음질 개선을 위한 돌발잡음 제거와 음성복원

손 백 권, 한 민 수
한국정보통신대학원대학교 음성/음향정보연구실

Abrupt Noise Cancellation and Speech Restoration for Speech Enhancement

BeakKwon Son, Minsoo Hahn
Speech and Audio Information Lab., Information and Communications University
E-mail: neverland@icu.ac.kr, mshahn@icu.ac.kr

Abstract

In this paper, speech quality is improved by removing abrupt noise intervals and then substituting the gaps with estimates of the previous speech waveform. An abrupt noise detection signal has been proposed as a prediction error signal by utilizing LP coefficients of the previous frame. Abrupt noise intervals are estimated by using spectral energy. After removing estimated noise intervals, we applied several waveform substitution techniques such as zero substitution, previous frame repetition, pattern matching, and pitch waveform replication. To prove the validity of our algorithm, the LPC spectral distortion test and the recognition test are executed and, the results show that the speech quality is fairly well improved.

I. 서 론

각종 신호처리에서 문제가 되는 잡음은 신호 전체구간에 골고루 분포하는 잡음과 특정 구간에 일시적으로 존재하는 돌발성 잡음으로 나눌 수 있다. 돌발성 잡음은 짧은 시간동안 존재하며 그 진폭이 매우 크고 단순한 파형을 갖는 임펄스성 잡음과 임펄스성 잡음에 비해 비교적 장시간 동안 존재하며 음성과 비슷한 진폭과 신호성분이 복잡한 돌발잡음으로 나눌 수 있다[1]. 예를 들면 키보드 치는 소리, 문 닫는 소리, 손뼉 치는 소리 등이 돌발잡음에 속한다. 돌발잡음이 음성 중간에 섞이게 되면 음질저하와 음성인식 오류의 발생 가능성이 커지기 때문에 돌발잡음을 제거하는 것은 음성신호처리를 위한 전처리 과정으로 중요한 의미를 갖는다. 그러나 돌발잡음은 확률적으로 모델링 하기가 힘들기 때문에 음성 중간에 섞인 돌발잡음을 제거하는 것은 기존의 잡음제거 기술로는 제거하기 어렵다. 본 논문에서는 음성 신호에 섞여 있는 돌발잡음을 검출하고 제거하는 방법을 제시였으며, 제거한 부분을 여러 종류의 파형대체 방법으로 복원하고, 성능평가를 위하여 LPC 스펙트럼 왜곡을 이용한 객관적 평가와 HTK를 이용한 인식률 평가를 수행하였다[2, 3].

II. 돌발잡음 제거 및 음성복원

2.1 검출신호

키보드 치는 소리와 책상 두드리는 소리 같은 돌발성 잡음의 특징은 많은 경우 처음에 짧은 구간에서 큰 진폭을 가지다 빠르게 그 크기가 감소하는 파형을 갖는다(그림 1). 비교적 진폭이 큰 돌발잡음에 대해서는 단순히 그 진폭만을 이용하여도 어느 정도 잡음의 위치를 알아낼 수 있으나, 돌발잡음의 크기가 음성신호와

