



(19)대한민국특허청(KR)
(12) 등록특허공보(B1)

(51) 。 Int. Cl.	(45) 공고일자	2007년04월25일
G10L 19/00 (2006.01)	(11) 등록번호	10-0711280
G10L 19/14 (2006.01)	(24) 등록일자	2007년04월18일

(21) 출원번호	10-2005-7006204	(65) 공개번호	10-2005-0049537
(22) 출원일자	2005년04월09일	(43) 공개일자	2005년05월25일
심사청구일자	2005년04월20일		
변역문 제출일자	2005년04월09일		
(86) 국제출원번호	PCT/CA2003/001571	(87) 국제공개번호	WO 2004/034379
국제출원일자	2003년10월09일	국제공개일자	2004년04월22일

(30) 우선권주장 60/417,667 2002년10월11일 미국(US)

(73) 특허권자 노키아 코포레이션
핀란드핀-02150 에스푸 카일알라텐티에 4

(72) 발명자 젤리넥 밀란
캐나다 퀘벡 제이1에이치 1케이4 서브룩 왈튼 925

(74) 대리인 리엔목특허법인

(56) 선행기술조사문헌
TAMMI, "Signal modification for voiced wideband spe WO01/22402
US05911128
* 심사관에 의하여 인용된 문헌

심사관 : 장호근

전체 청구항 수 : 총 84 항

(54) 소스 제어되는 가변 비트율 광대역 음성 부호화 방법 및장치

(57) 요약

음성 신호 분류 및 부호화 시스템 및 방법이 개시된다. 신호 분류는 3 단계에서 수행되고, 그들 각각은 특정 신호 클래스를 식별한다. 우선, 음성 활동 검출기(VAD)가 액티브(active) 및 인액티브(inactive) 음성 프레임들을 구별한다. 인액티브 음성 프레임이 검출되는 경우(배경 잡음 신호), 분류 단계가 종료하고 프레임은 키포트 잡음 생성(CNG)을 이용하여 부호화된다. 액티브 음성 프레임이 검출되는 경우, 프레임은 무성음 프레임을 식별하도록 구성된 제2 분류기에 제공된다. 분류기가 프레임을 무성음 음성 신호로서 분류하는 경우, 분류 단계가 종료하고, 프레임은 무성음 신호에 최적화된 부호화 방법을 이용하여 부호화된다. 그렇지 않은 경우, 음성 프레임은 "안정된 유성음(stable voiced)" 분류 모듈에 전달된다. 프레임이 안정된 유성음 프레임으로서 분류되는 경우, 상기 프레임은 안정된 유성음 신호에 최적화된 부호화 방법을 이용하여 부호화된다. 그렇지 않은 경우, 프레임은 유성음 온셋(onset) 또는 급속히 전개되는 유성음 음성 신호와 같은 비정적 음성 세그먼트를 포함할 것이다. 이 경우에 있어서, 좋은 주관적인 품질을 유지하기 위해 고 비트율을 갖는 일반 목적 음성 부호기가 이용된다.

대표도

도 2

특허청구의 범위

청구항 1.

음성 프레임들을 포함하는 샘플링된 음성 신호를 부호화하는 방법에 있어서,

상기 샘플링된 음성 신호의 현재 프레임이 액티브(active) 음성 프레임인지 인액티브(inactive) 음성 프레임인지를 결정하는 단계;

상기 현재 프레임이 액티브 음성 프레임인 경우, 상기 현재 프레임이 무성음 프레임인지 여부를 결정하기 위한 분류 절차를 수행하는 단계로서, 상기 분류 절차는 상기 현재 프레임이 무성음 프레임인지 여부를 결정하기 위하여 다음 매개변수들

- a) 유성음화 값(voicing measure)($r_x, \overline{r_x}$);
- b) 스펙트럼 틸트 값(spectral tilt measure)(e_{tilt}, e_t);
- c) 상기 현재 프레임내의 에너지 변동(dE);
- d) 상기 현재 프레임의 상대 에너지(E_{rel});

중 적어도 3개를 검사하는 단계를 포함하는 단계; 및

상기 현재 프레임이 상기 분류 절차에 의해 무성음 프레임으로서 분류되는 경우, 상기 현재 프레임을 무성음 신호 부호화 알고리즘을 이용하여 부호화하는 단계를 포함하는 것을 특징으로 하는 부호화 방법.

청구항 2.

제1항에 있어서,

상기 유성음화 값($\overline{r_x}$)은

$$\overline{r_x} = \frac{1}{3} (r_x(0) + r_x(1) + r_x(2))$$

와 같이 정의되고, $r_x(0)$ 은 상기 현재 프레임의 제1 하프의 정규화된 상관(normalized correlation)이고, $r_x(1)$ 은 상기 현재 프레임의 제2 하프의 정규화된 상관이며, $r_x(2)$ 는 상기 현재 프레임 다음의 프레임의 제1 하프의 정규화된 상관인 것을 특징으로 하는 부호화 방법.

청구항 3.

제2항에 있어서,

잡음 정정 인자(r_e)를 상기 유성음화 값($\overline{r^x}$)에 추가하는 단계를 더 포함하는 것을 특징으로 하는 부호화 방법.

청구항 4.

제1항에 있어서,

최저 주파수 범위를 나타내는 최초 인식 임계 대역(perceptual critical band)으로부터 최고 주파수 범위를 나타내는 최종 인식 임계 대역까지의 증가하는 주파수에 따라 배열된, 상기 현재 프레임의 에너지 스펙트럼내의 주파수 범위를 나타내는 다수의 인식 임계 대역들을 정의하는 단계; 및

상기 현재 프레임의 스펙트럼 분석을 수행하여 상기 인식 임계 대역들 중의 에너지의 분포를 결정하는 단계를 포함하는 것을 특징으로 하는 부호화 방법.

청구항 5.

제1항에 있어서,

상기 스펙트럼 틸트는 저주파에서의 상기 현재 프레임의 에너지 및 고주파에서의 상기 현재 프레임의 에너지간의 비(ratio)에 비례하는 것을 특징으로 하는 부호화 방법.

청구항 6.

제4항에 있어서,

최종 2개의 인식 임계 대역들의 에너지의 평균을 계산함으로써 고주파에서의 상기 현재 프레임의 에너지를 나타내는 값

($\overline{E_h}$)을 계산하는 단계를 포함하는 것을 특징으로 하는 부호화 방법.

청구항 7.

제4항에 있어서,

최초 i개의 인식 임계 대역들의 에너지의 평균을 계산함으로써 저주파에서의 상기 현재 프레임의 에너지를 나타내는 값

($\overline{E_l}$)을 계산하는 단계를 포함하는 것을 특징으로 하는 부호화 방법.

청구항 8.

제4항에 있어서,

최초 인식 임계 대역을 제외한 최초 i개의 인식 임계 대역들의 에너지의 평균을 계산함으로써 저주파에서의 상기 현재 프

레이프의 에너지를 나타내는 값($\overline{E_l}$)을 계산하는 단계를 포함하는 것을 특징으로 하는 부호화 방법.

청구항 9.

제7항에 있어서,

음성 피치 기간을 결정하는 단계, 및 소정 값보다 더 짧은 음성 피치 기간에 대해 상기 현재 프레임의 스펙트럼 분석에 기인하는 주파수 빈(bin)들내의 에너지를 합하여 저주파 에너지 값(\overline{E}_l)을 계산하는 단계를 더 포함하고, 음성 하모닉(harmonics)에 충분히 근접한 주파수 빈들만이 다음 수학식에 따라 상기 합계에 고려되며,

$$\overline{E}_l = \frac{1}{cnt} \sum_{k=K_{min}}^{24} E_{BIN}(k) w_h(k)$$

$E_{BIN}(k)$ 은 주파수 빈내의 에너지이고, K_{min} 은 상기 합계에서 고려된 제1 주파수 빈의 인덱스이며, cnt는 합계에서 제로가 아닌 항의 수이고, 주파수 빈 및 가장 근접한 하모닉간의 거리가 소정 주파수 임계값보다 더 크지 않은 경우 $w_h(k)$ 는 1로 세팅되고, 그 외의 경우는 $w_h(k)$ 는 0으로 세팅되는 것을 특징으로 하는 부호화 방법.

청구항 10.

제7항에 있어서,

음성 피치 기간을 결정하는 단계, 및 소정 값보다 더 큰 피치 값에 대해 다음 수학식에 따라 저주파 에너지 값(\overline{E}_l)을 계산하는 단계를 더 포함하고,

$$\overline{E}_l = \frac{1}{i} \sum_{k=0}^{i-1} E_{CB}(k)$$

$E_{CB}(k)$ 는 인식 임계 대역 k의 에너지인 것을 특징으로 하는 부호화 방법.

청구항 11.

제7항에 있어서,

$r_x(0)$ 은 상기 현재 프레임의 제1 하프의 정규화된 상관(normalized correlation)이고, $r_x(1)$ 은 상기 현재 프레임의 제2 하프의 정규화된 상관이며, r_e 은 잡음 정정 인자일 때에,

$r_x(0) + r_x(1) + r_e < 0.6$ 인 경우에 이전 무성음 사운드를 식별하는 단계, 및

다음 수학식에 따라 저주파 에너지 값(\overline{E}_l)을 계산하는 단계를 더 포함하고,

$$\overline{E}_l = \frac{1}{i} \sum_{k=0}^{i-1} E_{CB}(k)$$

$E_{CB}(k)$ 는 인식 임계 대역 k의 에너지인 것을 특징으로 하는 부호화 방법.

청구항 12.

제6항 내지 제11항 중의 어느 한 항에 있어서,

- 최종 2개의 인식 임계 대역의 잡음 에너지의 평균을 계산함으로써 고주파에서의 현재 프레임의 잡음 에너지를 나타내는 값(N_h)을 계산하는 단계;

- 최초 i 개의 인식 임계 대역의 잡음 에너지의 평균을 계산함으로써 저주파에서의 현재 프레임의 잡음 에너지를 나타내는 값(N_l)을 계산하는 단계;

- 상기 고주파 에너지 값($\overline{E_h}$)에서 상기 고주파 잡음 값(N_h)을 감산하여 고주파 에너지(E_h)를 획득하는 단계;

- 상기 저주파 에너지 값($\overline{E_l}$)에서 상기 저주파 잡음 값(N_l)을 감산하여 저주파 에너지(E_l)를 획득하는 단계; 및

- 상기 고주파 에너지(E_h)에 의해 나누어진 상기 저주파 에너지(E_l)의 비로서 스펙트럼 틸트 값(e_{tilt})을 계산하는 단계를 더 포함하는 것을 특징으로 하는 부호화 방법.

청구항 13.

제12항에 있어서,

제4항의 스펙트럼 분석을 상기 현재 프레임에 대해 2번, 상기 현재 프레임의 제1 하프에 대해 1번, 상기 현재 프레임의 제2 하프에 대해 1번 수행하는 단계, 및 추가로 상기 스펙트럼 틸트 값(e_{tilt})을 상기 현재 프레임에 대해 2번, 각 스펙트럼 분석에 대해 1번 계산하여 상기 현재 프레임의 제1 하프에 대해 제1 스펙트럼 틸트 값($e_{tilt}(0)$) 및 상기 현재 프레임의 제2 하프에 대해 제2 스펙트럼 틸트 값($e_{tilt}(1)$)을 획득하는 단계를 포함하는 것을 특징으로 하는 부호화 방법.

청구항 14.

제13항에 있어서,

다음 수학식에 따라 평균 스펙트럼 틸트(e_t)를 계산하는 단계를 더 포함하고,

$$e_t = \frac{1}{3} (e_{old} + e_{tilt}(0) + e_{tilt}(1))$$

e_{old} 는 이전 프레임의 제2 하프의 스펙트럼 분석으로부터 획득된 스펙트럼 틸트 값인 것을 특징으로 하는 부호화 방법.

청구항 15.

제1항에 있어서,

dB 단위의 프레임 에너지(E_t) 및 장기간 평균 에너지값($\overline{E_f}$) 간의 차로서 상기 현재 프레임의 상대 에너지(E_{rel})를 계산하는 단계를 포함하는 것을 특징으로 하는 부호화 방법.

청구항 16.

제15항에 있어서,

다음 수학적식에 따라 프레임 에너지(E_t)를 계산하는 단계를 포함하고,

$$E_t = 10 \log \left(\sum_{i=0}^{19} E_{CB}(i) \right), \quad dB$$

$E_{CB}(i)$ 는 임계 대역에 대한 평균 에너지인 것을 특징으로 하는 부호화 방법.

청구항 17.

제15항에 있어서,

다음 수학적식에 따라 상기 장기간 평균 에너지값을 갱신하는 단계를 포함하고,

$$\overline{E_f} = 0.99 \overline{E_f} + 0.01 E_t$$

$\overline{E_f}$ 는 45dB인 초기값을 갖는 것을 특징으로 하는 부호화 방법.

청구항 18.

제1항에 있어서,

한 세트의 이용가능한 부호화 비트율로부터 부호화 비트율을 선택하는 단계, 및 상기 선택된 부호화 비트율에 따라 상기 현재 프레임을 부호화하는 단계를 포함하는 것을 특징으로 하는 부호화 방법.

청구항 19.

제18항에 있어서,

상기 한 세트의 이용가능한 부호화 비트율은 풀 레이트(full-rate) 부호화 비트율, 하프 레이트(half-rate) 부호화 비트율, 1/4 레이트(quarter-rate) 부호화 비트율 및 1/8 레이트(eighth-rate) 부호화 비트율을 포함하는 것을 특징으로 하는 부호화 방법.

청구항 20.

제19항에 있어서,

상기 현재 프레임이 무성음 프레임으로서 분류되는 경우, 무성음 하프 레이트 부호화 알고리즘을 이용하여 상기 하프 레이트 부호화 비트율로 상기 현재 프레임을 부호화하는 단계를 포함하는 것을 특징으로 하는 부호화 방법.

청구항 21.

제19항에 있어서,

상기 현재 프레임이 무성음 프레임인지 여부를 결정하기 위한 분류 절차는 상기 현재 프레임이 유성음 및 무성음 간의 전이(transition)에 있는지 여부를 결정하는 단계; 상기 현재 프레임이 무성음 프레임으로서 분류되고 유성음 및 무성음 간의 전이에 있는 경우, 무성음 하프 레이트 부호화 알고리즘을 이용하여 상기 하프 레이트 부호화 비트율로 상기 현재 프레임을 부호화하는 단계; 및 상기 현재 프레임이 무성음 음성으로 분류되고 유성음 및 무성음 간의 전이에 있지 않은 경우, 무성음 1/4 레이트 부호화 알고리즘을 이용하여 상기 1/4 레이트 부호화 비트율로 상기 현재 프레임을 부호화하는 단계를 더 포함하는 것을 특징으로 하는 부호화 방법.

청구항 22.

제1항에 있어서,

상기 현재 프레임이 인액티브 음성 프레임인 것으로 결정된 경우 콤프트 잡음 생성 알고리즘을 이용하는 단계를 포함하는 것을 특징으로 하는 부호화 방법.

청구항 23.

제1항에 있어서,

상기 현재 프레임이 인액티브 음성 프레임인 것으로 결정된 경우 불연속 전송 모드를 이용하는 단계를 포함하는 것을 특징으로 하는 부호화 방법.

청구항 24.

제19항에 있어서,

각각의 동작 모드가 소정의 평균 비트율을 제공하는 한 세트의 동작 모드를 정의하는 단계, 동작 모드를 선택하는 단계, 및 상기 선택된 동작 모드에 따라 샘플링된 음성 신호를 부호화하는 단계를 포함하는 것을 특징으로 하는 부호화 방법.

청구항 25.

제24항에 있어서,

상기 한 세트의 동작 모드는 최고의 평균 비트율을 갖는 프리미엄(Premium) 모드, 중간 평균 비트율을 갖는 표준(Standard) 모드, 및 최저의 평균 비트율을 갖는 절약(Economy) 모드를 포함하는 것을 특징으로 하는 부호화 방법.

청구항 26.

제25항에 있어서,

상기 샘플링된 음성 신호가 프리미엄 모드에서 부호화되고 상기 현재 프레임이 무성음 프레임으로서 분류되는 경우, 상기 현재 프레임은 다음 조건이 충족되는 경우 상기 하프 레이트 부호화 비트율로 부호화되며,

다음 조건은

상기 유성음화 값이 소정의 제1 임계값보다 더 작고;

상기 스펙트럼 틸트 값이 소정의 제2 임계값보다 더 작으며;

상기 에너지 변동이 소정의 제3 임계값보다 더 작은 것을 특징으로 하는 부호화 방법.

청구항 27.

제25항에 있어서,

상기 샘플링된 음성 신호가 표준 모드에서 부호화되고 상기 현재 프레임이 무성음 프레임으로서 분류되는 경우, 상기 현재 프레임은 다음 조건이 충족되는 경우 상기 하프 레이트 부호화 비트율로 부호화되며,

다음 조건은

상기 유성음화 값이 소정의 제4 임계값보다 더 작고;

상기 스펙트럼 틸트 값이 소정의 제5 임계값보다 더 작으며;

상기 에너지 변동이 소정의 제6 임계값보다 더 작거나 상기 상대 에너지가 소정의 제7 임계값보다 더 작은 것을 특징으로 하는 부호화 방법.

청구항 28.

제27항에 있어서,

상기 제4 임계값은 0.695이고, 상기 제5 임계값은 4이며, 상기 제6 임계값은 40이고, 상기 제7 임계값은 -14인 것을 특징으로 하는 부호화 방법.

청구항 29.

제25항에 있어서,

상기 샘플링된 음성 신호가 절약 모드에서 부호화되고 상기 현재 프레임이 무성음 프레임으로서 분류되는 경우, 상기 현재 프레임은 다음 조건이 충족되는 경우 상기 하프 레이트 부호화 비트율로 부호화되며,

다음 조건은

상기 유성음화 값이 소정의 제8 임계값보다 더 작고;

상기 스펙트럼 틸트 값이 소정의 제9 임계값보다 더 작으며;

상기 에너지 변동이 소정의 제10 임계값보다 더 작거나 상기 상대 에너지가 소정의 제11 임계값보다 더 작은 것을 특징으로 하는 부호화 방법.

청구항 30.

제29항에 있어서,

상기 제8 임계값은 0.695이고, 상기 제9 임계값은 4이며, 상기 제10 임계값은 60이고, 상기 제11 임계값은 -14인 것을 특징으로 하는 부호화 방법.

청구항 31.

제25항에 있어서,

상기 샘플링된 음성 신호가 절약 모드에서 부호화되고 상기 현재 프레임이 무성음 프레임으로서 분류되는 경우, 상기 현재 프레임은 다음의 추가 조건이 충족되는 경우 상기 1/4 레이트 부호화 비트율로 부호화되며,

다음의 추가 조건은

미리보기(lookahead) 프레임에서의 정규화된 상관($r_x(2)$)이 소정의 제12 임계값보다 더 작고;

상기 현재 프레임의 제2 하프에 대한 제2 스펙트럼 틸트 값($e_{\text{tilt}}(1)$)이 소정의 제13 임계값보다 더 작은 것을 특징으로 하는 부호화 방법.

청구항 32.

제31항에 있어서,

상기 제12 임계값은 0.73이고, 상기 제13 임계값은 3인 것을 특징으로 하는 부호화 방법.

청구항 33.

음성 프레임들을 포함하는 샘플링된 음성 신호를 부호화하는 장치에 있어서,

상기 샘플링된 음성 신호의 프레임들이 액티브(active) 음성 프레임들인지 인액티브(inactive) 음성 프레임들인지를 결정하는 음성 활동 검출기; 및

상기 액티브 음성 프레임들이 무성음 프레임들인지 여부를 결정하기 위하여 액티브 음성 프레임들에 대해 분류 절차를 수행하도록 이루어진 분류 유닛을 포함하고,

상기 분류 절차는 현재 프레임이 무성음 프레임인지 여부를 결정하기 위하여 다음 매개변수들

a) 유성음화 값(voicing measure)($r_x, \overline{r_x}$);

b) 스펙트럼 틸트 값(e_{tilt}, e_t);

- c) 상기 현재 프레임내의 에너지 변동(dE);
- d) 상기 현재 프레임의 상대 에너지(E_{rel});

중 적어도 3개를 검사하는 단계를 포함하고,

상기 장치는 상기 분류 유닛이 상기 현재 프레임을 무성음 프레임으로서 분류한 경우, 상기 현재 프레임을 무성음 신호 부호화 알고리즘을 이용하여 부호화하도록 이루어진 것을 특징으로 하는 부호화 장치.

청구항 34.

제33항에 있어서,

상기 유성음화 값($\overline{r_x}$)은

$$\overline{r_x} = \frac{1}{3} (r_x(0) + r_x(1) + r_x(2))$$

와 같이 정의되고, $r_x(0)$ 은 상기 현재 프레임의 제1 하프의 정규화된 상관(normalized correlation)이고, $r_x(1)$ 은 상기 현재 프레임의 제2 하프의 정규화된 상관이며, $r_x(2)$ 는 상기 현재 프레임 다음의 프레임의 제1 하프의 정규화된 상관인 것을 특징으로 하는 부호화 장치.

청구항 35.

제34항에 있어서,

추가로 잡음 정정 인자(r_e)를 상기 유성음화 값($\overline{r_x}$)에 추가하도록 이루어진 것을 특징으로 하는 부호화 장치.

청구항 36.

제33항에 있어서,

최저 주파수 범위를 나타내는 최초 인식 임계 대역으로부터 최고 주파수 범위를 나타내는 최종 인식 임계 대역까지의 증가하는 주파수에 따라 배열된, 상기 현재 프레임의 에너지 스펙트럼내의 주파수 범위를 나타내는 다수의 인식 임계 대역들을 정의하고, 상기 현재 프레임의 스펙트럼 분석을 수행하여 상기 인식 임계 대역들 중의 에너지의 분포를 결정하도록 이루어진 것을 특징으로 하는 부호화 장치.

청구항 37.

제33항에 있어서,

상기 스펙트럼 틸트는 저주파에서의 상기 현재 프레임의 에너지 및 고주파에서의 상기 현재 프레임의 에너지간의 비(ratio)에 비례하는 것을 특징으로 하는 부호화 장치.

청구항 38.

제36항에 있어서,

최종 2개의 인식 임계 대역들의 에너지의 평균을 계산함으로써 고주파에서의 상기 현재 프레임의 에너지를 나타내는 값 $\overline{E_h}$ 을 계산하도록 이루어진 것을 특징으로 하는 부호화 장치.

청구항 39.

제36항에 있어서,

최초 i개의 인식 임계 대역들의 에너지의 평균을 계산함으로써 저주파에서의 상기 현재 프레임의 에너지를 나타내는 값 $\overline{E_i}$ 을 계산하도록 이루어진 것을 특징으로 하는 부호화 장치.

청구항 40.

제36항에 있어서,

최초 인식 임계 대역을 제외한 최초 i개의 인식 임계 대역들의 에너지의 평균을 계산함으로써 저주파에서의 상기 현재 프레임의 에너지를 나타내는 값 $\overline{E_i}$ 을 계산하도록 이루어진 것을 특징으로 하는 부호화 장치.

청구항 41.

제39항에 있어서,

추가로 음성 피치 기간을 결정하고, 소정 값보다 더 짧은 음성 피치 기간에 대해 상기 현재 프레임의 스펙트럼 분석에 기인하는 주파수 bin(들)내의 에너지를 합하여 저주파 에너지 값 $\overline{E_i}$ 을 계산하고, 음성 하모닉(harmonics)에 충분히 근접한 주파수 bin들만을 다음 수학적식에 따라 상기 합계에 고려하도록 이루어지며,

$$\overline{E_i} = \frac{1}{cnt} \sum_{k=K_{min}}^{24} E_{BIN}(k) w_h(k)$$

$E_{BIN}(k)$ 은 주파수 bin내의 에너지이고, K_{min} 은 상기 합계에서 고려된 제1 주파수 bin의 인덱스이며, cnt는 합계에서 제로가 아닌 항의 수이고, 주파수 bin 및 가장 근접한 하모닉간의 거리가 소정 주파수 임계값보다 더 크지 않은 경우 $w_h(k)$ 는 1로 세팅되고, 그 외의 경우는 $w_h(k)$ 는 0으로 세팅되는 것을 특징으로 하는 부호화 장치.

