

G.729 음성 보코더를 이용한 가변 전송율 보코더 구현

함 명 규, 배 명 진

송실대학교 정보통신공학과
156-743 서울시 동작구 상도동 1-1
mkham7@hanmail.net

Implementation of the Variable Bit Rate Vocoder Using G.729 Vocoder

MyungKyu Ham, MyungJin Bae

Dept. Information and Telecommunication Engr., Soongsil University
1-1 Sangdo-5Dong, Dongjak-Ku, Seoul 156-743, KOREA
mkham7@hanmail.net

요 약

본 논문에서는 8kbps의 전송율을 가진 ITU G.729 보코더와 PSOLA(Pitch Synchronized Overlap -Add) 알고리즘을 적용하여 전송율을 6kbps와 4kbp까지 낮출 수 있는 가변 전송율 보코더를 구현하였다. 제안한 방법은 4kbps일 경우에 G.729의 부호화전에 PSOLA를 적용하여 피치의 주기를 반으로 줄여 부호화한다. 이렇게 부호화된 데이터는 G.729의 복호화를 거치고 다시 PSOLA를 통해 음성의 피치 주기를 2배로 늘려주어 원음성을 합성하게 된다. 기존의 8kbp의 전송율을 갖는 G.729는 음성의 크기가 반으로 줄어 부호화되므로 전송율이 4kbp로 줄어들게 된다. 실험의 평가는 MOS 테스트를 통해 수행되었으며 4kbp에서 MOS값이 3.37정도로 측정되었다. 또한 처리해야할 음성의 길이가 줄어들게 되므로 계산시간도 줄어들게 된다.

I. 서 론

현재까지 발표된 음성부호화기 중 가장 많은 연

구가 이루어지고 있는 방식은 CELP(Code Excited Linear Prediction)구조이다. 이러한 구조의 보코더들은 낮은 전송율에서 양호한 음질을 얻을 수 있으며 ITU-T 국제표준화 기구를 통해 다양한 응용분야에서 표준화가 이루어지고 있다. 특히 PCS 및 전화기 라인상에서의 인터넷을 통한 화상회의를 위하여 낮은 전송율에서 고음질을 가지는 코덱이 많은 주목을 받고 있다. 이러한 CELP 계열 보코더들 중에서 G.729는 멀티미디어 통신 환경하의 음성 전송 표준 보코더로 개발되었다.

G.729는 8kbps의 전송율을 갖는 구조로 현재 별정 통신으로 상용화되는 인터넷폰과 그 외의 이동 통신용 보코더로 사용되어지고 있으며, 낮은 전송율에 비해서 우수한 음질을 제공하고 있다.

본 논문에서는 G.729의 전송율인 8kbp 전송율에 대해 PSOLA 알고리즘을 적용하여 6kbps와 4kbps의 전송율로 낮추어 주는 가변 보코더를 구현하였다. 또한 음질평가 결과 낮은 전송율에 비해 우수한 음질을 얻을 수 있었다.

II. PSOLA 피치변경 합성방식

PSOLA에 의한 피치변경은 시간 영역에서만 처리를 하여 계산시간이 적게 소모되는 장점을 가진 TD(Time Domain)-PSOLA와 주파수 영역에서 처리하는 FD(Frequency Domain)-PSOLA의 방식이 있다[1][2]. 본 논문에서는 TD-PSOLA를 사용하였다. 이 방식은 파라미터를 사용하는 합성 방식과는 달리 사전에 저장된 음성신호를 부드럽게 연결하여 합성을 하므로 합성음의 음질이 좋다.

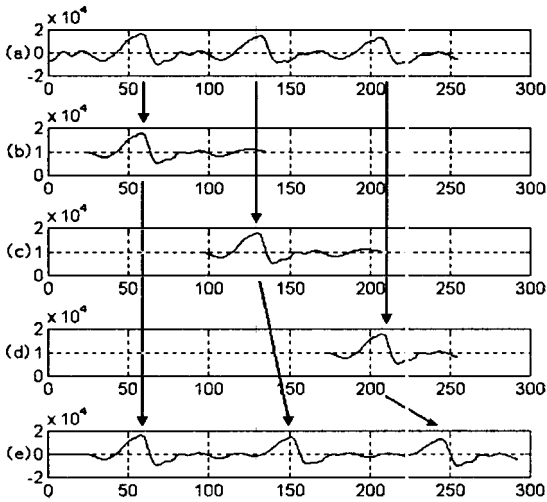


그림 2-1. TD-PSOLA의 피치 신장 합성법
(a) 원음성 신호, (b) 단구간 신호 1
(c) 단구간 신호 2, (d) 단구간 신호 3
(e) 피치가 신장된 합성 신호

PSOLA 피치변경 합성법의 과정은 다음과 같다.

