

論文2003-40SP-6-7

SMV와 G.723.1 음성부호화기를 위한 파라미터 직접 변환 방식의 상호부호화 알고리즘

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

徐晟豪*, 張達源*, 李善一*, 柳昌東*

(Seongho Seo, Dalwon Jang, Sunil Lee, and Chang D. Yoo)

요약

본 논문에서는 SMV와 G.723.1 음성부호화기를 위한 파라미터 직접 변환 방식의 상호부호화 알고리즘을 제안한다. 상호부호화를 위하여 부가적인 복호화, 부호화 과정을 거쳐야하는 Tandem 방식과 달리 제안된 방식에서는 양 음성부호화기가 음성을 부호화하는데 공통적으로 사용되는 파라미터들을 직접 변환한다. 제안된 알고리즘은 파라미터 복호화, LSP 변환, 피치 지연 변환, 여기신호 변환 그리고 비트율 결정으로 이루어진다. 제안된 알고리즘을 다양한 방법으로 평가해 본 결과 계산량과 지연시간을 줄이면서 tandem 방식과 동등한 수준의 음질을 구현함을 확인할 수 있었다.

Abstract

In this paper, a transcoding algorithm for the Selectable Mode Vocoder (SMV) and the G.723.1 speech coder via 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.

Keyword : transcoding, SMV, G.723.1, tandem, direct parameter transformation

I. 서론

오늘날에는 다양한 유/무선 통신망이 존재하며, 각 통신망은 고유의 특성에 적합한 음성 부호화 표준을 채택하여 사용하고 있다. 최근들어, 인터넷이 급격히 발

달함에 따라 IP기반의 통신망을 통한 Voice over Internet Protocol(VoIP) 서비스가 확산되고, 따라서, 여러 음성 통신망간의 효율적인 연동이 요구되고 있다. 그러나, 각 통신망은 서로 다른 음성 부호화 표준을 사용하므로 서로 호환되지 않는다. 따라서, 음성 통신망간의 효율적인 연동을 위해서는 한 음성 부호화기로 부호화된 비트열들을 다른 음성 부호화기에 맞는 비트열로 변환하는 상호부호화가 필요하다.

본 논문에서는 VoIP 망에서 널리 사용되는 G.723.1^[1]

* 正會員, 韓國科學技術院 電子電算學科

(Dept. of Electrical Engineering and Computer Science, Korea Advanced Institute of Science and Technology)

接受日:2003年6月18日, 수정완료일:2003年11月22日

음성부호화기와 CDMA2000 시스템을 위한 새로운 음성 표준인 Selectable Mode Vocoder (SMV)^[3] 음성부호화기 간의 상호부호화 알고리즘을 제안한다. 일반적으로, 서로 다른 음성 부호화기간의 호환성 문제를 해결하는 가장 간단한 방법은 한 음성 부호화기의 복호화 과정을 통해 비트열로부터 음성을 생성하고, 이 음성을 다시 다른 음성 부호화기를 이용하여 비트열로 바꾸는 것이다. 이 방법은 tandem 방식의 상호부호화라고 불린다. 그러나, tandem 방식의 상호부호화기는 중복된 부호화 과정에 의해 음질 저하와 계산량 및 시간 지연 증가라는 단점을 가진다. 따라서, 본 논문에서는 이러한 tandem 방식의 문제점을 개선할 수 있는 파라미터 직접 변환 방식의 상호부호화 알고리즘을 제안한다. 파라미터 직접 변환 방식의 상호부호화는 복호화를 통해 생성된 음성을 다시 부호화하는 Tandem 방식과 달리, 양 음성부호화기에서 음성을 부호화하는데 공통적으로 사용되는 파라미터를 직접 변환한다. 기존의 파라미터 직접 변환 방식에 기반한 상호부호화를 다룬 논문들로는 [5~7] 등이 있다.

본 논문의 구성은 다음과 같다. II 장에서는 상호부호화의대상이 되는 SMV와 G.723.1 음성부호화기에 대하여 간략히 알아본다. 그리고 III장과 IV장에서는 SMV에서 G.723.1로의, 그리고 G.723.1에서 SMV로의 상호부호화 알고리즘을 자세히 설명한다. V장에서는 제안된 파라미터 직접 변환 방식의 상호부호화 알고리즘을 다양한 방법으로 평가한 결과를 제시하며, 마지막으로 VI장에서 결론으로써 논문을 끝맺는다.

II. G.723.1과 SMV 음성부호화기

1. G.723.1 음성부호화기

G.723.1 음성부호화기는 5.3kbps 와 6.3kbps의 두 가지 비트율에서 동작한다. 음성부호화기의 입력은 8kHz로 표본화된 16bit 선형(Uniform) PCM 형식이 사용된다. 음성 프레임(Frame)의 길이는 240 샘플에 해당하는 30ms이며, 한 프레임은 60 샘플에 해당하는 4개의 부프레임(Subframe)으로 나뉘어진다. 매 부프레임마다 10차의 선형 예측 계수(Linear Prediction Coefficients)가 얻어지며, 선형 예측 분석 (Linear Prediction Analysis)을 위해서는 7.5ms 의 예견 구간 (Look-ahead)이 필요하다. 마지막 부프레임에서 계산된 선형

