

음성기반 오디오 융합코덱 기술동향

Trends of Speech-Based Audio Convergence Codec Technology

IT 융합기술 동향 및 전망 특집

김도영 (D.Y. Kim)	멀티미디어통신연구팀 책임연구원
성종모 (J.M. Sung)	멀티미디어통신연구팀 선임연구원
이미숙 (M.S. Lee)	멀티미디어통신연구팀 선임연구원
배현주 (H.J. Bae)	멀티미디어통신연구팀 팀장
이병선 (B.S. Lee)	융합서비스플랫폼연구부 부장

목 차

-
- I. 서론
 - II. G.711 · G.722 기반 융합코덱
 - III. G.729 · G.718 기반 융합코덱
 - IV. 시사점 및 결론

* 본 연구는 지식경제부 및 정보통신산업진흥원의 IT신성장동력핵심기술개발사업의 일환으로 수행된 결과의 일부임. [2008-S011-02, FMC 어커스틱 융합코덱 및 제 어기술 연구]

본 논문에서는 통신과 방송서비스가 하나의 기기 또는 단말장치 안에서 결합되고 단말 내부에서는 디바이스의 통합에 따라 코덱의 개수를 최소화하기 위한 음성기반 오디오 융합코덱의 기술동향에 대해 기술한다. 하지만 기술적으로 완전히 태생이 다른 음성과 오디오 코덱을 진정한 의미에서 융합할 수 있는 기술적 모델과 기법은 아직 개발되지 않고 있다. 본 고에서는 이러한 시도의 일환으로 ITU-T SG16을 중심으로 진행되고 있는 음성기반 코덱을 점진적 대역폭 확장 기술을 사용하여 광대역 음성, 슈퍼와이드 밴드 및 향후 오디오 대역까지 커버할 수 있는 임베디드 가변비트율 코덱기술을 중심으로 기술동향의 분석을 시도한다.

I. 서론

음성통신 서비스와 방송, 음악감상 등의 오디오 서비스는 (그림 1)에서와 같이 개별적으로 휴대폰과 D-TV, PMP 등에서 사용되어 전혀 별개의 기술적 기반으로 개발되었다. 음성 코덱의 경우 사람의 발성시스템을 모방한 LPC 모델에 기반한 CELP 기술, 오디오 코덱의 경우 사람의 청각전달 시스템을 모방한 심리음향 기술을 기반으로 음성과 오디오 정보를 각각 효과적으로 부호화하고 복호화 할 수 있었다.

하지만 MP3 플레이어와 DMB 기능을 내장한 이동전화 단말기의 경우, 하나의 기기 또는 단말장치에서 디바이스의 통합에 따라 시장에서는 코덱의 개수를 최소화하고 가능하다면 단일화를 요구한다. 예를 들어 GSM/WCDMA 휴대전화와 DMB, MP3/MPEG-4 음악을 재생하기 위한 스마트폰의 경우 선택의 여지가 있지만 반드시 사용하여야 하는 음성 코덱과 오디오 코덱은 <표 1>에서 나타낸 바와 같이 비디오 코덱을 제외하더라도 거의 8개에 달하기

<표 1> GSM/WCDMA 융합폰 사용 코덱의 예

코덱 구분	코덱명(비트율, 시스템)	적용 알고리즘
음성 코덱	- FR(13.0, GSM FR)	- RPE-LTP
	- HR(5.6, GSM HR)	- VSELP
	- EFR(12.2, GSM FR)	- ACELP
	- AMR(4.75~12.2, GSM FR/HR, 3G WCDMA)	- ACELP
	- AMR-WB(6.60~23.85 중 하위 7개 모드, GSM FR; 모든 9개 모드, EDGE 및 3G WCDMA)	- ACELP
오디오 코덱	- MP3 Decoder	- 심리음향, Filtet Bank
	- MPEG-4 AAC	- PNS, LTP, Twin VQ
	- MPEG-4 BSAC	- BSAC, 비트율 제어

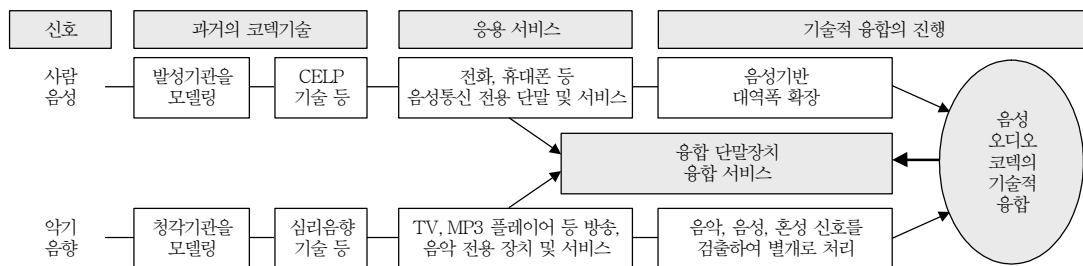
때문이다. <표 1>에서 AMR 코덱은 1999년 3GPP Release-99에서 WCDMA 음성서비스 필수 코덱으로 선정되었고, AMR-WB 코덱은 2002년 3GPP Release-5에서 광대역 음성 서비스를 위한 음성 코덱으로 선정되었다. MP3라 부르는 코덱인 MPEG-1 레이어 3 코덱과 MPEG-4 AAC 코덱은 음악재생을 위해, MPEG-4 BSAC 코덱은 DMB의 오디오 복호화시 사용한다.

