

G.729A에서 EVRC로의 상호부호화

*곽영진, *정지민, *권구락, **임정석, **황인호, **이경훈, *고성재
*고려대학교 전자컴퓨터공학과, **NSRI
전화 : 02-3290-3672 / 핸드폰 : 016-297-6907

A Transcoding Algorithm from G.729A to EVRC

Abstract

Communication between speech networks employing different speech codecs requires interoperability. The cascade connection of two different codecs, called tandem coding, not only degrades speech quality, but also produces high computational loads. These problems can be solved by using the transcoding algorithm. This paper presents an effective algorithm for transcoding from G.729A to EVRC and its simulation results.

I. 서론

멀티미디어 통신이 급속히 발전해감에 따라 VoIP, 디지털 이동통신 등에서 이용되는 다양한 종류의 음성 표준 간의 상호부호화가 중요시 되고 있다. 무선 통신망에서는 디지털 셀룰라 CDMA 시스템의 음성 부호화기로 사용되는 EVRC (Enhanced Variable Rate Codec)[1]가 있고, 유선 네트워크에서는 H.323 표준을 기반으로 하는 VoIP의 연구가 진행되고 있으며 이 기술의 표준 음성코덱이 G.729 (CS-ACELP)[2]이다.

각각의 통신망에서 사용되는 음성신호를 연동시키려면 두 번의 부/복호화 과정을 거치는 tandem coding이

필요한데, 이때 통화품질의 열화와 함께 많은 계산량이 요구되며 전달 시간이 지연되는 문제가 발생한다. 이러한 문제점들을 해결하기 위하여 두 번의 부/복호화 과정을 한 번의 상호부호화[3] 과정으로 바꾸어 계산량과 전달 시간을 줄이면서도 통화품질을 유지하는 기술이 중요한 문제로 대두되고있다.

본 논문은 CS-ACELP를 기반으로 하는 G.729A에서 RCELP를 기반으로 하는 EVRC로의 상호부호화 알고리즘을 제안하였다.

II. ITU-T G.729A와 IS-127 EVRC

2.1 ITU-T G.729A

ITU-T G.729 CS-ACELP (Conjugate Structure - Algebraic Code Excited Linear Prediction)는 8 kbps의 저전송율을 갖는 멀티미디어 통신용 표준 음성 부호화기이다. G.729는 8 kHz로 샘플링된 16 비트 PCM 음성 신호를 10 ms (80 샘플) 단위의 프레임으로 잘라서 부호화한다. G.729는 부호화 시에 5 ms의 예견구간을 포함하므로 총 15 ms의 알고리즘 지연이 생긴다. 하나의 프레임은 두 개의 서브프레임으로 구성되며 각 서브프레임은 동일하게 5 ms (40 샘플)의 길이를 갖는다. 부호화기에서는 프레임 단위로 10차의 선형예측분석 (Linear Predictive Analysis)을 거쳐 LPC (Linear Prediction Coefficients)를 얻고 안정하고 양자화 과정에서 효과적인 LSP (Line Spectral Pair)[4]로 바꾸어 전송한다. 또한 G.729는 프레임 단위로 개회로 피치

주기 (20~143 샘플)를 추정한 후 서브프레임 단위로 적응 코드북, 고정 코드북 탐색을 거쳐 비트열을 생성한다. 적응 코드북과 고정 코드북의 이득값은 결레 구조를 갖는 두 단계의 코드북을 사용하여 백터 양자화된다. G.729A[5]는 G.729에서 LPC 분석, 적응 코드북 탐색, 고정 코드북 탐색 과정의 계산량을 감소시킨 것이다.

