

Adaptive Comb Filtering을 이용한 이동 통신 환경에서의 효과적인 잡음 제거

박정식^o, 정규준, 오영환
한국과학기술원 전자전산학과

Effective Noise Reduction in Mobile Communication Environment using Adaptive Comb Filtering

Jeong-Sik Park^o, Gue-Jun Jung, Yung-Hwan Oh
Department of Electrical Engineering & Computer Science
Korea Advanced Institute of Science and Technology
E-mail : {dionpark, sylph, yhoh}@speech.kaist.ac.kr

Abstract

In this paper, we employ the adaptive comb filtering for effective noise reduction in mobile communication environment. Adaptive comb filtering is a well-known method for noise reduction, but requires the correct pitch period and must be applied just in voiced speech frames. To satisfy these requirements we use two kinds of information extracted from speech packets, one of which is the pitch period information measured precisely by a speech coder and the other is the frame rate information related to a decision on speech or silence frame. Experiments on speech recognition system confirm the efficiency of this method. Feature parameters employing this method give superior performance in noise environment to those extracted directly from output speech.

I. 서론

음성 인식 기술은 이제 전화망 뿐 아니라 유선 통신망에 이르기까지 응용 범위가 확대되고 있으며 이동 통신 환경으로 영역을 넓혀가고 있다. 이동 통신 환경에서의 음성 인식에 대한 연구는 특징 파라미터를 추

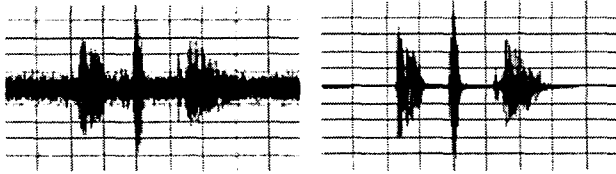
출하는 전처리부의 설계 방법에 따라 음성 코덱의 출력 음성으로부터 특징 파라미터를 추출하는 방법, 코덱의 부호화기에서 인식 파라미터를 추출하여 그것을 코딩 및 인식에 이용하는 방법, 그리고 음성 코딩에 사용되는 파라미터를 패킷에서 추출하여 특징 파라미터로 사용하는 방법으로 나뉜다. 이 중 두 번째 방법은 인식률은 좋으나 합성음의 음질 저하가 문제되며, 첫 번째 방법은 구현이 용이하나 세 번째 방법보다 인식 성능이 좋지 않다는 연구 결과가 있었다 [1]. 하지만 15dB 이하의 잡음 음성에 대해 인식 실험을 수행한 결과 세 번째 방법보다는 첫 번째 방법의 인식률이 높았으며, 이는 음성 코덱의 영향임을 확인하였다 [2].

본 논문에서는 이처럼 출력 음성에서 특징 파라미터를 추출하는 경우 출력 음성에 포함된 잡음을 제거하기 위해 Adaptive Comb Filtering을 이용하는 방법을 제안한다. 2장에서 잡음 환경에서 음성 코덱이 인식 성능에 미치는 영향을 살펴보고 3장에서 Adaptive Comb Filtering을 이용하여 잡음을 제거하는 방법에 대해 설명한다. 4장에서 실험 결과를 제시하며 5장에서 결론을 맺는다.