청구항 42.

제39항에 있어서,

추가로 음성 피치 기간을 결정하고, 소정 값보다 더 큰 피치 값에 대해 다음 수학식에 따라 저주파 에너지 값(\overline{E}_l)을 계산하도록 이루어지며,

$$\overline{E}_l = \frac{1}{i} \sum_{k=0}^{i-1} E_{CB}(k)$$

$E_{CB}(k)$ 는 인식 임계 대역 k 의 에너지인 것을 특징으로 하는 부호화 장치.

청구항 43.

제39항에 있어서,

$r_x(0)$ 은 상기 현재 프레임의 제1 하프의 정규화된 상관(normalized correlation)이고, $r_x(1)$ 은 상기 현재 프레임의 제2 하프의 정규화된 상관이며, r_e 은 잡음 정정 인자일 때에,

추가로 $r_x(0) + r_x(1) + r_e < 0.6$ 인 경우에 이전 무성을 사운드를 식별하고, 다음 수학식에 따라 저주파 에너지 값(\overline{E}_l)을 계산하도록 이루어지며,

$$\overline{E}_l = \frac{1}{i} \sum_{k=0}^{i-1} E_{CB}(k)$$

$E_{CB}(k)$ 는 인식 임계 대역 k 의 에너지인 것을 특징으로 하는 부호화 장치.

청구항 44.

제38항 내지 제43항 중의 어느 한 항에 있어서, 추가로

- 최종 2개의 인식 임계 대역의 잡음 에너지의 평균을 계산함으로써 고주파에서의 현재 프레임의 잡음 에너지를 나타내는 값(N_h)을 계산하고;

- 최초 i 개의 인식 임계 대역의 잡음 에너지의 평균을 계산함으로써 저주파에서의 현재 프레임의 잡음 에너지를 나타내는 값(N_l)을 계산하며;

- 상기 고주파 에너지 값(\overline{E}_h)에서 상기 고주파 잡음 값(N_h)을 감산하여 고주파 에너지(E_h)를 획득하고;

- 상기 저주파 에너지 값(\overline{E}_l)에서 상기 저주파 잡음 값(N_l)을 감산하여 저주파 에너지(E_l)를 획득하며;

- 상기 고주파 에너지(E_h)에 의해 나누어진 상기 저주파 에너지(E_l)의 비로서 스펙트럼 틸트 값(e_{tilt})을 계산하도록 이루어진 것을 특징으로 하는 부호화 장치.

청구항 45.

제44항에 있어서,

제36항의 스펙트럼 분석을 상기 현재 프레임에 대해 2번, 상기 현재 프레임의 제1 하프에 대해 1번, 상기 현재 프레임의 제2 하프에 대해 1번 수행하고, 추가로 상기 스펙트럼 틸트 값(e_{tilt})을 상기 현재 프레임에 대해 2번, 각 스펙트럼 분석에 대해 1번 계산하여 상기 현재 프레임의 제1 하프에 대해 제1 스펙트럼 틸트 값($e_{\text{tilt}}(0)$) 및 상기 현재 프레임의 제2 하프에 대해 제2 스펙트럼 틸트 값($e_{\text{tilt}}(1)$)을 획득하도록 이루어진 것을 특징으로 하는 부호화 장치.

청구항 46.

제45항에 있어서,

추가로 다음 수학적식에 따라 평균 스펙트럼 틸트(e_t)를 계산하도록 이루어지며,

$$e_t = \frac{1}{3} (e_{\text{old}} + e_{\text{tilt}}(0) + e_{\text{tilt}}(1))$$

e_{old} 는 이전 프레임의 제2 하프의 스펙트럼 분석으로부터 획득된 스펙트럼 틸트 값인 것을 특징으로 하는 부호화 장치.

청구항 47.

제33항에 있어서,

dB 단위의 프레임 에너지(E_f) 및 장기간 평균 에너지값($\overline{E_f}$) 간의 차로서 상기 현재 프레임의 상대 에너지(E_{rel})를 계산하도록 이루어진 것을 특징으로 하는 부호화 장치.

청구항 48.

제47항에 있어서,

다음 수학적식에 따라 프레임 에너지(E_f)를 계산하도록 이루어지며,

$$E_f = 10 \log \left(\sum_{i=0}^{19} E_{CB}(i) \right), \quad dB$$

$E_{CB}(i)$ 는 임계 대역에 대한 평균 에너지인 것을 특징으로 하는 부호화 장치.

청구항 49.

제47항에 있어서,

다음 수학적식에 따라 상기 장기간 평균 에너지값을 갱신하도록 이루어지며,

$$\overline{E_f} = 0.99\overline{E_f} + 0.01E_t$$

$\overline{E_f}$ 는 45dB인 초기값을 갖는 것을 특징으로 하는 부호화 장치.

청구항 50.

제33항에 있어서,

한 세트의 이용가능한 부호화 비트율로부터 부호화 비트율을 선택하고, 상기 선택된 부호화 비트율에 따라 상기 현재 프레임 부호화하도록 이루어진 것을 특징으로 하는 부호화 장치.

청구항 51.

제50항에 있어서,

상기 한 세트의 이용가능한 부호화 비트율은 풀 레이트(full-rate) 부호화 비트율, 하프 레이트(half-rate) 부호화 비트율, 1/4 레이트(quarter-rate) 부호화 비트율 및 1/8 레이트(eighth-rate) 부호화 비트율을 포함하는 것을 특징으로 하는 부호화 장치.

청구항 52.

제51항에 있어서,

상기 현재 프레임이 무성음 프레임으로서 분류되는 경우, 무성음 하프 레이트 부호화 알고리즘을 이용하여 상기 하프 레이트 부호화 비트율로 상기 현재 프레임을 부호화하도록 이루어진 것을 특징으로 하는 부호화 장치.

청구항 53.

제51항에 있어서,

추가로 현재 무성음 프레임이 유성음 및 무성음 간의 전이(transition)에 있는지 여부를 결정하며, 상기 현재 프레임이 무성음 프레임으로서 분류되고 유성음 및 무성음 간의 전이에 있는 경우, 무성음 하프 레이트 부호화 알고리즘을 이용하여 상기 하프 레이트 부호화 비트율로 상기 현재 프레임을 부호화하고, 상기 현재 프레임이 무성음 음성으로 분류되고 유성음 및 무성음 간의 전이에 있지 않은 경우, 무성음 1/4 레이트 부호화 알고리즘을 이용하여 상기 1/4 레이트 부호화 비트율로 상기 현재 프레임을 부호화하도록 이루어진 것을 특징으로 하는 부호화 장치.

청구항 54.

제33항에 있어서,

상기 현재 프레임이 인액티브 음성 프레임인 것으로 결정된 경우 킴포트 잡음 생성 알고리즘을 이용하도록 이루어진 것을 특징으로 하는 부호화 장치.

청구항 55.

제33항에 있어서,

상기 현재 프레임이 인액티브 음성 프레임인 것으로 결정된 경우 불연속 전송 모드를 이용하도록 이루어진 것을 특징으로 하는 부호화 장치.

청구항 56.

제51항에 있어서,

각각의 동작 모드가 소정의 평균 비트율을 제공하는 한 세트의 동작 모드를 정의하고, 동작 모드를 선택하며, 상기 선택된 동작 모드에 따라 샘플링된 음성 신호를 부호화하도록 이루어진 것을 특징으로 하는 부호화 장치.

청구항 57.

제56항에 있어서,

상기 한 세트의 동작 모드는 최고의 평균 비트율을 갖는 프리미엄(Premium) 모드, 중간 평균 비트율을 갖는 표준(Standard) 모드, 및 최저의 평균 비트율을 갖는 절약(Economy) 모드를 포함하는 것을 특징으로 하는 부호화 장치.

청구항 58.

제57항에 있어서,

상기 샘플링된 음성 신호가 프리미엄 모드에서 부호화되고 상기 현재 프레임이 무성음 프레임으로서 분류되며 다음 조건이 충족되는 경우 상기 현재 프레임은 상기 하프 레이트 부호화 비트율로 부호화되도록 이루어지고,

다음 조건은

상기 유성음화 값이 소정의 제1 임계값보다 더 작고;

상기 스펙트럼 틸트 값이 소정의 제2 임계값보다 더 작으며;

상기 에너지 변동이 소정의 제3 임계값보다 더 작은 것을 특징으로 하는 부호화 장치.

청구항 59.

제57항에 있어서,

상기 샘플링된 음성 신호가 표준 모드에서 부호화되고 상기 현재 프레임이 무성음 프레임으로서 분류되며 다음 조건이 충족되는 경우 상기 현재 프레임은 상기 하프 레이트 부호화 비트율로 부호화되도록 이루어지고,

다음 조건은

상기 유성음화 값이 소정의 제4 임계값보다 더 작고;

상기 스펙트럼 틸트 값이 소정의 제5 임계값보다 더 작으며;

상기 에너지 변동이 소정의 제6 임계값보다 더 작거나 상기 상대 에너지가 소정의 제7 임계값보다 더 작은 것을 특징으로 하는 부호화 장치.

청구항 60.

제59항에 있어서,

상기 제4 임계값은 0.695이고, 상기 제5 임계값은 4이며, 상기 제6 임계값은 40이고, 상기 제7 임계값은 -14인 것을 특징으로 하는 부호화 장치.

청구항 61.

제57항에 있어서,

상기 샘플링된 음성 신호가 절약 모드에서 부호화되고 상기 현재 프레임이 무성음 프레임으로서 분류되며 다음 조건이 충족되는 경우 상기 현재 프레임은 상기 하프 레이트 부호화 비트율로 부호화되도록 이루어지며,

다음 조건은

상기 유성음화 값이 소정의 제8 임계값보다 더 작고;

상기 스펙트럼 틸트 값이 소정의 제9 임계값보다 더 작으며;

상기 에너지 변동이 소정의 제10 임계값보다 더 작거나 상기 상대 에너지가 소정의 제11 임계값보다 더 작은 것을 특징으로 하는 부호화 장치.

청구항 62.

제61항에 있어서,

상기 제8 임계값은 0.695이고, 상기 제9 임계값은 4이며, 상기 제10 임계값은 60이고, 상기 제11 임계값은 -14인 것을 특징으로 하는 부호화 장치.

청구항 63.

제57항에 있어서,

상기 샘플링된 음성 신호가 절약 모드에서 부호화되고 상기 현재 프레임이 무성음 프레임으로서 분류되며 다음의 추가 조건이 충족되는 경우 상기 현재 프레임은 상기 1/4 레이트 부호화 비트율로 부호화되도록 이루어지며,

다음의 추가 조건은

미리보기(lookahead) 프레임에서의 정규화된 상관($r_x(2)$)이 소정의 제12 임계값보다 더 작고;

상기 현재 프레임의 제2 하프에 대한 제2 스펙트럼 틸트 값($e_{\text{tilt}}(1)$)이 소정의 제13 임계값보다 더 작은 것을 특징으로 하는 부호화 장치.

청구항 64.

제63항에 있어서,

상기 제12 임계값은 0.73이고, 상기 제13 임계값은 3인 것을 특징으로 하는 부호화 장치.

청구항 65.

음성 프레임들을 포함하는 샘플링된 음성 신호를 부호화하는 장치에 있어서,

상기 샘플링된 음성 신호의 현재 프레임이 액티브(active) 음성 프레임인지 인액티브(inactive) 음성 프레임인지를 결정하는 수단;

상기 현재 프레임이 액티브 음성 프레임인 것에 응답하여, 상기 현재 프레임이 무성음 프레임인지 여부를 결정하기 위한 분류 절차를 수행하는 수단으로서, 상기 분류 절차는 상기 현재 프레임이 무성음 프레임인지 여부를 결정하기 위하여 다음 매개변수들

a) 유성음화 값(voicing measure)($r_x, \overline{r^x}$);

b) 스펙트럼 틸트 값(e_{tilt}, e_t);

c) 상기 현재 프레임내의 에너지 변동(dE);

d) 상기 현재 프레임의 상대 에너지(E_{rel});

중 적어도 3개를 검사하는 것을 포함하는 수단; 및

상기 현재 프레임이 상기 분류 절차에 의해 무성음 프레임으로서 분류되는 경우, 상기 현재 프레임을 무성음 신호 부호화 알고리즘을 이용하여 부호화하는 수단을 포함하는 것을 특징으로 하는 부호화 장치.

청구항 66.

음성 부호기에 있어서,

현재 프레임이 액티브 음성 프레임으로서 분류된 것에 응답하여, 상기 현재 프레임을 무성음 신호 부호화 알고리즘을 이용하여 부호화하고,

액티브 음성 프레임은 추가로 다음 세트: 유성음화 값(voicing measure)($r_x, \overline{r^x}$), 스펙트럼 틸트 값(e_{tilt}, e_t), 상기 현재 프레임내의 에너지 변동(dE), 및 상기 현재 프레임의 상대 에너지(E_{rel})로부터 선택된 적어도 3개의 매개변수들을 검사함으로써 액티브 무성음 음성으로서 분류되는 것을 특징으로 하는 음성 부호기.

청구항 67.

음성 프레임들을 포함하는 샘플링된 음성 신호를 부호화하도록 하는 단계를 수행하기 위하여, 디지털 데이터 프로세서에 의해 실행가능한 기계 독출가능한 명령어들을 포함하는 프로그램을 저장하는 컴퓨터 독출가능 저장 매체에 있어서,

상기 단계는

상기 샘플링된 음성 신호의 현재 프레임이 액티브(active) 음성 프레임인지 인액티브(inactive) 음성 프레임인지를 결정하는 단계;

상기 현재 프레임이 무성음 프레임인지 여부를 결정하기 위한 분류 절차를 액티브 음성 프레임에 관해 수행하는 단계로서, 상기 분류 절차는 상기 현재 프레임이 무성음 프레임인지 여부를 결정하기 위하여 다음 매개변수들

a) 유성음화 값(voicing measure)($r_x, \overline{r_x}$);

b) 스펙트럼 틸트 값(e_{tilt}, e_t);

c) 상기 현재 프레임내의 에너지 변동(dE);

d) 상기 현재 프레임의 상대 에너지(E_{rel});

중 적어도 3개를 검사하는 단계를 포함하는 단계; 및

상기 현재 프레임이 상기 분류 절차에 의해 무성음 프레임으로서 분류되는 경우, 상기 현재 프레임을 무성음 신호 부호화 알고리즘을 이용하여 부호화하는 단계를 포함하는 것을 특징으로 하는 컴퓨터 독출가능 저장 매체.

청구항 68.

제67항에 있어서,

상기 유성음화 값($\overline{r_x}$)은

$$\overline{r_x} = \frac{1}{3} (r_x(0) + r_x(1) + r_x(2))$$

와 같이 정의되고, $r_x(0)$ 은 상기 현재 프레임의 제1 하프의 정규화된 상관(normalized correlation)이고, $r_x(1)$ 은 상기 현재 프레임의 제2 하프의 정규화된 상관이며, $r_x(2)$ 는 상기 현재 프레임 다음의 프레임의 제1 하프의 정규화된 상관인 것을 특징으로 하는 컴퓨터 독출가능 저장 매체.

청구항 69.

제68항에 있어서, 상기 단계는

잡음 정정 인자(r_e)를 상기 유성음화 값($\overline{r_x}$)에 추가하는 단계를 더 포함하는 것을 특징으로 하는 컴퓨터 독출가능 저장 매체.

청구항 70.

제67항에 있어서, 상기 단계는

최저 주파수 범위를 나타내는 최초 인식 임계 대역으로부터 최고 주파수 범위를 나타내는 최종 인식 임계 대역까지의 증가하는 주파수에 따라 배열된, 상기 현재 프레임의 에너지 스펙트럼내의 주파수 범위를 나타내는 다수의 인식 임계 대역들을 정의하는 단계; 및

상기 현재 프레임의 스펙트럼 분석을 수행하여 상기 인식 임계 대역들 중의 에너지의 분포를 결정하는 단계를 더 포함하는 것을 특징으로 하는 컴퓨터 독출가능 저장 매체.

청구항 71.

제67항에 있어서,

상기 스펙트럼 틸트는 저주파에서의 상기 현재 프레임의 에너지 및 고주파에서의 상기 현재 프레임의 에너지간의 비 (ratio)에 비례하는 것을 특징으로 하는 컴퓨터 독출가능 저장 매체.

청구항 72.

제70항에 있어서, 상기 단계는

최종 2개의 인식 임계 대역들의 에너지의 평균을 계산함으로써 고주파에서의 상기 현재 프레임의 에너지를 나타내는 값 ($\overline{E_h}$)을 계산하는 단계를 더 포함하는 것을 특징으로 하는 컴퓨터 독출가능 저장 매체.

청구항 73.

제70항에 있어서, 상기 단계는

최초 i개의 인식 임계 대역들의 에너지의 평균을 계산함으로써 저주파에서의 상기 현재 프레임의 에너지를 나타내는 값 ($\overline{E_l}$)을 계산하는 단계를 더 포함하는 것을 특징으로 하는 컴퓨터 독출가능 저장 매체.

청구항 74.

제70항에 있어서, 상기 단계는

최초 인식 임계 대역을 제외한 최초 i개의 인식 임계 대역들의 에너지의 평균을 계산함으로써 저주파에서의 상기 현재 프레임의 에너지를 나타내는 값($\overline{E_l}$)을 계산하는 단계를 더 포함하는 것을 특징으로 하는 컴퓨터 독출가능 저장 매체.

청구항 75.

제73항에 있어서, 상기 단계는

음성 피치 기간을 결정하는 단계, 및 소정 값보다 더 짧은 음성 피치 기간에 대해 상기 현재 프레임의 스펙트럼 분석에 기인하는 주파수 빈(bin)들내의 에너지를 합하여 저주파 에너지 값(\overline{E}_l)을 계산하는 단계를 더 포함하고, 음성 하모닉(harmonics)에 충분히 근접한 주파수 빈들만이 다음 수학식에 따라 상기 합계에 고려되며,

$$\overline{E}_l = \frac{1}{cnt} \sum_{k=K_{min}}^{24} E_{BIN}(k)w_h(k)$$

$E_{BIN}(k)$ 은 주파수 빈내의 에너지이고, K_{min} 은 상기 합계에서 고려된 제1 주파수 빈의 인덱스이며, cnt는 합계에서 제로가 아닌 항의 수이고, 주파수 빈 및 가장 근접한 하모닉간의 거리가 소정 주파수 임계값보다 더 크지 않은 경우 $w_h(k)$ 는 1로 세팅되고, 그 외의 경우는 $w_h(k)$ 는 0으로 세팅되는 것을 특징으로 하는 컴퓨터 독출가능 저장 매체.

청구항 76.

제73항에 있어서, 상기 단계는

음성 피치 기간을 결정하는 단계, 및 소정 값보다 더 큰 피치 값에 대해 다음 수학식에 따라 저 주파수 에너지 값(\overline{E}_l)을 계산하는 단계를 더 포함하고,

$$\overline{E}_l = \frac{1}{i} \sum_{k=0}^{i-1} E_{CB}(k)$$

$E_{CB}(k)$ 는 인식 임계 대역 k의 에너지인 것을 특징으로 하는 컴퓨터 독출가능 저장 매체.

청구항 77.

제73항에 있어서,

$r_x(0)$ 은 상기 현재 프레임의 제1 하프의 정규화된 상관(normalized correlation)이고, $r_x(1)$ 은 상기 현재 프레임의 제2 하프의 정규화된 상관이며, r_e 은 잡음 정정 인자일 때에,

상기 단계는

$r_x(0) + r_x(1) + r_e < 0.6$ 인 경우에 이전 무성음 사운드를 식별하는 단계, 및

다음 수학식에 따라 저 주파수 에너지 값(\overline{E}_l)을 계산하는 단계를 더 포함하고,

$$\overline{E}_l = \frac{1}{i} \sum_{k=0}^{i-1} E_{CB}(k)$$

$E_{CB}(k)$ 는 인식 임계 대역 k의 에너지인 것을 특징으로 하는 컴퓨터 독출가능 저장 매체.

청구항 78.

제72항 내지 제77항 중의 어느 한 항에 있어서, 상기 단계는

- 최종 2개의 인식 임계 대역의 잡음 에너지의 평균을 계산함으로써 고주파에서의 현재 프레임의 잡음 에너지를 나타내는 값(N_h)을 계산하는 단계;

- 최초 i 개의 인식 임계 대역의 잡음 에너지의 평균을 계산함으로써 저주파에서의 현재 프레임의 잡음 에너지를 나타내는 값(N_l)을 계산하는 단계;

- 상기 고주파 에너지 값($\overline{E_h}$)에서 상기 고주파 잡음 값(N_h)을 감산하여 고주파 에너지(E_h)를 획득하는 단계;

- 상기 저주파 에너지 값($\overline{E_l}$)에서 상기 저주파 잡음 값(N_l)을 감산하여 저주파 에너지(E_l)를 획득하는 단계; 및

- 상기 고주파 에너지(E_h)에 의해 나누어진 상기 저주파 에너지(E_l)의 비로서 스펙트럼 틸트 값(e_{tilt})을 계산하는 단계를 더 포함하는 것을 특징으로 하는 컴퓨터 독출가능 저장 매체.

청구항 79.

제78항에 있어서, 상기 단계는

제70항의 스펙트럼 분석을 상기 현재 프레임에 대해 2번, 상기 현재 프레임의 제1 하프에 대해 1번, 상기 현재 프레임의 제2 하프에 대해 1번 수행하는 단계, 및 추가로 상기 스펙트럼 틸트 값(e_{tilt})을 상기 현재 프레임에 대해 2번, 각 스펙트럼 분석에 대해 1번 계산하여 상기 현재 프레임의 제1 하프에 대해 제1 스펙트럼 틸트 값($e_{tilt}(0)$) 및 상기 현재 프레임의 제2 하프에 대해 제2 스펙트럼 틸트 값($e_{tilt}(1)$)을 획득하는 단계를 더 포함하는 것을 특징으로 하는 컴퓨터 독출가능 저장 매체.

청구항 80.

제79항에 있어서, 상기 단계는

다음 수학식에 따라 평균 스펙트럼 틸트(e_t)를 계산하는 단계를 더 포함하고,

$$e_t = \frac{1}{3} (e_{old} + e_{tilt}(0) + e_{tilt}(1))$$

e_{old} 는 이전 프레임의 제2 하프의 스펙트럼 분석으로부터 획득된 스펙트럼 틸트 값인 것을 특징으로 하는 컴퓨터 독출가능 저장 매체.

청구항 81.

제67항에 있어서, 상기 단계는

dB 단위의 프레임 에너지(E_t) 및 장기간 평균 에너지값($\overline{E_f}$) 간의 차로서 상기 현재 프레임의 상대 에너지(E_{rel})를 계산하는 단계를 더 포함하는 것을 특징으로 하는 컴퓨터 독출가능 저장 매체.

청구항 82.

제81항에 있어서, 상기 단계는

다음 수학적식에 따라 프레임 에너지(E_t)를 계산하는 단계를 더 포함하고,

$$E_t = 10 \log \left(\sum_{i=0}^{19} E_{CB}(i) \right), \quad dB$$

$E_{CB}(i)$ 는 임계 대역에 대한 평균 에너지인 것을 특징으로 하는 컴퓨터 독출가능 저장 매체.

청구항 83.

제81항에 있어서, 상기 단계는

다음 수학적식에 따라 상기 장기간 평균 에너지값을 갱신하는 단계를 더 포함하고,

$$\overline{E_f} = 0.99 \overline{E_f} + 0.01 E_t$$

$\overline{E_f}$ 는 45dB인 초기값을 갖는 것을 특징으로 하는 컴퓨터 독출가능 저장 매체.

청구항 84.

제67항에 있어서,

상기 저장 매체 및 상기 디지털 데이터 프로세서는 이동국내에 배치되는 것을 특징으로 하는 컴퓨터 독출가능 저장 매체.

