

음성 신호에서의 시간-주파수 축 충격 잡음 검출 시스템

Time-Frequency Domain Impulsive Noise Detection System in Speech Signal

최민석*, 신호선*, 황영수**, 강홍구*

(Min-Seok Choi*, Ho Seon Shin*, Young-Soo Hwang**, Hong-Goo Kang*)

*연세대학교 전기전자공학과, **관동대학교 전자정보통신공학부

(접수일자: 2010년 11월 29일; 수정일자: 2011년 1월 20일; 채택일자: 2011년 2월 1일)

본 논문에서는 음성 신호를 녹음하는 과정에서 발생하는 충격 잡음의 위치를 검출하는 새로운 알고리즘을 제안하였다. 제안한 방법은 충격 잡음의 주파수 축 특성을 반영하여 기존의 방법에 비해 높은 검출 정확도를 가지면서 음성의 피치를 충격 잡음과 구분하지 못하는 문제를 해결하였다. 또한, 시간 축, 주파수 축 파라미터의 단점을 상호 보완하여 false-alarm 문제를 최소화하는 시간-주파수 축 충격 잡음 검출 시스템을 제안하였다. 실제 녹음된 충격 잡음을 이용한 실험 결과, 제안한 시간-주파수 축 충격 잡음 검출기는 99.33 %의 가장 높은 검출 정확도와 1.49 %의 가장 낮은 false-alarm 비율을 나타내었다.

핵심용어: 음성 신호 처리, 충격 잡음, 충격 잡음 검출, 음질 개선

투고분야: 음성 처리 분야 (2,3)

This paper presents a new impulsive noise detection algorithm in speech signal. The proposed method employs the frequency domain characteristic of the impulsive noise to improve the detection accuracy while avoiding the false-alarm problem by the pitch of the speech signal. Furthermore, we proposed time-frequency domain impulsive noise detector that utilizes both the time and frequency domain parameters which minimizes the false-alarm problem by mutually complementing each other. As the result, the proposed time-frequency domain detector shows the best performance with 99.33 % of detection accuracy and 1.49 % of false-alarm rate.

Keywords: Speech signal processing, Impulsive noise, Impulsive noise detection, Speech enhancement

ASK subject classification: Speech Signal Processing (2,3)

I. 서론

녹음되는 음성 신호의 음질은 주변 잡음이나 녹음 기기 내부의 잡음에 의해 저하된다. 특히 휴대용 기기를 이용해 음성을 녹음하는 경우엔 기기의 버튼을 조작하거나 기기를 두드림으로써 충격 잡음이 발생할 수 있고 충격 잡음은 음성의 음질 및 명료도를 크게 저하시키므로 반드시 제거하여야 한다. 최근 user created contents (UCC) 문화의 확산과 휴대용 녹화/녹음 기기의 대중화로 인해 휴대용 기기를 이용한 녹음 활동이 일반화되면서 음성을 녹음하는 중에 발생하는 충격 잡음을 제거할 필요성이 점차 증가하고 있다.

일반적으로 대부분의 음질 향상 시스템은 가우시안 분포를 갖는 정적 배경 잡음을 가정하여 개발되었으며 충격 잡음과 같이 비 정적인 잡음은 고려하지 않는다 [1-3]. 충격 잡음은 임의의 시간에 발생하는 잡음으로 일반적으로 크기가 크고, 빠르게 변화하는 특성을 갖는다. 따라서, 그 크기와 특성을 예측하는 것이 불가능하고 잡음의 스펙트럼 정보를 이용하는 기존의 선형 잡음 제거 필터로는 충격 잡음을 제거할 수 없다. 지금까지 연구된 대부분의 충격 잡음 제거 알고리즘은 비선형 필터를 사용하는 방향으로 개발되었다 [4-7]. 충격 잡음의 제거를 위해 충격 잡음이 발생한 구간을 찾은 후, 해당 구간에 비선형 필터를 적용하여 신호의 크기를 감소시키는 방법을 주로 사용한다. 따라서, 충격 잡음을 제거하는 시스템에서 충격 잡음이 발생한 구간을 검출하는 알고리즘은 반드시 필요하며, 그 성능은 충격 잡음 제거 성능 및 음성의 왜곡 정도에

크게 영향을 미친다. 시스템이 충격 잡음이 존재하는 구간을 찾지 못하면 충격 잡음을 제거할 수 없고, 충격 잡음이 존재하지 않는 구간을 충격 잡음 구간으로 잘못 검출하면 불필요하게 음성이 왜곡된다 [5-7].

