

저전송율 보코더의 성능개선에 관한 연구

*박영호, 배명진
숭실대학교 정보통신공학과

On Improving the Performance of Low Bit-Rate Speech Coder

*Young-Ho Park, Myung-Jin Bae
Dept. of Telecommunication Engr. Soongsil Univ.
E-mail: yhpark@assp.soongsil.ac.kr, mjbae@saint.soongsil.ac.kr

요 약

본 논문은 5.6kbps의 전송율에서 fixed codebook으로 ISPP(60.5.2)의 dynamic sparse algebraic codebook을 이용한 ACELP 알고리즘을 제안한다. 저전송율에서 음질에 중대한 영향을 미치는 대수적 방식의 고정코드북이 가지는 문제점을 최소화하여 음질의 증진을 꾀하였다. 또한 추가 계산량이 필요없는 U/V 분리기를 도입하여 LSF 보간시 발생하는 천어구간에서의 지연을 최소화하였다. 구현된 5.6kbps ACELP는 전화선상의 음질을 시료로하여 주관적 음질면에서 6.3kbps MP-MLQ와 동등하였으며 MNRU 15dB에서 약간 낮았다.

1. 서 론

최근에 디지털 이동통신 및 개인통신망의 발달로 증가하는 이동통신 가입자를 보다 많이 수용하고 서비스의 질을 높이기 위해 음성부호화기에 대해 많은 연구가 진행되고 있다. 이 중 가입자의 수용율을 증가시키는데 있어서 효과적인 방법으로는 부호화기의 전송율을 낮추는 방법이 있다. 이론적으로 보코더의 전송율이 1/2로 낮아질 경우 약 2배의 가입자를 수용할 수 있다고 알려져 있다. 그런 목적으로 현재 600bps대까지 보코더의 전송율이 낮아지고 있다. 그러나 전송율과 그에 따른 음질의 저하는 가입자로 하여금 통신 서비스 질에 대한 불만을 초래한다. 따라서 전송율과 음질이라는 두 지표를 적정수준으로 맞추는데 있어 이동통신사업자의 어려움이 예상된다.

현재 국내에서 디지털 셀룰라 서비스에 사용되고 있는 보코더로는 Qualcomm사의 IS-96 QCELP, IS-127

EVRC가 있다. 이들 두 부호화기의 차이로는 인코더단에 모토롤라 소유권의 노이즈 억압 방식을 채용하고 있으며 Relaxation CELP의 피치 contour, Algebraic CELP의 dynamic sparse algebraic codebook을 적용하고 있다는 것이다. 또한 유선망을 통한 화상회의를 목적으로 한 G.723 Dual-rate speech coder 역시 algebraic codebook(5.3 kbps)을 사용하며 위 두 부호화기처럼 국내에서 많은 연구가 이루어지고 있다. 간차 신호를 모델링하는 데 통계적 방식대신 대수적 방식을 사용하는 이들 부호화기의 간차코드북은 알고리즘의 성능지표인 메모리요구량을 현저히 줄일 수 있으며 검색속도 역시 통계적 방식에 비해 향상시킬 수 있다. 더구나 기존의 Fast Nested-Loop Search 법 대신 DFS(Depth-First Search)방식은 약간의 음질저하로 검색시간을 현저히 향상시킬 수 있는 방식이다. 그러나 동일 트랙상에서 단지 하나의 펄스만이 선택 가능하다는 점, 각 펄스위치에서 최만의 제한된 부호비트로 코드어의 다양성을 제한한다는 단점이 있다.

본 논문에서는 상기 부호화기에서 사용되는 기존의 대수적 코드북을 가반으로 추가 계산량이 필요치 않은 U/V 분리를 사용하여 LSF의 보간 지연과 대수적 코드북을 개선한 5.6kbps ACELP를 구현하였다. 실험결과 6.3kbps MP-MLQ에 비해 주관적 음질평가에서는 동등하였으나 MNRU Q=15dB에서 약간의 음질저하를 느낄 수 있었다.

본 논문의 구성은 다음과 같다. 2장에서는 대수적 코드북 검색 법 중 널리 사용되는 두가지 방식의 검색법

저전송율 보코더의 성능개선에 관한 연구

에 대해 설명하고 3 장에서는 LSF 보간과 제안하는 dynamic sparse algebraic 코드북의 구조와 기존 Depth-First Nested-Loop 검색에 비해 개선된 검색 방법에 대해 논한다. 4 장에서는 5.6kbps ACELP 에 대한 주관적인 음질 평가 및 결과를 고찰하고 5 장에서 결론을 맺는다.