이와 같이 음성 오디오 융합코덱 기술은 서비스와 디바이스의 통합에 따라 시장으로부터 요구된 것이다. 그러나 기술적으로 완전히 태생이 다른 음성 코덱과 오디오 코덱을 진정한 의미에서 융합할 수 있는 부호화 및 복호화 모델, 음성과 오디오 정보의 공통적인 신호적 특징과 분리 기법 등의 핵심원천 기술은 현재까지 개발되지 않고 있다.

오디오 코덱이 처리하는 신호의 대역폭(20~20,000Hz)이 음성 코덱이 처리하는 영역(50~7,000Hz)을 포함하므로 오디오 코덱을 사용하면 음성 코덱을 당연히 대체할 수 있을 것처럼 보이지만, 동일한 조건에서 음성신호는 ACELP 코덱이, 오디오 신호는 심리음향 기법을 사용하는 오디오 코덱이 보다 좋은 성능을 나타낸다. 매우 높은 비트율의 오디오 코덱을 사용하면 음성정보에 대해서도 우수한 성능을 보

● 용어해설 ●

광대역 음성 코덱(Wideband Speech Codec): 현재 유선전화, 인터넷전화(VoIP), 휴대폰에서 사용하는 코덱은 정보량을 감축하기 위해 사람의 음성정보 중 절반 정도인 300~3400Hz만 처리하고 나머지는 버리지만, 사람이 발생 가능한 50~7000Hz 대역의 모든 정보를 처리하여 자연성과 명료도가 우수한 코덱



(그림 1) 음성 오디오 코덱의 응용분야 및 기술개발 진행 현황

〈표 2〉 코덱기술의 관점에서 본 서비스의 융합 및 정보교환 유형

서비스 유형			정보교환 유형				
서비스 구분	대표적인 장치/서비스	대표 미디어	실시간통신	양방향통신	Broadcast Multicast	스트리밍 다운로드	
개별 서비스	음성통신	- 휴대폰 - 유선전화 - 인터넷전화	음성	●	●		
	방송	- TV 방송 - IP-TV, DMB	영상 오디오	●		●	●
	영화·음악감상	- DVD - 오디오 CD - PMP	영상 오디오				●
융합 서비스	통방융합 멀티미디어	- 융합단말기(휴대폰/MP3/DMB/PMP) - 양방향 IP-TV - 영상전화 - 화상회의 - 온라인 게임 - 온라인 쇼핑 - 실시간 e-러닝	음성 오디오 비디오 음향	●	●	●	
	멀티미디어 재생	- 멀티미디어 메일 - e-북(Book) - 오디오북 - Off-line e-러닝	문자 음성 오디오				●

이다. 그러나 대부분의 MPEG 계열의 오디오 코덱은 고품질의 방송서비스와 정보 저장을 위해 개발하였기 때문에 알고리즘 지연이 크며, 복잡도가 상대적으로 높아서 <표 2>에서 나타낸 정보교환의 유형 중 실시간 양방향 통신을 요구하는 휴대전화나 인터넷전화 서비스에는 부적합하다고 볼 수 있다. 이는 오디오 코덱이 원천적으로 사람이 발생할 수 없는 7,000~14,000/20,000Hz 대역까지 부호화하여 보내는 오버헤드 외에 저가 단말장치에서의 구현성, 알고리즘의 지연특성, 통신망에서의 손실보상, VAD/CNG와 같은 기능적 측면에서도 그러하다.

II장에서는 이의 대안으로서 ITU-T SG16에서 진행중인 G.711과 G.722 코덱기반 광대역 및 슈퍼와이드밴드 코덱 기술과 표준화 동향을 살펴보고, III장에서는 역시 ITU-T SG16에서 진행중인 G.729과 G.718 코덱을 기반으로 임베디드 형태로 대역폭을 확장하여 응용서비스에 적합한 비트율을 사용할 수 있도록 고려한 가변비트율 기법의 코덱 기술을 중심으로 관련 기술동향의 분석을 시도하며, IV장에서는 시사점을 분석하고 결론을 도출한다.

II. G.711 · G.722 기반 융합코덱

G.711 코덱은 1972년에 표준화된 이래로 유선전화와 VoIP 등에서 필수적이고 가장 널리 사용되는 코덱이다. 그러나 G.711 코덱은 일반 전화망의 64kbit/s에 매핑되도록 300~3,400Hz에 해당하는 협대역 음성정보의 처리만을 지원함으로써 제한된 품질을 제공하는 단점을 가지고 있다.

고품질 음성에 대한 요구에 부응하기 위해 G.711 광대역 확장 코덱에 대한 개발은 2008년 3월 완료되었으며, 현재 유럽의 차세대 무선전화 규격인 G.722 코덱과 공통사용 가능한 슈퍼와이드밴드 대역 및 스테레오급 확장기술 개발이 진행 중이다.

G.711 광대역 확장 코덱 기술은 2007년 1월 일본 NTT가 제안하여 개발이 시작되었으며 유선전화 등 PCM 코덱을 사용하는 기존 단말 및 시스템과 끊임 없는 호환성을 제공하면서 고품질 음성 서비스, 특히 IP 전화와 음성 회의를 제공하는 데 있다.

