

# 음성의 변곡점 추출 및 전송에 기반한 가변 데이터율 음성 부호화 기법

## A Variable Data Rate Speech Coding Technique Based on the Inflection Point Detection of Speech

임 병 관<sup>\*</sup>  
(Byeong-Gwan Iem)

**Abstract** - A new variable rate speech coding technique is proposed. The method is based on the observation that the speech signal approximately looks linear for a very short period of time. The information transmitted is the location and data value of inflection points. If the distance between the inflection points is large, the mid point location and its data value are also delivered. Thus, the encoder transmits both the location and the data value for the inflection samples, but the location only for the non-inflection points. The location information is expressed using one bit for each sample, 0 for non-inflection and 1 for inflection point. At the receiver, using the interpolation, the decoder estimates the untransmitted sample values for non-inflection locations from the received sample values for the inflection samples. With 50 % of computational cost of the existing CVSD delta modulation, the proposed method is expected to achieve the data rate of 36 to 38 kbps and the SNR of 10 to 13 dB.

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

### 1. 서 론

음성 부호화 및 복호화는 신호처리의 고전적인 연구과제이며, 최근의 무선 이동통신의 발달로 지속적으로 관심을 받아온 분야이다. 고전적으로는 다양한 파형부호화 기법이 개발되어 왔으며, 보다 나은 데이터 압축을 위하여 선형예측분석에 기반한 다양한 부호화 기법이 연구되어 왔다 [1]. 파형부호화 방식으로는 펄스코드변조(Pulse Code Modulation, PCM), 델타변조(Delta Modulation) 등의 방법이 있다. 이러한 파형부호화 방식은 유선 공중전화망 등에 활용되고 있다. 파형부호화 기법은 16 ~ 64kbps의 데이터 속도를 보인다. 보다 나은 데이터 압축을 위하여 선형예측분석을 활용한 음성부호화 방식이 연구되어왔다 [1]. 보코딩(voice coding)으로 통칭되는 이러한 부호화 방식은 음성의 특성 파라미터를 추출하여 전송하는 기법으로 약 1.2 ~ 19.2 kbps의 데이터 전송율을 보이며, 최근 이동통신의 음성부호화 방식으로 활용되고 있다 [1]. 파형부호화에 비교하여 탁월한 데이터 압축율을 보이는 보코딩 방식은 파형부호화 보다 많은 계산량을 요구한다. 따라서 데이터 압축이라는 측면에서 음성부호화 분야는 성숙된 분야라 할 수 있지만 계산량 면에서는 개선의 여지가 있다고 볼 수 있다.

본 연구에서는 기존의 음성부호화 방식과는 다른 새로운 음성부호화 기법을 제안한다. 새로운 방법은 음성이 국소적으로 선형적인 형태에 가깝다는 관찰에 착안하여, 변곡점의

위치를 추출한다. 추출된 변곡점들은 그 위치 정보와 샘플 값의 이진값이 전송된다. 비변곡점들의 경우, 위치정보만을 전송한다. 따라서 변곡점과 비변곡점을 구분하는 위치정보를 위하여 1 bit가 할당되며, 변곡점의 경우 그 값을 양자화하기 위해 고정된 비트수가 할당된다. 따라서 모든 샘플값을 전송하는 PCM방식에 비교해서 우월한 데이터 압축율을 보인다. 아울러 비교적 적은 계산량으로도 변곡점들을 추출할 수 있기 때문에 보코딩 방식에 비교하여 계산량이 절감된다.

본 논문의 구조는 다음과 같다. 먼저 2장에서는 제안된 음성부호화/복호화 방식에 대한 자세한 소개를 한다. 3장에서는 음성신호에 대한 부호화 및 복호화 결과를 보인다. 4장에서는 Blue Tooth에서 활용되는 기존의 파형부호화 방식인 CVSD 델타변조 방식의 음성부호화 기법과 비교한다. 끝으로 5장에서 결론을 맺는다.

### 2. 변곡점 추출에 기반한 음성부호화 및 복호화 기법

아래 그림 1은 전형적인 음성 신호를 확대한 그림이다. 그림 1의 (b)는 0.5~0.505초 부분을 확대한 그림으로 유성음에 해당하고, 그림 1의 (c)는 0.7~0.705초 부분을 확대한 그림으로 무성음에 해당한다.

