
비트스트림의 구조 분석을 이용한 음성 부호화 방식 추정 기법

유훈* · 박철순** · 박영미** · 김종호***

Blind Classification of Speech Compression Methods using Structural Analysis of Bitstreams

Jongho Kim* · Cheol-Sun Park** · Young Mi Park** · Hoon Yoo***

이 논문은 국방과학연구소의 지원에 의한 연구 결과임

요 약

본 논문에서는 임의의 음성 압축 비트스트림의 구조를 분석하여 음성 신호의 부호화 방식을 추정 및 분류하는 기법을 제안한다. 저 비트율 전송 및 저장을 위하여 다양한 코덱 방식의 음성 압축 기법이 개발되었는데, 이들은 블록 구조를 반드시 포함하고 있다. 각 부호화 방식을 구분하는데 있어, 본 논문에서는 Measure of Inter-Block Correlation (MIBC)를 이용하여 블록 구조의 유무 및 신호 블록의 길이를 파악하고, 블록 길이가 동일한 부호화 방식의 경우 각 부호화 방식마다 압축 스트림 내의 각 비트 위치별로 상관도 분포가 다르다는 점을 이용하여 해당 부호화 방식을 정확하게 추정하는 기법을 제안한다. 실험 결과 제안한 비트스트림 분석 기법은 다양한 음성 신호의 종류, 음성 신호의 길이 및 잡음 환경에 강인한 검출 능력을 나타냄을 보인다.

ABSTRACT

This paper addresses a blind estimation and classification algorithm of the speech compression methods by using analysis on the structure of compressed bitstreams. Various speech compression methods including vocoders are developed in order to transmit or store the speech signals at very low bitrates. As a key feature, the vocoders contain the block structure inevitably. In classification of each compression method, we use the Measure of Inter-Block Correlation (MIBC) to check whether the bitstream includes the block structure or not, and to estimate the block length. Moreover, for the compression methods with the same block length, the proposed algorithm estimates the corresponding compression method correctly by using that each compression method has different correlation characteristics in each bit location. Experimental results indicate that the proposed algorithm classifies the speech compression methods robustly for various types and lengths of speech signals in noisy environment.

키워드

음성 신호 압축, MIBC, 상관도 분석, 비트스트림 구조 분석

Key word

Speech compression, MIBC, correlation analysis, structural bitstream analysis

* 중신회원 : 상명대학교 디지털미디어학부

** 정회원 : 국방과학연구소

*** 정회원 : 순천대학교 멀티미디어공학과 (교신저자, jhkim@sunchon.ac.kr)

접수일자 : 2011. 08. 25

심사완료일자 : 2011. 09. 15

I. 서 론

무선 통신 시스템의 음성 채널 대역폭은 매우 제한되어 있어서 효과적인 전송을 위해서는 압축 부호화 과정이 필수적이다. 음성 부호화 방식은 음성 신호를 디지털화하여 얻어진 64kbps의 PCM (Pulse Code Modulation) 방식부터 음성 신호의 특성을 분석하여 저 비트율로 압축하는 보코더(vocoder) 방식에 이르기까지 통신방식 및 환경에 따라 매우 다양한 기법이 존재한다[1]-[3]. 통신 시스템을 분석하기 위해서는 변조(modulation), 인터리빙/스크램블링(interleaving/scrambling), 채널 부호화(channel coding) 방식 등의 다양한 요소기술을 포함하여 음성 부호화 방식을 알아낼 필요가 있는데, 다양한 음성 압축 방식이 존재하는 환경에서 임의의 압축 비트스트림을 사전 정보 없이 그 부호화 방식을 추정하기는 매우 어렵다.

통신이 이루어지기 위해서는 송·수신 시스템에서 사용되는 기술이 서로 약속되어 있어야 하고, 이는 국제 표준 등으로 규정되어 있다[4]. 따라서 통신 시스템의 음성 부호화 방식을 알아내기 위해서는 이들 사이의 공통된 특성을 기반으로 음성 압축 방식을 추정할 수 있다. 즉 각 비트스트림의 구조를 분석함으로써 음성 부호화 방식에 공통적으로 사용되는 특성을 추정하고 그룹화하여 추정해야 할 부호화 방식의 개수를 줄이고, 각 그룹 내에서 해당 부호화 방식에서만 나타나는 특성을 찾아내어 최종 부호화 방식을 확정하는 계층적 방식으로 추정 오류를 줄이게 된다.