2007년 7월부터 품질시험을 통과한 ETRI, France Telecom, Huawei, NTT, 그리고 VoiceAge가 공동

으로 개발하였으며, 2008년 3월에 ITU-T G.711.1 코덱으로 최종 승인되었다.

G.711.1 코덱은 다음과 같이 3개의 계층으로 구성된다.

- 협대역 신호를 압축 및 복원하기 위한 G.711 코덱과 호환되는 로그 PCM 코어 계층
- 협대역 신호의 음질 향상을 위한 임베디드 PCM 확장 계층
- MDCT 기반의 광대역 향상 계층

이 3개 계층의 4가지 조합으로 이루어진 가변 비트율은 품질과 비트율, 통신망간 연동성 측면에서 응용서비스의 요구에 따라 선택할 수 있는 3종의 가변 비트율을 제공한다. 즉, 64kbit/s(협대역), 80kbit/s(협대역), 80kbit/s(광대역) 및 96kbit/s(광대역)를 지원한다. <표 3>은 G.711.1 코덱의 각 모드에 대한 비트스트림 구성을 나타낸 것이다.

인코더의 입력과 디코더의 출력은 기본적으로 16kHz 신호이지만, 협대역에 해당하는 8kHz 신호도 지원한다. G.711.1 코덱은 5ms 프레임 단위로 동작하고 최대 알고리즘 지연은 11.875ms이며 계산량은 8.70 WMOPS이다.

G.711.1과 G.722 광대역 코덱의 대역을 슈퍼와

<표 3> G.711.1 코덱의 비트스트림 조합

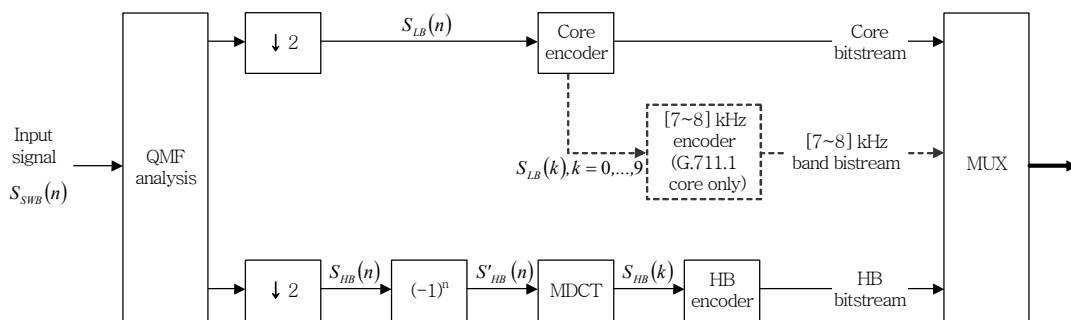
모드	계층 0	계층 1	계층 2	비트율(kbit/s)
R1	X	-	-	64
R2a	X	X	-	80
R2b	X	-	X	80
R3	X	X	X	96

이드밴드(20~14,000Hz)까지 확장하기 위한 기술개발과 표준화는 2007년 10월 시작하여 2008년 4월 회의에서 기술 규격이 승인되어 2009년 7월부터 ETRI, France Telecom, Huawei, NTT, Siemens, VoiceAge의 협력개발이 진행되고 있다.

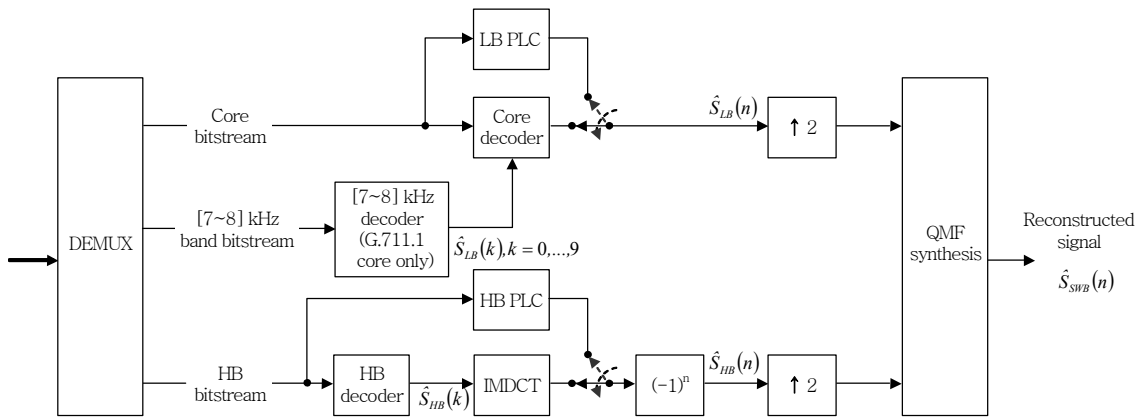
(그림 2)와 (그림 3)은 G.711 · G.722 기반의 슈퍼와이드밴드 코덱 구성을 위한 부호화기와 복호화기의 구성 예[1]를 보여준다. (그림 2)에서와 같이 부호화기는 5ms 마다 32kHz로 샘플링한 160 샘플을 입력하며, 입력 프레임은 64-tap QMF 분석 필터뱅크를 사용하여 2개의 밴드로 분리한다. 필터를 통과한 각 신호는 다운 샘플링하여 저대역(LB)과 고대역(HB) 신호를 구성한다. 저대역 신호는 2개 코덱의 코어코덱을 사용하여 각각 부호화하고, 코어 코덱의 모드에 따라 56, 64, 80kbit/s 코어 비트스트림을 만든다. 고대역 신호는 spectral folding된 다음, MDCT를 사용하여 주파수 영역의 계수로 변환된다. 이때 MDCT는 5ms 마다 수행하고 10ms 크기의 분석 윈도우를 사용한다.

