

가변 비트율 음성 부호화기의 성능분석

Performance Analysis of A Variable Bit Rate Speech Coder

임 병 관*
(Byeong-Gwan Iem)

Abstract - A variable bit rate speech coder is presented. The coder is based on the observation that a speech signal can be viewed as a combination of piecewise linear signals in a short time period. The encoder detects the sample points where the slope of the signal changes, which are called the inflection points in this paper. The coder transmits the location and value for the detected inflection sample, but only the location information for the noninflection samples. In the decoder, the noninflection samples are estimated with interpolation of the received information. Several factors affecting the performance of the coder have been tested through simulation. Simulation results show that the linear interpolation produces 1 ~ 5 dB improvement over the cubic spline interpolation. And the μ -law companding does not provide any benefit when it is applied before the inflection detection. With low threshold values in the inflection point detection, the coder shows better MOS and more than 16 dB improvement in SNR compared to the continuously variable slope delta modulation (CVSDM).

Key Words : Speech coding, Variable rate, Inflection point, Interpolation

1. 서 론

음성 부호화는 오랜 기간 연구되어온 분야이다. 디지털 통신의 진전과 데이터 압축 등의 발전과 함께 다양한 음성 부호화 방법이 연구되어 왔다. 음성 부호화는 크게 파형 부호화(waveform coding), 보코딩(vocoding), 하이브리드 부호화(hybrid coding)으로 구분된다. 파형 부호화는 음성의 매 표본을 부호화 하는 기법으로, 대표적인 예로는 μ -law 혹은 A-law PCM (Pulse Code Modulation), CVSDM (Continuously Variable Slope Delta Modulation), ADPCM (Adaptive Differential PCM) 등의 방법이 있다 [1,2]. 파형 부호화는 비교적 전송 비트율이 높아서 데이터 압축을 면에서 저조하나 객관적인 성능과 주관적인 성능이 대체로 탁월하다 [1,2]. 반면 보코딩은 음성의 매 표본을 부호화하는 것이 아니라, 음성 발생모델에 기초하여 블록 단위로 음성의 주요 인자들을 추출하여 전송한다. 대표적인 부호화 방식으로는 2.4~4.8kbps의 전송률을 보이는 미국 국방성 표준의 LPC-10 보코딩을 들 수 있다. 보코딩의 경우 우수한 데이터 압축율을 보이지만, 음질이 저하하며 부호화기가 복잡한 단점이 있다 [2]. 하이브리드 부호화는 최근의 이동통신의 발전과 함께 진화된 부호화 기법으로 파형 부호화와 보코딩의 조합으로 볼 수 있다. 대표적인 예로는 Subband Coding(SBC), Adaptive Transform Coding(ATC), Multiband Excited Vocoder (MBEV), Residual error

Excited Linear Predictive coder (RELPE), Code Excited LP (CELPE) 등을 들 수 있다 [2,3]. 하이브리드 부호화 기법은 앞의 두가지 기법의 장점을 취한 것으로 비교적 탁월한 데이터 압축을 구현하며 우수한 성능을 보인다. 단점으로는 복잡한 구조를 들 수 있다. 특히, 이동통신에 있어서는 복잡한 구조에 따른 시간지연 방지 문제가 하이브리드 부호화의 주요 연구과제이다 [2].

본 논문에서는 비교적 간단한 구조를 갖는 가변 비트율의 음성 부호화기를 소개한다 [4,5]. 본 논문에서 소개하는 가변 비트율 부호화기는 음성 파형이 국소적으로는 선형적인 특성을 보인다는 관찰에 기초한다. 따라서 부호화기에서는 선형적 특성이 교차하는 시점과 그때의 값을 추출하며, 이러한 정보를 부호화하여 전송한다. 본 논문에서는 선형적인 특성이 교차하는 지점을 변곡점이라고 칭한다. 부호화기에서는 수신된 정보를 활용하여 전송이 안 된 선형구간의 음성을 추정하여 음성신호를 복원한다. 본 논문의 부호화기는 음성의 구간에 따라 전송되는 데이터의 양이 달라지므로 가변 비트율을 보인다. 본 논문에서는 가변 비트율 부호화기의 구조와 주요 구성 알고리즘을 소개하고, 신호대 잡음비로 대표되는 객관적인 성능과 Mean Opinion Score (MOS) 로 측정되는 주관적인 성능을 검증한다. 아울러 가변 비트율 부호화기의 주요 구성부분의 변화에 따른 성능의 변화도 분석한다. 따라서 본 논문은 다음과 같은 구조로 이뤄진다. 먼저 2장에서는 가변 비트율 부호화기의 구조와 주요 구성 알고리즘을 소개한다. 3장에서는 성능에 영향을 줄 수 있는 인자들을 살펴본다. 4장에서는 실제 음성 데이터에 부호화기를 적용하여 성능을 검증한다. 성능 검증시에는 3장에서 고려한 인자들을 조합하여 시험기록 한다. 마지막으로 결론을 맺는다.