예측 계수만이 LSP로 변환된 후, 예측 분할 벡터 양자화 (Predictive Split Vector Quantizer, PSVQ)를 통해 양자화하여 전송되고, 양자화되지 않은 선형 예측 계수들은 각각 가중된 음성 신호를 얻는데 이용된다. 이렇게 얻어진 각각 가중된 음성 신호로부터, 두 부프레임마다 개회로 피치 분석(Open-loop Pitch Analysis)이 이루어진다. 그런 다음, 매 부프레임마다 적응 코드북 (Adaptive Codebook)과 고정 코드북 (Fixed Codebook)이 검색된다. 적응 코드북 검색에서는, 폐회로(Closed-loop) 피치 지연과 적응 코드북 이득이 개회로 피치 지연과 5차 피치 예측기(5th order pitch predictor)를 이용하여 계산된다. 마지막으로, 여기 신호의 비주기성 성분이 고정 코드북 벡터를 이용하여 근사화된다. 고정 코드북 검색을 제외한 모든 처리과정이 두 개의 비트율에 대해 동일하게 이루어진다. 고정 코드북 검색은, 고 비트율에서 동작하는 경우 MP-MLQ(Multi-pulse Maximum Likelihood Quantization) 방식, 저 비트율에서 동작하는 경우 ACELP(Algebraic Code-Excited Linear Prediction) 방식이 이용된다.

G.723.1 음성 부호화기는 평균 비트율을 줄이기 위한 방법으로 묵음 압축 기법(silence compression scheme)^[2]을 사용하며, 이 기법의 동작은 외부에서 입력되는 옵션(option)을 통해 결정된다. 이 기법은 Voice Activity Detection (VAD) 알고리즘과 Comfort Noise Generation(CNG) 알고리즘으로 이루어져 있다. G.723.1 음성부호화기는 VAD를 이용하여 입력 음성에서 묵음 구간을 찾아내고, 판별된 묵음 구간에 대해서는 CNG 알고리즘을 이용하여 5.3kbps보다 더 낮은 비트율로 음성 프레임을 부호화한다.

2. SMV 음성부호화기

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 4가지 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. 이때, 각 음성 프레임에 대한 비트율은 비트율 결정 알고리즘(Rate Decision Algorithm, RDA)에 의해 결정되며, 비트율 결정 알고리즘은 각 음성 프레임에 대한 프레임 클래스 분류 (Frame Classification)에 기

반을 두고 있으며, 통신망에 의해 결정된 모드에 따라 다르게 동작한다.

SMV 음성부호화기는 다음과 같이 음성을 부호화한다. 입력 음성은 먼저 묵음 증강(Silence enhancement), 고대역 필터 통과, 잡음 억제(Noise Suppression), 그리고 적응 틸트 보상 (Adaptive tilt compensation)으로 이루어진 전처리 과정을 통과한다. 그리고, 음성 프레임 단위로 선형 예측 분석, 개회로 피치 검색, 음성 변형 (Signal modification)과 프레임 클래스 분류 과정을 거친다. SMV 음성부호화기의 선형 예측 분석은 각 프레임마다 3번씩 수행되며, 각각의 선형 예측 분석에서는 모두 다른 형태의 윈도우(Window)가 사용된다. 3가지 선형 예측 분석을 통해 얻어진 선형 예측 계수들 중, 음성 프레임의 마지막 1/4부분에 중심을 둔 윈도우를 이용하여 계산된 선형 예측 계수만이 LSP로 변환된 다음, 양자화하여 전송된다. 나머지 2가지 선형 예측 계수들은 기중 음성 신호의 생성과 프레임 클래스 분류 등의 처리 과정에 사용된다. 선형 예측 분석을 위해 필요한 예전 구간은 10ms이다. 개회로 피치 검색은 매 프레임마다 두번씩 수행된다. 반 프레임의 기중 음성 신호에 대해 개회로 피치 검색을 통해 개회로 피치 지연을 찾아내므로, 매 프레임마다 두 개의 개회로 피치 지연이 얻어진다. 음성 변형에서는, 개회로 피치 지연으로부터 계산된 피치 컨투어(Pitch contour)에 맞도록 기중된 음성 신호를 변형시킨다. 클래스 분류에서는, 현재 음성 프레임이 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, 음성 프레임의 부호화를 위해서 외부에서 입력된 모드와 분류된 프레임 클래스에 따라서 4 가지 비트율 중 한가지를 선택한다. 또한, Rate 1이나 Rate 1/2로 결정된 프레임의 경우에는 타입(Type) 0와 타입1으로 다시 나누어진다. 타입 1 프레임은 일정한 유성음에 해당하는 음성을 나타내고, 타입 0는 그 외의 모든 프레임을 나타낸다. 타입 1 프레임은 타입0 프레임보다 고정 코드북에 더 많은 비트를 할당하고, 적응 코드북에는 좀 더 적은 비트를 할당한다. 세번째로, SMV 음성부호화기는 결정된 비트율에 따라 서로 다른 방식으로 합성 필터를 위한 여기 신호를 계산한다. Rate 1/4에 대해서, 여기신호는 난수 발생기 (random number generator)에 의해 생성된 후, 2ms의 부프레임

마다 이득을 곱하고 주파수 변형 필터 (frequency-shaping filter)를 통과시켜 얻어진다. Rate 1/8에 대해서는, 난수 발생기에 의해 얻어진 신호에 하나의 이득 값을 곱하여 여기신호를 생성한다. Rate 1과 Rate 1/2의 여기 신호 계산은 extended CELP(ex-CELP)^[4] 방식으로 이루어진다.