앞에서 언급한 바와 같이 충격 잡음은 크기가 크고 빠르게 변화하는 특징이 있으므로 신호의 단구간 에너지나 변화량을 관찰하여 충격 잡음이 존재하는 구간을 검출할 수 있다. 다시 말해, 신호의 단구간 에너지나 변화량이 일정 기준 값보다 큰 구간을 충격 잡음이 존재하는 구간으로 결정하는데, 이에 더하여 배경 신호의 크기를 추정하여 검출 파라미터를 조정하거나 신호를 사전에 whitening 하여 배경 신호의 영향을 제거하는 등의 추가적인 기법을 사용함으로써 충격 잡음 검출기의 성능을 높게 된다 [5-8]. 하지만 충격 잡음이 음성 신호와 함께 존재하는 경우 음성의 피치 부분이 충격 잡음과 시간 축에서 유사한 특성을 가지므로 충격 잡음의 시간 축 특성을 사용하는 기존의 충격 잡음 검출 알고리즘은 음성 신호와 충격 잡음을 구분할 수 없다. 또한, 음성의 피치는 linear prediction과 같은 auto-regressive (AR) 필터에 의한 whitening에 의해서도 제거되지 않고 충격 잡음과 같이 남기 때문에 whitening에 의한 성능 향상도 기대할 수 없다 [9]. 이를 해결하기 위해 모음의 주기성을 이용한 충격 잡음 검출 알고리즘이 제안되었으나, 충격 잡음이 반복하여 나타나는 환경에 대해 그 성능을 보장 받을 수 없다 [10].

본 논문은 음성 신호에 포함된 충격 잡음의 존재 구간을 검출하는 충격 잡음 검출 알고리즘을 제안한다. 충격 잡음의 위치를 찾기 위해서 신호의 시간 축 변화율이나 단구간 에너지와 같은 파라미터가 주로 쓰이지만, 이는 음성의 모음과 충격 잡음을 구분하기 어려운 문제가 있다. 본 논문에선 음성의 모음이 상대적으로 낮은 고주파 에너지를 갖는 특성을 이용한 주파수 축에서의 충격 잡음 검출 알고리즘을 제안한다. 주파수 축에서의 충격 잡음 검출기는 시간 축 알고리즘에 비해 높은 검출 성능을 가지면서 음성의 모음에서 피치를 충격 잡음으로 오인하는 false-alarm 문제를 해결한다. 하지만, 주파수 축에서의 알고리즘 만으로 음성과 충격 잡음을 완전하게 구분할 수 없으므로 충격 잡음 검출 성능을 유지한 채 false-alarm 문제를 최소화 할 수 있도록 시간 및 주파수 축 파라미터를 결합한 새로운 충격 잡음 검출 시스템을 제안한다.

본 논문은 다음과 같이 구성되어 있다. 먼저 기존의 시간 축 충격 잡음 검출 알고리즘을 소개한다 [5]. 다음으로

시간 축 충격 잡음 검출기의 문제점을 해결한 주파수 축에서의 파라미터를 제안하고, 두 가지의 파라미터를 모두 사용하는 시간-주파수 축 충격 잡음 검출기를 제안하여 충격 잡음 검출 오차를 최소화한다. 마지막으로 제안한 충격 잡음 검출기의 성능을 보이고 결론을 맺는다.

II. 시간 축 충격 잡음 검출기

충격 잡음은 큰 에너지를 가지고 있으며 짧은 시간 동안 빠르게 변화하는 특성이 있다. 따라서 입력 신호의 단구간 에너지나 이차 미분 계수의 크기로 충격 잡음의 존재 여부를 판단할 수 있다 [4, 6, 7, 8]. Kasparis는 입력 신호의 이차 미분 계수의 단구간 에너지를 이용한 충격 잡음 검출기를 제안하였다 [5]. 입력 신호를 $x[n]$ 으로 정의하면, 입력의 이차 미분 계수 $z[n]$ 은 식 (1)과 같다.

$$z[n] = D^2 \{x[n]\} = x[n-1] - 2x[n] + x[n+1] \quad (1)$$

또한, 2차 미분 계수의 단구간 에너지 $w[n]$ 은 식 (2)와 같다.

$$w[n] = \left[\frac{1}{N+1} \sum_{j=-N/2}^{N/2} z^2[n+j] \right]^{\frac{1}{2}} \quad (2)$$

2차 미분 계수의 단구간 에너지는 현재 샘플을 포함해서 앞과 뒤의 $N/2$ 개의 샘플의 제곱의 평균으로 구한다. 충격 잡음은 일반적으로 짧은 길이를 갖기 때문에 단구간 에너지도 적은 샘플을 이용하여 구하며 본 논문에선 8 kHz의 샘플링 주파수를 갖는 신호에 대해 N 을 10으로 설정하였다.