* Corresponding Author : Dept. of Electronic Engineering,
Gangneung-Wonju National University, Korea
E-mail : ibg@gwnu.ac.kr

Received : June 5, 2013; Accepted : November 21, 2013

2. 음성부호화기의 구조

2.1 부호화기 및 복호화기의 구조

음성신호는 인간의 마음 속에 있는 생각을 전달하기 때문에 시간경과에 따라 파형의 형태가 달라진다. 하지만 매우 짧은 시간구간에서 거의 직선에 근접한 형태를 보인다. 따라서, 구간별로 선형에 가까운 신호의 변곡점을 추출하고 그곳에서의 시간정보와 신호 값을 전송하면 정보의 양을 줄일 수 있다. 부호화기는 이러한 변곡점을 추출하는 부분과 변곡점 정보를 전송하기 위하여 비트 열을 형성하는 부분으로 구성된다 [5]. 변곡점에서는 신호의 미분 값이 변화하므로 부호화기에서는 신호의 미분을 구한다. 미분은 중앙차동 방정식을 이용하여 아래와 같이 근사적으로 구할 수 있다 [6].

$$y[n] = \frac{1}{4} (x[n+1] - x[n-1])$$

여기에서, $x[n]$ 은 음성 신호의 표본, $y[n]$ 은 미분의 근사치이다. 이렇게 얻어진 미분의 근사치를 부호화기에서 관찰하여 그 값이 급격하게 변하는 위치에서 변곡이 발생한 것으로 판단한다.

부호화기는 변곡이 발생한 표본 위치에서 변곡이 발생했음을 알리기 위하여 플래그를 '1'로 설정하고 그 뒤에 음성 표본 값을 미리 정해진 비트수로 양자화 하여 전송한다. 변곡점이 아닌 표본에서는 변곡점이 아님을 나타내기 위하여 플래그만을 '0'으로 설정하여 전송한다. 따라서, 부호화기의 구조는 그림 1과 같다. 그림 2는 이와 같은 방법으로 부호화된 음성신호의 전송비트열의 예를 보여준다. 변곡점 표본 값을 8비트 양자화 하여 전송하는 비트열의 예시이다 [5].

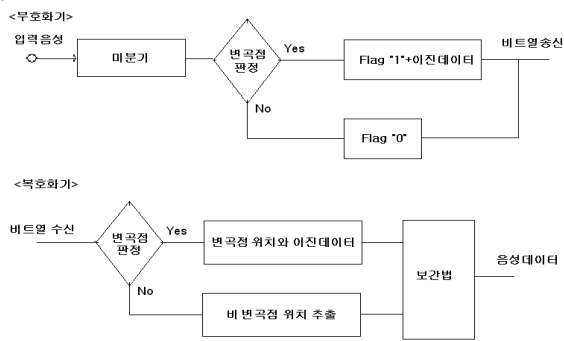


그림 1 음성 부호화기/복호화기 구조 [5]
Fig. 1 Structure of the speech coder [5]

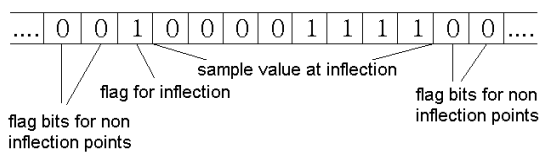


그림 2 전송 데이터 형식 [4]
Fig. 2 Bit stream for proposed encoding scheme [4]

복호화기에서는 수신된 비트 열 가운데 '1'로 설정된 비트 바로 뒤의 정해진 비트수의 값을 복호하여 변곡점의 위치와 값을 복원한다. 그리고 0으로 수신된 플래그에 해당하는 위치에서의 신호 값을 보간법을 이용하여 복원한다.

