

반향 필터 추정에서 성감 특성을 이용한 단일채널 음성반향제거 방법

A Single-Channel Speech Dereverberation Method Using Sparse Prior Imposition in Reverberation Filter Estimation

지민선¹⁾ · 박형민²⁾

Zee, Min-Seon · Park, Hyung-Min

ABSTRACT

Since a reverberation filter is generally much shorter than the corresponding dereverberation filter, a single-channel speech dereverberation method based on reverberation filter estimation has been developed to improve its performance. Unfortunately, a typical reverberation filter still requires too many coefficients to be accurately estimated using limited speech observations. In order to exploit sparseness of reverberation filter coefficients, in this paper, we present an algorithm to impose a sparse prior to the process of reverberation filter estimation. Simulation results demonstrate that the sparse prior imposition further improves performance of the speech dereverberation method based on reverberation filter estimation.

Keywords: speech dereverberation, sparse prior, reverberation filters, filter estimation

1. 서론

음성은 영상통화 또는 핸드프리 통화에서와 같이 화자가 마이크로부터 떨어져있을 때 음향반향에 의해 왜곡된다. 반향은 일반적으로 음성의 명료도와 자동 음성 인식의 성능을 크게 떨어뜨린다고 알려져 있다[1],[2]. 이러한 경우, 반향이 없는 음성신호를 복원하기 위해서 반향제거기술이 적용된다. 반향제거기술은 화상통화나 핸드프리 통화 외에도 원격 화상회의, 원거리 음성인식과 같이 음질향상과 통신 분야에서 널리 이용되고 있기 때문에, 음성신호처리 분야의 연구와 발전에 있어서 중요한 주제가 되어왔다.

대부분의 경우 반향이 없는 원래 음성 신호와 반향 환경을

알지 못하기 때문에, 마이크 신호만을 이용하여 음향반향의 역에 해당하는 필터를 추정하기 위해 암묵 반향제거방법들이 많이 제안되어왔다(예,[3]-[6]). 그 중 많은 연구들은 여러 개의 마이크를 동시에 사용하는 다채널 신호 획득을 기반으로 하였고, 그 대부분은 화자로부터 마이크까지의 다른 경로에 의해 발생한 신호 차이를 이용하였다(예,[7],[8]). 그러나 추가적인 마이크 부담이 없는 단일 마이크를 사용할 때는 이 신호 차이를 이용할 수 없고, 음향환경에서의 반향제거가 매우 복잡할 뿐만 아니라, 이웃한 표본 사이에 상관성이 있고 비정상적(non-stationary)인 음성을 이용하기 때문에, 단일 채널 음성 암묵반향제거는 어려운 문제로 알려져 있다.

단일채널 암묵반향제거 방법들 중에서 많은 알고리즘은 반향이 없는 음성의 확률밀도함수가 정규분포가 아니라는 가정에 기초한다[3]. 반향된 음성의 확률밀도함수가 중심극한정리에 의해서 정규분포함수에 근접해나가므로, 복원된 신호의 비정규성을 최소화시킴으로써 반향제거를 할 수 있다. 특히 이와 동일한 효과를 얻을 수 있는 정보 이론에 기반한 적응알고리즘은 간단한 수식들과 강인성으로 인해 많은 관심을 받고 있다[3],[4],[9].

비록 이 알고리즘이 반향 성분을 효과적으로 없앨 수 있다

1) 서강대학교, minseon.zee@gmail.com

2) 서강대학교, hpark@sogang.ac.kr, 교신저자

본 연구는 교육과학기술부의 재원으로 한국연구재단의 지원(지원번호: 2010-0025703) 및 미래창조과학부의 ETRI 연구개발지원사업의 일환으로 수행되었음 [11921-03001, "Beyond 스마트 TV 기술개발"].

