음성신호의 발성율과 PSOLA 기법을 적용한 음성 보코더 전송률 개선에 관한 연구

장 경 아, 서 지 호, 배 명 진
숭실대학교 정보통신공학과, *숭실대학교 전자공학과
전화 : (02) 824-0906 / 팩스: (02) 820-0018

Improvement of Bit Rate applying the Speaking Rate and PSOLA Technique of Speech in CELP Vocoder

KyungA Jang, JiHo Seo, MyungJin Bae
Dept. Information and Telecommunication Engr, *Dept. Electronic Engr,.
SoongSil University
kajang74@hotmail.com

Abstract

In general, speech coding methods are classified into the following three categories: the waveform coding, the source coding and the hybrid coding. Fast speaking is possiable to encode with a few information compared with slow speaking rate. In case of speakng rate, low frequency band is more important than high frequency band while listening. Speech vocoding technique is developing to way with low bit rate and complexity and high sound quality. the CELP type of vocoder support very good sound quality with low bit rate but these vocoders don't consider about the speaking rate. when we consider speaking rate and encode the frame depending on the speaking rate, the bit rate is able to reduce the bit rate than the converntional vocoder. We propose the technique to estimate the speaking rate and applied PSOLA technique in case of the frame of slow speaking rate. As a result of simulation bit rate can be reduced about 300 bps.

Keywords : 합성기, PSOLA, 발성율, 보코더

Ⅰ. 서론

현재까지 발표된 음성부호화기 중 가장 많은 연구가 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 인터넷폰 및 화상통신용 음성 부호화기로 ACELP/MP- MLQ(algebraic CELP/ Multipulse Maximum Likelihood Quantization)의 5.3/6.3kbps dual rate를 G.723.1 권고안으로 선정하였다. CELP계열의 부호화기인 G.723.1 5.3kbps ACELP를 기반으로 하여 음질을 유지하면서 전송률을 개선할 수 있는 방법으로 발성율과 이 발성율에 따라 PSOLA 기법을 적용시킨 알고리즘을 제안하고자 한다. 본 논문에서 적용한 파라미터는 흔히 음성인식 시스템에서 적용되는 발화속도와 음성합성시 고려되는 파라미터로써 사용되는 PSOLA 기법을 보코더 내에서 입력되는 음성의 발성율을 측정하여 발성율이 임의로 설정되어진 기준 발성율보다 높은 경우엔 PSOLA 방법으로 압축하여 전송률을 개선시키고자 한다. 논문의 2장에서는 발화속도의 정의와 관련된 연구들을 소개하고, 3장에서는 PSOLA 기법의 정의에 대해 간략한 소개를 하겠다. 4장에서는 본 논문에서 사용한 발성율 측정법과 PSOAL 기법을 포함한 알고리즘을 소개하고 5장에서는 실험 및 결과에 대해 소개하고 마지막으로 결론을 내리겠다.

Ⅱ. 발성율에 관한 관련된 연구

1. 발성율 정의

발성율에 대한 연구는 중요성이 부각되기 시작하면서 많이 이루어지고 있지만, 아직까지는 이에 대한 정의에 대해서 단일한 의견이 나와있지 않다. Pfitzinger는 Global Speech Rate, Local Speech Rate, Relative Speech Rate로 구별하여 설명하고 있다.

■ Global Speech Rate

전역 발화속도는 단위시간에 발화된 음절들이나 모라(Mora)와 같은 특정 음성학적 단위들의 개수에 의해서 정의된다.

■ Local Speech Rate

부분 발화속도는 전역 발화속도와는 다르게 명확히 정의 되어 있지 않으며, 일반적으로 음성파형이나 주파수 스펙트럼을 통해서 얻어지는 분절들의 지속시간으로 나타낸다.

■ Relative Speech Rate

상대적 발화속도 측정은 주어진 음성과 언어학적으로 동일한 기준 음성과의 대응되는 부분의 비율을 이용하는 방법이다.

2. 발성율 단위

▷ 단어속도 (Word Rate)

가장 단순한 발화 속도 측정 단위로써 1분 혹은 1초간 발화된 단어의 개수로 정의된다.

$$WordRate = \frac{N_w}{L_T} \quad (1)$$

N_w: 단어개수

L_T: 1분 혹은 1초동안 발화된 총 길이

단어속도는 각 단어에 대한 길이 및 구조의 예측이 불가능하고 단어들 사이의 단락(pause)의 불확정성 때문에 정확한 값을 얻기 힘들다.

