한국음향학회지 제28권 제3호 pp. 290-297 (2009)

AMR-WB 음성 부호화기를 이용한 TTS 데이터베이스의 효율적인 압축 기법

Efficient TTS Database Compression Based on AMR-WB Speech Coder

임 종 욱*, 김 기 출*, 김 경 선**, 이 항 섭*, 박 혜 영**, 김 무 영*
(Jongwook Lim*, Kichul Kim*, Kyeongsun Kim**, Hangseop Lee**, Haeyoung Park**, Moo Young Kim*)

*세종대학교 정보통신공학과, **(주)에이치씨아이랩
(접수일자: 2009년 2월 16일; 수정일자: 2009년 3월 11일; 채택일자: 2009년 3월 31일)

본 논문에서는 효율적으로 Text-To-Speech (TTS) 데이터베이스를 압축하기 위해서 개선된 adaptive multi-rate wideband (AMR-WB) 음성 부호화 알고리즘을 제안하고자 한다. 제안된 알고리즘은 불필요한 common bit-stream (CBS)을 제거하고, 파라미터의 델타 코딩 방식과 특정 화자에 종속적인 Huffman coding을 접목하여 음질 저하 없이 비트율을 낮추고자 하였다. 또한, 최소한의 음질 손실로 최대의 비트율 개선 효과를 볼 수 있는 손실 압축 방식도 제안하였다. 기존의 12.65 kbit/s AMR-WB 코덱에 CBS 제거를 포함한 무손실 압축 방식을 적용한 결과 음질 저하 없이 최대 12.40%의 비트율 개선 효과를 나타냈다. 또한, 손실 압축방식에서는 20.00% 비트율 개선 시 PESQ로 0.12 정도의 음질 저하가 발생했다.
핵심용어: TTS, AMR-WB, 하프만 코딩, 음성 부호화, 정보 이론
투고분야: 음성처리 분야 (2.1)

This paper presents an improved adaptive multi-rate wideband (AMR-WB) algorithm for the efficient Text-To-Speech (TTS) database compression. The proposed algorithm includes unnecessary common bit-stream (CBS) removal and parameter delta coding combined with speaker-dependent huffman coding to reduce the required bit-rate without any quality degradation. We also propose lossy coding schemes to produce the maximum bit-rate reduction with negligible quality degradation. The proposed lossless algorithm including CBS removal can reduce bit-rate by 12.40% without quality degradation compared with the 12.65 kbps AMR-WB mode. The proposed lossy algorithm can reduce bit-rate by 20.00% with 0.12 PESQ degradation.
Keywords: TTS, AMR-WB, Huffman coding, Speech Coding, Information theory
ASK subject classification: Speech Signal Processing (2.1)

I. 서 론

최신 디지털 기기들에서는 멀티미디어 데이터를 한정된 메모리에 효율적으로 저장하는 것이 중요한 이슈이다. 특히, Text-To-Speech (TTS) 시스템을 휴대용 멀티미디어 단말기 등에 탑재하기 위해서는 TTS 데이터베이스의 효율적인 압축이 필수적이다. 이를 위해서 대표적으로 화자에 의존한 피치-펄스 코드북을 사용하는 방식과 파형보간 음성 압축을 사용하는 방식 등이 연구되어 왔다 [1,2].

책임저자: 김 무 영 (mooyoung@sejong.ac.kr)
143-747 서울시 광진구 군자동 세종대학교 정보통신공학과
(전화: 02-3408-3695; 팩스: 02-3408-4330)

기존의 Moving Picture Experts Group (MPEG) 오디오 코더들은 discrete cosine transform (DCT) 등의 transform 계수들을 심리 청각 모델과 무손실 코딩 방식을 이용하여 효율적으로 압축하였다 [3,4]. 이런 transform 코더들은 음악 신호를 압축하는데 주로 사용되며, 비교적 높은 비트율에서 동작하도록 설계 되어 있다.

반면, 음성 부호화기는 Code-Excited Linear Prediction (CELP) 방식을 주로 사용하고 있으며, 비교적 낮은 비트율에서 동작하도록 설계 되어 있다. 예를 들어서 8kHz에서 동작하는 ITU-T G.729 conjugate-structure algebraic CELP (CS-ACELP) [5,6], 16kHz에서 동작하는 adaptive multi-rate wideband (AMR-WB) 방식 등이 있다 [7-9]. 음성 부호화기 설계에 있어서도 wavelet [10]이나 modulated

lapped transform (MLT) [11,12] 등과 같은 transform 계수들을 무손실 코딩 방식과 결합하여 압축하는 방식들도 제안되었지만, 비교적 높은 비트율에서 동작하는 단점이 있다.

이에 본 연구에서는 기존의 CELP 방식인 AMR-WB 에 무손실 압축방식인 Huffman coding을 사용하여 보다 더 낮은 비트율에서 동작하는 효율적인 음성 신호 압축방식에 대해 연구하고자 한다. 제안하는 TTS 시스템에서는 채널 에러가 거의 발생하지 않음을 가정하므로, 채널 에러에 취약한 무손실 압축 방식인 Huffman coding의 단점을 최소화하면서 동시에 원본 데이터의 손실 없이 비트율을 줄일 수 있을 것으로 예상된다. 즉, Huffman coding은 Entropy coding 방식이므로 채널 에러에 취약하여 통신용으로 개발된 speech coder들에는 사용되기 어려우며 [13], TTS 시스템이나 음성 저장 매체 등에서 유용하게 사용될 수 있다.

