한국음향학회지 제31권 제1호 pp. 29~38 (2012)
The Journal of the Acoustical Society of Korea Vol.31, No.1 (2012)

장구간 예측 필터를 이용한 음성 신호에서의 돌발 잡음 제거

Transient Noise Reduction in Speech Signal Utilizing a Long-term Predictor

최민석 · 강홍구

(Min-Seok Choi and Hong-Goo Kang)

연세대학교 전기전자공학과

(접수일자: 2011년 9월 29일; 채택일자: 2011년 11월 23일)

초 록: 본 논문에서는 음성 신호에 더해진 돌발 잡음을 제거하는 시스템을 제안한다. 제안한 돌발 잡음 제거 시스템은 중앙값 필터를 이용하여 돌발 잡음을 제거한다. 중앙값 필터는 잡음을 제거하는 과정에서 음성을 왜곡시킬 수 있기 때문에, 음성의 왜곡을 최소화하기 위하여 장구간 예측 필터를 전처리단으로 사용한다. 장구간 예측 필터로 보존된 음성 정보는 잡음이 제거된 후 다시 합성된다. 본 논문에서는 돌발 잡음이 존재하는 환경에서 음성의 정보를 보존하는데 있어 단구간 예측 필터의 문제점을 밝히고 장구간 예측 필터의 우수함을 보인다. 제안한 돌발 잡음 제거 시스템의 출력 신호는 입력 신호에 비해 음성이 존재하는 구간에서 신호 대 잡음비가 약 8dB 향상 되었으며, PESQ 점수가 약 1점 증가하였다.

핵심용어: 음성 신호 처리, 돌발 잡음, 돌발 잡음 제거, 음질 개선
투고분야: 음성처리 분야 (2.3)

ABSTRACT: This paper presents a transient noise reduction system in a speech signal. The proposed transient noise reduction system utilizes a median filter to reduce the transient noise. Since the median filter can distort speech during the noise reduction, a long-term prediction (LTP) filter is adopted as a pre-processor to minimize speech distortion. The speech information preserved by the LTP filter is re-synthesized after reducing the noise. This paper verifies the weakness of a linear prediction (LP) filter and the superiority of the LTP filter for preserving the speech component in transient noise presence environment. Applying the proposed system, the signal-to-noise ratio (SNR) of output is improved by 8dB in both speech and noise presence region, and PESQ score is increased by 1 point comparing with noisy input.

Key words: Speech signal processing, Transient noise, Transient noise reduction, Speech enhancement
ASK subject classification: Speech Signal Processing (2.3)

I. 서 론

음성 신호에 더해진 잡음을 제거하는 것은 음성을 사용하는 어플리케이션의 품질을 결정하는데 중요한 역할을 한다. 지금까지 대부분의 잡음 제거 알고리즘들은 정적 잡음이나 느리게 변화하는 잡음을 가정하여 개발되어 왔으며, 돌발 잡음과 같이 빠르게 변화하면서 예측할 수 없는 잡음 환경은 고려되지 않았다[1-3]. 돌발 잡음의 위와 같은 특성으로 인해, 음질 향상 시스템에서 일반적으로 쓰이는 선형적 필터링 기법으로는 쉽게 제거할 수 없다[4-9].

돌발 잡음은 일반적으로 녹음기나 휴대 전화와 같이 음성을 입력 받는 기기의 본체를 두드리거나 그 근처의 사물을 두드림으로써 발생한다. 또한, 녹음

***Corresponding author:** 최민석 (zzugie@dsp.yonsei.ac.kr)
서울시 서대문구 신촌동 연세대학교 제2공학관 B601호
(전화: 02-2123-4534; 팩스: 02-364-4870)

이나 통화 도중에 버튼을 조작한다거나 키보드를 두드리는 것도 돌발 잡음의 원인이 된다. 돌발 잡음은 중앙값 필터 (median filter)나 크기 제한 필터와 같은 비선형 필터들을 이용하여 쉽게 제거할 수 있는데, 이 중에서 중앙값 필터는 입력 신호에서 느리게 변화하는 성분은 보존한 채 빠르게 변화하는 성분만을 제거하기 때문에 돌발 잡음을 제거하는데 효과적일 뿐만 아니라 목적 신호 성분을 일부분 보존할 수 있는 장점이 있어 널리 쓰인다[4-9]. 하지만, 음성의 피치 (pitch) 성분은 시간축에서 돌발 잡음과 유사한 특성을 가지므로, 음성 신호가 중앙값 필터를 통과하는 경우 신호의 특성이 크게 손상되어 심각한 음질 저하의 원인이 된다[10].

중앙값 필터는 돌발 잡음을 제거하는 성능이 뛰어나고, 별도의 파라미터 설정 없이 입력 신호의 특성에 따라 적절한 출력 값을 선택하기 때문에 안정적이고 좋은 성능을 갖는다[6]. 하지만, 음성 신호의 경우 시간축에서 파형이 빠르게 변화하는 경향이 있어 중앙값 필터에 의해 그 특성이 손상될 가능성이 있다. 따라서, 음성 신호의 정보를 보존하기 위한 별도의 장치가 필요한데, 음성을 모델링하는 필터로는 단구간 선형 예측 (LP: linear prediction) 필터와 장구간 예측 (LTP: long-term prediction) 필터가 가장 대표적으로 쓰인다[10,11]. 지금까지 대부분의 돌발 잡음 제거 시스템에서는 음성을 모델링하여 보존하는 전처리단으로써 단구간 LP 필터를 주로 사용하였다[4,7,8,12]. 하지만, 단구간 LP 필터는 필터 계수를 추정하는 과정에서 상대적으로 큰 에너지를 갖는 돌발 잡음의 영향을 받기 때문에 음성을 효과적으로 모델링하지 못하고 오히려 돌발 잡음의 특성을 쫓는 문제가 있다. 이와 같은 문제는 결국 돌발 잡음 제거 시스템의 출력 신호에서 잔여 잡음의 양이 증가하는 원인이 된다. 반면, LTP 필터는 음성 신호의 장구간 특성인 주기성을 이용하기 때문에 지속 시간이 짧은 돌발 잡음의 영향을 적게 받고, 중앙값 필터에 의해 가장 많이 손상되는 음성 성분인 피치를 보존하는데 효과적이다[5,10,11].