▷ 평균 음소 속도 (Averate Phone Rate)

음소속도는 음소 지속시간의 역으로 나타낸다. 발화된 음성 내에서 음소들간의 경계를 자동으로 결정하는 것은 매우 어렵기 때문에 음소들의 개수에 대한 평균 음소 속도를 이용한다.

$$Average\ \ Syllable\ \ Rate = \frac{N}{\sum_{i=1...N} d_i} \quad (2)$$

N = 음소개수

L_T: i번째 음소구간 길이

정밀하게 발화속도를 측정할 수 있지만, 음소 경계를 결정하는 것은 인식과정과 비슷한 계산량을 필요로 한다.

▷평균음절속도 (Average Syllable Rate)

음절속도는 음소 속도와 마찬가지로 음절 지속시간의 역으로 나타낸다. 음절 속도의 경우 음절의 개수가 많고 각 음절마다의 길이가 틀리기 때문에 음절 속도 보다 평균 음절 속도를 이용하는 것이 더욱 효과적이다.

$$Average\ \ Syllable\ \ Rate = \frac{N_s}{L_T} \quad (3)$$

N_s: 음절개수

L_T: 전체 발화길이

음절은 발화시 인간이 인지할 수 있는 가장 기초적인 단위로써 하나의 음절에는 반드시 모음 하나를 포함하고 있다. 본 논문에서는 음절 속도를 이용하여 발화속도를 측정한다.

Ⅲ. PSOLA 피치변경 합성방식

본 논문에서는 음성신호를 복원할 때 스펙트럼 왜곡률과 복잡성이 적은 Pitch Synchronous Overlap and Add 방법이 적합하다. 전송 또는 압축된 파형과 진폭정보와 피치정보를 이용하여 PSOLA합성을 수행한다.

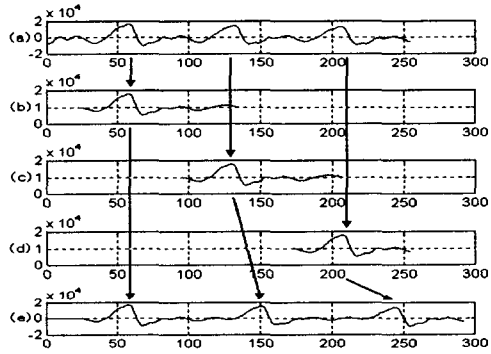


그림 1. TD-PSOLA의 피치 신장 합성법
(a) 원음성 신호, (b) 단구간 신호 1
(c) 단구간 신호 2, (d) 단구간 신호 3
(e) 피치가 신장된 합성 신호

VI. 제안한 알고리즘

본 논문에서 사용되는 파라미터는 LSP 변화도를 이용한 발성율과 PSOLA 기법의 압축 파라미이다. 여기서 발성율 측정하고자 하는 방법은 기존의 연구와 달리 처리시간과 알고리즘 자체가 간단하게 측정할 수 있도록 하였고 PSOLA 기법은 PSOLA 방법 중 TD-PSOLA 방법을 사용하였다. 보코더 자체 내부의 처리이나 계산량이 많으므로 인해 처리시간이 많이 소모되어, 적용한 알고리즘을 적용시킴으로써 보코더 내에서 계산량이나 처리시간을 더 부가시키지 않을 수 있는 알고리즘을 사용하였다.

1. 발성율 측정법

1.1 LSP 거리 계산을 통한 발성속도 측정

본 논문에서 고려하는 발성속도는 묵음 부분이 제거된 음성신호에서의 발성속도이다. 발성속도 측

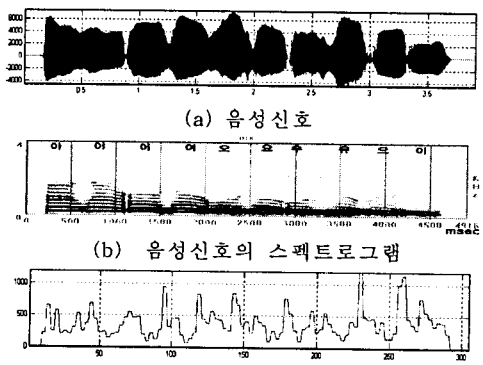
정에 있어 묵음 구간이 고려된다면 실제의 빠른 발성에 대해서도 다른 결과를 나타내게 된다. 따라서 유효한 발성속도를 측정하기 위해서는 음성구간의 검출이 먼저 선행되어야 한다. 본 논문에서는 먼저 묵음구간의 에너지와 LSP 파라미터를 정보를 이용하여 음성 검출을 수행하고, 파라미터를 추출하는 묵음구간은 발성시료의 처음부분을 이용하였다. 이 구간에서 음성 검출에 필요한 묵음 데이터를 추출하여 이후에 나타나는 음성구간과 비교하였다. 8KHz의 샘플링신호에서 처음 150msec의 신호를 묵음으로 간주하여 에너지와 LSP 데이터를 추출하고, 구해진 에너지의 200%를 에너지 문턱값으로 결정하였다.

