

# Source controlled variable bit-rate scheme을 이용한 파형 보간 부호화기의 음질 개선 기법

조근석, 양희식, 정상배, 한민수  
한국정보통신대학교 음성음향정보연구실

## Enhanced source controlled variable bit-rate scheme in a waveform interpolation coder

Keunseok Cho, Heesik Yang, Sangbae Jeong, Minsoo Hahn  
Speech and Audio Information Lab, Information and Communication Univ.

cks1223@icu.ac.kr, sheik@icu.ac.kr,  
sangbae@icu.ac.kr, mshahn@icu.ac.kr

### Abstract

This paper proposes the methods to enhance the speech quality of source controlled variable bit-rate coder based on the waveform interpolation. The methods are to estimate and generate the parameters that are not transmitted from encoder to decoder by the repetition and extrapolation schemes. For the performance evaluation, the PESQ(Perceptual Evaluation of Speech Quality) scores are measured. The experimental results shows that our proposed method outperforms the conventional source controlled variable bit-rate coder. Especially, the performance of the extrapolation method is better than that of the repetition method.

### I. 서론

오늘날 음성 부호화기 연구의 초점은 음성 신호에 대한 정보를 낮은 비트 할당을 통해 전송하면서도 음질의 저하를 최소화 하는 데에 중점이 되고 있다. 음성 부호화기 중의 하나인 파형 보간 부호화기(Waveform Interpolation coder)는 파형의 모양과 특징 파라미터를 동시에 전송하는 복합형 코딩(Hybrid coding)기법을 사용하는 코더로 낮은 전송률에서 비교적 좋은 음질을 얻을 수 있는 장점을 지니고 있다

[2][3].

일반적으로 파형 보간 부호화기에 대한 연구는 크게 음질을 개선하는 측면과 전송률을 줄이는 방향으로 이루어져 왔다. 그 중에서도 전송률을 줄이는 방법에 관한 연구로는 음성 신호를 음성의 특징에 따라 분류하고 그에 따른 가변적인 비트 할당을 통해 전송률을 줄이는 방법이 기존에 제안되어 졌다. 그러나 기존의 제안되어진 방법은 전송률을 줄이는 데 초점을 두어 비트 할당 측면에서는 큰 이득을 볼 수 있지만, 가변적인 비트할당에 의한 전송되는 정보의 손실로 인해 음질의 저하를 가져오는 단점이 있었다. 본 논문에서는 기존의 방법에서의 전송률을 유지하면서도 가변적 비트할당으로 인해 전송되지 않는 정보에 대해서 예측을 통한 복원을 해줌으로써 음질을 향상시켜 주는 방법을 제안하였다. 제안한 방법으로는 반복(Repetition)을 통한 정보 생성과 보외법(Extrapolation)을 통한 정보 생성이 있다. 그리고 제안된 방법사이에 음질 개선 결과 차이를 분석한다.

본 논문의 구성은 다음과 같다. 1장의 서론에 이어 2장에서는 기존에 제안된 source controlled variable bit-rate(SC-VBR) 음성 부호화기의 알고리즘에 대해 기술하고, 3장에서는 SC-VBR의 음질 개선을 위해 제안한 방법들을 설명하고, 4장에서는 제안한 방법을 통해 실험한 결과를 분석 및 검토한 후, 마지막으로 5장에서는 결론 및 향후 연구계획에 대해 논한다.

