

8 kbps G.729A에서 5.3 kbps G.723.1로의 상호부호화 알고리즘

윤성완*, 정성교*, 박영철**, 윤대희*

*연세대학교 전기·전자 공학과 음향, 음성, 신호처리 연구실

**연세대학교 신호 처리연구센터

Transcoding Algorithm from 8 kbps G.729A to 5.3 kbps G.723.1

Sung-Wan Youn*, Sung-Kyo Jung*, Young-Chul Park**, and Dae-Hee Youn*

*ASSP Lab. Department of Electric & Computer Engineering, Yonsei, University

**Center for Signal Processing Research, Yonsei, University

e-mail : yocello@radar.yonsei.ac.kr

요약문

유/무선 통신 시스템에서는 통신망마다 각각 다른 음성 부호화기를 사용하므로 음성신호는 두 번의 부/복호화 과정을 거치게 된다. 이로 인해 음질저하, 계산량 증가, 그리고 전달 지연 증가 등의 문제가 발생된다. 본 논문에서는 위의 문제점들을 개선하기 위하여 패킷 음성통신과 무선 이동 통신에 사용되는 음성 부호화기의 상호부호화를 위한 알고리즘을 제안한다. 효율적인 음성 패킷 변환 방법을 제안하였으며, 8 kbps G.729A 패킷을 5.3 kbps G.723.1 패킷으로 변환하는 방법을 제안한다.

제안된 음성 패킷 변환 방법은 LSP 변환과정, 적응 코드북 변환과정, 그리고 고정 코드북 고속 탐색 과정으로 구성된다.

여러 가지 음성 신호로 모의 실험한 결과, 본 논문에서 제안된 상호부호화 알고리즘이 두 번의 부/복호화 과정을 거친 경우보다 짧은 전달 지연 시간과 적은 계산량으로 동등한 음질의 음성신호로 복호화함을 확인하였다.

1. 서론

최근 들어 무선통신망과 유선네트워크망 사이에서의 음성을 저장, 전송하려는 시도가 많이 이루어지고 있다. 이때 각각의 통신망에서 사용되는 음성부호화기가 서로 다르므로 음성신호가 그림 1(a)과 같이 두 번의 부/복호화 과정을 거치게 되는데, 이를 tandem이라고 하며, 다음과 같은 문제점들을 갖게 된다. 첫째, 서로 다르게 표준화된 음성부호화기를 사용하는 단말기 사이에서 음성신호를 주고 받기 때문에 음성신호는 두 번의 부/복호화 과정을 거치게 되어 동일한 음성부호화기를 사용하는 단말기사이의 통신에서보다 통화품질이 떨어지게 된다. 둘째, 중계기에서 이중 부/복호화 과정이 이루어지므로 많은 계산량을 요구한다. 셋째, 음성신호를 두 번 압축하기 때문에 LPC 분석시에 예견구간 등 추가 버퍼링이 필요하게 되므로 전체적인 전달시간 지연이 발생한다.

이와 같은 문제점들의 근본적 원인은 음성신호가 음성부호화기를 두 번 통과하기 때문이다. 이중 부/복호화로 인해 발생하는 문제점들을 해결하기 위해서는 한쪽 통신망에서 음성신호를 부호화한 비트열 (Bit-stream)을 다른 쪽 통신망에서 그대로 복호화하는 방법이 필요하다. 즉, 한쪽 통신망의 압축된 비트열을 PCM 신호로 복호화한 후, 다른 쪽 통신망에서 다시 한번 부/복호화 과정을 거치는 것이 아니라, 그림 1(b)에서와 같이 전송되어온 비트열을 그대로 한번의 복호화 과정을 거치게 하는 것이다. 그러나 이런 방법은 같은 음성부호화기를 사용할 때에만 가능한 것이다. 따라서 서로 다른 음성부호화기끼리의 상호부호화를 통해 위에서 언급한 문제점을 해결할 수 있는 방법의 개발이 필요하다.