그림에서 알 수 있듯이 국소적으로 음성은 변곡점과 변곡점 사이에서 선형적인 형태를 보인다. 따라서 선형적인 음성파형의 기울기가 변화하는 위치를 추출하여 변곡점으로 정의한다. 송신측은 추출된 변곡점과 그곳에서의 값을 전송하고, 변곡점 사이의 신호 값을 전송하지 않으므로 데이터 전송율을 낮출 수 있다. 수신 측에서는 수신된 변곡점 위치

<sup>\*</sup> Corresponding Author : Department of Electronic engineering, Gangneung-Wonju National University, Korea  
E-mail : ibg@gwnu.ac.kr

Received : January 9, 2013; Accepted : March 25, 2013

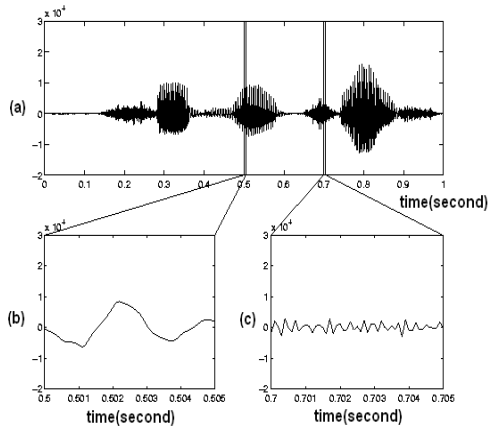


그림 1 확대된 음성신호의 예  
Fig. 1 Enlarged example of speech signal

와 값을 사용하여 전송이 안 된 음성샘플 값을 내삽법 등으로 추정할 수 있다. 기울기는 미분으로 주어지고, 미분은 이산시간 영역에서 이웃하는 샘플의 차인 차동방정식으로 근사화 되므로 아래와 같은 차동방정식을 이용한다 [2].

$$y[n] = x[n] - x[n-1] \quad (1)$$

여기에서  $x[n]$ 은  $n$ 이라는 시간에서의 음성샘플이며  $y[n]$ 은 미분의 근사치이다. 이러한 미분 근사치의 변화를 관찰하여 변곡점을 추출한다. 변곡점 추출의 조건은 아래와 같다.

- i) 연속하는 미분근사치의 곱  $y[n] \cdot y[n-1]$ 이 음수이면 변곡점이다.
- ii) 연속하는 미분 근사치의 곱이 양수이나 그 차가 임계값1 보다 크면 변곡점이다.
- iii) 위 조건 i)과 ii)에 해당이 안되면 비변곡점이다.
- iv) 변곡점과 변곡점 사이의 거리가 임계값2 보다 크면 그 중간 샘플도 전송한다.

조건 i)은 국소적인 최대값 또는 최소값의 발생에 따른 변곡점에 해당한다. 조건ii)는 기울기가 급격하게 변화하는 변곡점에 해당한다. 조건 iv)는 변곡점은 아니지만 비변곡점으로 판정된 샘플이 과다할 때 복호시 원신호를 충실히 복원 못하는 문제를 해결하기 위하여 변곡점 사이의 거리가 미리 정한 임계값 보다 클 때, 중간 샘플을 강제적으로 전송하는 조건이다. 조건 i), ii), iv)를 만족시키지 못하는 샘플은 비변곡점으로 판정한다. 아래 그림 2는 이러한 변곡점 추출 과정을 보이는 흐름도이다.

8비트로 양자화되는 경우 전송되는 데이터의 형식은 아래 그림 3과 같다. 변곡점으로 판정된 샘플은 변곡점 샘플임을 알리는 1로 설정된 하나의 플래그(flag) 비트와 이진수로 양자화된 샘플로 전송된다. 비변곡점 샘플들은 비변곡점 샘플임을 알리는 0으로 설정된 플래그 비트만 전송된다. 예를 들어, 비변곡점 샘플 3개와 변곡점 샘플 1개, 비변곡점 샘플 1개, 변곡점 샘플 1개, 비변곡점 샘플 2개 순서로 전송이 된다면 전송 데이터열은 아래 그림 4와 같다. 여기에서 변곡점 샘플 값은 편의상 10과 20이라고 가정하고 샘플당 8비트로 양자화 되었다고 가정한다.