2.2 IS-127 EVRC

IS-127 EVRC (Enhanced Variable Rate Codec)는 국내 및 북미의 CDMA 시스템에서 표준으로 제정하고 있는 음성 부호화기이다. EVRC는 8 kbps의 고정된 비트율을 갖는 G.729와 달리 음성 신호의 특성에 따라 8.55 kbps, 4 kbps, 800 bps의 가변적인 비트율을 갖는다. EVRC는 RCELP를 기반으로 하고 있는데, RCELP는 기존의 CELP 방식과 달리 피치 정보를 서브프레임 단위가 아닌 프레임 단위로 산출하여 피치에 할당되는 비트수를 줄이고 다른 중요 파라미터에 더 많은 비트수를 할당하는 방식을 취한다. EVRC의 한 프레임은 20 ms (160 샘플)로 구성되며 부호화시에 10 ms의 예견구간이 있어서 총 30 ms의 알고리즘 지연을 갖는다. 하나의 프레임은 세 개의 서브프레임으로 구성되며 두 개의 서브프레임은 53 샘플, 하나의 서브프레임은 54 샘플로 이루어진다. 프레임 단위로 10차 LPC 분석을 거쳐 LSP로 변환한 후에 개회로 피치 주기 (20~120 샘플) 추정을 한다. 그 후 서브프레임 단위로 적응 코드북 탐색과 고정 코드북 탐색 과정을 수행하여 비트열을 생성한다.

III. 제안된 상호부호화 알고리즘

본 논문에서 제안하는 G.729A에서 EVRC로의 상호 부호화 알고리즘은 LSP 변환과 피치 주기 변환 과정으로 이루어져 있다. 한 프레임이 10 ms로 이루어지는 G.729A의 비트열을 20 ms의 프레임으로 이루어진 EVRC 비트열로 변환하기 위하여 G.729A의 두개의 프레임을 EVRC 하나의 프레임으로 변환하였다. G.729A의 복호화 과정을 거친 PCM 신호와 함께 G.729A의 복호화 과정에서 추출된 LSP, 폐회로 피치 주기를 EVRC 부호화기에 전송함으로써 LPC 분석 과정과 피치 주기 추정 과정에서의 계산량을 감소시켰다. 제안하는 상호부호화 알고리즘은 그림 1과 같다.

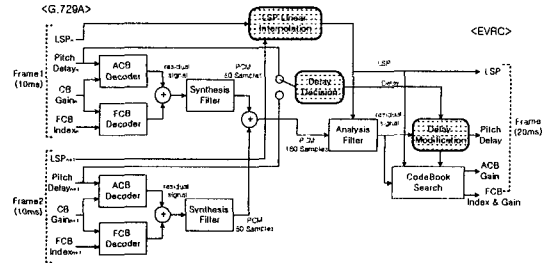


그림 1. G.729A에서 EVRC로의 상호부호화

3.1 LSP 변환

G.729A와 EVRC는 선형예측 부호화의 방법을 사용하고 있다. 각 프레임마다 LPC 분석을 수행하게 되는데 아래 그림2-(a)를 보면 G.729A와 EVRC는 LPC 분석을 위한 샘플 구간의 위치를 다르게 적용하고 있다. G.729A는 LPC 분석을 위하여 현재 프레임에 해당하는 80 샘플과 이전 프레임의 120 샘플 그리고 다음 프레임의 40 샘플, 총 240 샘플을 고려한다. 반면 EVRC는 그림 2-(b)에서처럼 현재 프레임의 절반에 해당하는 80 샘플과 다음 프레임의 80 샘플, 총 160 샘플을 고려하여 LPC 분석을 수행한다. LPC 분석 후 얻어진 10 개의 계수들은 안정하고 양자화에 효과적인 LSP로 변환된다.

