



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(51) Int. Cl.

G10L 19/06 (2006.01)

(45) 공고일자 2007년08월29일
 (11) 등록번호 10-0752797
 (24) 등록일자 2007년08월21일

| | | | |
|-------------|-------------------|-------------|-----------------|
| (21) 출원번호 | 10-2002-7000784 | (65) 공개번호 | 10-2002-0033737 |
| (22) 출원일자 | 2002년01월19일 | (43) 공개일자 | 2002년05월07일 |
| 심사청구일자 | 2005년05월10일 | | |
| 번역문 제출일자 | 2002년01월19일 | | |
| (86) 국제출원번호 | PCT/US2000/019672 | (87) 국제공개번호 | WO 2001/06495 |
| 국제출원일자 | 2000년07월19일 | 국제공개일자 | 2001년01월25일 |

(81) 지정국

국내특허 : 알바니아, 아르메니아, 오스트리아, 오스트레일리아, 아제르바이잔, 보스니아 헤르체고비나, 바베이도스, 불가리아, 브라질, 벨라루스, 캐나다, 스위스, 중국, 쿠바, 체코, 독일, 덴마크, 에스토니아, 스페인, 핀란드, 영국, 그루지야, 헝가리, 이스라엘, 아이슬란드, 일본, 케냐, 키르키즈스탄, 북한, 대한민국, 카자흐스탄, 세인트루시아, 스리랑카, 리베이라, 레소토, 리투아니아, 룩셈부르크, 라트비아, 몰도바, 마다가스카르, 마케도니아공화국, 몽고, 말라위, 멕시코, 노르웨이, 뉴질랜드, 슬로베니아, 슬로바키아, 타지키스탄, 투르크멘, 터키, 트리니아드토바고, 우크라이나, 우간다, 우즈베키스탄, 베트남, 폴란드, 포르투칼, 루마니아, 러시아, 수단, 스웨덴, 싱가포르, 아랍에미리트, 안티구와바부다, 코스타리카, 도미니카, 알제리, 모로코, 탄자니아, 남아프리카, 벨리제, 모잠비크, 그러나다, 가나, 감비아, 크로아티아, 인도네시아, 인도, 시에라리온, 세르비아 앤 몬테네그로, 짐바브웨,

AP ARIPO특허 : 케냐, 레소토, 말라위, 수단, 스와질랜드, 우간다, 시에라리온, 가나, 감비아, 짐바브웨, 탄자니아, 모잠비크,

EA 유라시아특허 : 아르메니아, 아제르바이잔, 벨라루스, 키르키즈스탄, 카자흐스탄, 몰도바, 러시아, 타지키스탄, 투르크멘,

EP 유럽특허 : 오스트리아, 벨기에, 스위스, 독일, 덴마크, 스페인, 프랑스, 영국, 그리스, 아일랜드, 이탈리아, 룩셈부르크, 모나코, 네덜란드, 포르투칼, 스웨덴, 핀란드, 사이프러스,

OA OAPI특허 : 부르키나파소, 베닌, 중앙아프리카, 콩고, 코트디브와르, 카메룬, 가봉, 기니, 말리, 모리타니, 니제르, 세네갈, 차드, 토고, 기니 비사우,

(30) 우선권주장

09/356,755

1999년07월19일

미국(US)

(73) 특허권자

활콤 인코포레이티드

미국 캘리포니아 샌디에고 모어하우스 드라이브5775 (우 92121-1714)

(72) 발명자

아낸더패드마나한,아라사니팔라이,케이.

미국92126캘리포니아샌디에고#127카미노루즈10187

만주나쓰,사라쓰

미국92126캘리포니아샌디에고실링애브뉴#57104

(74) 대리인

남상선

(56) 선행기술조사문현

LOO J H Y et al. 'Classified nonlinear predictive vector quantization of speech spectral parameters' In:ICASSP'96. IEEE,1996,Vol.2, pages 761-764.

SKOGLUND J et al. 'Predictive VQ for Noisy Channel Spectrum Coding: AR or MA?' In:ICASSP'97. IEEE,21.04.1997, pages 1351-1354.