본 논문에서는 보코더 방식의 음성 부호화 방식에서 공통적으로 사용되는 블록 구조의 유무를 판단하고, 블록의 길이를 추정하기 위하여 MIBC(Measure of Inter-Block Correlation)를 정의하고, 이를 음성 압축 비트스트림에 적용하는 방법을 제안한다. 또한 MIBC에 의해 그룹화된 부호화 방식들 내에서 비트스트림의 비트 위치별 상관도(correlation)를 이용하여 특정 부호화 방식을 확정하는 방법을 제안한다. 이는 보코더 방식에서 예측 신호를 만들기 위해서는 특정 길이를 갖는 블록 구조를 포함해야 하고, 각 블록 내에서는 표준 규격에 따라 특정 위치에서 특정 비트 패턴이 위치한다는 점을 이용한 것이다.

논문의 구성은 다음과 같다. 2장에서 보코더 방식의 주요 음성 압축 기법 및 그 특징에 대해서 간략하게 기술

한다. 3장에서는 제안하는 MIBC를 정의하고, MIBC 및 상관도 분석에 의해 임의의 음성 압축 스트림의 부호화 방식을 계층적으로 추정하는 방법을 제시한다. 4장에서 다양한 음성 압축 알고리즘에 의해 부호화된 여러 음성 데이터에 대해서 제안한 알고리즘의 성능 평가 결과를 보이고, 5장에서 결론을 맺는다.

II. 음성 신호의 부호화 기법

보코더 방식의 음성 부호화 기법은 인간의 발성 기관을 모델링하여 합성된 음성 신호 또는 파라미터를 부호화하는 방식으로 이루어진다. LPC (Linear Predictive Coding)의 경우 피치(pitch) 주기, 유성음/무성음 등의 정보 및 과거의 신호 샘플을 선형 예측하여 만든 신호와 현재 샘플의 오차가 가장 작도록 하는 필터의 계수 등을 부호화하여 표준 규격에 규정된 형식으로 압축 비트스트림을 구성한다[1], [2], [5]. CELP (Code Excited Linear Prediction)의 경우에는 LPC 필터에 의해 예측 신호를 생성하고, 발성 기관을 모델링하여 얻어진 예측 후보 신호의 모입인 부호 테이블(codebook)의 신호와 오차가 최소가 되는 부호어의 인덱스를 전송하는 방식으로 전송속도를 최소화하도록 구성된다. CELP 방식은 LPC에 비해 음질이 향상되고, 채널 오류에 강한 반면, 예측 신호를 구성하고 부호 테이블에서 최적의 신호 블록을 탐색하는데 복잡도가 증가하는 단점이 있다[6], [7]. 이러한 단점을 극복하기 위하여 최근 MELP (Mixed Excitation Linear Prediction) 방식이 개발되었는데, 이는 기존의 여기(excitation) 신호로 피치를 주기로 하는 임펄스 신호열(impulse train)을 사용함으로써 합성된 음질이 열화되는 단점을 보완하기 위하여 음성 신호의 특성에 따라 주기 또는 비주기 신호를 여기 신호로 사용하여 음질을 향상시키고, CELP와 비교하여 복잡도를 현저히 감소시키는 방법을 사용하였다[8].

이러한 보코더 방식의 예측 부호화 기법은 공통적으로 블록 구조를 포함하는데, 이는 부호화기 내부에 음성 신호의 특징을 분석하여 예측 신호를 생성하기 위하여 특정 길이의 블록 단위로 처리하기 때문이다. 압축 비트스트림 또한 블록 구조를 포함하고 있고, 각 블록을 구분하기 위하여 구분자(separator)를 포함하고 있는데, 이는 보코더 부호화 방식의 중요한 특징이 된다.

따라서 이러한 특징을 이용하여 본 논문에서 부호화 방식을 추정할 때 가장 먼저 비트스트림 내에 블록 구조를 포함하는지의 여부를 판별하고, 그 블록의 길이를 알아내어 추정해야 할 부호화 방식의 개수를 줄이게 된다.