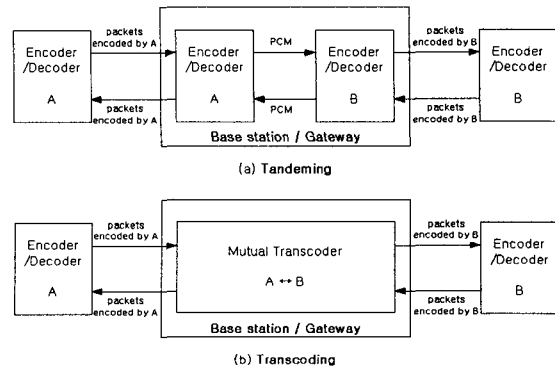


그림 1. 유/무선 통신망 (a) Tandem(이중 부/복호화) (b) 상호부호화기를 이용한 통신

본 논문에서는 8 kbps G.729A [3]에서 5.3 kbps G.723.1 [1]로의 상호부호화 알고리즘을 제안하였다. 제안된 상호부호화 알고리즘은 LSP 변환과정, 적응 코드북 변환과정, 그리고 고속 고정 코드북 탐색 과정으로 구성된다. 두 부호화기의 프레임 길이를 고려하여 10ms 프레임을 갖는 G.729A의 3개의 프레임이 30ms

프레임을 갖는 G.723.1의 프레임 1개로 변환된다.

여러 가지 음성 신호로 모의 실험한 결과, 본 논문에서 제안한 상호부호화 알고리즘을 이용한 패킷변환기가 두 번의 부/복호화 과정을 거친 경우보다 짧은 전달 지연 시간과 적은 계산량으로 동등한 음질의 음성 신호를 복호화하는 것을 확인하였다.

본 논문의 구성은 다음과 같다. 2장에서는 G.729A와 G.723.1 음성부호화기의 알고리즘을 간단히 설명하고, 3장에서는 본 논문에서 제안된 변환 알고리즘을 LSP 변환과정, 적응 코드북 변환과정, 고정코드북 탐색과정으로 나누어 살펴본다. 4장에서는 주/객관적 평가를 통하여 제안된 알고리즘과 기존의 이중 부/복호화의 성능을 비교하고, 5장에서는 결론을 맺는다.

2. ITU-T G.723.1&G.729A 음성부호화기

멀티미디어 통신용 음성 부호화 표준안으로 제정된 ITU-T G.723.1 음성부호화기는 5.3 kbps와 6.3 kbps의 두개의 전송율을 제공한다. 이 음성부호화기는 음성이나 오디오 신호를 30ms의 길이를 갖는 프레임 단위로 부호화하며 7.5ms의 예견구간이 있기 때문에 총 37.5ms의 알고리즘 지연이 생긴다.

부호화 과정을 간략히 살펴보면, 우선 7.5ms의 길이의 부프레임마다 10차의 LPC 분석을 하고, 마지막 부프레임의 LPC계수를 예측분할 벡터 양자화를 통하여 수신단에 전송한다. 나머지 LPC계수들은 단구간 지각가중필터를 만들 때 사용한다. 선형예측이 끝나면, 2개의 부프레임마다 개회로 피치 주기를 구하고, 이후 음성신호는 부프레임단위로 적응 코드북, 고정 코드북 탐색을 하게 된다. 적응 코드북 탐색에서는 5차의 피치 예측기를 이용하여 폐회로 피치추기와 이득값을 구한다. 고정 코드북은 6.3 kbps의 전송율에선 MP-MLQ (Multi Pulse - Maximum Likelihood Quantization) 방식, 5.3 kbps 전송율에선 ACELP (Algebraic Code Excited Linear Prediction) 방식으로 부프레임마다 4개의 펄스를 검색하게 된다.

G.729 음성부호화기 [2]는 CS-ACELP (Conjugate Structure Algebraic CELP)에 기반한 8 kbps의 전송율에서 동작한다. G.729 부호화기는 10ms의 프레임 단위로 음성신호를 부호화하며 5ms의 예견구간이 있기 때문에 총 15ms의 알고리즘 지연이 생긴다. 부호화 과정을 간략히 살펴보면, 우선 10ms의 프레임마다 10차의 LPC 계수를 구하고 이를 다단계 벡터 양자화를 이용해 수신단으로 전송한다. LPC 분석과 지각가중필터를 계산한후, 개회로 피치 주기를 프레임마다 한번씩 계산한다. 그 다음에 5ms의 부프레임마다 적응 코드북, 고정 코드북 탐색을 하게 된다. 적응 코드북에서는 피치 범위에 따라 1/3의 분해능을 갖는 실수 피치값을 찾게 되며, 고정 코드북 탐색에서는 부프레임마다 4개의 펄스를 검색한다. 적응 코드북과 고정 코드북의 이득값 양자화에는 검색의 효율을 위해 켈레 (conjugate)구조를 갖는 2개의 코드북을 사용한다.