이 MDCT 계수는 LPC 기반의 TPC 기법을 사용하여 고대역 정보로 부호화되는데, MDCT 계수의 고대역 정보를 고대역 이득계수로 정규화한 후 LPC 분석과 3단계로 구성된 펄스부호화를 선택적으로 수행하여 슈퍼와이드밴드 대역확장을 위한 고대역 부호를 산출한다.

(그림 3)에서 예시한 바와 같이 복호화기는 부호화기와 대응하는 코어 복호화기와 고대역 복호화기로 구성하며 역과정을 거쳐 신호를 재생하게 된다.



(그림 2) G.711 · G.722 기반 슈퍼와이드밴드 부호화기의 구성 예



(그림 3) G.711 · G.722 기반 슈퍼와이드밴드 복호화기의 구성 예

복호화기에서 패킷손실을 검출한 경우, 코어 계층에서는 각 코덱에서 사용하는 PLC 메커니즘을 사용하고 고대역 PLC는 정상적으로 수신된 직전 프레임의 역 MDCT 계수를 저장하였다가 이를 복사하여 사용할 수 있다. 알고리즘 지연은 G.722를 코어로 사용한 경우 약 12ms, G.711.1을 코어로 사용한 경우 14ms 이하로 매우 적은 지연특성을 제공한다.

Ⅲ. G.729 · G.718 기반 융합코덱

임베디드 가변비트율 코덱기술을 대표하는 대표적인 2개의 코덱인 G.729.1, G.718 코덱과 호환성을 제공하여 기존의 통신 시스템에 대한 영향을 최소화 하면서 슈퍼와이드밴드 음성 및 오디오 품질을 필요로 하는 응용분야에 사용하기 위해 2007년 7월부터 G.729.1과 G.718 슈퍼와이드밴드 확장코덱에 대한 표준화가 시작되었다.

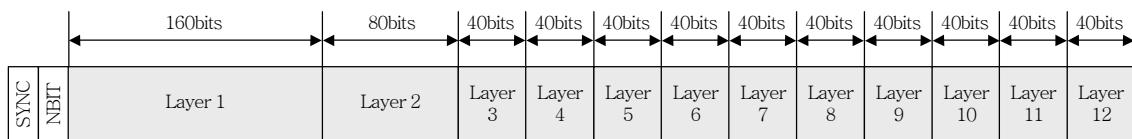
G.729.1과 G.718 코덱은 2006년 5월과 2008년 6월에 ITU-T 표준으로 각각 채택된 임베디드 가변비트율 광대역 코덱 기술로서, G.729와 AMR-WB로 널리 알려진 G.722.2 코덱의 12.65kbit/s 모드와

호환성을 각각 제공한다.

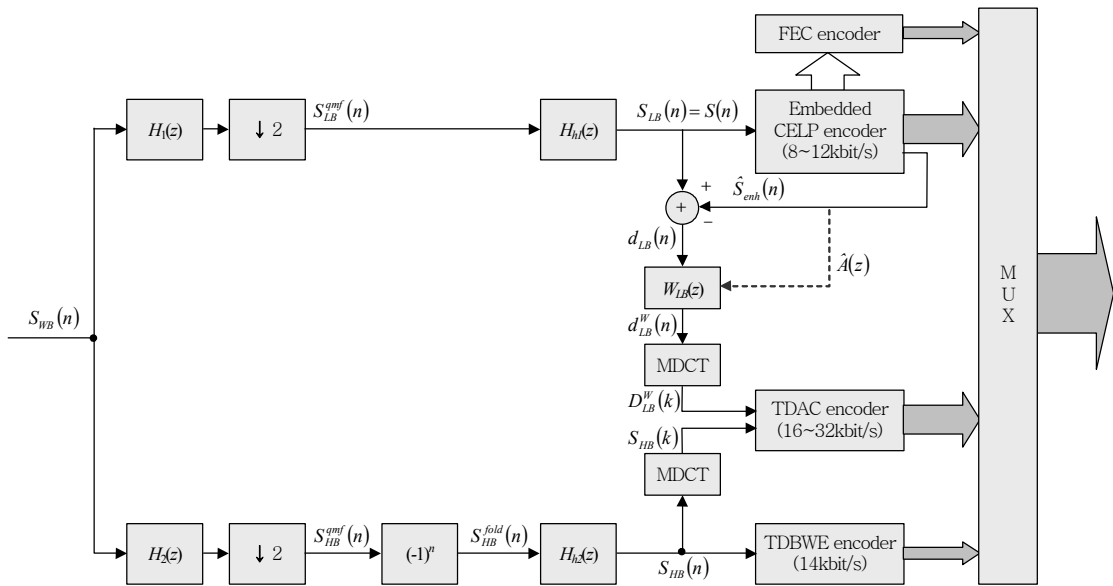
G.729.1 코덱 기술의 가장 큰 강점은 G.729a 코덱을 사용하는 VoIP 시스템과의 호환성을 유지하기 위해 임베디드 형태의 분할 대역 구조를 취한다는 점이며, (그림 4)에서와 같이 G.729 비트스트림을 코어계층(layer 1)으로 두고 그 비트열을 쌓아가면서 점진적으로 품질을 향상시킬 수 있는 구조이다.