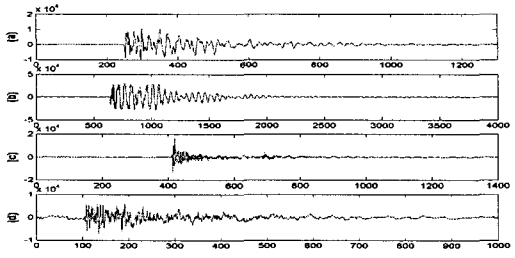

그림 1. 돌발잡음의 예, (a)노크소리, (b)문 닫는 소리, (c)펜으로 책상 치는 소리, (d) 키보드 치는 소리

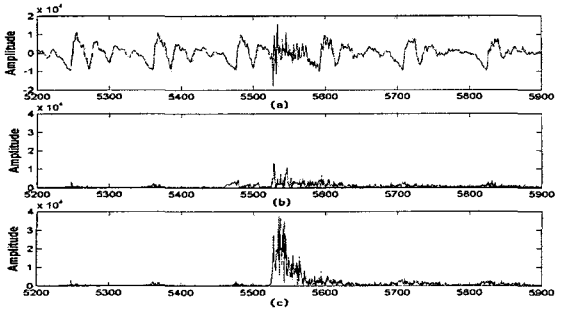

그림 2. (a) '펜으로 책상 치는 소리'가 섞인 음성파형, (b) 현재 프레임의 선형예측계수를 이용한 검출신호, (c) 이전 프레임의 선형예측계수를 이용한 검출신호

비슷하거나 작은 경우에는 잡음의 위치를 검출하기가 어려워진다. 본 논문에서는 임펄스 잡음을 다루는데 사용되는 검출신호를 돌발잡음 검출에 적용하였으며 [4], 음성의 손실은 최소화하면서 돌발잡음의 검출율을 높이기 위해 이전 프레임의 선형예측계수를 이용한 잔차신호를 검출신호로 이용하였다.

음성신호 x(n)는 AR(Autoregressive) 모델을 이용하여 다음과 같이 표현할 수 있다.

$$x(n) = \sum_{k=1}^{p} a_k x(n-k) + e(n) \qquad (1)$$

여기서 a_k는 선형예측계수이고, $e(n)$은 AR 모델로 예측시의 오류인 잔차신호이다. 잡음과 음성은 상관관계가 적기 때문에 음성 중간에 발생한 돌발잡음은 AR 모델로 표현되지 못하고 잔차신호가 크게 나타나게 된다. 이러한 잔차신호는 임펄스잡음 검출에 효과적이나, 비교적 지속시간이 큰 돌발잡음인 경우에는 선형예측계수를 구하는 과정에서 잡음구간의 정보가 사용될 수 있기 때문에, 본 논문에서는 돌발잡음에 대하여 검출율을 보다 향상시키기 위해 이전 프레임의 선형예측계수를 사용하였다[4, 5]. 이전 프레임의 선형예측계수를 이용한 검출신호, $d(n)$, 는 다음과 같이 정의한다.

$$d(n) = \left| x(n) - \sum_{k=1}^{p} a'_k x(n-k) \right| \qquad (2)$$

여기서 a'_k는 이전 프레임서 구한 선형예측계수이다.

그림 2에서 보는 바와 같이 이전 프레임의 선형예측계수를 이용한 검출신호가 현재 프레임의 선형예측계수를 사용한 것보다 더 큰 펄스로 표시되며 잡음의 시작 위치도 보다 정확히 나타난다.

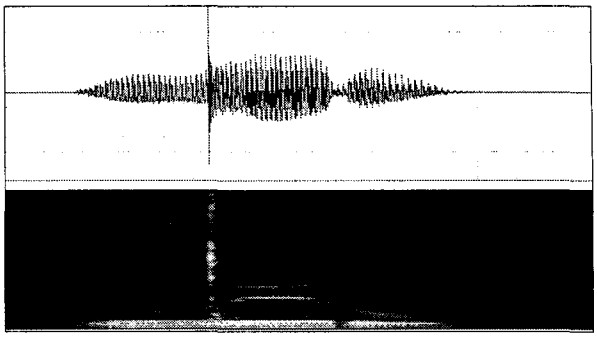

그림 3. 돌발잡음이 섞인 음성파형과 스펙트로그램

2.2 돌발잡음 구간 추정

돌발잡음은 많은 경우 짧은 시간에 큰 진폭을 가지며 생겼다가 빠르게 감쇠하는 특성을 지니며, 잡음의 시작부분에 많은 에너지가 존재한다. 또한 돌발잡음의 크기가 작은 뒷부분은 음성 파형에 심하게 영향을 주지 않기 때문에 돌발잡음을 어느 위치까지 제거해야 하는가가 관건이다. 본 논문에서는 돌발잡음의 에너지가 큰 부분을 제거하는데 중점을 두었으며 고정된 길이로 제거하는 경우와 비교하였다. 그림 3에서 보면 돌발잡음이 섞인 구간에서 스펙트로그램의 에너지가 전 주파수 범위에서 큰 에너지로 나타났다가 사라지는 것을 볼 수 있다.

