

다중레벨 진폭 코드북을 이용한 음성 부호화기에 관한 연구

⁰홍 성훈, 김 정진, *박 영호, 배 명진

송실대학교 정보통신공학과, *송실대 전자공학과
mjbae@saint.soongsil.ac.kr

On the Research of a Speech Coder Using a Multi-Level Amplitude Codebook

⁰SeongHoon HONG, JeongJin KIM, *YoungHo PARK, MyungJin BAE

Dept. of Telecomm. Engr., *Dept. of Electronic Engr., Soongsil Univ.

Abstract

This paper analyzes the dynamic sparse algebraic codebook used to model a residual signal and proposes a new algebraic codebook structure as well as a searching process with improved performance. The proposed algorithm improves the disadvantage of algebraic codebook without increased computation. First, this paper makes it possible to select various pulse amplitudes differently from the conventional method which looks up the sign bit simply. In addition, two pulses are made to be selected on the same track. For speech quality on the telephone line 5.6kbps speech coder using the proposed algorithm was equivalent to the 6.3kbps MP-MLQ in the viewpoint of subjective speech quality. However, speech degradation was caused a little compared to the MP-MLQ where MNRU Q=15dB.

1. 서론

현재까지 발표된 음성부호화기 중 가장 많은 연구가 이루어지고 있는 방식이 CELP(Code Excited Linear Prediction)방식이다. 이 방식은 4.8kbps내외의 전송율에서 양호한 음질을 얻을 수 있으며 ITU-T, TIA/EIA 등 여러 국제표준화 기구를 통해 다양한 응용분야에서 표준화가 이루어지고 있다. 특히 국내에서는 PCS의 출현으로 CELP부호화기에 대해 많은 관심이 집중되고 있다[1].

현재 국내에서 디지털 셀룰라 서비스에 사용되고 있

는 보코더로는 Qualcomm사의 IS-96 QCELP, IS-127 EVRC[2]가 있다. 이들 두 부호화기의 구조적인 차이로는 IS-127이 부호화단에 모토롤라 소유권의 노이즈 억압 방식을 채용하고 있으며 Relaxation CELP의 피치 contour, 대수CELP의 동적회박 대수코드북(dynamic sparse algebraic codebook)을 적용했다는 차이가 있다. 또한 유선망을 통한 화상회의를 목적으로 ITU-T에서 1996년에 표준화한 G.723.1 Dual-rate speech coder[3] 역시 대수 코드북(5.3kbps)을 사용하며 인터넷, 화상회의에 응용하기 위해 국내에서 많은 연구가 이루어지고 있다[4-6]. 잔차 신호를 모델링하는 데 기존의 통계적 방식대신 대수적 방식을 사용하는 이들 부호화기의 잔차코드북은 음성코딩알고리즘의 성능지표중의 하나인 메모리 요구량을 현저히 줄일 수 있으며 검색속도 역시 통계적 방식에 비해 향상시킬 수 있다. 그러나 대수적 코드북의 단점으로는 동일 트랙상에서 단지 하나의 펄스만이 선택 가능하다는 점, 각 펄스위치에서 진폭 검색시에 복잡도를 줄이기 위해 ± 1 만의 제한된 부호비트로 코드어의 진폭 다양성을 제한한다는 점이 있다.

따라서 본 논문에서는 이 같은 문제점을 효과적으로 극복할 수 있는 방식을 제안하였다.

2. 기존의 제안된 대수 코드북 검색법

2.1 Fast Search Nested-Loop 검색법

최적의 펄스위치를 검색하기 전에 먼저 코드북을 구성하는 테이블을 결정해야 한다. 특히 대수코드북은 트랙내의 구성샘플위치에 따라 음질차이가 발생하므로

일반적으로 각 트랙마다 이전 트랙과의 한 샘플위치씩 차이를 주는 구조를 가지고 있다. 그러나 부 프레임길이 길거나 트랙 당 전송할 펄스의 개수가 많을 경우 G.723과 같이 짝수위치/홀수위치 테이블을 따로 만들어 선택된 비트와 Grid 한 비트를 함께 전송하는 경우도 있다[3]. 멀티펄스방식에서 펄스를 선택할 위치와 진폭을 제한한 형식의 대수코드북은 ISPP(Interleaved Single Pulse Permutation)를 이용하여 구성되며 ISPP(F, \mathcal{E})으로 표현한다. 여기서 F 는 부프레임의 길이, \mathcal{E} 는 트랙의 수이다.