1.2 인접 구간의 LSP 거리 측정

본 논문에 LSP 거리를 구하기 전, 먼저 음성신호의 발성모델에 의해 음성신호의 스펙트럼은 무성음을 제외하고는 짧은 시간동안 변화하지는 않는다. 따라서 60msec동안의 평균 LSP 값을 사용하여 거리를 측정하기 위해 유클리디안 거리측정법을 사용하였다. (a)는 음성신호의 시간 영역 파형이고, (b)는 스펙트로그램으로 가로축은 시간이고 세로축은 주파수 값이다. (c)는 식 4.1과 같이 계산한 LSP 거리를 나타낸다. 발성한 음성시료는 /아야어여오요우유으이/로 유성음인 모음이다.

$$D(n) = \frac{1}{P}\sum_{i=0}^{P} |LSP_n(i) - LSP_{n+1}(i)|^2 \qquad (4)$$

D(n)는 n번째 분석구간과 n+1번째 분석구간과의 LSP 거리를 나타내고, P는 LSP 분석 차수이다.



(a) 음성신호

(b) 음성신호의 스펙트로그램

(c) LSP 거리 변화

그림 2. LSP 거리 측정

그림(b)에서 각 모음에 해당하는 영역에 텍스트를 나타내어 보았다. 그림(c)와 비교하면 거리가 크게 나타나는 영역이 각 모음의 경계 영역임을 알 수 있다. 따라서 음소의 변화를 추정할 때 큰 거리의 차이를 보이는 영역을 검출하게 된다.

1.3 발성속도 계산

입력음성의 발성속도를 계산하기 위해서는 먼저 현재 처리되는 분석구간이 묵음인지 판정해야 된다. 묵음의 판정은 미리 구한 에너지 문턱값과 LSP 파라미터를 이용한다. 묵음으로 간주된 구간은 발성속도 계산에서 제외된다. 묵음 판정이 끝난 후 인접 분석구간과의 LSP 거리를 측정한다. 측정된 거리 값이 문턱값을 넘는 경우는 음소의 변화가 일어난 것으로 판정하고 이전에 음소가 변화된 구간에서 진행된 시간을 계산한다.

$$SPR = \frac{Fs}{VST(n) - VST(n-1)} \qquad (5)$$

1.4 무성음으로 시작되는 구간의 처리

본 논문에서는 무성음으로 시작되는 음소에 대해 인접한 유성음과 함께 묶어 발성속도를 측정하였다. 이렇게 발성속도를 구하기 위해서는 먼저 현재의 분석구간이 무성음인지 유성음인지 구분하여야 한다. 가장 작은 간격을 나타내는 선스펙트럼쌍의 위치가 2KHz 이상의 고주파수 영역에서 나타나면 무성음으로 간주하였다. 단 사전에 묵음 판정이 모든 분석구간에서 수행되므로 묵음이 아닌 구간에 대해서 이에 해당되는 조건을 검사하게 되는 것이다.

2. 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. 본 논문에서는 TD-PSOLA를 사용하였다. 이 방식은 파라미터를 사용하는 합성 방식과는 달리 사전에 저장된 음성신호를 부드럽게 연결하여 합성을 하므로 합성음의 음질이 좋다[5][6]. PSOLA 피치변경 합성법의 자세한 과정은 크게 두 단계로 나누어진다. 첫째는 피치를 찾아 윈도우를 씌우는 분석과정과 이렇게 얻어진 단구간 신호열들을 피치 조절하고 다시 합성 과정이다[7].

V. 실험 및 결과

본 논문에서 제안한 방법을 시뮬레이션하기 위해 IBM-PC/PentiumIV(1GHz)에 마이크 입력이 가능한 16비트 A/D변환기를 인터페이스하여 8kHz의 표본화 율로 16비트 양자화하여 저장하였다. 처리결과의 성능을 측정하기 위해 다음의 대표적인 문장을 연령층이 다양한 남녀 5명의 화자가 각 5번씩 발성하여 시료로 사용하였다. 시료는 두드러진 피크를 가지지 않고 잡음이 30dB를 가진 방

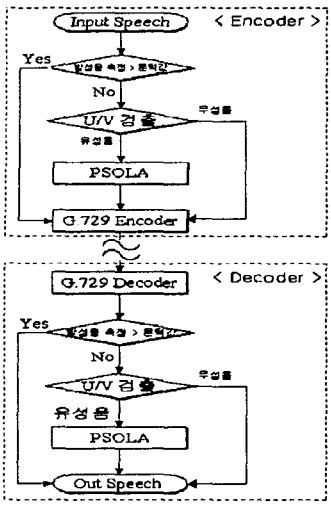


그림 2. 제안한 알고리즘

에서 녹음하였다.