명세서

기술분야

본 발명은 사운드 신호의 디지털 부호화에 관한 것으로, 특히 배타적으로 음성 신호만이 아니라, 사운드 신호를 전송하고 합성하는 것에 관한 것이다. 특히, 본 발명은 가변 비트율(VBR; variable bit-rate) 음성 부호화를 위한 신호 분류 및 레이트 선택 방법에 관한 것이다.

배경기술

원격지간 회의(teleconferencing), 멀티미디어, 및 무선 통신과 같은 다양한 응용 영역에서 주관적인 품질과 비트율간에 잘 트레이드-오프(trade-off)되는 효율적인 디지털 협대역 및 광대역 음성 부호화 기술에 대한 요구가 증가하고 있다. 최근까지 200-3400 Hz의 범위로 제한된 전화 대역폭은 주로 음성 부호화 애플리케이션에서 사용되었다. 하지만, 광대역 음

성 애플리케이션은 종래의 전화 대역폭에 비해 통신하는데 더 자연스럽게 명료함을 제공한다. 50-7000 Hz 범위의 대역폭은 직접(face-to-face) 통신의 느낌을 주는 좋은 품질을 제공하기에 충분하다는 것을 알았다. 일반적인 오디오 신호에 있어서, 상기 대역폭은 수용할 수 있는 주관적인 품질을 제공하지만, 각각 20-16000 Hz 및 20-20000 Hz의 범위에서 동작하는 FM 라디오 또는 CD의 품질보다 훨씬 낮다.

음성 부호기는 음성 신호를 디지털 비트 스트림으로 변환한다. 디지털 비트 스트림은 통신 채널을 통해 전송되거나 저장 매체에 저장된다. 음성 신호는 디지털화된다. 즉 샘플링되고 샘플당 일반적으로 16 비트로 양자화된다. 음성 부호기는 좋은 주관적인 음성 품질을 유지하면서 더 적은 수의 비트들을 가지고 디지털 샘플들을 나타내는 역할을 한다. 음성 복호기 또는 합성기(synthesizer)는 전송되거나 저장된 비트 스트림에 작용하여 사운드 신호로 변환한다.

코드 여기된 선형 예측(CELP; Code-Excited Linear Prediction) 부호화는 주관적인 품질 및 비트율간의 좋은 절충안을 달성할 수 있게 하는 공지된 기법이다. 이 부호화 기법은 무선 및 유선 애플리케이션에서 몇몇 음성 부호화 표준의 기초이다. CELP 부호화에 있어서, 샘플링된 음성 신호는 일반적으로 프레임으로 지칭되는 L개의 샘플들의 연속 블록으로 처리된다. L은 전형적으로 10-30 ms에 대응하는 소정 수이다. 선형 예측(LP; linear prediction) 필터가 매 프레임마다 계산되고 전송된다. LP 필터의 계산은 전형적으로 다음 프레임으로부터 5-15 ms 음성 세그먼트를 미리보기(lookahead)할 필요가 있다. L-샘플 프레임은 서브프레임으로 지칭되는 더 작은 블록들로 분할된다. 일반적으로 서브프레임의 수는 3개 또는 4개이고, 따라서 4-10 ms인 서브프레임이 된다. 각 서브프레임에 있어서, 여기 신호는 보통 2개의 구성요소, 과거 여기 및 혁신적인, 고정-코드북 여기로부터 획득된다. 과거 여기로부터 형성된 구성요소는 종종 적응성 코드북(adaptive codebook) 또는 피치 여기(pitch excitation)로서 지칭된다. 여기 신호를 특징화하는 매개변수들이 부호화되고 복호기에 전송된다. 재구성된 여기 신호는 LP 필터의 입력으로서 사용된다.

코드 분할 다중 액세스(CDMA; code division multiple access) 기술을 이용하는 무선 시스템에 있어서, 소스 제어되는 가변 비트율(VBR) 음성 부호화의 사용은 시스템 용량을 상당히 개선한다. 소스 제어되는 VBR 부호화에 있어서, 코덱은 몇몇 비트율로 동작하고, 레이트 선택 모듈은 음성 프레임의 성질(예를 들어, 유성음, 무성음, 경과음(transient), 배경 잡음)에 근거하여 각 음성 프레임을 부호화하는데 사용되는 비트율을 결정하는데 사용된다. 또한 평균 데이터율(ADR; average data rate)로도 지칭되는 주어진 평균 비트율에서 최선의 음성 품질을 획득하는 것이 목표이다. 코덱은 코덱 성능이 증가된 ADR에서 개선되는 상이한 모드에서 상이한 ADR을 획득하도록 레이트 선택 모듈을 조정하여 상이한 모드로 동작할 수 있다. 동작 모드는 채널 상태에 의존하여 시스템에 의해 정해진다. 이것은 코덱이 음성 품질 및 시스템 용량간의 트레이드 오프의 메커니즘을 갖도록 허용한다.

전형적으로, CDMA 시스템의 VBR 부호화에 있어서, 음성 활동이 없는 프레임(침묵 또는 잡음만의 프레임)을 부호화하는데 1/8 레이트(eighth-rate)가 사용된다. 프레임이 정적 유성음(stationary voiced) 또는 정적 무성음(stationary unvoiced)인 경우, 동작 모드에 따라 하프 레이트 또는 1/4 레이트가 사용된다. 하프 레이트가 사용될 수 있는 경우, 피치 코드북이 없는 CELP 모델은 무성음의 경우에 사용되고 유성음의 경우에 피치 인덱스에 대한 비트의 수를 감소시키고 주기성을 향상하는데 신호 수정(signal modification)이 사용된다. 동작 모드가 1/4 레이트를 지정하는 경우, 비트의 수가 불충분하고 몇몇 매개변수 부호화가 일반적으로 적용되기 때문에 과형 매칭이 일반적으로 가능하지 않다. 풀 레이트(full-rate)는 온셋(onsets), 과도 프레임(transient frame), 및 혼합 유성음 프레임에 대해 사용된다(전형적인 CELP 모델이 일반적으로 사용된다). CDMA 시스템에서 소스 제어되는 코덱 동작에 추가하여, 상기 시스템은 코덱의 강인함을 향상시키기 위하여 (셀 경계 근처와 같은) 나쁜 채널 조건 동안 또는 (딤-앤-버스트 시그널링(dim-and-burst signalling)으로 지칭되는) 대역내 시그널링 정보를 전송하기 위하여 몇몇 음성 프레임들에서 최대 비트율을 제한할 수 있다. 이것은 하프 레이트 맥스(half-rate max)로 지칭된다. 레이트 선택 모듈이 풀 레이트 프레임으로서 부호화되는 프레임을 선택하고 상기 시스템이 예를 들어 HR 프레임을 지정하는 경우, 전용 HR 모드가 온셋 및 과도 신호를 효율적으로 부호화할 없기 때문에 음성 성능은 저하된다. 다른 HR(또는 1/4 레이트(QR; quarter-rate)) 부호화 모델이 상기 특별한 경우에 대처하기 위하여 제공될 수 있다.

상기 설명으로부터 알 수 있는 바와 같이, 신호 분류 및 레이트 결정이 효율적인 VBR 부호화를 위해 매우 중요하다. 레이트 선택은 최선의 가능한 품질을 가지고 최저의 평균 데이터율을 획득하기 위한 중요한 부분이다.

발명의 상세한 설명

본 발명의 목적은 일반적으로 가변 레이트 광대역 음성 부호화를 위한 개선된 신호 분류 및 레이트 선택 방법을 제공하는 것이다. 특히, CDMA 시스템에 적합한 가변 레이트 다중모드 광대역 음성 부호화를 위한 개선된 신호 분류 및 레이트 선택 방법을 제공하는 것이다.

소스 제어되는 VBR 음성 부호화의 사용은 많은 통신 시스템, 특히 CDMA 기술을 이용하는 무선 시스템의 용량을 상당히 개선한다. 소스 제어되는 VBR 부호화에 있어서, 코덱은 몇몇 비트율로 동작할 수 있고, 레이트 선택 모듈은 음성 프레임의 성질(예를 들어, 유성음, 무성음, 경과음(transient), 배경 잡음)에 근거하여 각 음성 프레임을 부호화하는데 사용되는 비트율을 결정하는데 사용된다. 주어진 평균 비트율에서 최선의 음성 품질을 획득하는 것이 목표이다. 코덱은 코덱 성능이 증가된 ADR에서 개선되는 상이한 모드에서 상이한 ADR을 획득하도록 레이트 선택 모듈을 조정하여 상이한 모드로 동작할 수 있다. 몇몇 시스템에 있어서, 동작 모드는 채널 상태에 의존하여 시스템에 의해 정해진다. 이것은 코덱이 음성 품질 및 시스템 용량간의 트레이드 오프의 메커니즘을 갖도록 허용한다.

신호 분류 알고리즘은 입력 음성 신호를 분석하고 각 음성 프레임을 한 세트의 소정의 클래스들(예를 들어, 배경 잡음, 유성음, 무성음, 혼합 음성, 경과음 등) 중의 하나로 분류한다. 레이트 선택 알고리즘은 요망되는 평균 데이터율 및 음성 프레임의 클래스에 근거하여 사용될 부호화 모델 및 비트율을 결정한다.

다중 모드 VBR 부호화에 있어서, 상이한 평균 데이터율에 대응하는 상이한 동작 모드가 각 비트율의 이용 비율을 정의함으로써 획득된다. 따라서, 레이트 선택 알고리즘은 음성 프레임의 성질(분류 정보) 및 요구되는 평균 데이터율에 근거하여 어떤 음성 프레임에 대해 사용될 비트율을 결정한다.

몇몇 실시예에 있어서, 3가지 동작 모드가 고려된다. [7]에서 기술되는 바와 같은 프리미엄(Premium), 표준(Standard) 및 절약(Economy) 모드. 프리미엄 모드는 가장 높은 ADR을 이용하여 최고의 달성가능한 품질을 보장한다. 절약 모드는 여전히 고품질 광대역 음성을 허용하는 가장 낮은 ADR을 이용함으로써 시스템 용량을 최대화한다. 표준 모드는 시스템 용량 및 음성 품질을 절충하고 프리미엄 모드의 ADR 및 절약 모드의 ADR 사이의 ADR을 이용한다.

CDMA-1 및 CDMA-2000 시스템에서 동작하도록 제공되는 다중 모드 가변 비트율 광대역 코덱은 본 명세서에서 VMR-WB 코덱으로서 지칭될 것이다.

보다 상세하게는, 본 발명의 제1 태양에 따라, 사운드를 디지털로 부호화하는 방법에 있어서,

- i) 상기 사운드의 샘플링된 버전으로부터 신호 프레임을 제공하는 단계;
- ii) 상기 신호 프레임이 액티브(active) 음성 프레임인지 인액티브(inactive) 음성 프레임인지를 결정하는 단계;
- iii) 상기 신호 프레임이 인액티브 음성 프레임인 경우, 배경 잡음 저 비트율 부호화 알고리즘을 이용하여 상기 신호 프레임을 부호화하는 단계;
- iv) 상기 신호 프레임이 액티브 음성 프레임인 경우, 상기 액티브 음성 프레임이 무성음(unvoiced) 프레임인지 아닌지를 결정하는 단계;
- v) 상기 신호 프레임이 무성음 프레임인 경우, 무성음 신호 부호화 알고리즘을 이용하여 상기 신호 프레임을 부호화하는 단계;
- vi) 상기 신호 프레임이 무성음 프레임이 아닌 경우, 상기 신호 프레임이 안정된(stable) 유성음(voiced) 프레임인지 아닌지를 결정하는 단계;
- vii) 상기 신호 프레임이 안정된 유성음 프레임인 경우, 안정된 유성음 신호 부호화 알고리즘을 이용하여 상기 신호 프레임을 부호화하는 단계; 및
- viii) 상기 신호 프레임이 무성음 프레임이 아니고 상기 신호 프레임이 안정된 유성음 프레임이 아닌 경우, 일반 신호 부호화 알고리즘을 이용하여 상기 신호 프레임을 부호화하는 단계를 포함하는 것을 특징으로 하는 방법이 제공된다.

본 발명의 제2 태양에 따라, 또한 사운드를 디지털로 부호화하는 방법에 있어서,

- i) 상기 사운드의 샘플링된 버전으로부터 신호 프레임을 제공하는 단계;
- ii) 상기 신호 프레임이 액티브 음성 프레임인지 인액티브 음성 프레임인지를 결정하는 단계;

iii) 상기 신호 프레임이 인액티브 음성 프레임인 경우, 배경 잡음 저 비트율 부호화 알고리즘을 이용하여 상기 신호 프레임을 부호화하는 단계;

iv) 상기 신호 프레임이 액티브 음성 프레임인 경우, 상기 액티브 음성 프레임이 무성음(unvoiced) 프레임인지 아닌지를 결정하는 단계;

v) 상기 신호 프레임이 무성음 프레임인 경우, 무성음 신호 부호화 알고리즘을 이용하여 상기 신호 프레임을 부호화하는 단계; 및

vi) 상기 신호 프레임이 무성음 프레임이 아닌 경우, 일반 음성 부호화 알고리즘을 이용하여 상기 신호 프레임을 부호화하는 단계를 포함하는 것을 특징으로 하는 방법이 제공된다.

본 발명의 제3 태양에 따라, 무성음 신호를 분류하는 방법에 있어서,

다음 매개변수들

a) 유성음화 값(\overline{x});

b) 스펙트럼 틸트 값(e_t);

c) 상기 신호 프레임내의 에너지 변동(dE); 및

상기 신호 프레임의 상대 에너지(E_{rel})

중 적어도 3개가 무성음 프레임을 분류하는데 사용되는 것을 특징으로 하는 방법이 제공된다.

본 발명에 따른 방법들은 VBR 코덱이 IP-기반 시스템뿐 아니라 코드 분할 다중 액세스(CDMA) 기술에 근거한 무선 시스템내에서 효율적으로 동작할 수 있게 한다.

최종적으로, 본 발명의 제4 태양에 따라, 사운드 신호를 부호화하는 장치에 있어서,

상기 사운드 신호를 나타내는 디지털화된 사운드 신호를 수신하는 음성 부호기로서, 상기 디지털화된 사운드 신호는 적어도 하나의 신호 프레임을 포함하는 음성 부호기를 포함하고,

상기 음성 부호기는

액티브 및 인액티브 음성 프레임들을 구별하는 제1 레벨 분류기;

인액티브 음성 프레임들을 부호화하는 컴포트 잡음 생성기(comfort noise generator);

유성음 및 무성음 프레임들을 구별하는 제2 레벨 분류기;

무성음 음성 부호기;

안정 및 불안정 유성음 프레임들을 구별하는 제3 레벨 분류기;

유성음 음성 최적화 부호기; 및

일반 음성 부호기를 포함하고,

상기 음성 부호기는 부호화 매개변수들의 이진 표현을 출력하도록 이루어진 것을 특징으로 하는 장치가 제공된다.

본 발명의 상기 및 다른 목적들, 장점들 및 특징들이 첨부한 도면들을 참조하여 단지 예로서 제공되는 본 발명의 예시적인 실시예의 비제한적인 설명을 읽는 경우 더 명백하게 될 것이다.

실시예

이제 첨부된 도면들 중 도 1을 참조하면, 본 발명의 제1 태양의 예시적인 실시예에 따라 음성 부호화 및 복호화의 이용을 나타내는 음성 통신 시스템(10)이 도시된다. 음성 통신 시스템(10)은 통신 채널(12)을 통한 음성 신호(speech signal)의 전송 및 재생을 지원한다. 통신 채널(12)은 예를 들어, 와이어(wire), 광학 또는 섬유 링크, 또는 무선 주파수 링크를 포함할 수 있다. 통신 채널(12)은 또한 상이한 통신 매체의 조합, 예를 들어 부분적으로 섬유 링크 및 부분적으로 무선 주파수 링크일 수 있다. 무선 주파수 링크는 셀룰러 전화에서 발견될 수 있는 바와 같은 공유 대역폭 자원을 필요로 하는 다중 동시 음성 통신을 지원할 수 있다. 대안으로, 통신 채널은 나중 재생을 위해 부호화된 음성 신호를 기록하고 저장하는 통신 시스템의 단일 장치 구현에서 저장 장치(미도시)에 의해 대체될 수 있다.

통신 시스템(10)은 통신 채널(12)의 송신기 측에 마이크로폰(14), 아날로그 디지털 변환기(16), 음성 부호기(18), 및 채널 부호기(20)를 포함하는 부호기 장치, 및 수신기 측에 채널 복호기(22), 음성 복호기(24), 디지털 아날로그 변환기(26) 및 스피커(28)를 포함한다.

마이크로폰(14)은 아날로그 음성 신호를 생성한다. 상기 아날로그 음성 신호는 아날로그 디지털(A/D; analog-to-digital) 변환기(16)에서 디지털 형태로 변환된다. 음성 부호기(18)는 디지털화된 음성 신호를 부호화하여 2진 형태로 부호화된 한 세트의 매개변수들을 생성하여 채널 부호기(20)에 전달한다. 옵션의 채널 부호기(20)는 부호화 매개변수들의 2진 표현에 리던던시(redundancy)를 추가한 후 통신 채널(12)을 통해 전송한다. 또한, 패킷 네트워크 애플리케이션과 같은 몇몇 애플리케이션에 있어서, 부호화된 프레임들은 전송 전에 패킷화된다.

수신기 측에 있어서, 채널 복호기(22)는 수신된 비트스트림에서의 리던던시 정보를 이용하여 전송시에 야기된 채널 에러를 검출하고 정정한다. 음성 복호기(24)는 채널 복호기(22)로부터 수신된 비트스트림을 한 세트의 부호화 매개변수들로 변환하여 합성 음성 신호를 생성한다. 음성 복호기(24)에서 재구성된 합성 음성 신호는 디지털 아날로그(D/A; digital-to-analog) 변환기(26)에서 아날로그 형태로 변환되고 스피커 유닛(28)에서 재생된다.

마이크로폰(14) 및/또는 A/D 변환기(16)는 몇몇 실시예들에서 음성 부호기(18)를 위한 다른 음성 소스로 대체될 수 있다.

부호기(20) 및 복호기(22)는 후술되는 바와 같이 본 발명에 따라 음성 신호를 부호화하는 방법을 구현하도록 구성된다.

신호 분류

이제 도 2를 참조하면, 본 발명의 제1 태양의 제1 예시적인 실시예에 따라 음성 신호를 디지털로 부호화하는 방법(100)이 도시된다. 상기 방법(100)은 본 발명의 제2 태양의 예시적인 실시예에 따른 음성 신호 분류 방법을 포함한다. 음성 신호(speech signal)라는 표현은 음성 내용(음악 중의 음, 배경 음악을 갖는 음, 특별한 사운드 효과를 갖는 음, 등)을 갖는 오디오와 같은 음성 부분을 포함할 수 있는 어떤 멀티미디어 신호뿐 아니라 음성 신호(voice signal)를 나타낸다는 것을 유념한다.

도 2에 도시된 바와 같이, 신호 분류는 3 단계(102, 106 및 110)에서 수행되고, 그들 각각은 특정 신호 클래스를 식별한다. 우선, 단계 102에서, 음성 활동 검출기(VAD; voice activity detector)(미도시)의 형태를 갖는 제1 레벨 분류기가 액티브(active) 및 인액티브(inactive) 음성 프레임들을 구별한다. 인액티브 음성 프레임이 검출되는 경우, 부호화 방법(100)은 예를 들어 콤포트 잡음 생성(CNG; comfort noise generation)을 이용하여 현재 프레임의 부호화를 수행한다(단계 104). 액티브 음성 프레임이 단계 102에서 검출되는 경우, 상기 프레임은 무성음 프레임을 식별하도록 구성된 제2 레벨 분류기(미도시)에 제공된다. 단계 106에서, 상기 분류기가 상기 프레임을 무성음 음성 신호로서 분류하는 경우, 부호화 방법(100)은 단계 108에서 끝난다. 단계 108에서, 상기 프레임은 무성음 신호에 최적화된 부호화 기법을 이용하여 부호화된다. 그렇지 않은 경우, 음성 프레임은 "안정된 유성음(stable voiced)" 분류 모듈(미도시)의 형태를 갖는 제3 레벨 분류기(미도시)에 전달된다(단계 110). 현재 프레임이 안정된 유성음 프레임으로 분류되는 경우, 상기 프레임은 안정된 유성음 신호에 최적화된 부호화 기법을 이용하여 부호화된다(단계 112). 그렇지 않은 경우, 프레임은 유성음 온셋(onset) 또는 급속히 전개되는 유성음 음성 신호 부분과 같은 비정적(non-stationary) 음성 세그먼트를 포함할 것이고, 상기 프레임은 중

은 주관적인 품질을 유지하도록 허용하는 고 비트율을 갖는 일반 목적 음성 부호기를 이용하여 부호화된다(단계 114). 프레임의 상대 에너지가 어떤 임계값보다 더 낮은 경우, 이러한 프레임들은 추가로 평균 데이터율을 감소시키기 위하여 일반적인 저 레이트 부호화 유형을 이용하여 부호화될 수 있다는 것을 유념한다.

분류기들 및 부호기들은 전자 회로로부터 칩 프로세서까지 많은 형태를 가질 수 있다.

이하, 상이한 유형의 음성 신호의 분류가 더 상세하게 설명될 것이고, 무성음 및 유성음 음성의 분류 방법이 개시될 것이다.

인액티브 음성 프레임의 식별(VAD)

인액티브 음성 프레임은 단계 102에서 음성 활동 검출기(VAD; Voice Activity Detector)를 이용하여 식별된다. VAD 구조는 해당 기술 분야의 지식을 가진 사람에게 공지되어 있으므로 본 명세서에서 더 상세하게 기술되지 않을 것이다. VAD의 예는 [5]에 기술된다.

무성음 액티브 음성 프레임의 식별

음성 신호의 무성음 부분은 주기성이 없는 것을 특징으로 하고, 에너지 및 스펙트럼이 급격히 변화하는 불안정 프레임(unstable frames), 및 상기 특징이 비교적 안정되어 있는 안정된 프레임(stable frames)으로 분리될 수 있다.

단계 106에서, 무성음 프레임들은 다음 매개변수들 중의 적어도 3개를 이용하여 식별된다.

- 평균 정규화된 상관으로서 계산될 수 있는 유성음화 값(voicing measure)($\overline{r_x}$);
- 스펙트럼 틸트 값(spectral tilt measure)(e_t);
- 프레임내의 프레임 에너지 변동 및 프레임 안정성을 액세스하는데 사용되는 신호 에너지 비(dB); 및
- 프레임의 상대 에너지.

유성음화 값(voicing measure)

도 3은 본 발명의 제3 태양의 예시적인 실시예에 따라 무성음 프레임을 식별하는 방법(200)을 나타낸다.

유성음화 값을 결정하는데 사용되는 정규화된 상관은 개방 루프 피치 검색 모듈(open-loop pitch search module)(214)의 일부로서 계산된다. 도 3의 예시적인 실시예에 있어서, 20ms 프레임들이 사용된다. 개방 루프 피치 검색 모듈은 일반적으로 매 10 ms마다(프레임당 2번) 개방 루프 피치 추정값(p)을 출력한다. 방법(200)에 있어서, 상기 개방 루프 피치 검색 모듈은 또한 정규화된 상관 값(r_x)을 출력하는데 사용된다. 상기 정규화된 상관은 가중된(weighted) 음성 및 개방 루프 피치 지연으로 과거 가중된 음성에 대해 계산된다. 가중된 음성 신호($s_w(n)$)는 인식 가중치 필터(perceptual weighting filter)(212)에서 계산된다. 예시적인 실시예에 있어서, 광대역 신호에 적합한 고정 분모(denominator)를 갖는 인식 가중치 필터(212)가 사용된다. 다음 수학식은 인식 가중치 필터(212)를 위한 전달 함수의 예를 제공한다.

$$W(z) = A(z/\gamma_1) / (1 - \gamma_2 z^{-1}) \quad \text{여기서} \quad 0 < \gamma_2 < \gamma_1 \leq 1$$

A(z)는 다음 수학식에 의해 제공되는 모듈(218)에서 계산된 선형 예측(LP; linear prediction) 필터의 전달 함수이다.

$$A(z) = 1 + \sum_{i=1}^p a_i z^{-i}$$

유성음화 값은 수학적 식 1과 같이 정의되는 평균 상관($\overline{r_x}$)에 의해 주어진다.

