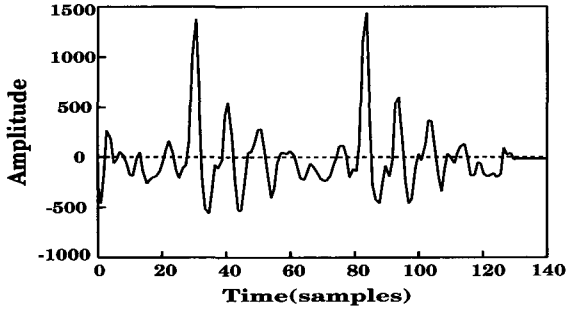


그림 8. 잡음이 없는 입력음성 신호의 파형($k = 0$)

Fig. 8. Input of clean speech signal (in the case of $k = 0$).

수 있다.

그림 8, 9, 10은 샘플 수를 128로 하였을 경우, "aioi"의 음성신호 /a/의 부분을 파형으로 표시한 것이다. 그림 8은 잡음이 부가되지 않는 경우의 입력음성 파형이고, 그림 9는 백색잡음이 부가된 입력음성 파형이다. 그림 10은 강조된 재생음성 파형을 나타낸 것인데, 그림의 파형에서 나타난 것과 같이 잡음제거의 모양을 알

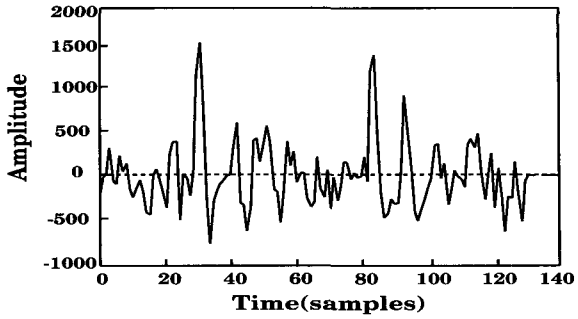


그림 9. 백색잡음이 부가된 입력음성 신호의 파형($k = 3$)

Fig. 9. Input of contaminated speech signal with white noise (in the case of $k = 3$).

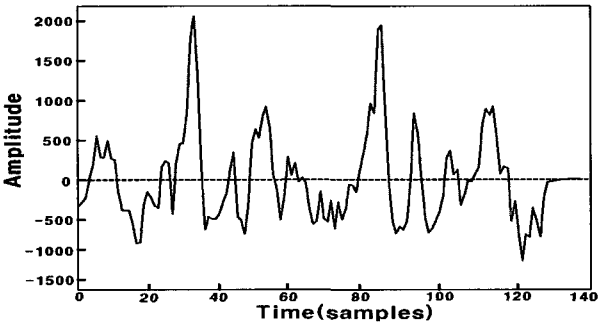


그림 10. 재생 음성신호의 파형($k = 3$)

Fig. 10. Waveform of output speech signal (in the case of $k = 3$).

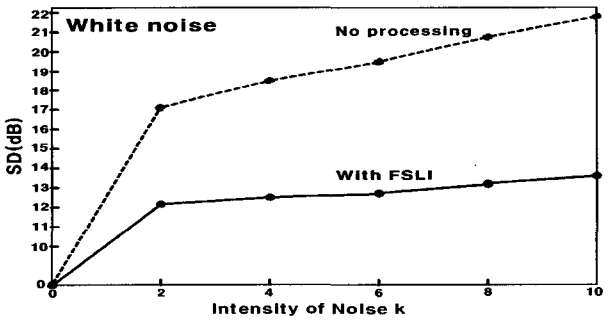


그림 11. 백색잡음의 경우의FSLI 효과($B_f = 8$)

Fig. 11. FSLI Effect when white noise is added(in the case of $B_f = 8$).

수 있다.