접수일자: 2013년 9월 25일

수정일자: 2013년 11월 19일

게재결정: 2013년 12월 10일

하더라도, 음향 환경에서 반향제거 필터들은 일반적으로 매우 길고 복잡하기 때문에, 한정된 개수의 관측값으로 반향제거 필터를 직접 추정할 때 오차가 크게 나타난다³⁾. 따라서 본 저자는 반향제거 필터를 직접 추정하는 대신에, 반향 필터를 추정하는 효율적인 정보 이론 기반 주파수영역 적응알고리즘을 제안하였다[10]. 반향 필터는 보통 반향제거 필터보다 훨씬 간단하기 때문에 한정된 개수의 입력 신호를 가지고 필터 추정 시에 오차를 줄여서 반향이 없는 원래 음성에 더 가까운 신호를 복원할 수 있게 된다. 그러나 일반적인 반향 필터는 여전히 일정한 음향 채널 구간 내에 표본 간 상관도가 매우 높은 음성신호의 한정된 데이터를 사용하여 정확히 추정하기에는 너무 많은 필터 계수를 필요로 한다.

일반적인 암묵 필터 추정알고리즘에서는 관측값만을 사용하여 필터 추정을 피하지만, 추정하고자 하는 필터의 일부 특성을 선형적 지식에 기반하여 미리 가정이 가능하다면 한정된 개수의 마이크 신호값으로도 더 정확한 필터 추정이 가능하다. 실제로 혼합 행렬이나 디콘볼루션(deconvolution) 필터에 대한 성김(sparsity) 특성을 부과하여 불충분한 양의 데이터로부터 영상특징이나 반향이 제거된 음성을 추정하는 성능을 향상시킬 수 있다[11],[12]. 방 임펄스 응답과 같이 일반적인 음향 환경에서의 반향 필터는 크기가 큰 계수가 드물게 나타난다. 따라서 본 논문에서는 미리 가정이 가능한 필터 특성으로서 반향 필터 추정 시에 성김 특성을 이용하여 반향 필터를 더 정확히 추정하는 방법을 제안하고 추정된 필터의 역필터를 사용함으로써 반향이 제거된 음성을 찾아내는 방법을 설명한다.

본 논문의 구성은 다음과 같다. 2장에서 반향제거 필터 추정을 위한 기존의 정보 이론에 기반한 주파수 영역 적응알고리즘에 대해 살펴보고, 3장에서는 반향 필터 추정을 위한 알고리즘을 설명한다. 그리고 4장에서 반향 필터에 성김 특성을 부과하는 방법에 대하여 설명하고, 5장에서는 기존방법과 4장에서 제안한 방법을 비교하기 위한 실험방법 및 결과를 기술한다. 마지막으로 6장에서 결론을 맺는다.

2. 반향제거 필터 추정을 위한 정보이론기반 적응알고리즘

화자와 멀리 떨어져있는 마이크로로부터 반향된 음성을 받아들인다고 가정하자. 이 때, 입력신호와 출력신호의 관계는 식 (1)과 같다.

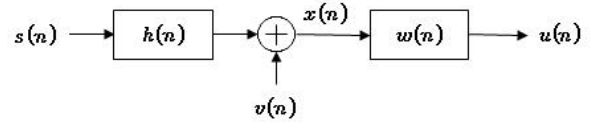


그림 1. 단일채널 반향 및 반향제거에 대한 전체적인 시스템.

Figure 1. Overall system of single-channel reverberation and dereverberation.

$$x(n) = \sum_{k=0}^{L-1} h(k)s(n-k) + v(n) \quad (1)$$

암묵반향제거 문제는 반향 필터 $h(n)$ 와 반향이 없는 원신호 $s(n)$ 을 알지 못한 채, 출력신호 $x(n)$ 을 이용하여 $s(n)$ 을 복원해내는 것이다. 보통, 잡음신호 $v(n)$ 은 평균이 0인 정규 잡음으로 가정한다.

대부분의 암묵반향제거 방법은 반향이 없는 원신호 $s(n)$ 을 복원하기 위해서 유한 임펄스 응답을 선형 반향제거 시스템으로 사용한다. 따라서 반향제거 모델은 식 (2)와 같이 나타낼 수 있다.

$$u(n) = \sum_{k=0}^{L_w-1} w(k)x(n-k) \quad (2)$$

여기서 $w(n)$ 는 반향제거 필터의 필터계수이다. 단일채널 반향 및 반향제거를 포함한 전체시스템은 <그림 1> 과 같다. 암묵반향제거 방법은 미리 약속된 훈련 신호를 따로 가지고 있지 않으므로, 보통 반향제거 신호 $u(n)$ 의 선형(a priori) 통계적인 지식을 이용하여 반향제거 필터를 추정한다.