식 (2)에서 구한 2차 미분 계수의 단구간 에너지가 일정 값보다 큰 경우 해당 샘플은 충격 잡음이 존재하는 샘플이라고 할 수 있다. 이때 충격 잡음 여부를 결정하는 기준 값은 배경 신호의 크기에 따라 다르기 때문에 입력 신호의 변화에 맞게 검출 기준을 바꾸어야 한다. Kasparis의 충격 잡음 검출 알고리즘에선 recursive median filter (RMF)로 배경 신호의 단구간 에너지를 추정하여 충격 잡음 검출에 반영하였다 [5]. 배경 신호의 2차 미분 계수의 단구간 에너지 $b[n]$ 은 식 (3)을 이용하여 추정한다.

$$b[n] = RMF_M \{b[n-1], w[n]\} \\ = med \{b[n-M], \dots, b[n-1], w[n], w[n+1], \dots, w[n+M]\} \quad (3)$$

식 (3)에서 M 은 필터의 좌, 우 길이를 나타낸다. Median filter (MF)는 주어진 샘플들의 중앙 값을 출력으로 선택하기 때문에, 충격 잡음과 같이 큰 에너지를 갖는 샘플을 배제한 단구간 에너지의 유사 평균값을 구할 수 있다. RMF는 MF의 입력 중 과거 M 개의 입력을 필터의 과거 출력으로 대체하는 재귀적 필터로써 중간 값을 찾는 과정에서 과거의 충격 잡음 성분이 배제되기 때문에 더 안전하게 배경 신호의 단구간 에너지를 추정할 수 있는 장점이 있다. 위의 식에서 RMF 필터의 길이는 충격 잡음의 영향을 받지 않을 정도로 길어야 한다. 본 논문에선 241 샘플의 RMF를 사용하여 배경 신호의 단구간 에너지를 추정하였다.

식 (2)에서 구한 입력 신호의 단구간 에너지와 식 (3)의 $b[n]$ 을 이용해 식 (4)와 같이 충격 잡음이 존재하는 구간을 검출한다.

$$d[n] = \frac{|w[n] - b[n]|}{b[n]}$$

$$g_s[n] = \begin{cases} 1, & d[n] > C_s \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

위의 식에서 $g_s[n]$ 은 시간 축 충격 잡음 검출 결과를 나타낸다. 충격 잡음이 존재하는 샘플에 대해 $g_s[n]$ 은 1이 되고, 그렇지 않은 경우 0이 된다. 시간 축에서의 충격 잡음 검출은 입력 신호의 단구간 에너지, $w[n]$,를 $b[n]$ 으로 정규화한 후 이를 상수 C_s 와 비교함으로써 이루어진다. 이때, $d[n]$ 은 배경 신호의 크기에 영향을 받지 않으므로 상수를 기준 값으로 하여 충격 잡음을 검출하는데 용이하다. 본 논문에서 C_s 는 2로 설정하였다.

위에서 설명한 알고리즘과 유사하게 신호의 단구간 에너지를 이용하거나 단구간 에너지의 미분 계수를 이용한 방법이 많이 제안되어 왔다. 또한 배경 신호의 영향을 줄이기 위한 방법으로 앞에서와 같은 배경 신호의 크기를 이용한 파라미터의 정규화 이외에도 AR 필터를 이용하여 신호를 whitening하는 것도 가능하다 [6-8]. 이와 같은 시간 축에서의 충격 잡음 검출 알고리즘은 간단하면서도 좋은 성능을 보인다. 하지만, 배경 신호로 음성 신호가 들어오는 경우 음성의 피치는 충격 잡음과 시간 축에서 유사한 특성을 가지며, linear prediction과 같은 AR 필터 기반의 whitening으로도 그 영향을 제거할 수 없다 [9]. 따라서 기존의 충격 잡음 검출 알고리즘들은 음성의 피치와 충격 잡음을 구별하지 못하며, 음성이 존재하는 환경에서 사용할 경우 음성을 크게 왜곡시킬 수 있다.

