

G.729A 와 SMV 음성부호화기를 위한 파라미터 직접 변환 방식의 상호부호화 알고리즘

장달원, 서성호, 이선일, 유창동
한국 과학 기술원 전자전산학과

A Novel Transcoding Algorithm for G.729A and SMV Speech Codec via Direct Parameter Conversion

Dalwon Jang, Seongho Seo, Sunil Lee, Chang D. Yoo
Department of Electrical Engineering and Computer Science,
Korea Advanced Institute of Science and Technology
E-mail : mcjang@eeinfo.kaist.ac.kr

Abstract

본 논문에서는 G.729A 와 SMV 음성 부호화기를 위한 새로운 파라미터 직접 변환 방식의 상호 부호화 알고리즘을 제안한다. 상호 부호화를 위하여 부가적인 복호화, 부호화 과정을 거쳐야하는 기존의 tandem 방식과 달리 제안된 파라미터 직접 변환 방식에서는 양 음성부호화기에서 공통적으로 사용하는 파라미터들이 직접 변환된다. SMV 에서 G.729A 로의 상호 부호화에서는 LSP 변환, 피치 지연 변환, 낮은 전송률에서의 상호 부호화 등의 알고리즘을 제안하고, G.729A 에서 SMV 로의 상호 부호화에서는 LSP 변환, 피치 지연 변환, 전송률 결정 등의 알고리즘을 제안한다. 제안된 알고리즘을 다양한 방법으로 평가해본 결과 기존의 tandem 방식과 비교하여 계산량과 지연 시간을 줄이면서도 동등한 음질 또는 향상된 음질을 구현함을 확인할 수 있었다.

이 과정을 상호 부호화라고 한다. 상호 부호화를 하는 가장 간단한 방법은 원천 음성 부호화기의 복호화기와 목적 음성 부호화기의 부호화기를 단순히 일렬로 연결하는 tandem 방식이다. Tandem 방식의 상호 부호화는 반복적인 부호화 과정을 거치게 되므로 많은 계산량, 긴 지연 시간, 음질 저하, 세 가지의 문제를 가진다. 상호 부호화의 대상이 두 음성 부호화기의 특성을 이용해서 파라미터 레벨에서 변환을 할 경우, 이 문제를 어느 정도 해결할 수 있다. 물론 파라미터 레벨에서 변환을 하려면 양 음성 부호화기가 같은 종류의 파라미터들을 사용하는 경우에만 가능하다. 본 논문에서 다루고 있는 G.729A[1]와 SMV[2]는 둘 다 기본적으로 CELP[3] 알고리즘을 기반으로 만들어졌다.

본 논문은 다음과 같이 구성되어 있다. 먼저 II 장에서는 두 음성부호화에 대하여 간략히 살펴본다. 그리고 III장과 IV장에서는 양방향 상호 부호화 알고리즘을 설명한다. V장에서는 제안된 알고리즘을 평가한 결과를 제시하며, 마지막으로 VI장에서는 이 논문에 대한 결론을 맺는다.

I. 서론

현대에는 여러 종류의 음성 부호화기가 다양한 통신망에서 사용되고 있다. 서로 다른 음성 부호화 표준을 사용하는 통신망끼리 음성 통화를 하려면 송신단에서 비트열로 부호화해서 보낸 신호를 다른 음성 부호화기의 포맷에 맞는 비트열로 변환하는 과정이 필요하다.

II. G.729A 와 SMV 음성 부호화기

G.729A 는 Conjugate Structure Algebraic Code Excited Linear Prediction(CS-ACELP)를 기반으로 하는 ITU-T 의

표준 음성부호화기이다. G.729A 음성부호화기는 G.729 음성부호화기를 계산량을 감소시키는 알고리즘을 적용하여 개선한 것이다.