정보 이론에 기반한 적응알고리즘을 유도하기 위해서 반향제거 신호의 이웃한 표본 간의 중복성(redundancy)을 없애도록 정보를 최대화한다[3]. 이를 위해서 반향제거 신호의 엔트로피가 최대가 되도록 반향제거 필터계수를 추정하기 위한 최대경사법은 주파수 영역에서 식 (3)으로 나타낼 수 있다.

$$\Delta W \propto (1/W^* - \text{fft}\{g(\underline{u})\} \cdot X^*) \quad (3)$$

여기서 W 와 X 는 각각 반향제거 필터와 마이크 녹음신호 $x(n)$ 의 고속 푸리에 변환값이고[3], \underline{u} 는 “유한 임펄스 응답 행렬의 대수학 (FIR matrix algebra)”[13]에 따라 반향제거 신호 $u(n)$ 으로 구성된 벡터이다⁴⁾. 비선형 함수 $g(\cdot)$ 는 $p(\cdot)$ 가 확률밀도함수를 의미할 때, $-p'(u)/p(u)$ 를 나타낸다. 수렴 성능을 개선하기 위해서 자연경사법(the natural gradient)을 사용하면 식 (3)은 식 (4)로 바뀐다.

3) 음향반향 필터는 시간에 따라서 변하는 경우도 있기 때문에, 반향 필터가 거의 일정한 구간의 입력 신호 개수는 반향제거 필터를 정확히 추정하는데 부족할 수 있다.

4) 위신호(aliasing)가 나타나지 않도록 벡터의 전반은 0으로 구성한다.

$$\Delta W \propto (W - \text{fft}\{g(\underline{u})\} \cdot U^* \cdot W) \quad (4)$$

여기서 U 는 복원된 반향제거 신호 $u(n)$ 의 고속푸리에 변환 값을 나타낸다[4],[9].

음성의 연속적인 표본들은 크게 상관되어있는 반면에, 이 알고리즘은 출력 신호를 시간적으로 독립화하려 하기 때문에, 이를 음성반향제거에 바로 적용하면 음성 고유의 시간상관이 제거된 결과를 얻게 된다. 본 논문에서는 [14]에서와 같이 음성의 고유의 시간상관을 제거하기 위해 반향이 없는 음성신호에 식 (4)의 알고리즘을 적용하여 독립화 필터를 미리 추정한다. 이 독립화 필터를 사용하여 반향이 있는 마이크 신호의 음성 고유의 시간적 상관을 제거한 후에 반향제거 필터를 추정하는 반향제거 알고리즘을 사용하게 된다. 이 때, 반향제거 음성신호는 반향이 있는 마이크 신호에 반향제거 필터를 적용하여 얻을 수 있다.

3. 반향 필터 추정방법

일반적으로 반향 필터의 역에 해당하는 음향반향제거 필터는 반향 필터보다 훨씬 더 길고 복잡하다. 3장에서는, 더 간단한 필터인 반향 필터를 추정하는 방법을 설명한다. 더 간단한 필터를 추정하기 때문에 한정된 개수의 입력 데이터에 대해 실제 반향 필터에 더 가까이 수렴할 수 있으며, 반향이 없는 원래 음성에 더 가까운 반향제거 신호를 찾을 수 있다.

<그림 1>에서 반향모델을 나타낸 식 (1)은 시간지연 연산자 z^{-1} 을 이용하면 식 (5)와 같이 나타낼 수 있다.

$$x(n) = H(z)s(n) + v(n) \quad (5)$$

이때, $H(z)$ 는 식 (6)과 같다.

$$H(z) = \sum_{k=0}^{L-1} h(k)z^{-k} \quad (6)$$

반향 필터 $h(k)$ 를 추정하기 위해 길이가 N 인 반향모델의 입출력 신호벡터를 다음과 같이 정의한다.

$$\mathbf{s} = [s(0), s(1), \dots, s(N-1)]^T \quad (7)$$

$$\mathbf{x} = [x(0), x(1), \dots, x(N-1)]^T \quad (8)$$

간단한 수식 유도를 위해 $n < 0$ 일 때 $s(n)$ 과 $x(n)$ 이 0 이고, $v(n)$ 은 무시할 수 있다고 가정한다.

