

SMV 와 G.723.1 음성부호화기를 위한 파라미터 직접 변환 방식의 상호부호화 알고리즘

서성호, 장달원, 이선일, 유창동

한국과학기술원 전자전산학과

Transcoding Algorithm for SMV and G.723.1 Vocoders via Direct Parameter Transformation

Seongho Seo, Dalwon Jang, Sunil Lee, Chang D. Yoo

Dept. of Electrical Engineering and Computer Science, Korea Advanced Institute of Science and Technology,
E-mail : wing@kaist.ac.kr

Abstract

In this paper, a transcoding algorithm for the Selectable Mode Vocoder (SMV) and the G.723.1 speech coder via a direct parameter transformation is proposed. In contrast to the conventional tandem transcoding algorithm, the proposed algorithm converts the parameters of one coder to the other without going through the decoding and encoding process. The proposed algorithm is composed of four parts: the parameter decoding, line spectral pair (LSP) conversion, pitch period conversion, excitation conversion and rate selection. The evaluation results show that the proposed algorithm achieves equivalent speech quality to that of tandem transcoding with reduced computational complexity and delay.

I. 서론

오늘날에는 다양한 유/무선 통신망이 존재하며, 각 통신망은 고유의 특성에 적합한 음성 부호화 표준을 채택하여 사용하고 있다. 음성 통신망 간의 효율적인 연동을 위해서는 한 음성 부호화기로 부호화된 비트열들을 다른 음성 부호화기에 맞는 비트열로 변환하는 상호

부호화가 필요하다. 일반적으로 가장 간단한 방법은 복호화를 통해 생성된 음성을 다시 부호화하는 tandem 방식의 상호부호화이며, 이것은 중복된 부호화 과정에 의해 음질 저하와 계산량 및 시간 지연 증가라는 단점을 가진다.

본 논문에서는 G.723.1^[1] 과 Selectable Mode Vocoder (SMV)^[2] 음성부호화기 간의 상호부호화 알고리즘으로 tandem 방식의 문제점을 개선할 수 있는 파라미터 직접 변환 방식의 상호부호화 알고리즘을 제안한다. 파라미터 직접 변환 방식의 상호부호화에서는 양 음성부호화기에서 음성을 부호화하는데 공통적으로 사용되는 파라미터를 직접 변환한다.

II. G.723.1 과 SMV 음성부호화기

G.723.1 음성부호화기는 5.3kbps 와 6.3kbps 의 두 가지 비트율에서 동작한다. 음성부호화기의 입력은 8kHz 로 표본화된 16bit 선형(Uniform) PCM 형식이 사용된다. 음성 프레임(Frame)의 길이는 240 샘플에 해당하는 30ms 이며, 한 프레임은 60 샘플에 해당하는 4 개의 부프레임(Subframe)으로 나뉘어진다. 매 부프레임마다 10 차의 선형 예측 계수(Linear Prediction Coefficients)가 얻어지며, 이때, 7.5ms 의 예견 구간(Look-ahead)이 필요하다. 마지막 부프레임의 선형 예측 계수만이 LSP 로 변환된 후 양자화하여 전송된다. 그런 다음, 개회로 피

치 지연 (Open-loop Pitch)이 지각 가중된 음성 신호로부터 계산되고, 그 주변에서 적응 코드북이 5 차 피치 예측기(5th order pitch predictor)를 이용하여 계산된다. 고정 코드북 검색은, 고 비트율에서는 MP-MLQ (Multi-pulse Maximum Likelihood Quantization) 방식, 저 비트율에서는 ACELP(Algebraic Code Excite -d Linear Prediction) 방식을 이용한다. 또한, G.723.1 음성 부호화기는 평균 비트율을 줄이기 위한 방법으로 묵음 압축 기법(silence compression scheme)^[2]을 사용한다. 이 기법의 동작은 외부에서 입력되는 옵션(option)을 통해 결정된다.