III. 제안한 충격 잡음 검출기

3.1. 주파수 축 충격 잡음 검출 알고리즘

일반적으로 충격 잡음은 시간 축에서 음성의 피치와 유사한 성질을 가지는 반면 주파수 축에서의 에너지 분포는 음성 신호와 다르다. 음성의 모음은 저대역에 에너지가 집중되어 있는데 반해 충격 잡음은 전체적으로 평탄한 주파수 응답을 갖는다 [4]. 따라서, 충격 잡음을 검출하기 위해 스펙트럼의 평탄도나 고대역과 저대역의 에너지 비율 등을 이용할 수 있다. 하지만, 음성 신호와 충격 잡음이 함께 존재하는 구간에선 음성 신호의 영향으로 스펙트럼 평탄도가 낮아지기 때문에 해당 구간이 충격 잡음이 존재하는 구간인지의 여부가 불분명해진다. 반면, 신호의 고대역 에너지는 음성 신호의 영향이 적어 충격 잡음의 특성이 명확하게 나타나기 때문에 음성의 피치와 충격 잡음을 구별하는 좋은 척도이다. 본 논문에선 입력 신호의 고대역 에너지를 이용하여 충격 잡음의 발생 위치를 추정하는 알고리즘을 제안한다.

일반적으로 충격 잡음이 존재하는 구간은 음성의 모음에 비해 2.5 kHz 이상 주파수 대역의 에너지가 크다. 이는 충격 잡음이 비교적 평탄한 주파수 특성을 갖는데 비해 음성 신호는 약 2.5 kHz 미만의 주파수 대역에 대부분의 에너지가 분포되어 있기 때문이다 [9]. 본 논문에선 8 kHz로 샘플링된 신호를 기준으로 2.5 kHz에서 3.5 kHz 사이의 에너지를 충격 잡음 검출을 위한 파라미터로 사용하였다. 즉, 입력 신호 $x[n]$ 의 고대역 에너지 $E_H(l)$ 은 다음 식 (5)와 같이 정의할 수 있다.

$$E_H(l) = \sum_{k=2.5\text{kHz}}^{3.5\text{kHz}} |X(k,l)|^2 \quad (5)$$

위의 식에서 $X(k,l)$ 은 $x[n]$ 의 Fourier 변환 계수를 나타내며, k 와 l 은 각각 주파수와 프레임 index이다. 이때, Fourier 변환을 위한 프레임의 길이가 충격 잡음의 길이에 비해 길면, 충격 잡음을 제거할 때 충격 잡음이 존재하지 않는 구간까지 함께 필터링하게 되기 때문에 추가적인 음성의 왜곡이 발생한다. 반면, 프레임의 길이가 충격 잡음의 길이보다 너무 짧으면 상대적으로 에너지가 작은 충격 잡음의 말단을 검출하지 못할 가능성이 높아진다. 본 논문에선 충격 잡음의 초기 펄스가 1~5 ms의 길이를 갖는 것을 감안해 8 ms 길이의 프레임 단위로 주파수 축 충격 잡음 검출을 수행하였으며 프레임 길이와 같은 point의 FFT를 수행하여 $X(k,l)$ 을 취득하였다 [4].

또한 검출기의 정확도를 높이기 위해 앞뒤 프레임간에 50 %의 overlap 구간을 주었다. 즉, 주파수 축 충격 잡음 검출기는 4 ms의 정확도를 갖는다. Overlap된 구간에선 두 프레임 중 하나라도 충격 잡음이 있는 것으로 판단하는 경우엔 충격 잡음이 존재하는 구간으로 결정하였다.

고대역 에너지를 이용한 충격 잡음 검출기는 입력 신호의 고대역 에너지가 일정 기준 값보다 큰 구간을 충격 잡음이 존재하는 구간으로 결정한다. 이때, 입력 신호의 고대역 에너지는 충격 잡음의 크기뿐만 아니라 배경 신호의 크기에 영향을 받으므로 배경 신호의 크기를 고려해야 한다. 배경 신호의 고대역 에너지 $\bar{E}(l)$ 은 식 (3)과 유사하게 RMF 필터를 이용하여 구할 수 있다.

$$\begin{aligned} \bar{E}(l) &= RMF_{r=4} \{ \bar{E}_H(l-1), E_H(l) \} \\ &= med \{ \bar{E}_H(l-4), \dots, \bar{E}_H(l-1), E_H(l), E_H(l+1), \dots, E_H(l+4) \} \end{aligned} \quad (6)$$

추정한 배경 신호의 고대역 에너지를 이용하여 현재 프레임의 고대역 에너지를 정규화하면 아래의 식 (7)과 같이 충격 잡음이 존재하는지의 여부를 결정할 수 있다.