그림 11과 그림 12는 여성화자에 의한 음성 데이터 "kondonoonseikenkyuukaio"에 대해서, "No processing"과 "With FSLI"의 경우의 백색잡음과 자동차의 주행잡음에 대한 SD를 비교한 것이다. 그림 11의 백색잡음에 대한 SD의 효과를 보면, FSLI를 사용한 경우가 "No processing"에 대해서 최대 약 8.5dB 개선되어 있다. 그림 12의 자동차의 주행잡음에 대해서도, FSLI를 사용한 경우가 "No processing"에 대해서 최대 약 7dB 개선되

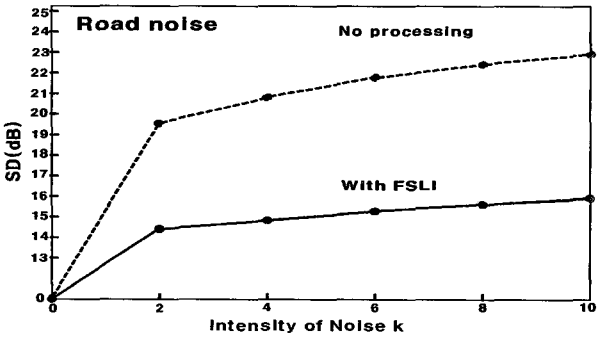


그림 12. 자동차의 주행잡음의 경우의 FSLI 효과 ($B_f = 8$)

Fig. 12. FSLI Effect when road noise is added(in the case of $B_f = 8$).

어 있다. 여기에서, $k = 10$인 경우, 백색잡음의 입력 SNR은 -10.7dB이며, 자동차의 주행잡음의 입력 SNR은 -30.5dB이다. 두 그림으로부터 알 수 있듯이 잡음량이 증가할수록 SD의 개선량이 증가하는 경향이 있으며, 특히 입력 SNR이 -30dB 이하의 극히 열악한 조건에서도 잡음 제거 효과가 높다는 것을 확인할 수 있다. 따라서, FSLI의 효과에 의해서 백색잡음 및 자동차의 주행잡음에 대해서 각각 8.5dB와 7dB 개선되어 있는 것을 알 수 있다.

이상의 결과로부터, 본 연구에 사용한 음성강조 시스템이 백색잡음 및 유색잡음에 대해서 효과적인 것을 말할 수 있다. 특히 단어의 음성데이터에 대해서는 입력 SNR이 -33dB 이하의 극히 열악한 조건에서도 잡음 제거 효과가 높은 것을 확인할 수 있었고, 문장의 음성데이터에 대해서는 입력 SNR이 -30dB 이하의 극히 열악한 조건에서도 충분히 잡음 제거 효과가 높다는 것을 확인하였다.

VI. 결 론

인간의 청각 시스템에서의 상호억제 기강이 잡음억제의 효과가 있다는 것에 착안하여, 이것을 공학적으로 응용하려는 하나의 시도로써 본 시스템을 제안하여, 이것이 스펙트럴 왜곡율(SD) 에서 유효하다는 것을 여러 종류의 잡음에 대해서 실험적으로 증명하였다.

실험 결과를 정리하면, 다음과 같은 결론을 얻는다.

1. 본 연구에서는 입력 SNR이 극히 열악한 조건에서의 잡음제거에 대해서도 높은 효과를 나타내고 있다. 특히 단어의 음성데이터에 대해서는 입력 SNR이 -33dB 이하에서, 문장의 음성데이터에 대해서는 입력 SNR이 -30dB 이하의 극히 열악한 조건에서도 충분히 잡음 제거 효과가 높다는 것을 확인하였다.

2. 각각의 잡음계수k에 대해서 최적인 진폭성분 조정계수 R이 존재한다. 실제로, 재생음성을 들어본 결과, SD값이 적으면 적을수록 양호한 음성이 얻어졌다. 또한, "No processing"의 음성은 SD값이 적어도 귀로 들었을 때의 거슬린 음성으로 들린다. 이것은 청각의 마스킹(masking) 효과에 의한 것으로 생각된다.