2.1. 분석 과정

원음성 신호를 피치단위의 윈도우를 씌워 단구간(ST:Short Term)의 신호 열로 분해한다. 분석을 위한 윈도우에는 Hanning 윈도우를 사용하고 있다. 그러나, 본 논문에서는 음성 신호의 특징을 잘 표현할 수 있는 비대칭 윈도우를 사용하였다[3]. 아래의 식 (2-1)은 피치 단위로 분해된 단구간의 신호열을 나타내는 식이다.

$$S_{analysis}(n) = W_{analysis}(m-n)S(n), \quad (2-1)$$

$S_{analysis}(n)$: 피치주기 단위의 단구간(ST) 신호

$W_{analysis}(n)$: 분석 윈도우 함수

m : m번째 피치

$S(n)$: 원음성 신호

2.2. 피치 조절 및 합성 과정

분석과정에 분해된 단구간의 신호 열들은 원래 음성샘플의 피치단위로 배열되어 있다. 따라서, 피치를 낮거나 높게 합성하기 위해서는 이 간격들을 넓거나 좁게 재배열하면 된다. 아래의 식 (2-2)는 피치 변경된 신호를 나타내는 식이다.

$$S_{synthesis}(n) = S_{analysis}(n-m_a) \quad (2-2)$$

$S_{analysis}(n)$: 피치가 변경된 단구간 신호

m_a : 변경할 피치 간격

따라서, 피치를 높일 경우에는 단구간 신호의 간격을 작게 배열하고, 피치를 낮출 경우에는 단구간 신호의 간격을 크게 배열하면 된다. 여기서 음질을 좋게 하기 위해서는 피치를 동기화 시켜주는 것이 아주 중요하게 작용한다. 이렇게 재배열된 단구간 신호 열에서 겹쳐지는 부분들을 더해주면 된다.

III. 실험 및 결과

제안한 방법의 성능을 평가하기 위해 IBM-PC/PentiumIV(1GHz)에서 실험하였고, 알고리즘은 C언어로 구현하였다. 음성 시료로는 남성과 여성에 대한

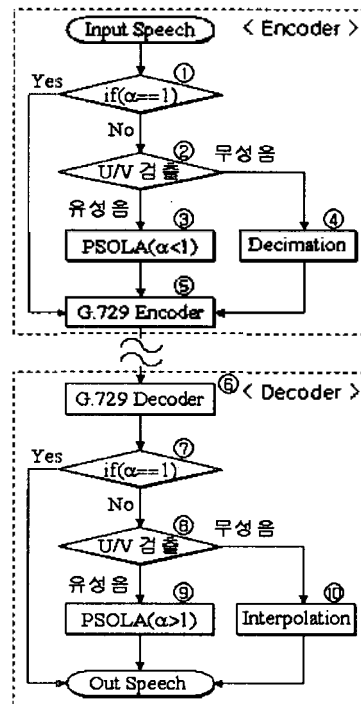


그림 3-1. 제안한 알고리즘의 블록도

한국어와 영어의 2가지 언어를 사용하였다. 음성시료는 8kHz로 표본화하고 16bit로 양자화하여 저장하였다. 그림 3-1은 본 논문에서 제안한 알고리즘을 나타낸 블록도이다.

그림 3-1에서 윗쪽의 점선부분은 부호화단을 나타내고 있으며, 아래쪽의 점선부분은 복호화단을 나타낸 것이다.

3.1. 부호화단

① 원음성 시료가 입력되면 3가지의 전송율을 결정하는 블록단이다. 속도 변화율 $a(Ss/Sa)$ 의 값이 1일 경우에는 기존의 G.729의 전송율인 8kbps를 위해 그대로 인코더 단에서 처리하게 된다. 여기서, Ss 는 합성을 위한 단위이고, Sa 는 분석을 위한 단위를 나타낸다. 그러나, a 의 값이 1보다 작을 경우에는 6kbps와 4kbps의 전송율을 위해 아래의 블록에서 PSOLA를 적용하게 된다.

② 유성음과 무성음을 결정하는 블록단이다. 유/무성음의 검출을 위해 에너지와 ZCR을 사용하였다. 만약 유성음이 검출되었을 경우에는 PSOLA 알고리즘을 적용하여 처리를 수행하게 된다. 그러나, 무성음이 검출되었을 경우에는 음질 향상과 처리 시간의 단축을 위해 다른 블록에서 처리한다.