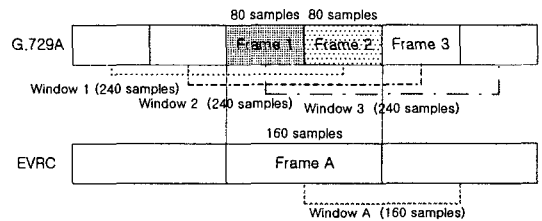


그림 2. G.729A와 EVRC의 LPC 분석을 위한 윈도우

본 논문에서는 G.729A의 두 프레임에서 생성된 LSP를 EVRC의 한 프레임에서 생성되는 LSP와 유사하도록 선형보간법을 이용하여 처리해주었다. 그림 2에서 볼 수 있듯이 EVRC의 Frame A로 변환되어야 할 G.729A의 프레임은 Frame 1과 Frame 2이지만 Frame 3에 해당되는 윈도우가 EVRC의 Frame A에서 LPC 분석을 위한 윈도우 구간과 더 유사함을 볼 수 있다. 따라서 G.729A의 두 프레임인 Frame 1과 Frame 2의 LSP를 선형보간하여 Frame 3에서의 LSP 값과 유사하도록 만들어주고 이 값을 EVRC에 넘겨주어 Frame A에서의 LSP를 대체하였다.

G.729A의 Frame 1과 Frame 2의 LSP를 식 (1)과 같이 선형보간 하였다.

$$0.69LSP_1 + 0.31LSP_2 = LSP_3 \quad (1)$$

식 (1)의 계수들은 프레임이 인접할수록 유사성이 커지는 LSP의 특징에 의하여 결정되었다. 인접한 정도에 따른 평균적인 MSE를 실험을 통해 계산한 후 MSE의 역수에 해당하는 만큼의 가중치를 주어 효과적인 선형보간을 수행할 수 있도록 하였다.

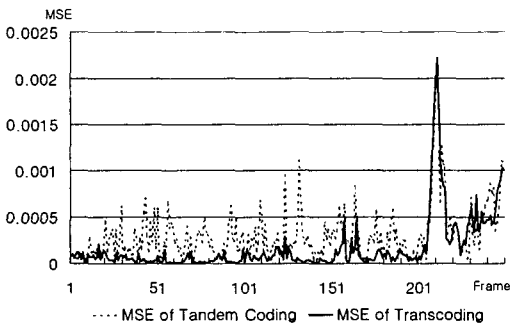


그림 3. tandem coding과 transcoding LSP의 MSE 비교

그림 3에서 점선이 가리키는 것은 tandem coding을 할 때 생성되는 LSP와 EVRC로 부호화할 때 생성되는 LSP 간의 MSE이다. 실선이 가리키는 것은 LSP 선형보간을 하였을 때 생성되는 LSP와 EVRC로 부호화할 때 생성되는 LSP 간의 MSE이다. 제안된 변환을 통하여 얻어진 LSP가 G.729A와 EVRC로 이중 부/복호화한 경우보다 EVRC의 부호화를 통해 얻어진 LSP와 더 유사함을 알 수 있다. 이것은 tandem coding이 두 번의 부/복호화를 거침으로써 LSP도 두 번의 양자화를 거쳐 그 값에 큰 오차가 발생하는 데에 반해 LSP 선형보간을 이용하는 경우에는 한번의 양자화를 거치기 때문이다. LSP 선형보간 알고리즘을 적용함으로써 tandem coding에 비해 음질을 개선하였을 뿐만 아니라 LPC 분석 과정을 생략하여 계산량을 감소시켰다.

3.2 피치 주기 변환

G.729A는 10 ms 프레임 단위로 개회로 분석을 거친 후에 5 ms 서브프레임 단위로 폐회로 분석을 수행하여 한 프레임당 2 개의 피치 주기를 전송한다. EVRC는 RCELP 알고리즘을 기반으로 폐회로 분석 없이 개회로 피치 탐색만으로 결정된 피치 주기를 20 ms 한 프레임 단위로 전송한다. 따라서 5 ms로 이루어진 G.729A 서브프레임의 4 개 피치를 20 ms EVRC 한 프레임의 피치로 변환하는 과정이 필요하다.