$$\begin{aligned} R_H(l) &= \frac{|E_H(l) - \bar{E}_H(l)|}{\bar{E}_H(l)} \\ g_H(l) &= \begin{cases} 1, & R_H(l) > C_f \\ 0, & otherwise \end{cases} \end{aligned} \quad (7)$$

위의 식에서 $g_H(l)$ 은 주파수 축 충격 잡음 검출 결과를 나타내며, 시간 축에서와 마찬가지로 정규화된 고대역 에너지, $R_H(l)$,를 일정 기준 값과 비교하여 충격 잡음이 존재하는 구간을 찾는다. 기준 값 C_f 는 2로 설정하였다. 그림 1은 입력 신호의 고대역 에너지를 이용하여 충격 잡음의 발생 위치를 검출한 결과이다.

그림 1에서 입력 신호는 깨끗한 음성 데이터에 실제 환경에서 녹음한 충격 잡음을 더하여 만들었다. 충격 잡음은 휴대 전화의 녹음 기능을 이용하여 휴대 전화의 버튼을 누르거나 휴대 전화 본체를 가볍게 두드리는 소리를 녹음하였다. 위의 그림에서 볼 수 있듯이 고대역 에너지를 이용한 충격 잡음 검출 알고리즘은 매우 적은 오차를 가진다. 특히, 음성의 피치를 충격 잡음으로 오인하는 false-alarm 문제가 발생하지 않는 것을 그림 1을 통해 확인할 수 있다. 반면, 음성의 자음 구간이나 음절이 바뀌는 부분을 충격 잡음으로 잘못 검출하는 false-alarm이 존재하였다.

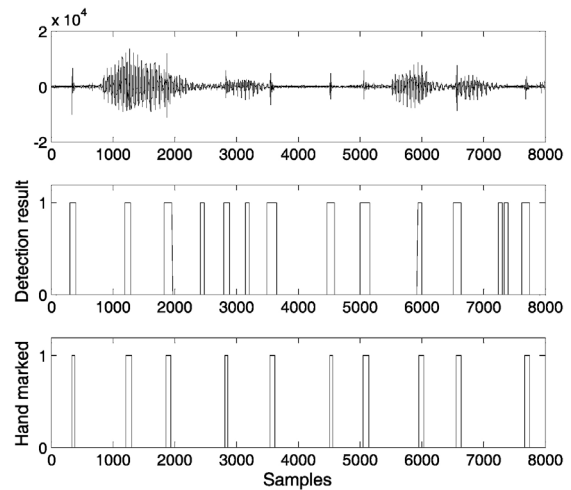


그림 1. 주파수 축 충격 잡음 검출 결과
Fig. 1. Result of frequency-domain impulsive noise detection.

3.2. 시간-주파수 축 충격 잡음 검출 알고리즘

지금까지 시간 축에서의 충격 잡음 검출 알고리즘을 소개하고, 그 단점을 해결하기 위한 주파수 축 충격 잡음 검출 알고리즘을 제안하였다. 시간 축 충격 잡음 검출 알고리즘은 입력 신호의 이차 미분 계수를 이용하여 신호의 변화율이 큰 구간을 충격 잡음이 존재하는 구간으로 결정한다. 하지만, 음성의 피치 성분이 시간 축에서 충격 잡음과 유사한 파형을 갖기 때문에 이를 충격 잡음이 존재하는 구간으로 잘못 검출하는 문제가 발생한다. 주파수 축에서의 충격 잡음 검출 알고리즘은 상대적으로 높은 고대역 에너지를 갖는 구간을 충격 잡음 구간으로 결정하는데, 음성의 피치와 충격 잡음을 효과적으로 구분하는 반면 자음 구간이나 음절이 바뀌는 구간을 충격 잡음으로 오인하는 경우가 발생한다.

결론적으로 충격 잡음의 시간 축 특성, 혹은 주파수 축 특성만으로는 충격 잡음과 음성을 완벽하게 구분할 수 없다. 본 논문에선 충격 잡음을 검출하기 위한 시간 축 파라미터와 주파수 축 파라미터가 각각 다른 구간에서 false-alarm이 발생하는 것에 착안하여 두 가지 파라미터를 모두 사용한 충격 잡음 검출기를 제안한다. 이때 추정 단위는 주파수 축에서의 충격 잡음 검출 방식과 같이 프레임 단위로 결정하는데, 이는 충격 잡음의 말단 부분이 충격 잡음으로 추정되지 않는 문제를 해결하기 위한 것이다.