III. SMV에서 G.723.1 음성부호화기로의 상호부호화

1. 전체적인 구조

II장에서 살펴본 것처럼, SMV 음성부호화기와 G.723.1 음성부호화기는 서로 다른 음성 프레임 길이를 가진다. 따라서, 두 음성 부호화기의 음성 프레임간의 직접적인 비교는 힘들기 때문에, 같은 길이의 음성을 비교할 수 있도록 SMV 음성부호화기의 음성 프레임 3 개와 G.723.1 음성부호화기의 음성 프레임 2개를 비교하여 상호부호화를 수행한다.

<그림 1>은 SMV에서 G.723.1 음성부호화기로의 상호부호화에 대한 제안된 알고리즘의 구조도를 나타낸 것이다. <그림 1>에서 볼 수 있듯이, 제안된 SMV에서 G.723.1 음성부호화기로의 상호부호화 알고리즘은 파라미터 복호화, LSP 변환, 피치 지연 변환 및 여기 신호 변환(적응 코드북 및 고정 코드북의 변환)으로 이루어지며, 입력은 SMV 음성부호화기에 의해 부호화 된 비

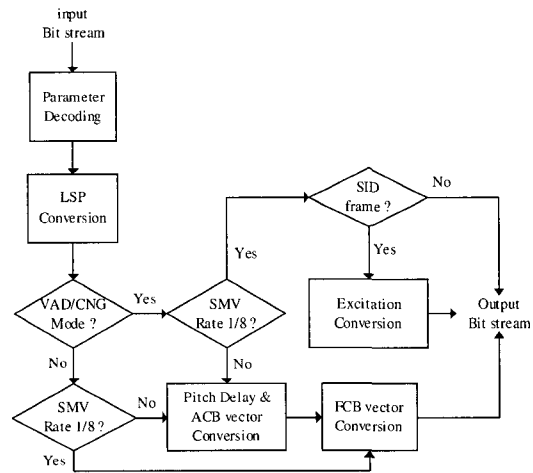


그림 1. 제안된 파라미터 직접 변환 방식의 SMV에서 G.723.1로의 상호부호화 알고리즘의 블록도.

Fig. 1. Block diagram of the proposed transcoding algorithm from SMV to G.723.1.

트열(Input Bit stream)이다. 입력 비트열은 먼저 파라미터 복호화(Parameter Decoding) 과정을 거친다.

파라미터 복호화 과정에서는 상호부호화 과정에서 변환되어야 하는 파라미터들을 입력 비트열로부터 복호화한다. 또한 G.723.1 음성부호화기에 맞는 여기 신호의 생성을 위한 목적 신호(Target signal)도 계산한다. 이 부분은 SMV 음성부호화기의 복호화기를 부분적으로 이용하여 구현하였다. 복호화되는 파라미터들은 LSP, 피치 지연, SMV의 부호화 비트율이다. 파라미터 복호화가 끝나고 나면 LSP 변환을 통해, SMV 부호화기를 통해 부호화된 LSP를 G.723.1의 형태에 맞도록 변환한다.

SMV 음성부호화기는 4가지 비트율로 동작하며, 낮은 비트율로 동작하는 경우에는 다른 비트율에 비해 비교적 간단한 방법으로 상호부호화가 가능하다. 따라서 상호부호화 알고리즘은 SMV 음성부호화기의 비트율에 따라 다르게 동작한다. 파라미터 복호화 과정과 LSP 변환 과정은 SMV의 비트율에 상관없이 모든 비트율에 대해 동일하게 수행되며, 선형 합성 필터를 위한 여기 신호를 계산하는 과정이 비트율에 따라 달라진다. G.723.1 음성부호화기를 위한 여기 신호는 파라미터 복호화 과정에서 얻어진 SMV 음성부호화기의 파라미터와 목적 신호를 이용하여 얻어진다. SMV 음성부호화기에서, Rate 1/8은 배경 잡음이나 묵음으로 분류된 음성 신호에 대해 결정된다. 적응 코드북 벡터는 음성신호의 주기적 특성을 표현하는 것이므로, SMV 음성부호화기에서 Rate 1/8로 결정한 음성 프레임에 해당하는 G.723.1의 음성프레임의 경우에는 적응 코드북 검색이 적용되지 않아도 된다. 따라서, 이 경우에는, 제안된 알고리즘은 선형 예측 합성 필터를 위한 여기신호를 생성하기 위해서 고정 코드북 검색만을 수행하고 적응 코드북은 검색하지 않는다. 복호화된 SMV의 비트율이 Rate 1/8이 아닌 경우에는 적응 코드북 변환과 고정 코드북 변환을 모두 수행한다. 고정 코드북 변환은 파라미터 복호화 과정을 통해 얻어진 목적 신호로부터 G.723.1의 고정코드북 검색 방법을 이용하여 수행된다.

또한, G.723.1 음성부호화기는 외부 옵션에 의해서, 묵음 압축 기법을 사용할 수 있기 때문에, 제안된 알고리즘에서는 이러한 외부 입력이 주어진 경우에는 CNG를 포함한 조금 다른 처리과정을 수행한다. 만약 외부 옵션에 의해서 묵음 압축 기법의 동작이 결정되고, 파

라미터 복호화 과정에서 얻어진 SMV의 비트율이 Rate 1/8이라고 판별되면, 제안된 알고리즘은 CNG를 이용하여 부호화를 수행한다. CNG 알고리즘에서는 음성 프레임이 SID(Silence Insertion Descriptor) 프레임인지 아닌지를 판별하고, 만약 SID 프레임이면 5.3kbps의 경우보다 더 낮은 비트율로 여기신호를 부호화하여 전송하고, SID 프레임이 아닌 경우에는 전송하지 않는다^[2]. SMV의 비트율이 Rate 1/8인 경우에는 묵음 압축 기법 옵션이 지정되지 않는 경우와 마찬가지로 적응 코드북 변환과 고정 코드북 변환을 모두 수행한다.