본 논문에서는 검출신호로 돌발잡음의 시작점을 찾아낸 후, 잡음 시작점 이후 프레임의 스펙트럼 에너지(E_a)와 이전 프레임의 스펙트럼 에너지(E_b)의 차(E_d)를 계산한다. 이렇게 구한 에너지의 10 % 크기와 잡음 이전 프레임 에너지(E_b)의 합을 스펙트럼 에너지 문턱값(SET)으로 설정하였다. 그리고 프레임을 이동하면서 스펙트럼 에너지를 구하고 SET와 같아지는 프레임까지를 잡음구간으로 설정하여 제거하였다.

$$E_d = E_a - E_b = \frac{1}{N}\sum_{k=0}^{N-1} X_a(k) - \frac{1}{N}\sum_{k=0}^{N-1} X_b(k)$$
$$SET = E_b + E_d \times 0.1 \qquad (3)$$

여기서 N는 FFT(Fast Fourier Transform) 차수이고, $X_a(k)$는 잡음 이후의 신호 $x_a(n)$을 FFT하여 구한 k번째 스펙트럼 크기이고, $X_b(k)$는 잡음 이전의 신호 $x_b(n)$을 FFT하여 구한 k번째 스펙트럼 크기이다.

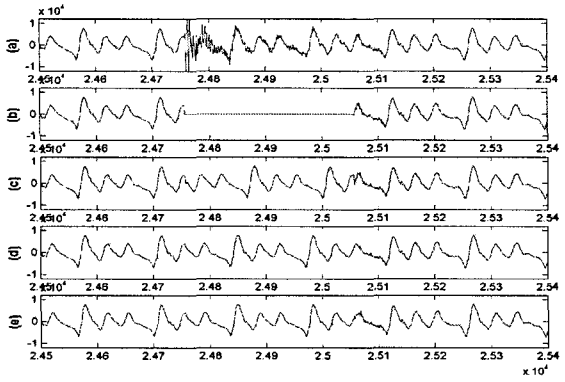

그림 4. Speech Waveform Samples (a) Noisy Speech, (b) ZERO, (c) REPT, (d) OSPM, (e) OSPW

2.3 손실구간 대체

돌발잡음을 제거함으로써 발생한 음질 저하는 제거되지 않는 음성의 일부를 대체함으로써 음질 개선과 함께 제거된 구간의 음성정보 보상 효과를 얻을 수 있다. 여기에서는 음성합성 과정이 필요 없는 파형 대체 방법들을 이용하여 잡음이 제거된 구간을 대체하였다. 본 논문에서 사용한 파형대체 방법으로는 zero substitution, 이전 프레임 파형 반복, pattern matching 그리고 pitch waveform replication 방법을 이용하였다.

돌발잡음이 섞인 음성 파형을 비롯하여 ZERO(zero substitution), RETP(이전 프레임 파형 반복), OSPM(one sided pattern matching), 그리고 OSPW(one sided pitch waveform) 방법으로 복원한 음성 파형을 그림 4에 나타내 보았다.

Ⅲ. 실험 및 결과

먼저 본 논문에서 제안한 검출신호의 잡음 검출 효과를 평가하였고, 위의 4가지 복원방법에 제거 구간 이후의 음성샘플까지 이용한 TSPM(two sided pattern matching) 방식까지 총 5가지 복원 방법을 이용하여 실험하였다. 실험에 사용한 돌발잡음은 '펜으로 책상 두드리는 소리'를 이용하였다.

3.1 돌발잡음 검출 및 제거 비교

표 1과 표 2는 SNR에 따라 잡음의 위치를 놓치는 경우와 잘못 판단한 경우를 나타낸 것이다.