제안한 충격 잡음 검출기는 아래의 식 (8)을 이용하여 충격 잡음이 존재하는 프레임을 결정한다.

$$g(l) = \begin{cases} 1, & g_H(l) = 1 \text{ and } \sum_{n \in l} g_t(n) \geq C_n \\ 0, & otherwise \end{cases} \quad (8)$$

식 (8)에서 $g(l)$ 은 최종적인 충격 잡음 검출 결과를 나타낸다. 제안된 충격 잡음 검출기는 입력 신호의 각 프레임에 대해 식 (7)의 주파수 축 충격 잡음 검출 결과, $g_H(l)$ 가 1이 되고, 해당 프레임의 샘플들에서 식 (4)를 이용해 추정한 $g_t(n)$ 의 합이 일정 값, C_n , 이상이 되는 경우 충격 잡음이 존재하는 것으로 결정한다. 본 논문에서 C_n 은 프레임 길이의 1/4로 설정하였다. 시간 축에서의 충격 잡음 검출 알고리즘은 진동하는 충격 잡음 파형의 영교차점 주변이나 잡음의 말단을 충격 잡음이 존재하는 구간으로 검출하지 못하는 문제가 있다. 다시 말해, 충격 잡음이 존재하는 프레임 내에서도 시간 축 충격 잡음 검출 결과, $g_t(n)$ 가 0이 되는 샘플이 빈번하게 나타나기 때문에 C_n 은 프레임의 길이에 비해 작은 것이 좋다. 실험 결과 C_n 을 프레임 길이의 1/4로 하는 것이 충격 잡음 검출에 적합한 것으로 나타났다.

IV. 충격 잡음 검출 성능 평가

제안한 충격 잡음 검출 알고리즘의 성능을 평가하기 위해 실제 환경에서 휴대 전화의 버튼을 누르거나 본체를 두드리며 녹음한 충격 잡음의 위치를 찾는 실험을 진행하였다. 실험에 사용된 음성 신호는 4명의 여성 화자와 4명의 남성 화자에 의해 녹음된 8개의 문장이며 총 16초의 길이를 갖는다. 위의 음성 신호의 전체에 걸쳐 임의의 위치에 녹음한 충격 잡음을 더하였으며, 16초의 입력 신호에 존재하는 충격 잡음의 개수는 약 100개이다. 음성 및 잡음은 모두 8 kHz의 샘플링 주파수를 갖는다.

기존 알고리즘인 시간 축 충격 잡음 검출기를 이용한 실험 결과에선 그림 2와 같이 음성의 피치를 충격 잡음으로 잘못 검출하는 false-alarm이 빈번하게 발생하였다. 이는 뒷단에서 충격 잡음을 제거할 때 음성을 크게 손상시키는 원인이 된다. 반면, 주파수 축 충격 잡음 검출 알고리즘을 이용할 경우 음성의 모음 구간에서의 false-alarm이 크게 줄어들고 음성의 자음 구간과 음절이 바뀌는 부분에서 새로운 false-alarm이 발생하였다. 그림 3은 주파수 축 충격 잡음 검출기에 의해 자음 구간에서 발생한 false-alarm의 예를 보여주고 있다. 두 가지의 충격 잡음 검출 알고리즘에 의한 false-alarm의 경향을 보면 주파수 축 충격 잡음 검출기에 의한 false-alarm이 시간 축 충격 잡음 검출 알고리즘에 의한 것에 비해 음성에 미치는 영향이 상대적으로 적다고 볼 수 있다. 하지만, 주파수 축 충격 잡음 검출기의 결과에서도 여전히 음성이

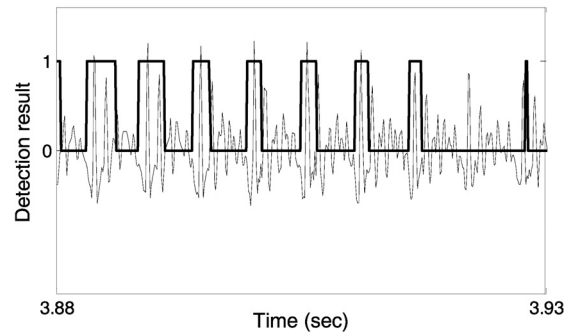


그림 2. 시간 축 충격 잡음 검출기에 의한 false-alarm의 예
Fig. 2. Example of false-alarm caused by time-domain impulsive noise detector.