이렇게 Rate 1/8의 경우에 몇 가지 처리과정을 줄임으로써, 전체적인 계산량을 줄일 수 있었다. 다음 절들에서 그 자세한 상호부호화 과정 및 알고리즘이 설명된다.

2. LSP 변환

제안된 상호부호화 알고리즘에서, SMV 음성부호화기의 LSP는 간단한 선형 보간(Linear Interpolation)을 통해 G.723.1 음성부호화기의 LSP로 변환된다. SMV 음성부호화기는 프레임 당 3번의 선형 예측 분석을 하고, 두 음성부호화기 모두 부프레임마다 음성 합성을 위해 LSP를 계산한다. 그러나, 두 음성부호화기는 프레임 당 하나의 LSP 세트만을 전송한다. 따라서 전송에 이용되는 LSP 세트와 그에 해당하는 선형 예측 분석 윈도우만을 고려한다. <그림 2>는 SMV에서 G.723.1 음성 부호화기로의 상호부호화에서 LSP가 변환되는 과정을 나타내며, 전송되는 LSP 세트에 해당하는 선형 분석 윈도우만이 표시되어 있다. 각 음성 부호화기는 서로 다른 길이의 음성 프레임을 가질 뿐 아니라, 선형

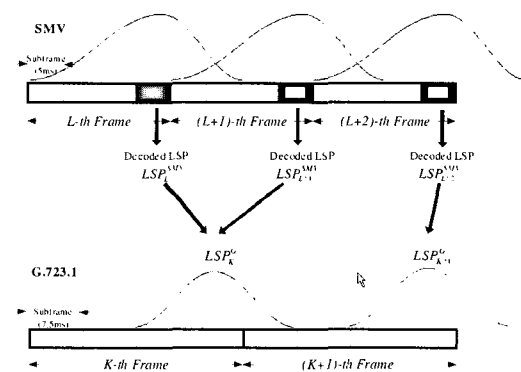


그림 2. SMV에서 G.723.1로의 LSP 변환

Fig. 2. LSP conversion from SMV to G.723.1 via linear interpolation.

예측 분석을 위해 서로 다른 길이와 모양을 가지는 윈도우를 사용하므로 이 차이를 고려하여 선형 보간을 위한 계수들을 구하였다. 선형 보간의 식은 다음과 같다.

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

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

여기서 LSP_L^{SMV} 과 LSP_K^G 는 각각 SMV 음성 부호화기의 L 번째 프레임의 LSP와 G.723.1 음성부호화기의 K 번째 프레임의 LSP이다. 식 (1)을 살펴보면, <그림 2>에서 처럼 SMV 음성부호화기의 L 번째와 $L+1$ 번째 프레임의 LSP를 이용하여 G.723.1 음성부호화기의 K 번째 프레임의 LSP를 얻어낼 수 있다. 또한, 식 (2)에서는 SMV 음성 부호화기의 $L+2$ 번째 프레임의 LSP가 바로 G.723.1 음성부호화기의 $K+1$ 번째 프레임의 LSP로 사용됨을 알 수 있다. 제안된 알고리즘을 사용하면 추가적인 선형 예측 분석 과정이 필요하지 않게 되므로 예견 구간에 의한 알고리즘 지연이 없어져 전체 지연 시간이 짧아지고, 계산량 또한 줄어드는 장점이 있다.

3. 피치 지연 및 적응 코드북 이득 변환

<그림 3>은 SMV에서 G.723.1로의 상호부호화에서 피치 지연 변환을 위해 제안된 알고리즘의 블록도이다. G.723.1 음성부호화기는 매 두 부프레임마다 개회로 피치 지연을 지각 가중된 음성 신호를 이용하여 검색하는데 반해, 제안된 알고리즘에서는 G.723.1의 매 두 부프레임마다 개회로 피치 지연을 과거값을 이용하여 예측한다. 예측된 피치 지연은 파라미터 복호화 과정을 통해 복호화된 SMV의 폐회로 피치 지연과 비교된다. 만약 두 피치 지연의 차가 문턱값보다 작다면, SMV의 폐회로 피치 지연이 G.723.1의 개회로 피치 지연으로 선택되어 개회로 피치 지연 검색과정을 생략한다. 그렇지 않고 피치 지연의 차가 문턱값보다 크다면, 좀 더 정확한 피치 지연을 얻기 위해 G.723.1의 개회로 피치 지연 검색을 수행한다. 이렇게 얻어진 개회로 피치 지연 주변에서 폐회로 피치 지연과 적응 코드북 이득이 얻어진다. 이 과정은 G.723.1 음성부호화기의 처리 과정과 동일하게 수행된다.

제안된 선형 예측을 이용한 피치 지연 변환 알고리

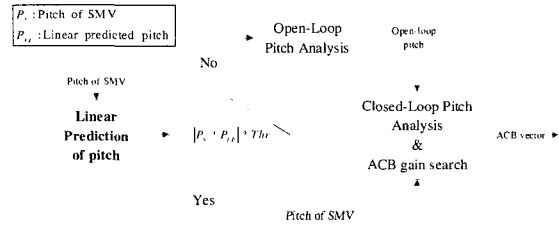


그림 3. SMV에서 G.723.1로의 피치 지연 변환의 블록도
Fig. 3. Block diagram of pitch conversion from SMV to G.723.1.

