2층 코드북 구조를 통한 CELP 음성부호화기의 성능 향상에 관한 연구

"김중우", "김응곤", "한승조"
*조선대학교 전자정보통신공학부, **순천대학교 컴퓨터과학과

A Study on the performance improvement of the CELP coder by the structure of dual codebook

"Jong-Woo Kim", "Eung-Kon Kim", "Seung-Jo Han"
* School of Electronics and Information Communication Eng, Chosun Univ.
** Dept. of Computer Science, Sunchon National Univ.

요 약

본 논문에서는 CELP 부호화기의 재산량을 줄이면서도 고음질의 음성을 할성할 수 있는 코드북 구조를 제안한다. 제안한 코드북 구조는 복수의 코드북과 희박 중첩형 코드북 두 개의 코드북의 합으로 저지신호를 표현한다. codebook 1에서 전무신호와 오차가 적은 저지신호를 구한 후, 이 저지신호에 codebook II의 저지신호를 합하여 저덕의 저지신호를 구한다. 또한 이로 인한 전송 비트수의 증가를 막기위해 효율적인 프레임에서는 두 개 코드북의 index 를, 짧은 프레임에서는 codebook I의 저지신호를 그대로 사용하고 codebook II에서만 절약하여 전송하는 방법을 사용하였다. 이와 같은 2층 코드북 구조는 2개의 저지신호를 함으로 표현되고 각각의 서로 다른 코드북 이득을 사용하기 때문에 정확한 이득을 표현할 수 있어 기존의 개선 알고리트미보다 더 나은 음질을 제공할 수 있다. 검색시간이 끝나고, 본 코드북 구조를 갖는 4.8kbs CELP형 부호화기를 설계하여 컴퓨터 모의 실험한 결과, 같은 전송률을 갖는 DOD CELP 부호화기보다 segSNR가 0.53dB 더 높게 나타났다.

1. 서 론


CELP 음성 부호화기는 Analysis-by-Synthesis 방식인, 입력 음성을 분석하여 필요한 파라미터를 추출하고 이를 이용, 코드북내에 저장된 코드워드와 합성하여 입력 음성의 오차가 가장 적은 코드워드로 선택하는 방식에 기초를 둔다. 이러한 CELP 음성 부호화기는 저지신호_excitation signal의 부호화와 정족의 코드워드를 찾는 코드북 구조에 따라 부호화기의 종류가 나뉘고 성능이 달라진다.

본 논문에서는 기존의 CELP 음성 부호화기의 코드북 구조를 개선함으로써 적은 비트수로 더 나은 음질을 갖는 코드북 구조를 제안한다. 제안된 코드북 구조는 기존의 코드북 구조는 달리 2층 코드북 구조를 적용함으로써 저지신호를 더욱 잘 표현할 수 있다.

2. CELP 음성 부호화기의 구조

CELP(Code Excited Linear Prediction) 음성 부호화기는 코드북내에 저장되어 있는 입력 저지신호를 두 개의 시변 선형 회귀 필터를 통과시키며 얻은 신호 중, 주어진 손실 도 관정을 최적화 시키는 데 선택하도록 구성되어 있다. 음성신호를 두 개의 시변 선형 회귀 필터로 필터링한 후 남아 있는 신호를 잔류신호(저지신호)라고 하는데 이 신호는 음성 신호보다 낮은 분산을 가지고 있어서 음성신호보다 더 쉽게 양자화할 수 있다. 그러나 낮은 전송률을 갖는 부호화기는 잔류신호를 위해 적합한 비트수가 적으므로 직접 양자화하기 힘들기 때문에 코드북을 이용하여 양자화 하였다. 이 코드북의 구조에 따라 부호화기의 성능에 차이가 있다. 잔류신호와 코드북내의 여타신호들과의 오차를 최소화시킨다고 해서 입력 음성신호와 잔류신호의 오차가 최소화된다고 말할 수 없다. 그래서, CELP 음성 부호화기는 코드북내의 여타신호와 입력 음성신호로부터 구한 파라미터를 이용하여 합성한 음성신호와 입력 음성신호를 비교하여 오차를 최소화하는 저지신호 잔류신호를 양자화하는 소위 함성에 의한 분석(Analysis-by-Synthesis)법을 사용한다. 그러
1999년도 한국정보과학회 가을 학술발표논문집 Vol. 26. No. 2