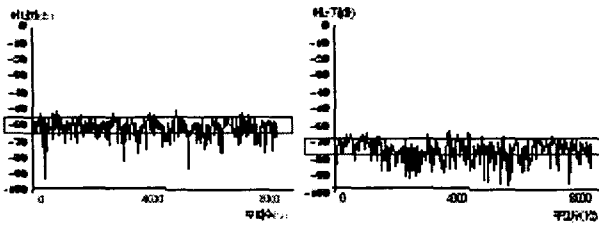
II. 잡음환경에서 음성 코덱의 영향

출력 음성과 패킷에서 추출한 특징 파라미터의 성능을 비교한 기존 연구의 실험 환경은 배경 잡음이 첨가

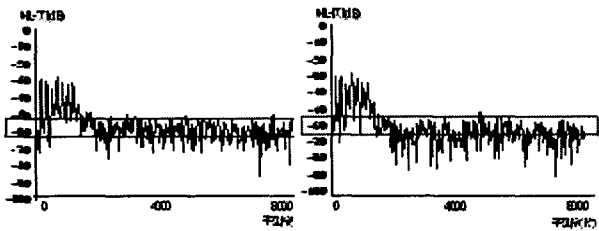
되지 않은 깨끗한 실험실 환경이었다. 그러나 동일한 실험 조건에서 배경 잡음이 첨가된 음성을 대상으로 인식 실험을 수행한 결과 기존의 연구와 달리 출력 음성에서 추출한 파라미터가 더 좋은 성능을 나타냈다 [2]. 이와 같은 결과는 실험에 사용한 QCELP 음성 코덱의 영향과 관련이 있다.



(a) 입력 음성 (b) 출력 음성
그림 1. 음성 코덱의 영향 (백색 잡음, SNR 5dB)



(a) 입력 음성 (b) 출력 음성



(a) 입력 음성 (b) 출력 음성

그림 2. 음성(上)/비음성(下) 프레임의 스펙트럼

그림 1은 백색 잡음이 첨가된 입력 음성과 출력 음성의 파형을 비교한 것으로 디코딩 후 비음성 구간의 에너지가 상당히 감소되었음을 확인할 수 있다. 그림 2는 동일한 음성의 음성의 음성 구간과 비음성 구간의 스펙트럼으로 비음성 구간의 경우 디코딩 후 전 주파수 대역의 에너지가 감소되었지만, 음성 구간은 에너지의 변화가 거의 없었다.

위의 그림으로부터 알 수 있듯이 잡음 환경에서 출력 음성에서 추출한 파라미터의 성능이 높은 이유는 음성 코덱으로 인한 비음성 구간의 에너지 감소 때문이다. 일반적으로 음성에 잡음이 첨가되면 음성 구간의 검출이 어렵다. 그러나 코덱을 통과한 출력 음성의 경우 비음성 구간의 에너지 감소로 인해 음성 구간의 검출이 용이해지며 이로 인해 인식 성능은 향상된다.

III. Adaptive Comb Filtering을 적용한 잡음 제거

앞서 살펴 본 바와 같이 음성 코덱은 비음성 구간의 에너지를 감소시킨다. 이와 같은 상황에서 출력 음성의 음성 구간에 여전히 존재하는 잡음 성분을 제거함으로써 인식 성능을 더욱 향상시킬 수 있을 것이다. 이를 위해 본 연구에서는 Adaptive Comb Filtering을 음성 구간의 잡음 제거에 이용하고자 한다.

3.1 Adaptive Comb Filtering을 이용한 출력 음성의 잡음 제거

잡음 차감법과 같은 대부분의 음질 개선 기법은 입력 신호에서 비음성 구간을 검출한 후 잡음 성분을 추정하고 그 결과를 음성 구간의 잡음 처리에 이용한다. 그러나 디코딩 후 생성된 출력 음성의 비음성 구간은 코덱의 영향으로 에너지가 감소된 상태이기 때문에 정확한 잡음 추정이 불가능하며 음성 구간의 잡음을 제거하는데 한계가 있다. 따라서 출력 음성의 잡음 제거를 위해서는 비음성 구간의 잡음 추정 없이 음성 구간의 잡음을 제거할 수 있는 음질 개선 기법을 적용해야 한다. Adaptive Comb Filtering은 이러한 특성을 지닌 음질 개선 기법으로 음성 구간에서 측정된 기본 주파