본 논문은 LTP 필터를 전처리단으로 사용하고 중앙값 필터로 잡음을 제거하는 돌발 잡음 제거 시스템을 제안한다. 먼저, 지금까지 흔히 사용되었던 단구간 LP 필터가 돌발 잡음이 존재하는 환경에서 음성을 모델링하기에 적합하지 않음을 밝히고, 실험을 통해 이를 증명한다. 또한, 돌발 잡음 환경에서 LTP 필터가 효과적으로 음성 정보를 모델링하는 것을 보이고, 이를 이용한 돌발 잡음 제거기를 제안하고 그 성능을 보인다.

본 논문의 구성은 다음과 같다. 먼저, II 절에서 중앙값 필터를 이용한 돌발 잡음 제거 알고리즘을 간략히 서술한다[6]. III 절에선 단구간 LP 필터가 돌발 잡음 제거 시스템의 전처리단으로써 적합하지 않음을 밝힌다. 이어서, IV 절에선 제안한 돌발 잡음 제거 시스템의 구조와 알고리즘을 서술하고, V절에서 그 성능을 평가한다. 마지막으로 VI 절에서 논문을 요약하고 결론을 맺도록 한다.

II. 중앙값 필터를 이용한 돌발 잡음 제거 알고리즘

돌발 잡음은 임의의 시간에 나타나며, 예측할 수 없는 충격 응답 (impulse response)을 갖는다. 돌발 잡음 $d(n)$은 다음 식 (1)과 같이 모델링 할 수 있다[5].

$$d(n)=\sum_{k}\left(h_k(n)*\delta(n-T_k)\right)g_k(n). \tag{1}$$

식 (1)에서 $h_k(n)$과 $g_k(n)$은 각각 k 번째 돌발 잡음의 충격 응답과 크기 이득을 나타내며, T_k는 k 번째 돌발 잡음이 발생하는 시간을 나타낸다. 돌발 잡음이 발생하는 시간은 예측할 수 없으며, $h_k(n)$과 $g_k(n)$은 k에 따라 불규칙하게 변화하는 값이다. 돌발 잡음 제거 시스템의 입력 신호 $x(n)$은 음성 신호 $s(n)$에 식 (1)의 돌발 잡음이 더해진 형태로 정의할 수 있다.

$$x(n)=s(n)+d(n). \tag{2}$$

위의 식 (2)에서 돌발 잡음의 크기와 특성은 시간에 따라 빠르게 변화하므로 기존의 선형 필터로는 제거하는 것이 어렵다. 반면, 중앙값 필터는 입력 신호에서 빠르게 변화하는 성분을 제거하는 역할을 하기 때문에 돌발 잡음을 쉽게 제거할 수 있다[6]. 현재의 샘플을 중심으로 앞뒤 w 개의 샘플들의 중앙값

을 구하는 필터는 식 (3)과 같이 정의된다.

$$\mathrm{med}_w\left[x(n)\right] = \mathrm{median}\left[x(n-w), x(n-w+1), \cdots, x(n), x(n+1), \cdots, x(n+w-1), x(n+w)\right]. \quad (3)$$

결과적으로 $\mathrm{med}_w[\cdot]$ 필터의 길이는 총 $2w+1$ 샘플이다. 중앙값 필터의 출력은 입력 샘플들의 값들 중에 크기 서열이 중간 순위인 값으로 결정된다. 따라서, 중앙값 필터의 입력에 매우 큰 크기를 갖는 샘플이 존재하더라도 필터의 출력은 그 영향을 받지 않는다. 그러므로, 돌발 잡음이 존재하는 신호를 중앙값 필터에 통과시키는 경우, 필터가 돌발 잡음의 영향 없이 목적 신호의 유사 평균 값을 출력으로 결정하게 되어 잡음 성분을 효과적으로 제거할 수 있다. 하지만, 중앙값 필터에 의한 신호의 평탄화 효과로 인하여 필터의 출력 신호에서 잡음뿐만 아니라 목적 신호 역시 왜곡되는 문제가 있다. 이와 같은 문제는 목적 신호의 파형이 빠르게 변화할수록 심화된다.

일반적으로, 돌발 잡음 제거 시스템은 목적 신호의 왜곡을 최소화하기 위하여 돌발 잡음이 존재하는 구간에 대해서만 잡음 제거 작업을 수행하도록 한다 [6,7,8,9,13]. 돌발 잡음이 존재하는 구간을 알고 있다고 가정할 때, 돌발 잡음 제거 시스템의 출력 신호는 식 (4)와 같다.

$$y(n) = \begin{cases} x(n), & H_T(n) = 0 \\ \mathrm{med}_w\left[x(n)\right], & H_T(n) = 1. \end{cases} \quad (4)$$

식 (4)에서 $H_T(n)$은 n 번째 입력 샘플에 돌발 잡음이 존재하는지의 여부를 나타내는 지표로써, $H_T(n)$이 1일 때 돌발 잡음이 존재하는 것으로 정의한다. $y(n)$은 돌발 잡음 제거 시스템의 출력 신호로써 돌발 잡음이 존재하지 않는 구간에서 입력 신호와 같은 값이 되고, 잡음이 존재하는 구간에서 중앙값 필터의 출력과 같은 값이 된다. 돌발 잡음이 존재하는 구간은 입력 신호의 시간축 단구간 에너지나, 주파수축 단구간 에너지, 혹은 입력 신호의 상관 계수 등을 이용하여 검출할 수 있다 [4,6-9,12-14]. 음성 신호가 존재하는 환경에서 시간-주파수축 단구간 에너지를 사용한 돌발 잡음 검출 알고리즘을 사용할 경우 99.3 %의 검출 정확도를 나타냈으며, 1.49 %의 낮은 확률로 돌발 잡음이 존재하지 않는 구간을 돌발 잡음 구간으로 잘못 검출하였다 [14].

식 (4)의 돌발 잡음 제거 알고리즘은 목적 신호인 음성 신호가 왜곡되는 구간을 최소화하면서 돌발 잡음을 효과적으로 제거할 수 있다. 하지만, 돌발 잡음과 음성 신호가 함께 존재하는 구간에서는 음성의 왜곡을 피할 수 없다. 돌발 잡음 제거기에 의한 음성의 왜곡을 더 줄이기 위해서 중앙값 필터의 앞 단에서 음성의 정보를 모델링하여 저장하는 단계와 중앙값 필터로 잡음을 제거한 후에 모델링 된 음성 정보를 복원하는 단계를 추가할 필요가 있다. 음성의 특성을 모델링 하기 위하여 가장 널리 쓰이는 알고리즘은 단구간 LP 필터와 LTP 필터이다 [10,11]. 지금까지 대부분의 돌발 잡음 제거 알고리즘에서는 음성 신호의 정보를 보존하기 위하여 단구간 LP 필터를 사용하였다 [4,7,8,12].