나 매번 음성을 합성해서 비교해와므로 매우 복잡한 구조를 가지며 많은 제한량을 필요로 한다. 또, 제한된 코드북구조 때문에 여기서의 제로 표현하기 힘든 단점을 있다. 그림 1은 CELP 음성 부호화기의 기본 구조를 나타낸 것이다.

\[ P(z) = \frac{1}{1 - b_z z^{-1}} \]
\[ A(z) = \frac{1}{1 - a_z z^{-1} - a_z z^{-2} - \ldots - a_z z^{-m}} \]

식(1)의 피처 필터 파라미터 b와 M은 개회로(Open Loop) 구조나 폐회로(Close Loop) 구조 중의 하나를 이용해 얻어질 수 있는데, 폐회로 구조를 이용하여 구한 경우의 음성은 정립된 일반적으로 기본적으로 개회로 구조를 사용한다. 식(2)의 LPC 계수는 음성 압축시 양자화에 부적절하다고 알려져 있다. 그래서 LPC 계수와 같은 스펙트럼에 관한 정보를 가지며, 안정성 이 압축시에도 보장되고 성능이 매우 좋은 특성을 가진 LSP 계수로 변환하여 사용한다.

3. 2 째 코드북 구조

본 논문에서 제안한 코드북 구조는 기존의 코드북 구조와는 달리 2중으로된 코드북으로 전처리신호와 합성음성신호를 2번 비교함으로서 여기서의 제로 표현할 수 있다. 그림 2는 제안한 코드북 구조를 나타낸 것이다.

2중 코드북 구조는 codebook Ⅰ은 불규칙코드북, codebook Ⅱ는 확률 전처리코드북으로 구성된다. codebook Ⅰ내의 여기서 신호의 \( x'(n) \)을 \( b_k \)를 통해 적절하게 이득을 조절한 \( u'(n) \)과 전처리신호\( u(n) \)의 차이를 오차 가중 필터로 통해 오차가 가장 적은 여기서 신호를 이용한다. 그리고 codebook Ⅱ내의 여기서 신호의 \( x'(n) \)을 \( b_k \)를 통해 이득을 조절한 \( u'(n) \)과 codebook Ⅰ에서 구한 \( u(n) \)을 더하여 합성 여기서 신호\( u'(n) \)을 구한다. 그 후 기존의 CELP 음성부호화와 마찬가지로 이 여기서 신호\( u'(n) \)을 두 개의 시간 전처리 필터를 통과시켜 음성을 합성하여 임력음성과 비교하여 오차를 최소화하는 여기서 신호\( x'(n) \)을 구한다.

2중 코드북 구조는 codebook Ⅰ은 \( b_k \)의 음성신호에 \( 0.1 \)의 신호감을 갖는 크기가 N인 불규칙코드북으로 구성되어 있고, codebook Ⅱ는 \( b_k \)의 1과 80%이하의 시간을 갖고 입출력 코드워드간에 상관성을 갖는 크기가 M인 확률 전처리코드북으로 구성되어 있다.

여기서 신호은 codebook Ⅰ과 codebook Ⅱ의 합으로 구하기 때문에 이론적으로 \( N \times M \)의 크기의 불규칙 코드북을 가지는 CELP 음성부호화의 성능을 낼다. 또한 코드북 검색시 N번 진한 신호로 비교하고 M번 음성으로 합성하여 비교하므로, 코드북내 모든 여기서 신호로 음성을 합성하여 비교하는 기존의 방

법보다 코드북 검색시간이 단축된다. 즉 확률 전처리 코드북을 가지는 부호화기에 비해 검색 속도가 더 빠르고 더 나은 음

질을 가지는 불규칙 코드북의 성능을 가진다는 장점이 있다.