이에 따라 G.729.1 코덱은 모두 12개의 계층을 갖는 가변비트율을 제공함으로써, 유무선 인터넷 간의 서로 다른 단말장치와 시스템 간의 연동성을 제공한다.

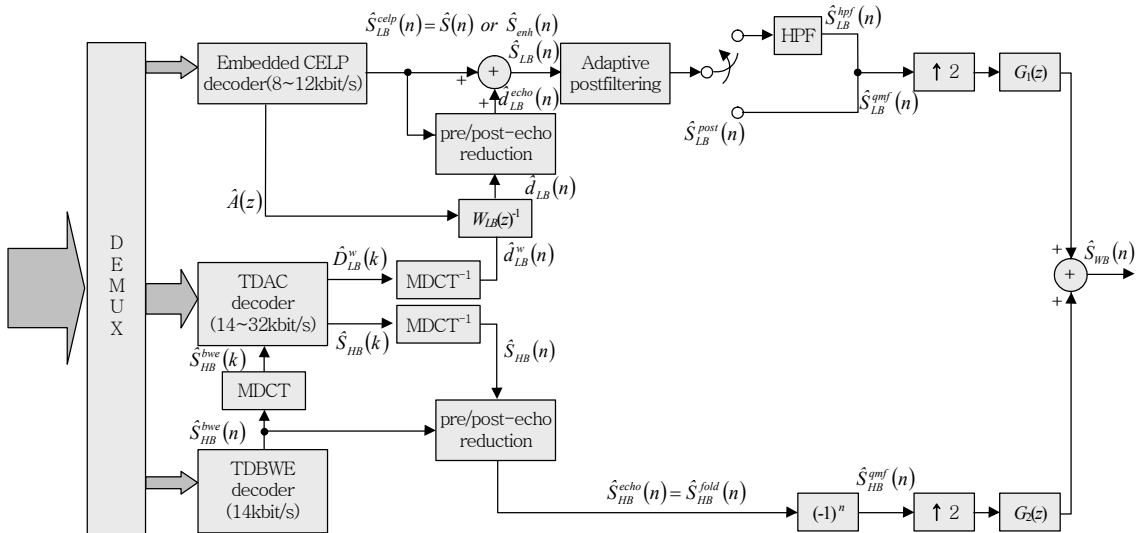
(그림 5)에 G.729.1 코덱의 전체 부호화기, (그림 6)에 복호화기로 각각 도시[2],[3]하였다. (그림 5)의 입력신호는 20ms 단위로 16kHz 16비트 PCM 신호를 입력으로 처리한다. 8kHz 샘플링 주파수도 지원하지만 출력은 8과 16kHz의 샘플링 주파수를 갖는 16비트 PCM 신호이다. G.729.1 코덱은 CELP, TDBWE, TDAC 세 개의 모듈로 구성된다. CELP 모듈은 8, 12kbps(layer 1, 2)에서 협대역 신호를 생성하고, TDBWE 모듈은 14kbps(layer 3)에서 광대역 신호를 생성한다. TDAC 모듈은 14~32kbps(layer



(그림 4) G.729.1의 계층형 비트스트림 형식



(그림 5) G.729.1 부호화기



(그림 6) G.729.1 복호화기

4~12)에서 2kbps 단위로 음질을 향상시키기 위해 사용된다. 50~4,000Hz의 입력신호와 CELP 모듈에서 발생한 신호 사이의 차이 신호, 4,000~7,000Hz의 입력신호를 MDCT 영역에서 코딩한다. CELP 모듈은 G.729와 마찬가지로 10ms마다 동작한다. G.729.1의 프레임 길이는 20ms이므로 CELP 모듈은 한 프레임마다 두번씩 동작한다.

입력신호는 QMF 필터를 사용해서 고대역과 저

대역 신호로 분리된다. 저대역 신호는 2만쯤 decimation을 한 후 50Hz 이하 주파수 성분을 제거하기 위해서 고대역 필터를 통과시킨다. 이 신호는 8과 12kbps로 CELP 부호화기에 의해 코딩된다. 전처리된 저대역 신호와 CELP 부호화기에 의해 재합성된 신호의 차이는 인지가중필터를 거쳐 차 신호와 고대역 입력 신호 사이의 스펙트럼 연속성을 보장하기 위해 이득 보상한다. 이 신호를 MDCT를 사용하여 변

환한다. 고대역 입력 신호는 2만배 decimation하고 주파수 대칭을 시킨다. 이 신호는 3kHz 이상 성분을 제거하도록 저대역 필터를 사용하여 전처리하고 이 신호를 TDBWE 기법을 사용하여 부호화한다.

MDCT 계수들은 TDAC 부호화기로 처리된다. 이 외에 프레임 손실이 발생시 음질 저하를 막기 위한 몇 개의 파라미터가 전송된다. (그림 6)에는 다중화된 비트스트림을 수신하여 처리하는 복호화기의 구조를 도시하였다.