본 연구에서는 화자에 상관없이 첫 번째 서브 프레임과 세 번째 서브 프레임을 직전 프레임과 델타 코딩하는 방식으로 바꿈으로써 비트율 개선을 이루었다. 이상의 제안 알고리즘들은 모두 무손실 압축 방식이며 어떠한 왜곡도 발생시키지 않는다. 여기에 추가적으로 최소한의 왜곡을 감수하면서 최대의 효과를 볼 수 있는 손실 압축 방식도 연구하였다. 손실 압축 방식은 크게 두 가지로 하나는 TTS 시스템의 특성 상 제거해도 크게 왜곡이 생기지 않는 파라미터 2가지를 제거하여 비트율을 낮추는 방식이며 다른 하나는 algebraic codebook search에 사용되는 펄스의 수를 줄임으로써 음질 저하는 최소로 하면서 최대의 이득을 얻고자 하는 방식이다.

또한, 제안하는 TTS 시스템은 채널 에러가 거의 없는 환경에서 동작함을 가정하였으므로, 네트환경에 따라서 비트율을 바꿔가면서 동작하는 AMR-WB의 다양한 모드들을 모두 사용할 필요가 없다. 따라서, 기본적인 비트 포맷에서 payload를 제외한 헤더 정보들은 저장할 필요가 없게 된다. 따라서, 헤더 정보들을 제외한 알고리즘을 제안하여 비트율 절감 효과를 얻을 수 있다.

참고문헌 [1]에서는 화자가 한 명일 경우 기존에 사용하던 적응 코드북 대신 화자에 의존해서 음소별로 디자인된 고정 코드북을 사용하고 있다. 이 방식은 적응 코드북을 사용하지 않으므로, partial segment decoding이나 random access capability를 향상시킬 수 있었다. 본 논문에서는 [1]의 방식을 적용하지는 않았지만, [1]의 음소별 코드북에 Huffman coding을 적용한다면 좀 더 향상된 성능을 얻을 수 있을 것으로 기대된다.

본 논문의 구성은 다음과 같다. 2장에서는 기존 AMR-WB 알고리즘에 대한 설명이 있겠고, 3장에는 제안된 알고리즘들에 대한 자세한 설명이 있겠다. 그 후 4장에서는 기존 알고리즘과 제안된 알고리즘들에 대한 비교 실험 결과가 나타나 있다. 마지막으로 5장에는 최종 결론이 포함된다.

II. AMR-WB Algorithm

AMR-WB은 16 kHz sampling rate에서 동작한다. 한 프레임은 20ms에 해당하며 복잡도를 줄이고자 50-6400Hz 밴드는 CELP 방식으로, 6400-7000Hz 밴드는 band-width extension (BWE) 기법 등으로 분리해서 코딩한다. AMR WB은 네트워크 환경에 따라서 23.85, 23.05, 19.85, 18.25, 15.85, 14.25, 12.65, 8.85, 6.60 kbit/s의 총 9가지 비트율에서 동작하며, 본 연구에서는 12.65 kbit/s 모드를 선택하였다. 9가지의 모드들 중 12.65 kbit/s 이상의 비트율을 가지는 모드들은 광대역 음성에 대해서 높은 음질을 제공하며, 그 이하의 모드들은 네트워크 congestion등 특정 목적에 쓰이고자 디자인되었다. 따라서, 본 연구에서는 TTS 시스템 구현을 위해 최소한의 음질 수준을 보장하면서 비트율 관점에서는 최

표 1. 제안한 알고리즘의 실험 결과
Table 1. Experiment results of the proposed algorithm.

	bits /subframe	bits /frame	제안하는 알고리즘의 메모리 증가량 (* 4bytes)	
			개선 전	개선 후
VAD		1	2	2
LSP0		8	256	256
LSP1		8	256	256
LSP2		6	64	64
LSP3		7	128	128
LSP4		7	128	128
LSP5		5	32	32
LSP6		5	32	32
ADCB1	9+9	18	1024	512
ADCB2	6+6	12	128	64
LTPfil	1	4	2	2
ACELP0	8+1	36	512	256+2
ACELP1	8+1	36	512	
ACELP2	8+1	36	512	
ACELP3	8+1	36	512	
GAIN	7	28	128	128
total		253	4228	1862

대한의 이득을 얻을 수 있는 12.65 kbit/s 모드에서 실험을 진행하였다. 또한, TTS 데이터베이스는 음소별로 압축을 진행하게 되므로 별도의 voice activity detector (VAD)를 사용하지 않아도 되며, 모든 프레임을 음성구간으로 가정하게 되므로 비음성 구간을 코딩하기 위한 저전송율 압축 방식은 별도로 필요하지 않게 된다. 12.65 kbit/s 모드에서 사용되는 파라미터별 비트 할당 방식은 표 1에 제시 되어 있다.

AMR-WB 인코더는 입력신호가 들어오면, 우선 전처리 과정을 거친 후 linear predictive coding (LPC) analysis를 통해서 LPC 파라미터를 구하게 된다. LPC 파라미터는 line spectrum pairs (LSP)로 변환된 후 split-multistage vector quantization (SMVQ) 방식으로 양자화 되며, 표 1에서는 각 단계별 파라미터를 LSP0-LSP6으로 나타내고 있다.

또한, 입력음성으로부터 인코더의 target 신호를 얻기 위해서 perceptual weighting 필터를 거치게 된다. 이 필터는 스펙트럼상에서 pole의 에너지를 끌어내려 zero 부분을 더 정확하게 합성하게 됨으로써 더욱 자연스러운 합성음을 재생할 수 있게 해준다. 인코더의 적응 코드북 탐색 과정은 다음과 같다.