2.2 변곡점 검출 알고리즘

변곡점을 검출하기 위하여 신호의 기울기에 해당하는 미분의 근사값을 이용한다. 국소적인 최대값 혹은 최소값 전후로는 미분 값의 부호가 바뀐다는 점을 활용한다. 아울러, 단순 감소 구간 혹은 단순 증가 구간에서도 기울기가 변화할 수 있다는 것에 유의한다. 따라서 변곡점 검출 알고리즘을 아래와 같이 정리할 수 있다. 즉, 음성 표본은 아래의 조건 가운데 하나라도 만족하면 변곡점으로 판정한다 [4,5].

조건1) 연속하는 미분 근사값의 곱이 음수이다. 즉,

$$y[n]y[n-1] < 0$$

조건2) 연속하는 미분 근사값의 차가 미리 설정된 임계값 (thr) 보다 크다. 즉, $|y[n] - y[n-1]| > thr$

첫번째 조건은 국소적인 최대값 혹은 최소값 발생에 따른 변곡점을 검출하기 위한 조건이다. 두 번째 조건은 음성 신호의 기울기가 갑자기 변화하는 위치를 검출하는 조건이다. 이와 같은 변곡점 검출 알고리즘을 흐름도로 나타내면 아래 그림 3과 같다 [4,5].

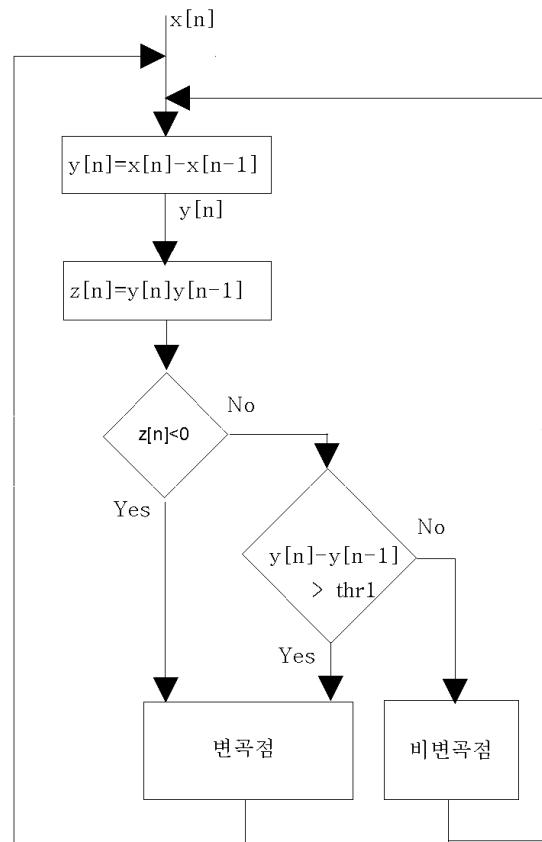


그림 3 변곡점 추출 흐름도 [4,5]
Fig. 3 Flow chart for the inflection point detection [4,5]

3. 가변 비트율 부호화기의 주요 인자

2장에서 소개한 부호화기의 구조를 고려할 때 아래와 같이 성능에 영향을 줄 수 있는 인자를 고려할 수 있다.

3.1 보간법에 따른 성능의 차이

복호화기에서는 보간법을 사용하여 비변곡점에서의 표본값을 추정한다. 대표적인 보간법으로는 선형 보간법, 큐빅 스플라인 보간법 등이 있다 [7]. 보간법에 따라서 다른 복호 결과를 보일 것이므로 보간법에 따른 복호된 음성의 품질을 검증한다.

3.2 변곡점 검출 알고리즘의 임계값

변곡점을 추출하기 위하여 이웃하는 미분 근사값을 비교하기 때문에 임계값의 설정에 따라서 조건 2의 만족 여부가 달라질 수 있다. 임계값이 너무 크면 작은 기울기의 변화는 검출되지 않고 무시된다. 따라서, 검출되는 변곡점의 수가 감소하며, 데이터 압축율이 좋아질 것이다. 그러나, 수신단에서 복호된 신호의 품질이 저하되는 문제가 있다. 반면에 임계값이 너무 작으면, 작은 기울기의 변화도 검출되어 많은 변곡점이 추출될 것이다. 따라서 데이터의 압축율은 낮은 단점이 있으나, 음성의 품질은 원음성에 근접할 것이다. 따라서 조건 2의 임계값을 이용하여 데이터의 압축율과 품질을 조정할 수 있다.