G.729A의 프레임 길이는 10ms 이며 한 개의 프레임은 5ms 길이의 두 개의 부프레임을 가지고 있다. G.729A 음성 부호화기의 입력은 8kHz 로 샘플링되어 16bit 로 양자화된 PCM 신호이다. 입력신호는 전처리 과정을 통해서 크기가 반으로 줄어들고 DC 성분이 제거된다. 전처리 과정을 거친 입력음성은 선형 예측 분석을 거치게 되고, 각 프레임마다 10 차의 선형 예측 계수(Linear-Prediction Coefficient, LPC)가 구해진다. 선형 예측 분석 과정에서는 30ms 비대칭 윈도우를 입력 음성에 씌우는데, 이 때, 5ms 의 예견 구간을 필요로 한다. 선형 예측 계수는 양자화 에러 및 전송 과정에서의 에러를 줄이기 위해서 LSP 의 형태로 변환되어 전송된다. 선형 예측 분석 후에는 여기 신호를 구하게 된다. 여기 신호는 적응 코드북 벡터(피치 지연)와 고정 코드북 벡터에 각 코드북의 이득을 곱하여 더한 형태로 모델링된다. 피치 지연은 개회로 피치 검색과 폐회로 피치 검색 과정을 통해서 각 부프레임마다 하나씩 구한다. 개회로 검색에서 피치 지연의 범위를 정하고, 폐회로 피치 검색 과정에서 정확한 피치 지연을 찾는다. 고정 코드북은 Interleaved Single-Pulse Permutation(ISPP) 구조에 기반하고 각 부프레임당 4 개의 펄스를 가진다.

SMV 는 extended Code Excited Linear Prediction (eX-CELP)을 기반으로 만들어졌다. SMV 는 8.55, 4.0, 2.0, 0.8kbps, 총 네 개의 전송률을 가지고 있으며, 프레임 단위로 전송률을 결정해서 부호화한다. 그리고 8.55kbps 와 4.0kbps 의 전송률의 경우 type 0 와 type 1, 두 가지의 type 을 가지고 있다. 전송률과 type 을 결정하는데 중요한 요소로 프레임 클래스(frame class)가 있다. SMV 는 각 프레임을 입력 음성의 특성에 따라서, 묵음(silence), 잡음(noise-like), 무성음(unvoiced), 변화(onset), 일정하지 않은 유성음(non-stationary voiced), 일정한 유성음(stationary voiced), 총 6 가지의 프레임 클래스로 분류한다. 프레임 클래스에 따라서 결정되는 전송률이 달라지고, 일정한 유성음으로 분류되었을 경우, type 1 로 결정되어서 type 0 과는 다른 부호화 과정을 거치게 된다. SMV 는 mode 0,1,2,3, 총 4 개의 동작모드를 가진다. 이것은 외부의 통신 환경에 의해서 결정되는 것이다. 모드 번호가 낮을수록 높은 전송률을 많이 사용해서 좋은 품질의 음성을 전송한다.

SMV 에서 사용하는 프레임의 길이는 20ms 이고 부프레임의 개수와 그 길이는 프레임의 전송률에 따라서 달라지게 된다. 선형 예측 분석 과정에서는 10ms 의 예견 구간을 가지는 윈도우를 사용하고 10 차의 선형 예측 계수를 구해서 LSF 로 변환해서 양자화한다. 전송률과 type 결정 후 여기 신호에 대한 부호화를 한다.

III. SMV 에서 G.729A 로의 상호 부호화

3.1 전체적인 구조

그림 1 은 제안된 파라미터 직접 변환 방식의 상호 부호화 알고리즘의 구조도를 나타낸 것이다. 전송률에 따라서 다르기는 하지만, SMV 는 CELP 의 구조를 따르고 있기 때문에 SMV 의 신호를 파라미터로 변환하면 LSP 에 관련된 값들과 여기 신호에 관련된 값들이 얻어진다. LSP 는 직접적으로 변환 가능하고 여기 신호는 SMV 의 파라미터들을 복호화하여 생성한 음성신호 - 후처리 과정을 거치지 않기 때문에 거친 음성이고 이 블록 다이어그램에서는 target signal 로 표시 □ 를 분석해서 찾는 과정을 거친다. 물론 이 분석 과정에서 SMV 의 파라미터들을 이용해서 효율성을 높인다. 이렇게 얻어진 G.729A 의 파라미터는 비트열로 변환 후 전송된다.