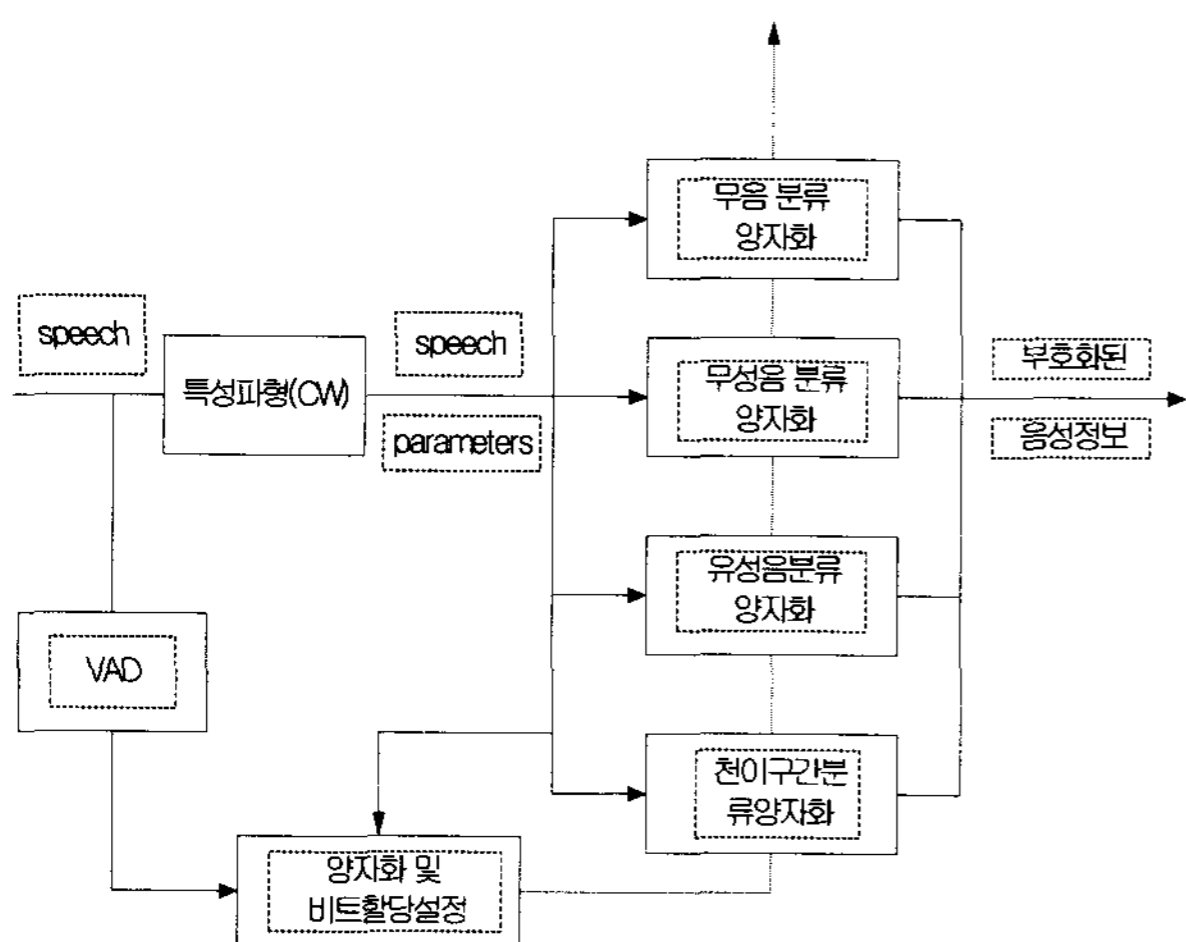


그림 1. SC-VBR 부호기의 구성도

## II. SC-VBR 음성 부호화기

기존에 제안된 SC-VBR 음성 부호화기는 음성 신호를 네 가지 형태로 분류하여 이를 각각 다른 전송률로 전송하는 방식의 부호화기이다. 그림 1은 SC-VBR 부호기의 구성도이다. 추출된 특성 파형(Characteristic waveform)을 이용하는 SC-VBR 음성 부호화기는 음성 끝점 검출기(voicing activity detection)와 특성 파형을 SEW(Slowly evolving waveform)와 REW(Rapidly evolving waveform)로 분리한 후의 SEW/REW 비율과 인접한 특성파형들의 상호상관계수(cross-correlation coefficient)를 이용하여 네 가지 형태로 분류할 수가 있다[1]. 네 가지 분류 형태는 무음, 유성음, 무성음, 천이구간으로 정의되어진다. 네 가지 분류를 통해 각 프레임의 분류 형태가 결정되어진다. 음성 끝점 검출기에 의해 무음 구간이 결정되어 지는데 기존의 제안된 SC-VBR 음성 부호화기에서는 표준화된 음성부호화기인 G.729 음성부호화기의 음성 끝점 검출기를 채택해 사용하였다. 유성음과 무성음, 천이구간은 SEW/REW의 에너지 비율값과 상호상관계수를 분류 요소로 사용하여 분류 되어진다. 유성음 구간의 경우 SEW의 에너지 비율이 REW에 비해 높고 유성음구간의 상호상관계수가 높게 나오는데 비해 무성음 구간에서는 REW의 에너지 비율이 높고 상호상관계수가 낮게 나오기 때문에 이 두 가지 특성을 이용하여 분류되게 된다. 천이구간은 무성음과 유성음의 특징이 같이 나오는 구간이므로 분류 요소의 값이 유성음과 무성음의 중간정도의 값으로 나타나게 된다. 유성음과 무성음, 천이구간을 분류하는 기준인 분류 요소의 경계 값은 실험에 따른 값으로 결정되어진다[1].

표 1. 분류에 따라 전송되는 파라미터 정보들과 프레임 당 할당되는 비트율

Parameters	분류 형태 (bits)			
	무음	무성음	유성음	천이구간
