

# 차세대 음성통신 시스템을 위한 상호부호화 알고리즘

이문근, 강홍구\*, 박영철\*\*, 윤대희\*

LG 전자 디지털 미디어 연구소, \*연세대학교 전기전자공학부, \*\*연세대학교 정보기술학부

## A Transcoding Algorithm for the Next Generation Speech Communication System

Moon-Kuen Lee, Hong-Goo Kang\*, Youngcheol Park\*\*, Dae Hee Youn\*

Digital Media Research Lab. LG Electronics, \*Dept. of Electric & Electronic Eng., Yonsei University,

\*\*Division of Information Technology, Yonsei University

E-mail : samalim@mcspl.yonsei.ac.kr

### Abstract

본 논문에서는 비동기식 3세대 이동통신망인 WCDMA의 표준 음성 부호화기인 AMR(Adaptive Multi-Rate)[1]과 VoIP(Voice over Internet Protocol) 응용분야에 최근 널리 활용되고 있는 ITU-T 8kbit/s G.729A[2]의 효율적인 연동을 위한 상호부호화(transcoding) 알고리즘을 제안한다. AMR은 통신 채널 환경에 따라 4.75kbit/s부터 12.2kbit/s까지 가변하여 통화품질을 보장한다. 따라서, 제안된 상호부호화 알고리즘은 순방향 8모드, 역방향 8모드를 합하여 총 16모드를 지원한다.

제안된 알고리즘의 성능 평가를 위해 지연 추정, 연산량 측정과 주/객관적 음질평가를 수행한 결과, 제안된 알고리즘은 기존의 tandem보다 최소 5ms의 짧은 지연, 평균 50.2%의 적은 연산량으로 우수한 음질의 복호화 음성 신호를 제공함을 확인하였다.

### I. 서론

통신 기술 및 기기의 발달과 네트워크 환경이 비약적으로 발전함에 따라 정보화 사회는 급속한 발전과 폭발적인 통신 인구의 증가를 가져왔으며, 다양한 네트워크 환경의 출현으로 인한 서로 다른 네트워크 간 통신의 필요성이 발생하게 되었다. 네트워크 간의 통신을 위해서는 gateway와 같은 중계기에서 서로 다른 네트워크의 연동이 이루어져야 한다. 고유 음성부호화기가 다

른 네트워크의 연동시, 중계기에서는 전송단 음성부호화기의 복호화 과정을 통해 음성 신호를 복호화한 후, 수신단 음성부호화기의 부호화 과정을 통해 재부호화를 실시하는 이중 부/복호화 방식으로 음성신호를 전송하게 되는데, 이를 cross-tandem이라고 한다[3]. 이러한 cross-tandem은 네트워크 고유의 음성부호화기에 의한 양자화 잡음 등으로 인한 왜곡과 과도한 계산량, 그리고 전달 지연 증가 등의 문제가 발생하므로 음성 왜곡을 최소화하는 효율적인 상호부호화기의 역할이 매우 중요하다[3][4].

본 논문에서는 현재 VoIP 서비스에서 널리 사용되고 있는 G.729A와 3세대 이동통신 방식의 하나인 WCDMA에서 표준 음성부호화기로 채택된 AMR의 상호부호화 알고리즘을 제안한다. 제안된 상호부호화 알고리즘은 부호화 파라미터들을 직접적으로 변환시키는 비트열 사상(direct mapping in bitstream domain) 방법에 기반한다. 이후에 사용할 '순방향 상호부호화'는 AMR에서 G.729A로의 상호부호화를, '역방향 상호부호화'는 G.729A에서 AMR로의 상호부호화를 의미하고, tandem은 cross-tandem을 의미한다.

### II. 상호부호화 알고리즘

제안된 알고리즘은 음성부호화기의 프레임 길이를 고려하여 한 프레임 길이가 10ms인 G.729A의 두 프레임과 20ms인 AMR의 한 프레임이 상호부호화된다.

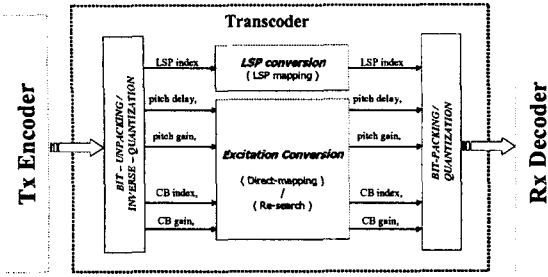


그림 1. 제안된 상호부호화 알고리즘의 블럭도

그림 1은 제안된 상호부호화 알고리즘의 블럭도로써 크게 LSP(Line Spectrum Pairs) 파라미터 변환 과정과 여기신호 변환 과정으로 구성된다.