수학적 식 1

$$\overline{r_x} = \frac{1}{3} (r_x(0) + r_x(1) + r_x(2))$$

$r_x(0)$ 는 현재 프레임의 제1 하프의 정규화된 상관(normalized correlation)이고, $r_x(1)$ 는 현재 프레임의 제2 하프의 정규화된 상관이며, $r_x(2)$ 는 (다음 프레임의 시작) 미리 보기(look-ahead)의 정규화된 상관이다.

배경 잡음의 존재를 고려하기 위하여 잡음 정정 인자(r_e)가 수학적 식 1의 정규화된 상관에 추가될 수 있다. 배경 잡음의 존재로, 평균 정규화된 상관이 감소된다. 하지만, 신호 분류를 위하여, 상기 감소가 유성음-무성음 결정에 영향을 미치지 않아야 한다. 따라서 이것은 r_e 의 추가에 의해 보상된다. 좋은 잡음 감소 알고리즘이 사용되는 경우 r_e 는 실제로 제로가 되는 것을 유념해야 한다. 방법(200)에 있어서, 13ms의 미리보기가 사용된다. 정규화된 상관($r_x(k)$)이 수학적 식 2와 같이 계산된다.

수학적 식 2

$$r_x(k) = \frac{r_{xy}}{\sqrt{r_{xx} \cdot r_{yy}}}$$

여기서,

$$r_{xy} = \sum_{i=0}^{L_k-1} x(t_k+i)x(t_k+i-p_k)$$

$$r_{xx} = \sum_{i=0}^{L_k-1} x^2(t_k+i)$$

$$r_{yy} = \sum_{i=0}^{L_k-1} x^2(t_k+i-p_k)$$

방법(200)에 있어서, 상관의 계산은 다음과 같다. 상관($r_x(k)$)은 가중된 음성 신호($s_w(n)$)에 대해 계산된다. 순간(t_k)은 현재 하프 프레임 시작(half-frame beginning)에 관련되고 12800 Hz 샘플링 레이트로 $k = 0, 1$ 및 2 에 대해 각각 0, 128 및 256과 같다. 값($p_k = T_{OL}$)은 하프 프레임에 대해 선택된 개방 루프 피치 추정값이다. 자동 상관 계산의 길이(L_k)는 피치 기간에 의존한다. 제1 실시예에 있어서, L_k 의 값은 (12.8 kHz 샘플링 레이트에 대해) 다음과 같이 요약된다.

$p_k \leq 62$ 샘플들에 대해 $L_k = 80$ 샘플들

$62 < p_k \leq 122$ 샘플들에 대해 $L_k = 124$ 샘플들

$p_k > 122$ 샘플들에 대해 $L_k = 230$ 샘플들

상기 길이는 상관된 벡터 길이가 적어도 하나의 피치 기간을 포함한다는 것을 보장하고, 강인한 개방 루프 피치 검출에 도움이 된다. 긴 피치 기간($p_1 > 122$ 샘플들)에 있어서, $r_x(1)$ 및 $r_x(2)$ 는 동일하다. 즉, 미리 보기에 대한 분석이 더 이상 필요하지 않을 만큼 상관된 벡터들이 충분히 길기 때문에 단 하나의 상관만이 계산된다.

대안으로, 가중된 음성 신호는 개방 루프 피치 검색을 간략하게 하기 위하여 2로 데시메이션(decimation)될 수 있다. 가중된 음성 신호는 데시메이션 이전에 저역 통과 필터링될 수 있다. 이 경우에 있어서, L_k 의 값은 다음에 의해 주어진다.

$p_k \leq 31$ 샘플들에 대해 $L_k = 40$ 샘플들

$31 < p_k \leq 61$ 샘플들에 대해 $L_k = 62$ 샘플들

$p_k > 61$ 샘플들에 대해 $L_k = 115$ 샘플들

상관을 계산하기 위해 다른 방법들이 사용될 수 있다. 예를 들어, 단 하나의 정규화된 상관 값이 몇몇 정규화된 상관을 평균하는 대신에 전체 프레임에 대해 계산될 수 있다. 또한, 잔류(residual) 신호, 음성 신호, 또는 저역 통과 필터링된 잔류, 음성, 또는 가중된 음성 신호와 같은 가중된 음성 이외의 다른 신호들에 대해 상관이 계산될 수 있다.

스펙트럼 틸트(Spectral tilt)

스펙트럼 틸트 매개변수는 에너지의 주파수 분포에 대한 정보를 포함한다. 방법(200)에 있어서, 스펙트럼 틸트는 저주파에 집중된 에너지 및 고주파에 집중된 에너지 간의 비로서 주파수 영역에서 추정된다. 하지만, 상기 스펙트럼 틸트는 또한 음성 신호의 2개의 제1 자동 상관 계수들간의 비와 같이 상이한 방식으로 추정될 수 있다.

방법(200)에 있어서, 도 10의 모듈(210)에서 스펙트럼 분석을 수행하기 위하여 이산 푸리에 변환(Fourier Transform)이 사용된다. 주파수 분석 및 틸트 계산은 프레임당 2번 수행된다. 256 포인트 고속 푸리에 변환(FFT; Fast Fourier Transform)이 50 퍼센트 중첩되어 사용된다. 전체 미리 보기가 이용되도록 분석 윈도우(analysis windows)가 위치된다. 제1 윈도우의 시작은 현재 프레임의 시작 이후에 24 샘플들에 위치한다. 제2 윈도우는 더 나아가 128 샘플들에 위치한다. 주파수 분석을 위해 입력 신호를 가중하기 위해 상이한 윈도우들이 사용될 수 있다. (사인(sine) 윈도우와 동등한) 해밍(Hamming) 윈도우의 제공근이 사용된다. 이 윈도우는 특히 중첩-추가(overlap-add) 방법에 매우 적합하다. 따라서 이 특정 스펙트럼 분석은 스펙트럼 삭감(subtraction) 및 중첩-추가 분석/합성에 근거한 옵션의 잡음 억제 알고리즘에서 사용될 수 있다. 잡음 억제 알고리즘은 해당 기술에 공지되어 있기 때문에, 본 명세서에서 더 상세하게 기술하지 않을 것이다.

고주파 및 저주파에서의 에너지는 인식 임계 대역(perceptual critical bands)에 따라 계산된다[6].

임계 대역 = {100.0, 200.0, 300.0, 400.0, 510.0, 630.0, 770.0, 920.0, 1080.0, 1270.0, 1480.0, 1720.0, 2000.0, 2320.0, 2700.0, 3150.0, 3700.0, 4400.0, 5300.0, 6350.0} Hz

고주파에서의 에너지는 최종 2개의 임계 대역의 에너지 평균으로서 계산된다.

$$\overline{E_k} = 0.5(E_{CB}(18) + E_{CB}(19))$$

$E_{CB}(i)$ 는 다음과 같이 계산되는 임계 대역에 대한 평균 에너지이다.

$$E_{CB}(i) = \frac{1}{N_{CB}(i)} \sum_{k=0}^{N_{CB}(i)-1} (X_R^2(k+j_i) + X_I^2(k+j_i)), \quad i=0, \dots, 19$$

$N_{CB}(i)$ 는 i 번째 대역에서의 주파수 빈(bin)들의 수이다. $X_R(k)$ 및 $X_I(k)$ 는 각각 k 번째 주파수 빈의 실수부 및 허수부이다. j_i 는 i 번째 임계 대역에서의 제1 빈의 인덱스이다.

저주파에서의 에너지는 최초 10개의 임계 대역에서의 에너지의 평균으로서 계산된다. 중간 임계 대역은 저주파에서의 높은 에너지 집중을 갖는 프레임(일반적으로 유성음) 및 고주파에서의 높은 에너지 집중을 갖는 프레임(일반적으로 무성음) 간의 구별을 개선하기 위하여 계산에서 제외되었다. 그 사이에서는, 에너지 내용이 어떤 클래스에 대해 특징을 가지지 않고 결정에 혼동을 증가시킨다.

저주파에서의 에너지는 긴 피치 기간 및 짧은 피치 기간에 대해 상이하게 계산된다. 유성음 여성 음성 세그먼트에 있어서, 스펙트럼의 하모닉(harmonic) 구조가 유성음-무성음 구별을 증가시키는데 이용된다. 따라서, 짧은 피치 기간에 대해, E_l 은 빈-와이즈(bin-wise)로 계산되고 음성 하모닉에 충분히 근접한 주파수 빈만이 합계에 고려된다.

$$\overline{E_l} = \frac{1}{cnt} \sum_{k=0}^{24} E_{BIN}(k) w_h(k)$$

$E_{BIN}(k)$ 은 최초 25 주파수 빈에서의 빈 에너지이다(DC 성분은 고려되지 않음). 상기 25빈은 최초 10개의 임계 대역에 대응한다는 점을 유념한다. 상기 합계에 있어서, 피치 하모닉에 근접한 빈에 관련된 항만이 고려되고, 따라서 빈 및 가장 근접한 하모닉간의 거리가 어떤 주파수 임계값(50 Hz) 보다 더 크지 않은 경우 $w_h(k)$ 는 1로 세팅되고, 그 외의 경우는 0으로 세팅된다. 카운터(cnt)는 합계에서 제로가 아닌 항의 수이다. 가장 근접한 하모닉에 50 Hz보다 더 가까운 빈들만이 고려된다. 따라서, 구조가 저주파에서 하모닉인 경우, 고-에너지 항만이 합에 포함될 것이다. 다른 한편, 상기 구조가 하모닉이 아닌 경우, 항의 선택은 랜덤할 것이고 그 합은 더 작을 것이다. 따라서 저주파에서의 고 에너지 내용을 갖는 무성음조차 검출될 수 있다. 이러한 프로세싱은 주파수 해상도가 충분하지 않기 때문에 더 긴 피치 기간에 대해 수행될 수 없다. 128보다 더 큰 피치 값에 대해 또는 이전(priori) 무성음 사운드에 대해 저주파 에너지가 다음과 같이 임계 대역에 대해 계산된다.

$$\overline{E_l} = \frac{1}{10} \sum_{k=0}^9 E_{CB}(k)$$

이전 무성음 사운드는 $r_x(0) + r_x(1) + r_e < 0.6$ 인 경우에 결정된다. 값(r_e)은 상술된 바와 같이 정규화된 상관에 추가된 보정값이다.

결과적인 저주파 및 고주파 에너지는 상기 계산된 값들($\overline{E_l}$ 및 $\overline{E_h}$)로부터 추정된 잡음 에너지를 감산함으로써 획득된다. 즉,

$$E_h = \overline{E_h} - N_h$$

$$E_l = \overline{E_l} - N_l$$

N_h 및 N_l 은 각각 최종 2개의 임계 대역 및 최초 10개의 임계 대역에서의 평균 잡음 에너지이다. 추정된 잡음 에너지는 배경 잡음의 존재를 고려하기 위해 틸트 계산에 추가되었다.

최종적으로, 스펙트럼 틸트는 다음에 의해 주어진다.

$$e_{tilt}(i) = \frac{E_l}{E_h}$$

스펙트럼 틸트 계산은 프레임에 대한 스펙트럼 분석에 대응하는 $e_{\text{tilt}}(0)$ 및 $e_{\text{tilt}}(1)$ 를 획득하기 위하여 프레임당 2번 수행된다. 무성음 프레임 분류에 사용되는 평균 스펙트럼 틸트는 다음에 의해 주어진다.

$$e_t = \frac{1}{3} (e_{\text{old}} + e_{\text{tilt}}(0) + e_{\text{tilt}}(1))$$

e_{old} 는 이전 프레임의 제2 스펙트럼 분석으로부터의 틸트이다.

에너지 변동(dE)

에너지 변동(dE)은 잡음 제거된 음성 신호($s(n)$)에 대해 계산된다. $n=0$ 은 현재 프레임 시작에 대응한다. 신호 에너지는 32 샘플 길이의 단기 세그먼트에 근거하여 서브프레임당 2번, 즉 프레임당 8번 계산된다. 또한, 이전 프레임으로부터 최종 32 샘플들 및 다음 프레임으로부터 최초 32 샘플들의 단기간 에너지가 또한 계산된다.

단기 최대 에너지는 다음과 같이 계산된다.

$$E_{st}^{(1)}(j) = \max_{i=0}^{31} (s^2(i+32j)), \quad j=-1, \dots, 8$$

$j=-1$ 및 $j=8$ 은 이전 프레임의 끝 및 다음 프레임의 시작에 대응한다. 다른 한 세트의 9 최대 에너지는 음성 인덱스를 16 샘플만큼 시프트함으로써 계산된다.

$$E_{st}^{(2)}(j) = \max_{i=0}^{31} (s^2(i+32j-16)), \quad j=0, \dots, 8$$

연속 단기간 세그먼트들간의 최대 에너지 변동(dE)은 다음의 최대로서 계산된다.

$$E_{st}^{(1)}(0)/E_{st}^{(1)}(-1) \quad E_{st}^{(1)}(0) > E_{st}^{(1)}(-1) \text{인 경우}$$

$$E_{st}^{(1)}(7)/E_{st}^{(1)}(8) \quad E_{st}^{(1)}(7) > E_{st}^{(1)}(8) \text{인 경우}$$

$$\frac{\max(E_{st}^{(1)}(j), E_{st}^{(1)}(j-1))}{\min(E_{st}^{(1)}(j), E_{st}^{(1)}(j-1))} \quad j=1 \text{ 내지 } 7 \text{에 대해}$$

$$\frac{\max(E_{st}^{(2)}(j), E_{st}^{(2)}(j-1))}{\min(E_{st}^{(2)}(j), E_{st}^{(2)}(j-1))} \quad j=1 \text{ 내지 } 8 \text{에 대해}$$

대안으로, 프레임내의 에너지 변동을 계산하기 위하여 다른 방법들이 사용될 수 있다.

상대 에너지(E_{rel})

프레임의 상대 에너지는 프레임 에너지(dB) 및 장기간 평균 에너지간의 차에 의해 주어진다. 프레임 에너지는 다음과 같이 계산된다.

$$E_t = 10 \log \left(\sum_{i=0}^{19} E_{CB}(i) \right), \text{ dB}$$

$E_{CB}(i)$ 는 상술된 바와 같이 임계 대역에 대한 평균 에너지이다. 장기간(long-term) 평균 프레임 에너지는 다음에 의해 주어진다.

$$\overline{E_f} = 0.99 \overline{E_f} + 0.01 E_t$$

초기값은 $\overline{E_f} = 45 \text{ dB}$

따라서 상대 프레임 에너지는 다음에 의해 주어진다.

$$E_{rel} = E_t - \overline{E_f}$$

상대 프레임 에너지는 배경 잡음 프레임 또는 무성음 프레임으로서 분류되지 않은 낮은 에너지 프레임을 식별하는데 사용된다. 상기 프레임들은 ADR을 감소시키기 위하여 일반 HR 부호기를 이용하여 부호화될 수 있다.

무성음 음성 분류

무성음 음성 프레임(unvoiced speech frame)의 분류는 상술된 매개변수들, 즉 유성음화 값($\overline{r^x}$), 스펙트럼 틸트(e_t), 프레임내의 에너지 변동(dE), 및 상대 프레임 에너지(E_{rel})에 근거한다. 분류의 결정은 상기 매개변수들 중 적어도 3개에 근거하여 수행된다. 결정 임계값은 동작 모드(요구되는 평균 데이터율)에 근거하여 세팅된다. 기본적으로 더 낮은 요망되는 데이터율을 갖는 동작 모드에 대해, 임계값은 더 많은 무성음 분류를 선호하도록 세팅된다(하프 레이트 또는 1/4 레이트 부호화가 프레임을 부호화하는데 사용되기 때문에). 무성음 프레임들은 일반적으로 무성음 HR 부호기를 이용하여 부호화된다. 하지만, 절약 모드의 경우, 추가 어떤 조건이 충족되는 경우 ADR을 더 감소시키기 위해 무성음 QR이 또한 사용될 수 있다.

프리미엄 모드에 있어서, 프레임은 다음 조건이 충족되는 경우 무성음 HR로서 부호화된다.

$$(\overline{r^x} < th_1) \text{이고 } (e_t < th_2) \text{이며 } (dE < th_3)$$

이때, $th_1=0.5$ 이고, $th_2=1$ 이며,

$$th_3 = \begin{cases} -4 & \overline{E_f} > 34 \\ 0 & 21 < \overline{E_f} \leq 34 \\ 4 & \text{그외} \end{cases} \text{이다.}$$

음성 활동 결정에 있어서, 결정 행오버(decision hangover)가 사용된다. 따라서, 액티브 음성 기간 이후, 알고리즘이 프레임이 인액티브 음성 프레임인 것으로 결정하는 경우, 로컬 VAD는 제로로 세팅되지만 실제 VAD 플래그는 어떤 수의 프레임이 경과(행오버 기간)된 이후에만 제로로 세팅된다. 이것으로 음성 오프셋의 클리핑(clipping)을 피할 수 있다. 표준 및 절약 모드 양자에 있어서, 로컬 VAD가 제로인 경우, 프레임은 무성음 프레임으로서 분류된다.

표준 모드에 있어서, 프레임은 로컬 VAD=0인 경우 또는 다음 조건이 충족하는 경우 무성음 HR로서 부호화된다.

$$\overline{r_x} < th_4 \text{이고 } (e_t < th_5) \text{이며 } ((dE < th_6) \text{ 또는 } (E_{rel} < th_7))$$

이때, $th_4 = 0.695$, $th_5 = 4$, $th_6 = 40$, $th_7 = -14$ 이다.

절약 모드에 있어서, 로컬 VAD=0인 경우 또는 다음 조건이 충족하는 경우 프레임은 무성음 프레임으로서 선언된다.

$$\overline{r_x} < th_8 \text{이고 } (e_t < th_9) \text{이며 } ((dE < th_{10}) \text{ 또는 } (E_{rel} < th_{11}))$$

이때 $th_8 = 0.695$, $th_9 = 4$, $th_{10} = 60$, $th_{11} = -14$ 이다.

절약 모드에 있어서, 무성음 프레임들은 일반적으로 무성음 HR로서 부호화된다. 하지만, 다음의 추가 조건이 또한 충족되는 경우 무성음 프레임들은 또한 무성음 QR을 가지고 부호화될 수 있다. 최종 프레임이 무성음 또는 배경 잡음 프레임인 경우, 및 프레임의 종단에서 에너지가 고주파에 집중되고 미리 보기에서 어떠한 잠재 유성음 온셋이 검출되지 않은 경우 프레임은 무성음 QR로서 부호화된다. 최종 2가지 조건은 다음과 같이 검출된다.

$$(r_x(2) < th_{12}) \text{이고 } (e_{tilt}(1) < th_{13})$$

이때, $th_{12} = 0.73$, $th_{13} = 3$ 이다.

$r_x(2)$ 는 미리 보기(lookahead)에서의 정규화된 상관이며 $e_{tilt}(1)$ 은 미리 보기 및 상기 신호 프레임의 종단을 확장하는 제2 스펙트럼 분석에서의 틸트인 것을 유념한다.

물론 무성음 프레임을 식별하기 위하여 방법(200) 이외의 다른 방법들이 사용될 수 있다.

안정된 유성음 음성 프레임들의 식별

표준 및 절약 모드의 경우, 안정된 유성음 프레임들은 유성음 HR 부호화 유형을 이용하여 부호화될 수 있다.

유성음 HR 부호화 유형은 안정된 유성음 프레임들을 효율적으로 부호화하기 위해 신호 수정(signal modification)을 이용한다.

신호 수정 기법은 신호의 피치(pitch)를 소정의 지연 윤곽선(contour)에 조절한다. 장기간 예측은 과거 여기(excitation) 신호를 상기 지연 윤곽선을 이용하고 이득 매개변수에 의해 스케일링하여 현재 서브프레임에 매핑한다. 지연 윤곽선은 2개의 개방 루프 피치 추정, 이전 프레임에서 획득된 제1 추정 및 현재 프레임에서 획득된 제2 추정 사이에 보간에 의해 직접 획득된다. 보간(interpolation)은 프레임의 모든 시간 순간에 대한 지연 값을 제공한다. 지연 윤곽선이 이용가능한 이후, 현재 부호화되는 서브프레임에서의 피치는 신호의 시간 스케일을 변경하고 워핑(warping)함으로써 인공 윤곽선(artificial contour)을 따르도록 조정된다. 불연속 워핑(discontinuous warping) [1, 4, 5]에 있어서, 신호 세그먼트는 세그먼트 길이를 변경하지 않고 좌측으로 또는 우측으로 시프트된다. 불연속 워핑은 결과적인 중첩되거나 손실(missing)된 신호 부분을 처리하기 위한 절차를 필요로 한다. 상기 단계에 있어서 인공물(artifacts)을 감소시키기 위하여, 시간 스케일에서의 허용되는 변경은 작게 유지된다. 더욱이, 워핑은 일반적으로 결과적인 왜곡을 감소시키기 위하여 가중된 음성 신호 또는 LP 잔류 신호를 이용하여 수행된다. 음성 신호 대신에 상기 신호의 이용은 또한 피치 펄스들 및 피치 펄스들 사이의 저전력 영역

의 검출을 용이하게 하고, 따라서 워핑을 위한 신호 세그먼트의 결정을 촉진시킨다. 실제 수정된 음성 신호는 역필터링에 의해 생성된다. 신호 수정이 현재 서브프레임에 대해 수행된 후, 적응성의 코드북 여기(adaptive codebook excitation)가 소정의 지연 윤곽선을 이용하여 생성된다는 점을 제외하고 종래의 방식으로 부호화가 진행될 수 있다.

예시적인 실시예에 있어서, 신호 수정은 피치 및 프레임에 동시에 수행된다. 즉, 다음의 음성 프레임이 원래 신호에 대한 완전한 시간 정렬에서 시작하도록 현재 프레임의 시간에 하나의 피치 사이클 세그먼트를 적합하게 한다. 피치 사이클 세그먼트는 프레임 경계에 의해 제한된다. 이것은 프레임 경계를 넘어 시간 시프트 변형을 방지하고 부호기 구현을 간단하게 하며 수정된 음성 신호에서의 인공물의 위험을 감소시킨다. 모든 신규 프레임이 원래 신호에 대한 시간 정렬에서 시작하기 때문에 또한 신호 수정 인에이블(enabled) 및 디스에이블(disabled) 부호화 유형간에 가변 비트율 동작을 간단하게 한다.

도 2에 도시된 바와 같이, 프레임이 인액티브 음성 프레임 또는 무성음 프레임으로 분류되지 않은 경우 상기 프레임이 안정된 유성음 프레임인지를 검사한다(단계 110). 안정된 유성음 프레임의 분류는 안정된 유성음 프레임을 부호화하는데 사용되는 신호 수정 절차와 관련하여 폐쇄 루프 접근법(closed-loop approach)을 이용하여 수행된다.

도 4는 본 발명의 제4 태양의 예시적인 실시예에 따라 안정된 유성음 프레임을 식별하는 방법(300)을 나타낸다.

신호 수정에서의 서브 절차는 현재 프레임에서의 장기간 예측의 얻을 수 있는 성능을 양자화하는 표시자를 제공한다. 상기 표시자들 중의 어느 것이 허용된 한계를 벗어나는 경우, 신호 수정 절차는 로직 블록들 중의 하나에 의해 종료된다. 이 경우에 있어서, 원래 신호는 그대로 유지되고 프레임은 안정된 유성음 프레임으로서 분류되지 않는다. 이러한 집적 로직은 저 비트율에서의 부호화 및 신호 수정 이후에 수정된 음성 신호의 품질을 최대화할 수 있다.