III. MIBC 및 상관도에 기반한 음성 압축 스트림의 구조 분석

본 논문에서는 임의의 음성 압축 스트림으로부터 해당 부호화 방식을 추정하기 위해서 다음과 같은 두 가지 조건을 고려한다.

- 부호화 방식을 추정하기 위하여 특정 파라미터를 이용하거나 부분적 복호화 과정을 거치지 않고 음성 압축 스트림만을 이용하여 해당 부호화 방식을 추정한다.
- 음성 압축 스트림의 분석 및 해당 복호화 과정의 실시간성을 보장하기 위하여 부호화 방식 추정 알고리즘은 저 복잡도를 갖도록 한다.

음성 압축 스트림의 부호화 방식을 추정하는데 있어 특정 부호화 파라미터를 이용하지 않기 때문에 먼저 음성 신호 부호화에 대한 일반적인 정보에 기반하여 추정할 부호화 방식의 경우의 수를 줄여 나간다. 즉, 보코더의 경우 부호화 과정이 블록 단위로 이루어지고 압축 스트림 또한 블록 구조를 갖기 때문에 각 블록을 구분하기 위한 구분자가 필요하게 되는데, 압축 스트림을 분석하여 이러한 구분자가 존재할 경우 보코더 방식으로 부호화된 스트림으로 판단한다.

이후 각 방식 내에서 구체적인 부호화 방식을 추정하게 되는데, 보코더 방식의 해당 부호화 방식 및 비트율에 따라 블록의 길이가 결정되고, 이러한 블록 길이는 각 부호화 방식의 가장 뚜렷한 특징이므로 압축 스트림을 분석하여 블록 길이를 적절히 추정하면 해당 부호화 방식을 결정할 수 있다. 또한 각 블록 내에서는 부호화 방식에 따라 특정 비트 위치에서의 상관도의 분포가 달라지는데, 이러한 상관도 분포를 분석하여 최종적으로 부호화 방식을 확정한다.

3.1. MIBC를 이용한 sync bit 검출

보코더 방식에 의한 음성 압축 스트림은 각 블록을 구분하는 구분자로 sync bit를 삽입하는데, 구체적으로 매 블록의 끝에 0과 1을 교대로 삽입한다(0과 1의 순서는 부호화 방식에 따라 다르나, 그 순서가 특정 부호화 방식을 의미하지는 않는다). 따라서 본 논문에서 제안하는 압축 비트스트림의 구조 분석에 의한 음성 부호화 방식 추정을 위해서는 이러한 sync bit의 유무와 블록의 길이를 올바르게 추정해야 한다.

음성 압축 스트림 $s(n)$ 에 대해서, i 번째 블록 $B_i(n)$ 는 다음과 같이 표현할 수 있다.

$$B_i(n) = s(i \times L_B + n), n = 0, \dots, L_B - 1 \quad (1)$$

이때 L_B 는 블록의 길이를 의미하고, n 은 스트림에서 각 비트의 위치를 나타낸다. 식 (1)과 같이 정의된 블록에 대해서 sync bit는 0과 1이 규칙적으로 반복되는 비트이므로 식 (2)와 같이 MIBC를 정의하여 그 결과값에 따라 sync bit의 유무와 블록의 길이를 추정한다.

$$C_B(n) = \frac{1}{N} \sum_{i=0}^{N-1} (B_i(n) \oplus B_{i+1}(n)) \quad (2)$$

$(n = 0, \dots, L_B - 1)$

여기서, \oplus 은 exclusive-OR 연산을 나타내고, N 은 블록의 개수, L_B 는 블록의 길이, n 은 블록에서의 각 비트의 위치를 의미한다. 식 (2)의 MIBC는 0과 1 사이의 값을 갖고, 각 값에 따라 의미하는 바는 표 1에 나타난 것과 같다.

표 1. MIBC 값과 그 의미
Table. 1 MIBC values and their meanings

$C_B(n)$	의미
0	각 블록의 n 번째 비트가 동일한 값
0.5	각 블록의 n 번째 비트가 임의의 값
1	각 블록의 n 번째 비트가 0과 1이 반복되는 패턴의 값

즉, 특정 블록 길이에 대하여 MIBC를 적용하였을 때, 그 값이 1이면 sync bit가 존재하므로 보코더 방식으로 부호화된 스트림이고, 그 때 블록의 길이는 L_B 임을 알 수 있다.