따라서 출력신호벡터 \mathbf{x} 는 식 (9)와 같이 나타낼 수 있다.

$$\mathbf{x} = \mathbf{H}\mathbf{s} \quad (9)$$

이때, \mathbf{H} 는 식 (10)과 같다.

$$\mathbf{H} = \begin{bmatrix} h(0) & 0 & \dots & 0 \\ h(1) & h(0) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ h(N-1) & h(N-2) & \dots & h(0) \end{bmatrix} \quad (10)$$

여기서 필터 길이 L 이 N 보다 훨씬 작다고 가정하여 $h(L+1) = h(L+2) = \dots = h(N-1)$ 은 모두 0이 된다.

출력신호벡터 \mathbf{x} 의 확률밀도함수는 식 (11)과 같다.

$$p(\mathbf{x}|\mathbf{H}) = \frac{p(\mathbf{s})}{|\det(\mathbf{H})|} \quad (11)$$

식 (11)의 로그우도는 식 (12)와 같다.

$$\log L(\mathbf{H}) = \log p(\mathbf{s}) - \log |\det(\mathbf{H})| \quad (12)$$

\mathbf{H} 에 대하여 로그우도를 최대화하는 자연경사법은 [15]의 유도를 따라 식 (13)과 같이 나타낼 수 있다.

$$\Delta \mathbf{H} \propto \mathbf{H}\mathbf{H}^T \nabla_{\mathbf{H}} \log L(\mathbf{H}) = \mathbf{H}g(\mathbf{s})\mathbf{s}^T - \mathbf{H} \quad (13)$$

따라서 주파수영역에서 반향 필터를 추정하는 식은 식 (14)와 같다.

$$\Delta H \propto (H \cdot \text{fft}\{g(\underline{s})\} \cdot S^* - H) \quad (14)$$

여기서 H 와 S 는 각각 반향 필터 $h(k)$ 와 반향모델의 입력신호 $s(n)$ 의 고속 푸리에 변환 값이고, \underline{s} 는 앞 장의 \underline{u} 와 같이 반향모델의 입력신호 $s(n)$ 으로 구성된 벡터이다.

실제 반향제거 문제에서는 $s(n)$ 을 반향 필터 $H(z)$ 의 역 필터를 사용하여 구할 수 있다. 반향 필터 $h(k)$ 를 알고 있을 때 역필터 $h_{\text{inv}}(k)$ 를 찾는 알고리즘은 많지만(예.[16],[17]), 본 논문에서는 식 (15), (16)과 같이 고속 푸리에 변환에 기반한 간단한 방법을 사용한다.

$$\begin{aligned} & [h_{\text{inv}}(0) h_{\text{inv}}(1) \dots h_{\text{inv}}(L_{\text{inv}}-1)] \\ & = \text{ifft}([1/h(0) \ 1/h(1) \dots \ 1/h(L_{\text{inv}}-1)]) \end{aligned} \quad (15)$$

$$\begin{aligned} & [h(0) h(1) \dots h(L_{\text{inv}}-1)] \\ & = \text{fft}([0 \ 0 \dots \ 0 \ h(0) h(1) \dots \ h(L-1) \ 0 \ 0 \dots \ 0]) \end{aligned} \quad (16)$$

여기서 위신호(aliasing)를 피하기 위해서는 고속 푸리에 변환의 입력인 반향 필터 벡터의 양쪽에 있는 0의 개수를 크게 해야 한다.