1) Open-loop 피치탐색: 적응 코드북 탐색 과정의 복잡도를 최소화하기 위해서 perceptual weighting 필터를 통과한 신호로부터 최적의 피치를 찾는다. AMR-WB의 6.60 kbit/s 모드에서만 프레임 당 한 번의 과정을 거치며 나머지 다른 모드에서는 프레임 당 두 번씩 수행한다.

2) 적응 코드북 탐색: 서브 프레임 당 한 번씩 수행하며 closed loop pitch search와 computing the adaptive codevector 두 과정으로 이루어진다. AMR-WB의 6.60 kbit/s 모드와 8.85 kbit/s 모드를 제외한 나머지 모드들에서는, 첫 번째 서브 프레임과 세 번째 서브 프레임에는 fractional pitch delay [34, 127 (3/4)], [128, 159 (1/2)], [160, 231] 범위에 각각 1/4, 1/2, 1의 resolution을 사용한다. 두 번째와 네 번째 서브 프레임에는 [T_1-8, T_1+7 (3/4)] 범위에 항상 1/4 resolution을 사용하며, 여기서 T_1은 이전 서브 프레임의 fractional pitch delay와 가장 가까운 정수 값이다. closed loop pitch 과정은 원음과 합성음과의 weighted mean-squared error를 최소화하는 과정으로 수행된다. adaptive codevector는 결정된 fractional pitch delay에 의해서 구할 수 있으며, AMR-WB 12.65 kbit/s 모드에서는 표-1과 같이 첫 번째와 세 번째 서브 프레임에 각각 9비트가 할당되고 이 두 개를 합한 18비트를 ADCB1이라 명명하였다. 두 번째와 네 번째 서브 프레임에는 각각 6비트가 할당되고 그것을 합한 12비트를 ADCB2라 명명하였다. adaptive codevector는 경우에 따라서 과거 excitation과 interpolation으로 구할 수도 있으며 적응 코드북 필터링 파라미터인 LTP-filtering은 LTPfilt 라 명명하였다.

표 2. AMR-WB 12.65 kbit/s 모드 Algebraic codebook의 트랙 당 펄스 포지션
Table 2. Valid pulse position of the algebraic codebook of AMR-WB 12.65 kbit/s.

Track	Valid pulse positions in subframe
T_0	0,4,8,12,16,20,24,28,32,36,40,44,48,52,56,60
T_1	1,5,9,13,17,21,25,29,33,37,41,45,49,53,57,61
T_2	2,6,10,14,18,22,26,30,34,38,42,46,50,54,58,62
T_3	3,7,11,15,19,23,27,31,35,39,43,47,51,55,59,63

다음 과정은 고정 코드북 탐색이다. AMR-WB 12.65 kbit/s 모드에서는 고정 코드북을 표 2와 같이 algebraic codebook 구조로 설계하여 사용하고 있다. 해당 구조는 한 서브 프레임에 해당하는 64-dimension의 codevector에 대해서 총 8개의 펄스를 심을 수 있도록 되어 있다. 64-dimension의 codevector는 트랙 당 16개씩, 총 4개의 트랙으로 나누어지며, 트랙 당 2개씩 총 8개의 +1 또는 -1 값을 가지는 non-zero 펄스를 심을 수 있다. AMR-WB 12.65 kbit/s 모드에서는 서브 프레임 당 36비트 씩 총 144비트가 할당되며, 본 연구에서 각 트랙 당 할당된 36비트를 각각 ACELP0-ACELP3으로 명명하였다. 그밖에, GAIN이라 명명한 codebook gain에 28비트가 할당되고 과거 신호를 저장하기 위한 메모리 업데이트 과정을 거치게 된다. AMR-WB 23.85 kbit/s 모드에서는 예외적으로 high band gain이 필요하며 6400-7000Hz구간을 코딩하는 역할을 한다. 이외에 음성구간의 유무를 판단하는 파라미터를 VAD 라 명명하였다.

III. 제안 알고리즘

본 절에서는 제안 된 AMR-WB 개선사항에 대한 설명을 하도록 하겠다. 본 연구는 AMR-WB에 기반을 두고 TTS 시스템에서 효율적으로 사용 할 수 있도록 기존 AMR-WB에 비해 비트율을 낮추는 방향으로 진행하였다. 제안된 알고리즘은 다음과 같다.

3.1. Cmmon Bit-Stream (CBS) 제거

AMR-WB에서는 한 프레임을 전송할 때 핵심 파라미터들에 해당하는 비트 스트림 이외에 Frame Type, Frame Quality Indicator (FQI), Mode Indication, Mode Request, Codec CRC 등에 관련된 부가 정보를 헤더로 전송한다. 여기서 Frame Type이란 9가지 AMR-WB 모드 중 어느 모드를 사용하는지 지시하는 부분으로 총 4비트이다. FQI는 데이터 프레임에서 에러 발생 여부를 확인하는 부분으로 1비트를 가진다. Mode Indication과 Mode Request는 Frame Type과 동일한 역할과 비트할당을 가진다. Codec CRC는 에러를 detection하는 기능을 하며 총 8비트를 가진다. 이외에 spare 부분과 undefined 부분까지 모두 합치면 헤더 정보에 총 27 비트가 할당되어 있다. AMR-WB 12.65 kbit/s 모드의 경우, 헤더 27 비트는 매 프레임마다 payload에 해당하는 253 비트와 함께 전송되므로, 해당 모드의 실제 비트율은 14.00 kbit/s가 된다.