Salami는 선택 확률이 적은 펄스 트리를 제거하는 방식을 사용하여 계산량을 감소시키는 방법을 제안하였다. 일명 Focused 검색이라고 하는 이 검색법은 마지막 펄스 검색 전 N-1번째까지 펄스 검색을 행한 후 미리 결정된 문턱값을 넘는 경우에만 마지막 N번째 펄스를 검색하는 검색법이다.[8]

그림 2-1은 두개의 문턱값을 설정하여 불필요한 하부 트리를 검색하는 단점을 개선하였다.

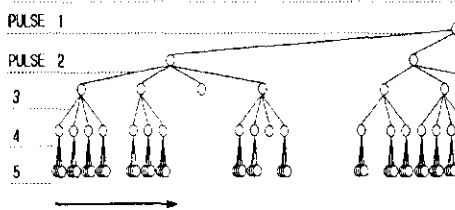


그림2-1. First Search Nested-loop 검색법

2.2 Depth First 검색법

Adoul에 의해 제안된 Depth First 검색법은 Fast Search Nested-Loop법에 비해 약간의 음질을 희생하며(G.729A의 경우 SNR : 0.2dB감소) 계산량을 감소시킨 방식이다.[9]

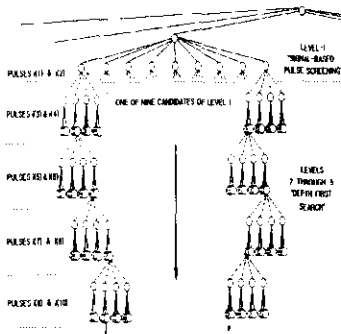


그림2-2. Depth First Search 검색법

N_p 개의 여기펄스는 N_m 개의 펄스로 구성된 M부분집합으로 나누어진다. 검색은 부분집합 #1에서부터 시작되며 트리구조로 검색을 실시하므로 다음 부분 집합이

검색되며 검색은 위치트레안에 할당된 펄스순서를 변경하면서 반복적으로 검색된다. 먼저 펄스는 두 펄스쌍의 부분집합안에서 검색된다. 트랙 T_{k_0} 와 $T_{(k_0+1) \text{ modulo } \mathcal{E}}$ 트랙의 개수로 구성된 쌍 $(i_{0,1})$ 에 대해 최적의 펄스위치를 검색한다. 그다음에 $T_{(k_0+2) \text{ modulo } \mathcal{E}}$ 트랙의 개수와 $T_{(k_0+3) \text{ modulo } \mathcal{E}}$ 트랙의 개수안의 펄스들의 가능한 위치결합 $(i_{2,3})$ 에 대해 검색한다. 같은 과정을 $(i_{4,5}), (i_{6,7})$...등에 대해 수행한다. 각 레벨에서 검색 판별식은 그 레벨에서 이용 가능한 펄스에 대해서만 수행한다. 트랙에 할당된 펄스의 수를 변경하여 N_k 번의 반복을 수행한다. 그림2-2는 검색에 대한 구조를 나타내고 있다. Focused검색법과는 다르게 처음 부분집합부터 후보펄스를 선택하기때문에 다른 검색이 가능하나 잘못된 패스선택은 음질의 저하를 가져올 수 있다.