즘은 두가지 장점을 가진다. 첫째로, tandem 방식에 비해 적은 계산량으로 개회로 피치 지연을 얻을 수 있다. G.723.1의 개회로 피치 지연 검색은 지각 가중된 음성의 상호 상관도(Cross-Correlation)를 계산하고, 그것의 최대값을 찾는 과정이 필요하지만, 제안된 알고리즘에서는 피치 지연의 예측 과정과 예측된 피치 지연과 부호화된 피치 지연간의 비교 과정만이 필요하므로 계산량이 줄어든다. 둘째로, 두 음성부호화기에서 얻어지는 피치 지연의 큰 차이로 인해 발생하는 음질 저하 문제를 해결할 수 있다.

IV. G.723.1에서 SMV 음성부호화기로의 상호부호화

1. 전체적인 구조

<그림 4>는 G.723.1에서 SMV 음성부호화기로의 상호부호화 알고리즘의 구조를 나타낸다. G.723.1에서 SMV 음성부호화기로의 상호부호화는 파라미터 복호화, LSP 변환, 여기 신호의 변환 등의 과정을 거친다는 면에서 SMV에서 G.723.1로의 상호부호화와 유사한 구조를 가진다. 그러나 SMV 음성부호화기는 각 프레임마다 비트율과 타임을 결정하기 위해 비트율 결정 과정을 거치므로, G.723.1에서 SMV 음성부호화기로의 상호부호화에서도 비트율 결정을 위한 추가적인 과정이 필요하다.

G.723.1에서 SMV 음성부호화기로의 상호부호화 과정은 다음과 같다. 입력은 G.723.1 음성부호화기에서 부호화된 비트열이다. 입력된 비트열은 먼저 파라미터 복호화 과정을 거친다. 복호화되는 파라미터들은 LSP, 피치 지연, 적응 코드북 이득 및 고정 코드북 이득이다. G.723.1에서 SMV 음성부호화기로의 상호부호화의 경우 파라미터 복호화 과정에서 상호부호화를 위한 파라

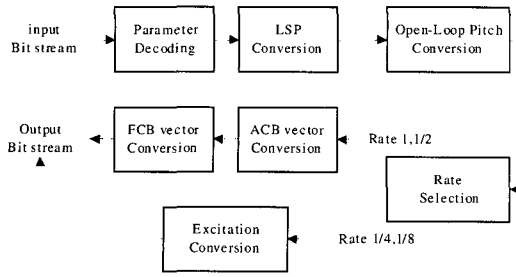


그림 4. 제안된 파라미터 직접 변환 방식의 G.723.1에서 SMV로의 상호부호화 알고리즘의 블록도
Fig. 4. Block diagram of the proposed transcoding algorithm from G.723.1 to SMV.

미터 뿐만이 아니라 비트율 결정을 위한 음성 파라미터도 계산한다. LSP 변환과 피치 지연 변환을 거친 후, 제안된 상호부호화 알고리즘은 간략화된 비트율 결정 알고리즘을 이용하여, 음성 프레임의 클래스를 결정하고 결정된 클래스와 외부에서 입력된 모드 정보에 따라 비트율과 타입을 결정한다. 마지막으로, 앞서 결정된 비트율과 타입에 따라서 합성 필터를 위한 여기신호를 구해낸다. Rate 1과 Rate 1/2에 대해서는 피치 변환을 통해 얻어진 개회로 피치 지연을 기초로 하여 SMV의 적응 코드북 검색 방식을 이용하여 폐회로 피치 지연과 적응 코드북 이득을 구한다. 고정 코드북의 검색 또한 SMV 음성부호화기와 같은 방법으로 수행한다. Rate 1/4과 Rate 1/8의 경우에는 SMV 음성부호화기에서처럼 난수 발생기를 이용하여 여기신호를 발생시킨다. 각 과정은 다음 절 등에서 좀더 상세히 설명된다.

2. LSP 변환

G.723.1에서 SMV 음성부호화기로의 LSP 변환도 SMV에서 G.723.1 음성부호화기로의 상호부호화의 경우와 마찬가지로 간단한 선형 보간을 이용하여 이루어진다. G.723.1에서 SMV 음성부호화기로의 LSP 변환 과정은 <그림 5>에 나타나있다. <그림 2>에서와 마찬가지로, 전송되는 LSP 세트에 해당하는 선형 분석 윈도우만이 표시되어 있다. 구체적인 선형 보간의 식은 다음과 같다.

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

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

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

(492)

여기서도 LSP_L^{SMV} 과 LSP_K^G 는 각각 SMV 음성 부호화기의 L 번째 프레임의 LSP와 G.723.1 음성부호화기의 K 번째 프레임의 LSP이다. 식 (3)을 살펴보면, <그림 5>에서처럼 G.723.1 음성 부호화기의 $K-1$ 번째와 K 번째 프레임의 LSP를 이용하여 SMV 음성부호화기의 L 번째 프레임의 LSP를 얻어냄을 알 수 있다. 또한, 식 (4)에서는 G.723.1 음성 부호화기의 K 번째와 $K+1$ 번째 프레임의 LSP로부터 SMV 음성부호화기의 $L+1$ 번째 프레임의 LSP가 얻어지는 것을 알 수 있고, 식 (5)에서는 G.723.1 음성부호화기의 $K+1$ 번째 프레임의 LSP가 SMV 음성부호화기의 $L+2$ 번째 프레임의 LSP로 바로 이용되는 것을 볼 수 있다.