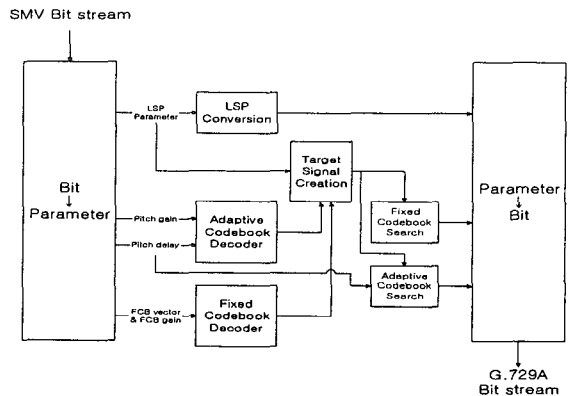


그림 1 SMV → G.729A 상호 부호화의 블록도

3.2 LSP 변환

SMV 와 G.729A 에서 선형 예측 계수는 똑같이 10 차의 계수를 가지고, Levinson-Durbin 알고리즘으로 계산하지만 입력 음성에 적용하는 윈도우는 다르다. 이 윈도우를 보상해 줄 수 있는 선형 보간법으로 LSP 변환

을 할 수 있다. 그림 2 는 SMV 와 G.729A 의 선형 예측 분석에서 사용하는 윈도우를 프레임과 같이 그린 것이다. 그림에서와 같이 SMV 의 k 번째, $(k-1)$ 번째 프레임의 LSP 계수를 선형 보간해서 G.729A 의 $(2k-1)$ 번째 프레임의 LSP 계수를 만들고, SMV k 번째 프레임의 LSP 계수는 G.729A 의 $(2k)$ 번째 프레임의 LSP 계수로 그대로 사용된다. 이 것을 수식으로 나타내면 아래와 같다.

$$LSP_{2k-1}^{G.729A} = \frac{1}{2}(LSP_{k-1}^{SMV} + LSP_k^{SMV}) \quad (1)$$

$$LSP_{2k}^{G.729A} = LSP_k^{SMV} \quad (2)$$

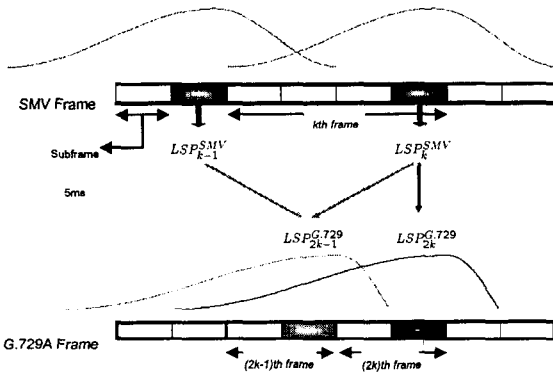


그림 2 SMV 에서 G.729A 로의 LSP 변환

3.3 피치 지연 변환

피치 지연은 음성의 유성음적 특성을 나타낸 값으로 유성음의 주기를 잰 값이다. 당연히 이 값이 두 음성 부호화기에서 비슷한 경향을 보이고 실험으로도 확인할 수 있다. 그래서 SMV 비트열에서 나온 피치 지연을 그대로 사용해 보았으나 결과음성의 음질이 상당히 떨어지는 것을 확인할 수 있었기에, SMV 의 피치 지연을 그대로 사용할 수 없다. 하지만, 두 음성부호화기에서 나오는 피치 지연이 상당히 비슷하다는 점을 최대한 이용하여 SMV 의 피치 지연 부근에서 피치 지연의 패치로 검색을 하여 피치 지연을 구하는 알고리즘을 사용하였다. 즉 SMV 의 피치 지연이 G.729A 에서 피치 지연을 찾기 위해서 수행하는 개회로 검색의 결과를 대신한다고 생각할 수 있다. 이 방법으로 개회로 피치 검색 과정을 제거하여 개회로 피치 검색 과정만큼의 계

산량의 감소를 가져올 수 있다.