2. 일반적인 Algebraic 코드북 검색법

Adoul et. al에 의해 개발된 대수적 검색법은 통계적 방식의 잔차신호 모델링법보다 적은 메모리 사용량, 빠른 검색법으로 널리 사용되고 있다. 특히 G.729, G.723, EVRC 에 적용되었으며 DFS 의 대수적 코드북이 ITU-T 에 제출된 여러 4kb/s 보코더 및 6.3kbps CS-ACELP, 11.8kbps CS-ACELP 에 적용되고 있다. 본 장에서는 저전송율 부호화기에 적용되는 대수적 코드북의 검색법 중 Fast Search Nested-Loop 법과 DFS 법에 대해 논한다.

2.1 일반적인 대수적 코드북 구조

고정 코드북은 가중된 입력음성과 가중화 된 합성음 사이의 MSE(mean squared error)를 최소로 하는 펄스위치를 선택한다. 코드어를 결정하는 변수가 펄스의 위치와 이득이므로 이득을 일정하게 한 후 각 펄스위치에서 오차를 구하여 위치를 결정한후 이득을 구한다.

$$\begin{aligned} \min_k \|\Delta\|^2 &= \min_k \|S - S^k\|^2 = \min_k \|S - (P - gA_k H^T)\|^2 \\ &= \min_k \|X - gA_k H^T\|^2 \end{aligned} \quad (2-1)$$

식(2-1)을 다시 쓰면

$$\min_k \|\Delta\|^2 = \min_k \left\{ \|X\|^2 - \frac{(X(A_k H^T)^T)^2}{\|A_k H^T\|^2} \right\} \quad (2-2)$$

여기서 H 는 1xL 하삼각 Toeplitz 행렬이고 h(n)의 임펄스 응답이다. A_k 는 인덱스 k 의 코드벡터, X 는 목적벡터이다. (2-2)식의 오른쪽 두번째항이 최대가 될때 (2-2)식이 최적이 되므로

$$\begin{aligned} \max_k \frac{(X(A_k H^T)^T)^2}{\|A_k H^T\|^2} &= \max_k \frac{((X H^T) A_k^T)^2}{\alpha^2 k} \\ &= \max_k \frac{(DA_k^T)^2}{\alpha^2 k} \end{aligned} \quad (2-3)$$

과 같이 쓸수 있다

여기서 D 는 후필터링된 목적벡터이다. 여기서 식(2-1)을 만족하는 기본 criterion 이 필요하다.

N 개의 영이 아닌 펄스를 가지는 코드벡터의 집합 중 최적의 코드벡터를 결정하는 기본적인 criterion 은 다음과 같이 나타낸다.

$$k = \max_i^{-1} [Q_i(N)] \quad (2-4)$$

여기서

$$Q_i(N) = \left[\frac{(DA_k^T)^2}{\alpha_i^2 k} \right] \text{ 이다.}$$

criterion 의 분모는

$$\alpha_i^2 = \sum_{m=1}^N S_m^2 U(P_m, P_m) + 2 \sum_{m=1}^{N-1} \sum_{j=m+1}^N S_m S_j U(P_m, P_j) \quad (2-5)$$

분자는

$$\begin{aligned} DA_k^T &= \sum D_m S_m \text{ 이며} \\ U(i, j) &= \sum_{k=1}^i h(k-i+1)h(k-j+1) \end{aligned} \quad (2-6)$$

S_m S_n 는 펄스의 진폭이다.

2.2 Fast Search Nested-Loop 검색법[3][7][8]

criterion 을 만족하는 최적의 펄스위치를 검색하기위해서 먼저 코드북을 결정한다. 멀티펄스방식에서 펄스의 선택위치와 진폭을 제한한 대수코드북은 ISPP(Interleaved Single Pulse Permutation)를 이용하여 구성한다. 표 2.1 은 ISPP(40,5)의 특별한 경우이다.

ISPP 를 이용하여 구성된 코드북은 많은 계산량을 요구하는 식(2-5)을 계산하기위해 N 개의 루프를 순회하면서 최적의 펄스를 찾는다. 일명 focused 검색이라고 하는 이 검색법은 N-1 번째 펄스 검색을 행한 후 미리 결정된 문턱값을 넘는 경우에만 N 번째 펄스를 검색하는 검색법이다.