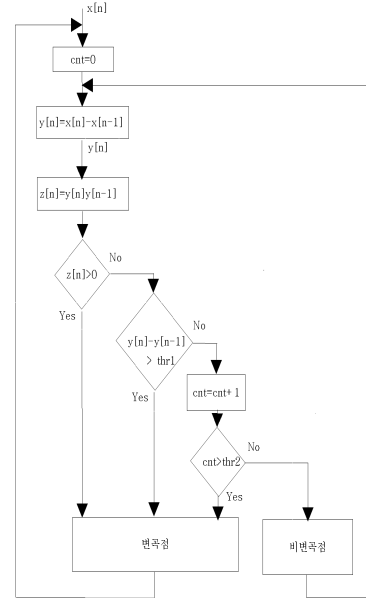


그림 2 변곡점 추출 흐름도  
Fig. 2 Flow chart for the inflection point detection

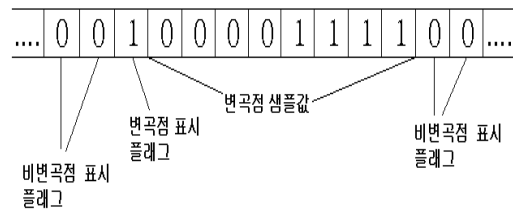


그림 3 전송 데이터 형식  
Fig. 3 Transmitted data format

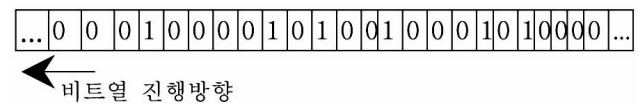


그림 4 제안된 방식에 따른 전송 데이터열 예  
Fig. 4 Transmitted data stream example

복호화기에서는 0로 수신되는 비트열을 비변곡점 샘플들로 판정하며, 1이 수신될 때, 그 비트 뒤의 8비트를 변곡점 샘플값으로 판정한다. 그리고 판정된 변곡점 샘플값들을 활용하여 비변곡점 샘플 위치에서의 값을 내삽법을 이용하여 복호한다. 따라서, 부호화기와 복호화기는 아래 그림 5와 같은 구조를 갖는다.

### 3. 음성 부호화 및 복호화 예제

본 장에서는 실제 음성신호에 제안된 방법을 적용하였다. 제안된 부호화 방법에는 10kHz로 표본화된 음성신호를 사용하였다 [3]. 음성은 약 2.1248초 분량이 사용되었다. 따라

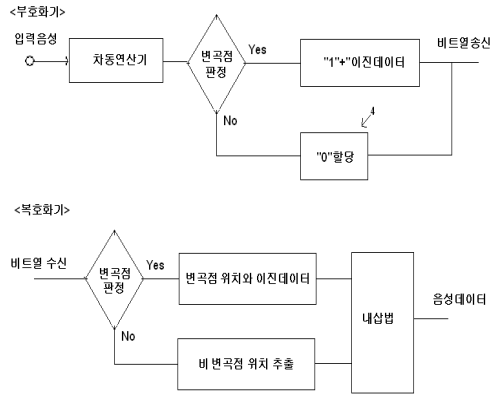


그림 5 제안된 음성 부호화기/복호화기 구조  
Fig. 5 Structure of the encoder and decoder

서 총 샘플 수는 21248 샘플이다. 원음성을 8bit로 양자화된 PCM방식으로 부호화하면 총 데이터의 양은 169,984 bits이다. 원음성에 제안된 변곡점 추출을 이용한 방법을 적용한 결과는 다음과 같다. 변곡점으로 9,447 샘플이 추출되었고, 비변곡점은 11,801 샘플이 결정되었다. 따라서 전송되는 전체 데이터의 양은 변곡점샘플 9,447 x 9 비트/ 샘플 + 비변곡점샘플 11,801x 1 비트/샘플 = 96,824 비트이다. 따라서 데이터 압축율은 43% 임을 알 수 있다. 본 예제에서 알 수 있듯이 제안된 방법은 가변 비트율 부호화 방식이다. 즉 기존의 부호화 방식이 고정된 데이터 전송율의 부호화 방식인 것과 비교하여, 제안된 변곡점 추출 방식의 부호화는 음성 샘플에 따라서 가변의 데이터 전송율을 보일 것이다. 아래 그림 6과 7은 음성신호를 본 논문에서 제안한 변곡점 추출 방식으로 부호화하고 내삽법으로 복호한 신호와 CVSD 방식 델타변조로 부호화와 복호화를 한 결과를 보인다. CVSD 방식 델타 변조는 32 kHz로 표본화된 데이터 샘플을 사용하여 부호화 하였다. 그림 6은 원음성 신호, CVSD 델타변조에서 복호된 음성신호, 그리고 원음성, 변곡점 추출결과와 복호된 음성 신호를 확대한 그림이다. 수신단에서 변곡점 샘플들로부터 음성을 복호하기 위하여 사용된 내삽법은 큐빅 스플라인(cubic spline) 내삽법이다 [4]. 확대된 그림에서