Power	8	8	8	8
Pitch	-	-	7	7
LSFs	-	30	30	30
SEW (amplitude)	-	-	28	28
REW (amplitude)	12	12	-	12
Modes	2	2	2	2
Bit rate(bit/s)	1100	2600	3750	4350

분류 요소에 따라 분류 형태가 정해진 각 프레임은 표 1과 같이 분류 형태에 따라 전송되는 파라미터와 전송되지 않는 파라미터가 정해지고 그에 따른 비트 할당도 달라진다. 천이 구간으로 분류된 프레임인 경우에는 모든 파라미터가 전송된다. 그러나 무음, 유성음, 무성음으로 분류된 프레임일 경우엔 각 분류 형태의 특징에 따라 음질열화를 고려하여 전송하지 않을 파라미터들을 정하여 전송률을 줄일 수가 있다. 전송률을 줄이는 측면에서는 이득을 얻을 수 있지만 분류 선택의 오류와 전송되지 않는 파라미터에 의해서 음질 열화가 어느 정도 발생하게 되는 문제점을 가지고 있다.

## III. 제안한 알고리즘

본 논문에서는 SC-VBR scheme에 의해 발생하는 음질 열화의 문제점을 해결하기 위해 부호기에서 복호기로 전송되어지지 않은 파라미터들을 예측 및 복원하기 위해 repetition과 extrapolation을 이용한 파라미터 정보 복원 방법을 제안하였다.

### 3.1 Repetition method

부호기에서 복호기로 전송되지 않은 파라미터 정보를 생성하기 위해서 이전의 프레임에 전송된 파라미터 정보를 반복하여 사용하는 방법이다. LSF, SEW, REW를 인접한 프레임의 정보를 이용하여 반복적으로 생성하여 복원한다. 특히, SEW 파라미터 정보는 프레임과 프레임 사이에 변화가 크지 않기 때문에 이전에 전송된 프레임의 SEW 파라미터 정보를 이용하여 반복하는 방식으로 생성시키면 음질의 개선을 가져올 수가 있다. Pitch정보는 분류를 위한 할당된 비트 정보를

통해 무음 또는 무성음인지를 판단하여 정의 되어진 피치 최소 값을 갖게 된다. 그림 2는 Repetition 방법을 통해 이전 프레임의 파라미터 정보로 현재 프레임의 전송되지 않은 파라미터를 복원하는 것을 보여준다.