하지만, 본 연구에서는 TTS 데이터베이스 압축을 위해서 AMR-WB 12.65 kbit/s 모드만을 사용하므로 헤더에 대한 정보는 저장하지 않아도 된다. 우선 spare 부분 3 비트와 undefined 부분 3 비트는 전송하는 정보가 없는 불필요한 부분이므로 제거할 수 있다. 또한, 채널 코딩을 위한 Frame Type, FQI, Codec CRC 등은 TTS 시스템에는 사용 할 필요가 없으므로 제거가 가능하다. 소스 코딩을 위한 헤더 정보인 Mode Indication, Mode Request 등은 다양한 AMR-WB 비트율 중에서 어떤 비트율을 선택하여 전송할지 결정하게 된다. 하지만, 제안하는 TTS 시스템의 특성 상 여러 가지 모드를 사용하지 않고 하나의 모드만을 고정해서 사용하므로 소스 코딩과 관련된 헤더 정보도 저장할 필요가 없이 제거할 수 있게 된다.

즉, 우리가 사용하려는 TTS 시스템에서는 사용하는 비트율이 고정되어 있고 채널 에러가 거의 발생하지 않기 때문에 모든 헤더 부분을 제거해서 사용하는 것이 보다 더 효율적이다. 따라서 본 연구에서는 핵심 파라미터에 해당하는 (예를 들어, 12.65 kbit/s에서는 253 비트) 부분만을 전송하고자 한다.

3.2. AMR-WB + Huffman Coding

기존 양자화기에는 크게 고정 비트율 관점에서 평균 왜곡을 최소화하는 Resolution-Constrained Quantization (RCQ) 방식과 평균 비트율 관점에서 평균 왜곡을 최소화하는 Entropy-Constrained Quantization (ECQ) 방식이 있다. 우리가 사용하려는 분야인 TTS 시스템에는 채널 에러가 심하지 않기 때문에 RCQ 방식보다는 ECQ 방식으로 양자화하는 것이 바람직하다. 즉, RCQ 방식이 고정 비트율 관점에서 평균 왜곡을 최적화하는데 비해 ECQ 방식은 평균 비트율 관점에서 평균 왜곡을 최적화하기 때문에, ECQ 방식이 보다 효율적으로 평균 비트율을 떨어뜨릴 수 있다. 단, ECQ 방식은 고정된 비트율이 아닌 가변 비트율로 데이터를 전송하기 때문에 채널 에러가 심한 환경에서는 사용이 제한될 수 있으며 RCQ 방식에 비해서 시스템의 복잡도를 증가시킬 수 있는 단점을 가지고 있다. 따라서 채널 에러가 심하지 않은 TTS 시스템에 사용이 가능해지는 것이며 복잡도 대비 최대의 비트율의 효율을 가질 수 있는 방법을 연구하였다.

High-Rate theory 에 의하면 가장 이상적인 ECQ 시스템은 Uniform Quantization과 무손실 압축 방식을 결합한 시스템 형태라고 알려져 있다 [14]. 본 연구에서 사용하는 AMR-WB은 uniform quantization을 사용하고 있지는 않지만, 비트율 성능 개선을 위해 AMR-WB에 기반을 두고 AMR-WB이 생성하는 파라미터들을 Huffman coding하는 방식으로 어느 정도의 개선이 있는지를 실험하였다. 그림 1은 제안하는 알고리즘을 도식화한 것이다. 제안 방식의 장점은 AMR-WB의 비트 스트림을 변형 없이 Huffman coding의 인코더와 디코더 단에서 사용할 수 있다는 점이다. 즉, 기존 AMR-WB 코딩 방식의 변형 없이 플러그인 방식으로 동작하므로, Huffman coding을 사용할 수 없는 응용처에서 생성된 비트 스트림과의 backward compatibility가 완벽하게 보장 된다는 장점이 있다.

3.3. 적응 코드북 파라미터의 무손실 압축 방식

AMR-WB 이 전송하게 되는 파라미터들과 파라미터별 비트 할당은 표 1에 나타나 있다 (12.65 kbit/s 기준). 이 중 하나인 적응 코드북 search 파라미터는 서브 프레임 단위로 이루어진다. AMR-WB은 4개 서브 프레임으로 이루어져 있으며, 첫 번째와 세 번째 서브 프레임에 해당하는 delay 값인 ADCB1을 9비트로 직접 양자화하여 전송하게 되고, 두 번째와 네 번째 서브 프레임에 대해서는 각각 첫 번째와 세 번째 서브 프레임에서 찾은 delay값과의 차이 값인 ADCB2를 양자화하여 전송하게 된다. 본 연구에서는 첫 번째와 세 번째 서브 프레임에 해당하는 ADCB1 값도 이전 값과의 차이만을 델타 코딩하는 방식으로 실험하였다. 그것은 다음의 가정에 근거한 것이다.

- TTS 시스템에는 채널 에러가 거의 없다.
- TTS 시스템에서 음성 압축은 동일 화자에 대한 음소 단위로 진행된다.
- 대부분의 음소가 모음에 해당한다. (자음에 비해

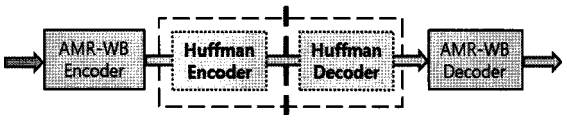


그림 1. 제안하는 알고리즘의 블럭도
Fig. 1. Block diagram of proposed algorithm.