여기서 음성신호와 돌발잡음의 SNR은 다음과 같은 방식으로 정의하였다.

$$SNR(dB) = 20\log\left(\frac{Es}{En}\right) \quad (4)$$

표 1. 검출 실패율 (%)

	-17dB	-6dB	6dB	15dB
Current LPC	0.73	0.96	7.96	31.19
Previous LPC	0.00	0.22	1.54	13.00

표 2. 오류 잡음 검출율(False detection rate) (%)

	-17dB	-6dB	6dB	15dB
Current LPC	0.07	0.81	20.65	69.98
Previous LPC	0.07	0.51	17.92	59.21

표3. 고정된 잡음구간과 제안한 잡음구간 사용하였을 때, LPC Spectral Distortion 비교 (SNR=-6dB, OSPW)

	Removal Duration	LPC Spectral Distortion
Fixed	4 ms	0.879 dB
	8 ms	0.875 dB
	12 ms	0.865 dB
	20 ms	0.848 dB
	28 ms	0.828 dB
	36 ms	0.817 dB
	48 ms	0.822 dB
Proposed		**0.816 dB**

여기서 En은 돌발잡음의 90% 에너지이고 Es는 돌발잡음의 90% 되는 구간의 길이의 음성신호의 평균 에너지이다.

표 1과 표 2에서 보는 바와 같이 이전 프레임의 선형예측계수를 사용한 검출신호가 돌발잡음의 크기가 작아져도 더 정확히 찾아내는 것을 볼 수 있다. 표 3에서는 본 논문에서 제안한 방법으로 잡음구간을 추정하였을 때, 고정된 구간을 사용한 경우보다 OSPW방식으로 복원하였을 때, 식 (5)의 LPC spectral distortion이 제일 작게 나타났다. 이것은 돌발잡음이 위치한 음성 에너지에 따라 잡음의 길이를 가변적으로 추정하여 제거한 것이 음성의 손실을 보다 줄일 수 있는 방법임을 시사한다.

3.2 LPC 스펙트럼 왜곡

LPC spectral distortion[2]은 잡음 없는 음성과 돌발잡음이 섞인 음성 신호, 그리고 복원된 음성 사이의 LPC envelope difference(d)로 정의하였다.

$$d = \frac{1}{F}\int_0^F |P - P'|df \quad (5)$$

$$D=\frac{1}{N}\sum_{0}^{N}d$$

여기서 P와 P'은 잡음 없는 음성과 복원된 음성의 spectral envelope이고 F는 upper limit frequency, N은 프레임 개수, 그리고 D는 LPC 스펙트럼 왜곡이다.

표 4는 각각의 파형 대체 방법으로 복원한 음성에 대해서 LPC 스펙트럼 왜곡을 보여준다. TSPM의 경우 제거구간 이후의 잡음이 남아있는 음성을 복원하는데 이용했기 때문에 LPC 스펙트럼 왜곡이 다른 방법보다 크게 나왔다. OSPW 방법이 모든 경우에 대해서 왜곡이 가장 작았으며 돌발잡음 2개가 섞인 경우 OSPW 방법이 1개가 섞인 경우의 OSPM과 같게 나왔고, 돌발잡음이 3개가 섞인 경우 2개가 섞였을 때의 REPT 결과와 비슷하게 나타나 OSPW가 음성 제거구간을 복원하는데 제일 좋은 파형 대체 방법임을 보여주었다.

3.3 인식률 평가

돌발잡음이 섞인 음성구간을 제거함으로써 생기는 음성정보의 손실 효과와 제거구간 복원 후에 보상되는 음성정보를 조사하기 위해 인식실험을 하였다. 사용한 음성 DB는 원광대에서 제작한 PBW452 DB 중 남성 화자 34명분을 사용하였다. 인식시스템은 HMM 기반의 HTK를 사용하였으며 특징 파라미터로는 MFCC 39차를 사용하였다. 돌발잡음으로는 '펜으로 책상 치는 소리'를 사용하였으며 -17 dB의 SNR의 돌발잡음 개수를 증가시켜 가면서 인식률을 조사하였다.