### 2.1 LSP 상호부호화

상호부호화기 설계시 LSP 변환에서 가장 중요하게 고려되어야 할 요소는 할당된 비트 수 차이에 따른 두 부호화기 간의 양자화 거리차이다. 이러한 문제는 LSP 파라미터의 상호부호화를 합성 영역이 아닌 LSP 영역에서 실시함으로써 줄일 수 있다. 제안된 알고리즘에서는 두 음성부호화기의 대응 프레임간의 LSP 거리를 고려하여 다음의 3 가지 방법을 통하여 상호부호화를 실시하였다.

#### 2.1.1 직접 변환 (Direct-mapping)

G729A 와 AMR 12.2kbit/s 는 LPC 분석을 위해 필요한 버퍼 길이와 사용되는 윈도우가 모두 동일하며, LSP 벡터 양자화 간격도 동일하므로 LSP 상호부호화는 직접 변환 방법을 사용한다.

#### 2.1.2 선형 보간 (Linear Interpolation)

AMR 10.2 kbit/s 이하 순방향 상호부호화에서는 AMR 의 한 프레임에 해당하는 LSP 정보를 식 (1)과 같이 선형 보간법을 이용하여 G729A 의 두 프레임에 해당하는 LSP 정보로 변환한다. 여기서,  $\omega_{1,2}^{(G)}$ ,  $\omega_{1,4}^{(G)}$  는 각각 G729A 의 첫번째와 두번째 프레임의 LSP 를 나타내고,  $\omega_{k-1,4}^{(A)}$ ,  $\omega_{k,4}^{(A)}$  는 각각 AMR 의 과거 프레임과 현재 프레임의 LSP 를 나타낸다.

$$\begin{cases} \omega_{1,2}^{(G)} = 1/2 (\omega_{k-1,4}^{(A)} + \omega_{k,4}^{(A)}) \\ \omega_{1,4}^{(G)} = \omega_{k,4}^{(A)} \end{cases} \quad (1)$$

#### 2.1.3 선형 보간을 고려한 변형된 LSP 상호부호화

역방향 상호부호화에서 G729A 첫번째 부프레임

LSP 를 고려하여 변환을 수행하는 방법이다. 이러한 방법은 [3]에서 사용한 방법이다. 그러나, 본 논문에서는 각 부프레임의 스펙트럼 왜곡도에 대한 평균과 분산 측정 실험을 통하여, 선형 보간된 부프레임의 LSP 양자화 오차는 그 부프레임의 선형 보간에 사용되는 각 프레임의 LSP 양자화 오차에 의존하며, 선형 보간된 부프레임의 LSP 양자화 오차의 동적 범위는 선형 보간에 사용되는 LSP 양자화 오차의 동적 범위 이하임을 확인하였다. 이와 같은 사실에 기인하여 오차의 동적 범위가 큰 G729A 의 2, 4 번째 부프레임의 LSP 를 고려하여 식 (2) 를 최소화 시키는 유도식 (3)으로부터 AMR 의 LSP 계수를 추정하는 변형된 LSP 상호부호화를 제안한다.

$$\epsilon = \sum_{s=1}^4 W_s \left\| \omega_{1,s}^{(G)} - \omega_{k,s}^{(A)} \right\|^2 \quad (2)$$

$$\bar{\omega}_{k,s}^{(A)} = (4W_4 \omega_{1,4}^{(G)} + 2W_2 \omega_{1,2}^{(G)} - W_2 \omega_{k-1,4}^{(A)}) / (4W_4 + W_2) \quad (3)$$

위 식에서,  $\bar{\omega}_{k,s}^{(A)}$  는 AMR 의  $k$  번째 프레임의  $s$  번째 부프레임 LSP 를,  $\omega_{1,s}^{(G)}$  는 이와 상응하는 G729A 2 개 프레임의  $s$  번째 부프레임의 LSP 를 나타내고,  $W_s$  는 가중치를 나타낸다. 제안된 알고리즘에서는 객관적, 주관적 음질 평가를 통하여  $W_4 = 0.61$  로 고정하였고, 그에 따라  $W_2 = 1 - W_4$  로 고정하였다.