모음의 종류가 많고 길이도 길다.)

- 동일 화자에 대한 모음 음소의 경우 프레임 간 피치 변화가 적다.

ADCB1 값을 델타 코딩함으로써 ADCB1 값이 가질 수 있는 값의 범위는 +와 - 두 가지 변이를 모두 포함하므로 오히려 증가하지만, pdf는 0 근처 값에 밀집되어 있으므로 최종적으로 델타 값에 Huffman coding을 적용한 결과 성능 개선을 확인 할 수 있었다.

3.4. 손실 압축 방식

앞에서 제안한 알고리즘들은 모두 무손실 압축 방식 관점에서 비트율을 낮추는 방식들을 제안하고 있다. 반면, 본 절에서는 앞에서 제시한 방식들에 손실 압축방식을 추가함으로써 손실대비 최대의 효율을 내고자 한다.

AMR-WB 파라미터 중 VAD 값은 MR-DTX 모드에서 주로 사용된다. 예를 들어, AMR-WB 소스 코딩에 할당하는 비트율을 12.65 kbit/s로 고정해도 discontinuous transmission (MR-DTX) 모드를 on 시키면 음성이 없는 프레임 (VAD=0) 에 대해서는 12.65 kbit/s 모드가 아닌 MR-DTX 모드로 동작하게 되며, 디코더 단에서 comfort noise를 재생할 수 있는 최소한의 비트만을 전송하게 된다. 하지만, TTS 시스템은 음소단위로 음성코딩이 이루어지며 음성이 없는 구간을 데이터베이스화 할 필요는 거의 없으므로 대부분 프레임을 음성이 있는 프레임 (VAD=1) 으로 인식하고 동작한다는 사실을 알 수 있다. 따라서, VAD 파라미터에 할당된 1비트를 제거하여도 음질열화는 거의 발생하지 않는다.

12.65 kbit/s의 경우에는 적응 코드북 search로 찾은 excitation signal 을 두 가지 path로 관리하고 있다. 첫 번째 path는 excitation signal을 그대로 사용하는 것이고, 두 번째 path는 excitation signal을 필터링하여 불필요한 고주파 성분을 낮춰 주는 것이다. 보다 적합한 path는 target 신호와 excitation signal의 mean-squared error (MSE) 를 최소화 하는 과정에서 찾게 된다. 하지만, informal listening test를 통해서 살펴보면 그 효과는 그다지 크지 않았다. 따라서, 본 연구에서는 excitation signal은 필터링 없이 그 자체를 사용함으로써 서브 프레임 당 1비트, 즉 프레임 당 4비트를 줄이고자 한다. AMR-WB 의 여러 가지 모드 중 6.60 kbit/s와 8.85 kbit/s 모드에서는 이미 해당 부분에 비트 할당 없이 excitation signal 자체를 사용하고 있다.

또한, AMR-WB 파라미터들 중 algebraic codebook search에 사용되는 일부 펄스를 제거하여 비트율을 떨어뜨리는 방식을 연구하였다. AMR-WB 설계 시 고려된 통신 시스템 환경은 다양한 배경 잡음을 고려하고 있지만, TTS 시스템에서 압축하고자 하는 데이터베이스는 노이즈가 없는 무향실 환경에서 숙련된 성우에 의해서 녹음이 이루어지므로 informal listening test를 통해서 살펴보면 펄스의 수를 줄여도 성능에 큰 차이가 없었다. 기존 12.65 kbit/s 모드에서는 하나의 서브 프레임을 4개의 트랙으로 나누어 트랙 당 2개씩 총 8개의 펄스를 찾아서 탐색하고 있다 (기존의 AMR-WB을 8-pulse version 이라 하겠다). 하지만, 본 연구에서는 네 번째 트랙에서는 하나의 펄스만을 찾음으로써 총 7개의 펄스만으로 algebraic codebook search를 수행하도록 하였다 (7-pulse version). 여기에 추가적으로 비트율을 더 떨어뜨리기 위해서 세 번째와 네 번째 트랙에 대해서 하나의 펄스만을 선택함으로써 총 6개 펄스만으로 algebraic codebook search를 하는 버전도 구현하였다 (6-pulse version).

IV. 실험 및 결과

제안된 알고리즘은 총 8가지 변형이 존재하며 그것들은 다음과 같다.

1) CBS 제거 version
2) CBS 제거 + Huffman coding version
3) CBS 제거 + Huffman coding + ADCB deltaQ version
4) CBS 제거 + Huffman coding + ADCB deltaQ + VAD, LTPfiltering 파라미터 제거 version
5) CBS 제거 + Huffman coding + 7pulse version
6) CBS 제거 + Huffman coding + 7pulse + ADCB deltaQ + VAD, LTPfiltering파라미터 제거 version
7) CBS 제거 + Huffman coding + 6pulse version
8) CBS 제거 + Huffman coding + 6pulse + ADCB deltaQ + VAD, LTPfiltering파라미터 제거 version

본 절에서는 앞에서 제안한 알고리즘들에 대한 성능을 평가해 보고자 한다. 실험에는 TTS 시스템에 사용되는

표 3. 제안 알고리즘의 실험 결과
Table 3. The experiment result of proposed algorithm.