표 2.1 8kbps CS-ACELP codebook[4]

펄스	부호	위치
I ₀	+1/-1	0,5,10,15,20,25,30,35
I ₁	+1/-1	1,6,11,16,21,26,31,36
I ₂	+1/-1	2,7,12,17,22,27,32,37
I ₃	+1/-1	3,8,13,18,23,28,33,38
		4,9,14,19,24,29,34,39

2.3 Depth First 검색법[2]

Depth First 검색법은 Fast Search Nested-Loop 법에 비해

약간의 음질을 희생하며(G.729A의 경우 SNR: 0.2dB 감소) 계산량을 감소시켰다. N_p 개의 여기펄스는 N_m 펄스의 M 부분집합으로 나누어진다. 검색은 부분집합#1에서부터 시작되며 트리구조로 검색을 실시하므로 다음 부분 집합이 검색된다. 검색은 위치트랙안에 할당된 펄스순서를 변경하면서 반복적으로 검색된다. 먼저 펄스는 두펄스의 부분집합안에서 검색된다. 트랙 T_{k_0} 안에 최대값을 가지는 펄스 i_0 와 $T_{(k_0+1) \bmod \# \text{ of track}}$ 안의 최대값을 가지는 펄스 i_1 을 설정한다. 다음 $T_{(k_0+2) \bmod \# \text{ of track}}$ 와 $T_{(k_0+3) \bmod \# \text{ of track}}$ 안의 펄스들의 가능한 위치결합 (i_2, i_3) 에 대해 검색한다. 같은 과정을 $(i_4, i_5), (i_6, i_7), \dots$ 등에 대해 수행한다. 각 레벨에서 검색 criterion은 그 레벨에서 이용가능한 펄스에 대해서만 수행한다. 트랙에 할당된 펄스의 수를 변경하여 N_t 번의 반복을 수행한다. 그림 1에 검색법의 개념도를 나타내었다.

부프레임 길이가 40이며 한트랙안에 펄스위치의 수가 8이고 트랙이 5개인경우 3번의 트랙안에 펄스위치를 변경하여 수행할경우 $3 \times 4 \times 8 \times 8 = 768$ 번을 수행한다. 전체검색에 비해 2.34%의 검색만을 필요로 한다.

표 2.2 6.4kbps CS-ACELP track table

펄스	부호	위치
i_0	+1/-1	1, 3, 6, 8, 11, 13, 16, 18, 21, 23, 26, 28, 31, 33, 36, 38
i_1	+1/-1	0, 1, 2, 4, 5, 6, 7, 9, 10, 11, 12, 14, 15, 16, 17, 19, 20, 21, 22, 24, 25, 26, 27, 29, 30, 31, 32, 34, 35, 36, 37, 39

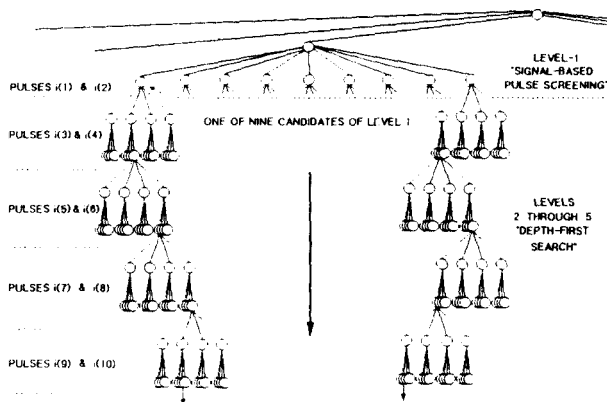


Fig 1. Depth First Search 검색법

3. 개선된 알고리즘

3.1 유/무성음 분리에 의한 LSF 보간

CELP 형의 보코더는 20-30ms의 프레임 구간에 대해 LSF 계수를 구한 후 인접 프레임간의 불연속에 의한 불협음을 제거하기위해 부프레임 단위로 보간을 행한다. 그러나 음의 형태에 따른 구분없이 일률적인 soft interframe 보간으로 무성음에서 유성음으로의 천이구간에서 원치않는 LP 값의 지연이 발생한다. 그 결과 재생음에서 유성음의 시작시 왜곡발생의 원인이 된다. 본 논문에서는 Durbin 알고리즘 동작시 부가적으로 발생하는 LAR(log area ratio)계수값을 가지고 별도의 추가 계산량을 필요치 않는 간단한 유/무성음/천이 구간 분리를 행 하였다. LAR(1), LAR(2)의 문턱값을 결정하기 위해 영교차율, 에너지 분포를 가지고 기준값을 결정한 후 실험적으로 결정하였다. Soft interframe 보간은 무성음에서 유성음으로의 천이시에 hard interframe 보간으로 대체된다.