LPC 스펙트럼 비교를 통해, 제안된 방법의 스펙트럼이 tandem 의 경우보다 송신단 부호화기의 스펙트럼과 더 유사함을 확인하였고, SD(Spectral Distance) 측정을 통해 제안된 방법이 우수한 성능을 보임을 확인하였다.

### 2.2 이득 정보 상호부호화

이득 정보를 상호부호화하기 위해서는 이득 영역, 여기신호 영역, 합성 영역에서의 변환 방법을 이용할 수 있다. 본 논문에서는 고정코드북 인덱스 상호부호화에 따라 계산량과 성능을 고려하여 여기신호 영역에서의 변환 방법(Excitation-Domain mapping)을 사용하였다.

$$C_{exc} = \underset{0 \leq i \leq 5, 0 \leq j \leq 6}{\operatorname{argmin}} \left\| x_e - (CB_{g,p,i}^n + CB_{g,p,j}^n) \cdot y_e - [(CB_{g,r}^n + CB_{g,i}^n) \cdot g_e] \cdot z_e \right\| \quad (4)$$

식(4)와 같이 송신단 부호화기의 적응 코드북 벡터와 고정 코드북 벡터를 복호화하여 여기신호를 생성한 후, 수신단 이득 양자화 테이블을 이용하여 여기신호 영역에서의 거리차가 최소화되는 테이블 인덱스를 찾는 방법이다. 이 방법은 [3]에서 적용한 방법과 유사하다.

### 2.3 적응 코드북 지연 상호부호화

지연을 위한 양자화 영역이 서로 수용 가능한 범위 인 경우는 직접 변환이 가능하다. 즉, 역방향의 상호부 호화인 경우, 버퍼를 사용하여 직접 변환된 적응 코드 북 지연을 AMR 에 적합한 비트열로 생성, 복호화기로 전송한다. 그러나, 순방향 상호부호화(7.95kbit/s)에서와 같이 양자화 범위가 다른 경우에는 직접변환 시 음질 저하가 발생한다. 이러한 현상은 특히 무성음이나 천이 구간에서 심하게 발생된다. 이런 경우, 적응 코드북 지연을 halving 시키는 방법과, 이전 프레임의 적응 코드북 지연에 따라 완만화 시키는 방법을 적용한다.

제한된 비트수로 인해 양자화 알고리즘이 서로 다 른 경우는 직접변환이 불가능하므로 다음과 같은 개회 로 피치 추정 방법을 이용한다. 상호부호화를 위한 해 당 프레임의 G.729A 적응 코드북 지연 정보를 이용해 지각 가중된 음성 영역에서 개회로 피치를 추정한다. 즉, 현재 프레임에 해당되는 G.729A 의 지연을 AMR 의 개회로 피치로 대체시켜 재추정하여 양자화한다.

2.4 고정 코드북 인덱스 상호부호화

두 부호화기는 고정 코드북이 대수 코드북 구조를 갖는 ACELP 모델에 기초를 두고 있으며, 고정 코드북 탐색 알고리즘과 파라미터 추출 방법이 유사하다.

AMR 7.95/7.4kbit/s 는 G.729A 와 펄스 탐색을 위한 트랙의 구조와 트랙당 찾는 펄스의 개수가 동일하다. 따라서, 음질에 큰 왜곡없이 직접변환이 가능하다.

비트수의 제한으로 인해 트랙 구조와 펄스의 개수 가 다른 경우는 직접 변환을 불가능하므로, 고정 코드 북 인덱스를 재검색하는 과정이 필요하다.

III. 성능 평가

제안된 알고리즘의 성능 평가를 위하여 객관적, 주 관적 음질 평가 및 계산량 측정을 수행하여 기존의 tandem 과 성능을 비교하였다. 성능평가는 한국인 남성/ 여성 화자 각각 4 명에 대한 8kHz 로 표본화된 8 초 길 이의 96 개 문장에 대하여 수행하였다. 이후 표에서 'TR'은 상호부호화를 'TD'는 tandem 을 의미한다.