4. 반향 필터에 성감 특성 부과

앞 장에 설명한 방법과 같이 반향제거 필터보다 상대적으

로 간단한 반향 필터를 추정함으로써 성능이 향상될 수 있지만, 여전히 반향 필터에 대한 어떤 가정도 하지 않았다. 만약 반향 필터에 대한 선험적인 지식을 이용할 수 있다면, 한정된 개수의 마이크 신호로 더욱 정확하게 필터를 추정할 수 있다.

반향 필터의 확률밀도함수 $p(H(z))$ 를 미리 알고 있다고 가정하자. 통계적 관점에서 반향 필터의 최적 추정이 사후확률과 관계되어있기 때문에 사후확률밀도함수 $p(H(z)|\mathbf{x})$ 의 로 그 값은 Bayes 정리를 이용해 식 (17)과 같이 나타낼 수 있다.

$$\log p(H(z)|\mathbf{x}) = \log p(\mathbf{x}|H(z)) + \log p(H(z)) - \log p(\mathbf{x}) \quad (17)$$

여기서, $H(z)$ 에 대해서 $\log p(H(z)|\mathbf{x})$ 를 최대화함으로써 선험적 지식을 부과한 반향 필터 추정방법을 유도할 수 있다. 식 (17)의 우변 중 세 번째 항은 필터와 관계없기 때문에 적용알고리즘에 영향을 주지 않고, 첫 번째 항을 $H(z)$ 에 대해서 최대화하는 것은 식 (14)의 반향 필터 추정방법과 같다. 식 (14)에 더해, 두 번째 항 $\log p(H(z))$ 을 $H(z)$ 에 대해서 식 (18)과 같이 최대화 시킨다.

$$\frac{d \log p(H(z))}{dH(z)} = \frac{dp((H(z)))}{p(H(z))} \quad (18)$$

방의 임펄스 응답과 같은 일반적인 음향반향 필터에서는 크기가 큰 필터 계수가 드물게 나타나기 때문에, 반향 필터에 성감 특성을 부과할 수 있다. 성감 특성을 반영할 수 있는 분포로서 간단하고 일반적인 확률분포함수인 0의 평균값을 갖는 라플라스분포를 사용하면, 식 (18)은 다음과 같아진다.

$$\begin{aligned} \Delta h(k) &\propto \frac{\partial \log p(H(z))}{\partial h(k)} = \frac{\partial \sum_{l=0}^{L-1} \log p(h(l))}{\partial h(k)} \\ &= \frac{\partial \sum_{l=0}^{L-1} -|h(l)|}{\partial h(k)} = -\text{sgn}(h(k)) \end{aligned} \quad (19)$$

여기서 수식 유도를 간단히 하기 위해 반향 필터계수는 상호 독립인 동일분포를 따른다고 가정한다.

사후확률밀도함수를 최대화함으로써, 마이크 입력신호의 우도를 최대화한 식 (14)와 사전확률밀도함수를 최대화한 식 (19), 이 두 가지 공식을 얻을 수 있다. 식 (14)가 2L개 고속 푸리에 변환으로 주파수영역에서 효율적으로 구현되더라도 추정된 필터는 순환 컨볼루션에 의한 오차를 피하기 위해서 길이가 L이거나 그보다 더 짧아야 한다. 따라서 3장에서 설명한 반향 필터를 추정하는 방법은 시간 영역에서 마지막 L개의 값들을 0으로 하기 위해 필터 값이 갱신될 때 고속 역 푸리에 변환이 필요하다. 결국 식 (19)에 따른 성감 특성 최대화는 무시할만한 계산량의 증가로 쉽게 수행이 가능하다.

5. 실험 결과

3장과 4장에서 기술한 반향 필터를 추정하는 방법과 그 필터에 성감 특성을 부과한 효과를 평가하기 위해 반향이 없는 음성을 반향 필터에 통과시킨 반향입력신호로 모의실험을 하였다. 반향이 없는 음성신호로는 TIMIT database[18]에서 특정 화자의 문장들을 연결하여 사용하였다. 화자를 바꿔가며 실험을 반복하기 위해 10명의 남자 화자를 임의로 선택하였다⁵⁾. 그 신호들의 길이는 평균적으로 25초 정도이고, 표본화속도는 계산량을 줄이기 위해 8 kHz로 변환하였다. 음성신호는 대략적으로 라플라스분포를 따른다는 사실에 근거하여 식 (4)와 (14)의 $g(\cdot)$ 으로 $\text{sgn}(\cdot)$ 을 사용하였다. 음성 고유의 시간적 상관성을 없애기 위한 필터 추정을 위해 이전에 선택한 10명의 남자 화자를 제외하고 다시 10명의 남자 화자를 임의로 선택하였다⁶⁾. 이 화자들의 반향이 없는 음성신호로부터 길이가 256인 독립화 필터를 생성하였다.