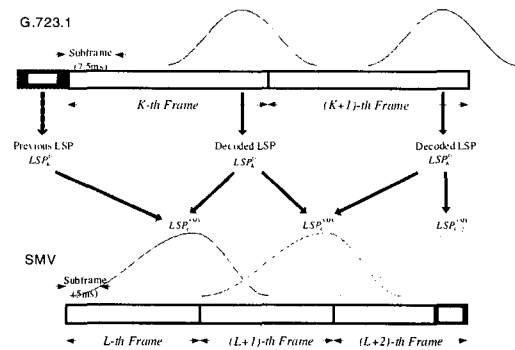


그림 5. G.723.1에서 SMV로의 LSP 변환
Fig. 5. LSP conversion from G.723.1 to SMV via linear interpolation.

3. 피치 지연 및 적응 코드북 이득 변환

G.723.1에서 SMV로의 상호부호화에서는, SMV에서 G.723.1로의 상호부호화에 제안된 선형 예측을 이용한 피치 지연 변환 알고리즘이 적용되지 않았고, 두 음성부호화기의 개회로 피치 지연 검색 영역을 일치시키기 위한 선형 보간만이 이용되었다. 파라미터 복호화 과정에서 얻어진 G.723.1의 개회로 피치 지연은 <그림 6>의 선형 보간법을 이용하여 SMV의 개회로 피치로 변환된다. <그림 6>에서 $P_K^G[i]$ 와 $P_L^S[j]$ 는 각각 G.723.1과 SMV의 피치 지연을 나타내고, L 과 K 는 프레임 번호를, i, j 는 부프레임 번호를 나타낸다. G.723.1과 SMV의 프레임 크기가 다르므로, 이것을 고려하여 G.723.1의 두 프레임과 SMV의 세 프레임을 비교하여 피치 지연을 변환한다. SMV는 각 프레임마다 2개의 개회로 피치 지연을 구하므로, 8개의 G.723.1의 피치 지연 값으로부터 6개의 SMV의 개회로 피치 지연 값을

언는다. 개회로 피치 지연이 결정되면, 변환된 개회로 피치 지연을 이용하여 SMV의 폐회로 피치 탐색 알고리즘을 이용하여 앞서 결정된 현재 프레임의 전송률과 타입에 따라서 폐회로 피치 지연 및 적응 코드북 이득을 계산한다.

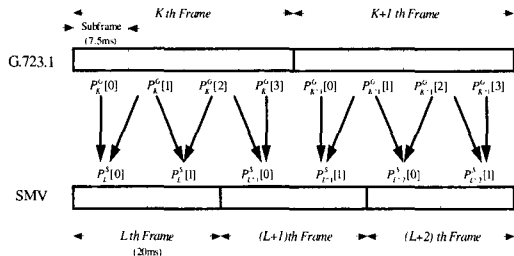


그림 6. G.723.1에서 SMV로의 피치 지연
Fig. 6. Pitch conversion from G.723.1 to SMV via linear interpolation.

4. 비트율 결정

SMV는 입력 음성 프레임의 6가지 프레임 클래스로 분류하고, 분류된 결과와 외부에서 입력된 모드 정보를 이용하여 비트율과 타입을 결정한다. 그러나 이러한 처리과정들은 다양한 음성 파라미터들을 필요로 하며, 이 파라미터들은 입력 음성을 이용하여 계산된다. 또한 이 음성 파라미터는 선형 예측 분석과정, VAD 과정, 개회로 피치 분석 과정 등을 통해 얻어진다. 그런데, G.723.1에서 SMV로의 상호부호화 과정에서는 선형 예측 분석과정, 개회로 피치 분석 과정이 이루어지지 않으므로, 비트율과 타입을 결정하기 위한 음성 파라미터들을 얻어낼 수 없다. 따라서, 이러한 음성 파라미터를 사용하지 않고 비트율과 타입을 결정하기 위해, 간략화된 비트율 결정 알고리즘을 제안하였다. 제안된 비트율 결정 알고리즘에서는 파라미터 복호화 과정에서 얻어진 G.723.1의 파라미터들을 몇가지 선형적 계산을 통해 변형하고, 변형된 파라미터들과 고정된 문턱값의 비교를 통해 비트율과 타입의 결정을 위한 프레임 클래스 분류를 수행한다. 또한 SMV 음성부호화기의 비트율 결정 알고리즘과는 달리 잡음 클래스를 생략하고, 프레임 클래스를 5가지로 줄여 분류한다. 프레임 클래스 분류에 사용되는 여러 문턱값들은 다양한 실험 음성을 대상으로 평균 전송률이 크게 달라지지 않는 범위 내에서 음질이 가장 좋은 값으로 결정하였다.

<그림 7>은 제안된 비트율 결정 알고리즘의 블록도

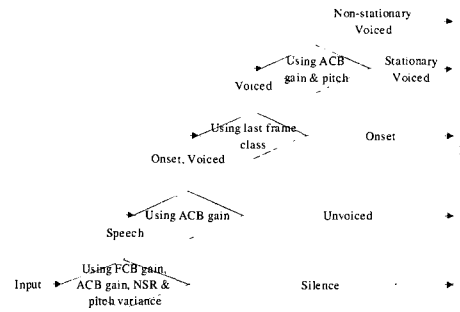


그림 7. 간략화된 비트율 결정 알고리즘의 블록도
Fig. 7. Block diagram of simplified rate selection algorithm.