3.1 지연 시간

표 1은 두 알고리즘의 지연시간을 추정하여 비교한 것이다. 표에서  $p_m^E, p_m^D$  ( $m = A \text{ or } G$ )는 각 시스템의 부

표 1. 지연시간 추정 비교

	Total Delay(ms)	
	TD	TR
Forward	$30 + p_A^E + p_A^D + 2(p_G^E + p_G^D)$	$25 + p_A^E + p_{AG}^{TR} + 2p_G^D$
Backward	$30 + p_A^E + p_A^D + 2(p_G^E + p_G^D)$	$25 + p_G^E + p_{GA}^{TR} + 2p_A^D$

표 2. 제안된 알고리즘의 Average PESQ

	TD		TR	
	Forward	Backward	Forward	Backward
12.2k	3.509	3.445	3.486	3.511
10.2k	3.485	3.414	3.456	3.446
7.95k	3.396	3.325	3.453	3.467
7.4 k	3.391	3.312	3.522	3.412
6.7 k	3.306	3.268	3.396	3.300
5.9 k	3.277	3.205	3.350	3.218
5.15k	3.194	3.162	3.273	3.176
4.75k	3.147	3.134	3.223	3.141

호화기와 복호화기의 처리시간을 나타내며,  $p_{AG}^{TR}, p_{GA}^{TR}$  는 순방향과 역방향 상호부호화기의 처리시간을 나타낸 다. 제안된 상호부호화 알고리즘은 재 부호화를 위한 5ms 의 지연(look-ahead)시간이 필요 없기 때문에 기존의 방법보다 최소한 5ms 의 지연시간을 단축함을 알 수 있 다. 또한, 처리시간에 있어서  $p_{AG}^{TR}$  는  $p_A^D$  와  $p_G^E$  의 합보 다 작기 때문에 전체적인 지연시간의 이득은 실제로 5ms 보다 더 단축될 수 있다.

3.2 객관적 음질평가

ITU-T 에서 권고한 PESQ (Perceptual Evaluation of Speech Quality )를 측정하였다[5]. 표 2는 남성/여성 화자 에 대한 PESQ 측정 후, 그 결과를 평균한 것이다.

AMR 12.2/10.2 kbit/s 은 단일 음성부호화 성능이 우 수하다. 따라서, 복호화된 음성 신호와 원 음성과의 오 차가 크지 않기 때문에, tandem 을 사용한 경우에도 성 능이 우수함을 확인할 수 있다. AMR 12.2/10.2 kbit/s 를 제외한 남녀 화자 모든 경우에 제안된 알고리즘의 객관 적 성능이 우수한 것을 알 수 있다. 제안된 알고리즘에 서는 LSP 영역에서 변환된 LSP 정보를 이용해 부호화 파라미터들을 모델링하므로 실제로 상호부호화된 여기 신호는 전송부호화기의 여기신호와 서로 유사한 특성을 지니게 됨에 따라 얻어진 결과라고 생각된다.

순방향 상호부호화에서 7.95 kbit/s 가 7.4 kbit/s 보다 성능이 저하되는 이유는 AMR 의 적응 코드북 지연의 동적 범위를 G.729A 가 수용하지 못하기 때문에 발생한 것이다. 그러나, 기존의 tandem 보다는 우수한 성능을보 인다. AMR 5.9 kbit/s 의 경우 새로 검색한 고정 코드북

표 3. 제안된 알고리즘의 선호도 평가

	Forward(%)			Backward(%)		
	TD	TR	X	TD	TR	X
12.2k	24.4	33.3	42.3	25.0	37.5	37.5
10.2k	25.0	36.9	38.1	27.4	40.5	31.5
7.95k	18.7	34.0	47.4	21.1	27.1	51.8
7.4 k	24.6	41.2	34.2	21.7	23.0	55.3
6.7 k	25.5	33.3	41.2	22.0	27.1	50.9
5.9 k	23.7	33.3	43.0	24.6	34.2	41.2
5.15k	28.6	38.1	33.3	20.2	35.9	43.9
4.75k	23.0	37.5	39.5	23.7	42.9	33.4

표 4. 제안된 알고리즘의 Average WMOPS

	Forward			Backward		
	TD	TR	감소(%)	TD	TR	감소(%)
12.2k	9.867	6.966	29.40	13.449	8.230	38.81
10.2k	9.824	5.779	41.17	13.739	7.964	42.03
7.95k	9.860	2.805	71.55	13.788	3.188	76.88
7.4 k	9.620	2.749	71.42	13.257	3.165	76.13
6.7 k	9.863	5.720	42.01	14.021	9.005	35.77
5.9 k	9.844	5.706	42.04	11.708	6.692	42.84
5.15k	9.842	5.703	42.06	10.395	7.105	32.52
4.75k	9.849	5.712	42.01	12.330	8.952	27.38