단계 302의 피치 펄스 검색 절차는 현재 프레임의 주기성에 대한 몇몇 표시자들을 생성한다. 따라서 그에 따른 로직 블록은 분류 로직의 중요한 구성요소이다. 피치 사이클 길이의 발달이 관찰된다. 상기 로직 블록은 이전 검출된 피치 펄스의 거리에 대해서 뿐 아니라 보간된 개방 루프 피치 추정에 대해 검출된 피치 펄스 위치의 거리를 비교한다. 신호 수정 절차는 개방 루프 피치 추정 또는 이전 피치 사이클 길이로의 차이가 너무 큰 경우 종료된다.

단계 304에서 지연 윤곽선의 선택은 현재 음성 프레임의 주기성 및 피치 사이클의 발달(evolution)에 대한 추가 정보를 제공한다. 조건 $|d_n - d_{n-1}| < 0.2d_n$ 이 충족되는 경우, 신호 수정 절차는 이 블록에서 계속된다. d_n 및 d_{n-1} 은 현재 및 과거 프레임에서의 피치 지연이다. 이것은 본질적으로 현재 프레임을 안정된 유성음 프레임으로서 분류하기 위해 작은 지연 변경만이 허용된다는 것을 의미한다.

신호 수정이 수행된 프레임이 저 비트율로 부호화되는 경우, 피치 사이클 세그먼트의 형태는 신뢰할 만한 신호가 장기간 예측에 의해 모델링되고 따라서 주관적인 품질을 저하시키지 않으면서 저 비트율로 부호화하도록 허용하기 위해 상기 프레임에 대해 유사하게 유지된다. 단계 306의 신호 수정에 있어서, 연속 세그먼트의 유사성은 현재 세그먼트 및 최적 시프트된 타깃 신호간의 정규화된 상관에 의해 양자화될 수 있다. 타깃 신호와의 상관을 최대화하는 피치 사이클 세그먼트의 시프팅(shifting)은 주기성을 향상시키고 신호 수정이 유용한 경우 높은 장기간 예측 이득을 제공한다. 모든 상관 값이 미리 정의된 임계값보다 더 크지 않아야 한다는 것을 요구함으로써 상기 절차의 성공이 보장된다. 이 조건이 모든 세그먼트에 대해 충족되지 않는 경우, 신호 수정 절차는 종료하고 원래의 신호가 그대로 유지된다. 일반적으로, 약간 더 낮은 이득 임계값 범위가 동일한 부호화 성능을 갖는 남성 음성에 대해 허용될 수 있다. 신호 수정을 적용하고 타깃 평균 비트율을 변경하는 부호화 모드의 사용을 조정하기 위하여 이득 임계값들은 VBR 코덱의 상이한 동작 모드에서 변경될 수 있다.

상술된 바와 같이, 방법(100)에 따른 완전한 레이트 선택 로직은 3 단계를 포함하고, 각 단계는 특정 신호 클래스를 식별한다. 단계들 중의 하나는 필수 부분으로서 신호 수정 알고리즘을 포함한다. 우선, VAD는 액티브 및 인액티브 음성 프레임들을 구별한다. 인액티브 음성 프레임이 검출되는 경우, 분류 방법은 프레임이 배경 잡음으로 간주되고 예를 들어 콤포트 잡음 생성기를 이용하여 부호화되는 것으로 종료한다. 액티브 음성 프레임이 검출되는 경우, 상기 프레임에 대해 무성음 프레임을 식별하는 제2 단계가 수행된다. 프레임이 무성음 음성 신호로서 분류되는 경우, 분류 단계가 종료하고, 상기 프레임은 무성음 프레임에 대한 전용 모드를 가지고 부호화된다. 최종 단계로서, 이 서브 섹션에서 상술된 조건이 검증되는 경우 수정을 가능하게 하는 제안된 신호 수정 절차를 통해 음성 프레임이 처리된다. 이 경우에 있어서, 프레임은 안정된 유성음 프레임으로서 분류되고, 원래 신호의 피치는 인공적이고 잘 정의된 지연 윤곽선으로 조정되며, 상기 유형의 프레임에 대해 최적의 특정 모드를 이용하여 프레임이 부호화된다. 그렇지 않은 경우, 상기 프레임은 유성음 온셋 또는 급속히 전개되는 유성음 음성 신호와 같은 비정적 음성 세그먼트를 포함할 것이다. 상기 프레임은 전형적으로 보다 일반적인 부호화

모델을 필요로 한다. 상기 프레임들은 일반적으로 일반 FR 부호화 유형을 이용하여 부호화된다. 하지만, 프레임의 상대 에너지가 어떤 임계값보다 더 낮은 경우, 상기 프레임들은 추가로 ADR을 감소시키기 위하여 일반 HR 부호화 유형을 가지고 부호화될 수 있다.

CDMA 다중 모드 VBR 시스템에 대한 레이트 선택 및 음성 부호화

레이트 세트 II에서 동작할 수 있는 CDMA 다중 모드 VBR 시스템에서 사운드의 디지털 부호화 및 레이트 선택 방법이 이제 본 발명의 예시적인 실시예에 따라 기술될 것이다.

상기 코덱은 몇몇 광대역 음성 서비스를 위한 국제 전기 통신 연합 전기 통신 표준화 섹터(ITU-T; International Telecommunications Union - Telecommunication Standardization Sector)에 의해 그리고 GSM 및 W-CDMA 제3 세대 무선 시스템을 위한 제3 세대 협력 프로젝트(3GPP; third generation partnership project)에 의해 최근 선택된 적응성 다중 레이트 광대역(AMR-WB; adaptive multi-rate wideband) 음성 코덱에 근거한다. AMR-WB 코덱은 9개의 비트율, 즉 6.6, 8.85, 12.65, 14.25, 15.85, 18.25, 19.85, 23.05, 및 23.85 kbit/s로 구성된다. CDMA 시스템용 AMR-WB 기반 소스 제어되는 VBR 코덱은 AMR-WB 코덱을 이용한 다른 시스템 및 CDMA간의 상호 동작을 가능하게 한다. 레이트 세트 II의 13.3kbit/s 풀 레이트(full-rate)에 적합하고 가장 근접한 레이트인 12.65 kbit/s의 AMR-WB 비트율이 (음성 품질을 저하시키는) 트랜스코딩(transcoding)을 필요로 하지 않고 상호 운용성을 가능하게 하는 AMR-WB 및 CDMA 광대역 VBR 코덱간의 공통 레이트로서 사용될 수 있다. 특히 CDMA VBR 광대역 솔루션이 레이트 세트 II 프레임워크에서 효율적인 동작을 할 수 있도록 더 낮은 레이트 부호화 유형이 제공된다. 이때 상기 코덱은 모든 레이트를 이용하는 소수의 CDMA 특정 모드에서 동작할 수 있지만, AMR-WB 코덱을 이용하는 시스템과 상호 운용성을 가능하게 하는 모드를 가질 것이다.

본 발명의 실시예에 따른 부호화 방법들은 표 1에 요약되고 일반적으로 부호화 유형(coding types)으로서 지칭될 것이다.

[표 1] 대응하는 비트율을 갖는 예시적인 실시예에서 사용되는 부호화 유형

부호화 유형	비트율[kbit/s]	비트 / 20ms 프레임
일반 FR	13.3	266
상호 운용 FR	13.3	266
유성음 HR	6.2	124
무성음 HR	6.2	124
상호 운용 HR	6.2	124
일반 HR	6.2	124
무성음 QR	2.7	54
CNG QR	2.7	54
CNG ER	1.0	20

풀 레이트(FR; full-rate) 부호화 유형은 12.65 kbit/s의 AMR-WB 표준 코덱에 근거한다. AMR-WB 코덱의 12.65 kbit/s 레이트의 사용은 AMR-WB 코덱 표준을 이용하는 다른 시스템과 상호 동작할 수 있는 CDMA 시스템을 위해 가변 비트율 코덱의 설계를 가능하게 한다. 프레임당 여분의 13 비트들이 CDMA 레이트 세트 II의 13.3 kbit/s 풀 레이트에 적합하도록 추가된다. 상기 비트들은 삭제된 프레임들의 경우 코덱의 강인함을 향상시키고 본질적으로 일반 FR 및 상호 운용 FR 부호화 유형(상기 비트들이 상호 운용 FR에서 사용되지 않는다)간의 차이를 만드는데 사용된다. FR 부호화 유형은 일반 광대역 음성 신호에 최적화된 대수 코드 여기 선형 예측(ACELP; algebraic code-excited linear prediction) 모델에 근거한다. 상기 유형은 16 kHz의 샘플링 주파수를 갖는 20 ms 음성 프레임에서 동작한다. 추가 프로세싱 이전에, 입력 신호는 12.8 kHz 샘플링 주파수로 다운 샘플링되고 전처리된다. LP 필터 매개변수들은 46 비트를 이용하여 프레임당 한번 부호화된다. 이때 상기 프레임은 4개의 서브프레임으로 분할되고, 적응성의 고정된 코드북 인덱스 및 이득이 서브프레임당 한번 부호화된다. 고정 코드북은 대수 코드북 구조를 이용하여 구성되고 서브 프레임의 64 위치들이 인터리빙된 위치의 4 트랙으로 분할되며 2개의 부호 펄스(signed pulses)가 각 트랙에 위치한다. 트랙당 2개의 펄스들은 9비트를 이용하여 부호화되고 서브 프레임 당 전체 36 비트를 제공한다. AMR-WB 코덱에 대한 보다 상세한 설명은 참고문헌 [1]에서 발견될 수 있다. FR 부호화 유형에 대한 비트 할당은 표 2에서 제공된다.

[표 2] 12.65 kbit/s의 AMR-WB 표준에 근거한 일반 및 상호 운용 풀 레이트 CDMA2000 레이트 세트 II의 비트 할당

	프레임당 비트

매개변수	일반 FR	상호 운용 FR
클래스 정보	-	-
VAD 비트	-	1
LP 매개변수	46	46
피치 지연	30	30
피치 필터링	4	4
이득	28	28
대수 코드북	144	144
FER 보호 비트	14	-
미사용 비트	-	13
합계	266	266

안정된 유성음 프레임의 경우, 하프 레이트 유성음 부호화(Half-Rate Voiced coding)가 이용된다. 하프 레이트 유성음 비트 할당은 표 3에 주어진다. 이 통신 모드에서 부호화되는 프레임들이 매우 주기적인 특징을 가지기 때문에, 실질적으로 더 낮은 비트율이 예를 들어 전이 프레임들(transition frames)에 비해 좋은 주관적인 품질을 유지하는데 충분하다. 20 ms 프레임당 9개의 비트만을 이용하여 지연 정보의 효율적인 부호화를 허용하는 신호 수정이 사용되고 다른 신호 부호화 매개변수들에 대한 상당 부분 비트 예산(bit budget)을 덜어준다. 신호 수정에 있어서, 신호는 프레임당 9비트를 가지고 전송될 수 있는 어떤 피치 윤곽선을 따르도록 강요된다. 장기간 예측의 좋은 성능은 주관적인 음성 품질을 희생하지 않고 고정 코드북 여기에 대해 5-ms 서브프레임당 12 비트만을 이용하도록 허용한다. 고정 코드북은 대수 코드북이고 하나의 펄스를 갖는 2개의 트랙을 포함한다. 각 트랙은 32개의 가능한 위치를 갖는다.

[표 3] CDMA2000 레이트 세트 II에 따른 하프 레이트 일반, 유성음, 무성음의 비트 할당

매개변수	프레임당 비트			
	일반 HR	유성음 HR	무성음 HR	상호 운용 HR
클래스 정보	1	3	2	3
VAD 비트	-	-	-	1
LP 매개변수	36	36	46	46
피치 지연	13	9	-	30
피치 필터링	-	2	-	4
이득	26	26	24	28
대수 코드북	48	48	52	-
FER 보호 비트	-	-	-	-
미사용 비트	-	-	-	12
합계	124	124	124	124

무성음 프레임의 경우, 적응성 코드북(또는 피치 코드북)이 사용되지 않는다. 13-비트 가우스 코드북이 각 서브프레임에 사용되고 코드북 이득은 서브 프레임당 6비트를 가지고 부호화된다. 평균 비트율이 추가로 감소될 필요가 있는 경우 무성음 1/4 레이트가 안정된 무성음 프레임의 경우에 사용될 수 있다는 것을 유념한다.

일반적인 하프 레이트 모드는 낮은 에너지 세그먼트에 대해 사용된다. 이러한 일반 HR 모드는 또한 후술되는 바와 같이 최대 하프 레이트 동작에서 사용될 수 있다. 일반 HR의 비트 할당은 표 3에 표시된다.

예로서, 상이한 HR 부호기들에 대한 분류 정보에 대해, 일반 HR의 경우, 프레임이 일반 HR인지 다른 HR인지를 나타내기 위해 1 비트가 사용된다. 무성음 HR의 경우, 2 비트가 분류를 위해 사용된다. 제1 비트는 프레임이 일반 HR이 아닌 것을 나타내고 제2 비트는 프레임이 무성음 HR이고 유성음 HR이 아니거나 상호 운용 HR(후술되는)인 것을 나타낸다. 유성음 HR인 경우, 3비트가 사용된다. 처음 2개의 비트는 프레임이 일반 또는 무성음 HR이 아닌 것을 나타내고, 제3 비트는 프레임이 무성음 HR인지 상호 운용 HR인지를 나타낸다.

절약 모드에 있어서, 대부분의 무성음 프레임은 무성음 QR 부호기를 이용하여 부호화될 수 있다. 이 경우에 있어서, 가우스 코드북 인덱스는 랜덤하게 생성되고 이득은 서브 프레임당 5비트만을 이용하여 부호화된다. 또한, LP 필터 계수들이 더 낮은 비트율을 가지고 양자화된다. 1 비트는 2가지 1/4 레이트(quarter-rate) 부호화 유형을 구별하는데 사용된다: 무성음 QR 및 CNG QR. 무성음 부호화 유형에 대한 비트 할당은 6에서 제공된다.

상호 운용 HR 부호화 유형은 프레임이 폴 레이트로서 분류된 경우 CDMA 시스템이 특정 프레임에 대한 최대 레이트로서 HR을 지정하는 경우에 대응하도록 허용한다. 상호 운용 HR은 프레임이 폴 레이트 프레임으로서 부호화된 이후에 고정 코드북 인덱스를 드롭(drop)함으로써 폴 레이트 부호기로부터 직접 유도된다(표 4). 복호기 측에서, 고정 코드북 인덱스는 랜덤하게 생성될 수 있고 복호기는 폴 레이트인 것처럼 동작할 것이다. 이러한 설계는 (이동 GSM 시스템 또는 W-CDMA 제3 세대 무선 시스템과 같은) AMR-WB 표준을 이용하는 다른 시스템 및 CDMA 시스템간의 탠덤 프리 동작(tandem free operation) 동안 강요된 하프 레이트 모드의 영향을 최소화하는 장점을 갖는다. 상술된 바와 같이, 상호 운용 FR 부호화 유형 또는 CNG QR은 AMR-WB를 이용한 탠덤-프리 동작(TFO; tandem-free operation)에 사용된다. CDMA2000으로부터 AMR-WB 코덱을 이용하는 시스템으로의 방향을 갖는 링크에 있어서, 다중 서브 계층이 하프 레이트 모드 요청을 나타내는 경우, VMR-WB 코덱은 상호 운용 HR 부호화 유형을 이용할 것이다. 시스템 인터페이스에서, 상호 운용 HR 프레임이 수신되는 경우, 랜덤하게 생성된 대수 코드북 인덱스가 12.65 kbit/s 레이트를 출력하도록 비트 스트림에 추가된다. 수신기 측의 AMR-WB 복호기는 비트 스트림을 원래의 12.65 kbit/s 프레임으로 해석할 것이다. 다른 방향으로, 즉 AMR-WB 코덱을 이용하는 시스템에서 CDMA2000으로의 링크에 있어서, 시스템 인터페이스에서 하프 레이트 요청이 수신되는 경우, 대수 코드북 인덱스가 드롭되고 상호 운용 HR 프레임 유형을 나타내는 모드 비트들이 추가된다. CDMA2000 측의 복호기는 VMR-WB 부호화 솔루션의 일부인 상호 운용 HR 부호화 유형으로서 동작한다. 상호 운용 HR 없이, 강제 하프 레이트 모드는 프레임 삭제(frame erasure)로서 해석될 것이다.

인액티브 음성 프레임의 처리를 위해 컴포트 잡음 생성(CNG; Comfort Noise Generation) 기법이 사용된다. CDMA 시스템내에서 동작하는 경우 인액티브 음성 프레임을 부호화하기 위해 CNG 1/8 레이트(ER; eighth rate) 부호화 유형이 사용된다. AMR-WB 음성 부호화 표준을 갖는 상호 운용이 요구되는 콜(call)에 있어서, 그 비트율이 AMR-WB의 CNG 복호기에 대한 갱신 정보를 전송하는데 필요한 비트율보다 더 낮기 때문에 CNG ER이 항상 사용될 수 있는 것은 아니다[3]. 이 경우에 있어서, CNG QR이 사용된다. 하지만, AMR-WB 코덱은 종종 불연속 전송 모드(DTX; Discontinuous Transmission Mode)에서 동작한다. 불연속 전송 동안, 배경 잡음 정보는 매 프레임마다 갱신되지 않는다. 전형적으로, 8 개의 연속 인액티브 음성 프레임 중에서 하나의 프레임만이 전송된다. 이러한 갱신 프레임은 침묵 기술자(SID; Silence Descriptor)로 지칭된다[4]. DTX 동작은 모든 프레임이 부호화되는 CDMA 시스템에서 사용되지 않는다. 따라서, SID 프레임들만이 CDMA 측의 CNG QR을 가지고 부호화될 필요가 있고 나머지 프레임들은 AMR-WB 대응부에 의해 사용되지 않기 때문에 ADR을 감소시키기 위해 CNG ER을 가지고 부호화될 수 있다. CNG 부호화에 있어서, LP 필터 매개변수들 및 이득만이 프레임당 한번 부호화된다. CNG QR에 대한 비트 할당이 표 4에서 제공되고 CNG ER의 비트 할당은 표 5에서 제공된다.

[표 4] 무성음 QR 및 CNG QR 부호화 유형에 대한 비트 할당

매개변수	무성음 QR	CNG QR
선택 비트	1	1
LP 매개변수	32	28
이득	20	6
미사용 비트	1	19
합계	54	54

[표 5] CNG ER에 대한 비트 할당

매개변수	CNG ER 비트/프레임
LP 매개변수	14
이득	6
미사용	-
합계	20

프리미엄 모드에서의 레이트 선택 및 신호 분류

본 발명의 제2 태양의 제2 예시적인 실시예에 따라 사운드 신호를 디지털로 부호화하는 방법(400)이 도 5에 도시된다. 방법(400)은 이용가능한 비트율이 주어진 최대 합성 음성 품질을 위해 제공되는 프리미엄 모드에서의 방법(100)의 특정 응용인 점을 유념한다(시스템이 특정 프레임에 대한 최대 이용가능한 레이트를 제한하는 경우는 별도의 서브섹션에서 기술 될 것임을 유념한다). 따라서, 대부분의 액티브 음성 프레임들은 폴 레이트로, 즉 13.3 kb/s로 부호화된다.

도 2에 도시된 방법(100)에 유사하게, 음성 활동 검출기(VAD)는 액티브 및 인액티브 음성 프레임을 구별한다(단계 102). VAD 알고리즘은 모든 모드의 동작에서 동일할 수 있다. 인액티브 음성 프레임(배경 잡음 신호)이 검출되는 경우 분류 방법은 중단되고 프레임은 CDMA 레이트 세트 II에 따라 1.0 kbit/s로 CNG ER 부호화 유형을 가지고 부호화된다(단계 402). 액티브 음성 프레임이 검출되는 경우, 프레임은 무성음 프레임을 식별하는 제2 분류기에 제공된다(단계 404). 프리미엄 모드가 최선의 가능한 품질을 목적으로 하기 때문에, 무성음 프레임 식별은 매우 엄격하고 매우 정적인 무성음 프레임만이 선택된다. 무성음 분류 규칙 및 결정 임계값은 상술된 바와 같다. 제2 분류기가 프레임을 무성음 음성 신호로서 분류하는 경우, 분류 방법이 중단되고 프레임은 무성음 신호에 최적인(CDMA 레이트 세트 II에 따라 6.2 kbit/s) 무성음 HR 부호화 유형을 이용하여 부호화된다(단계 408). 다른 모든 프레임들은 12.65 kbit/s의 AMR-WB 표준에 근거하여 일반 FR 부호화 유형을 가지고 처리된다.

표준 모드에서의 레이트 선택 및 신호 분류

본 발명의 제2 태양의 제3 예시적인 실시예에 따라 사운드 신호를 디지털로 부호화하는 방법(500)이 도 6에 도시된다. 방법(500)은 표준 모드에서의 부호화 및 음성 신호의 분류를 허용한다.

단계 102에서, VAD는 액티브 및 인액티브 음성 프레임을 구별한다. 인액티브 음성 프레임이 검출되는 경우, 분류 방법은 중단되고 프레임은 CNG ER 프레임으로서 부호화된다(단계 510). 액티브 음성 프레임이 검출되는 경우, 프레임은 무성음 프레임을 식별하는 제2 레벨 분류기에 제공된다(단계 404). 무성음 분류 규칙 및 결정 임계값은 상술되었다. 제2 레벨 분류기가 프레임을 무성음 음성 신호로서 분류하는 경우, 분류 방법은 중단되고, 프레임은 무성음 HR 부호화 유형을 이용하여 부호화된다(단계 508). 그렇지 않은 경우, 음성 프레임은 "안정된 유성음(stable voiced)" 분류 모듈에 전달된다(단계 502). 유성음 프레임의 식별은 상술된 바와 같은 신호 수정 알고리즘의 고유 특징이다. 프레임이 신호 수정에 적합한 경우, 상기 프레임은 안정된 유성음 프레임으로서 분류되고 안정된 유성음 신호에 최적인 모듈에서(CDMA 레이트 세트 II에 따라 6.2kbit/s) 유성음 HR 부호화 유형을 이용하여 부호화된다(단계 506). 그렇지 않은 경우, 상기 프레임은 유성음 온셋 또는 급속히 전개되는 유성음 음성 신호와 같은 비정적 음성 세그먼트를 포함할 것이다. 상기 프레임은 전형적으로 좋은 주관적인 품질을 유지하기 위해 고 비트율을 필요로 한다. 하지만, 프레임의 에너지가 어떤 임계값보다 더 낮은 경우, 상기 프레임들은 일반 HR 부호화 유형을 가지고 부호화될 수 있다. 따라서, 단계 512에서 제4 레벨 분류기가 낮은 에너지 신호를 검출하는 경우 프레임은 일반 HR을 이용하여 부호화된다(단계 514). 그렇지 않은 경우, 음성 프레임은 일반 FR 프레임으로서 부호화된다(CDMA 레이트 세트 II에 따라 13.3 kbit/s)(단계 504).

절약 모드에서의 레이트 선택 및 신호 분류

본 발명의 제1 태양의 제4 예시적인 실시예에 따라 사운드 신호를 디지털로 부호화하는 방법(600)이 도 7에 도시된다. 4 레벨 분류 방법인 방법(600)은 절약 모드에서의 부호화 및 음성 신호의 분류를 허용한다.

절약 모드는 여전히 고품질 광대역 음성을 생성하는 최대 시스템 용량을 허용한다. 레이트 결정 로직은 또한 무성음 QR 부호화 유형이 사용되고 일반 FR 사용이 감소된다는 점을 제외하고 표준 모드와 유사하다.