G.729A 음성부호화기는 G.729의 계산량을 감소시킨 것으로, 압축된 비트열은 G.729와 상호호환 가능하다 [3]. G.729A에서 계산량이 감소된 부분은, LPC분석, 적응 코드북, 고정 코드북 탐색, 그리고 후처리 필터이

다.

3. 제안된 상호부호화 알고리즘

제안된 G.729A에서 G.723.1 (5.3 kbps)로의 상호부호화 알고리즘들은 크게 LSP 변환과정, 적응 코드북 변환과정, 고속 고정 코드북 탐색과정으로 구성된다. 그리고 두 부호화기의 프레임 길이를 고려하여 10ms 프레임임을 갖는 G.729A의 3개의 프레임이 30ms 프레임을 갖는 G.723.1의 프레임 1개로 변환된다.

3.1 선형보간을 이용한 LSP 변환

제안된 LSP 변환과정에선, G.729A의 3개 프레임의 LSP 정보를 선형보간법을 통하여 (1)과 같이 G.723.1의 1개의 프레임에 해당하는 LSP 값을 구해준다.

$$P_A(i) = 0.05 \cdot P_{B1}(i) + 0.15 \cdot P_{B2}(i) + 0.80 \cdot P_{B3}(i), 1 \leq i \leq 10 \quad (1)$$

여기서 $P_A(i)$ 는 G.723.1의 LSP값, $P_{B1}(i), P_{B2}(i), P_{B3}(i)$ 는 G.729A 각 프레임의 LSP값이다.

그림 2는 유성음부분의 LPC 스펙트럼을 나타내고 있다. 그림에서 보는 바와 같이 음질에 중요한 저주파 부분에서 원래의 G.723.1의 LPC 스펙트럼에 이중 부/복호화한 LSP보다 본 논문에서 제안한 LSP 변환과정을 통해 얻은 LSP가 더 유사함을 알 수 있다.

또한, 이중 부/복호화의 경우 재생된 음성신호로 LPC분석을 다시 한번 해야 하는 반면, 제안한 LSP 변환법은 LPC 분석을 거치지 않아 적은 계산량으로 좋은 성능을 얻는다.

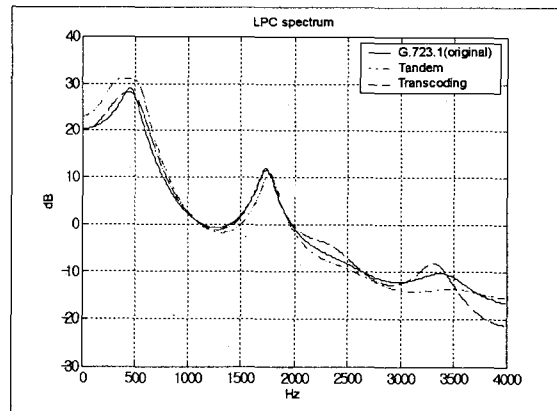


그림 2. 이중 부/복호화 방법과 LSP변환의 LPC 스펙트럼비교.

3.2 후방향 추적을 이용한 개회로 피치 변환

후방향 피치추적을 이용한 개회로 피치계산은 그림 3에 나타나 있다. 우선 첫번째 개회로 피치를 구할 때에는 이전 프레임의 마지막 부프레임과 G.729A의 첫번째 부프레임의 폐회로 피치의 ± 5 샘플 범위에서 (2)를 최대화하는 지역 최대 지연값을 찾은 후, (3)과 같이 G.723.1의 지역 최대값이 G.729A의 지역 최대값의 3/4배 보다 클 경우 G.723.1의 지연값을, 반대의 경우

에는 G.729A의 지연값을 개회로 피치값으로 결정한다. 두 번째 개회로 피치의 후보값으로는 이전 개회로 피치와 G.729A의 2번째 프레임의 2번째 부프레임의 폐회로 피치가 쓰이며, 역시 ± 5 샘플 범위에서 지역최대값을 구하고, G.723.1의 지역최대값이 G.729A의 지역최대값의 3/4배보다 클 경우 G.723.1의 지연값을, 반대의 경우에는 G.729A의 지연값을 개회로 피치로 결정한다.