펄스 수가 2 개 밖에 되지 않기 때문에, G.729A 여기신호의 특성을 따라가지 못하는 경우가 발생하게 된다. 따라서 상호부호화 경우에도 적응 코드북 지연, 고정 코드북 펄스와 이득을 재추정하는 것이 효과적이다. 순방향 상호부호화에서는 모든 경우에 있어서 제안된 알고리즘의 객관적인 성능이 우수한 것을 확인할 수 있다. 그 이유 중 하나는 G.729A 적응 코드북 지연의 동적 범위가 AMR 범위에 비해 좁기 때문에 직접 변환방법에 의한 성능을 보장할 수 있기 때문이다.

3.3 주관적 음질평가

두 알고리즘에 의해 합성된 음성의 선호도 평가 (preference test)를 수행하였다. 30 명의 비전문가 청취자에게 8kHz 로 표본화된 8 초 길이의 남/여 각각 4 명 화자의 문장에 대하여 두 방법에 의해 최종 복호화된 신호를 랜덤한 순서로 2 번씩 들려주고 음질의 우열을 가리도록 하였다. 표 3은 남성/여성 화자에 대한 선호도 평가를 실시하여 그 결과를 평균한 것이다. 제안된 방법은 기존의 tandem 보다 평균적으로 우수한 선호도를 보인다. 그러나, 남성화자의 경우 일부 비트율에서는 tandem 의 선호도가 높았다. 이는 순방향 하위 2 개 비트율에서 고정 코드북 펄스 수가 2 개 밖에 되지 않기 때문에, G.729A 여기신호의 특성을 따라가지 못하는 경우가 발생하게 되기 때문이다.

3.4 계산량 측정

제안된 알고리즘과 기존의 tandem 에 대한 계산량 비교를 위하여 WMOPS(Weighted Million Operations Per Second)를 측정하였다[6]. 표 4는 그 결과를 보여준다. 제안된 알고리즘은 기존 tandem 보다 비트율에 따라 순방향 상호부호화의 경우, 최대 71.55%, 최소 29.40%의 평균 계산량 감소율을 보이며, 역방향의 경우, 최대 76.88%, 최소 27.38%의 평균 계산량 감소율을 보인다.

V. 결론

본 논문에서는 AMR 과 G.729A 의 시스템 연동을 위한 상호부호화 알고리즘을 제안하였다. 제안된 알고리즘의 성능 평가를 위해, 실시간 통신을 위한 지연 추정, 연산량 추정을 위한 WMOPS 측정, 주관적 음질 평가를 위한 청취자 선호도 평가와 객관적 음질 평가를 위한 PESQ 측정을 수행하였다. 그 결과, 제안된 알고리즘은 기존의 tandem 방식보다 최소 5ms 의 짧은 지연과 평균 50.2%의 적은 연산량으로 우수한 음질의 복호화 음성신호를 제공함을 확인하였다.

참고문헌

- [1] Mandatory Speech Codec speech processing functions AMR Speech Codec; General Description(3G TS 26.071 version 3.0.1 Release 1999)
- [2] ITU-T Rec. G.729 Annex A "Reduced Complexity 8 kbit/s CS-ACELP Speech Codec," 1996.
- [3] Hong-Goo Kang, Hong Kook Kim, Cox, R.V. "Improving Transcoding Capability of Speech Coders," submitted to *IEEE Transaction, Multimedia*, Mar. 2003.
- [4] K. T Kim, S. K Jung, Y.C Park, Y. S Choi, D. H Youn, "An Efficient Transcoding Algorithm for G.723.1 and EVRC Speech Coders," in Proc. *IEEE VTS 54th(VTC 2001)*, vol.3, pp.1561-1564, Oct.7-10, 2001.
- [5] ITU-T Draft Rec P.862 "Perceptual evaluation of speech quality (PESQ), an objective method of end-to-end speech quality assessment of narrowband telephone networks and speech codecs," May. 2000.
- [6] ITU-T Draft Rec P.191 "Software tools for speech and audio coding standardization," Nov. 2000.