3.2 개선된 코드북 검색법

기존의 G.729 보코더의 대수코드북이 표 2-1에 나타나 있다. 대수적 코드북의 문제점은 첫째, 멀티펄스에 대해 가능성이 있는 일부 펄스위치만을 검색하는 부분적인 하이퍼큐브방식을 사용한다. 둘째, 각 트랙에서 단지 하나의 펄스만을 검색하므로 동일 트랙상에 다른위치에 있는 두펄스가 최적의 값을 갖더라도 준 최적을 이루는 펄스를 검색한다. 셋째, 발견된 모든 펄스가 ± 1 이라는 유준화된 진폭을 가지며 이득을 음하더라도 부프레임 내에서 동일한 크기를 가진다. 따라서 본 논문에서는 표 3-1과 같은 코드북을 사용한다. 트랙 1,2에서는 하나의 펄스를 각 트랙에서 검색하여 발견하며 트랙 3에서는 두개의 펄스를 검색한다. 이렇게 하는 이유는 위에 설명한바와 같이 트랙 3에서 다른위치에 있는 펄스가 최적으로 판단되었을 경우 검색 불가능한것을 방지하기 위함이다. 또한 동일 위치에 두 펄스가 최적이라 판단되었을 경우 진폭부호는 두펄스의 합인 진폭 ± 2 를 전송한다.

부호비트의 측면에서 보았을 경우 한 트랙에서 두개의 펄스를 발견할 경우 부호 1비트는 잉여성분이다. 펄스 a, b 가 부호가 다르다 할지라도 절대적값으로 보아 두펄스의 위치는 교환이 가능하다. 만약 $P_a \leq P_b$ 이면 $S_1 = S_2$ 이고 그렇지 않다면 S_1 과 S_2 는 다르다고 본다. 그렇다면

저전송율 보코더의 성능개선에 관한 연구

단지 한비트의 부호만을 저장하여 동일트랙내의 두펄스의 부호를 표현할수 있다.

기존대수적 코드북의 단점은 펄스진폭비트를 ± 1 로만 나타내어 다양한 펄스 진폭 변화를 나타낼 수 없었다. 또한 각 펄스마다 N의 진폭변화를 설정할 경우 트랙이 L인경우 Σ 펄스위치+(N×L)의 급격한 비트증가로 검색시 많은 어려움이 따른다.

이를 개선하기위해서 먼저 부프레임내에서의 펄스진폭을 미리 설정한다. 펄스설정을 위해 목적벡터와 잔차신호의 r.m.s 값을 이용한다. 설정된 펄스는 부프레임 구간내 최대피크로 정규화한다.

$$S = Q \left(\frac{B_p}{\max_k |B_k|} \right) \quad (3-1)$$

여기서 $\max |B_k|$ 는 역이 아닌 진폭펄스의 최대치를 나타내는 정규화인자이고 Q(·)는 양자화 함수이다.

본 논문에서 사용하는 진폭의 경우 $\pm 1/\pm 2$ 크기의 진폭을 사용하므로 트랙 1의 경우 선택된 진폭에서 부호비트만을 사용한다. 트랙 2와 트랙 3에서는 S 함수가 미리 설정된 문턱값에 따라 $\pm 1/\pm 2$ 의 진폭을 결정한다.

검색과정은 기존의 Depth-First Search 법과 동일하다. 단, 행렬 U(i,j)를 계산하는데 미리 할당된 펄스 진폭을 결함하여 비록 0이 아닌 펄스가 q개의 경우의 수를 가지더라도 펄스 위치 검색시 모든 진폭은 고정시킨다.

$$U(i,j) = S, S, U(i,j) \quad (3-2)$$

표 3-1. 본 논문의 대수적 코드북 구조

트랙	부호	위치
T ₀	$\pm 1, \pm 2$	0,8,16,24,32,40,48,56
T ₁	$\pm 1, \pm 2$	2,10,18,26,34,42,50,58
T ₂	± 1	4,12,20,28,36,44,52,(60) 6,14,22,30,38,46,54,(62)

4. 실험 및 결과

본 논문에서 개선한 알고리즘을 평가하기위하여 컴퓨터 시뮬레이션을 실행하였다. 이용된 장비로는 IBM Pentium(166MHz)과 음성신호를 입출력 하기 위한 16비트 AD/DA 변환기를 인터페이스로 하여 8KHz로 샘플링

하였다.

이상의 방법에 대한 성능을 평가하기위해 아래와 같은 대표적인 문장을 20대 중,후반의 남녀화자가 각각 5번씩 발생하여 시료로 사용하였다.