를 나타낸다. <그림 7>에서 볼 수 있듯이, 제안된 알고리즘에서는 파라미터 복호화 과정을 통해 얻어진 파라미터와 새롭게 계산된 음성 파라미터인 잡음 대 신호비 (Noise to Signal Ratio, NSR)을 이용하여 모두 5가지의 프레임 클래스로 분류 하고, 분류된 클래스 정보를 비트율과 타입 결정에 이용한다. 프레임 클래스로부터 비트율과 타입을 결정하는 과정은 SMV 음성부호화기와 거의 동일한 과정을 거친다. 다양한 실험을 통해, 유성음의 경향을 보이는 음성 프레임의 경우에는 적응 코드북 이득값이 크게 나타나는 것을 관찰할 수 있었다. 따라서, 적응 코드북 이득을 적절한 선형 변형을 통해 변형시키고 그 결과를 문턱값과 비교하여, 음성 프레임을 묵음, 무성음, 유성음으로 분류하도록 하였다. 또한, 피치 지연의 편차(Pitch variance)가 유성음 프레임에서는 작은데 비해, 잡음이 나 묵음 프레임에서는 크다는 사실도 프레임 클래스 분류에 이용될 수 있다. 피치 지연의 편차는 음성 프레임과 묵음 프레임의 구분에 이용된다. 고정 코드북 이득 역시 유성음 특성과 유사한 경향을 보이는 것을 알 수 있었다. 하지만, 고정 코드북 이득의 경우에는 잡음에 쉽게 영향을 받기 때문에 고정 코드북만을 프레임 클래스 분류에 이용할 수는 없다. 잡음 문제를 해결하기 위해서, SMV 음성부호화기에서 프레임 클래스 분류를 위해 사용하는 알고리즘을 이용하여 잡음 대 신호비를 구하고, 이것을 고정 코드북 이득과 함께 고려한다.

<그림 7>에서 볼 수 있는 것처럼, 간략화된 비트율 결정 알고리즘의 프레임 클래스 분류 과정은 다음과 같다. 먼저, 고정 코드북 이득과 적응 코드북 이득을 선형 변형하여 얻은 값과 잡음 대 신호비, 피치 지연의 편차를 문턱값과 비교하여 묵음과 음성으로 프레임을

구분한다. 음성으로 구분된 프레임은 다시 적응 코드북을 이용하여 무성음과 유성음으로 구분한다. 유성음 클래스에 대해서, 무성음에서 유성음으로의 전이가 있는 경우에는 해당 프레임은 변화 클래스로 분류된다. 변화가 아닌 유성음의 경우에는 일정한 유성음과 일정하지 않은 유성음으로 나뉘어지며, 이 구분에는 현재 음성 프레임 내의 모든 부프레임에서 결정되는 적응 코드북 이득값과 피치 지연의 변화가 사용된다. 만약 각 적응 코드북 이득이 문턱값보다 크고 피치 지연의 변화가 크지 않으면, 그 프레임은 일정한 유성음 클래스로 분류된다. 프레임 클래스 분류가 끝나면, 비트율과 타입이 결정된다.

V. 성능 평가

본 장에서는 제안된 파라미터 직접 변환 방식의 상호부호화 알고리즘의 성능을 다양한 지표를 이용하여 평가한다. 성능 평가를 위해 사용된 지표는 계산량, 객관적 음질 평가, 지연 시간이다. 각 모의 실험에 대해 제안된 알고리즘의 성능을 tandem 알고리즘의 성능과 비교 평가한다. 성능 평가를 위하여 제안된 파라미터 직접 변환 방식의 상호부호화 알고리즘을 C 언어를 이용하여 고정 소수점 형태로 구현하였다. 그리고 한국인 남성 화자와 여성 화자가 발성한 약 60초 길이의 문장을 모든 모의 실험에서 공통적으로 사용하였다. 각 음성 파일은 8kHz로 표본화 되었으며 16 비트 선형 PCM 형식이다. 전송 과정의 오류는 없다고 가정하였다.

1. 계산량

제안된 알고리즘에 의해 소요되는 계산량을 측정하기 위해서 Weighted Million Operations Per Second (WMOPS)를 이용하였다. <표 1>과 <표 2>는 제안된 파라미터 직접 변환 방식 알고리즘의 계산량과 Tandem 방식 알고리즘의 계산량을 비교한 것이다. <표 1>, <표 2>의 결과를 살펴보면, 제안된 파라미터 직접 변환 방식의 계산량이 Tandem 방식에 비하여 최저 20%에서 최고 35%까지 줄어드는 것을 확인할 수 있다. 이러한 계산량 감소는 부가적인 선형 예측 분석의 생략, 개회로 피치 지연 검색의 생략, 잠음 프레임에서의 적응코드북 검색 생략 및 간략화된 비트율 결정 알고리즘의 도입 등으로 가능하였다.

표 1. 계산량 비교(SMV→G.723.1)

Table 1. Comparison of computational complexity(SMV→G.723.1).

Speaker	Mode	Tandem	Proposed
Male	0	19.95	12.59
	1	20.04	12.42
	2	20.24	12.32
	3	20.26	12.30
Female	0	20.01	13.95
	1	20.20	13.80
	2	20.30	13.79
	3	20.30	13.77

표 2. 계산량 비교(G.723.1→SMV)

Table 2. Comparison of computational complexity(G.723.1→SMV).

Speaker	Mode	Tandem	Proposed
Male	0	29.46	23.30
	1	27.69	21.90
	2	27.44	21.53
	3	27.52	21.52
Female	0	30.22	24.97
	1	29.47	22.81
	2	28.47	22.65
	3	29.24	22.67

표 3. 객관적 음질 평가 결과(SMV→G.723.1)

Table 3. Comparison of PESQ score(SMV→G.723.1).