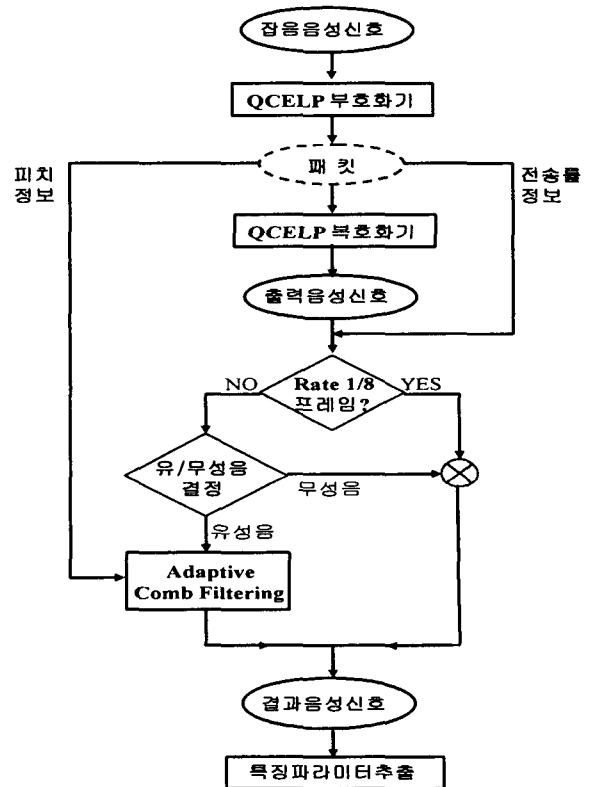


그림 3. 제안한 잡음 제거 과정

수(또는 피치 주기)를 이용하여 잡음 성분을 제거하기 때문에 비음성 구간의 잡음 추정이 필요 없다 [3][4]. 그림 3은 본 논문에서 제안한 잡음 제거 방법을 나타낸 것으로, 출력 음성의 음성 구간에 Adaptive Comb Filtering을 적용하는 과정이다.

3.2 음성 프레임 및 유성음 프레임 선별

Adaptive Comb Filtering은 피치 주기를 갖는 음성 구간에 대해서만 적용할 수 있으므로 음성 프레임을 선별해야 한다. 이 과정에서 프레임의 전송률 정보를 패킷에서 추출하여 이용하였다. QCELP 코덱은 각 프레임의 speech activity를 측정하여 전송률을 결정하는데 activity가 작은 비음성 구간은 전송률이 1 kbps로 결정되는 특성이 있다 [5].

음성 구간은 유성음과 무성음으로 이루어지는데 비주기성이 강한 무성음 구간에 대해 필터링을 수행하면 무성음의 신호 성분이 제거될 우려가 있다. 따라서 음성 구간 중 유성음 프레임만을 선별하는 과정이 필요하다. 본 연구에서는 자기상관계수(autocorrelation coefficient)를 유성음 프레임 선별에 사용하였다 [6].

유성음 프레임에 대해 필터링을 수행하면 잡음 성분이 제거되어 프레임의 전체 에너지가 감소한다. 따라서 필터링이 적용되지 않는 비음성 또는 무성음 구간은 상대적으로 큰 에너지를 갖게 되므로 부적절하게 강조될 수 있다. 이러한 문제를 해결하기 위해 비음성과 무성음 프레임에 대해 간단한 감쇠 처리를 하였다.

3.3 유성음 프레임의 필터링 처리

위의 과정을 통해 선별된 유성음 프레임에 대해 Adaptive Comb Filtering을 적용한다. 필터링을 효과적으로 수행하기 위해서는 각 프레임의 정확한 피치 주기가 요구된다. 캡스트럼 기반 또는 LPC기반의 피치 측정 방법 등 피치 주기를 측정하는 다양한 방법들이 존재하지만 정확한 피치 주기를 측정하는 것은 쉽지 않으며 잡음 음성의 경우는 더욱 어렵다. 이와 같은 제약을 해결하기 위해 본 연구에서는 음성 코덱에서 측정된 피치 주기를 음성 패킷에서 추출하여 직접 이용하였다. QCELP 음성 코덱은 프레임마다 피치 주기가 최대 네 차례 갱신될 정도로 음성의 피치 주기를 정확히 측정한다. 특히 패킷에서 추출한 피치 주기는 복호화기에서 출력 음성의 합성에 이용하는 정보로서 출력 음성에 정확히 부합하는 피치 주기이기 때문에 피치 측정이 어려운 잡음 음성에 대해서도 필터링을 효과적으로 수행할 수 있다.