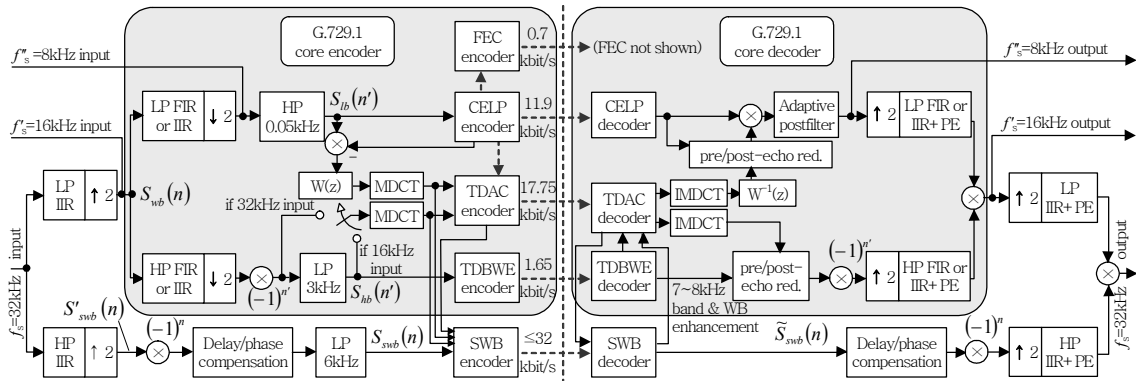
복호화 처리를 위해 CELP, TDBWE, TDAC 세 단계의 역과정을 거치며 복호화 과정은 수신된 비트스트림의 헤더 정보를 해석하여 전송된 비트율에 따라 처리한다. 즉, CELP 단계는 8, 12kbps에서, TDBWE 단계는 14kbps에서, TDAC 단계는 14kbps 이상에서 복호화를 수행한다. 8, 12kbps에서는 저대역(50~4,000Hz) 신호가 복원된다. 14kbps에서는 TDBWE 복호화기에 의해 고대역 신호를 발생한다. 이 신호는 3,000Hz 이상의 신호는 0으로 놓고 MDCT에 의해 주파수 영역으로 변환된다. 14kbps 이상에서는 TDAC 복호화기로 저대역에서 인지가중된 차이 신호와 고대역 신호를 재합성한다. TDAC에서 전송받지 못한 대역은 TDBWE에서 생성한 신호로 대체된다. 이 때 스펙트럼 연속성을 위해 레벨은 조절된다.

G.729.1의 경우, 통신망에서 발생할 수 있는 프레임 손실보상을 위한 강력한 프레임 손실 은닉 기능이 함께 사용된다. 실제 네트워크에서 패킷 손실이 발생한 빈도를 고려해 보면 손실 은닉 알고리즘의 성능이 음질에 많은 영향을 미친다. G.729.1의 프레임 손실보상은 부호화기에서 이루어지는데 저대역 신호에서 음성 신호 종류, 에너지, 위상 정보를 코딩하여 전송한다. 음성 신호 종류는 12kbps 계층에서 2bit를, 위상은 14kbps 계층에서 7bit를, 에너지는 16kbps 계층에서 5bit를 전송한다. 핵심 계층(8kbps)에서는 어떠한 정보도 전달하지 않고 복호화기에서 음성 신호를 예측해서 사용한다. 음성 서비스 시스템에서 지연은 알고리즘 지연, 프로세싱 지연, 통신 지연으로 나뉜다. 총 지연이 200ms 이하가 되어야 소비자는 지연을 느끼지 못한다고 알려져 있다.

알고리즘 지연은 프레임 크기에 해당하는 지연과 look-ahead 처럼 성능 향상을 위해 프레임 밖의 데이터를 사용하는 지연을 말한다. 이 지연은 다른 지연과 달리 구현방식에 의해 줄어들 수 없다. G.729.1의 알고리즘 지연은 총 48.9375ms이다. 프레임 길이에 의하여 20ms, MDCT를 위한 look-ahead 20ms, 협대역 LPC 분석을 위한 look-ahead로 5ms, QMF 필터뱅크를 위해 3.9375ms 지연이 발생한다. 만약 입력이 8kHz 샘플링 주파수이거나 저지연 모드로 사용한다면 알고리즘 지연은 25ms이다. G.729.1 복잡도는 35.79 WMOPS이다. 슈퍼와이드밴드 확장 모듈은 G.729.1이나 G.718[4]과의 호환성과 대역 확장성을 제공한다.

즉, 기존의 G.729.1에 슈퍼와이드밴드 대역폭 확장을 적용하여 협대역, 광대역, 슈퍼와이드밴드 대역 신호를 점진적으로 제공하여 고품질의 통화연결을 제공, 현장감있는 원격회의 서비스, 음악광고, e-러닝 등 응용서비스의 다양화를 가능하게 한다. 또한 스테레오 기능을 통해 음악품질을 향상시킬 수 있다.