3.3 변곡점 표본을 양자화하는 방법

변곡점으로 검출된 표본은 이진부호로 양자화 되기 때문에 양자화 오차에 따른 상대적인 품질 저하를 개선하기 위하여, PCM에서와 같은 로그 컴펜딩의 활용 가능성을 검토한다 [1]. 즉, 변곡점 추출 전 혹은 추출 후, 로그 컴펜딩을 적용할 때의 효과를 검증한다.

4. 음성신호로의 응용 및 성능 분석

본 장에서는 먼저 음성신호에 가변 비트율 부호화기를 적용하여 부호화기로서 유효함을 보인다. 아래 그림 4와 5는 유성음과 무성음 구간에서 원음성, 가변 부호화기로 검출한 변곡점 표본, 복호된 신호를 보인다. 그림에서 보이듯이 부호화기는 적절하게 변곡점을 검출하며, 복호화기에서는 보간법을 통해 비변곡점에서의 표본값을 적절하게 추정함을 알 수 있다. 복호화기의 보간법으로는 선형보간법을 사용하였다.

4.1 보간법에 따른 성능 비교

아래 표 1은 복호된 신호의 신호대 잡음비의 보간법에 따른 차이를 보인다. 여기에서 잡음은 원 음성표본에 대한 복호된 신호의 차이를 말한다. 사용된 음성은 4가지 표본 음성으로 성인 남녀의 음성이 사용되었다. 음성 표본 #1 ~ 3은 유성음과 무성음이 적절히 조합된 음성이며, 음성 표본 #4는 유성음으로 구성되었다. 신호대 잡음비로 측정한 객관적인 성능평가에 따르면, 선형 보간법이 큐빅 스플라인 보간법보다 0.9 ~ 5.7 dB 까지 우수한 성능을 보임을 알 수 있다. 이는 음성 신호가 변곡점과 변곡점 사이의 구간에서 선형적인 특성에 가깝기 때문으로 사료된다. 즉, 선형 보간법은 두 표

본 사이에서 선형적으로 중간 값을 추정하지만, 큐빅 스플라인은 3차 함수를 이용하여 비선형적으로 중간 값을 추정하기 때문에, 결과 과형에서 선형 보간 결과가 원음성에 더 근

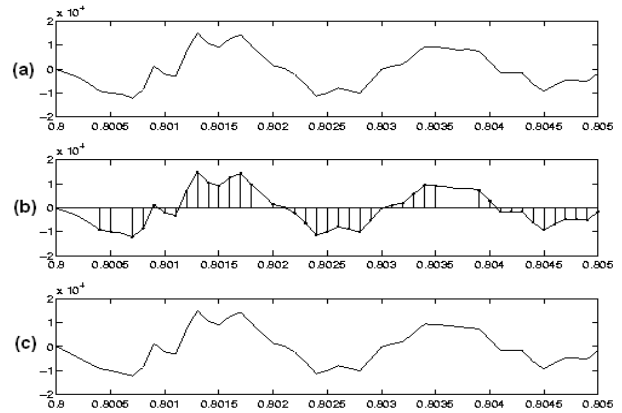


그림 4 유성음 적용예제 (a)원음성 (b) 추출된 변곡점 (c) 복호된 신호

Fig. 4 Application example of the proposed method for the voiced part (a) original speech (b) detected inflection points with original speech (c) decoded speech through interpolation