우선, 단계 102에서, VAD는 액티브 및 인액티브 음성 프레임을 구별한다. 인액티브 음성 프레임이 검출되는 경우, 분류 방법이 중단되고 프레임은 CNG ER 프레임으로서 부호화된다(단계 402). 액티브 음성 프레임이 검출되는 경우, 프레임은 모든 무성음 프레임을 식별하는 제2 분류기에 제공된다(단계 106). 무성음 분류 규칙 및 결정 임계값은 상술되었다. 제2 분류기가 프레임을 무성음 음성 신호로서 분류하는 경우, 음성 프레임은 제1 제3-레벨 분류기에 전달된다(단계 602). 상기 제3 레벨 분류기는 프레임이 상술된 규칙을 이용하여 유성음-무성음 전이(voiced-unvoiced transition) 중에 있는지 여부를 검사한다. 특히, 상기 제3 레벨 분류기는 최종 프레임이 배경 잡음 프레임의 무성음인지 여부를 검사하고, 프레임의 중단에서 에너지가 고주파에 집중되어 있는지 및 잠재적 유성음 온셋이 미리 보기에서 검출되지 않는지를 검사한다. 상술된 바와 같이, 최종 2개의 조건은 다음으로서 검출된다.

$$(r_x(2) < th_{12}) \text{이고 } (e_{\text{tilt}}(1) < th_{13})$$

이때, $th_{12} = 0.73$, $th_{13} = 3$ 이다. $r_x(2)$ 는 미리 보기에서의 상관이며 $e_{\text{tilt}}(1)$ 은 미리 보기 및 프레임의 종단을 확장하는 제2 스펙트럼 분석에서의 틸트이다.

프레임이 유성음-무성음 전이를 포함하는 경우, 프레임은 무성음 HR 부호화 유형을 가지고 부호화된다(단계 508). 그렇지 않은 경우, 음성 프레임은 무성음 QR 부호화 유형을 가지고 부호화된다(단계 604). 무성음으로서 분류되지 않은 프레임들은 제2 제3-레벨 분류기인 "안정된 유성음" 분류 모듈에 전달된다(단계 110). 유성음 프레임의 식별은 상술된 바와 같은 신호 수정 알고리즘의 고유 특징이다. 프레임이 신호 수정에 적합한 경우, 상기 프레임은 안정된 유성음 프레임으로서 분류되고 단계 506에서 유성음 HR을 가지고 부호화된다. 표준 모드와 유사하게, (무성음 또는 안정된 유성음으로서 분류되지 않은) 나머지 프레임들은 저(low) 에너지 내용에 대해 테스트된다. 저 에너지 신호가 단계 512에서 검출되는 경우, 프레임은 일반 HR을 이용하여 부호화된다(단계 514). 그렇지 않은 경우, 음성 프레임은 일반 FR 프레임으로서 (CDMA 레이트 세트 II에 따라 13.3 kbit/s) 부호화된다(단계 504).

상호 운용 모드(interoperable mode)에서의 레이트 선택 및 신호 분류

본 발명의 제2 태양의 제5 예시적인 실시예에 따라 사운드 신호를 디지털로 부호화하는 방법(700)이 도 8에 도시된다. 방법(700)은 상호 운용 모드에서의 부호화 및 음성 신호의 분류를 허용한다.

상호 운용 모드는 12.65 kbit/s(또는 더 낮은 레이트)의 AMR-WB 표준을 이용하는 다른 시스템 및 CDMA 시스템간의 탠덤 프리 동작을 허용한다. CDMA 시스템에 의해 가해지는 레이트 제한이 없는 경우, 상호 운용 FR 및 컴포트 잡음 생성기만이 사용된다.

우선, 단계 102에서, VAD는 액티브 및 인액티브 음성 프레임을 구별한다. 인액티브 음성 프레임이 검출되는 경우, 상기 프레임이 SID 프레임으로서 부호화되어야 하는지 여부에 대한 결정이 수행된다(단계 702). 상술된 바와 같이, SID 프레임은 DTX 동작 동안 AMR-WB 측의 CNG 매개변수들을 갱신하도록 기능한다[4]. 전형적으로, 8개의 인액티브 음성 프레임만이 침묵 기간 동안 부호화된다. 하지만, 액티브 음성 세그먼트 이후, SID 갱신은 제4 프레임에서 이미 전송되어야 한다(더 상세한 설명을 위해 참고문헌 [4] 참조). ER이 SID 프레임을 부호화하는데 충분하지 않기 때문에, SID 프레임은 CNG QR을 이용하여 부호화된다(단계 704). SID 인액티브 프레임이외의 다른 프레임들은 CNG ER을 이용하여 부호화된다(단계 402). 탠덤 프리 동작(TFO; Tandem Free Operation)에서 CDMA VMR-WB로부터 AMR-WB로의 방향을 갖는 링크에서, CNG ER 프레임은 AMR-WB가 상기 프레임들을 이용하지 않기 때문에 시스템 인터페이스에서 폐기된다. 반대 방향에 있어서, 상기 프레임들은 이용가능하지 않고(AMR-WB가 SID 프레임들만을 생성하고 있다) 프레임 삭제로서 선언된다. 모든 액티브 음성 프레임들은 본질적으로 12.65 kbit/s에서 AMR-WB 부호화 표준인 상호 운용 FR 부호화 유형을 가지고 처리된다(단계 706).

하프 레이트 맥스 동작(Half-Rate Max operation)에서의 레이트 선택 및 신호 분류

본 발명의 제2 태양의 제6 예시적인 실시예에 따라 사운드 신호를 디지털로 부호화하는 방법(800)이 도 9에 도시된다. 방법(800)은 프리미엄 및 표준 모드에 대한 하프 레이트 맥스 동작에서의 부호화 및 음성 신호의 분류를 허용한다.

상술된 바와 같이, CDMA 시스템은 특정 프레임에 대해 최대 비트율을 부과한다. 종종, 시스템에 의해 부과되는 최대 비트율은 HR로 제한된다. 하지만, 상기 시스템은 또한 더 낮은 레이트를 부과할 수 있다.

종래 일반 동작 동안 FR로서 분류된 모든 액티브 음성 프레임들이 이제 HR 부호화 유형을 이용하여 부호화된다. 분류 및 레이트 선택 메커니즘은 (단계 506에서 부호화되는) 유성음 HR을 이용하는 모든 유성음 프레임들 및 (단계 408에서 부호화되는) 무성음 HR을 이용하는 모든 무성음 프레임들을 분류한다. 일반 동작 동안 FR로서 분류되는 모든 나머지 프레임들은 상호 운용 HR 부호화 유형이 사용되는 상호 운용 모드(도 11의 단계 908)를 제외하고 단계 514에서 일반 HR 부호화 유형을 이용하여 부호화된다.

도 9에서 볼 수 있는 바와 같이, 신호 분류 및 부호화 메커니즘은 표준 모드에서의 일반 동작과 유사하다. 하지만, 일반 HR (단계 514)은 일반 FR 부호화(도 5에서 단계 406) 대신에 사용되고 무성음 및 유성음 프레임들을 구별하는데 사용되는 임계값은 무성음 HR 및 유성음 HR 부호화 유형을 이용하여 부호화되기 위해 가능한 한 많은 프레임들을 허용하도록 더 관대하다. 기본적으로, 절약 모드에 대한 임계값은 프리미엄 또는 표준 모드 하프 레이트 맥스 동작의 경우에 사용된다.

본 발명의 제1 태양의 제7 예시적인 실시예에 따라 사운드 신호를 디지털로 부호화하는 방법(900)이 도 10에 도시된다. 방법(900)은 절약 모드에 대한 하프 레이트 맥스 동작에서의 부호화 및 음성 신호의 분류를 허용한다. 도 10의 방법(900)은 일반 FR을 가지고 부호화된 모든 프레임들이 이제 일반 HR을 가지고 부호화된다(하프 레이트 맥스 동작에서 저 에너지 프레임 분류에 대해 불필요함)는 점을 제외하고는 도 7의 방법(600)과 유사하다. 본 발명의 제1 태양의 제8 예시적인 실시

예에 따라 사운드 신호를 디지털로 부호화하는 방법(920)이 도 11에 도시된다. 방법(920)은 하프 레이트 맥스 동작 동안 상호 운용 모드에서의 레이트 결정 및 음성 신호의 분류를 허용한다. 방법(920)이 도 8의 방법(700)과 매우 유사하기 때문에, 상기 방법들 간의 차이만이 본 명세서에서 설명될 것이다.

방법(920)의 경우에 있어서, 부호화 유형들이 AMR-WB 상대편에 의해 이해될 수 없기 때문에 어떠한 신호 특정 부호화 유형(무성음 HR 및 유성음 HR)도 사용될 수 없고, 또한 일반 HR 부호화도 사용될 수 없다. 따라서, 하프 레이트 맥스 동작에서의 모든 액티브 음성 프레임들은 상호 운용 HR 부호화 유형을 이용하여 부호화된다.

시스템이 HR보다 더 낮은 최대 비트율을 부과하는 경우, 특히 상기 경우들이 극히 드물고 상기 프레임들이 프레임 삭제로 선언될 수 있기 때문에, 어떠한 일반적인 부호화 유형도 상기 경우들을 처리하는데 제공되지 않는다. 하지만, 최대 비트율이 시스템에 의해 QR로 제한되고 신호가 무성음으로서 분류되는 경우, 무성음 QR이 사용될 수 있다. 하지만 이것은 AMR-WB 상대편이 QR 프레임을 해석할 수 없기 때문에, CDMA 특정 모드(프리미엄, 표준, 절약)에서만 가능하다.

AMR-WB 및 레이트 세트 II VMR-WB 코덱간의 효율적인 상호 운용

본 발명의 제4 태양의 예시적인 실시예에 따라 VMR-WB 및 AMR-WB 코덱들 간의 상호 운용을 위해 사운드 신호를 디지털로 부호화하는 방법(1000)이 도 12를 참조하여 이제 설명될 것이다.

보다 상세하게는, 방법(1000)은 예를 들어 (VMR-WB 코덱으로 지칭되는) CDMA2000 시스템을 위해 설계된 소스 제어되는 VBR 코덱 및 AMR-WB 표준 코덱간의 탠덤-프리 동작을 가능하게 한다. 방법(1000)에 의해 허용되는 상호 운용 모드에 있어서, VMR-WB 코덱은 AMR-WB 코덱에 의해 해석될 수 있고 예를 들어 CDMA 코덱에서 사용되는 레이트 세트 II 비트율에 적합한 비트율을 이용한다.

레이트 세트 II의 비트율이 FR 13.3, HR 6.2, QR 2.7, 및 ER 1.0 kbit/s이기 때문에, 사용될 수 있는 AMR-WB 코덱 비트율은 풀 레이트에서 12.65, 8.85, 또는 6.6 및 1/4 레이트에서 1.75 kbit/s의 SID 프레임이다. 12.65 kbit/s의 AMR-WB는 비트율이 CDMA2000 FR 13.3 kbit/s에 가장 근접하고 예시적인 실시예에서 FR 코덱으로서 이용된다. 하지만, AMR-WB가 GSM 시스템에서 사용되는 경우, 링크 적응 알고리즘은 (더 많은 비트들을 채널 부호화에 할당하기 위하여) 채널 조건에 의존하여 비트율을 8.85 또는 6.6 kbit/s까지 낮출 수 있다. 따라서, AMR-WB의 8.85 및 6.6 kbit/s 비트율은 상호 운용 모드의 일부일 수 있고 상기 비트율 중 하나를 이용하여 결정된 GSM 시스템의 경우 CDMA2000 수신기에서 이용될 수 있다. 도 12의 예시적인 실시예에 있어서, 12.65, 8.85, 및 6.6 kbit/s의 AMR-WB 레이트에 대응하여 3가지 유형의 I-FR이 사용되고 각각 I-FR-12, I-FR-8, 및 I-FR-6으로 표시될 것이다. I-FR-12에는, 13개의 사용되지 않은 비트들이 있다. 처음 8개의 비트들은 (프레임 삭제 은폐를 개선하기 위하여 추가 비트들을 이용하는) 일반 FR 프레임들 및 I-FR 프레임들을 구별하는데 사용된다. 다른 5개의 비트들은 3가지 유형의 I-FR 프레임 신호를 전송하는데 사용된다. 보통 동작에 있어서, I-FR-12가 사용되고 GSM 링크 적응에 의해 필요한 경우 더 낮은 레이트가 사용된다.

CDMA2000 시스템에 있어서, 음성 코덱의 평균 데이터율은 시스템 용량에 직접 관련된다. 따라서 음성 품질이 최소 손실을 가지고 가능한 최저의 ADR을 획득하는 것이 상당히 중요하다. AMR-WB 코덱은 GSM 셀룰러 시스템 및 GSM 발달 (evolution)에 근거하는 제3 세대 무선을 위해 주로 설계되었다. 따라서 CDMA2000 시스템을 위한 상호 운용 모드는 특히 CDMA2000 시스템을 위해 설계된 VBR 코덱에 비해 더 높은 ADR을 야기한다. 주요 원인은 다음과 같다.

- AMR-WB에서의 6.2 kbit/s의 하프 레이트 모드의 결여;
- AMR-WB에서의 SID의 비트율이 레이트 세트 II 1/8 레이트(ER)에 적합하지 않은 1.75 kbit/s이다;
- AMR-WB의 VAD/DTX 동작은 SID_FIRST 프레임을 계산하기 위하여 (음성 프레임으로서 부호화되는) 행오버의 몇몇 프레임들을 사용한다.

AMR-WB 및 VMR-WB 코덱 간의 상호 운용을 위해 음성 신호를 부호화하는 방법은 상술된 제한을 극복하게 할 수 있고, 비교할만한 음성 품질을 가지고 CDMA2000 특정 모드에 동등하도록 상호 운용 모드의 ADR을 감소시킨다. 양방향 동작을 위한 방법이 후술된다: VMR-WB 부호화 - AMR-WB 복호화, 및 AMR-WB 부호화 - VMR-WB 복호화.

VMR-WB 부호화 - AMR-WB 복호화

CDMA VMR-WB 코덱 측의 부호화의 경우, AMR-WB 표준의 VAD/DTX/CNG 동작이 요구되지 않는다. VAD는 VMR-WB 코덱에 적합하고 다른 CDMA2000 특정 모드에서와 같은 방식으로 작용한다. 즉 사용된 VAD 행오버는 무성음 중단을 미스하지 않도록 필요한 만큼 길고, VAD_flag=0인 경우(분류된 배경 잡음) CNG 부호화가 동작한다.

VAD/CNG 동작은 AMR DTX 동작에 가능한 한 근접하도록 수행된다. AMR-WB 코덱에서의 VAD/DTX/CNG 동작은 다음과 같이 작용한다. 액티브 음성기간 이후에 7개의 배경 잡음 프레임이 음성 프레임으로서 부호화되지만, VAD 비트는 제로로 세팅된다(DTX 행오버). 그 다음 SID_FIRST 프레임이 전송된다. SID_FIRST 프레임에서 신호가 부호화되지 않고 CNG 매개변수들이 복호기의 (7 음성 프레임들) DTX 행오버로부터 유도된다. DTX 행오버 오버헤드를 감소시키기 위하여 24 프레임들보다 더 짧은 액티브 음성기간 이후에 AMR-WB는 DTX 행오버를 이용하지 않는다는 것을 유념한다. SID_FIRST 프레임 이후에, 2개의 프레임이 NO_DATA 프레임(DTX), 그 다음 SID_UPDATE 프레임(1.75 kbit/s)으로서 전송된다. 그 다음, 7개의 NO_DATA 프레임이 SID_UPDATE 프레임에 이어 전송된다. 이것은 액티브 음성 프레임이 검출될 때까지 계속된다(VAD_flag = 1).[4]

도 12의 예시적인 실시예에 있어서, VMR-WB 코덱에서의 VAD는 DTX 행오버를 이용하지 않는다. 액티브 음성 기간 이후에 제1 배경 잡음 프레임은 1.75 kbit/s로 부호화되고 QR로 전송되며, 1 kbit/s(1/8 레이트)로 부호화된 2개의 프레임이 있고 QR에서 전송되는 1.75 kbit/s의 다른 프레임이 있다. 그 이후, 7개의 프레임이 하나의 QR 프레임 등에 이어 ER로 전송된다. 이것은 ADR을 감소시키기 위하여 어떠한 DTX 행오버도 사용되지 않는다는 점을 제외하고 AMR-WB DTX 동작에 대체로 대응한다.

상기 예시적인 실시예에서 기술된 VMR-WB 코덱에서의 VAD/CNG 동작이 AMR-WB DTX 동작에 근접하지만, 추가로 ADR을 감소시킬 수 있는 다른 방법들이 사용될 수 있다. 예를 들어, QR CNG 프레임이 덜 빈번하게, 예를 들어 12 프레임마다 한번 전송될 수 있다. 또한, 잡음 변동(noise variations)이 부호기에서 평가될 수 있고 잡음 특징이 변경되는 경우에만(8 또는 12 프레임마다 1번이 아니고) QR CNG 프레임이 전송될 수 있다.

AMR-WB 부호기에서 6.2 kbit/s의 하프 레이트의 부존재의 제한을 극복하기 위하여, 풀 레이트 프레임으로서 프레임을 부호화하고 대수 코드북 인덱스에 대응하는 비트들(12.65 kbit/s의 AMR-WB에서의 프레임당 144 비트)을 드롭하는 것을 포함하는 상호 운용 하프 레이트(I-HR; interoperable half rate)가 제공된다. 이것은 비트율을 CDMA2000 레이트 세트 II 하프 레이트에 적합한 5.45 kbit/s로 감소시킨다. 복호화 이전에, 드롭된 비트들은 랜덤하게(즉 랜덤 생성기를 이용하여) 또는 의사-랜덤하게(즉 현존 비트스트림의 부분을 반복하여) 또는 어떤 소정의 방식으로 생성될 수 있다. 덤-앤-버스트(dim-and-burst) 또는 하프 레이트 맥스 요구가 CDMA2000 시스템에 의해 시그널링되는 경우 I-HR이 사용될 수 있다. 이것은 음성 프레임을 손실된 프레임으로서 선언하는 것을 피한다. 또한 합성된 음성 품질에 대한 대수 코드북 기여가 최소화되는 프레임 또는 무성음 프레임을 부호화하기 위하여 상호 운용 모드에서 VMR-WB 코덱에 의해 I-HR이 사용될 수 있다. 이것은 ADR을 감소시킨다. 이 경우에 있어서, 부호기는 I-HR 모드에서 부호화되는 프레임을 선택할 수 있고 따라서 그러한 프레임의 사용에 의해 야기되는 음성 품질 저하를 최소화할 수 있다.

도 12에 도시된 바와 같이, VMR-WB 부호화/AMR-WB 복호화 방향으로, 음성 프레임들은 VMR-WB 부호기(1002)의 상호 운용 모드를 가지고 부호화되고, 다음의 가능한 비트율 중 하나를 출력한다. 액티브 음성 프레임들에 대한 I-FR(I-FR-12, I-FR-8, 또는 I-FR-6), 덤-앤-버스트 시그널링의 경우 또는 옵션으로서 어떤 무성음 프레임들 또는 합성된 음성 품질에 대한 대수 코드북 기여가 최소화되는 프레임들을 부호화하기 위한 I-HR, (상술된 바와 같은 8개의 배경 잡음 프레임 중 하나, 또는 잡음 특징에서의 변동이 검출되는 경우) 관련된 배경 잡음 프레임을 부호화하기 위한 QR CNG, 및 (배경 잡음 프레임이 QR CNG 프레임으로서 부호화되지 않은) 대부분의 배경 잡음 프레임에 대한 ER CNG 프레임들. 게이트웨이의 형태를 갖는 시스템 인터페이스에서, 다음 동작이 수행된다.

우선, VMR-WB 부호기로부터 게이트웨이에 의해 수신된 프레임의 유효성이 테스트된다. 유효한 상호 운용 모드 VMR-WB 프레임이 아닌 경우, 상기 프레임은 삭제로서 전송된다(AMR-WB의 음성 손실된 유형). 상기 프레임은 예를 들어 다음 조건들 중 하나가 발생하는 경우 무효한 것으로 고려된다.

- 모두 제로(all-zero) 프레임이 수신되는 경우(버스트 및 공백(blank)의 경우에서의 네트워크에 의해 사용되는) 상기 프레임은 삭제된다.

- FR 프레임의 경우, 3개의 프리앰블(preamble) 비트가 I-FR-12, I-FR-8 또는 I-FR-6에 대응하지 않는 경우, 또는 미사용 비트들이 제로가 아닌 경우, 상기 프레임은 삭제된다. 또한 I-FR이 VAD 비트를 1로 세팅하고 따라서 수신된 프레임의 VAD 비트가 1이 아닌 경우 프레임은 삭제된다.

- FR에 유사한 HR 프레임의 경우, 프리앰블 비트가 I-HR-12, I-HR-8 또는 I-HR-6에 대응하지 않는 경우, 또는 미사용 비트들이 제로가 아닌 경우, 상기 프레임은 삭제된다. VAD 비트에 대해서도 동일하다.

- QR 프레임의 경우, 프리앰블 비트가 CNG QR에 대응하지 않는 경우 프레임은 삭제된다. 또한, VMR-WB 부호기는 SID_UPDATE 비트를 1로 세팅하고 모드 요구 비트를 0010으로 세팅한다. 이것이 그 경우가 아닌 경우 프레임이 삭제된다.

- ER 프레임의 경우, 모두 1(all-one) ER 프레임이 수신되는 경우 프레임이 삭제된다. 또한 VMR-WB 부호기는 공백 프레임을 시그널링하기 위하여 (최초 14개 비트) 모두 제로 ISF 비트 패턴을 사용한다. 이 패턴이 수신되는 경우 프레임은 삭제된다.

수신된 프레임이 유효한 상호 운용 모드 프레임인 경우, 다음 동작이 수행된다.

- I-FR 유형에 의존하여 I-FR 프레임은 12.65, 8.8, 또는 6.6 kbit/s 프레임으로서 AMR-WB 복호기에 전송된다.

- QR CNG 프레임은 SID_UPDATE 프레임으로서 AMR-WB 복호기에 전송된다.

- ER CNG 프레임은 NO_DATA 프레임으로서 AMR-WB 복호기에 전송된다.

- I-HR 프레임은 단계 1010에서 손실된(missing) 대수 코드북 인덱스를 생성함으로써 (프레임 유형에 근거하여) 12.65, 8.85, 또는 6.6 kbit/s 프레임으로 변환된다. 상기 인덱스는 랜덤하게, 또는 현존 부호화 비트들의 부분을 반복함으로써, 또는 몇몇 소정의 방식으로 생성될 수 있다. 또한 I-HR 유형을 나타내는 비트들(VMR-WB 코덱에서의 상이한 하프 레이트 유형을 구별하는데 사용된 비트들)을 폐기한다.

AMR-WB 부호화 - VMR-WB 복호화

이 방향에 있어서, 방법(1000)은 AMR-WB DTX 동작에 의해 제한된다. 하지만, 액티브 음성 부호화 동안, (DTX 행오버 기간 동안 0, 액티브 음성에 대해 1) VAD_flag를 나타내는 (제1 데이터 비트) 비트스트림에 하나의 비트가 있다. 따라서 게이트웨이에서의 동작은 다음과 같이 요약될 수 있다.

- SID_UPDATE 프레임은 QR CNG 프레임으로서 전달된다.

- SID_FIRST 프레임 및 NO_DATA 프레임은 ER 공백 프레임으로서 전달된다.

- 삭제된 프레임(음성 손실)은 ER 삭제 프레임으로서 전송된다.

- (단계 1012에서 검증된) VAD_flag = 0을 갖는 액티브 음성 이후 최초 프레임은 FR 프레임으로서 유지되지만 VAD_flag = 0을 갖는 다음 프레임들은 ER 공백 프레임으로서 전달된다.

- FR 프레임을 수신하는 경우 단계 1014에서 게이트웨이가 하프-레이트-맥스 동작(프레임 레벨 시그널링) 요구를 수신하는 경우, 프레임은 I-HR 프레임으로 번역된다. 이것은 대수 코드북 인덱스에 대응하는 비트를 드롭하고 I-HR 프레임 유형을 나타내는 모드 비트를 추가하는 것으로 구성된다.