③ 이 블록에서는 피치의 주기를 줄여주기 위한 PSOLA 알고리즘을 적용하는 블록단이다. 4kbps의 전송율을 위해서는 $a=0.5$ 이 필요하고, 6kbps의 전송율을 위해서는 $a=0.75$ 의 값이 필요하다. 이렇게 a 의 값에 대한 배율만큼 피치의 주기를 줄여준다.

④ 무성음 구간에서는 피치가 존재하지 않으므로 원하는 a 의 값에 대한 배율만큼씩 Decimation을 수행하여 처리한다.

⑤ G.729의 부호화단에 대한 처리는 그림 3과 같다. 프레임의 크기는 80샘플을 사용하고, 부분적으로 이것을 다시 40샘플의 서브프레임으로 나누어 처리를 수행한다. 그리고, LPC는 10차를 사용한다 [4].

3.2. 복호화단

⑥ 전송 또는 저장된 파라미터들을 다시 G.729로 부호화 시키는 블록단이다. 80bit의 파라미터를 입력으로 하여 복호화되고, 80샘플로 음성을 출력한다 [4].

⑦ a 의 값이 1일 경우에는 8kbps의 전송율을 의미하므로 그대로 복호화된 음성 파일을 출력시킨다.

⑧ 복호화된 데이터를 유성음과 무성음으로 나누어 따로 처리하기 위해 유성음과 무성음을 검출한다.

⑨ 여기서는 부호화단과는 반대의 개념으로 전송율이 4kbps인 경우에는 $a=2$ 가 필요하고, 6kbps인 경우에는 $a=1.25$ 가 필요하게 된다. G.729로 복호화된 음성은 피치가 줄어든 상태이므로 이를 다시 복원하기 위해서 PSOLA 알고리즘을 이용하여 피치를 늘려준다.

⑩ 이 블록단 역시 무성음에 대해서만 처리를 하기 위해 a 의 비율만큼 Interpolation을 수행하게 된다.

그림 3-2는 원음성 신호의 파형을 나타낸 것이다. 그리고, 그림 3-3은 $a=0.75$ 인 경우에 PSOLA 알고리즘을 통해 피치가 줄어 G.729의 부호화단으로 입력되는 음성을 나타낸 것이다.

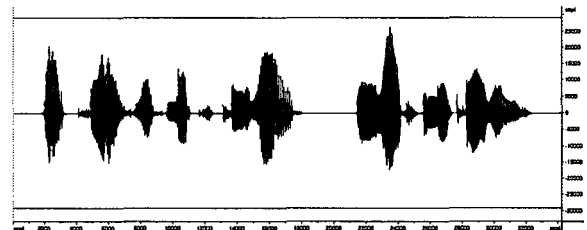


그림 3-2. 원음성 신호 파형

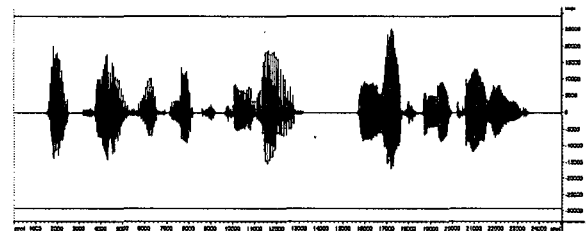


그림 3-3. $a=0.75$ 인 PSOLA 합성의 파형

그림 3-4은 G.729로 복호화된 음성의 파형이다. 또한 그림 3-5은 $a=1.25$ 만큼의 배율로 피치를 늘려준 최종 출력 파형을 보여준다.

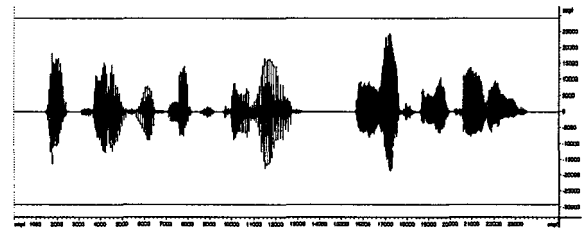


그림 3-4. G.729로 복호화된 파형

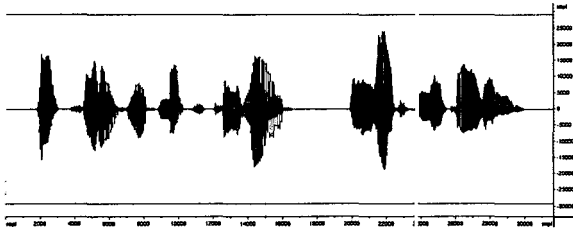


그림 3-5. $\alpha=1.25$ 인 PSOLA 합성의 파형

그림 3-6는 $\alpha=0.5$ 로 피치가 줄어든 파형을 보여 주고 있으며, 그림 3-7은 G.729의 복호화단을 거친 파형이고, 그림 3-8은 $\alpha=2$ 로 피치가 늘어난 최종 음성 파일을 나타내고 있다.