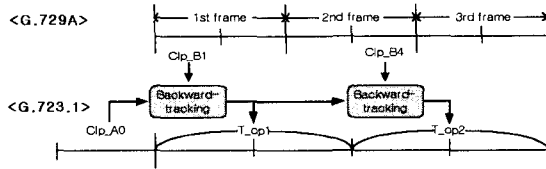


그림 3. 후방향 피치 추적.

여기서 Clp_B1 과 Clp_B4 는 G.729A의 폐회로 피치이며 Clp_A0 는 G.723.1의 이전 프레임 중 마지막 부프레임의 폐회로 피치, T_op1 과 T_op2 는 현재 프레임의 개회로 피치이다. 만약, 후방향 피치추적에 쓰이는 두 후보값의 차이가 10샘플 미만일 때 Clp_B1 이 T_op1 로, Clp_B4 가 T_op2 로 쓰인다.

$$C_{ol}(j) = \frac{\left(\sum_{n=0}^{119} f(n) \cdot f(n-j) \right)^2}{\sum_{n=0}^{119} f(n-j) \cdot f(n-j)}, T-5 \leq j \leq T+5 \quad (2)$$

여기서 T 는 지역최대값의 후보이며, $f(n)$ 은 지각가중된 음성신호이다.

$$\begin{aligned} T_{op} &= t_1 \\ C_{ol}(T_{op}) &= C_{ol}(t_1) \\ \text{if } C_{ol}(t_2) &\geq 0.75 \cdot C_{ol}(T_{op}) \\ C_{ol}(T_{op}) &= C_{ol}(t_2) \\ T_{op} &= t_2 \\ \text{end} \end{aligned} \quad (3)$$

여기서 t_1 과 t_2 는 각각 G.729A와 G.723.1의 지역최대값의 피치이다.

3.3 적응코드북 변환

G.723.1 음성부호화기에서 기존의 5차 피치 예측기를 사용한 적응 코드북 탐색은 지연과 이득을 동시에 검색하기 때문에 많은 계산량이 든다. 반면, [4]에서 제안된 고속 적응 코드북 탐색은 지연과 이득을 순차적으로 구한다. 즉, 지연은 G.729처럼 1차 피치예측기와 같은 방법으로 찾고 구해진 지연에 대해 5차 피치 예측기의 이득을 구한다.

5.3 kbit/s의 전송율에서 G.723.1은 부프레임마다 적응 코드북 이득값표의 170개의 후보값을 탐색하게 되어 많은 계산량을 요구된다. 본 논문에서는 [4]의 고속

적응 코드북 탐색을 기반으로 하여 (4)에서와 같이 G.729A의 적응코드북 이득값에 따라 5차 피치예측기의 이득값 탐색범위를 제한시켜 주었다.

$$\begin{aligned} \text{if } 0 \leq g_p < 0.3 \\ bnd0 &= 0, \quad bnd1 = 60 \\ \text{if } 0.3 \leq g_p < 0.5 \\ bnd0 &= 0, \quad bnd1 = 60 \\ \text{if } 0.5 \leq g_p < 0.7 \\ bnd0 &= 0, \quad bnd1 = 60 \\ \text{if } 0.7 \leq g_p < 0.9 \\ bnd0 &= 0, \quad bnd1 = 60 \\ \text{if } 0.9 \leq g_p < 1.05 \\ bnd0 &= 0, \quad bnd1 = 60 \\ \text{else} \\ bnd0 &= 0, \quad bnd1 = 170 \end{aligned} \quad (4)$$

여기서 g_p 는 G.729A의 적응코드북 이득값, $bnd0$, $bnd1$ 는 각각 g_p 에 따른 적응코드북 이득값표의 탐색범위의 아래쪽, 위쪽 경계값이다.