3.2. 비트 위치별 상관도 분석에 의한 추정

MIBC는 블록 내 각 비트 위치에서의 상관도를 나타내는 것으로 해석할 수 있다. 즉 MIBC 값이 0에 가까울수록 해당 비트 위치에서 같은 값이 반복되므로 상관도가 높고, 0.5에 가까울수록 임의의 패턴으로 비트가 발생하므로 상관도가 낮다. 이러한 특성에 기반하여 보코더 부호화 범주 내의 특정 부호화 방식을 추정할 수 있다. 53bits의 블록 길이를 갖는 LPC 및 MELP 스트림의 경우 표준에서 정의하는 비트스트림 포맷, 즉 신텍스(syntax)의 배열 특성 및 부호화 방식의 특성에 의해서 각 비트 위치별 상관도가 달라지는데, 이에 따라 본 논문에서는 두 부호화 방식을 구분하기 위해 다음과 같이 식 (3)을 정의한다.

$$\begin{cases} LPC(L_B = 54): (C_B(3) < C_B(6)) \wedge \\ (C_B(3) < C_B(4)) \wedge (C_B(3) < C_B(2)) \\ MELP(L_B = 54): (C_B(6) < C_B(3)) \wedge \\ (C_B(6) < C_B(7)) \wedge (C_B(6) < C_B(4)) \end{cases} \quad (3)$$

이때 \wedge 는 논리적 AND 연산을 나타내고, $C_B(n)$ 는 L_B 를 블록 길이로 할 때 n 번째 비트 위치에서의 MIBC 값을 의미한다. 그림 1은 LPC와 MELP의 각 비트 위치별 MIBC 값의 분포를 나타낸다.

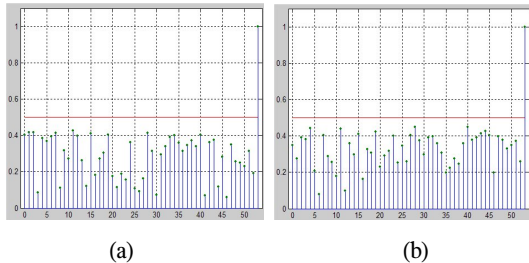


그림 1. MIBC 값의 분포: (a) LPC, (b) MELP
Fig. 1 Distribution of MIBC values: (a) LPC, (b) MELP

본 논문에서 제안하는 MIBC를 이용하여 블록 구조 및 블록 길이를 추정하고, 블록 내 각 비트 위치에서의 상관도를 분석하여 부호화 방식을 확정하는 방식의 비트스트림의 구조 분석에 의한 계층적 부호화 방식 추정 알고리즘을 정리하면 그림 2와 같이 나타낼 수 있다.

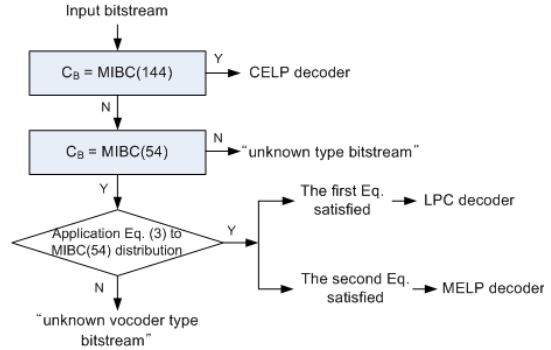


그림 2. 제안하는 비트스트림 구조 분석에 의한 부호화 방식 추정 알고리즘
Fig. 2 Block diagram of the proposed structural analysis algorithm for classification of the speech compression method

IV. 실험 및 결과

본 논문에서 제안한 음성 신호 부호화 방식 추정 알고리즘의 성능을 검증하기 위하여 8kHz 샘플링, 샘플당 16bits로 균일 양자화된 5개의 음성 신호를 표 2에 정리한 각 부호화 방식으로 압축한 스트림을 이용하였다.