표 5의 결과를 보면 돌발잡음을 섞었을 경우 인식률이 잡음 없는 음성의 인식률 보다 각각 4.65%, 15.18%, 29.16% 저하되었다. ZERO 방식에서는 돌발잡음을 섞었을 때보다 더 나쁜 인식률을 나타냈으며, 이는 제거된 구간에서 어떠한 음성 정보도 사용하지 못했기 때문이다. 그러나 복원된 음성의 경우 제거구간의 음성정보를 상관관계가 매우 높은 이전 또는 이후의 음성정보로 대체되기 때문에 대다수가 잡음 없는 음성의 인식률 수준으로 나타났으며 OSPW과 OSPM 방식이 인식률 향상에 좀 더 좋은 결과를 보였다.

Ⅳ. 결 론

본 논문에서는 음성 중간에 섞인 돌발잡음을 검출, 제거하고 제거한 부분을 잡음 이전의 음성 파형으로 대체함으로써 음질 및 인식률 향상의 효과를 보이고자 하였다. 이전 프레임의 선형예측계수로 구한 잔차신호를 돌발잡음 검출을 위한 검출신호로 이용하여 잡음의 위치를 검출하였고 잡음의 스펙트럼 에너지 변화를 조사하여 돌발잡음의 길이를 추정하였다. 이렇게 추정한

표 4. LPC 스펙트럼 왜곡 (dB), (SNR =-6 dB)

	NOISY	REPT	OSPM	TSPM	OSPW
1개	0.851	0.826	0.820	0.830	0.816
2개	0.967	0.830	0.828	0.866	0.820
3개	1.025	0.848	0.842	0.891	0.831

표 5. 인식률 (%) (SNR =-17 dB)

	Noisy	ZERO	REPT	OSPM	TSPM	OSPW
1개	95.13	75.72	99.78	99.78	99.78	99.34
2개	84.60	57.61	98.67	99.12	98.89	99.12
3개	70.62	39.10	98.23	98.67	98.45	98.58
Without Noise 99.78						

잡음구간을 음성신호와 함께 제거하고 그 부분을 여러 파형 대체방법으로 복원하였을 때 LPC 스펙트럼 왜곡은 잡음 섞인 음성파형 보다 줄이고 인식률은 향상시키는 효과를 얻었다.

보다 향상된 음질을 얻기 위해서는 제거된 부분에 더욱 정교한 보상 기법이 요구된다. 손실된 부분이 길어지면 이전 음성 정보를 이용한 복구에는 한계가 있기 때문에 잡음의 길이를 보다 정확히 추정하여 잡음 이후의 음성 정보까지 제거된 구간을 복구하는데 이용해야 할 필요가 있다.

참고문헌

[1] 한현배, "음성신호특성을 이용한 사무실환경에서의 돌발잡음제거방법에 대한 연구", 박사학위논문, 한국정보통신대학원대학교, 2001

[2] O.J. Wasem, D.J. Goodman, C.A. Dvorak, H.G. Page, "The effect of waveform substitution on the quality of PCM packet communications.", *IEEE Transactions on*, Vol.36, pp.342-348, Mar., 1988

[3] J. Tang, "Evaluation of double sided periodic substitution(DSPS) method for recovering missing speech in packet voice communication.", *Computers and Communications,* pp. 454-458, Mar., 1991

[4] I. Kauppinen, "Methods for detecting impulsive noise in speech and audio signals.", *Digital Signal Processing, 2002.* Vol .2 pp. 967-970, July, 2002

[5] S.V. Vaseghi, P.J.W. Rayner, "Detection and suppression of impulsive noise in speech communication systems.", *IEE Proceedings I*, Vol. 137, pp. 38-46, Feb., 1990