SMV 음성부호화기는 네가지 비트율 8.55(Rate 1), 4.0(Rate 1/2), 2.0(Rate 1/4) 그리고 0.8(Rate 1/8) kbit/s로 동작한다. 또한, SMV 음성부호화기는 통신망의 상황에 따라 결정되는 4 가지 모드(Mode)에서 동작하여, 평균 데이터율과 음질의 균형을 적절히 조절할 수 있다. 8kHz로 표본화된 160 샘플에 해당하는 20ms의 음성 프레임이 네가지 비트율 중 한가지로 부호화된다. 매 프레임마다 10 차의 선형 예측 분석이 수행되며, 이를 위해 10ms의 예견 구간이 필요하다. 각 음성 프레임은 묵음(silence), 잡음(noise like unvoiced), 무성음(unvoiced), 변화 (onset), 일정하지 않은 유성음(non-stationary voiced), 일정한 유성음(stationary voiced)의 여섯가지 프레임 클래스(Frame class) 중 하나로 분류되고, 외부에서 입력된 모드와 분류된 프레임 클래스에 따라서 비트율 결정 알고리즘(Rate Decision Algorithm, RDA)을 이용하여 비트율을 결정한다. 또한, Rate 1 이나 Rate 1/2 로 결정된 프레임의 경우에는 타입(Type) 0 와 타입 1 으로 다시 나누어진다. 타입 1 프레임은 일정한 유성음에 해당하는 음성을 나타내고, 타입 0 는 그외의 모든 프레임을 나타낸다. Rate 1 과 Rate 1/2 의 여기 신호 계산은 extended CELP(ex-CELP)^[4] 방식으로 이루어지고, Rate 1/4 과 Rate 1/8 의 여기 신호는 난수를 발생시켜 계산한다.

III. SMV 에서 G.723.1 로의 상호부호화

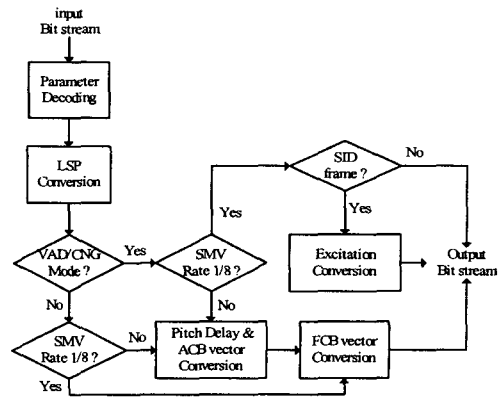


그림 1. SMV→G.723.1 상호부호화의 블록도.

3.1 전체적인 구조

그림 1 은 SMV 에서 G.723.1 로의 상호부호화에 대한 제안된 알고리즘의 구조도이다. 먼저, SMV 음성부호화기에 의해 부호화 된 입력 비트열(Input Bit stream)로부터 파라미터 복호화를 통해 LSP, 피치 지연, SMV 의 부호화 비트율 등의 파라미터들을 계산한다. 파라미터 복호화가 끝나고 나면 LSP 변환이 이루어진다. SMV 에서 Rate 1/8 로 결정된 음성 프레임은 배경 잡음이나 묵음에 해당하므로 경우에는 고정 코드북 검색만을 수행하고, 적응 코드북은 검색하지 않는다. Rate 1/8 이 아닌 경우에는 적응 코드북 변환과 고정 코드북 변환을 모두 수행한다. 또, G.723.1 이 외부 옵션에 의해서, 묵음 압축 기법을 사용하는 경우에는 SMV 의 비트율이 Rate 1/8 이라고 판별되면, 제안된 알고리즘은 묵음 압축 기법을 이용하여 부호화를 수행한다. 다음 절 들에서 그 자세한 상호부호화 과정 및 알고리즘이 설명된다.

3.2 LSP 변환

제안된 상호부호화 알고리즘에서, 간단한 선형 보간(Linear Interpolation)을 통해 LSP 로 변환한다. 그림 2 는 SMV 에서 G.723.1 로의 LSP 변환 과정을 나타낸 것이다. 각 음성 부호화기는 선형 예측 분석을 위해 서로 다른 길이와 모양을 가지는 윈도우를 사용하므로, 이

차이를 고려하여 다음과 같은 선형 보간을 위한 식을 구하였다.

$$LSP_K^G = \frac{2}{3}LSP_L^{SMV} + \frac{1}{3}LSP_{L+1}^{SMV} \quad (1)$$

$$LSP_{K+1}^G = LSP_{L+2}^{SMV} \quad (2)$$

여기서 LSP_L^{SMV} 과 LSP_K^G 는 각각 SMV 음성 부호화기의 L 번째 프레임의 LSP 와 G.723.1 음성부호화기의 K 번째 프레임의 LSP 이다.