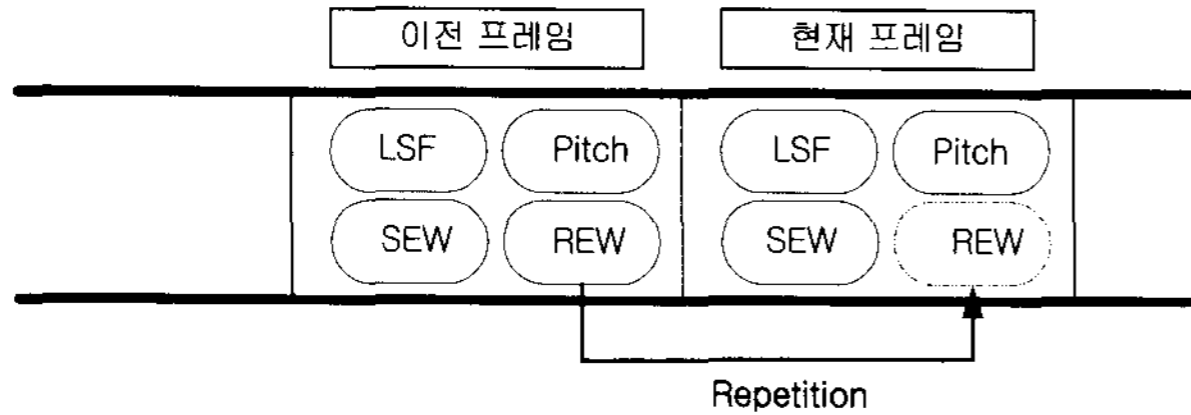


그림 2. Repetition(반복) method

### 3.2 Extrapolation method

부호기에서 복호기로 전송되지 않은 파라미터 정보를 생성하기 위해서 이전의 프레임에 전송된 파라미터 정보를 가지고 선형 보외법을 써서 예측 및 복원하는 방법이다. 이전의 값을 가지고 현재의 값을 선형으로 예측하는 방법이다.

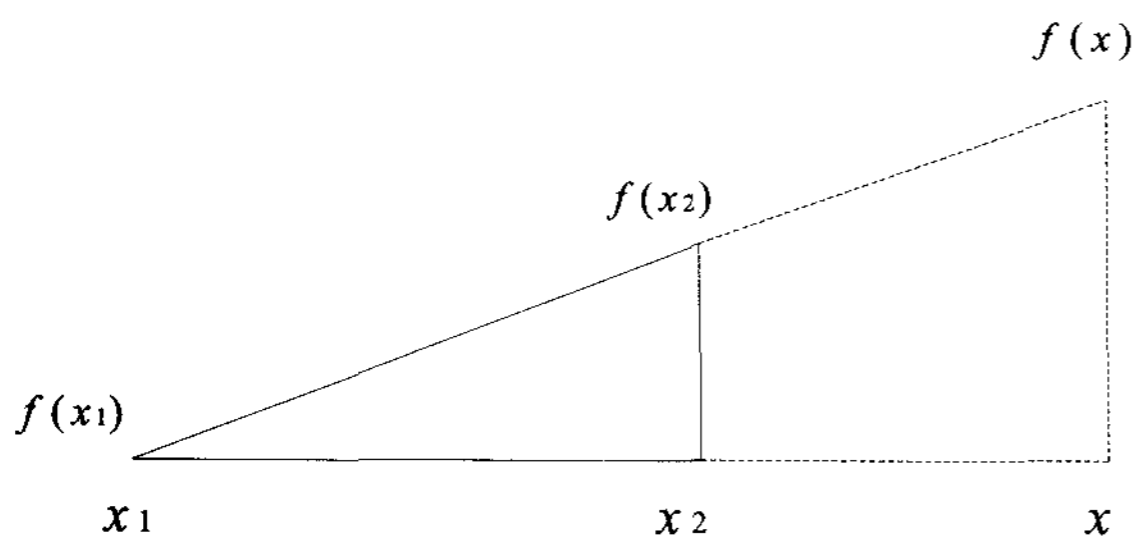


그림 3. Linear extrapolation(선형 보외법)