표 2. 성능 평가를 위한 대상 부호화 방식
Table. 2 Target coding types for performance evaluation

부호화 방식	비트율 (kbps)	블록 길이 (bits)
LPC	2.4	$L_B = 54$
CELP	4.8	$L_B = 144$
MELP	2.4	$L_B = 54$

성능평가는 두 가지 측면에서 수행하였는데, 먼저 음성 신호의 길이에 따른 성능 변화를 살펴보기 위하여 각각 30.0sec, 15.0sec, 7.5sec, 3.75sec, 1.875sec, 0.514sec 길이를 갖는 5가지 종류의 음성 신호를 표 2의 방식으로 압축한 스트림을 이용하여 부호화 방식을 추정하였고, 다음으로 잡음의 영향에 따른 성능을 검증하기 위하여 각 길이의 음성 압축 스트림에 균일 분포 (uniform distribution) 잡음을 10^{-3} , 10^{-2} , 2×10^{-2} , 10^{-1} 의 비율로 첨가

하여 부호화 방식을 추정하였다. 그림 3에 실험에 사용한 30.0sec 길이의 음성 신호의 예를 부분 확대하여 나타내었다.

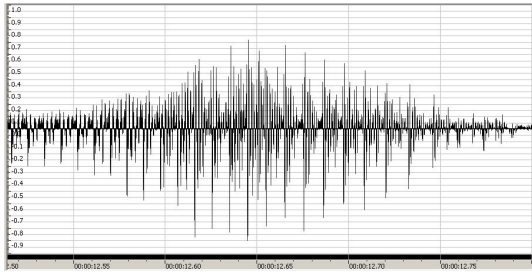


그림 3. 성능 평가에 사용한 음성 신호의 예
Fig. 3 One of speech signals used in the performance evaluation

실험결과 5개의 음성 신호의 길이가 각각 30.0sec, 15.0sec, 7.5sec, 3.75sec, 1.875sec의 경우에 대해서는 앞에서 언급한 모든 비율의 잡음 환경에서 100%의 검출율을 보였다. 이는 MIBC에 평균 개념이 포함되어 잡음의 영향을 줄여주기 때문이다. 또한 분석 대상이 되는 부호화 방식에서 블록의 길이가 최대 144 비트로 매우 짧은 길이의 음성 신호인 1.875sec에 비해서도 매우 짧기 때문이다. 따라서 제안한 알고리즘은 다양한 잡음의 영향 및 음성 신호의 길이와 같은 조건에 강한 검출율을 나타낼 수 있다.

잡음의 영향이 극대화된 환경에서의 검출 능력을 실험하기 위하여 음성 신호의 길이가 0.514sec일 때, 즉 4,112개의 샘플을 각 부호화 방식으로 압축한 비트스트림에 제안한 알고리즘을 적용하였고, 각 잡음 환경에서의 검출 결과는 표 3에 나타난 바와 같다.

표 3. 0.514sec 길이의 음성 신호에 대한 검출 결과
Table. 3 Classification results for the bitstreams of 0.514sec-length

에러율	Classification Ratio (correct/total)		
	LPC	CELP	MELP
error-free	5/5	5/5	5/5
10^{-3}	5/5	5/5	5/5
10^{-2}	3/5	5/5	CELP
2×10^{-2}	CELP	5/5	CELP
10^{-1}	1/5	3/5	CELP

표 3에서 LPC 및 MELP의 검출 결과에서 CELP라 표시된 부분은 입력 스트림이 LPC 또는 MELP 방식으로 부호화되었으나 CELP로 인식한다는 의미한다. 0.514sec 길이의 음성 신호를 각 방식으로 압축하면 매우 적은 양의 데이터를 얻게 되고, 이에 잡음을 첨가하게 되면 정보량에 비해 잡음의 양이 많아서 비록 MIBC에 평균 개념이 포함되어 있더라도 잡음의 영향을 줄이기 어려워지게 된다. 또한 매우 짧은 길이의 신호에 대해서는 MIBC보다는 비트 위치별 상관도 분석의 기법의 성능이 더 떨어지는데, 이는 0.514sec의 신호 길이가 지나치게 짧아 각 비트 위치에서 상관도가 충분히 나타나지 않기 때문이다. 따라서 MIBC와 상관도 분석 기법을 동시에 적용하는 LPC와 MELP에 대한 검출율이 CELP보다 더 낮음을 알 수 있다. 이와 같은 결과에도 불구하고, 0.514sec 길이의 음성 신호는 제안한 알고리즘의 성능 검증을 위한 극단적인 경우이고, 보통 이러한 길이의 음성 신호에는 충분한 양의 정보를 포함할 수 없으므로 제안한 알고리즘은 다양한 종류, 형태 및 환경에서 음성 부호화 방식을 추정하는데 매우 효율적인 알고리즘이라 할 수 있다.