III. 돌발 잡음 환경에서의 단구간 선형 예측 (LP) 필터

단구간 LP 필터는 현재의 음성 샘플을 이전 샘플들의 선형 조합 (linear combination)으로 모델링 하는 기법이다. 즉, LP 필터로 추정된 음성 신호 $\tilde{x}(n)$은 아래 식 (5)와 같다 [10,11].

$$\tilde{x}(n) = \sum_{k=1}^{p} a_k x(n-k). \quad (5)$$

식 (5)에서 p는 LP 필터의 길이를 나타낸다. LP 필터의 계수 a_k는 주어진 프레임에서 추정한 음성 신호와 실제 음성 사이의 평균 오차를 최소화하도록 결정된다.

$$\tilde{r}(n) = x(n) - \tilde{x}(n) = x(n) - \sum_{k=1}^{p} a_k x(n-k). \quad (6)$$

식 (6)은 LP 필터의 잔여 신호를 나타낸다. LP 필터의 추정 오차는 잔여 신호 $\tilde{r}(n)$의 제곱의 평균으로 정의된다. 일반적으로 LP 필터는 음성 신호의 포

먼트 (formant)를 모델링 하기 때문에 식 (6)의 잔여 신호는 음성의 모음에 포함된 피치 성분과 자음의 백색 잡음 성분으로 구성된다.

돌발 잡음은 식 (1)과 같이 특정한 충격 응답을 잡음의 특성으로 갖는다. 일반적으로, 돌발 잡음의 충격 응답은 약 1~5 ms 길이의 초기 펄스를 가지며, 그 뒤에 작은 에너지를 갖는 진동 구간이 최대 50 ms 가량 지속된다 [5]. 식 (5)와 같이 음성 신호를 모델링하기 위한 단구간 LP 필터는 일반적으로 5 ms 미만의 길이를 갖도록 설계하며, 단구간 LP 필터의 계수는 약 20~30 ms 정도 길이의 프레임 단위로 추정한다 [10,11]. 즉, LP 필터는 짧은 필터 길이로 단구간에서의 신호 특성을 추정한다. 따라서, 돌발 잡음이 존재하는 구간에서 LP 필터의 계수를 추정하는 경우 필터 계수는 상대적으로 에너지가 큰 돌발 잡음의 충격 응답을 모델링 한다. 결과적으로, 단구간 LP 필터는 음성의 정보뿐만 아니라 돌발 잡음의 정보도 함께 저장하게 되어 잡음을 제거한 후 음성을 재합성 할 때 잔여 잡음의 양이 증가하는 원인이 된다.

그림 1은 돌발 잡음이 더해진 음성 신호를 LP 필터로 모델링한 후의 잔여 신호와 잔여 신호에서 돌발 잡음을 제거한 후 음성을 다시 합성한 결과를 나타낸다. 실험에 사용한 입력 신호는 깨끗한 음성에 실제 환경에서 녹음한 돌발 잡음을 더해서 만들었으며, 음성을 모델링하기 위한 LP 필터의 차수는 16차, 그리고 LP 필터의 계수는 15 ms 길이의 프레임 단위로 추정하였다. 그림 1 (a)는 깨끗한 음성 신호의 파형을 나타내며, 그림 1 (b)는 돌발 잡음이 포함된 입력 신호의 파형, 그리고 그림 1 (c)는 LP 필터로 음성을 모델링 한 후의 잔여 신호를 그린 것이다. 그림 1 (c)에 나타난 것과 같은 잔여 신호에서 돌발 잡음을 완전히 제거하면 그림 1 (d)와 같은 잡음이 제거된 잔여 신호를 얻을 수 있다. 마지막으로 잡음이 제거된 잔여 신호와 LP 필터의 계수를 이용하여 음성 신호를 다시 합성한 결과는 그림 1 (e)와 같다. 먼저, 그림 1 (b)와 그림 1 (c)를 비교하면 LP 필터의 잔여 신호에 포함되어있는 돌발 잡음 성분이 입력 신호에 더해진 돌발 잡음에 비해 에너지가 작고 파형이 변화한 것을 확인할 수 있다. 특히, 돌발 잡음에서 초기 펄스를 제외한 뒷부분의 크기가 크게 감쇄된다. 이는 돌발 잡음의 일부가 단구간 LP 필터에 의해 모델링 되어 저장된 것을 의미한다. 따라서, 그림 1 (d)와 같이 잔여 신호에서 돌발 잡음을 완전히 제거하더라도 LP 필터로 음성을 재합성 하면 그림 1 (e)와 같이 잡음 성분이 함께 복원된다.

그림 1에서 볼 수 있듯이 돌발 잡음이 존재하는 환경에서 단구간 LP 필터를 이용하여 음성을 모델링 할 경우 돌발 잡음 제거 시스템의 출력에서 잔여 잡음이 증가하는 문제가 있다. 또한, 단구간 LP 필터는 음성의 피치를 모델링할 수 없기 때문에 LP 필터의 잔여 신호를 중앙값 필터로 처리하면 그림 1 (d)와 같이 피치가 손상되어 음성의 음질이 저하된다. 따라서, 돌발 잡음을 제거하는 과정에서 잔여 잡음을 최소화하고 음성 손실을 방지하기 위한 전처리단으로써 단구간 LP 필터를 사용하는 것 보다 입력 신호의 장구간 특성을 이용하여 음성의 피치를 모델링 하는 방법인 LTP 필터를 사용하는 것이 적합하다 [10,11].