표 1은 그림 4 PCM 음성신호 400 ms~600 ms의 시

간영역에 대해 G.729A와 EVRC의 부호기에서 구한 피치주기 값이다. 피치 주기는 유성음 구간 (약 440 ms~560 ms)에서는 프레임 사이에서 완만한 변화를 보이지만 무성음 구간에서는 불규칙한 변화를 보인다. 피치주기는 자기 상관함수의 에너지가 최대가 되는 지연 값을 취하게 되는데, 무성음 구간에서는 성대를 통과하며 발생하는 특정한 피치 주기가 없기 때문에 피치 주기의 심한 변화가 발생하게 된다. 따라서 피치 주기가 변환될 때 무성음 구간보다 유성음 구간의 피치가 좀더 정확히 전달되어야 한다.

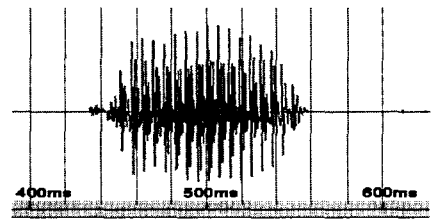


그림 4. PCM 음성신호

표 1. 그림 4에서의 시간별 G.729A와 EVRC의 피치 주기 값

Time (ms)	G.729A Frame 1 Average Delay (SubDelay/SubDelay ₂)	G.729A Frame 2 Average Delay (SubDelay/SubDelay ₂)	EVRC Frame Delay
380~400	106 (107/104)	112 (112/112)	56
400~420	106 (107/104)	112 (112/112)	56
420~440	139 (142/135)	107 (103/108)	55
440~460	125 (125/124)	43 (41/44)	44
460~480	50 (50/49)	49 (49/49)	50
480~500	50 (49/50)	50 (49/50)	49
500~520	49 (49/49)	49 (49/49)	49
520~540	49 (49/49)	49 (49/48)	49
540~560	48 (48/48)	47 (49/44)	48
560~580	54 (52/56)	49 (49/49)	49
580~600	133 (133/130)	133 (132/133)	48
600~620	138 (141/134)	122 (121/122)	29

제안된 피치 주기 변환과정은 무성음과 유성음이 바뀌는 부분에서 적응적으로 피치 주기를 선택하게 된다. 그림 5에서 알 수 있듯이, G.729A 두 프레임의 피치 주기가 EVRC의 한 프레임의 피치 주기로 변환될 때 유성음이 시작되는 구간에서는 G.729A 두 번째 프레임의 피치 주기가 EVRC의 피치 주기로 변환되고, 유성음이 끝나는 구간에서는 G.729A 첫 번째 프레임의 피치 주기가 EVRC의 피치 주기로 변환된다. 결정된 G.729A의 프레임내의 두 서브프레임의 피치 주기를 평균하여 EVRC의 피치 주기로 대체한다.

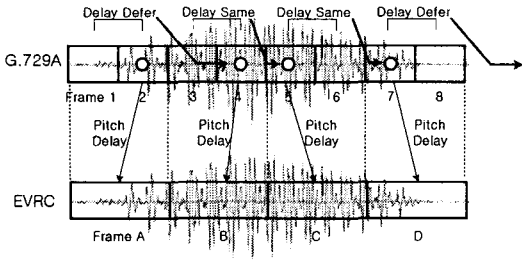


그림 5. 피치 주기 결정

적응적으로 피치 주기를 변환해주기 위해서 이전 EVRC 프레임에 대응하는 G.729A의 두 프레임의 평균 피치 주기의 차이가 5이상일 때 EVRC 현재 프레임의 피치 주기를 G.729A 두 번째의 프레임의 피치 주기로 결정하였고 5미만일 때는 G.729A 첫 번째 프레임의 피치 주기로 결정하였다. 결정된 피치 주기의 ± 3 샘플구간 내에서 자기상관 함수를 취하여 최대값을 갖는 지연값을 피치 주기로 보정한 후 EVRC의 피치 완만화 과정을 거쳐 최종 피치 주기 값을 얻는다.