3. 개선된 알고리즘

기존의 대수코드북의 구조를 보면 다음과 같은 문제점을 발견할 수 있다. 첫째, 멀티펄스에 대해 가능성이 있는 일부 펄스위치만을 검색하는 부분적인 하이퍼큐브 방식을 사용한다. 둘째, 각 트랙에서 단지 하나의 펄스만을 검색하므로 동일 트랙상에 다른 위치에 있는 두 펄스가 최적의 값을 갖더라도 준 최적을 이루는 펄스를 검색한다. 셋째, 발견된 모든 펄스가 일정한 진폭을 가지고 있다고 가정하고 ± 1 이라는 규준화된 진폭만을 검색한다. 따라서 본 논문에서는 이를 개선하기 위하여 표 3-1과 같은 코드북을 구성하였다. 검색방법은 DFS검색을 사용한다. 첫 번째와 두 번째 문제를 해결하기 위하여 트랙 1,2에서는 한 펄스를 각 트랙에서 검색하여 발견하며 트랙3에서는 두개의 펄스를 검색한다. 이렇게 하는 이유는 위에 설명한 바와 같이 트랙3에서 다른 위치에 있는 펄스가 최적으로 판단되었을 경우 검색 불가능한 것을 방지하기 위함이다. 또한 동일 위치에 두 펄스가 최적이라 판단되었을 경우 진폭부호는 두 펄스의 합인 진폭 ± 2 를 전송한다. 부호비트의 측면에서 한 트랙에서 두개의 펄스를 발견할 경우 부호 1비트는 잉여성분으로 볼 수 있다. 즉, 펄스 a, b가 서로 다른 부호를 가지고 있다 할지라도 절대적 값으로 보아 두 펄스의 위치는 교환이 가능하다. 만약 두 펄스의 부호가 같다면($S_1=S_2$) 낮은 위치의 펄스는 P_a 에, 높은 위치의 펄스는 P_b 에 둔다. 그리고 두 부호가 같지 않다면 낮은 위치의 펄스는 P_b 에, 높은 위치의 펄스는 P_a 에 둔다. 따라서 두 펄스의 위치와 S_1 의 부호만 전송하면 수신단에서 두펄스의 위치및 부호를 복원할 수 있다.

세 번째 문제를 해결하기 위하여 다음과 같은 방식을 제안하였다. 기존대수적 코드북의 단점은 펄스진폭 비트를 ± 1 의 부호성분만 나타내어 다양한 펄스 진폭 변화를 나타낼 수 없었다. 또한 각 펄스마다 N의 진폭

변화를 설정할때 트랙이 L개인경우 \sum 펄스위치 $\cdot (N \times L)$ 만큼의 비트가 필요하다. 즉, 코드북 크기의 급격한 증가로 검색시 많은 계산량이 필요로 된다.

표3-1. 본 논문에서 제안한 대수적 코드북 구조

트랙	부호	위 치
T_0	$\pm 1, \pm 0.5$	0,8,16,24,32,40,48,56
T_1	$\pm 1, \pm 0.5$	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)

이를 개선하기 위해서 먼저 부프레임 내에서의 펄스 진폭을 미리 설정한다. 식(3.1)로 진폭을 결정하기 위한 목적진폭신호를 만든다. 식(3.1)은 펄스설정을 위해 목적벡터와 피치성분을 제거한 원음성 그리고 LPC성분을 제거한 원음성과의 적절한 혼합비율로 구성되어 있다. 여기서 v_1, v_2, v_3 는 0과 1사이의 값이어야 하며 본 논문에서는 $v_1 = v_2 = 0.5, v_3 = 0.1$ 을 사용하였다. 그 결과 추출된 펄스 진폭은 부프레임 구간내에서 식(3.2)을 이용하여 정규화한다.

$$Y = v_1 \frac{D}{|D|} + v_2 \frac{P}{|P|} + v_3 \frac{S}{|S|} \quad (3.1)$$

$$\Delta = Q \left(\frac{V_p}{\max |V_n|} \right) \quad (3.2)$$

여기서 $\max |V_n|$ 는 영이 아닌 진폭펄스의 최대치를 나타내는 정규화 인자이고 $Q(\cdot)$ 는 양자화 함수이다.