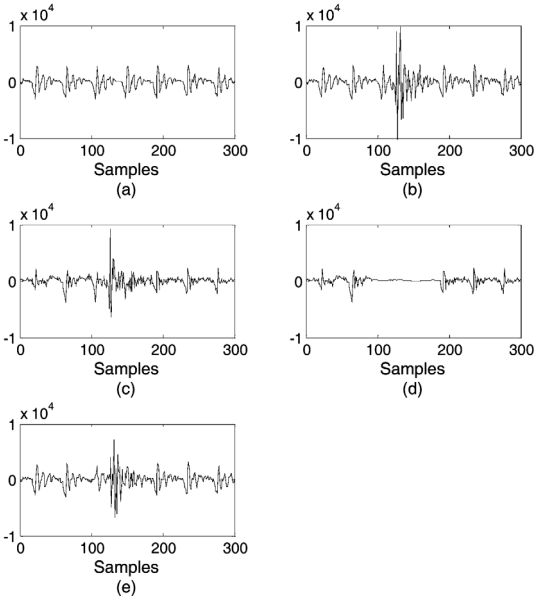

그림 1. 단구간 LP 필터를 이용한 돌발 잡음 제거 실험 결과 (a): 목적 음성 신호, (b): 입력 신호, (c): LP 필터의 잔여 신호, (d): 중간값 필터 출력 신호, (e): 최종 출력 신호

Fig. 1. Transient noise reduction result using short-term LP filter.
(a): desired speech signal, (b): input signal, (c): residual signal of LP filter, (d): output signal of median filter, (e): final output signal

IV. 제안한 돌발 잡음 제거 시스템

본 논문은 음성과 돌발 잡음이 함께 존재하는 환경에서 음성의 왜곡을 최소화 한 채 돌발 잡음을 효과적으로 제거하기 위한 시스템을 제안한다. 제안한 시스템은 잡음을 제거하기 위하여 중앙값 필터를 사용하는데, 중앙값 필터는 돌발 잡음을 제거하는데 가장 효과적인 기법 중 하나이지만 음성의 피치와 같이 시간 축에서 빠르게 변화하는 특성을 갖는 신호를 크게 손상시키는 문제가 있다 [6]. 따라서, 중앙값 필터의 앞 단에 음성의 정보를 모델링하여 저장하는 필터를 사용하여 음성 신호의 왜곡을 줄여야 한다.

음성을 모델링하기 위한 대표적인 필터에는 LP 필터와 LTP 필터가 있는데, LP 필터는 음성을 모델링하는 과정에서 돌발 잡음의 특성을 함께 저장하기 때문에 잔여 잡음의 원인이 된다. 반면, LTP 필터는 음성 신호의 주기성을 이용하기 때문에 돌발 잡음의 영향을 상대적으로 적게 받을 뿐만 아니라 중앙값 필터에 의해 가장 크게 손상되는 음성의 피치를 모델링하는데 적합하다 [10,11]. 본 논문에서는 LTP 필터를 전처리단으로 사용하는 돌발 잡음 제거 시스템을 제안하고, 그 우수성을 증명한다.

4.1 돌발 잡음 환경에서의 장구간 예측 (LTP) 필터

LTP 필터는 음성 신호의 피치를 이전의 피치 성분으로 모델링하는 필터이다 [10]. 음성의 피치는 주기성을 가지며 인근에 위치한 피치와 매우 높은 상관도를 갖기 때문에, 적은 탭 수의 필터로도 쉽게 모델링할 수 있다. 일반적으로 LTP 필터는 LP 필터의 잔여 신호를 입력으로 받지만, 돌발 잡음이 존재하는 환경에서는 LP 필터가 잔여 잡음의 원인이 되므로 제안한 시스템은 입력 신호에 LTP 필터를 바로 사용하도록 한다 [10,11]. LTP 필터에서 현재 프레임의 음성 신호는 아래의 식 (7)과 같이 추정한다.

$$\hat{x}(m,l) = g_p(l)\,x\big(m-\tau_p(l),l\big), \quad \text{where } 0 \le m \le M-1. \tag{7}$$

식 (7)에서 l과 m은 각각 프레임 인덱스와 프레임 내에서의 샘플의 인덱스를 나타내며, M은 한 프레임의 길이를 정의한다. 즉, (m,l)은 l번째 프레임의 m번째 샘플을 의미하므로 전체 입력 신호에서 $(m+(l-1)M)$번째 샘플을 가리킨다. 식 (7)의 $\tau_p(l)$과 $g_p(l)$은 각각 l번째 프레임에서 추정한 이전 피치까지의 시간 지연 값과 피치를 추정하기 위한 크기 이득 값이다. LTP 필터는 $\tau_p(l)$ 샘플 이전의 음성 샘플에 $g_p(l)$ 만큼의 크기 이득을 곱하여 현재 음성 샘플을 추정한다. LTP 필터에 의해 추정된 음성 신호는 $\hat{x}(m,l)$로 정의한다.

LTP 필터에서의 피치 주기 $\tau_p(l)$은 각 프레임에서 상관도가 최대가 되는 시간 지연 값으로 추정하며,

$$\tau_p(l) = \arg\max_{\tau_{\min}\le\tau\le\tau_{\max}} \frac{\sum_{m=0}^{M-1} x(m,l)\,x(m-\tau,l)}{\sqrt{\sum_{m=0}^{M-1} x^2(m-\tau,l)}}, \tag{8}$$

추정된 피치 주기를 기반으로 피치 추정 오차가 최소가 되는 크기 이득 $g_p(l)$을 아래 식 (9)와 같이 추정한다.

$$\hat{g}_p(l) = \frac{\sum_{m=0}^{M-1} x(m,l)\,x\big(m-\tau_p(l),l\big)}{\sqrt{\sum_{m=0}^{M-1} x^2\big(m-\tau_p(l),l\big)}},$$
$$g_p(l) = \begin{cases} \hat{g}_p(l), & \hat{g}_p(l) < g_{p\max} \\ g_{p\max}, & \text{otherwise.} \end{cases} \tag{9}$$

최적의 피치 주기 $\tau_p(l)$은 일반적인 사람의 피치 주기를 넘지 않는 범위 내에서 검색한다. 본 논문에선 식 (8)의 $\tau_{\min}$을 2.5 ms로, $\tau_{\max}$를 18 ms로 설정하였다. LTP 필터의 크기 이득 또한 필터가 추정 오차에 강인하도록 하기 위해 식 (9)와 같이 일정 상수 값, $g_{p\max}$,을 넘지 않도록 제한하는 것이 보통인데, 이는 돌발 잡음이 존재하는 환경에서 돌발 잡음에 의하여 피치가 잘못 추정되는 문제를 완화하는데 역시 도움이 된다 [11]. 즉, 돌발 잡음이 존재하는 구간에서 잡음의 큰 에너지 때문에 피치 추정을 위한 크기 이득이 매우 큰 값으로 잘못 결정 될 수 있는데,

크기 이득의 최대 값을 제한함으로써 이와 같은 문제가 방지된다. LTP 필터의 잔여 신호는 아래의 식 (10)과 같이 정의한다.