EVRC 부호화의 개회로 피치 추정 과정에서는 잔차 신호의 자기상관 함수를 최대로 만드는 지연값을 20~120 샘플사이에서 찾아 피치 주기로 정하지만 제안된 피치 변환 알고리즘에서는 G.729A에서 얻어온 피치 주기 주변의 6 샘플사이에서만 자기상관 함수를 최대로 만드는 지연값을 찾도록 하여 계산량을 감소시켰다.

IV. 실험 결과 및 분석

4.1 음질 평가

본 논문에서 제안하는 상호부호화 알고리즘에 의해 생성된 음성신호에 대하여 음질 평가를 하기 위해 성인 남녀 각각 20 명에게 선호도 평가를 실시하였다. 각 음성신호에 대하여 이중 부/복호화와 상호부호화한 결과음성을 무작위 순서로 들려주고 두 가지 중에서 더 음질이 좋다고 생각되는 것을 선택하도록 하였다.

표 2. 선호도 평가

	Tandem Coding	Transcoding	No Preference
Female	30 %	20 %	50 %
Male	25 %	30 %	45 %

조사 결과 표 2와 같이 남녀 모두 이중 부/복호화와 상호부호화에 대하여 음질의 우열을 가리지 못하여 상호부호화가 이중 부/복호화 수준의 음질을 제공하는 것을 확인하였다.

4.2 연산량 측정

제안된 알고리즘을 통해 LPC 분석과정을 생략하였고, 개회로 탐색시간을 단축하였다. 연산량 비교를 위해 펜티엄4 1.7 GHz CPU, 512 M RAM 환경에서 80 초의 PCM 음성신호에 대하여 이중 부/복호화 방법과 제안된 상호부호화 알고리즘을 Visual C++ 컴파일러로 실행을 하며 시간 측정을 하였다. 표 3과 같이 제안된 알고리즘에서 완전 복호화와 부분 부호화를 수행하므로 G.729A 복호화 시간은 tandem coding과 동일하고 EVRC 부호화 과정은 tandem coding에 비해 시간 단축이 가능하다. EVRC 부호화 과정에서 453 ms의 시간을 단축하여 6.8 %의 계산량 감소를 확인하였다.

표 3. 계산량 비교

	G.729A Decoding	EVRC Encoding
Tandem Coding	Same	6656 ms
Transcoding	Same	6203 ms

V. 결론

본 논문에서는 G.729A에서 EVRC로의 상호부호화 알고리즘을 제안하였다. 서로 다른 음성코덱간의 이중 부/복호화에서 발생했던 통화품질 열화와 계산량 증가의 문제점을 제안된 LSP와 피치 주기 변환 알고리즘을 통해 최소화하였다. 음질평가와 계산량 실험결과 동등한 음성품질을 유지하면서도 계산량이 감소됨을 확인하였다.

참고문헌

- [1] TIA/EIA/IS-127 Enhanced variable rate codec, speech service option 3 for wideband spread spectrum digital systems, 1996.
- [2] ITU-T Rec. G.729, "Coding of Speech at 8kbit/s CS-ACELP Speech Coder," 1996.
- [3] K. T. Kim, S. K. Jung, Y. C. Park, Y. S. Choi, and D. H. Youn, "An Efficient Transcoding Algorithm for G.723.1 and EVRC Speech Coders," *proc. of IEEE Vehicular Technology Conference*, vol. 3, pp. 1561-1564, 2001.
- [4] N. Sugamura and N. Farvardin, "Quantizer design in LSP speech analysis-synthesis," *IEEE J.Select. Areas in Commun.*, vol. 6, pp. 432-440, Feb. 1988.
- [5] ITU-T Rec. G.729 Annex A, "Reduced Complexity 8kbit/s CS-ACELP Speech Codec," 1996.