(a)혼수 부프레임 (b)학수 부프레임

그림 2. 2중 코드북 구조

2중 코드북 구조는 두 개 코드북 각각의 이득을 사용하기 때문에 정확한 이득을 표현할 수 있어 더 나은 음질을 합성할 수 있다. 그러나 이를 인한 코드북에 필요한 비트수와 증가하게 된다. 이러한 문제를 해결하기 위해서 다음과 같은 방법을 사용하였다. 임력음성 분석적 재 동일한 프레임들을 사용한 결과, codebook Ⅰ, Ⅱ의 여기서 신호\( x'(n) \)와 각각의 이

득 두개를 전송한다. 동시에, 전신 부프레임에서는 앞 프레임에서 구한 codebook Ⅰ의 여기서 신호\( x'(n) \)은 끝내 이용하고 이득을 조

절한 후 codebook Ⅱ에서만 검색하여 두 개 코드북의 이득과 codebook Ⅱ의 여기서 신호\( x'(n) \)만 전송한다. 음성신호는 각

부프레임간에는 비교적 유사한 성질을 가지고 있다. 음질은 codebook Ⅰ에서의 최적 여기서 신호\( x'(n) \) 선택 여부보다는 이득과
codebook Ⅱ에서의 여기서 신호\( x'(n) \) 선택에 따라 더욱 영향을 미친다. 그러므로 이와 같은 방법을 사용하여 음질의 저하를 최소화하면서 많은 비트수를 줄일 수 있다. 또한 학습방법에선 codebook Ⅰ의 검색을 생략하기 때문에 검색속도도 항상한다.

表 1. 4.8kbps CELP 음성부호화기의 비트수 합산 예

<table>
<thead>
<tr>
<th>frame</th>
<th>subframe</th>
<th>subframe</th>
<th>subframe</th>
<th>subframe</th>
</tr>
</thead>
<tbody>
<tr>
<td>codebook Ⅰ</td>
<td>codebook Ⅰ</td>
<td>codebook Ⅰ</td>
<td>codebook Ⅰ</td>
<td></td>
</tr>
<tr>
<td>index gain</td>
<td>i</td>
<td>g</td>
<td>i</td>
<td>g</td>
</tr>
<tr>
<td>bit</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td>4</td>
</tr>
</tbody>
</table>

4. 결론 및 성능 평가

제안된 방법의 성능 평가를 위하여 제안한 코드북 구조의 4.8kbps 음성 부호화기를 C 언어로 프로그래밍하여 컴퓨터 모

의 실험을 하였고, 4.8kbps 음성 부호화기의 성능은 표 2와 같

고 표 3와 같이 비트수 합산하였으나.

272
표 2. 4.8kbps 음성 부호화기의 규격

<table>
<thead>
<tr>
<th>항목</th>
<th>값</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Rate</td>
<td>4800 bit/sec</td>
</tr>
<tr>
<td>Sampling Rate</td>
<td>8kHz</td>
</tr>
<tr>
<td>Frame Size</td>
<td>240 samples(30ms)</td>
</tr>
<tr>
<td>Pitch Subframe Size</td>
<td>60 samples(7.5ms)</td>
</tr>
<tr>
<td>Codebook Subframe Size</td>
<td>60 samples(7.5ms)</td>
</tr>
<tr>
<td>LPC Filter Order</td>
<td>10</td>
</tr>
<tr>
<td>Pitch Predictor Order</td>
<td>1</td>
</tr>
<tr>
<td>Codebook Order</td>
<td>32×64 =2048</td>
</tr>
</tbody>
</table>