심사관 : 장호근

전체 청구항 수 : 총 20 항

(54) 음성 코더에서 선 스펙트럼 정보 양자화법을 인터리빙하는 방법 및 장치

(57) 요약

음성 코더에서 선 스펙트럼 정보 양자화 방법을 인터리브하기 위한 방법과 장치는 두 가지 벡터 양자화 기술로 선 스펙트럼을 양자화하는 것을 포함하고 있다. 제 1 양자화 기술은 비 이동 평균 예측 기반 기술이고 제 2 기술은 이동 평균 예측 기반 기술이다. 선 스펙트럼 정보 벡터는 제 1 기술로 양자화된 벡터이다. 제 1 기술에 대한 등가 이동평균 코드벡터들이 계산된다. 코드 벡터의 이동평균 코드북 메모리는 음성 코더에 의해 이전에 처리된 일정한 수의 프레임에 대한 등가 이동 평균 코드벡터로 갱신된다. 제 2 기술에 대한 목표 양자화 벡터가 갱신된 이동 평균 코드북 메모리를 바탕으로 계산된다. 목표 양자화 벡터는 양자화된 목표 코드 벡터를 발생하기 위해 제 2 기술을 이용해 양자화된다. 이동 평균 코드 북 메모리는 양자화된 목표 코드벡터를 이용해 갱신된다. 양자화된 선 스펙트럼 정보 벡터는 양자화된 목표 코드벡터로부터 유도된다.

대표도

도 7

특허청구의 범위

청구항 1.

음성 코더로서,

프레임을 분석하여 상기 분석에 기초한 선 스펙트럼 정보 코드벡터를 생성하도록 구성된 선형 예측 필터; 및

상기 선형 예측 필터에 연결되며 비이동-평균-예측-기반 벡터 양자화 방식을 사용하는 제 1 벡터 양자화 기술로 상기 선형 스펙트럼 정보 벡터를 벡터 양자화하도록 구성된 양자화기를 포함하며,

상기 양자화기는:

제 1 벡터 양자화 기술에 대한 등가 이동 평균 코드벡터를 계산하며,

등가 이동 평균 코드벡터로 상기 음성 코더에 의해 미리 처리된 사전결정된 수의 프레임에 대해 코드벡터의 이동 평균 코드북의 메모리를 갱신하며,

상기 갱신된 이동 평균 코드북 메모리에 기초하여 제 2 벡터 양자화 기술에 대한 목표 양자화 벡터를 계산하며,

양자화된 목표 코드벡터를 생성하기 위해 제 2 벡터 양자화 기술로 상기 목표 양자화 벡터를 벡터 양자화하며 -상기 제 2 벡터 양자화 기술은 이동-평균-예측-기반 방식을 사용하며-,

상기 이동 평균 코드북의 메모리를 상기 양자화된 목표 코드벡터로 갱신하며, 그리고

상기 양자화된 목표 코드벡터로부터 양자화된 선 스펙트럼 정보 벡터를 계산하도록 구성되는 음성 코더.

청구항 2.

제1항에 있어서, 상기 프레임은 음성 프레임인 음성 코더.

청구항 3.

제1항에 있어서, 상기 프레임은 선형 예측 잔여 프레임인 음성 코더.

청구항 4.

제1항에 있어서, 목표 양자화 벡터는 다음과 같은 식으로 계산되며:

$$U_M \equiv \left\{ U_M^n = \frac{(L_M^n - \alpha_1^n \hat{U}_{M-1}^n - \alpha_2^n \hat{U}_{M-2}^n - \dots - \alpha_P^n \hat{U}_{M-P}^n)}{\alpha_0^n}; \quad n = 0, 1, \dots, N-1 \right\}$$

여기서 L_M^n 은 상기 프레임에 대한 입력 벡터이고, $\{\hat{U}_{M-1}^n, \hat{U}_{M-2}^n, \dots, \hat{U}_{M-P}^n; \quad n = 0, 1, \dots, N-1\}$ 은 상기 프레임 바로 이전에 처리된 사전결정된 수의 프레임의 선 스펙트럼 정보 파라미터에 대응하는 코드북 입력이고, $\{\alpha_1^n, \alpha_2^n, \dots, \alpha_P^n; \quad n = 0, 1, \dots, N-1\}$ 은 $\{\alpha_0^n + \alpha_1^n + \dots + \alpha_P^n = 1; \quad n = 0, 1, \dots, N-1\}$ 의 조건을 만족하는 개별 파라미터 가중치인 음성 코더.

청구항 5.