(그림 7)에 현재 개발중인 G.729.1과 G.718 코덱을 오디오대역까지 확장한 모노(mono)급 슈퍼와이드밴드 코덱의 부호화기와 복호화기의 구조를 예시[5]하였으며, 시스템의 복잡도를 낮추기 위해 G.729.1 코덱의 기능블록을 가능한 한 재사용하기 위해 노력함을 알 수 있다. (그림 7)의 복호화기에서 32kHz로 샘플링된 입력 신호는 IIR QMF 필터를 사용하여 2개의 16kHz로 샘플링된 부대역 신호로 나누어진다. 광대역 신호는 G.729.1 코어코덱을 거쳐 부호화하되 2개의 예외를 둔다. 선택적으로 4kHz 대역분할을 위해 G.729.1 코덱에서 사용하는 IIR QMF 분석 필터뱅크를 사용할 수 있으며, 32kHz 입력신호의 경우 4,000~8,000Hz의 3kHz 저대역 전처리 과정은 TDAC 전단에서 바이패스 한다. 8,000~16,000Hz 처리과정에서는 spectral mirroring 기법, 지연 또는 위상보상과 6kHz 저대역 필터를 사용하여 전처리한다. 이 신호가 SWB 부호화기에서 처리되고 코어코덱의 TDAC에서 처리된 0~8,000Hz 변환계수와 함께 전송된다.



(그림 7) G.729.1과 G.718 슈퍼와이드밴드 코덱의 예

SWB 부호화기는 ATE, 파라미터 세부구조 부호화 기법, gain-shape 기반의 MDCT 계수의 부호화를 사용하며, 특기사항으로 G.729.1 코덱이 잘 처리하지 못하는 톤과 같은 정보처리를 위해 부가적인 TDAC 정보를 송신하는 방법으로 광대역 향상 기능을 부가하였다.

SWB 복호화기는 세부구조 재발생 기법과 스펙트럼 셰이핑, GLCVQ에 의한 품질 정밀화와 후처리, 적응적인 temporal 역정규화 기법, TDAC의 부가 정보를 처리하는 광대역 향상 기술을 사용한다. (그림 7)의 복호화기에서는 정상모드에서는 G.729.1 코어계층 신호가 (그림 6)의 G.729.1 복호화기를 사용하여 처리하고 예외적으로 부가적인 변환계수가 TDAC 복호화기로 바이패스 될 수 있도록 한다. IIR QMF 모드에서는 합성필터 뱅크는 위상등화기를 사용한 IIR QMF 기법으로 대체한다. 8,000~16,000Hz 영역의 신호는 지연 또는 위상보상기를 사용하여 처리한다. 0~8,000Hz, 8,000~16,000Hz 대역의 신호는 위상등화기를 갖는 IIR QMF 합성기법을 사용하여 결합하여 복호화를 수행한다.

복잡도는 32kHz 샘플링의 모노신호인 경우, 14.6 WMOPS(G.729.1 제외)이며, 알고리즘 지연은 위상 등화기를 갖는 IIR QMF 필터뱅크에서 2.1875ms 추가된다. ITU-T SG16은 2009년 10월 회의에서 모노급 슈퍼와이드밴드 코덱에 대한 표준 승인을 추진하고, 2010년 7월에 스테레오급 슈퍼와이드밴드 코덱 기술에 대한 표준을 승인할 예정이다.

IV. 시사점 및 결론

본 논문에서는 ITU-T에서의 표준코덱 기술을 중심으로 통신과 방송서비스가 결합되고 단말 내부에서는 디바이스의 통합에 따라 코덱의 개수를 최소화하기 위한 노력인 음성코덱 기반 오디오 융합코덱의 기술동향에 대해 살펴보았다.

이러한 기술개발은 ITU-T 뿐만 아니라 3GPP와 MPEG 오디오 그룹에서도 활발히, 그리고 경쟁적으로 진행되고 있다.

3GPP에서는 SA4 산하에 EVS SWG을 구성하여 AMR-WB, AMR-WB+ 코덱에 후속하여 LTE 서비스를 위한 EVS의 표준화를 위해 기술적 요구 사항을 도출하고 있으며 이는 기본적으로 음성정보와 오디오 대역(20~20,000Hz)의 정보를 모두 처리할 수 있는 음성 오디오 융합코덱 형태의 기술개발과 표준화가 진행될 것으로 예상된다. ETRI를 비롯한 LG전자, 노키아, 보이스에이지, 에릭슨, 퀄컴, 화웨이, 모토로라, 프랑스텔레콤, 오렌지, 프라운호퍼, 돌비 등이 참여하고 있다.

MPEG에서는 저지연 저복잡도와 고효율 오디오 코덱 기술을 활용하여 음성 및 오디오, 이들의 혼성신호를 하나의 코덱으로 처리할 수 있는 USAC 기술을 2007년 10월부터 개발하고 표준화를 진행하고 있으며, 2008년 7월 보이스에이지와 프라운호퍼의 코덱이 이 성능기능을 만족하는 RMO로 선정되었다. USAC 코덱 RMO의 성능 기준은 MPEG의 HE-AACv2 코

텍과 3GPP의 AMR-WB+ 코덱을 기준으로 음성, 음악, 혼성신호에 대해 우수한 성능을 제공하는 것으로 선정되었다. 선정된 USAC RMO 코덱은 입력신호의 주 특성(음성, 음악, 혼성)을 구분하여 음성신호는 AMR을 사용하고, 오디오신호는 AAC 기술을, 저전송률 혼성신호는 TCX 기술을 사용한다. 여기에 부가적으로 고대역 신호를 파라미터로 부호화하는 SBR 기술, 스테레오화를 위한 MPEG 서라운드 기술, 서로 다른 기술 적용시 발생하는 신호 왜곡을 최소화하기 위한 연결기술을 부가하여 사용한다. 이러한 MPEG USAC 코덱 기술은 여러 종류의 코덱을 내부에 병합한 하이브리드 구조로서 융합형 코덱은 아니지만 음성과 오디오 정보를 현재 기술을 사용하여 잘 처리하는 현실적 대안이 될 수 있을 것으로 보인다.