$$r(m,l) = x(m,l) - \hat{x}(m,l). \tag{10}$$

그림 2는 돌발 잡음이 존재하는 환경에서 LTP 필터가 모델링한 음성 신호와 LTP 필터의 잔여 신호를 그린 것이다. 그림 2 (a)와 그림 2 (b)는 각각 깨끗한 음성 신호와 돌발 잡음이 더해진 입력 신호를 나타낸다. 그림 2 (c)는 LTP 필터에 의해 모델링 된 음성의 피치 성분을 시간에 따라 나열한 것이다. 그림 2 (a)와 그림 2 (c)의 파형을 비교하면 LTP 필터가 돌발 잡음이 존재하는 환경에서도 목적 음성을 매우 효과적으로 모델링 하는 것을 확인할 수 있다. 그림 2 (d)는 LTP 필터의 잔여 신호인데, 그림 1 (c)의 잔여 신호와 달리 입력 돌발 잡음과 매우 유사한 형태를 가지며 음성의 정보가 거의 남아있지 않다. 이는 LTP 필터가 돌발 잡음의 영향을 받지 않고 음성 신호의 정보만을 추출했다는 것을 의미한다. 제안한 돌발 잡음 제거 시스템은 그림 2 (d)와 같은 잔여 신호 $r(m,l)$에서 돌발 잡음을 제거한 후, 그림 2 (c)의 음성 정보를 다시 합성한다.

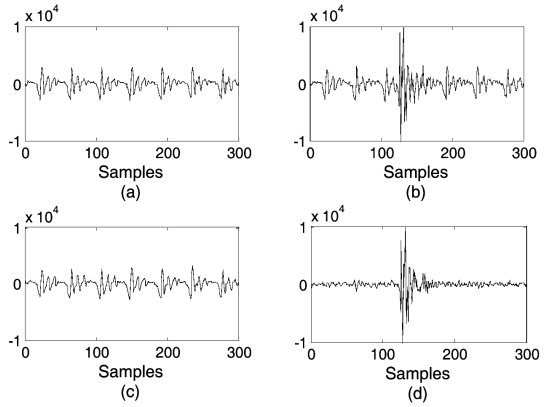

그림 2. 돌발 잡음 환경에서의 LTP 필터의 성능 평가
(a): 목적 음성 신호, (b): 입력 신호, (c): LTP 필터가 모델링 한 음성 신호, (d): LTP 필터의 잔여 신호
Fig. 2. Performance evaluation of LTP filter in transient noise environment.
(a): desired speech signal, (b): input signal, (c): speech signal estimated by LTP filter, (d): residual signal of LTP filter

4.2 장구간 예측 (LTP) 필터와 중앙값 필터를 이용한 돌발 잡음 제거 시스템

제안한 돌발 잡음 제거 시스템의 블록도는 그림 3과 같다.

제안된 시스템은 중앙값 필터를 이용하여 LTP 필터의 잔여 신호 $r(m,l)$에서 돌발 잡음을 제거한다. 이 과정에서 발생할 수 있는 음성의 왜곡을 최소화하기 위하여 식 (4)와 같이 돌발 잡음이 존재하는 구간에만 중앙값 필터를 적용한다. 잡음을 제거한 후에 재합성 된 음성 신호를 $\hat{y}(m,l)$로 정의하면, 제안된 시스템의 출력은 아래의 식 (11)과 같다.

$$y(m,l) = \begin{cases} x(m,l), & H_T(m,l) = 0 \\ \hat{y}(m,l), & H_T(m,l) = 1. \end{cases} \tag{11}$$

즉, 제안된 시스템의 출력 신호 $y(m,l)$은 돌발 잡음이 존재하지 않는 구간에서는 입력 신호와 같은 값을 갖고, 돌발 잡음이 존재하는 구간에서는 LTP 필터와 중앙값 필터로 처리한 출력을 받는다. 식 (11)에서 $\hat{y}(m,l)$은 잡음이 제거된 잔여 신호 $\hat{r}(m,l)$과 식 (7)에서 LTP 필터가 모델링한 음성 신호 $\hat{x}(m,l)$의 합으로 구할 수 있다.

$$\hat{y}(m,l) = \hat{r}(m,l) + \hat{x}(m,l). \tag{12}$$

또한, $\hat{r}(m,l)$은 다음 식 (13)과 같이 정의된다.

$$\hat{r}(m,l) = \mathrm{med}_w\left[r(m,l)\right]. \tag{13}$$

그림 3에서 볼 수 있듯이, 음성의 모델링 및 돌발 잡음 제거, 그리고 음성을 재합성하는 작업은 돌발

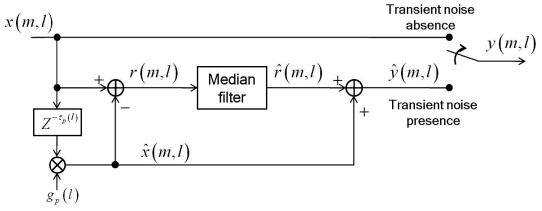

그림 3. LTP 필터를 전처리단으로 하는 돌발 잡음 제거 시스템
Fig. 3. Transient noise reduction system which utilize a LTP filter as a pre-processor.

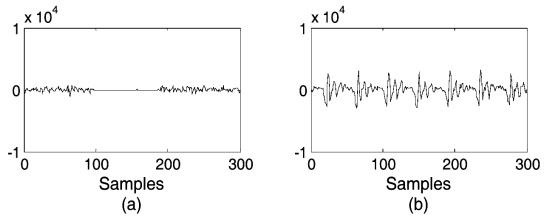

그림 4. LTP 필터를 이용한 돌발 잡음 제거 실험 결과 (a): 돌발 잡음이 제거된 잔여 신호, (b): 잡음을 제거 한 후 재합성된 음성 신호

Fig. 4. Transient noise reduction result using LTP filter. (a): residual signal after transient noise reduction, (b): re-synthesized speech signal

잡음이 존재하는 구간에만 필요하다. 다시 말해, 식 (7)의 LTP 필터를 이용한 음성의 모델링, 식 (13)의 중앙값 필터링, 그리고 식 (12)와 같은 음성의 재합성 과정은 모두 돌발 잡음이 존재하지 않는 경우에는 동작하지 않고 시스템의 입력을 그대로 출력으로 내보낸다.