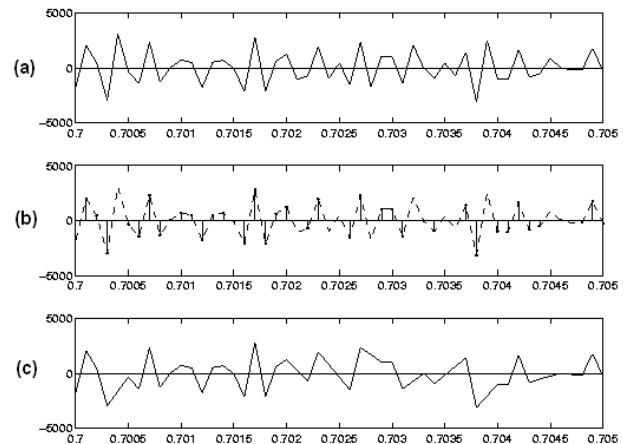


그림 5 무성음 적용예제 (a)원음성 (b) 추출된 변곡점 (c) 복호된 신호

Fig. 5 Application example of the proposed method for the unvoiced part (a) original speech (b) detected inflection points with original speech (c) decoded speech through interpolation

표 1 선형 보간법과 큐빅 스플라인 보간법의 비교

Table 1 Comparison of linear and cubic spline interpolation

음성표본	비트율 (kbps)	신호대 잡음비 (dB)	
		선형보간법	큐빅스플라인
#1	44	17	11.3
#2	41.4	21.1	16.1
#3	44.2	16.8	14.3
#4	38.6	16.7	15.8

접하는 것으로 여겨진다. 따라서 여타의 다른 모의시험에서는 음성 복호시 선형 보간법을 사용한다.

4.2 변곡점 검출시 임계값에 따른 압축율과 신호대 잡음비

3장에서 전술했듯이 임계값이 크면 검출되는 변곡점은 주로 국소적인 최대값과 최소값으로만 이뤄지며 따라서 최대의 데이터 압축율을 실현할 수 있다. 그러나 복호된 음성의 질의 저하된다. 반면 임계값이 작으면 대부분의 변곡점이 검출되어 음질은 개선되고 압축율은 저하된다. 아래 표 2는 임계값에 따른 데이터 압축율과 신호대 잡음비를 보여준다. 참고로 파형 부호화 방식 데이터 압축기법인 32 kbps Continuous Variable Slope Delta Modulation (CVSDM)의 신호대 잡음비 결과를 비교를 위하여 제시하였다. 대체로 임계값이 작은 경우 40 kbps 이상의 비트율을 보이며 임계값이 큰 경우 35 kbps 이하의 비트율을 보인다. 신호대 잡음비 역시 임계값과 관련이 있음을 볼 수 있다. 가변 비트율 음성부호화기가 작은 임계값에서 대체로 32 kbps CVSDM 보다 우수한 신호대 잡음비를 보인다.

4.3 주관적인 성능평가

음성신호는 실제 사용자가 청취하는 신호이기 때문에 신호대 잡음비 같은 객관적인 성능지표와 함께 주관적인 성능지표 역시 중요하다. 대표적인 주관적인 성능지표로는 Mean Opinion Score (MOS) 가 있다. MOS는 탁월한 경우 5, 매우 듣기 거북한 음질의 경우 1로 점수를 매겨 평균을 구한다. 아래 표 3은 표 2에서와 같이 두 가지 다른 임계값에 대한 MOS 시험 결과이다. 20세에서 25세 사이의 성인 남녀 134명이 참여하여 평가한 결과이다. 대체로 임계값이 작은 경우 CVSDM 보다 탁월한 음질을 보이며, 작은 임계값으로 데이터 압축을 많이 한 경우 CVSDM과 유사한 음질을 보인다.

4.4 로그 컴팬딩의 효과

음성표본을 이진 부호화하는 PCM 부호화에서는 양자화에 앞서서 음성을 비선형적으로 확장하는 로그 컴팬딩을 사용한다 [1]. 즉, 0에서 최대값 사이에서 진폭이 작은 신호는 더 크게, 진폭이 큰 신호는 적게 확장하여 진폭이 작은 신호의 상대적인 양자화 오차를 감소시킨다. 이와 같은 로그 컴

팬딩을 사용하면 PCM의 경우 신호대 잡음비가 약 24 dB 개선되어 양자화로 환산하면 약 4 bit를 더 사용하는 효과를 볼 수 있다 [1].