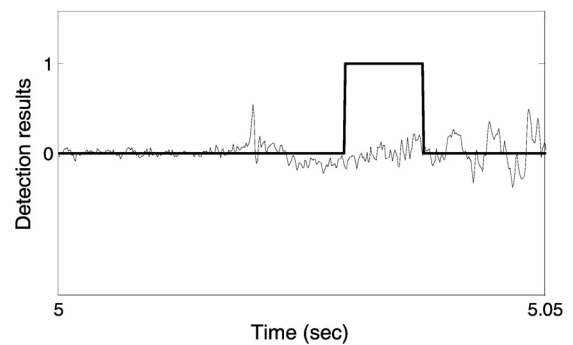


그림 3. 주파수 축 충격 잡음 검출기에 의한 false-alarm의 예
Fig. 3. Example of false-alarm caused by frequency-domain impulsive noise detector.

표 1. 충격 잡음 검출기의 성능 비교

Table 1. Performance evaluation of various impulsive noise detection algorithms.

	전체 구간		음성이 존재하는 구간	
	Detection rate	False-alarm rate	Detection rate	False-alarm rate
시간 축 검출기	46.39 %	2 %	46.30 %	2.77 %
주파수 축 검출기	96.77 %	5.94 %	95.47 %	6.36 %
시간-주파수 축 검출기	95.48 %	1.23 %	94.11 %	1.69 %

존재하는 구간에서의 false-alarm이 존재한다. 반면, 시간-주파수 축 충격 잡음 검출기는 false-alarm이 발생하는 구간이 거의 나타나지 않았다. 표 1은 시간 축, 주파수 축, 그리고 두 파라미터를 혼합한 시간-주파수 축에서의 충격 잡음 검출 성능을 비교한 것이다.

위의 표 1에서 ‘Detection rate’는 충격 잡음이 존재하는 샘플들 중에서 충격 잡음 검출기가 충격 잡음으로 맞게 검출한 샘플들의 비율을 계산한 것이며, ‘False-alarm rate’는 충격 잡음이 존재하지 않는 샘플들 중에서 검출 결과가 충격 잡음이 존재하는 구간으로 잘못 나타난

샘플들의 비율을 구한 것이다. 위의 모든 성능은 샘플 단위로 계산하였으며, 성능 확인을 위한 reference 데이터로는 음성에 더해지기 전의 충격 잡음을 이용해 손으로 표시한 충격 잡음 구간 정보를 이용하였다.

위의 표에 의하면 시간 축에서의 충격 잡음 검출기의 'Detection rate'는 매우 낮게 나타나는데, 이는 시간 축 충격 잡음 검출기가 대부분의 충격 잡음에 대해 충격 잡음이 끝나는 부분을 찾지 못하기 때문이다. 반면, 시간 축 충격 잡음 검출기에 의한 false-alarm이 매우 빈번하게 일어남에도 불구하고 수치상으로 주파수 축에서의 검출기보다 낮은 'False-alarm rate'를 갖는 것은 시간 축 충격 잡음 검출기의 false-alarm은 비교적 짧은 구간에서 나타나기 때문이다. 하지만 false-alarm이 발생하는 위치가 그림 2과 같이 모음의 피치가 있는 부분이기 때문에 음성 신호에 주는 영향은 매우 크다. 제안한 주파수 축 충격 잡음 검출기는 월등히 높은 'Detection rate'를 갖는 반면 'False-alarm rate'도 같이 상승하였는데, 상대적으로 음성의 품질에 영향이 적은 구간에서의 false-alarm이기 때문에 체감되는 검출 성능은 시간 축 충격 잡음 검출기에 비해 우수하다. 최종적으로 두 파라미터를 결합한 시간-주파수 축 충격 잡음 검출기를 사용하면 높은 검출 성능을 가지면서도 가장 낮은 'False-alarm rate'를 나타냈으므로 음성 신호에서 충격 잡음 구간을 검출하는데 가장 적합하다. 위의 실험 결과를 통해 주파수 축 충격 잡음 검출기가 기존의 알고리즘에 비해 음성이 존재하는 환경에서 충격 잡음이 존재하는 구간을 찾는 데 월등한 성능을 갖는 것을 확인할 수 있으며, 시간-주파수 축 충격 잡음 검출기가 높은 검출 성능을 유지하면서 false-alarm을 최소화한 것을 볼 수 있다.