본 논문에서 사용하는 진폭의 경우 $\pm 1/\pm 0.5$ 크기의 진폭을 사용하므로 트랙2의 경우 선택된 진폭에서 부호비트만을 사용한다. 트랙 1,2에서는 Δ 함수가 미리 설정된 문턱값에 따라 $\pm 1/\pm 0.5$ 의 진폭을 결정한다. 검색과정은 기존의 Depth-First Search법과 동일하다. 단, 행렬 $\Phi(i,j)$ 를 계산하는데 제안된 알고리즘으로 미리 계산되어진 진폭을 식(3.3)과 같이 결합시킨다. 이렇게 함으로써 진폭이 q개의 경우의 수를 가지더라도 진폭을 검색하기 위한 계산량의 증가는 발생하지 않는다.

$$\Phi(i,j) = S, S, \Phi(i,j) \quad (3.3)$$

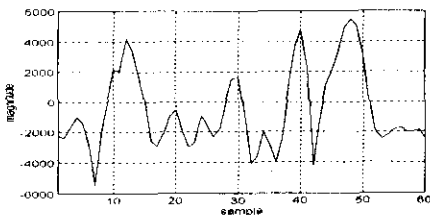


그림 3-1. 목적벡터 파형 (부 프레임 : 60샘플)

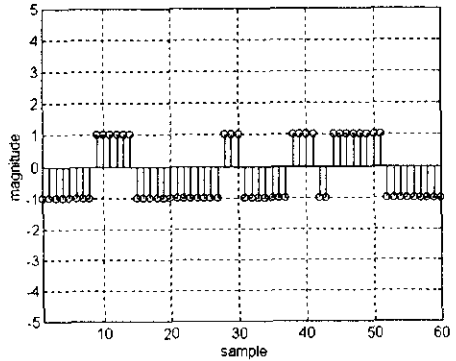


그림 3-2. 기존방식의 목적벡터 진폭추출

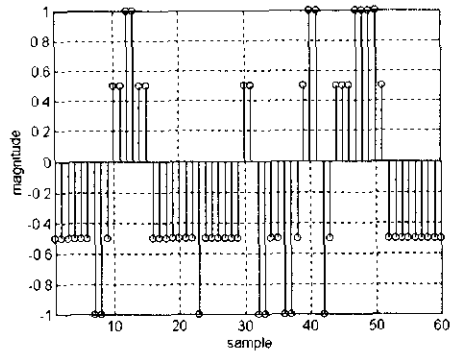


그림 3-3. 제안한 방식을 이용해 추출한 진폭파형

4. 실험 및 결과

본 논문에서 개선한 알고리즘을 평가하기 위하여 컴퓨터 시뮬레이션을 실행하였다. 이용된 장비로는 IBM Pentium(166MHz)과 음성신호를 입출력 하기 위한 16 비트AD/DA변환기를 인터페이스로 하여 8kHz로 샘플링 하였다.

이상의 방법에 대한 성능을 평가하기 위해 아래와 같은 대표적인 문장을 20대 중,후반의 남녀화자가 각각 5번씩 발성한 음성을 시료로 사용하였다. 시료는 두드러진 피크를 가지지 않고 노이즈가 30dBa를 가진 방에서 녹음하였다.

1. 인수네 꼬마는 천재소년을 좋아한다.
2. 예수님께서 천지창조의 교훈을 말씀하셨다.
3. 송실대 정보통신과 음성통신 연구실이다.
4. 창공을 헤쳐 나가는 인간의 도전은 끝이 없다.
5. 공일이삼사오육칠팔구