$$f(x) = f(x_1) + \frac{f(x_2) - f(x_1)}{x_2 - x_1} (x - x_1) \quad (1)$$

그림 3은 선형 보외법을 통해 현재의 값을 구하는 방법을 나타내고 있다. 그림 3과 식(1)에서  $x_1$ 과  $x_2$ 는 차례로 과거의 값을 나타내고  $x$ 는 구하고자 하는 현재의 값을 나타낸다. 식(1)을 통해 과거의 값으로부터 현재의 값을 예측하여 구할 수가 있다. 전송되지 않은 파라미터 정보를 이전의 프레임의 파라미터 정보를 통해 선형 보외법을 이용하여 현재의 파라미터 정보를 예측하여 생성한다. SEW 경우 프레임 사이의 변화가 크지 않으며 완곡한 파형의 형태를 가지고 있기 때문에 선형 보외법으로 근사하게 추정할 수 있다. 그러나 REW 경우 프레임 내에서도 변화가 크고 노이즈 성분과 같은 파형의 형태를 가지고 있기 때문에 선형 보외법을 이용하여 예측 및 복원할 경우 최종 복원된

음성 파형이 발산하는 형태로 나타날 수 있기 때문에 선형보외법은 SEW 파라미터 정보를 예측 및 복원하는데 이용할 수 있다.

## IV. 실험 및 결과

실험을 위해 10초에서 20초 사이 정도의 발성 길이를 가지는 한국어 낭독체 음성들이 사용되었다. 이 음성 데이터들은 여성 화자가 발성한 것이고, 8KHz로 샘플링 되었다. 실험을 위해 총 50문장의 음성 데이터가 사용되었다. 제안한 방법으로 얻은 음성의 음질평가를 위해 객관적인 척도인 PESQ score를 측정하였다. 성능 평가 실험은 SC-VBR scheme을 적용하지 않고 프레임마다 모든 파라미터 정보를 보내는 파형보간 부호화기의 음성결과와 기존의 SC-VBR을 이용한 파형보간 부호화기의 음성 결과, 그리고 제안한 두 가지 방법을 이용하여 얻은 각각의 음성 결과를 가지고 하였다. 제안된 두 가지 방법 중 repetition method를 이용하여 실험한 결과는 전송되지 않은 파라미터중 REW도 포함하여 복원한 실험과 REW를 포함하지 않고 복원한 실험으로 나누어 수행하였다. 그리고 extrapolation method를 이용한 실험은 발산의 가능성을 배제시키기 위해 전송되지 않은 파라미터중 REW를 포함하지 않고 나머지 파라미터들을 복원하는 실험을 수행하였다.

표 2. SC-VBR scheme을 적용하지 않은 기존의 파형 부호기의 프레임 당 파라미터 전송률

WI Parameters	전송률			
	Bits/update	Update Rate	Bits/Frame	Bit/s
Pitch	7	50 Hz	7	350
Power	4	100 Hz	8	400
LSFs	30	50 Hz	30	1500
SEW (amplitude)	14	100 Hz	28	1400
REW (amplitude)	3	200 Hz	12	1200
총 전송률:			85	4250

표 2와 같이 SC-VBR scheme을 적용하지 않고 프레임마다 모든 파라미터 정보를 전송하게 되면 4.25 KHz의 비트 할당이 필요하다[2]. SC-VBR scheme을 적용하면 표 1에서와 같이 파라미터 정보를 보내지 않음으로써 비트 할당을 줄일 수 있다. 50문장에 대해서 실험한 결과 SC-VBR scheme를 적용했을 때 적용하지 않았을 때보다 22.9%의 비트 감소율을 보였다.