3.4 고속 고정코드북 탐색

5.3 kbps G.723.1의 고정 코드북 탐색은 ACELP 구조를 기반으로 이루어지며, 각 부프레임을 4개의 트랙으로 나누고 각 트랙마다 1개의 펄스를 검색하는 nested-loop search를 사용한다. 이로 인해 펄스위치를 검색할 때 $8 \times 8 \times 8 \times 8 = 4096$ 의 경우의 수에 대해 분석-합성 방법으로 위치를 검색하며, 실제적인 구현에서는 마지막 펄스의 검색을 제한함으로써 계산량을 줄인다. 본 논문에서는 G.729A에서 사용한 depth-first tree search에 기반한 고속 고정 코드북 탐색방법을 적용하였다. 이로써 펄스 위치의 경우의 수는 $2 \times \{(8 \times 8) + (8 \times 8)\} = 256$ 으로 줄었다.

4. 상호부호화 알고리즘 성능 평가

4.1 객관적 음질평가

제안된 상호부호화 알고리즘에 대해 객관적 음질평가와 주관적 음질평가를 수행하였다. 객관적 음질평가로는 8초길이의 한국어인 남녀 각각 2문장에 대해 LPC-CD (LPC-Cepstral Distance)와 PSQM (Perceptual Speech Quality Measure) [6]을 사용하였다.

일반적으로 시간 영역보다 주파수 영역의 객관적 평가방법이 주관적 평가와 높은 상관관계가 있으며, LPC-CD는 (5)와 같이 정의된다 [5].

$$LPC_CD = 10 / \log_{10} \sqrt{2 \sum_{i=1}^p \{C_x(i) - C_y(i)\}^2} \quad (5)$$

여기서 $C_x(i)$ 와 $C_y(i)$ 는 각각 음성부호화기의 입력과 출력 음성신호의 켈프스트럼 계수이고 p 는 켈프스트럼 차수이다.

기존의 이중 부/복호화 방법과 본 논문에서 제안한 상호부호화 방법을 사용해 음성신호를 부/복호화 했을 때 LPC-CD가 표 1에 정리되어 있다. 표 1에서 보는 바와 같이 여성화자와 남성화자 모두 LPC-CD값이 이

중 부/복호화의 경우보다 본 논문에서 제안한 상호부호화 방법이 더 작은 수치를 나타냈다. 이는 그림 3의 LPC 스펙트럼 비교와 일치하는 결과로 볼 수 있다. 실험결과, LPC-CD의 차이는 여성화자의 경우 0.14dB, 남성화자의 경우 0.12(dB) 였다.

표 1. LPC-CD결과.

| LPC-CD | Female | Male |
|-------------|----------|----------|
| Tandem | 4.34(dB) | 4.36(dB) |
| Transcoding | 4.20(dB) | 4.24(dB) |

또 다른 객관적 음질평가의 척도로 PSQM을 사용했다. PSQM은 300-3400Hz의 전화 대역폭상에서 음성 부호화기의 객관적 음질 평가방법으로 주관적 평가와 높은 상관도를 갖고 있으며 ITU-T의 권고안이다.

기존의 이중 부/복호화 방법과 본 논문에서 제안한 상호부호화 방법을 사용해 음성신호를 부/복호화 했을 때의 PSQM 결과가 표 2에 정리되어 있다. 표 2에서 보는 바와 같이 여성, 남성 화자 모두 이중 부/복호화 방법보다 상호부호화 방법이 더 좋은 결과를 얻었다.

표 2. PSQM 결과.

| PSQM | Female | Male |
|-------------|--------|------|
| Tandem | 3.12 | 3.27 |
| Transcoding | 3.02 | 3.01 |

4.2 주관적 선호도 음질평가

주관적 음질평가로는 남녀 각각 10명에게 남성, 여성 화자의 문장을 2번씩 들려주고 이중 부/복호화 방법과 본 논문에서 제안한 방법의 선호도 조사를 하였다. 표 3은 기존의 이중 부/복호화와 제안된 상호부호화 방법의 선호도 평가 결과를 보여준다. 여성화자의 경우, 45%가 기존의 이중 부/복호화의 음질을 선호하였고 나머지 60%는 음질의 우열을 가리지 못하거나 제안된 알고리즘의 음질을 선호하였다. 남성화자의 경우 55%가 음질의 우열을 가리지 못하였고, 나머지 20%와 45%는 각각 제안된 알고리즘과 기존 방법의 음질을 선호하였다. 주관적 음질 평가를 통하여 제안된 상호부호화 알고리즘은 음질 저하가 거의 없음을 확인하였다.