음성 오디오 융합코덱 기술은 음성과 오디오 신호를 동시에 처리하는 단일 코덱으로 발전하여 실시간 양방향 통신과 멀티캐스트형 서비스를 효과적으로 지원하기 위한 구조 기술, 부호화 효율 향상 기술, 대역폭 확장시 품질저하 보상 기술 확보를 통해 이루어질 것으로 전망한다. 또한 사람의 청각모델과 발생모델을 통합할 수 있는 모델링 기술이 주요 원천 기술일 것으로 판단한다.

아울러 이러한 핵심원천 기술의 확보는 이동 융합 멀티미디어 서비스를 지향하는 스마트폰, 양방향 IPTV, e-러닝 등과 같은 양방향 멀티미디어 서비스를 위한 필수기술로서 지적재산권과 표준권리의 확보는 차기 성장동력 확보와 직결되어 있어, 관련 기술의 개발 및 표준화를 위한 기관간, 국가간 경쟁이 ITU-T, MPEG, 3GPP를 중심으로 코덱 전쟁이라고 부를 만큼 치열해지고 있다. 따라서 휴대폰과 스마트폰 등 국내의 성장동력을 지속적으로 유지하고 확대하기 위해서는 음성 오디오 융합코덱 기술의 확보는 매우 중요하며 기술의 확보뿐 아니라 표준권리 확보를 위한 국제협력 네트워크 또한 더욱 강화하여야 한다. 이는 ITU-T와 MPEG의 JVT와 같은 글로벌 표준화에 의해 2003년 5월 H.264/MPEG-4 AVC 비디오 코덱의 성공적인 사례가 있고 음성 오디오 기술 분야에서도 가능할 것으로 예상하기 때문이다.

기술적으로 코덱간의 호환성 문제를 근원적으로 해결하고 음성서비스와 음악서비스를 대역폭 확장을 통해 동시에 제공할 수 있는 임베디드 가변비트를 음성기반 오디오 융합코덱 기술은 매우 유력하다. 이는 대역폭과 비트 레벨의 용량의 선택과 연동성이 우수하므로 유무선통합(FMC) 환경, 디바이스 통합 환경, 이중 통신망간 서로 다른 QoS 협상 및 제어를 매우 동적으로 수행할 수 있다. 음성기반 오디오 융합코덱 기술은 앞으로 와이브로, 무선랜, 4G 이동망 등에서 인터넷전화(VoIP) 서비스, 오디오 컨퍼런싱, 원격교육, 실시간 인터넷 게임, 대화형 IPTV와 같은 서비스에서 융합멀티미디어 서비스를 제공하는 데 널리 사용될 것으로 전망한다.

약어 정리

AAC	Advanced Audio Coding
ACELP	Algebraic CELP
AMR	Adaptive Multi-Rate
AMR+	Adaptive Multi-Rate Plus
ATE	Adaptive Temporal Envelope
AVC	Advanced Video Codec
CELP	Code-Excited Linear Prediction
DMB	Digital Multimedia Broadcasting
EVS	Enhanced Voice Codecs
FEC	Frame Erasure Concealment
FMC	Fixed Mobile Convergence
FR/HR	Full Rate/Half Rate
GLCVQ	Gossett Low Complexity VQ
GSM	Global Systems for Mobile
HB	Higher Band
HE-AAC	High Efficiency AAC
IIR	Infinite Impulse Response
ITU	International Telecommunication Union
JVT	Joint Video Team
LB	Lower Band
LPC	Linear Predictive Coding
MDCT	Modified Discrete Cosine Transform
PMP	Portable Multimedia Player
QMF	Quadrature Mirror Filter bank
RM	Reference Model

SBR	Spectral Band Replication
TCX	Transform-Coded eXcitation
TDAC	Time-Domain Aliasing Cancellation
TDBWE	Time-Domain Bandwidth Extension
TPC	Track-Pulse Coding
USAC	Unified Speech and Audio Codec
VoIP	Voice over IP
VQ	Vector Quantization
WMOPS	Weighted Million Operations Per Second

참 고 문 헌

- [1] ETRI, "High-level Description of ETRI Candidate for G.722/G.711 SWB Extension," ITU-T WP3/SG16 AC-0907-Q10-08, Geneva, July 2009.
- [2] 김현우 외, "G.729.1 광대역 멀티코덱 표준 기술 동향," 전자통신동향분석, 제21권 제6호, 2006. 12., pp.77-85.
- [3] ITU-T Rec. G.729.1, "An 8-32kbit/s Scalable Wideband Coder Bit stream Interoperable with G.729," June 2006.
- [4] ITU-T G.718, "Frame Error Robust Narrow-band and Wideband Embedded Variable Bit-rate Coding of Speech and Audio from 8-32kbit/s," 2008. 6.
- [5] Bernd Geiser et al., "Candidate Proposal for ITU-T Super-wideband Speech and Audio Coding," *ICASSP 2009*, Apr. 2009, pp.4121-4124.