그림 4는 그림 3 (d)에서 보인 LTP 필터의 잔여 신호에 중앙값 필터를 적용하여 돌발 잡음을 제거한 후, 그림 3 (c)와 같이 LTP 필터가 모델링한 음성 신호의 정보를 재합성 한 결과를 나타낸다. 그림 3 (d)의 신호에서 돌발 잡음을 제거한 결과는 그림 4 (a)와 같고, 그림 4 (a)의 신호와 LTP 필터가 모델링한 그림 3 (c)의 신호를 이용하여 음성 신호를 복원한 결과가 그림 4 (b)와 같다. 목적으로 하는 음성 신호, 그림 3 (a),와 비교하였을 때, 제안된 시스템이 음성의 왜곡이 거의 없이 돌발 잡음을 효과적으로 제거한 것을 확인할 수 있다.

V. 돌발 잡음 제거 성능 평가

제안한 시스템의 돌발 잡음 제거 성능을 평가하기 위하여 실제 환경에서 녹음한 데이터를 이용한 잡음 제거 실험을 진행하였다. 실험에 사용된 돌발 잡음은 휴대 전화의 녹음 기능을 켠 상태로 전화의 본체를 두드리거나 전화의 버튼을 조작하며 녹음하였다. 녹음한 잡음 데이터를 깨끗한 음성 신호의 임의의 위치에 더해서 입력 신호를 생성하였으며, 이때 사용한 음성 신호는 4명의 남성 화자와 4명의 여성 화자에 의해 발화된 총 8개의 문장이다. 각 문장의 길이는 약 2초 정도이며, 입력 신호의 샘플링 주파수는 8 kHz이다.

실험에 사용된 중앙값 필터의 길이는 8 kHz 샘플링된 신호를 기준으로 101 샘플이다. 중앙값 필터의 전처리단으로 LP 필터와 LTP 필터를 사용하며 음성 모델링 성능과 돌발 잡음 제거 성능을 비교하였다. LP 필터의 길이는 16탭이고, LP 필터의 계수를 추정하기 위한 프레임의 길이는 15 ms이다. LTP 필터는 1탭의 길이를 가지며, LTP 필터의 파라미터를 추정하기 위한 프레임의 길이는 15 ms이다. LTP 필터의 피치 주기 $\tau_p(l)$은 2.5 ms에서 18 ms 사이의 범위에서 추정하였으며, 식 (8)의 상관 계수를 3배로 오버샘플링 (oversampling) 하여 샘플링 주파수보다 3배 높은 정확도를 갖는 피치 주기를 찾도록 하였다. 이는 LTP 필터의 음성 추정 오차를 감소시키는 역할을 한다. LTP 필터의 크기 이득 $g_p(l)$의 최대 값 $g_{p\max}$는 1.2로 설정하였다[11]. 돌발 잡음 제거 실험은 돌발 잡음이 존재하는 구간을 정확하게 알고 있는 환경에서 진행하였다. 하지만, 시간-주파수축 단구간 에너지를 기반으로 하는 돌발 잡음 검출 알고리즘을 사용하여도 본 논문에서 보이는 실험 결과와 유사한 성능의 출력을 얻을 수 있다[14].

먼저, 돌발 잡음이 존재하는 환경에서 단구간 LP 필터와 LTP 필터의 음성 모델링 성능을 확인하기 위해 각 필터의 잔여 신호와 입력 잡음 사이의 상호 상관 계수를 측정하였다. 두 음성 모델링 필터들의 목적은 돌발 잡음을 잔여 신호에 남겨두고 음성 정보만을 추출하여 저장하는 것이기 때문에 입력된 돌발 잡음과의 상관 계수가 클수록 돌발 잡음 제거에 더 유리하다고 할 수 있다. 상호 상관 계수를 구하는 식은 아래 (14)와 같다[15].

$$\frac{\sum_n \{r(n)d(n)\}H_T(n)}{\left\{\sum_n |r(n)|H_T(n)\right\}\left\{\sum_n |d(n)|H_T(n)\right\}} \tag{14}$$

식 (14)에서 $r(n)$은 단구간 LP 필터 혹은 LTP 필터의 잔여 신호를 나타낸다. 상관 계수를 구하는 과정에서 돌발 잡음의 존재 여부 $H_T(n)$을 참조하여 돌발 잡음이 존재하는 구간에서만 상관 계수를 측정

표 1. LP 필터와 LTP 필터의 잔여 신호와 입력 잡음 사이의 상호 상관 계수
Table 1. Cross-correlation coefficient between residual signals of LP filter and LTP filter and input noise.

	LP 필터의 잔여 신호	LTP 필터의 잔여 신호
상관 계수	0.8267	0.9858

하였다. 표 1은 상호 상관 계수를 측정한 결과이다.

위의 표 1에 나타난 바와 같이 LTP 필터의 잔여 신호는 입력 돌발 잡음과 1에 가까운 상관 계수를 갖는다. 따라서, LTP 필터의 잔여 신호에는 대부분의 돌발 잡음 성분이 남아있고 음성 성분은 제거되었다고 볼 수 있다. 반면, LP 필터의 잔여 신호는 입력 돌발 잡음과 상대적으로 낮은 상관 계수를 갖는다.

제안된 돌발 잡음 제거기의 성능을 평가하기 위해 시스템의 출력과 목적으로 하는 음성 신호 사이의 신호 대 잡음비 (SNR: signal-to-noise ratio)와 로그 스펙트럼 오차 (LSD: log spectral distance)를 측정하였다. 각 성능을 측정하기 위한 식은 다음 식 (15)와 같다 [16].

$$SNR = 10\log_{10}\left(\frac{E_n\left\{s(n)^2\right\}}{E_n\left\{\left(s(n)-y(n)\right)^2\right\}}\right),$$
$$LSD = E_l\left\{\sqrt{E_f\left\{\left(20\log_{10}\frac{|S(f,l)|}{|Y(f,l)|}\right)^2\right\}}\right\}. \quad (15)$$

식 (15)에서 $E_n\{\cdot\}$은 샘플들의 시간축 평균을 의미하며, $E_f\{\cdot\}$는 주어진 주파수 응답의 한 프레임 내에서의 주파수축 평균, $E_l\{\cdot\}$은 프레임들 간의 평균을 의미한다. $S(f,l)$과 $Y(f,l)$은 각각 목적 음성 신호와 시스템의 출력 신호의 주파수 응답을 정의하며, f는 주파수 인덱스이다. 제안한 시스템의 잡음 제거 성능 및 음성 왜곡 정도를 정확하게 판단하기 위하여 SNR과 LSD는 돌발 잡음과 음성 신호가 모두 존재하는 구간에서만 측정하였다. SNR과 LSD를 측정한 결과는 표 2와 같다.