발성1: "인수네 꼬마는 천재소년을 좋아한다."
발성2: "창공을 날으는 인간의 도전은 끝이 없다."
발성3: "예수님께서 천지창조의 교훈을 말씀하셨다. "
발성4: 일기예보 아나운서 음성시료

제안한 방법을 C-언어와 MATLAB으로 구현하여 5.3kbps ACELP (ITU-T 표준안 G.723.1) 보코더에 적용하였다. 표 1은 G.723.1의 보코더의 전송률과 LSP 파라미터를 이용한 발성율을 고려하여 PSOLA 기법을 적용시켜 음성부호화한 보코더의 보코더의 전송률을 비교한 것이다. 전송률은 제안한 방법이 기존의 5.3kbps ACELP보다 5.6%(280bps) 감소하였고 음질 열하는 거의 없었다.

표 1. 전송률 비교

	G.723.1 (5.3kbps)	Proposed Method	Degradation bps
발성 1	5.299	5.019	0.28
발성 2	4.759	4.457	0.302
발성 3	5.245	4.944	0.301
발성 4	5.019	4.748	0.271

VI. 결 론

CELP 부호화기는 선형 예측 합성에 의한 분석 부호화의 원칙에 기본을 두고 있다. 이 중 G.723.1은 5.3/6.3kbps의 이중 전송률을 갖는 구조로 되어있다. 그러나 G.723.1 역시 음성신호를 성분 분리하여 합성하는 방식인 CELP 보코더 계열의 합성에 의한 분석방법을 사용하기 때문에 많은 계산량으로 인한 처리 시간의 소모를 피할 수 없다는 문제점을 갖고 있다. 논문에서는 G.723.1 5.3kbps ACELP를 기반으로 하여 음질을 유지하면서 전송률을 5kbps정도로 낮출 수 있는 새로운 부호화 방법을 제안한다. 음성의 발성속도가 빠른 경우에는 발성속도가 느린 경우보다 적은 정보만으로도 부호화가 가능하다. 현재 상용화되고 있는 CELP형 보코더는 낮은 전송률에 비해 우수한 음질을 제공하지만, 기존 방식은 음성의 발성속도에 대해서 처리를 달리하지 않고 사용하고 있다. 본 논문에서는 CELP 부호화기의 전송률 감소를 위해 LSP 파라미터를 사용한 발성율 측정법을 이용하고, 발성율 따라 PSOLA 방식을 적용하여 간단한 알고리즘으로 음성을 압축하는 제안하였다. 제안한 방법을 G.723.1 5.3kbps ACELP에 사용하여 약 300bps의 전송률을 낮출 수 있었고 음질의 열하는 거의 없었다.

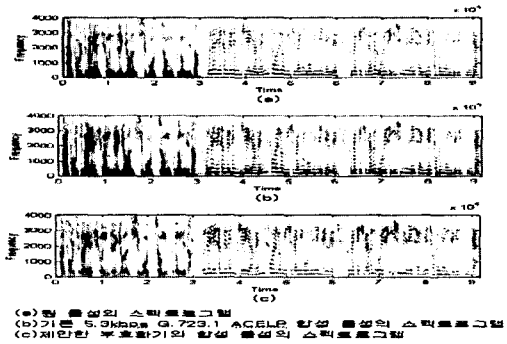


그림 3. 스펙트로그램 비교
/일기예보 아나운서 음성시료/

[참고문헌]

[1] H. R. Pfitzinger, " Local Speech Rate as A Combination of Syllable and Phone Rate", Proc., LCSLP' 98, vol. 3, pp. 1087-1090, Sydney. 1998

[2] H. R Pfitzinger, " Two Approaches to Speech Rate Estimation", Proc., SST' 96, PP. 421-426, Adelaide. 1996

[3] 박준배 외 3명, "음절 발화속도 보상을 이용한 한국어 연속음 음성인식 성능 향상", 신호처리 합동학술대회, 전자공학회, 2002

[4] 강경아, "LSP 파라미터를 이용한 발성속도 측정에 관한 연구", 석사학위 논문, 숭실대, 2001

[5] 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.

[6] 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.

[7] 함명규, 나덕수, 정찬중, 배명진, "비대칭 Weighting을 사용한 -음성 피치변경법" 전자공학회 하계종합학술대회, 1998년6월27일.