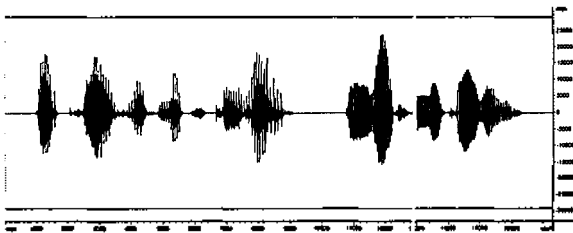


그림 3-6. $\alpha=0.5$ 인 PSOLA 합성의 파형

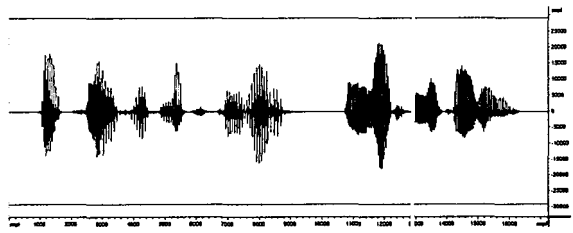


그림 3-7. G.729로 복호화된 파형

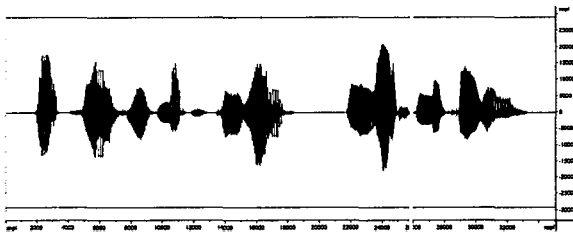


그림 3-8. $\alpha=2$ 인 PSOLA 합성의 파형

이렇게 하여 얻은 4개의 합성음을 사용하여 MOS 테스트를 수행한 결과가 아래의 표와 같다.

논문의 응용 범위는 음성 압축방법에 모두 적용 가능하다. 이러한 코덱들로는 (G.723.1, G.726, AMR, IMBE, Dod-CELP, QCELP, G.728, EVRC, VSELP 등이 있다. 본 논문의 알고리즘을 적용할 경우 위와 같은 코덱들이 기존의 전송율에서 1/2

배와 1/4배로 압축이 가능하다. 또한 원음성이 시간축에서 반으로 줄어들기 때문에 하드웨어 구현 시에도 계산시간이 줄어드는 장점이 있다.

표 3.1. MOS 점수

	8kbps	6kbps	4kbps
음성1	3.89	3.64	3.33
음성2	3.92	3.57	3.47
음성3	3.93	3.73	3.43
음성4	3.90	3.75	3.24
평균	3.91	3.67	3.37

V. 결 론

음성의 장시간 저장과 재생 등에서 현대에는 갈수록 많은 양의 메모리와 전송율을 요구하고 있다. 따라서, 본 논문에서는 이러한 문제를 해결하면서도 음질을 좋게 유지할 수 있는 방법을 제안하였다. 먼저 PSOLA 알고리즘을 통해 음성의 형태를 유지하면서 피치의 범위를 줄여 음성의 길이를 줄이면, G.729의 부호화기를 통해 저장되는 데이터의 양이 줄어든 상태에서 전송되거나 저장되어진다. 이렇게 저장되거나 전송된 데이터는 다시 G.729의 복호화기를 통해 음성이 출력되며, 이렇게 출력되는 음성을 다시 PSOLA 알고리즘을 통해 피치를 늘려주어 원음성을 합성하게된다.

MOS 테스트 결과 8kbps의 음성을 기준으로 6kbps에서는 0.24정도가 차이가 나며, 4kbps에서는 0.54정도의 음질 차이가 발생했다. 그러나, 모든 기존의 음성 압축 기법에 적용이 가능하며, 연산량도 줄어드는 큰 장점이 있다.

참고 문헌

- [1] M. Bae, "On a Pitch Alteration Method using Scaling the Harmonics Compensated with the Phase for Speech Synthesis," J., Acoust., Society, Korea, Vol.15, No.6, pp.99-103, December 1996.
- [2] M. BAE, "On the Pitch Alteration Methods for a High Quality Speech Synthesis", J., Acoust., Soc., Korea, Vol.12, No.2, pp.66-77, April 1993.
- [3] 함명규, 나덕수, 정찬중, 배명진, "비대칭 Weighting을 사용한 음성 피치변경법" 전자공학회 하계종합학술대회, 1998년6월27일.
- [4] Coding of speech at 8kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP), ITU-T Recommendation G.729