그림 4는 유성음 프레임에 대해 필터링을 처리하는 과정을 나타낸 것이다. QCELP에서는 각 프레임의 전송률에 따라 피치 주기의 갱신 횟수가 다르기 때문에 부프레임 단위로 필터링을 수행한다. 그림 4는 전송률이 8 kbps인 프레임으로 피치 주기가 네 차례 갱신되기 때문에 네 개의 부프레임에 대해 각각의 피치 주기(T1-T4)를 이용하여 Comb Filter를 구성하고 부프레임 별로 Comb Filter를 적용하여 필터링을 수행한다.

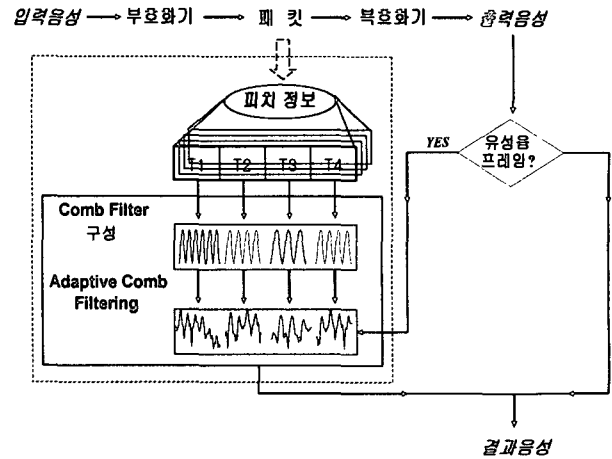


그림 4. 유성음 프레임의 필터링 처리

IV. 실험 및 결과

4.1 실험 환경

제안한 잡음 제거 방법의 효율성을 검증하기 위해 QCELP 코덱의 부호화기에서 생성된 패킷과 복호화기에서 생성된 출력 음성을 사용하여 특징 파라미터를 추출하였으며, 연속 HMM을 이용한 50 단어 규모의 고립 단어 인식 실험을 수행하였다. 실험에 사용한 DB는 한국어 낭독 음성 PBW (Phonetically Balanced Word) DB이며, 학습 데이터로는 20명의 화자가 두 차례씩 발성한 2000개의 발화를, 실험 데이터로는 10명의 화자가 두 차례씩 발성한 1000개의 발화를 사용하였다 [7]. Adaptive Comb Filtering을 적용한 후 추출한 파라미터 외에 출력 음성에서 직접 추출한 파라미터와 기존의 음질 개선 기법 중 가장 흔히 사용되는 스펙트럼 차감법을 적용한 후 추출한 파라미터를 성능 비교의 목적으로 사용하였다. HTK를 이용하여 추출한 39차 MFCC 파라미터를 특징 파라미터로 사용하였으며 13차 MFCC, 차분, 가속 MFCC로 구성하였다. 잡음 환경을 고려하기 위해 실험 데이터에 첨가한 배경 잡음

음은 NoiseX-92에 포함된 백색 잡음과 균중 잡음이며, 실험실 환경(clean speech) 및 SNR 15dB, 10dB, 5dB, 0dB 각각의 잡음 환경에 대하여 인식 성능을 비교하였다 [8].

QCELP에서는 각 프레임의 전송률에 따라 피치 주기의 갱신 횟수가 다르기 때문에 부프레임 단위로 필터링을 수행한다. 가령 전송률이 'Rate 1'인 프레임은 피치 주기가 4차례 갱신되므로 4개의 부프레임에 대한 각각의 피치 주기를 이용하여 Comb Filter를 구성하고 부프레임 별로 필터링을 수행한다.