표 2의 첫번째 열은 돌발 잡음 제거 시스템에서 중앙값 필터의 전처리단으로 사용한 음성 모델링 필터의 종류를 나타낸다. LTP 필터의 우수성을 증명하기 위하여 전처리단으로 LP 필터만을 사용한 경우, LP 필터와 LTP 필터를 모두 사용한 경우, 그리고 제안한 시스템인 LTP 필터만을 사용한 경우의 출력 신호를 모두 비교하였다.

표 2. 시스템 출력 신호와 목적 음성 신호 사이의 SNR과 LSD 측정 결과
Table 2. SNR and LSD between system output and desired speech.

	SNR	LSD
입력 신호	-1.31	15.57
LP 필터	-0.04	15.50
LP 필터 + LTP 필터	-0.09	15.25
LTP 필터	7.94	9.07

표 3. 시스템 출력 신호의 PESQ 점수 측정 결과
Table 3. PESQ score of system output.

	PESQ 점수
입력 신호	1.86
LP 필터	2.08
LP 필터 + LTP 필터	2.07
LTP 필터	2.82

돌발 잡음은 일반적으로 에너지가 매우 크기 때문에 돌발 잡음이 존재하는 구간에서 입력 신호의 SNR이 매우 낮게 나타난다. 하지만, 제안된 돌발 잡음 제거 시스템으로 돌발 잡음을 제거할 경우 SNR이 8 dB 가량 향상 되었으며, LSD도 6 dB 이상 감소하였다. 반면 LP 필터가 전처리단에 포함된 시스템의 출력 신호의 경우에는 입력 신호에 비해 약간의 성능 향상은 있었지만 잔여 잡음이 발생하는 문제로 인하여 제안한 알고리즘보다 매우 적은 향상 폭을 나타내었다.

제안한 시스템의 출력 신호의 음질을 객관적으로 평가하기 위해 PESQ (perceptual evaluation of speech quality) 점수를 측정하였다 [17]. PESQ 점수는 표 3에 정리되어 있다.

표 2에 사용된 출력 신호와 같은 데이터를 이용하여 PESQ 점수를 측정하였다. 표 2의 결과와 마찬가지로 LP 필터를 전처리단으로 사용하는 경우에는 PESQ 점수의 향상 폭이 적다. 하지만, 제안된 시스템의 출력 신호는 입력 신호에 비해 PESQ 점수가 1점

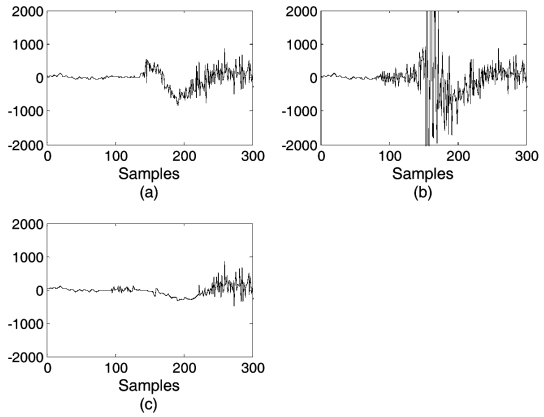

그림 5. 파열음 구간에서의 돌발 잡음 제거 실험 결과
(a): 목적 음성 신호, (b): 입력 신호, (c): 시스템의 출력 신호

Fig. 5. Transient noise reduction result in plosive sound region.
(a): desired speech, (b): input signal, (c): system output

가량 높았고, 인지적으로 평범한 수준의 음질을 의미하는 3점에 가까운 점수를 얻게 되었다 [17].

제안된 시스템은 음성의 주기성을 이용하기 때문에 자음 구간에서 음성을 보존하는 성능이 상대적으로 떨어진다. 즉, 음성의 자음이 돌발 잡음 제거 시스템에 의해 왜곡될 수 있다. 그림 5는 제안된 시스템의 자음 구간에서의 성능을 확인하기 위하여 음성의 자음 성분 중에 돌발 잡음과 특성이 가장 유사한 파열음 구간에 돌발 잡음을 더한 후 이를 제거한 실험의 결과이다.

그림 5 (a)는 목적으로 하는 음성 신호의 파형이며, 그림 5 (b)는 잡음이 포함된 입력 신호, 그리고 그림 5 (c)는 잡음을 제거한 후 LTP 필터를 이용하여 음성을 복원한 시스템의 최종 출력 신호를 나타낸다. 그림 5에서 100번째 샘플과 200번째 샘플 사이의 신호가 파열음 성분인데, 그림 5 (a)와 그림 5 (c)를 비교하면 제안된 시스템으로 잡음을 제거하는 과정에서 파열음 성분이 일부분 감쇄된 것을 확인할 수 있다. 하지만, 그림 5 (c)에 나타나듯이 출력 신호에서 파열음 성분의 일부가 보존되기 때문에 파열음 성분의 감쇄가 음성의 인지적 음질에는 크게 영향을 주지 않는다. 제안된 시스템의 출력 신호에서 파열음의 특성이 크게 왜곡 되지 않는 것은 목적 음성 신호와 출력 신호의 스펙트로그램을 비교함으로써 확연하게 확인할 수 있다. 아래 그림 6은 그림 5와 같은 구간에 대해 목적 음성 신호와 시스템의 출력 신호의 스펙트로그램을 그린 것이다.

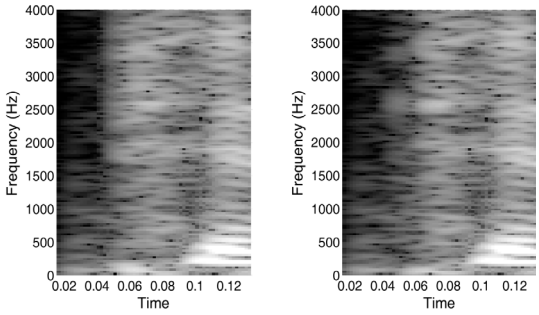

그림 6. 파열음 구간에서의 목적 음성 신호와 시스템 출력 신호의 스펙트로그램
(왼쪽: 목적 음성 신호, 오른쪽: 시스템의 출력 신호)

Fig. 6. Spectrograms of desired speech and system output in plosive sound region.
(Left: desired speech, right: system output)

그림 6의 왼쪽은 목적 신호의 스펙트로그램이고, 오른쪽은 출력 신호의 스펙트로그램이다. 그림에서 0.04초와 0.09초 사이의 구간이 파열음이 존재하는 구간인데, 출력 신호에서도 파열음의 특성이 크게 훼손되지 않는 것을 확인할 수 있다. 이와 같이 자음 성분이 보존되는 것은 입력 신호의 특성에 따라 출력 값이 변화하는 중앙값 필터의 특성 때문이다. 결론적으로, 제안한 시스템은 자음 구간에서 음성을 보존하는 성능이 저하되지만, 이로 인해 음성의 인지적 음질이나 명료도가 훼손되지는 않는다.