V. 결 론

본 논문에서는 임의의 음성 압축 스트림의 구조를 분석하여 계층적으로 해당 음성 신호의 부호화 방식을 추정하는 알고리즘을 제안하였다. 코덱 방식의 핵심적인 특징인 블록 구조의 여부를 파악하고, 추정 부호화 방식의 가짓수를 제한하기 위한 블록 길이를 추정하기 위하여 MIBC를 정의하고, 이를 음성 압축 비트스트림의 구조를 파악하는데 적용하였다. 또한, 각 부호화 방식마다 블록 내의 비트 위치별 상관도 분포가 다르다는 점을 이용하여 각 비트 위치에 따른 MIBC 값의 분포를 분석하고, 이를 바탕으로 최종 부호화 방식을 정확하게 추정하였다. 실험 결과 제안한 MIBC는 평균 개념을 포함하고 있어 음성 신호의 길이 및 다양한 잡음 환경에 매우 강한 검출율을 보였고, 저 복잡도로 구현될 수 있어 다양한 분야에 효과적으로 응용될 수 있다.

참고문헌

[1] A.M.Kondoz, *Digital Speech Coding for Low Bit Rate Communications Systems*, Chichester, U.K.: John Wiley & Sons, 1994.

[2] L.R.Rabiner and R.W.Schafer, *Digital Processing of Speech Signals*, Englewood Cliffs, NJ: Prentice Hall, 1978.

[3] CCITT Rec. G.711, "Pulse Code Modulation (PCM) of Voice Frequencies," *CCITT Blue Book*, vol. III, Fascicle III.4, pp. 175-184, Nov. 1988.

[4] A.Spanias, "Speech coding: a tutorial review," *Proceedings of the IEEE*, vol. 82, no. 10, pp. 1541-1582, Oct. 1994.

[5] Wikipedia, Linear Predictive Coding (LPC), [Online]. Available: http://en.wikipedia.org/wiki/Linear_predictive_coding.

[6] M.R.Schroeder and B.S.Atal, "Code-excited linear prediction (CELP): high quality speech at very low bit rates," *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, pp. 25.1.1-25.1.4, 1985.

[7] Wikipedia, Code-Excited Linear Prediction (CELP), [Online]. Available: http://en.wikipedia.org/wiki/Code-excited_linear_prediction.

[8] Wikipedia, Mixed Excitation Linear Prediction (MELP), [Online]. Available: http://en.wikipedia.org/wiki/Mixed_Excitation_Linear_Prediction.

저자소개

유 훈(Hoon Yoo)

한국정보통신학회 논문지
제 13권 제1호 참조



박철순(Cheol-Sun Park)

1989년: 경기대학교 전자계산학과
학사
1991년: 인하대학교 전자계산
공학과 석사

1991년~현재: 국방과학연구소 책임연구원
1997년: 전자계산조직응용 기술사
2007년: 충남대학교 정보통신공학과 공학박사
※관심분야: 신호처리, 통신응용



박영미(Young Mi Park)

1991년: 충남대학교 전산학과 학사
1991년~현재: 국방과학연구소
1991년: C3I 개발
1995년~2012년: 육해공 전자전
장비 개발

※관심분야: 신호처리, 통신시스템



김종호(Jongho Kim)

2008년: 한양대학교 전자통신
공학과 공학박사
2008년~2009년: 삼성전자통신
연구소 책임연구원

2009년~현재: 순천대학교 멀티미디어공학과 교수
※관심분야: 영상/음성처리 및 압축, 디지털 신호처리