표 3 MOS를 사용한 주관적 성능의 비교

Table 3 Subjective performance in MOS

음성표본	변곡점 기반 가변비트율 부호화기			32kbps CVSD MOS
	임계값	비트율(kbps)	MOS	
#1	0.016	44	4.05	2.71
	0.16	32.3	2.77	
#2	0.016	41.4	4.02	3.03
	0.16	34.4	2.83	
#3	0.016	44.2	3.88	2.63
	0.16	34.8	2.59	
#4	0.016	38.6	3.72	3.38
	0.16	26.6	2.39	

가변 비트율 음성 부호화기에서는 로그 컴팬딩을 변곡점 추출 전의 원시신호에 적용하거나, 변곡점 추출 후 변곡점의 음성표본에 적용하는 두 가지 가능성을 생각할 수 있다. 먼저, 변곡점으로 검출된 음성표본에 적용하는 경우는 검출된 표본을 이진 부호화 하므로 진폭이 작은 음성표본은 로그 컴팬딩으로 PCM에서와 같이 상대적인 양자화 오차를 감소시키는 효과를 볼 것이다. 다만 변곡점 추출 전에 원음성을 로그 컴팬딩을 하는 것은 변곡점 추출에 영향을 주므로 적절치 않을 것으로 예상된다.

표 4와 그림 6은 원음성에 로그 컴팬딩을 적용하고 변곡점을 검출한 후, 전송했을 때의 성능 분석 결과를 보인다. 복호에는 선형보간법이 사용되었으며, 로그 컴팬딩으로는 μ -law 컴팬딩을 사용하였다 [1]. 비교를 위하여 표 2의 임계값이 작은 경우의 결과와 같은 조건에서 시험평가 한다. 따라서 표 4의 둘째, 셋째 열은 표 2의 결과에서 차용했다. 표 4의 넷째와 다섯째 열에서 보이듯이 신호대 잡음비는 μ -law 컴팬딩 미적용시와 유사한 결과를 보인다. 그러나 데이터율은 μ -law 컴팬딩 미적용시에 비교해서 매우 높은 결과를 보인다. μ -law 컴팬딩 적용시 높은 비트율 결과는, 그림 6에서 보이듯이 μ -law 컴팬딩 후에 변곡점을 추출하면, 신호 파형의 기하학적 특성이 변화해서 더 많은 변곡점이 추출되기 때문으로 사료된다. 그림 6은 (a)에 원음성, (b)에 μ -law 컴팬딩된 음성, (c)에 원음성의 변곡점 추출 결과, 그리고 (d)에 컴팬딩된 신호의 변곡점 추출결과를 보인다. 따라서, μ -law 컴팬딩은 변곡점을 추출한 후에 변곡점으로 검출된 음성표본에만 적용하는 것이 바람직할 것으로 판단된다.

표 4 μ -law 컴팬딩 사용유무에 따른 객관적 성능의 비교

Table 4 Objective performance comparison on μ -law companding

음성표본	μ -law 컴팬딩 미적용시		μ -law 컴팬딩 적용시	
	비트율(kbps)	SNR (dB)	비트율(kbps)	SNR (dB)
#1	44	17	51.4	15.5
#2	41.4	21.1	46.0	17.6
#3	44.2	16.8	50.1	15.7
#4	38.6	16.7	51.7	19.5

표 2 신호대 잡음비에 따른 객관적 성능의 비교

Table 2 Objective performance in SNR

음성표본	변곡점 기반 가변비트율 부호화기			32kbps CVSD 신호 대 잡음비
	임계값	비트율 (kbps)	신호 대 잡음비(dB)	
#1	0.016	44	17	12.22
	0.16	32.3	11.1	
#2	0.016	41.4	21.1	12.38
	0.16	34.4	11.4	
#3	0.016	44.2	16.8	11.88
	0.16	34.8	9.6	
#4	0.016	38.6	16.7	16.3
	0.16	26.6	9.8	