원음성은 일반인을 이용하여 채취하였으며 그 이유는 훈련된 발화자보다는 일반 전화 사용자의 음성을

정확히 반영한다고 볼 수 있기 때문이다. 녹음된 원 시료는 전화기 특성을 나타내기 위하여 전화기 특성을 모델링한 변형된 IRS (Modified-Intermediate Reference System) 가중 필터를 통과시켰다. IRS 필터를 통과한 음성신호는 P.56 음성 레벨 측정 도구를 이용하여 $-26 \pm 1 \text{dBov}$ 로 정규화하였다. 그리고 주관적 음질평가를 하기 위해 MOS(Mean Opinion Score)를 사용하였으며 MOS의 신뢰성을 위해 기준 신호의 사용타입으로 P.81을 준수한 MNRU Ver.2.0를 사용하였다. MNRU의 Q값은 5dB간격으로 입력 음성에 첨가하였다. 본 논문에서는 진폭파형의 양자화 스텝을 네 단계($\pm 1, \pm 0.5$)로 하였으며 식(3.1)의 결정인자는 $v_1 = v_2 = 0.5, v_3 = 0.1$ 으로 실험적인 값을 사용하였다. 그림 3-1~3-3은 제안된 알고리즘의 진폭파형 추출에 대한 기존 방식과의 비교파형이다. 기존 방식에 의한 두 레벨(± 1) 진폭파형을 나타내는 그림 3-2보다 극소의 추가 계산량으로 네 레벨($\pm 1, \pm 0.5$)에 대해 그림 3-3과 같이 진폭파형을 추출하였다. 표 4-1은 주관적 음질평가를 나타낸다. 비교 대상으로 유사한 전송율을 가진 6.3kbps MP-MLQ 와 비교하였다. 수행결과 주관적 평가에서 MP-MLQ와 차이가 없었으나 MNRU Q=15dB에서 약간의 음질의 저하를 느낄 수 있었다.

표 4-1. 주관적 음질평가결과 [MNRU : dB]

코덱 \ MNRU	15	20	25	30
MP_MLQ	2.88	3.28	3.53	3.73
제안한 방식을 이용한 5.6kbps CELP	2.84	3.31	3.54	3.76

표 4-2. 5.6kbps ACELP의 비트할당

파라미터	부프레임 0/2	부프레임 1/3
LPC 코드북	24	
적용 코드북	7	2
잔차 코드북	20	20
이득 코드북	12	12
합계	170	

5. 결론

본 논문에서는 대수적 코드북을 이용한 음성부호화기의 성능을 개선시켰다. 로그영역비를 이용한 무성음에서 유성음으로의 천이검출과 제안된 대수 코드북을 이용하여 펄스진폭의 다양성을 개선하였으며 위치선택에 있어서의 제한을 비트증가없이 완화시켰다. 대수 코드북에서는 코드북의 구조를 개선하여 2개의 펄스를 한 트랙에서 검출하여 대수코드북이 가지는 단점을 보완하였다. 진폭면에서는 펄스진폭의 다양성을 주기위

해 네가지 경우에 대한 부호비트를 설정하였다. 결과적으로 코드북의 크기가 증가되었으나 진폭의 사전 설정에 의한 검색법의 단순화로 계산량의 증가없이 성능을 향상시켰다. 제안된 알고리즘에 대한 음질검사를 실시한 결과 주관적 음질에서 MP-MLQ와 차이가 없었으며 MNRU Q=15dB에서 제안된 알고리즘을 이용한 부호화기의 약간의 음질저하를 느낄 수 있었다.

6. 참고문헌

- [1] A. M. Kondoz, *Digital Speech - Coding for Low bit rate communication system*, 1994
- [2] TIA/EIA/IS-127, *Enhanced Variable Rate Codec, Speech Service Option 3 for Wideband Spread Spectrum Digital Systems*, 1997
- [3] ITU-T Recommendation G.723.1, March, 1996
- [4] 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.*
- [5] S.H. Hong, M.J. Bae, K.J. Byun and J.J. Cha, "On a Pitch Search Technique with Correlation Characteristic of Quantization Error in CELP Type Vocoder," *The Sixth Western Pacific Regional Acoustics Conference, Proceeding of WESTPRAC-VI'97, Vol.1, pp.106-111, Nov.19-21, 1997.*
- [6] M.J. BAE and J.D. KIM, "On a Smoothing Method by Time-Axis Control with the Zero-Crossing of Speech Signal," *IEEE Communication Society, Proceeding of MILCOM'97, Vol.1, pp.-, Nov.02-05, 1997.*
- [7] Matsushita Electronics Inc.co., *High Level description of proposed 4kbps Speech coder and Qualification Test Results*, ITU-T SG16 Delayed-133, Feb,1998
- [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] C. Laflame, J-P. Adoul et. Al., "16kbps wideband speech coding technique based on algebraic CELP", *Proc. of ICASSP, pp13-16, 1991*