VI. 결 론

본 논문은 음성 신호가 존재하는 환경에서 음성의 왜곡을 최소화한 채 돌발 잡음을 효과적으로 제거하는 알고리즘을 제안하였다. 제안된 알고리즘은 별도의 파라미터를 설정할 필요 없이 돌발 잡음을 효과적으로 제거할 수 있는 중앙값 필터를 이용하여 잡음을 제거한다. 중앙값 필터는 입력 신호에서 빠르게 변화하는 성분은 제거하면서 입력 신호의 특성을 반영하여 적응적으로 필터의 출력이 변화하는 장점이 있다 [5]. 따라서, 중앙값 필터는 돌발 잡음을 제외한 목적 신호의 특성을 일부 보존하면서 잡음을 효과적으로 제거할 수 있다. 하지만, 목적 신호가 음성

신호인 경우에는 음성 신호에 피치 성분과 같이 시간축에서 빠르게 변화하는 성분이 존재하기 때문에 중앙값 필터가 그 특성을 크게 손상시키는 문제가 있다. 그러므로, 입력 신호에서 음성 성분을 추출하여 저장하고 잡음을 제거한 후 이를 복원하는 별도의 프로세스가 반드시 필요하다. 지금까지 대부분의 돌발 잡음 제거 시스템들은 입력 신호에서 음성 성분을 분리하기 위하여 단구간 LP 필터를 주로 사용하였지만, 단구간 LP 필터는 돌발 잡음의 특성도 함께 모델링하는 경향이 있기 때문에 잔여 잡음이 증가하는 원인이 된다 [4,7,8,10-12]. 반면, 제안된 시스템은 LTP 필터를 중앙값 필터의 전처리단으로 사용하는데, LTP 필터는 돌발 잡음의 영향을 적게 받으면서 음성의 피치를 효과적으로 모델링하여 돌발 잡음을 제거한 후 음성 신호만을 복원하는데 탁월한 성능을 보인다 [10,11]. 제안된 시스템을 통하여 출력 신호의 SNR이 입력 신호에 비해 8 dB 증가하였으며, LSD가 6 dB 낮아졌다. 또한, 객관적 음질 평가에서 시스템의 출력은 인지적으로 보통 수준의 음질을 나타내는 3점에 가까운 PESQ 점수를 얻었다.

참고문헌

1. S. F. Boll, “Suppression of acoustic noise in speech using spectral subtraction,” *IEEE Trans. on Acoust., Speech, and Signal Process.*, vol. 27, no. 2, pp. 113-120, 1979.
2. Y. Ephraim and D. Malah, “Speech enhancement using a minimum mean-square error log-spectral amplitude estimator,” *IEEE Trans. on Acoust., Speech, and Signal Process.*, vol. 33, no. 2, pp. 443-445, 1985.
3. P. C. Loizou, *Speech enhancement, Theory and practice*, CRC Press, 2007.
4. I. Cohen and B. Berdugo, “Speech enhancement for non-stationary noise environments,” *Signal Process.*, vol. 81, Issue 11, pp. 2401-4218, 2001.
5. S. V. Vaseghi, *Advanced digital signal processing and noise reduction*, 2nd ed., John Wiley & Sons, 2000.
6. T. Kasparis and J. Lane, “Suppression of impulsive disturbances from audio signals,” *Electronics letters*, vol. 29, no. 22, pp. 1926-1927, 1993.
7. A. J. Efron and H. Jeen, “Detection in impulsive noise based on robust whitening,” *IEEE Trans. on Signal Process.*, vol. 42, no. 6, pp. 1572-1576, 1994.
8. S. R. Kim and A. Efron, “Adaptive robust impulse noise filtering,” *IEEE Trans. on Signal Process.*, vol. 43, no. 8, pp. 1855-1866, 1995.
9. I. Kauppinen, “Methods for detecting impulsive noise in speech and audio signals,” *in Proc. IEEE Int. Conf. on Digital Signal Process. 2002*, vol. 2, pp. 967-970, 2002.
10. A. M. Kondoz, *Digital speech −coding for low bit rate communication systems*, John Wiley & Sons, 1994.
11. ITU-T, *ITU-T recommendation G. 729*, 1996.
12. R. Talmin, I. Cohen, and S. Gannot, “Speech enhancement in transient noise environment using diffusion filtering,” *in Proc. IEEE Int. Conf. on Acoust., Speech, Signal Process. 2010*, pp. 4782-4785, 2010.
13. J. Beh, K. Kim and H. Ko, “Noise estimation for robust speech enhancement in transient noise environment,” *KSCSP 2007*, vol. 24, no. 1, pp. 35-36, 2007.
14. M. S. Choi, H. S. Shin, and Y. S. Hwang, and H. G. Kang, “Time-frequency domain impulsive noise detection system in speech signal,” *The journal of the Acoust. Society of Korea*, vol. 30, no. 2, pp. 73-79, 2011.
15. A. Papoulis and S. U. Piillai, *Probability, random variables and stochastic processes, forth edition*, Mc Graw Hill, 2002.
16. J. Benesty, S. Makino, and J. Chen, *Speech enhancement*, Springer, 2005.
17. ITU-T, *ITU-T Recommendation P.862. Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs*, 2001.

저자 약력

▸ 최 민 석 (Min-Seok Choi)

2004년: 연세대학교 전기전자공학과 학사
2006년: 연세대학교 전기전자공학과 석사
2006년 ~ 현재: 연세대학교 전기전자공학과 박사과정 재학중

▸ 강 홍 구 (Hong-Goo Kang)

1989년: 연세대학교 전기전자공학과 학사
1991년: 연세대학교 전기전자공학과 석사
1995년: 연세대학교 전기전자공학과 박사
1996년 ~ 2002년: Senior Technical Staff Member, AT&T Labs-Research
2002년 ~ 2005년: 연세대학교 전기전자공학과 조교수
2005년 ~ 현재: 연세대학교 전기전자공학과 부교수