3.3 피치 지연 및 적응 코드북 변환

제안된 알고리즘에서는 G.723.1 의 매 두 부프레임마다 개회로 피치 지연을 과거값을 이용하여 예측한다. 예측된 피치 지연은 파라미터 복호화 과정을 통해 복호화된 SMV 의 폐회로 피치 지연과 비교되어, 만약 두 피치 지연의 차가 문턱값보다 작다면, SMV 의 폐회로 피치 지연이 G.723.1 의 개회로 피치 지연으로 선택되고, 피치 지연의 차가 문턱값보다 크다면, 좀 더 정확한 피치 지연을 얻기 위해 G.723.1 의 개회로 피치 지연 검색을 수행한다. 이렇게 얻어진 개회로 피치 지연 주변에서 G.723.1 의 알고리즘을 이용하여 폐회로 피치 지연과 적응 코드북 이득이 얻어진다.

IV. G.723.1 에서 SMV 로의 상호부호화

4.1 전체적인 구조

그림 3 은 G.723.1 에서 SMV 음성부호화기로의 상호부호화 알고리즘의 구조를 나타낸다. 복호화되는 파라미터들은 LSP, 피치 지연, 적응 코드북 이득 및 고정 코드북 이득이다. 그리고 비트율 결정을 위한 음성 파라미터도 계산한다. LSP 변환과 피치 지연 변환을 거친 후, 제안된 상호부호화 알고리즘은 간략화된 비트율 결정 알고리즘을 이용하여 외부에서 입력된 모드 정보에 따라 비트율과 타입을 결정한다.

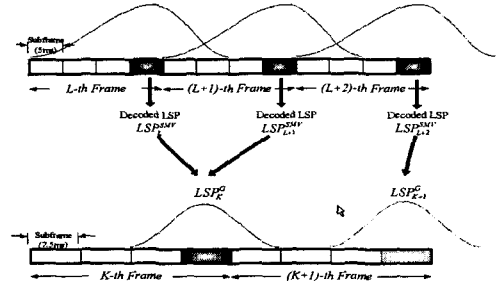


그림 2. SMV 에서 G.723.1 로의 LSP 변환

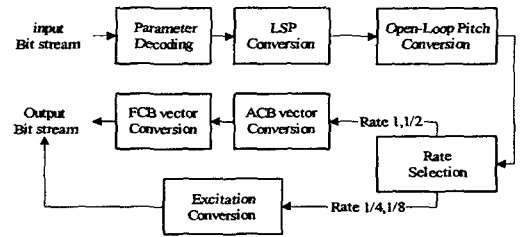


그림 3. G.723.1 → SMV 상호부호화의 블록도.

마지막으로, 앞서 결정된 비트율과 타입에 따라서 합성 필터를 위한 여기신호를 구해낸다. 여기신호의 계산은 SMV 에서 구현된 알고리즘을 이용하였다.

G.723.1 에서 SMV 로의 LSP 변환도 다음과 같은 선형 보간을 통해 이루어진다.

$$LSP_L^{SMV} = \frac{7}{24}LSP_{K-1}^G + \frac{17}{24}LSP_K^G \quad (3)$$

$$LSP_{L+1}^{SMV} = \frac{2}{3}LSP_K^G + \frac{1}{3}LSP_{K+1}^G \quad (4)$$

$$LSP_{L+2}^{SMV} = LSP_{K+1}^G \quad (5)$$

피치 지연의 경우에는, SMV 에서 G.723.1 로의 상호부호화에서 제안된 피치 지연 변환 알고리즘이 적용되지 않았고, 두 음성부호화기의 개회로 피치 지연 검색 영역을 일치시키기 위한 선형 보간만이 이용되었다.

4.2 비트율 결정

파라미터 디코딩 과정에서 얻어진 G.723.1 의 파라미터들을 이용하여 비트율을 결정하기 위하여 그림 4 에 나타난 간략화된 비트율 결정 알고리즘을 제안하였다. 제안된 알고리즘에서는 6 가지 클래스를

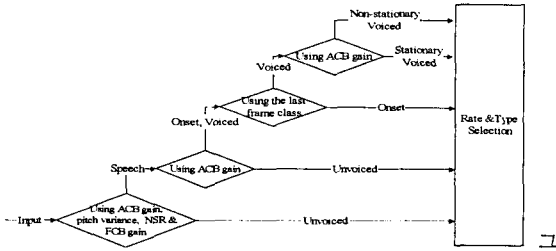


그림 4. 간략화된 비트율 결정 알고리즘의 블록도.