반향 필터에 의한 반향 성분이 얼마나 잘 추정되고 제거되었는지 파악하기 위한 직접적인 성능 척도로 심별 간 간섭 (ISI)과 신호 대 반향잡음비(SRR)를 사용하였다[12],[19]. 이때, 심별 간 간섭은 식 (20)과 같이 정의된다.

$$\text{ISI (dB)} = 10 \log \left(\frac{\sum_k |t(k)|^2 - \max_k |t(k)|^2}{\max_k |t(k)|^2} \right) \quad (20)$$

여기서 $t(k)$ 는 반향제거 필터추정에 대해서는 $w(k) * h(k)$, 반향 필터 추정에 대해서는 $h_{\text{inv}}(k) * h(k)$ 이다. $h(k)$ 는 실제 반향 필터를 의미한다. 반향이 없는 원래 음성신호 $s(n)$ 과 복원된 반향제거 신호 $u(n)$ 사이의 차이로 잔존반향성분을 추정하여 신호 대 반향잡음비는 식 (21)과 같이 나타낼 수 있다.

$$\text{SRR (dB)} = 10 \log \left(\frac{\sum_n s^2(n)}{\sum_n (s(n) - u(n))^2} \right) \quad (21)$$

음성반향제거에서는 심별 간 간섭이 낮거나 신호 대 반향잡음비가 높을수록 성능이 더 좋다고 할 수 있다.

반향음성신호를 생성하기 위해 <그림 2>에서 나타내었듯이 이미지 방법(image method)[20]으로 직육면체 방을 가정하고 두 지점 사이에서의 반향 필터를 생성하였다. 방의 크기는 $5\text{m} \times 4\text{m} \times 3\text{m}$ 로, 화자와 마이크 간의 거리는 1m로 설정하였다. 반향시간 RT_{60} 가 0.3초가 되도록 반사계수 값을 결정하였다. 보다 실제적인 환경을 고려하여 반향이 없는 음원신호를 반향 필터에 통과시켜서 반향음성신호를 생성한 후에 신

5) 화자 리스트 : MDAC0, MEDR0, MGRL0, MJWT0, MKLS0, MMGG0, MPGH0, MRAI0, MTJS0, MWAD0.

6) 화자 리스트 : MCPM0, MDPK0, MJEB1, MKLW0, MMRP0, MPGR0, MPSW0, MRDD0, MRSO0, MRWS0.

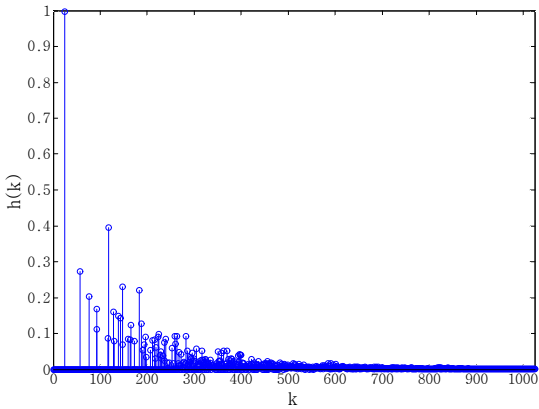


그림 2. 이미지 방법으로 생성한 음향반향 필터.

Figure 2. A reverberation filter generated by the image method.