이러한 예시적인 실시예에 있어서, ER 공백 프레임에서 최초 2개의 바이트는 0x00으로 세팅되고, ER 삭제 프레임에서 최초 2개의 바이트는 0x04로 세팅된다. 기본적으로, 최초 14개의 비트들이 ISF 인덱스에 대응하고 2개의 패턴들이 공백 프레임(모두 제로) 또는 삭제 프레임(16진수로 0x04인 1로 세팅된 14번째 비트를 제외하고 모두 제로)을 나타내기 위해 유보된다. VMR-WB 복호기(1004)에서, 공백 ER 프레임들이 검출되는 경우, 상기 프레임들은 CNG 복호기에 의해 최종 수신된 우량한 CNG 매개변수를 이용하여 처리된다. 최초 수신된 공백 ER 프레임의 경우 예외이다(CNG 복호기 초기화; 구 CNG 매개변수들이 아직 알려지지 않음). VAD_flag=0을 갖는 최초 프레임이 FR로서 전송되는 경우, 최종 CNG 매개변수 뿐 아니라 상기 프레임으로부터의 매개변수가 CNG 동작을 초기화하는데 사용된다. ER 삭제 프레임의 경우, 복호기는 삭제 프레임에 사용되는 은폐 절차(concealment procedure)를 이용한다.

도 12에 도시된 예시적인 실시예에 있어서 12.65 kbit/s가 FR 프레임에 대해 사용된다는 점을 유념한다. 하지만, 불량 채널 조건의 경우 더 낮은 레이트의 이용을 필요로 하는 링크 적응 알고리즘에 따라 8.85 및 6.6 kbit/s가 또한 사용될 수 있다. 예를 들어, CDMA2000 및 GSM 시스템 간의 상호 동작을 위해, GSM 시스템의 링크 적응 모듈은 불량 채널 조건의 경우 비트율을 8.85 또는 6.6 kbit/s로 낮출 것을 결정할 수 있다. 이 경우에 있어서, 상기 더 낮은 비트율은 CDMA VMR-WB 솔루션에 포함될 필요가 있다.

레이트 세트 I 에서 동작하는 CDMA VMR-WB 코덱

레이트 세트 I 에서, 사용되는 비트율은 FR에 대해 8.55 kbit/s, HR에 대해 4.0 kbit/s, QR에 대해 2.0 kbit/s, 및 ER에 대해 800 bit/s이다. 이 경우에 있어서, 6.6 kbit/s의 AMR-WB 코덱만이 FR에서 사용될 수 있고 CNG 프레임들은 (상술된 레이트 세트 II 동작에 유사하게) 다른 배경 잡음 프레임에 대해 ER 또는 QR(SID_UPDATE)로 전송될 수 있다. 6.6 kbit/s 레이트의 낮은 품질의 제한을 극복하기 위하여, AMR-WB 코덱의 8.85 kbit/s 비트율과 상호 동작하는 8.55 kbit/s 레이트가 제공된다. 이것은 레이트 세트 I 상호운용 FR(I-FR-I)로 지칭될 것이다. I-FR-I의 2개의 가능한 구성 및 8.85 kbit/s 레이트의 비트 할당이 표 6에 표시된다.

[표 6] 레이트 세트 I 구성의 I-FR-I 부호화 유형의 비트 할당

매개변수	8.85 kbit/s의 AMR-WB	8.55 kbit/s의 I-FR-I (구성 1)	8.55 kbit/s의 I-FR-I (구성 2)
	비트/프레임	비트/프레임	비트/프레임
하프-레이트 모드 비트	-	-	-
VAD 플래그	1	0	0
LP 매개변수	46	41	46
피치 지연	26 = 8 + 5 + 8 + 5	26	26
이득	24 = 6 + 6 + 6 + 6	24	24
대수 코드북	80 = 20 + 20 + 20 + 20	80	75
합계	177	171	171

I-FR-I에서, VAD_flag 비트 및 추가 5 비트들이 8.55 kbit/s 레이트를 획득하기 위해 드롭된다. 드롭된 비트들은 8.85 kbit/s 복호기가 사용될 수 있도록 복호기 또는 시스템 인터페이스에서 용이하게 도입될 수 있다. 음성 품질에 거의 영향을 미치지 않는 방식으로 몇몇 방법들이 5 비트들을 드롭하는데 사용될 수 있다. 표 6에 표시된 구성 1에 있어서, 5 비트들은 선형 예측(LP) 매개변수 양자화로부터 드롭된다. AMR-WB에서, 46 비트들이 (평균 제거(mean removal) 및 이동 평균 예측을 이용하여) 이미턴스 스펙트럼 페어(ISP; immitance spectrum pair) 영역에서의 LP 매개변수들을 양자화하는데 사용된다. (예측 이후) 16 디멘션(dimension) ISP 잔류 벡터는 분할-다단(split-multistage) 벡터 양자화를 이용하여 양자화된다. 벡터는 2개의 서브벡터 디멘션 9 및 7로 각각 분할된다. 2개의 서브벡터는 2 단계에서 양자화된다. 제1 단계에서 각 서브벡터는 8 비트를 가지고 양자화된다. 양자화 에러 벡터는 제2 단계에서 각각 3 및 2 서브벡터들로 분할된다. 제2 단계 서브벡터들은 디멘션 3, 3, 3, 3, 및 4이고, 각각 6, 7, 7, 5 및 5 비트를 가지고 양자화된다. 제안된 I-FR-I 모드에 있어서, 최종 제2 단계 서브벡터의 5 비트들이 드롭된다. 상기 비트들은 스펙트럼의 고주파 부분에 대응하기 때문에 최소의 영향을 갖는다. 상기 5 비트들을 드롭하는 것은 실제로 전송될 필요가 없는 어떤 값에 최종 제2 단계 서브벡터의 인덱스를 고정함으로써 수행된다. 이러한 5-비트 인덱스가 고정되어 있다는 사실은 VMR-WB 부호기에서의 양자화 동안 용이하게 고려된다. 고정 인덱스는 시스템 인터페이스에서(즉 VMR-WB 부호기/AMR-WB 복호기 동작) 또는 복호기에서(즉 AMR-WB 부호기/VMR-WB 복호기 동작) 추가된다. 이런 식으로 8.85 kbit/s의 AMR-WB 복호기는 레이트 세트 I I-FR 프레임을 복호화하는데 사용된다.

예시적인 실시예의 제2 구성에 있어서, 상기 5 비트들은 대수 코드북 인덱스들로부터 드롭된다. 8.85 kbit/s의 AMR-WB 에서, 프레임은 4개의 64-샘플 서브프레임으로 분할된다. 대수 여기 코드북은 서브프레임을 16 위치의 4 트랙으로 분할하고 각 트랙에 부호 펄스(signed pluse)를 위치시키는 것으로 구성된다. 각 펄스는 5 비트로 부호화된다: 위치에 대해 4 비트 및 부호에 대해 1 비트. 따라서, 각 서브프레임에 대해, 20-비트 대수 코드북이 사용된다. 5 비트들을 드롭하는 한가지 방법은 어떤 서브프레임으로부터 한 펄스를 드롭하는 것이다. 예를 들어, 제4 서브프레임에서의 제4 위치-트랙에서의 제4 펄스. VMR-WB 부호기에서, 각 펄스는 코드북 검색 동안 소정의 값(위치 및 부호)으로 고정될 수 있다. 이러한 알려진 펄스 인덱스는 시스템 인터페이스에 추가될 수 있고 AMR-WB 복호기에 전송될 수 있다. 다른 방향에 있어서, 이러한 펄스의 인덱스는 시스템 인터페이스에서 드롭되고, CDMA VMR-WB 복호기에서, 펄스 인덱스는 랜덤하게 생성될 수 있다. 다른 방법들이 또한 상기 비트들을 드롭하는데 사용될 수 있다.

CDMA2000 시스템에 의한 덤-앤-버스트 또는 하프-레이트 요구에 대처하기 위하여, 또한 레이트 세트 I 코덱에 대한 상호 운용 HR 모드가 제공된다(I-HR-I). 레이트 세트 II의 경우와 유사하게, 몇몇 비트들은 AMR-WB 부호화/VMR-WB 복호화 동작 동안 시스템 인터페이스에서 드롭되어야 하고, 또는 VMR-WB 부호화/AMR-WB 복호화 동안 시스템 인터페이스에서 생성되어야 한다. I-HR-I의 예시적인 구성 및 8.85 kbit/s 레이트의 비트 할당이 표 7에 표시된다.

[표 7] 레이트 세트 I 구성에서의 I-HR-I 부호화 유형의 예시적인 비트 할당

매개변수	8.85 kbit/s의 AMR-WB 비트/프레임	4.0 kbit/s의 I-HR-I 비트/프레임
하프-레이트 모드 비트	-	-
VAD 플래그	1	0
LP 매개변수	46	36
피치 지연	26 = 8 + 5 + 8 + 5	20
이득	24 = 6 + 6 + 6 + 6	24
대수 코드북	80 = 20 + 20 + 20 + 20	0
합계	177	80

제안된 I-HR-I 모드에서, LP 필터 매개변수의 양자화에서의 최종 2 제2 단계 서브벡터의 10 비트들이 상술된 레이트 세트 II와 유사한 방식으로 시스템 인터페이스에서 드롭되거나 생성된다. 피치 지연은 4 서브프레임에서 7, 3, 7, 3 비트들의 비트 할당을 가지고 그리고 정수 해상도(integer resolution)를 가지고 부호화된다. 이것은 시스템 인터페이스에서 피치의 분수부분을 드롭하고 제2 및 제4 서브프레임에 대해 3 비트로 차등 지연(differential delay)을 클리핑하도록 AMR-WB 부호기/VMR-WB 복호기 동작에서 변환한다. 대수 코드북 인덱스들은 레이트 세트 II의 I-HR 솔루션에서와 유사하게 드롭된다. 신호 에너지 정보는 그대로 유지된다.

나머지의 레이트 세트 I 상호 운용 모드의 동작은 도 12에서 (VAD/DTX/CNG 동작에 의해) 상술된 레이트 세트 II 모드의 동작과 유사하고 여기서 더 상세하게 기술되지 않을 것이다.

비록 본 발명이 본 발명의 예시적인 실시예에 의해 상술되었지만, 청구된 범위에서 정의된 바와 같이 본 발명의 정신 및 본질을 벗어나지 않으면서 수정될 수 있다. 예를 들어, 본 발명의 예시적인 실시예는 음성 신호의 부호화와 관련하여 기술되었지만, 상기 실시예들은 또한 음성과는 다른 사운드 신호에 적용할 수 있다는 것을 유념해야 한다.

참고문헌

[1] ITU-T 권고 G.722.2 "적용 다중 레이트 광대역(AMR-WB)을 이용한 16 kbit/s에서의 광대역 음성 부호화 (Wideband coding of speech at around 16 kbit/s using Adaptive Multi-Rate Wideband(AMR-WB))", 제네바, 2002년

[2] 3GPP TS 26.190, "AMR 광대역 음성 코덱; 트랜스코딩 기능(AMR Wideband Speech Codec; Transcoding Functions)", 3GPP 기술 사양.

[3] 3GPP TS 26.192, "AMR 광대역 음성 코덱; 컴포트 잡음 태양(AMR Wideband Speech Codec; Comfort Noise Aspects)", 3GPP 기술 사양.

[4] 3GPP TS 26.193, "AMR 광대역 음성 코덱; 소스 제어되는 레이트 동작(AMR Wideband Speech Codec; Source Controlled Rate operation)", 3GPP 기술 사양.

[5] M. Jelinek 및 F. Labonte, "광대역 음성 및 오디오 부호화를 위한 강인한 신호/잡음 구별(Robust Signal/Noise Discrimination for Wideband Speech and Audio Coding)", Proc. IEEE 음성 부호화에 대한 워크샵, 151-153 페이지, USA, 위스콘신, 델라반, 2000년 9월.

[6] J.D.Johnston, "인식 잡음 기준을 이용한 오디오 신호의 부호화 변환(Transform Coding of Audio Signals Using Perceptual Noise Criteria)", IEEE Jour. 통신의 선택된 영역에 대한. vol. 6, no. 2, 314-323 페이지.

[7] 3GPP2 C.S0030-0, "광대역 확산 스펙트럼 통신 시스템을 위한 선택가능한 모드 보코더 서비스 옵션(Selectable Mode Vocoder Service Option for Wideband Spread Spectrum Communication Systems)", 3GPP2 기술 사양.

[8] 3GPP2 C.S0014-0, "개선된 가변 레이트 코덱(Enhanced Variable Rate Codec(EVRC))", 3GPP2 기술 사양.

[9] TIA/EIA/IS-733, "광대역 확산 스펙트럼 통신 시스템을 위한 하이 레이트 음성 서비스 옵션 17(High Rate Speech Service option 17 for Wideband Spread Spectrum Communication Systems)", 또한 3GPP2 기술 사양 C.S0020-0.

도면의 간단한 설명

도 1은 본 발명의 제1 태양에 따라 음성 부호화 및 복호화 장치의 사용을 나타내는 음성 통신 시스템의 블록도이다.

도 2는 본 발명의 제2 태양의 제1 예시적인 실시예에 따라 사운드 신호를 디지털로 부호화하는 방법을 나타내는 흐름도이다.

도 3은 본 발명의 제3 태양의 예시적인 실시예에 따라 무성음 프레임을 식별하는 방법을 나타내는 흐름도이다.

도 4는 본 발명의 제4 태양의 예시적인 실시예에 따라 안정된 유성음 프레임을 식별하는 방법을 나타내는 흐름도이다.

도 5는 본 발명의 제2 태양의 제2 예시적인 실시예에 따라 프리미엄 모드에서 사운드 신호를 디지털로 부호화하는 방법을 나타내는 흐름도이다.

도 6은 본 발명의 제2 태양의 제3 예시적인 실시예에 따라 표준 모드에서 사운드 신호를 디지털로 부호화하는 방법을 나타내는 흐름도이다.

도 7은 본 발명의 제2 태양의 제4 예시적인 실시예에 따라 절약 모드에서 사운드 신호를 디지털로 부호화하는 방법을 나타내는 흐름도이다.

도 8은 본 발명의 제2 태양의 제5 예시적인 실시예에 따라 상호 운용 모드(interoperable mode)에서 사운드 신호를 디지털로 부호화하는 방법을 나타내는 흐름도이다.

도 9는 본 발명의 제2 태양의 제6 예시적인 실시예에 따라 하프 레이트 맥스 동안 프리미엄 또는 표준 모드에서 사운드 신호를 디지털로 부호화하는 방법을 나타내는 흐름도이다.

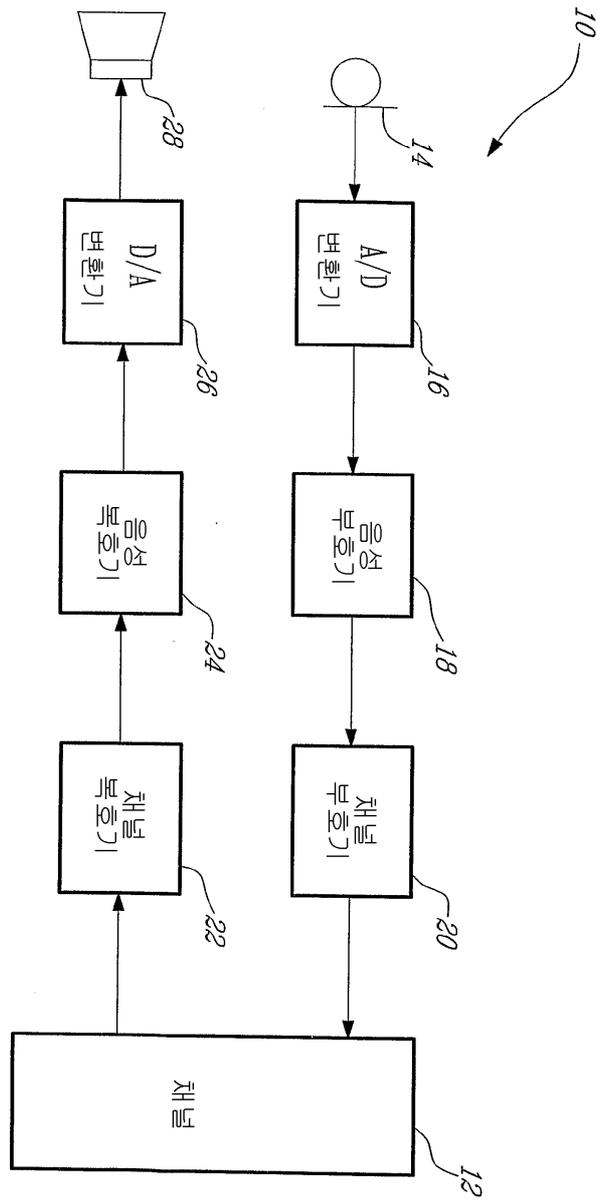
도 10은 본 발명의 제2 태양의 제7 예시적인 실시예에 따라 하프 레이트 맥스 동안 절약 모드에서 사운드 신호를 디지털로 부호화하는 방법을 나타내는 흐름도이다.

도 11은 본 발명의 제2 태양의 제8 예시적인 실시예에 따라 하프 레이트 맥스 동안 상호 운용 모드에서 사운드 신호를 디지털로 부호화하는 방법을 나타내는 흐름도이다.

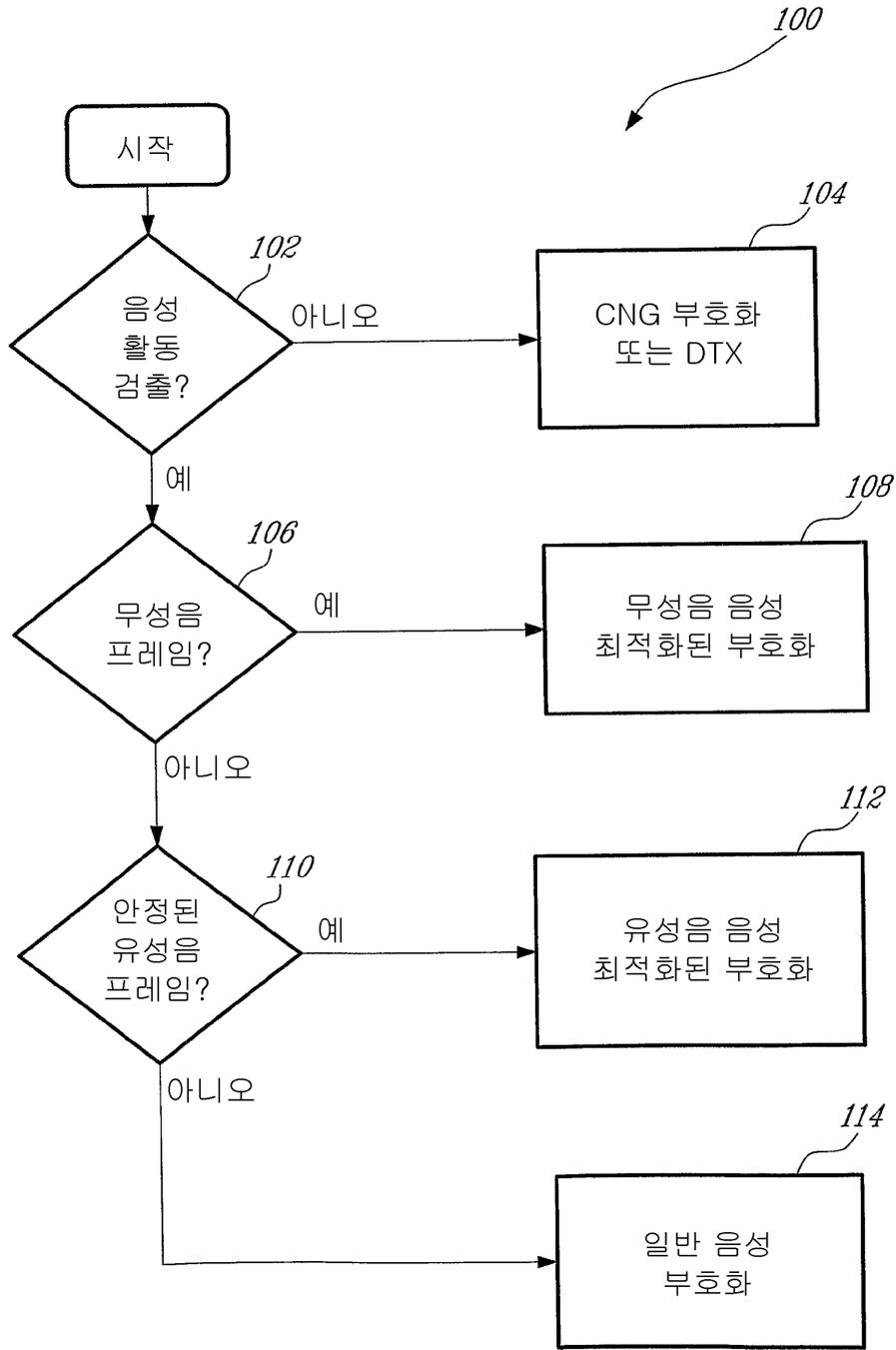
도 12는 본 발명의 제5 태양의 예시적인 실시예에 따라 VMR-WB 및 AMR-WB 코덱들 간의 상호 운용을 허용하기 위하여 사운드 신호를 디지털로 부호화하는 방법을 나타내는 흐름도이다.