concat30 (12.65 kbit/s)		AMR-WB (Original Bit-Format)	CBS 제거	8-pulse			7-pulse		6-pulse	
				+huffman	ADCB DeltaQ	VAD=1, exitation filtering	+huffman	VAD=1, exitation filtering	+huffman	VAD=1, exitation filtering
entropy	CBS+ unused bits	27	0	0	0	0	0	0	0	0
	VAD	1	1	0.67	0.67	0.00	0.67	0.00	0.67	0.00
	LSP0	8	8	6.84	6.84	6.84	6.84	6.84	6.84	6.84
	LSP1	8	8	7.22	7.22	7.22	7.22	7.22	7.22	7.22
	LSP2	6	6	5.82	5.82	5.82	5.82	5.82	5.82	5.82
	LSP3	7	7	6.87	6.87	6.87	6.87	6.87	6.87	6.87
	LSP4	7	7	6.75	6.75	6.75	6.75	6.75	6.75	6.75
	LSP5	5	5	4.89	4.89	4.89	4.89	4.89	4.89	4.89
	LSP6	5	5	4.94	4.94	4.94	4.94	4.94	4.94	4.94
	ADCB1	18	18	17.65	15.69	15.68	17.07	15.69	17.07	15.69
	ADCB2	12	12	11.11	11.11	11.13	11.14	11.15	11.15	11.14
	LTPfilt	4	4	3.78	3.78	0.00	3.77	0.00	3.76	0.00
	ACELP0	36	36	35.70	35.70	35.70	35.65	35.61	35.63	35.59
	ACELP1	36	36	35.78	35.78	35.77	35.72	35.69	35.72	35.71
	ACELP2	36	36	35.81	35.81	35.78	35.82	35.81	19.58	19.55
	ACELP3	36	36	35.79	35.79	35.79	19.56	19.53	19.57	19.54
	GAIN	28	28	26.29	26.29	26.36	26.29	26.37	26.25	26.36
total entropy		280.00	253.00	245.91	243.94	239.52	229.01	223.18	212.72	206.91
Lossless/Lossy Modification		Lossless	Lossless	Lossless	Lossless	Lossy	Lossy	Lossy	Lossy	Lossy
total bits		280.00	253.00	246.63	245.27	240.33	230.32	224.00	214.04	207.72
Reduction in bits		0.00	27.00	33.37	34.73	39.67	49.68	56.00	65.96	72.28
Reduction in ratio		0.00	9.64%	11.92%	12.40%	14.17%	17.74%	20.00%	23.56%	25.81%
PESQ score		3.59	3.59	3.59	3.59	3.56	3.50	3.47	3.37	3.35

무향실에서 녹음된 여성 화자 1인에 대한 음성을 사용하였다. 아래 실험 결과는 평균 5~7초 가량의 음성 30 문장을 이용한 결과이며, 300 문장에 대한 실험 결과도 30 문장의 경우와 거의 유사함을 확인하였다.

표 3은 제안하는 압축 알고리즘별 파라미터 엔트로피, 총 엔트로피, 실제 압축에 의해 얻어진 비트율, 그리고 Perceptual Evaluation of Speech Quality (PESQ) 점수를 나타내고 있다. 실제 코더를 구현하기 전에 이론치인 엔트로피를 측정해 봄으로써 코더 성능을 미리 예측해 보았으며, total bits는 실제로 구현된 코더를 통해 측정한 실험값으로 이론치인 엔트로피와 유사한 결과를 나타내었다. 또한, 제안하는 알고리즘이 Lossless/Lossy Modification 중 어떤 방식인지를 구분하였다. Reduction in Bits는 제안 알고리즘 사용 시 기존 AMR-WB 12.65 kbit/s 모드와 비교해 어느 정도 비트를 줄일 수 있는지를 나타내며 그것을 비율로 표현한 값이 Reduction in Ratio가 된다. 제안된 알고리즘을 사용해서 만든 합성음과 기존 AMR-WB을 사용해 만든 합성음과의 비교는 PESQ를 통해 확인할 수 있다.

4.1. common bit-stream (CBS) 제거

12.65 kbit/s AMR-WB 표준 비트 스트림에서 불필요한 헤더 정보와 undefined parts들을 제외 할 경우 다음과 같은 비트율 개선을 기대할 수 있었다: 280비트 → 253비트 (9.64%개선). 또한 불필요한 헤더 정보와 undefined parts만을 제거했기 때문에 음성 정보에는 어떤 손실도 없으며, 따라서 제안한 알고리즘 사용 시 기존 AMR-WB을 사용해 만든 합성음과 동일한 PESQ 점수를 얻을 수 있다.

4.2. AMR-WB + Huffman coding

CBS 제거 알고리즘을 수행한 이 후 각 파라미터 별로

Huffman coding을 실행한 결과를 표 3에 정리하였다. Huffman coding은 무손실 압축 방식이기 때문에 음질 저하 없이 비트율을 떨어뜨리는 효과를 기대할 수 있다. 실험 결과 대부분의 파라미터들이 거의 uniform하게 분포함을 알 수 있었으며, 이 경우 high rate theory에 근거하면 RCQ와 ECQ bound가 근접하게 되고, 따라서 RCQ에 Huffman coding 방식을 적용하였음에도 rate-distortion 관점에서 큰 이득을 얻기는 어렵다. 실험 결과에 의하면 엔트로피 관점에서는 7.09 비트, 구현 알고리즘 관점에서는 6.37 비트의 개선효과를 얻을 수 있었다. Huffman coding 방식의 비트율은 엔트로피로 예측된 비트율에 비해서 높다고 알려져 있으므로 (1 비트 이내), 구현된 Huffman coding 알고리즘은 정확히 동작하고 있다고 할 수 있다. 하지만, CBS reduction 과 Huffman coding 을 동시에 사용할 경우, 기존 AMR-WB 비트 스트림에 비해서 프레임 당 33.37비트를 절감할 수 있었다 (11.92%개선).