3.3 0.8kbps 의 신호 변환

SMV 에서 0.8kbps 로 부호화되어 전송되는 신호는 간단한 방법으로 상호 부호화가 가능하다. 이 신호는 SMV 내에게 목음이라고 인정한 프레임이며, LSP 와 에너지 정보만을 전송한다. 목음이라는 성질을 이용해서 LSP 변환과 이득 설정만으로 간단한 상호 부호화가 가능하다. 목음이기에, 신호의 크기가 작고 따라서 이득도 작다. G.729A 에서는 이득을 7 비트로 벡터 양자화하고, 그에 따라서 이득이 가질 수 있는 값은 128 개뿐이다. 그런데 목음이 입력으로 들어갔을 때는 특정 이득 한 가지만 선택되었다. 상호 부호화기에서 이것을 이용해서 목음일 때는 항상 그 이득으로 설정해주었다. 고정 코드북 벡터나 피치 지연을 변환하지 않아도 이득설정만으로 상호 부호화가 이루어졌다.

IV. G.729A 에서 SMV 로의 상호 부호화

그림 3 에 나타나 있는 G.729A 에서 SMV 로의 상호 부호화는 그 반대의 경우와 매우 유사하다. 하지만, 한 가지 크게 다른 점이 여기서는 전송률과 type 을 결정하는 과정이 있다는 것이다. G.729A 의 파라미터들을 이용해서 간단히 결정하는 알고리즘을 사용한다.

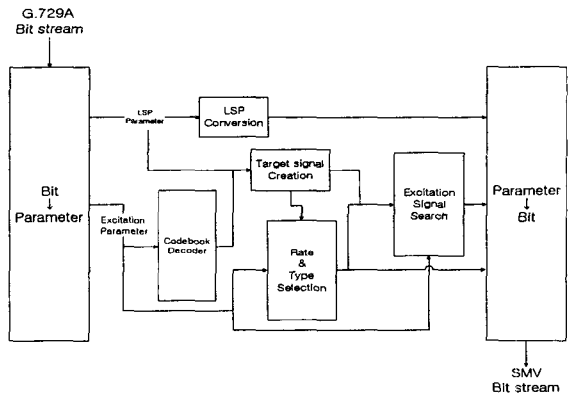


그림 3 G.729A → SMV 상호 부호화의 블록도

SMV 에서 G.729A 로의 상호 부호화와 마찬가지로 LSP 는 선형 보간을 이용해서 변환한다.

$$LSP_k^{SMV} = \frac{1}{2}(LSP_{2k-1}^{G.729A} + LSP_{2k}^{G.729A}) \quad (3)$$

피치 지연 변환은 피치 지연이 개회로 검색의 결과를

대신하도록 한다. 역시 SMV 에서 G.729A 로의 상호 부호화와 마찬가지로이다.

전송률과 type 결정은 CELP 파라미터를 이용해서 프레임을 간단히 분류하는 방법을 사용했다. 알고리즘의 단순성을 위해 프레임 클래스의 개수를 5 개로 줄였고, 그림 4 와 같이 분류를 하였다.