한 프레임의 길이는 30ms(240samples)로 하였으며, 각 프레임의 부프레임은 7.5ms(60samples)씩 네 개로 구성하였다. LPC 분석을 위해 37.5ms(300samples)의 Hamming 창 함수를 7.5ms(60samples)씩 중첩하여 사용하였다. 포먼트 필터는 자기 상관관계 방법으로 10개의 LPC를 구하고 전송을 위해서 LSP(Line Spectrum Pair)로 변환하여 프레임간의 LSP 끝이 차이를 34Bit로 스칼라 양자화를 사용하여 부호화 하였다. 피저 않은 첫 번째, 세 번째 부프레임에서는 부프레임 모든 구간에서 피저의 지연값을 7Bit로 부호화 하였고, 두 번째, 네 번째 부프레임에서는 바로 앞 부프레임의 상대적 차를 5Bit로 부호화 하였다. 피저 이득은 각 부프레임당 5Bit로 코드부는 흰부 부프레임 11(=5+6)Bit, 빈부 부프레임 6Bit 전부 34Bit씩으로 부호화 하였으며, 코드부 이득은 각 부프레임당 8(=4+4)Bit씩 32Bit로 부호화 하였다. 총 한 프레임당 144Bit로 부호화 하였으며, 4.8kbps의 전송률을 필요하다.

제안된 코드북 구조의 4.8kbps 음성 부호화기의 성능을 평가하기 위해 비교 기준 부호화로 4.8kbps DoD CELP 음성 부호화기를 선정하였다. 성능평가에는 판독가 평가 방법인 segSNR(segmental SNR)을 사용하였다. 평가 시리즈는 20개 남녀차각 각각 2명의 다음 문장을 받은 시료를 사용하였다. 표 4는 평가 결과를 나타낸다.

1. 영, 말, 이, 삼, 사, 오, 육, 칠, 팔, 구, 십
2. 여보세요. 안녕하십니까?
3. 지금은 휴식을 받을 수 있으니, 다음에 다시 잡아 주십시오.

표 4. segSNR 비교표

<table>
<thead>
<tr>
<th></th>
<th>4.8kbps DoD CELP</th>
<th>제안된 코드북 구조의 CELP</th>
</tr>
</thead>
<tbody>
<tr>
<td>문장 1</td>
<td>12.67</td>
<td>13.15</td>
</tr>
<tr>
<td>문장 2</td>
<td>13.04</td>
<td>13.59</td>
</tr>
<tr>
<td>문장 3</td>
<td>12.31</td>
<td>12.87</td>
</tr>
<tr>
<td>평균</td>
<td>12.67</td>
<td>13.20</td>
</tr>
</tbody>
</table>

표 4에서 알 수 있듯이 제안된 코드북 구조를 가지는 음성 부호화기는 4.8kbps DoD CELP 음성 부호화기보다 segSNR가 다 높게 나왔다. 제안된 코드북 구조를 이용한 음성부호화기의 segSNR는 13.20dB로 4.8kbps DoD CELP 음성 부호화기의 12.67dB보다 0.53dB 더 높게 나왔다. 제안된 코드북 구조를 갖는 CELP 음성 부호화기는 동일한 조건에서 4.8kbps DoD CELP 음성 부호화기보다 더 우수한 성능을 보였다.

5. 결론

본 논문에서는 202 코드북 구조를 이용하여 코드북 집합 웨이 구간을 줄이면서도 더 나은 음질을 할당할 수 있는 코드북 구조를 제안하였다. 이 코드북 구조는 평균적인 코드북과 최악 중첩 형 코드북 두 개로 구성되어 있어 두 개의 코드북의 합으로 음성의 여기서입률을 표현함으로써 더 나은 음질을 할당할 수 있다. 또한 압축 부프레임에서는 첫 번째 코드북의 여기서입률은 바로 앞 부프레임에서 구한 값을 그대로 사용하고 두 번째 코드북에서 감소하여 전송함으로써 성능의 개선 및 두 개의 코드북이 사용할 수 있는 비트수의 증가를 줄였다. 202 코드북 구조는 두 개의 코드북과 두 개의 코드북 이득을 사용하여 여기서입률을 표현함으로써 다양한 특성을 가지고 있는 음성 신호를 잘 표현할 수 있었다. 이러한 코드북 구조를 가지는 음성 부호화기는 실시간 구현이 가능한 연산량이 적고 성능을 분석한 결과 4.8kbps DoD CELP 음성 부호화기에 비해 성능이 우수함을 확인하였다.

참고 문헌