제안하는 방식의 성능 개선은 다수의 불특정 화자가 아닌 특정 화자 1인에 대하여 Huffman coding 테이블을 디자인 했을 때만 얻어질 수 있다. 따라서, 본 방식은 불특정 화자를 대상으로 하는 통신용 시스템보다는 특정 화자 1인을 이용하여 구현되는 TTS 시스템에 보다 효율적으로 사용될 수 있다.

4.3. AMR-WB 파라미터 내 무손실 압축 방식

AMR-WB의 파라미터 중 ADCB1을 델타 코딩방식에 의해 무손실 압축하는 알고리즘 (표 3의 ADCB DeltaQ)에 대한 성능평가 결과, 기존 AMR-WB에 대비해 PESQ 점수의 저하 없이 프레임 당 1.36 비트를 추가적으로 절감할 수 있었다.

4.4. 손실 압축 방식

앞서 설명했듯이 손실 압축 방식으로 제안된 알고리즘은 크게 두 가지 방식이다. 하나는 TTS 시스템의 특성상 불필요한 VAD와 LTPfiltering 2개 파라미터를 제외하는 방식이고, 다른 하나는 algebraic codebook searching에서 기존 8개 펄스를 찾는 대신 각각 7펄스와 6펄스만을 구해서 압축하는 방식이다. 8-pulse version의 경우 VAD와 LTPfiltering 2개 파라미터를 제외한 경우, 비트율 관점에서는 프레임당 4.94 비트를 추가적으로 줄일 수 있었으며, 이때 PESQ 점수는 거의 느낄 수 없는 수준인 0.03 점 정도 감소함을 알 수 있다. 7-pulse version 시는 최대 20.00%의 비트 효율을 나타냈으며, 6-pulse version 시는 최대 25.81%의 비트 효율을 볼 수 있었다. 각각의 경우에 대해서 기존 AMR-WB와 비교 시 PESQ 점수 저하는 0.12점 그리고 0.24점으로 나타났다.

4.5. Complexity

제안한 알고리즘들의 계산량 증가는 무시할 만한 수준이었으며, 메모리 증가에 대하여 논해 보고자 한다. Huffman coding 사용 시 각 파라미터별 발생 확률을 테이블로 가지고 있어야 하며 이에 필요한 메모리 증가는 표 1에 나타나 있다. 제안하는 알고리즘 사용 시 4228 units의 메모리 증가가 요구되며, unit 당 4 bytes가 할당되었다고 가정할 경우 16.91 kbytes의 메모리 증가가 요구된다. 하지만, 가장 많은 메모리를 차지하는 algebraic codebook 파라미터의 경우 서브 프레임별 pdf가 거의 동일하므로 서브 프레임마다 서로 다른 pdf 테이블을 저장할 필요가 없다. 또한, algebraic codebook 파라미터 9비트는 sign 1 비트와 pulse position 8 비트로 나눌 수 있으며, 파라미터 간에 독립적인 특징을 이용하면 테이블도 독립적으로 저장할 수 있다. 따라서, 기존에 512*4 units가 요구되던 algebraic codebook 파라미터를 258 units로 축소해서 저장할 수 있었다. 또한, 적응 코드북 파라미터인 ADCB1과 ADCB2도 각각에 대해서 서브 프레임별 pdf 분포가 유사하므로 독립적으로 테이블을 저장하면 필요한 units 양을 반으로 줄일 수 있었다. 개선된 방식을 적용한 경우, 1862 units의 메모리 증가가 요구되며, unit 당 4 bytes가 할당되었다고 가정할 경우 7.45 kbytes의 메모리 증가가 요구된다. 즉, 기존 방식에 비해 55.96%의 메모리 감소 효과를 얻을 수 있었다. 7.45 kbytes의 메모리 증가는 제안하는 시스템을 모바일 단말에서 사용하는 경우에도 무시할만한 수준이다.

4.6. 그 외 모드

AMR-WB 12.65 kbit/s 모드 외에 8.85 kbit/s 모드에 대해서도 동일한 실험을 진행하였다. 무손실 압축을 적용한 경우 8.85 kbit/s 모드에서는 음질 저하 없이 최대 18.25% 비트율 개선 효과를 얻을 수 있었다. 그 외 다른 모드들에서도 비슷한 결과가 나올 것으로 예상된다.

V. 결 론

본 논문에서는 TTS 시스템에 효율적으로 사용할 수 있는 AMR-WB 기반의 음성 부호화 알고리즘에 관해 연구하였다. 제안된 알고리즘은 불필요한 헤더 정보들을

제거하고, 특정 화자에 종속적인 Huffman coding 등을 사용함으로써 추가 왜곡 없이 압축 효율을 증가시킬 수 있었다. 또한, 개선효과가 적은 파라미터 제거 등을 통한 손실 압축 기법도 제안하였다. 실험 결과 무손실 압축 방식은 기존의 12.65 kbit/s AMR-WB에 비해 음질 저하 없이 최대 12.40%의 비트율 개선 효과를 나타냈다. 손실 압축 방식은 PESQ 0.12 저하 시 20.00% 비트율 개선 효과를 나타냈다. 본 논문에서 제안된 알고리즘은 TTS 데이터베이스 압축 시스템뿐만 아니라 tapeless answering device (TAD) 등 음성 저장 장치들에 다양하게 응용될 수 있으리라 기대된다. 또한, 향후에는 [1]과 제안하는 방식을 결합하여 추가 왜곡 없이 비트율을 개선한 수 있을 것으로 기대된다.