표 1의 결과는 충격 잡음이 존재하는 구간에 대한 레퍼런스 데이터와 비교하여 한 샘플이라도 어긋나면 모두 오차로 처리한 결과로서 많은 검출 오차가 충격 잡음이 존재하는 구간 주변의 샘플들에서 발생한 것이다. 일반적으로 검출기의 성능을 측정할 때 검출하고자 하는 위치 주변의 작은 오차는 허용하므로 이를 허용한 결과를 측정하였다. 표 2는 충격 잡음의 시작점과 끝점을 기준으로 앞뒤 4 ms의 오차는 허용하였을 때의 성능을 나타낸 것이다. 표 2의 충격 잡음 검출 성능은 음성이 존재하는 구간에 대해서만 측정하였다.

충격 잡음 주변 4 ms 이내의 오차를 제외하면 제안한 주파수 축 검출기와 시간-주파수 축 검출기는 99.33%의 높은 detection rate를 갖는다. 특히, 시간-주파수 축 충격 잡음 검출기는 1.49%의 낮은 비율로 false-alarm이

표 2. 충격 잡음 검출기의 성능 비교 (주변에서의 오차 허용)
Table 2. Performance evaluation of various impulsive noise detection algorithms (while ignoring errors near the impulsive noise).

	Detection rate	False-alarm rate
시간 축 검출기	85.1 %	3.09 %
주파수 축 검출기	99.33 %	5.78 %
시간-주파수 축 검출기	99.33 %	1.49 %

나타나기 때문에 음성 신호의 왜곡 없이 충격 잡음을 제거하는데 적합하다.

V. 결론

본 논문에선 음성 신호가 존재하는 환경에서 오차를 최소화하는 충격 잡음 검출 알고리즘을 제안하였다. 기존의 시간 축 충격 잡음 검출기는 음성의 모음을 충격 잡음으로 오인하는 문제가 있으므로 음성 신호가 존재하는 환경에서 사용하기에 적합하지 않다. 또한, 충격 잡음의 말단을 검출하지 못하기 때문에 낮은 검출 정확도를 갖는다. 제안한 주파수 축 충격 잡음 검출기는 음성과 충격 잡음이 갖는 특성의 차이를 감안하여 설계되어 false-alarm에 의해 음성이 왜곡될 수 있는 위험을 줄였다. 특히, 두 파라미터를 결합한 시간-주파수 축 충격 잡음 검출기는 매우 높은 검출 정확도를 보이면서도 false-alarm의 비율을 최소화하여 음성이 존재하는 환경에서 충격 잡음이 발생하는 위치를 찾는 데 적합하다.

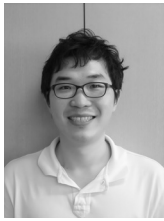
참고 문헌

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, pp. 443-445, 1985.
3. I. Cohen and B. Berdugo, "Speech enhancement for non-stationary noise environments," *Signal Process.*, vol. 81, pp. 2401-2428, 2001.
4. S. V. Vaseghi, *Advanced digital signal processing and noise reduction*, 2nd ed., John Wiley & Sons, 2000.
5. T. Kasparis and J. Lane, "Suppression of impulsive disturbances from audio signals," *Electronics letters*, vol. 29, no. 22, pp. 1926-1927, 1993.
6. A. J. Efron and H. Jeon, "Detection in impulsive noise based on robust whitening," *IEEE Trans. on Signal Process.*, vol. 42, no. 6, pp. 1572-1576, 1994.

7. S. R. Kim and A. Efron, "Adaptive robust impulse noise filtering," *IEEE Trans. on Signal Process.*, vol. 43, no. 8, pp. 1855-1866, 1995.
8. 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.
9. T. F. Quatieri, *Discrete-time speech signal processing*, Prentice Hall, 2001.
10. J. Beh, K. Kim and H. Ko, "Noise estimation for robust speech enhancement in transient noise environment," in *Proc. KSCSP 2007*, vol. 24, no. 1, pp. 35-36, 2007.

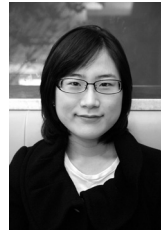
저자 약력

• **최민석 (Min-Seok Choi)**



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

• **신호선 (Ho Seon Shin)**



2008년: 연세대학교 전기전자공학과 학사
 2010년: 연세대학교 전기전자공학과 석사
 2010년 ~ 현재: 연세대학교 전기전자공학과 박사과정 재학중

• **황영수 (Young-Soo Hwang)**



1982년: 연세대학교 전자공학과 학사
 1984년: 연세대학교 전기전자공학과 석사
 1990년: 연세대학교 전기전자공학과 박사
 1989년 ~ 현재: 광동대학교 전자정보통신공학부 교수

• **강홍구 (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