가지는 SMV 와는 달리, 클래스를 5 가지로 줄여 분류하고, 클래스를 이용한 비트결정은 SMV 와 유사한 방법으로 수행된다.

V. 성능평가

본 장에서는 제안된 파라미터 직접 변환 방식의 상호부호화 알고리즘의 성능을 계산량, 객관적 음질 평가, 지연 시간의 지표를 이용하여 평가한다.

계산량의 평가를 위해서 Weighted Million Operations Per Second (WMOPS)를 이용하였다. 표 1의 결과를 살펴 보면, 제안된 파라미터 직접 변환 방식의 계산량이 Tandem 방식에 비하여 최저 20%에서 최고 35%까지 줄어드는 것을 확인할 수 있다.

객관적 음질 평가에는 ITU P.862 Perceptual Evaluation of Speech Quality (PESQ)를 이용하였다. 표 2는 제안된 파라미터 직접 변환 방식의 상호부호화 알고리즘이 Tandem 방식과 전체적으로 동등한 음질을 구현함을 보여주고 있다.

SMV 에서 G.723.1 로의 상호부호화의 경우, 선형 예측 분석을 위한 예견 구간이 필요하지 않으므로, 알고리즘 지연이 7.5ms 만큼 줄어들게 된다. 반면에, G.723.1 에서 SMV 로의 상호부호화의 경우에는 선형 예측 분석을 위한 예견 구간이 필요하지는 않지만, 이 예견 구간에 해당하는 음성 구간이 비트율 결정 같은 다른 처리 과정에 필요하기 때문에 알고리즘 지연이 예견 구간만큼 줄어들지는 않는다. 양방향의 상호부호화에서, 불필요한 처리과정은 생략되므로 처리 지연 시간 또한 줄어들게 된다.

VI. 결론

본 논문에서는 G.723.1 음성부호화기와 Selectable Mode Vocoder(SMV) 음성부호화기를 위한 파라미터 직접 변환 방식의 상호부호화 알고리즘을 제안하였다. 앞 장에서 제시된 성능 평가 결과는 제안된 알고리즘이 Tandem 방식에 비하여 적은 계산량과 짧은 지연 시간으로도 동등한 수준의 음질을 구현함을 보여주고 있다.

표 1. 계산량 비교

화자	모드	SMV→G.723.1		G.723.1→SMV	
		Tandem	Proposed	Tandem	Proposed
남성	0	19.95	12.59	29.46	23.30
	1	20.04	12.42	27.69	21.90
	2	20.24	12.32	27.44	21.53
	3	20.26	12.30	27.52	21.52
여성	0	20.01	13.95	30.22	24.97
	1	20.20	13.80	29.47	22.81
	2	20.30	13.79	28.47	22.65
	3	20.30	13.77	29.24	22.67

표 2. 객관적 음질 평가 결과

화자	모드	SMV→G.723.1		G.723.1→SMV	
		Tandem	Proposed	Tandem	Proposed
남성	0	3.237	3.239	3.303	3.244
	1	3.180	3.174	3.204	3.166
	2	3.098	3.135	3.112	3.050
	3	3.026	3.091	3.095	2.982
여성	0	3.066	3.023	3.081	3.102
	1	2.996	2.997	2.959	3.069
	2	2.942	2.960	2.904	2.863
	3	2.915	2.973	2.957	2.817

References

- [1] ITU-T Rec. G.723.1, "Dual-rate Speech Coder For Multimedia Communications Transmitting at 5.3 and 6.3 kbit/s", 1996
- [2] ITU-T Rec. G.723.1 - Annex A, "Silence Compression Scheme", 1996
- [3] 3GPP2 Spec., "Selectable Mode Vocoder Service Option for Wideband Spread Spectrum Communication Systems", 3GPP2-C.S0030-0 v2.0, Dec. 2001.
- [4] Yang Gao, A. Benyassine, J. Thyssen, Huan-yu Su, E. Shlomot, "EX-CELP : A Speech Coding Paradig", In Proc. ICASSP 2001, vol. 2, pp. 689-692, 2001.