도면

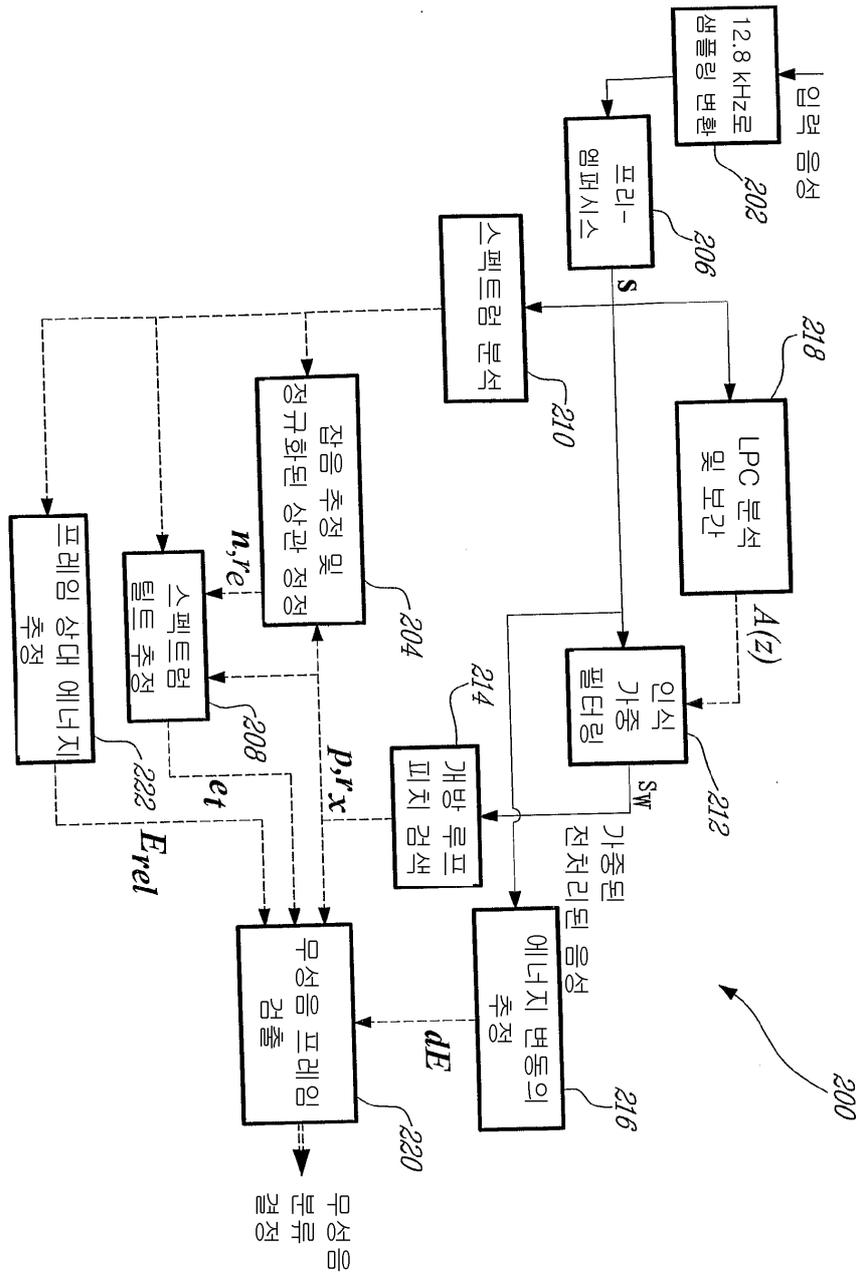
도면1



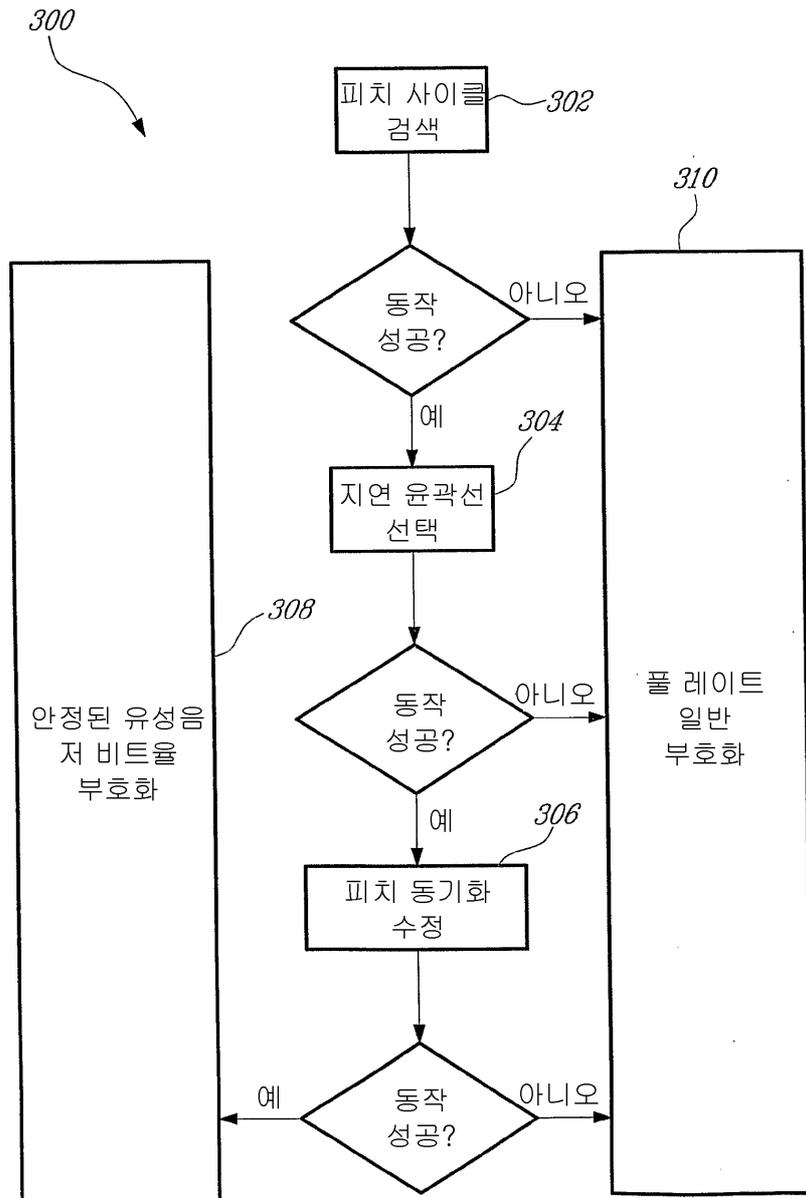
도면2



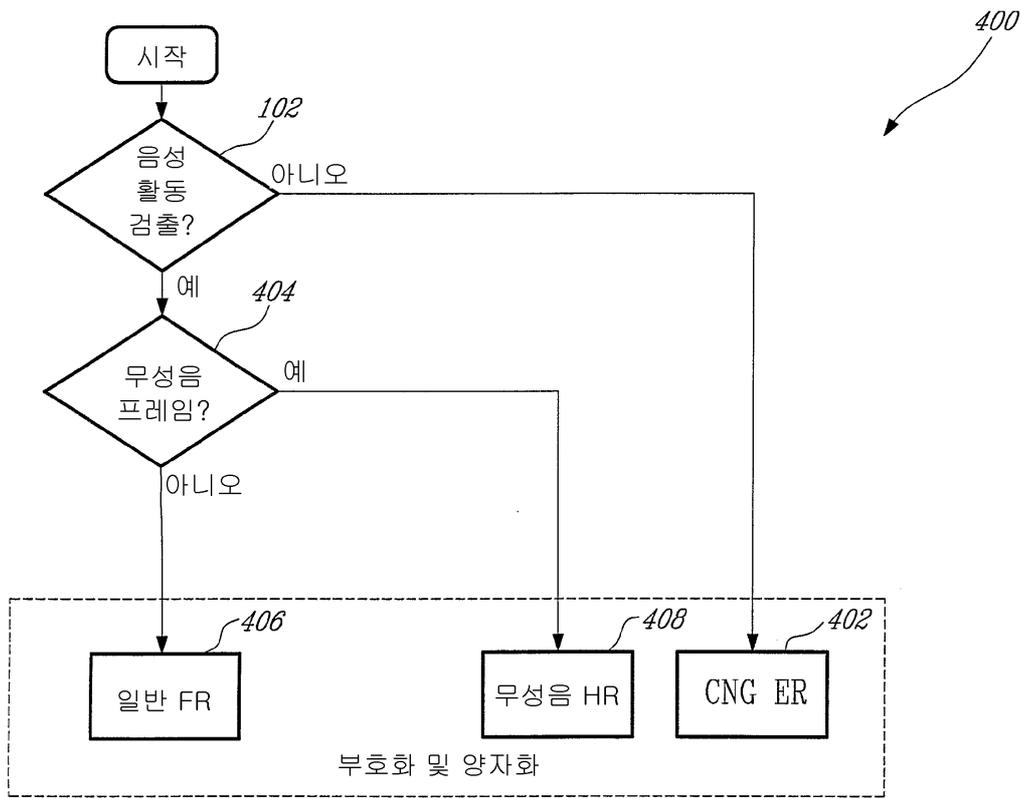
도면3



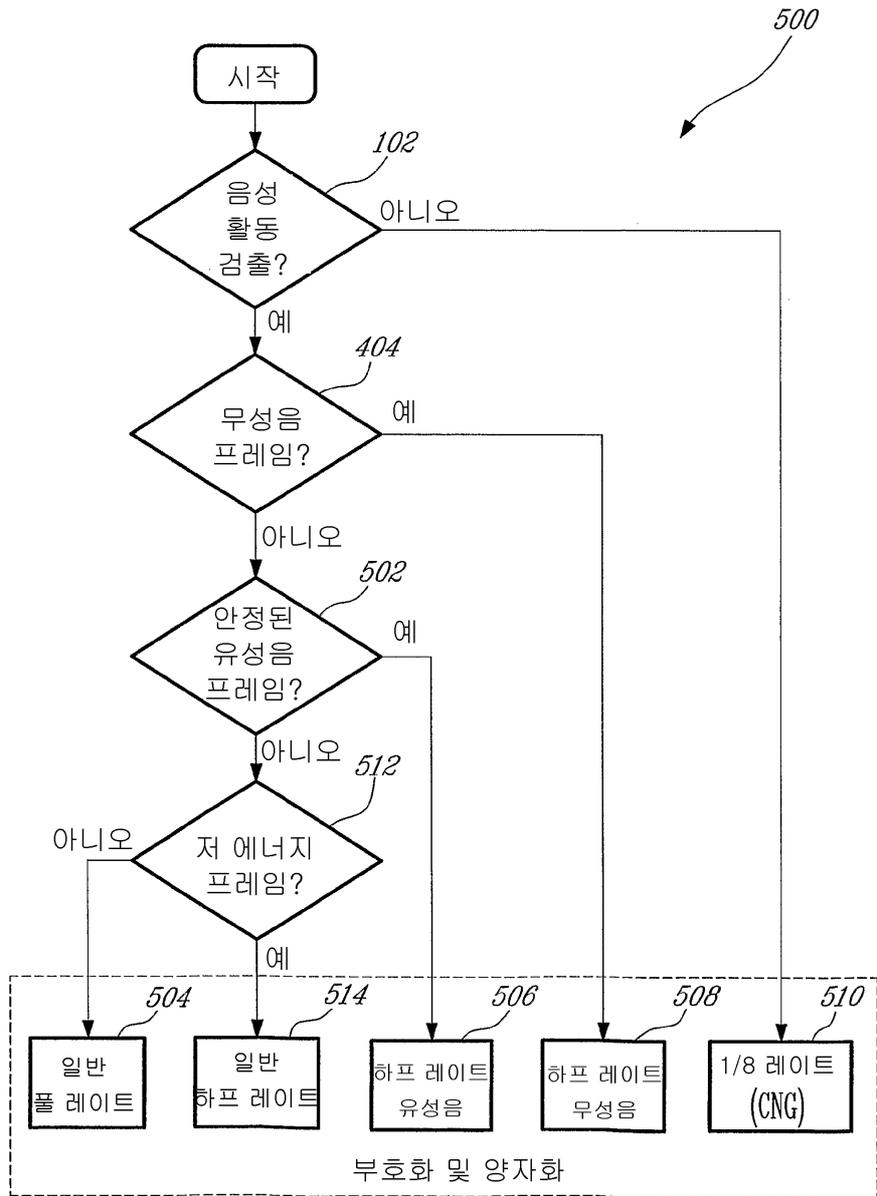
도면4



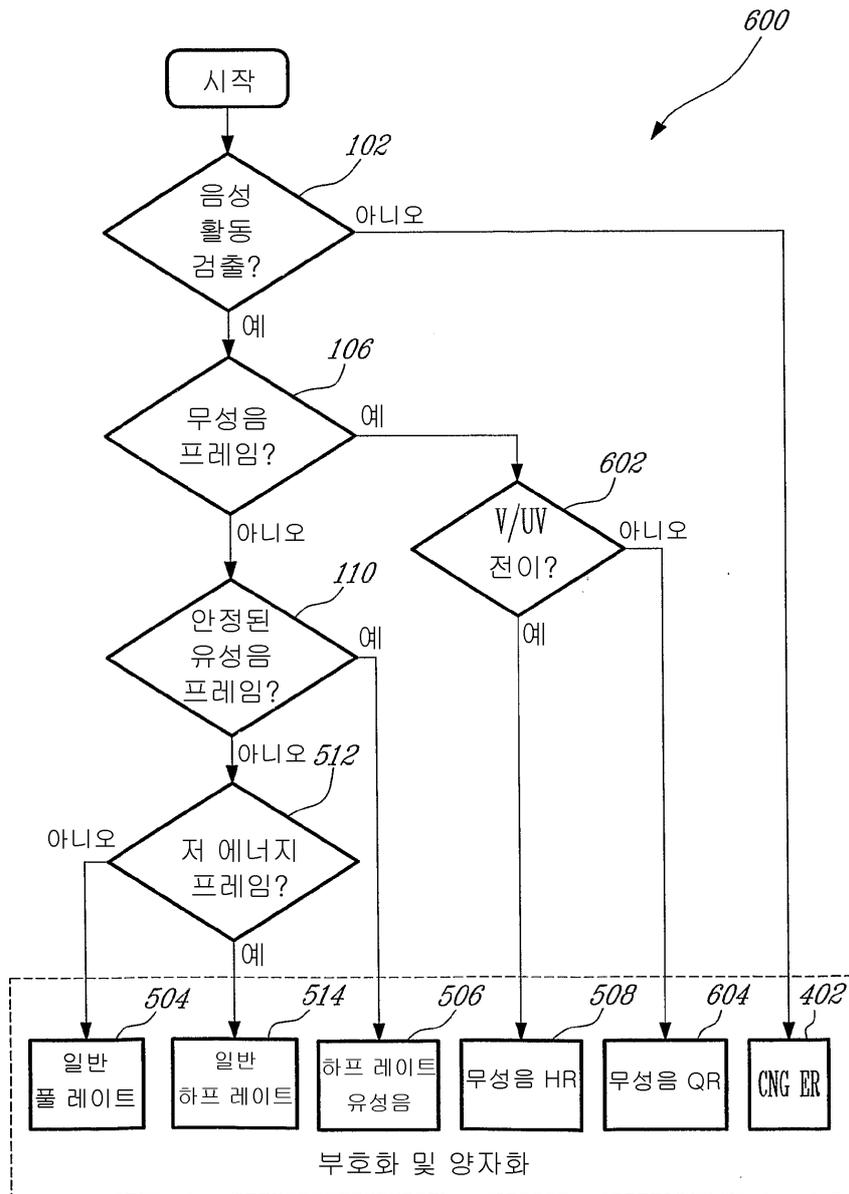
도면5



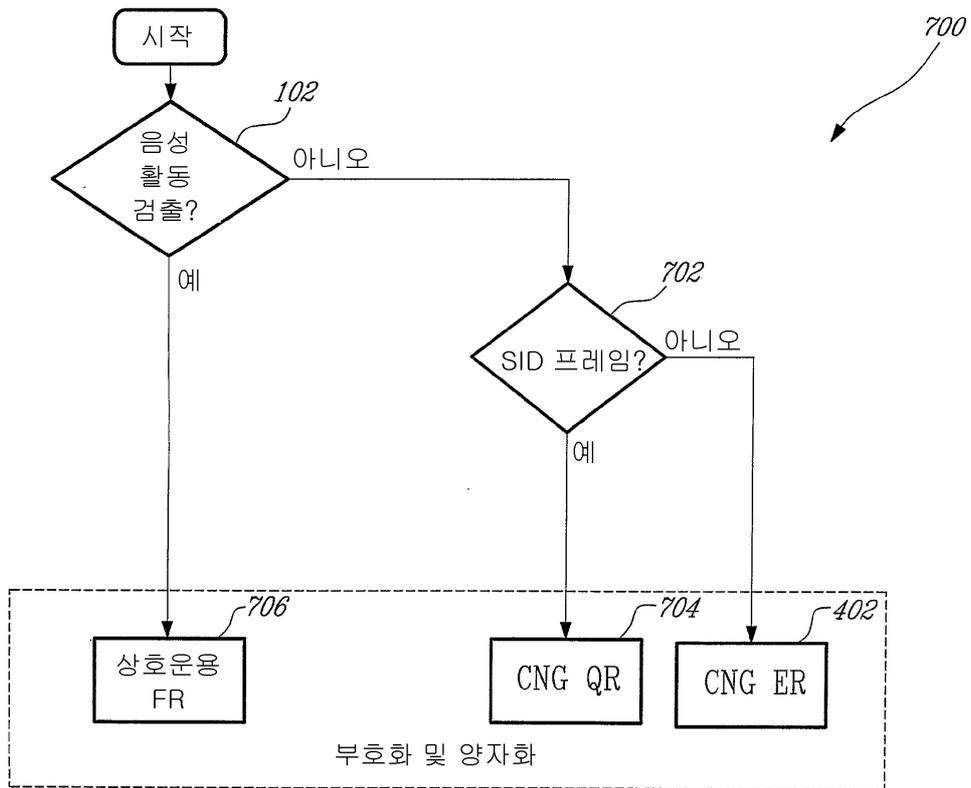
도면6



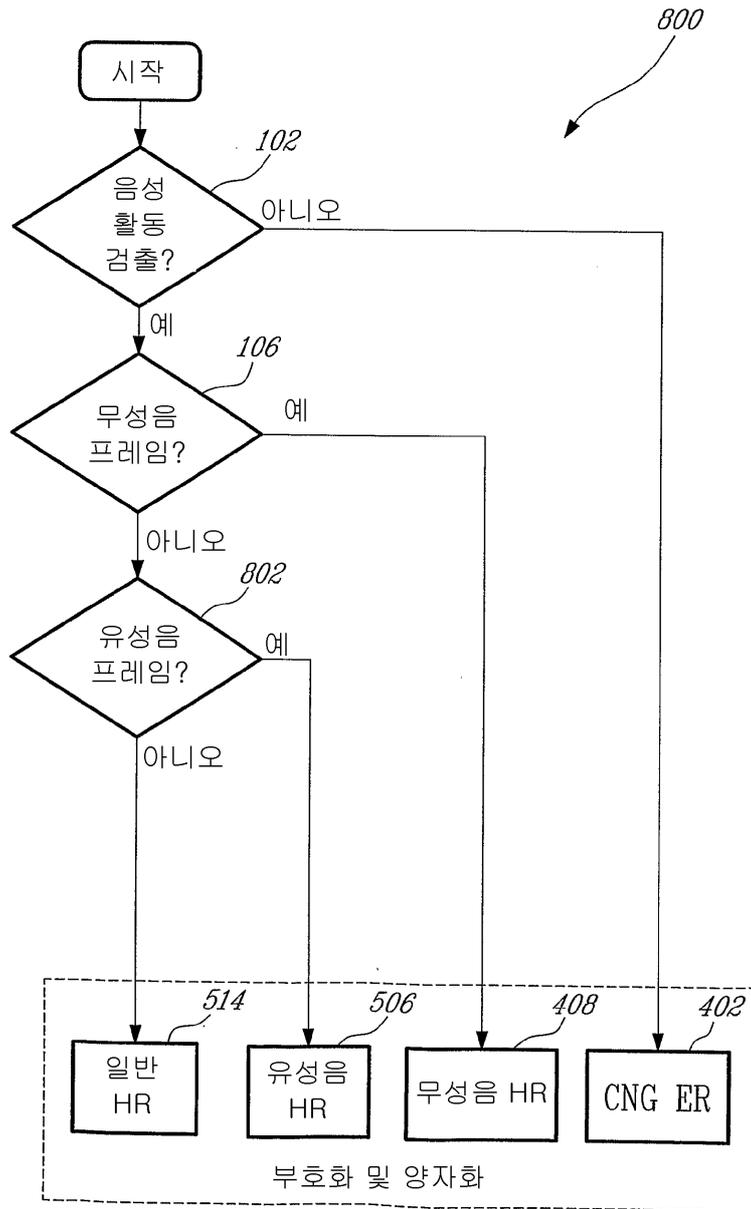
도면7



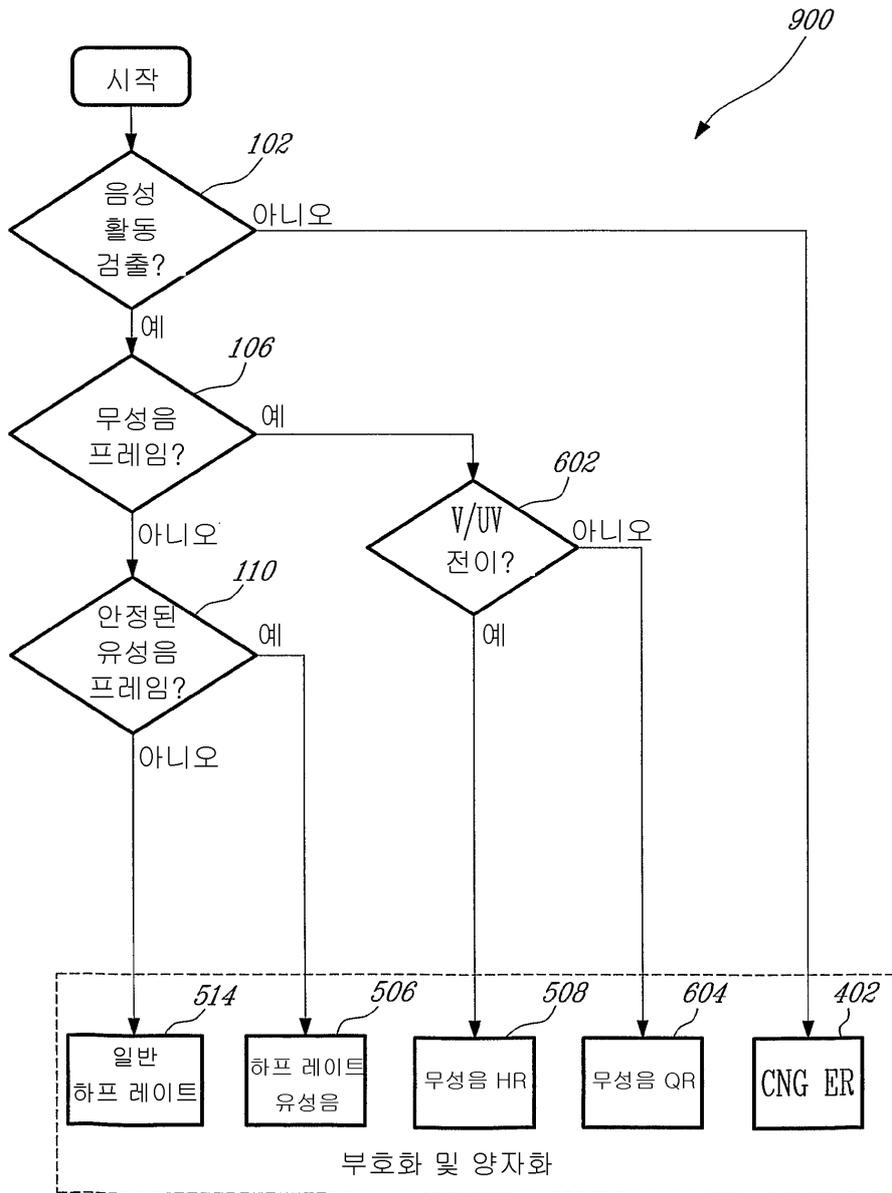
도면8



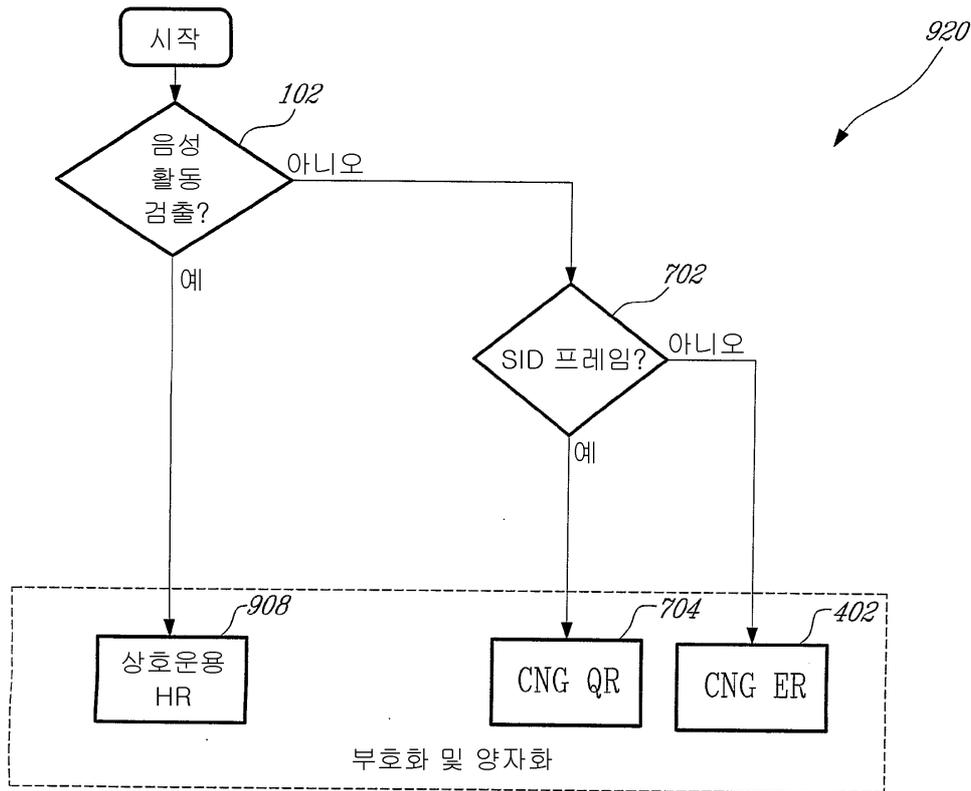
도면9



도면10



도면11



도면12