4.2 실험 결과

표 1. 각 파라미터 간 성능 비교 (백색 잡음)

SNR	∞ (clean)	15dB	10dB	5dB	0dB
출력음성	99.4	89.2	70.4	40.6	13.1
잡음차감법	99.3	90.9	72.1	42.4	14.3
제안한방법	96.5	87.7	77.3	48.3	25.7

표 2. 각 파라미터 간 성능 비교 (균중 잡음)

SNR	∞ (clean)	15dB	10dB	5dB	0dB
출력음성	99.4	96.7	89.2	60.6	16.2
잡음차감법	99.3	96.4	90.3	60.4	16.0
제안한방법	96.5	96.2	92.2	83.4	35.3

각 실험에 대한 인식 성능 결과는 표 1, 2와 같다. 실험실 환경과 SNR 15dB의 잡음 환경에서는 제안한 방법으로 추출한 파라미터의 성능이 낮았지만, SNR 10dB 이하의 잡음 환경에서는 다른 파라미터에 비해 성능이 크게 향상되었으며 잡음 정도가 심해질수록 성능 차이가 두드러졌다. 또한 기존의 잡음 차감법이 적용된 파라미터의 성능은 출력 음성에서 추출한 파라미터보다는 향상되었지만 제안한 방법으로 추출한 파라미터의 성능에는 미치지 못했다. 이와 같은 결과는 Adaptive Comb Filtering을 통해 음성 구간의 잡음이 효과적으로 제거되었음을 의미한다.

V. 결론

본 논문에서는 음성 코덱의 출력 음성에 대해 Adaptive Comb Filtering을 적용함으로써 음성에 포함된 잡음을 효과적으로 제거하는 방법을 제안하였다. 이 방법은 코덱에서 제공되는 정보를 이용함으로써 피치 추출 등의 과정을 생략할 수 있으며 필터링을 보다 정확히 수행할 수 있다는 장점을 갖는다.

QCELP 코덱에서 생성된 출력 음성을 사용하여

SNR 15dB 이하의 잡음 환경에서 인식 실험을 수행한 결과, 제안한 방법으로 추출한 파라미터의 성능 향상을 확인할 수 있었다. 그러나 실험실 환경에서는 다른 파라미터에 비해 낮은 성능을 보였는데 그 이유는 필터링에 의한 음성 신호의 왜곡 때문이다. 이를 해결하기 위해 왜곡을 최소화하는 개선된 Adaptive Comb Filtering을 연구해야 할 필요가 있다.

참고 문헌

- [1] Hong-Kook K., "Bitstream-based feature extraction for wireless speech recognition," Proc. of International Conference on ASSP, Vol. 3, pp. 1607-1610, 2000
- [2] 박정식, 정규준, 오영환, "Adaptive Comb Filtering을 이용한 이동 통신 환경에서의 강인한 음성 인식," 2002년 한국음향학회 추계학술발표대회 논문집, 제21권 제2(s)호, pp.35-38, 2002
- [3] R.H.Frazier, "An adaptive filtering approach toward speech enhancement," S.M.thesis, M.I.T., Cambridge, 1975
- [4] R.H.Frazier, S.Samsam, "Enhancement of speech by adaptive filtering," in Proc. IEEE Int. Conf. on ASSP, Philadelphia. PA, Apr.12-14, pp.251-253, 1976
- [5] W.Gardner, "QCELP: A variable rate speech coder for CDMA digital cellular," Speech and audio coding for wireless and network applications, pp.77-84, B.S. Atal et. al.(eds), Kluwer Academic Pub., 1993
- [6] B.S.Atal, L.R.Rabiner, "A pattern recognition approach to voiced-unvoiced-silence classification with applications to speech recognition," IEEE Trans. Acoust., Speech, Signal Processing, vol.24, pp.201-212, June 1976
- [7] 이용주, "음성언어코퍼스," 정보과학회지, 16(2), 1998
- [8] A.P.Varga, H.J.M Steenken, M.Tomlinson and D.Jones, "The NOISEX-92 study on the effect of additive noise on automatic speech recognition," Technical Report, DRA Speech Research Unit, 1992