표 3. 선호도 평가 결과.

| A-B Test | Tandem (A) Transcoding (B) | |
|---------------|-------------------------------|-------|
| | Female | Male |
| Preference A | 40(%) | 45(%) |
| Preference B | 20(%) | 20(%) |
| No Preference | 40(%) | 35(%) |

4.3 계산량

이중 부/복호화 방법과 본 논문에서 제안한 상호부호화 알고리즘의 계산량 비교를 위하여 펜티엄 II-400MHz CPU의 PC 환경에서 Visual C 콤파일러로 프로파일을 해보았다. 기존의 이중 부/복호화 과정에는 포함되어 있으나 상호부호화 방법의 연산에서 생략된 부분은 LPC 분석, LPC의 LSP로의 전환이며, 개회로

피치 예측, 적응 코드북, 고정 코드북 탐색 부분은 각각 후방향 피치 추적과 [4]에서 제안한 방법을 사용하여 계산량을 감소하였다. 표 3에는 각 모듈별로 기존의 방법과 제안한 방법의 Visual C 콤파일러의 프로파일 결과가 정리되어 있다.

표 4. 계산량 비교.

| | G729 Decoding | [POLSP] Analysis | LSP Conversion | Open-loop Pitch | ACBK | ROBK |
|-------------|------------------|---------------------|-------------------|--------------------|----------|---------|
| Tandem | same | 285(ms) | X | 335(ms) | 2854(ms) | 453(ms) |
| Transcoding | same | X | 231(ms) | 174(ms) | 1389(ms) | 370(ms) |

이중 부/복호화 방식과 상호부호화 방법을 사용하여 G.729A의 비트열에서 G.723.1의 압축된 비트열을 생성하는 데에 소요되는 시간은 각각 21.46ms, 16.80ms 이었다. 따라서 본 논문에서 제안한 상호부호화 방법이 기존의 방법보다 약 25% 적은 계산량으로 동등한 음질을 얻어냄을 확인하였다.

5. 결론

본 논문에서는 유/무선 통신망에서의 음성신호의 이중 부/복호화로 야기되는 계산량 증가, 전달시간 지연 등의 문제를 최소화할 수 있는 상호부호화 알고리즘을 제안하였다. 제안된 알고리즘은 8 kbps G.729A로 부호화된 패킷을 5.3 kbps G.723.1의 음성패킷으로 상호부호화 해줌으로써 이중 부/복호화를 할 경우보다 훨씬 적은 계산량으로 동등한 음질과 보다 짧은 전달시간 지연을 얻을 수 있었다.

현재는 제안된 알고리즘과 유사한 5.3 kbps G.723.1로부터 G.729A로의 상호부호화 알고리즘을 개발중에 있다.

참고 문헌

- [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.729 "Coding of Speech at 8 kbit/s CS-ACELP Speech Coder," 1996
- [3] ITU-T Rec. G.729 Annex A "Reduced Complexity 8 kbit/s CS-ACELP Speech Codec," 1996
- [4] 정성교, 박영철, 윤성완, 차일환, 윤대희, " 효율적인 다채널 구현을 위한 ITU-T G.723.1 음성부호화기 고속 알고리즘 제안", 한국음향학회 학술발표대회 논문집 제19권 제1호, pp67-70, 2000
- [5] N. Kitawaki, H. Nagabuchi, and K. Itoh, "Objective Quality Evaluation for Low-Bit-Rate Speech Coding System," in *IEEE Journal On Selected Areas In Commun.* vol. 7. No.2, pp.242-248, Feb. 1998
- [6] ITU-T Rec. P.861 "Objective Quality Measurement Of Telephoneband(300-3400Hz) Speech Codecs," 1996