3. 모음이 많은 음성이 자음이 많은 음성보다 잡음경감의 효과가 크다(즉, SD의 값이 적다).

4. 프린터의 구동잡음, 자동차의 주행잡음과 같이 잡음의 종류에 따라서 FSLI의 잡음경감 효과가 다르다. 특히, 백색잡음에 대한 FSLI의 잡음경감 효과가 현저하다.

5. 상호억제의 계수B_f에는 최적치가 존재한다.

이상과 같이, 음성신호의 잡음경감을 위해서 FLSI가 특히 백색잡음에 대해서 효과적이라는 것을 실험적으로 확인하였지만, 향후의 연구과제로서는 자동차의 주행잡음 등의 유색잡음에 의해서 열화된 음성에 대해서도 더욱 강화하는 방법의 검토가 필요하다고 생각된다.

본 연구는 청각생리의 상호억제라는 기본적인 특성을 이용하여 음성강조에의 응용의 가능성을 검증하였다. 향후 더욱 검토를 함으로써 더욱 효과적인 응용법이 전개될 것이라고 생각하고 있다.

참 고 문 헌

[1] Yadong Wu, Yan Li, "Robust speech/non-speech detection in adverse conditions using the fuzzy polarity correlation method", Systems, Man, and Cybernetics, 2000 IEEE International Conference on ,Vol. 4 ,8-11 Oct. pp. 2935-2939, 2000.

[2] Sreenivas and P. Kirnapure, "Codebook constrai -ned wiener filtering for speech enhancement," IEEE Trans. Speech and Audio Processing, Vol.4, No.5, pp. 383-389, 1996.

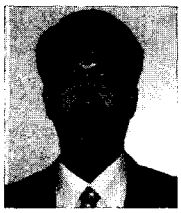
[3] B. Widrow et al., "Adaptive noise cancelling: Principles and applications," Proc. IEEE, Vol. 63, No. 12, pp. 1692-1716, 1975.

[4] W. G. Knecht, M. E. Schenkel, and G. S. Moschytz, "Neural network filters for speech enhancement," IEEE Trans. Speech and Audio Processing, Vol.3, No.6, pp. 433-438, 1995.

[5] Y. M. Cheng and D. O'Shaughnessy, "Speech enhancement based conceptually on auditory evidence," IEEE Trans. Signal Processing, Vol.

39, No. 9, pp. 1943-1953, 1991.
[6] V. C. Dang, R. Carre, D. Tuffelli, "Speech signal preprocessing taking into account lateral inhibi -tion", Signal Processing III: Theories and Applications., Vol. 1, pp. 533-536, 1986.
[7] Shihab A. Shamma, "'Speech processing in the auditory system II: Lateral inhibition and the central processing of speech evoked activity in the auditory nerve", J. Acoust. Soc Amer., pp. 1622-1632, 1985.
[8] Steven. Boll, "Suppression of acoustic noise in speech using spectral subtraction,"IEEE Trans. Acoust., Speech Signal Processing, Vol. ASSP-27, No.2, pp. 113~120, 1979.
[9] J. H. L. Hansen and S. Nandkumar, "Robust estimation of speech in noisy backgrounds based on aspects of the auditory process," J. Acoust. Soc. Am. 97(6), pp. 3833-3849, June, 1995.
[10] K. Itoh, et al., "A Study of Objective Quality Measures for Digital Speech Waveform Coding Systems", The Institute of Electronics, Informa -tion and Communication Engineers(IEICE), Vol. J 66-A, No. 3, pp. 274-281, 1983.

저 자 소 개



최 재 승(정회원)
1989년 조선대학교 전자공학과 졸업(공학사)
1995년 일본 오사카시립대학 정보통신공학과(공학석사)
1999년 일본 오사카시립대학 정보통신공학과(공학박사)
2000년 일본 마쯔시타 전기산업주식회사 AVC사 연구원
<주 관심분야: 음성신호처리, 잡음제거, 신경망 등>