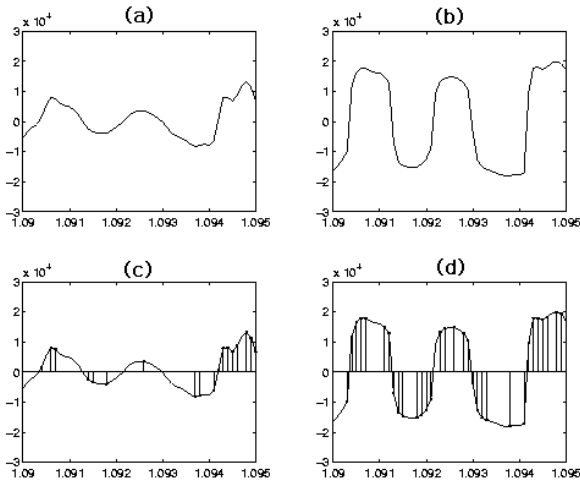


그림 6 유성음에 μ -law 컴팽딩 적용예제 (a)원음성 (b) μ -law 컴팽딩된 신호 (c)원음성에서 추출된 변곡점 (d) μ -law 컴팽딩된 신호에서 추출된 변곡점
Fig. 6 μ -law companding application example of a voiced part (a) original speech (b) μ -law companded signal (c) detected inflection points with original speech (d) detected inflection points with companded signal

5. 결 론

변곡점 추출에 기반한 가변 비트율 음성부호화기가 소개되었다. 가변 비트율 음성부호화기는 음성파형이 국소적으로는 선형적인 특성을 보인다는 관찰에 기초한다. 따라서 부호화기는 음성파형의 기울기가 변화하는 변곡점을 추출하며, 추출된 변곡점의 위치정보는 1 bit의 플래그로, 표본 값은 이진부호로 전송한다. 비변곡점 음성표본 값은 전송되지 않으며, 그 위치정보만 1 bit의 플래그로 전송된다. 복호화기에서는 수신된 변곡점 표본의 값과 위치정보, 그리고 비변곡점의 위치정보를 이용하여 보간법으로 비변곡점 표본 값을 추정한다. 변곡점 검출시에 설정하는 임계값을 조정하여 비트율과 복호된 음성의 품질을 조절할 수 있다. 즉, 임계값을 낮게 설정하면 40 kbps를 초과하는 비교적 높은 비트율과 16 dB 이상의 높은 신호대 잡음비를 얻으며, 임계값을 높게 설정하면 35 kbps 미만의 비교적 낮은 비트율과 12 dB 미만의 낮은 신호대 잡음비를 얻는다. 주관적인 성능평가의 경우, 낮은 임계값에서는 대체로 CVSDM 보다 우수한 MOS 점수를 보이며 높은 임계값에서는 CVSDM 보다 낮은 MOS 점수를 보인다. 아울러 복호화기에서 적용되는 보간법의 유형, μ -law 컴팽딩의 적용 등의 조건을 변화시키며 가변 비트율 부호화기의 성능을 분석하였다. 분석결과 보간법으로는 선형 보간법이 큐빅 스플라인 보간법에 비교해서 신호대 잡음비가 약 1~5 dB 개선된다. μ -law 컴팽딩은 변곡점 검출 후에, 변곡점으로 검출된 음성표본에 적용하는 것이 바람직한 것으로 판단된다. 원음성에 μ -law 컴팽딩을 적용하면, 신호대 잡음비는 컴팽딩이 없는 부호화기의 신호대 잡음비와 유사하나 비트율 면에서 4 ~ 14 kbps 까지 증가한다.

References

- [1] L. Rabiner and R. Schafer, Digital Processing of Speech Signals, Prentice-Hall, 1978.
- [2] A. M. Kondoz, Digital Speech, Wiley, 1994.
- [3] ITU-T Study Group 16, ITU-T Recommendation G.728 Annex H, July, 1997.
- [4] B. Iem, "Variable bit-rate speech coding technique based on inflection point detection," submitted to Journal of Acoustical Society of America, 2013.
- [5] B. Iem, "A Variable Data Rate Speech Coding Technique Based on the Inflection Point Detection of Speech," Trans. of KIEE, vol. 62, no. 4, pp.562-565, 2013.
- [6] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal-Part 2: algorithms and applications," Proc. of the IEEE, vol. 80, pp.540-568, 1992.
- [7] P.J. Davis, Interpolation and Approximation, Dover, 1975.

저 자 소 개



임 병 관 (林炳寬)

1988년 연세대 전자공학과 졸업, 1998년 Univ. of Rhode Island 전기공학과 졸업 (공학박사), 2002년 ~ 현재 강릉원주대학교 전자공학과 부교수
 Tel : 033-640-2426
 Fax : 033-646-0740
 E-mail : ibg@gwnu.ac.kr