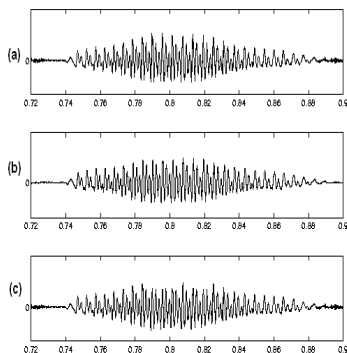


그림 6 음성 복호화 예제 (a)원음성 (b) CVSD 델타변조의 복호된 신호 (c) 제안된 복호된 신호

Fig. 6 Decoding example (a) original speech (b) decoded signal by CVSD (c) decoded signal by the proposed method

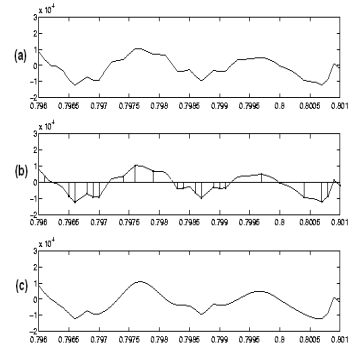


그림 7 음성 부호화/복호화 예제 (a)원음성 (b) 추출된 변곡점 신호 (c) 복호된 신호

Fig. 7 Encoding/decoding example (a) original speech (b) detected inflection points (c) decoded signal by the proposed method

알 수 있듯이 제안된 부호화 방식이 원음성 신호를 성공적으로 복호함을 알 수 있다. 그림 7은 제안된 알고리즘으로 변곡점을 추출한 결과와 내삽법으로 복호된 신호를 확대하여 보여주고 있다. 그림 (b)는 추출된 변곡점 신호와 원음성을 중첩해서 보여준다. 확대된 그림에서 알 수 있듯이 주요 변곡점들이 추출되었으며 수신단에서 내삽법을 통하여 성공적으로 음성이 복호된다.

#### 4. 실험 및 고찰

##### 4.1 소요 계산량

이절에서는 CVSD 델타변조 부호화 방식과 제안된 가변 비트율 부호화 방식의 계산량을 비교한다. CVSD의 경우는 음성 표본의 변화 정도에 따라서 계산량이 차이가 날 수 있으며, 구체적으로는 부호화되는 음성 N 표본에 대하여 2N 혹은 3N의 곱셈, 2N 혹은 3N의 덧셈과 2N의 비교가 필요하다. 제안된 가변 부호화 방법은 N의 곱셈, 2N의 덧셈 그리고 3N 이하의 비교가 요구된다. 따라서 CVSD 델타 변조에 비교해서 계산량을 절감할 수 있다. 아래 표 1은 두 방식 사이의 계산량을 음성 표본 당 정리한 것이다.

표 1 음성 표본당 계산량의 비교

Table 1 Comparison of computational cost

음성부호화 방식	곱셈	덧셈	비교
CVSD 델타변조	2~3	2~3	2
제안된 가변부호화	1	2	<=3

##### 4.2 객관적인 성능의 비교

복호된 신호와 원음성의 차이를 부호화 잡음으로 정의하여, 객관적인 성능의 지표인 신호대 잡음비 (SNR)를 구했다. 아래 표 2는 두가지 음성 표본에 대하여 SNR을 구한 결과이다. 음성표본 1은 무성음으로, 음성표본 2는 유성음으로 적절히 조합된 영어문장이다. SNR(dB) 값을 구하면 제안된 방법은 약 10 ~ 13 dB의 신호대 잡음비를 보인다. 동일한 데이터를 이용하여 CVSD방식의 델타변조의 신호대 잡음비를 구하였을 때 약 10.77 ~ 12.23 dB의 결과가 얻어졌