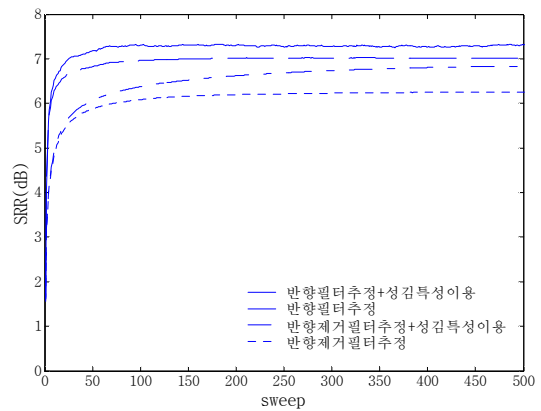


그림 4. 10명의 화자 음성 데이터에 대한 신호 대 반향잡음비의 평균.

Figure 4. SRRs averaged over 10-speaker data.

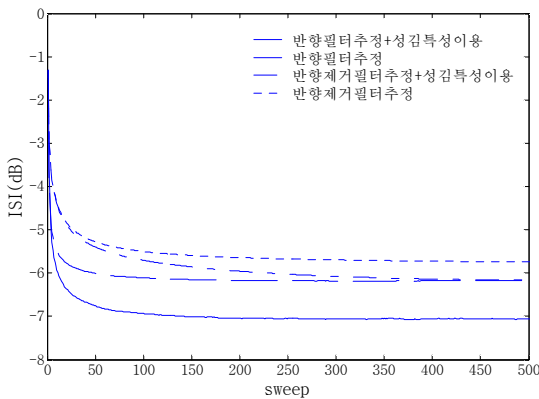


그림 3. 10명의 화자 음성 데이터에 대한 심벌 간 간섭의 평균.

Figure 3. ISIs averaged over 10-speaker data.

호 대 잡음비가 20dB가 되도록 백색정규잡음을 가산해 주었다.

<그림 3, 4>는 반향이 없는 원래 음성신호를 <그림 2>의 반향 필터에 통과시켜 얻은 가산 잡음 하 반향음성신호에 대한 실험 결과를 10명의 화자 음성 데이터에 대해서 평균한 결과이다. 심벌 간 간섭, 신호 대 반향잡음비 모두 반향 필터를 추정하는 방법이 반향제거 필터를 추정하는 방법보다, 그리고 필터의 성감 특성을 이용한 방법이 그렇지 않은 방법보다 더 좋은 성능을 나타내었다. 무엇보다도 필터의 성감 특성을 이용한 반향 필터 추정방법이 가장 좋은 성능을 나타냄을 알 수 있다.

6. 결론

본 논문에서는 하나의 마이크로 녹음된 음성신호로부터 반향성분을 효과적으로 제거하기 위하여 반향 필터 추정에 있

어서 성감 특성을 적용하는 방법을 제안하였다. 음향 환경에서는 일반적으로 반향 필터 자체도 매우 길고 복잡하다. 따라서 이웃한 표본 사이에 상관성이 있고 비정상적인 음성을 이용하여 반향 필터를 추정한다면, 한정된 개수의 관측값으로 그 필터를 정확히 추정하기 어렵다. 이런 문제점을 개선하기 위해서 선형적 지식에 기반하여 미리 가정이 가능한 필터의 특성을 고려하고자, 반향 필터에서 크기가 큰 계수가 드물게 나타나는 성감 특성을 이용하여 필터를 정확히 추정하는 방법을 제안하였다. 반향이 없는 원래 음성을 반향 필터에 통과시켜서 얻은 반향 신호에 대하여 모의실험을 한 결과, 성감 특성을 이용한 반향 필터 추정 방법이 비교 방법보다 더 좋은 성능을 나타내었다. 향후 본 논문에서 제안한 음성반향제거 방법을 다양한 음향반향 필터 및 음성 데이터에 대해서 검증하고자 한다.