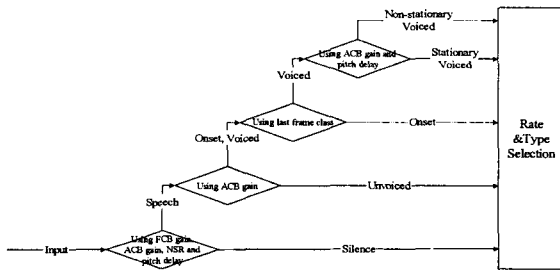


그림 4 제안된 전송률 결정 과정의 블럭도

V. 성능 평가 결과

본 절에서는 제안된 알고리즘의 성능을 계산량, 지연 시간, 음질의 세 가지 지표를 통해서 분석한다. 계산량은 Weighted Million Operations Per Second (WMOPS)를 측정해서 제시하고, 지연 시간은 알고리즘적인 지연 시간만을 살펴본다. 그리고 음질은 PESQ[4] 점수를 측정해서 제시한다.

표 1, 2 에서는 tandem 방식과 비교해서 제안된 알고리즘이 계산량이 감소되었다는 것을 보여준다. SMV 에서 G.729A 로의 상호 부호화에서는 약 45%, G.729A 에서 SMV 로의 상호부호화에서는 약 20% 정도의 계산량 감소가 있었다.

선형 예측 분석 과정에서 예견 구간을 필요로 하는데, 이것이 사라졌으므로 지연 시간이 준다. 지연 시간은 양방향에 대해서 5ms 의 감소가 있었다.

표 3, 4 에 tandem 방식과 제안된 알고리즘의 객관적 음질 평가를 제시하였다. G.729A 에서 SMV 로의 상호 부호화에서는 tandem 방식과 비슷한 음질을 보여주지만, SMV 에서 G.729A 로의 상호 부호화에서는 음질의 향상이 있었다.

표 1. 계산량 비교(SMV → G.729A)

Mode	남성 음성		여성 음성	
	Tandem	Proposed	Tandem	Proposed
0	10.25	5.19	10.32	5.34
1	10.47	5.56	10.71	5.90
2	10.79	5.87	10.95	6.13

3	1.092	6.01	10.99	6.18
---	-------	------	-------	------

표 2. 계산량 비교(G.729A → SMV)

Mode	남성 음성		여성 음성	
	Tandem	Proposed	Tandem	Proposed
0	27.69	23.19	29.24	24.29
1	25.90	20.26	27.06	21.40
2	25.55	20.02	26.80	21.16
3	25.51	20.00	26.82	21.16

표 3. 객관적 음질 평가 결과(SMV → G.729A)

Mode	남성 음성		여성 음성	
	Tandem	Proposed	Tandem	Proposed
0	3.461	3.554	3.135	3.210
1	3.303	3.475	3.019	3.168
2	3.219	3.423	2.991	3.119
3	3.188	3.367	2.995	3.108

표 4. 객관적 음질 평가 결과(G.729A → SMV)

Mode	남성 음성		여성 음성	
	Tandem	Proposed	Tandem	Proposed
0	3.498	3.498	3.233	3.227
1	3.445	3.395	3.178	3.112
2	3.375	3.343	3.136	3.057
3	3.368	3.342	3.123	3.049

VI. 결론

본 논문은 G.729A 와 SMV 음성부호화기를 위한 파라미터 직접 변환 방식의 상호부호화 알고리즘을 제안하였다. 성능 평가 결과, 제안된 알고리즘은 tandem 방식에 비해서 적은 계산량과 짧은 지연 시간으로도 동등한 수준의 음질을 보여주고 있다.

참고문헌

- [1] ITU-T Rec. G.729 Annex A, "Reduced Complexity 8 kbit/s CS-ACELP Speech Codec", 1996
- [2] 3GPP2 C.S0030-0 V2.0, Selectable Mode Vocoder Service Option for Wideband Spread Spectrum Communication System, 2001
- [3] Manfred R. Schroeder, Bishnu S. Atal, "Code Excited Linear Prediction(CELP): High-Quality Speech at Very Low Bit Rates," Proc. Of ICASSP, pp. 937-940, 1985
- [4] 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.