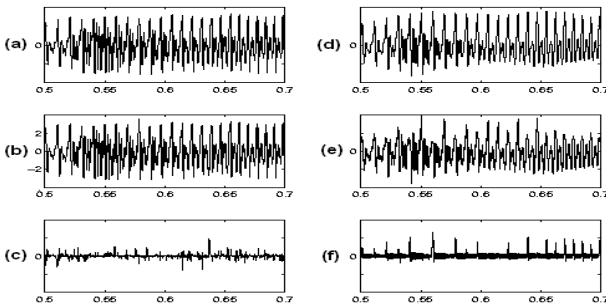
다. 본 실험에 사용된 데이터가 제한적이기 때문에 다양한 데이터에 대한 정량적인 분석이 필요하지만, 본 논문에서 제안된 방법이 CVSD 델타 변조에 비교해서 곱셈의 경우 약 50%의 계산량만으로도 우수한 데이터 압축을 실현하며, CVSD 방식의 델타 변조와 유사한 객관적인 성능을 보이는 것을 알 수 있다.

**표 2** 신호대 부호화 잡음비와 데이터 전송속도 결과

**Table 2** Performance of SNR and Data Rate

구분	제안된 방식		32kbps CVSD
	SNR	Data Rate	ADM
표본 1	10.3 dB	38.4 kbps	12.23 dB
표본 2	13.2 dB	36.48 kbps	10.77 dB

아래 그림 8은 음성의 일부 구간에서 원음성, 제안된 가변 데이터를 부호화 방법과 CVSD 델타변조방식으로 복호된 신호, 그리고 각각의 부호화 잡음을 보여준다. 그림에서 제안된 방법이 신호를 충실하게 복호함을 알 수 있다. 그리고 부호화 잡음도 CVSD델타 변조와 비교해서 그 값이 적음을 알 수 있다.



**그림 8** 제안된 방법의 (a) 원음성 (b) 복호된 신호 (c) 부호화 잡음 신호와 CVSD의 (d) 원음성 (e) 복호된신호 (f) 부호화 잡음

**Fig. 8** Encoding/Decoding example (a) original speech (b) decoded signal (c) coding noise by the proposed method (d) original speech (e) decoded signal (f) coding noise by CVSD

### 5. 결 론

본 논문에서는 음성이 국소적으로 선형에 가까운 특성을 보인다는 관찰에 기초하여 새로운 가변전송율의 음성부호화 방식을 제안하였다. 제안된 음성부호화 방식은 음성의 변곡점을 추출하여 변곡점의 위치를 1 비트('1')로 표시하고 변곡점에서의 샘플값을 이진부호로 전송한다. 변곡점이 아닌 샘플의 값은 전송되지 않으며, 그 위치만 1 비트('0')로 표시된다. 변곡점이 아닌 샘플이 일정 시간 이상 지속되면 중간위치에서의 샘플값도 전송된다. 제안된 가변전송율의 음성부호화 방식은 CVSD 델타변조 방식에 비교하여 계산량을 최대 50% 가량 절감할 수 있으며, 객관적인 성능으로는 CVSD 델타변조와 유사한 전송율과 SNR 성능을 보인다. 향후 과제로는 음성신호는 최종적으로 인간이 실제 소리를 듣

는 것으로 신호전달이 종료되기 때문에 객관적인 성능의 평가뿐 아니라 MOS (Mean Opinion Score)를 사용한 주관적인 성능평가가 요구된다 [5]. 또한 본 논문에서는 복호화기에서 큐빅 스플라인 방식의 내삽법이 사용되었으나 다른 내삽법 등을 사용한 경우의 복호화 성능을 분석할 필요가 있다.

### References

[1] A. M. Kondoz, Digital Speech, Wiley, 1994.  
 [2] 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, Apr. 1992.  
 [3] D. G. Childers, Speech Processing and Synthesis Toolboxes, Wiley, 2000.  
 [4] P. J. Davis, Interpolation and Approximation, Dover, 1975.  
 [5] L.R. Rabinar and R.W. Schafer, Digital Processing of Speech Signals, Prentice-Hall, 1978.

## 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