1. 인수내 꼬마는 전재소년을 좋아한다.
2. 예수님께서 전지장조의 교훈을 말씀하셨다.
3. 숭실대 정보통신과 음성통신 연구실이다.
4. 창공을 헤쳐나가는 인간의 도전은 끝이 없다.
5. 공일이삼사오육칠팔구

원음성을 일반인을 이용하여 채취한것은 훈련된 발화자보다는 일반 전화 사용자의 음성을 정확히 반영한다고 볼수있기때문이다. 사용된 시료에 대한 주관적 음질검사를 실시하였다.

표 4-1은 주관적 음질검사 결과이다. 전화기 특성을 모델링한 IRS(Intermediate Reference System) 가중 필터를 통과한 음성신호를 SV56 음성 레벨 측정 도구를 이용하여 -26dBov로 정규화하였다. 기준 신호의 사용타입으로 P.81을 수행한 MNRU Ver.2.0를 사용하였으며 5dB 간격으로 입력 음성에 첨가하였다.

수행결과 주관적 평가에서 MP-MLQ와 차이가 없었다. 그러나 MNRU 15dB에서 약간의 음질의 저하를 느낄수 있었다.

표 4-1. 주관적 음질평가결과

코덱 \ MNRU	15dB	20dB	25dB	30dB
MP-MLQ	2.88	3.28	3.53	3.73
5.6kbps ACELP	2.84	3.31	3.54	3.76

표 4-2. 5.6kbps ACELP의 비트할당.

파라미터	부프레임 0/2	부프레임 1/3
LPC 계수	24	
적응코드북 지연	7	2
이득코드북	12	12
고정코드북	20	20
합계	170	

5. 결론

본 논문에서는 LAR 에 의한 유/무성음분리와 개선된 algebraic 코드북을 이용한 5.6kbps ACELP 를 구현하였다. 먼저 무성음에서 유성음으로의 천이시 발생할수있는 원치않는 지연을 제거하기위해 추가 전송율을 요구하지 않는 LAR 를 사용하여 hard interframe 보간을 하였다. 대수적 코드북에서 코드북의 구조를 변경하여 2 개의 펄스를 한 트랙에서 발견하여 대수코드북이 가지는 단점을 보완하였으며 펄스진폭의 다양성을 주기위해 네가지 경우에 대한 부호비트를 설정하여 코드북의 크기를 증가시켰으며 검색법의 단순화로 큰 계산량의 요구없이 성능을 증가시켰다. 음질검사를 실시한 결과 주관적 음질에서 MP-MLQ 와 차이가 없었으며 MNRU Q=15dB 에서 약간의 음질저하를 느낄 수 있었다.

coding for Telecommunications, Saint-Adele, Quebec, Canada, Oct, 13-15,1993

[10] C.Laflamme, J-P Adoul et. Al., "16KBPS WIDEBAND SPEECH CODING TECHNIQUE BASED ON ALGEBRAIC CELP", Proc. of ICASSP, pp.13-16,1991

참고문헌

- [1] A.M.Kondoz, "Digital Speech - Coding for Low bit rate communication system.", 1994
- [2] ITU-T Draft description of Annex D to Recommendation G.729 - 6.4kbps CS-ACELP Speech coding algorithm, May, 1998
- [3] ITU-T Recommendation G.723.1, March, 1996
- [4] ITU-T Recommendation G.729, July, 1996
- [5] 박영호,외 4 명, " 피치동기를 이용한 대수 CELP 의 성능개선", 제 13 회 음성통신 및 신호처리 워크샵, 1996. 8. pp 388 - 392
- [6] Y.H. Park, J.C. Yang, S.M. Sohn and M.J. Bae, "On a Time Reduction of Pitch Searching by the Regular Pulse Search Technique in the CELP Vocoder," IEEE Comm. Society, Proc. of MILCOM'97, Vol.1, Nov.02-05, 1997.
- [7] C.Laflamme, J-P. Adoul, H.Y. Su, and S. Morissette, "On Reducing Computational Complexity of codebook Search in CELP Coder Through the Use of Algebraic Codes", Proc. ICASSP pp.177-181, 1990
- [8] R. Salami et al., "A Toll Quality 8kbps speech Codec for the Personal Communications System(PCS)", IEEE Trans. Vehic. Tech., vol 43, no 3, Aug. 1994, pp 806-816.
- [9] V. K. Varma, "Testing Speech Coders for usage in Wireless Communication system." Proc. IEEE Workshop on speech