표 3. 기존의 방법과 제안된 방법의 PESQ값

	PESQ 값
SC-VBR scheme을 적용하지 않은 과형보간 부호화기의 음성 결과	2.588
SC-VBR scheme을 적용한 과형보간 부호화기의 음성결과	2.437
SC-VBR scheme을 적용한 과형보간 부호화기에 Repetition method를 통해 얻은 음성결과(REW 포함)	2.296
SC-VBR scheme을 적용한 과형보간 부호화기에 Repetition method를 통해 얻은 음성결과(REW 제외)	2.531
SC-VBR scheme을 적용한 과형보간 부호화기에 Extrapolation method를 통해 얻은 음성결과(REW 제외)	2.552

표 3은 기존의 방법과 제안된 방법으로 얻은 결과를 객관적 성능평가인 PESQ값을 나타낸 것이다. SC-VBR scheme을 적용하지 않은 과형보간 부호화기의 음성결과는 모든 파라미터의 정보를 전송하기 때문에 가장 좋은 결과를 보임을 알 수 있다. 표 3의 결과와 같이 SC-VBR scheme을 적용한 기존의 방법은 적용하지 않은 기존의 방법에 비해 PESQ가 감소되는 것을 볼 수 있다. REW의 복원을 제외한 실험에서 두가지 제안한 방법을 통한 실험 결과는 기존의 SC-VBR scheme을 적용한 결과보다 음질평가 결과가 나아짐을 확인할 수가 있다. 전송되지 않은 파라미터를 예측하여 실제 값과 유사하게 생성시킴으로써 음질을 향상시키는 결과를 얻을 수가 있다. Repetition method를 통해 실험한 결과 중 REW도 복원한 실험의 경우 기존의 결과보다 안 좋은 결과를 보이는 것은 노이즈 성분과 같은 특성을 지닌 REW를 반복하여 생성시킴으로써 강제로 주기적인 특성을 지니게 되므로 음질 열화가 발생하는 것이다. 따라서 실험 결과와 같이 REW의 복원은 제외한 두 가지 제안한 방법을 사용하였을 경우 음질의 향상을 가져올 수 있음을 확인할 수 있다.

## V. 결론 및 향후 연구 계획

본 논문에서는 SC-VBR scheme을 적용한 과형 보간 부호화기의 비트 감소율과 음질 열화 정도를 확인 및 분석하였고, 음질 향상을 위해 repetition method와 extrapolation method를 제안하였다. 그리고 제안한 방법을 통해 SC-VBR scheme을 적용한 과형 보간 부호화기의 비트 감소율을 유지하면서 음질을 향상시키는

결과를 확인하였다. 두 가지 제안한 방법 모두 기존의 방법의 부호화기의 음질을 향상 시키는 데에 적용할 수 있고, 두 가지 방법 중에선 extrapolation method가 근소하게 좀 더 나은 결과를 보임을 확인할 수가 있다. 제안된 방법에선 REW의 특성상 복원하기 어렵다. 따라서 extrapolation method를 선형적인 방법이 아닌 spline을 이용한 방법을 적용하여 발산의 가능성을 막아 REW 정보도 함께 복원하는 방법에 대해 향후 연구할 예정이다.

## 참고문헌

- [1] F. Plante and B.M.G. Cheetham, D. Marstom, P.A. Barrett, "Source controlled variable bit-rate speech coder based on waveform interpolation," *Proc. ICSLP*, pp.848-851, 1998
- [2] Eddie L. T. Choy, "Waveform Interpolation Speech Coder at 4kb/s," MS. Thesis, McGill University
- [3] 주기현, 김남수, "4.0 kb/s 과형 보간 부호화기," *신호처리합동학술대회*, Vol.12, NO.1, pp.323-326, 1999