감사의 글

이 논문은 2008년 (주)에이치씨아이랩의 지원을 받아 수행된 연구이며, 이에 감사드립니다.

참 고 문 헌

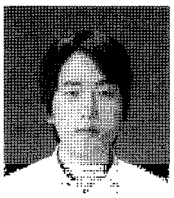
1. C.-H. Lee, S.-K. Jung, and H.-G. Kang, "Applying a Speaker-Dependent Speech Compression Technique to Concatenative TTS Synthesizers," *IEEE Trans. Audio Speech Language Processing*, vol. 15, no. 2, pp. 632-640, 2007.
2. 양희식, 한민수, "TTS DB 압축을 위한 광대역 파형보간 부호기 구현," *대한음성학회지*, 말소리 55호, 143-158쪽, 2005.
3. ISO/IEC JTC1/SC29/WG11 No.71, Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5Mbit/s: Part 3-Audio, 1993.
4. O. Derrien, P. Duhamel, M. Charbit, and G. Richard, "A New Quantization Optimization Algorithm for the MPEG Advanced Audio Coder using a Statistical Subband Model of the Quantization Noise," *IEEE Trans. Audio Speech Language Processing*, vol. 14, no. 4, pp. 1328-1339, 1998.
5. ITU-T Recommendation G.729, Coding of Speech at 8kbit/s using Conjugate-Structure Algebraic-Code-Excited Linear Prediction (CS-ACELP).
6. R. Salami, C. Laflamme, J. P. Adoul, A. Kataoka, S. Hayashi, T. Moriya, C. Lamblin, D. Massaloux, S. Proust, P. Kroon, and Y. Shoham, "Design and Description of CS-ACELP: A Toll Quality 8kb/s Speech Coder," *IEEE Trans. Speech Audio Processing*, vol. 6, no. 2, pp. 116-130, 1998.
7. B. Bessette, R. Salami, R. Lefebvre, M. Jelinek, J. Rotola-Pukkila, J. Vainio, H. Mikkola, and K.Jarvinen, "The Adaptive Multirate wideband Speech Codec(AMR-WB)," *IEEE Trans. Speech Audio Processing*, vol. 10, no. 8, pp. 620-636, 2002.
8. 3GPP TS 26.190, *Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; Transcoding functions*, v.7.0.0., 2007
9. 3GPP TS 26.201, *Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; Frame structure*, v.7.1.0., 2008
10. I. Singh, P. Agathoklis, and A. Antoniou, "Wavelet-based Compression of Speech Signals on the TMS320C30 Digital Signal Processor," in *Proc. IEEE Symposium on Advances in Digital Filtering Signal Processing*, pp. 178-182, 1998.
11. ITU-T Recommendation G.722.1, Coding at 24 and 32 kbit/s for Hands-Free Operation in Systems with Low Frame Loss.
12. X. Minjie, D. Lindbergh, and P. Chu, "ITU-T G.722.1 Annex c: A New Low-Complexity 14 kHz Audio Coding Standard," in *Proc. IEEE Conf. Acoust., Speech, Signal Processing*, pp. 173-176, 2006.
13. Y. Shoham, "Variable-size vector entropy coding of speech and audio," in *Proc. IEEE Conf. Acoust., Speech, Signal Processing*, vol.2, pp.769-772, 2001.
14. W. B. Kleijn, A Basis for Source Coding: Course Notes. KTH, Stockholm, 2008.

저자 약력

•임 종 욱 (Jongwook Lim)

2001.3~2008.8
세종대학교 정보통신공학과, 학사
2008.8~현재
세종대학교 정보통신공학과, 석사과정

•김 기 출 (Kichul Kim)



2002.3~2009.2
세종대학교 정보통신공학과, 학사
2009.3~현재
세종대학교 정보통신공학과, 석사과정

•김 경 선 (Kyeongsun Kim)

1987.3~1991.2 : 포항공대 전기전자과, 학사
1991.3~1993.2 : 포항공대 전기전자과, 석사
1993.3~2001.7 : 삼성종합기술원 전문연구원
2001.8~현재 : (주)에이치씨아이랩 연구소장

•이 항 섭 (Hangseop Lee)

1986.3~1990.2 : 굉운대학교 컴퓨터공학과, 학사
1990.3~1992.2 : 광운대학교 컴퓨터공학과, 석사
1992.1~2000.4 : 한국전자통신연구원 선임연구원
2003.8~현재 : (주)에이치씨아이랩 수석연구원

•박 혜 영 (Haeyoung Park)

1997.3~2001.2 : 창원대학교 제어계측공학과, 학사
2001.3~2003.2 : 부산대학교 전자공학과, 석사
2003.3~현재 : (주)에이치씨아이랩 책임연구원

•김 무 영 (Moo Young Kim)

1989.3~1993.2 : 연세대학교 전자공학과, 학사
1993.3~1995.2 : 연세대학교 전자공학과, 석사
1995.2~2000.12 : 삼성종합기술원 전문연구원
2001.1~2004.11 : Royal Institute of Technology (KTH, 스웨덴) Dept. Signals, Sensors, Systems, 박사
2004.12~2005.2 : Royal Institute of Technology (KTH, 스웨덴) Dept. Signals, Sensors, Systems, PostDoc
2005.2~2006.8 : Ericsson Research (스웨덴), Senior Research Engineer
2006.8~현재 : 세종대학교 정보통신공학과, 조교수
※관심분야 : 음성/오디오/비디오 신호처리, 패턴인식, 정보이론.