Speaker	Mode	Tandem	Proposed
Male	0	3.237	3.239
	1	3.180	3.174
	2	3.098	3.135
	3	3.026	3.091
Female	0	3.066	3.023
	1	2.996	2.997
	2	2.942	2.960
	3	2.915	2.973

2. 객관적 음질 평가

객관적 음질 평가 방법 중의 하나인 ITU P.862 Perceptual Evaluation of Speech Quality (PESQ)^[8]를 사용하여 제안된 상호부호화 알고리즘의 객관적 음질

을 평가하였다. <표 3>과 <표 4>는 Tandem 방식과 제안된 파라미터 직접 변환 방식의 PESQ 점수를 정리하여 비교한 것이다. <표 3>, <표 4>를 보면 전체적으로 Tandem 방식과 제안된 파라미터 직접 변환 방식의 PESQ 점수 차가 PESQ의 오차 범위 내일 정도로 미미하고, 어떤 경우에는 제안된 방식의 음질이 조금 더 나은 것을 알 수 있다. 이 결과는 제안된 파라미터 직접 변환 방식의 상호부호화 알고리즘이 Tandem 방식과 전체적으로 동등한 음질을 구현함을 보여주고 있다.

표 4. 객관적 음질 평가 결과(SMV→G.723.1)
Table 4. Comparison of PESQ score(SMV→G.723.1).

Speaker	Mode	Tandem	Proposed
Male	0	3.303	3.244
	1	3.204	3.166
	2	3.112	3.050
	3	3.095	2.982
Female	0	3.081	3.102
	1	2.959	3.069
	2	2.904	2.863
	3	2.957	2.817

3. 지연 시간

시간 지연의 평가에서는, 전송 지연은 통신망의 구조와 상태에 따라 달라지기 때문에 고려하지 않고 알고리즘 지연과 처리 지연 시간만을 고려하여 그 합을 총 지연 시간으로 가정한다. SMV에서 G.723.1 음성부호화 기로의 상호부호화의 경우, 선형 예측 분석을 위한 예견 구간이 필요하지 않으므로, 알고리즘 지연이 7.5ms 만큼 줄어들게 된다. 반면에, G.723.1에서 SMV로의 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. 결 론

본 논문에서는 VoIP 망에서 널리 사용되는 G.723.1 음성부호화기와 CDMA2000 시스템을 위한 새로운 음

성 표준인 Selectable Mode Vocoder(SMV) 음성부호화기를 위한 파라미터 직접 변환 방식의 상호부호화 알고리즘을 제안하였다. 제안된 알고리즘은 C 언어로 구현되어 다양한 성능 지표를 통하여 평가되었다. 앞 장에서 제시된 성능 평가 결과는 제안된 알고리즘이 Tandem 방식에 비하여 적은 계산량과 짧은 지연 시간으로도 동등한 수준의 음질을 구현함을 보여주고 있다. 추가적으로, 본 논문에서는 고려되지 않은 잡음이 첨가된 음성이나 전송 과정에서의 오류가 있는 경우 등의 상황에서의 상호부호화 알고리즘의 동작에 대한 연구가 필요할 것이다.

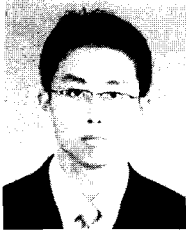
참 고 문 헌

- [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.
- [5] Hong-Goo Kang, Hong-Kook Kim, R.V. Cox, "Improving transcoding capability of speech coders in clean and frame erased channel environments," In Proc. IEEE Workshop on Speech Coding, 2000, pp. 78-80, Jan., 2000.
- [6] Sung Wan Yoon, Sung Kyo Jung, Young Cheol Park, and Dae Hee Youn, "An efficient transcoding algorithm for G.723.1 and G.729A speech coders", In Proc. Eurospeech 2001, vol. 4, pp. 2499-2502, 2001.
- [7] Sunil Lee, Seongho Seo, Dalwon Jang, Chang D. Yoo, "A novel transcoding algorithm for AMR and EVRC speech coders via direct parameter transformation," In. Proc. ICASSP 2003, 2003.

[8] ITU-T Rec. P.862, "Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of

narrowband telephone networks and speech codecs," 2000.

저 자 소 개



徐 晟 豪(正會員)

2002년 2월 : 한국과학기술원 전자전산학과 전기 및 전자공학전공 학사. 2002년 3월~현재 : 한국과학기술원 전자전산학과 전기 및 전자공학전공 석사과정

李 善 一(正會員)

2001년 2월 : 연세대학교 기계전자공학부 전자공학전공 학사. 2002년 8월 : 한국과학기술원 전자전산학과 전기 및 전자공학전공 석사. 2002년 9월~현재 : 한국과학기술원 전자전산학과 전기 및 전자공학전공 박사과정

張 達 源(正會員)

2002년 2월 : 한국과학기술원 전자전산학과 전기 및 전자공학전공 학사. 2003년 8월 : 한국과학기술원 전자전산학과 전기 및 전자공학전공 석사. 2003년 9월~현재 : 한국과학기술원 전자전산학과 전기 및 전자공학전공 박사과정

柳 昌 東(正會員)

1986년 : California Institute of Technology 학사. 1988년 : Cornell University 석사. 1996년 : MIT 박사. 1997년 1월~1999년 3월 : KT Senior Researcher. 1999년 3월~현재 : 한국과학기술원 전자전산학과 조교수