참고문헌

- [1] Haas, H. (1972). The influence of a single echo on the audibility of speech. *Journal of the Audio Engineering Society*, 20(2), 146-159.
- [2] Furui, S. (2001). *Digital speech processing, synthesis, and recognition*. Marcel Dekker.
- [3] Lambert, R. H. & Bell, A. J. (1997). Blind separation of multiple speakers in a multipath environment. *Proc. IEEE ICASSP*, 423-426.
- [4] Douglas, S. C., Sawada, H., & Makino, S. (2005). Natural gradient multichannel blind deconvolution and speech separation using causal FIR filters. *IEEE Transactions on Speech and Audio Processing*, 13, 92-104.
- [5] Hopgood, J. R. & Rayner, P. J. W. (2003). Blind single

- channel deconvolution using nonstationary signal processing. *IEEE Transactions on Speech and Audio Processing*, 11, 476-488.
- [6] Nakatani, T., Kinoshita, K., & Miyoshi, M. (2007). Harmonicity-based blind dereverberation for single-channel speech signals. *IEEE Transactions on Audio, Speech, and Language Processing*, 15, 80-95.
- [7] Furuya, K. & Kataoka, A. (2007). Robust speech dereverberation using multichannel blind deconvolution with spectral subtraction. *IEEE Transactions on Audio, Speech, and Language Processing*, 15, 1579-1591.
- [8] Miyoshi, M., Delcroix, M., & Kinoshita, K. (2008). Robust speech dereverberation using multichannel blind deconvolution with spectral subtraction. *IEICE Transactions on Fundamentals*, E91-A(6), 1579-1591.
- [9] Kokkinakis, K. & Nandi, A. K. (2006). Multichannel blind deconvolution for source separation in convolutive mixtures of speech. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(1), 200-212.
- [10] Zee, M.-S. & Park, H.-M. (2009). Speech dereverberation based on blind estimation of a reverberation filter. *IEICE Electronics Express*, 6(20), 1456-1461.
- [11] Hyvärinen, A. & Raju, K. (2002). Imposing sparsity on the mixing matrix in independent component analysis. *Neurocomputing*, 49, 151-162.
- [12] Park, H.-M., Oh, S.-H., & Lee, S.-Y. (2006). Blind deconvolution with sparse priors on the deconvolution filters. *Proc. International Conference on Independent Component Analysis and Blind Signal Separation*, 658-665.
- [13] Lambert, R. H. (1996). *Multichannel blind deconvolution: FIR matrix algebra and separation of multipath mixtures*. Ph.D. Dissertation, University of Southern California, Los Angeles.
- [14] Torrkola, K. (1997). Blind deconvolution, information maximization and recursive filters. *Proc. IEEE ICASSP*, 3301-3304.
- [15] Lewicki, M. S. & Sejnowski, T. J. (2000). Learning overcomplete representations. *Neural Computation*, 12, 337-365.
- [16] Radlović, B. D. & Kennedy, R. A. (2000). Nonminimum-phase equalization and its subjective importance in room acoustics. *IEEE Transactions on Speech and Audio Processing*, 8(6), 476-488.
- [17] Mourjopoulos, J. (1985). On the variation and invertibility of room impulse response functions. *Journal of Sound and Vibration*, 102(2), 217-228.
- [18] Garofolo, J. S., Lamel, L. F., Fisher, W. M., Fiscus, J. G., Pallett, D. S., & Dahlgren, N. L. (1993). *DARPA TIMIT acoustic-phonetic continuous speech corpus*. Gaithersburgh, MD: Tech. Rep. NISTIR 4930.
- [19] Sklar, B. (Ed.). (1988). *Digital communications (fundamentals and applications)*. Englewood Cliffs, NJ: Prentice-Hall.
- [20] Allen, J. B. & Berkley, D. A. (1979). Image method for efficiently simulating small-room acoustics. *Journal of the Acoustical Society of America*, 65(4), 943-950.

• **지민선(Zee, Min-Seon)**

서강대학교 전자공학과
서울시 마포구 백범로 35 (신수동)
Tel: 02-711-8916
Email: minseon.zee@gmail.com
관심분야: 음성신호처리, 음향반향제거, 적응필터 등

• **박형민(Park, Hyung-Min), 교신저자**

서강대학교 전자공학과
서울시 마포구 백범로 35 (신수동)
Tel: 02-705-8916
Email: hpark@sogang.ac.kr
관심분야: 음성신호처리, 음성인식, 잡음제거 등
2011 ~ 현재 서강대학교 부교수