제1항에 있어서, 양자화된 선 스펙트럼 정보 벡터는 다음과 같은 식으로 계산되며:

$$\hat{L}_M \equiv \left\{ \hat{L}_M^n = \alpha_0^n \hat{U}_M^n + \alpha_1^n \hat{U}_{M-1}^n + \dots + \alpha_P^n \hat{U}_{M-P}^n; \quad n = 0, 1, \dots, N-1 \right\}$$

여기서 $\{\hat{U}_{M-1}^n, \hat{U}_{M-2}^n, \dots, \hat{U}_{M-P}^n; \quad n = 0, 1, \dots, N-1\}$ 은 상기 프레임 바로 이전에 처리된 사전결정된 수의 프레임의 선 스펙트럼 정보 파라미터에 대응하는 코드북 엔트리이고, $\{\alpha_1^n, \alpha_2^n, \dots, \alpha_P^n; \quad n = 0, 1, \dots, N-1\}$ 은 $\{\alpha_0^n + \alpha_1^n + \dots + \alpha_P^n = 1; \quad n = 0, 1, \dots, N-1\}$ 의 조건을 만족하는 개별 파라미터 가중치인 음성 코더.

청구항 6.

제1항에 있어서, 등가 이동 평균 코드벡터는 다음과 같은 식으로 계산되며:

$$\tilde{U}_{M-K} \equiv \left\{ \tilde{U}_{M-K}^n = \frac{(L_{M-K}^n - \beta_1^n \hat{U}_{M-K-1}^n - \beta_2^n \hat{U}_{M-K-2}^n - \dots - \beta_K^n \hat{U}_{M-K-P}^n)}{\beta_0^n}; \quad n = 0, 1, \dots, N-1 \right\}$$

여기서 \hat{L}_{M-R}^n 은 양자화된 선 스펙트럼 정보 파라미터들의 벡터이고, $\{\beta_1^n, \beta_2^n, \dots, \beta_p^n; n = 0, 1, \dots, N-1\}$ 은 $\{\beta_0^n + \beta_1^n + \dots + \beta_p^n = 1; n = 0, 1, \dots, N-1\}$ 의 조건을 만족하는 개별 등가 이동 평균 코드 벡터 엘리먼트 가중치이고, 여기서 $\{\tilde{U}_{-1}, \tilde{U}_{-2}, \dots, \tilde{U}_{-p}\}$ 의 초기 조건이 설정되는 음성 코더.

청구항 7.

제1항에 있어서, 상기 음성 코더는 무선 통신 시스템의 가입자 유닛내에 존재하는 음성 코더.

청구항 8.

제 1 및 제 2 벡터 양자화 기술을 이용하여 프레임의 선 스펙트럼 정보 벡터를 벡터 양자화하는 방법 – 상기 제 1 기술은 비이동-평균-예측-기반 벡터 양자화 방식을 사용하고 제 2 기술은 이동-평균-예측-기반 벡터 양자화 방식을 사용함 –으로서,

상기 제 1 벡터 양자화 기술로 선 스펙트럼 정보 벡터를 벡터 양자화하는 단계;

상기 제 1 기술에 대한 등가 이동 평균 코드 벡터를 계산하는 단계;

상기 등가 이동 평균 코드 벡터로 음성 코더에 의해 미리 처리된 사전결정된 수의 프레임에 대한 코드 벡터의 이동 평균 코드북 메모리를 갱신하는 단계;

상기 갱신된 이동 평균 코드북 메모리에 기초하여 상기 제 2 기술에 대해 목표 양자화 벡터를 계산하는 단계;

양자화된 목표 코드 벡터를 생성하기 위해 상기 제 2 벡터 양자화 기술로 상기 목표 양자화 벡터를 벡터 양자화하는 단계;

상기 양자화된 목표 코드 벡터로 상기 이동 평균 코드북의 메모리를 갱신하는 단계; 및

상기 양자화된 목표 코드 벡터로부터 양자화된 선 스펙트럼 정보 벡터를 유도하는 단계를 포함하는 방법.

청구항 9.

제8항에 있어서, 상기 프레임은 음성 프레임인 방법.

청구항 10.

제8항에 있어서, 상기 프레임은 선형 예측 잔여 프레임인 방법.

청구항 11.

제8항에 있어서, 상기 계산 단계는 다음의 식에 따라 상기 목표 양자화 벡터를 계산하는 것을 포함하며:

$$U_M \equiv \left\{ U_M^n = \frac{(L_M^n - \alpha_1^n \hat{U}_{M-1}^n - \alpha_2^n \hat{U}_{M-2}^n - \dots - \alpha_p^n \hat{U}_{M-p}^n)}{\alpha_0^n}; n = 0, 1, \dots, N-1 \right\}$$

여기서 $\hat{\mathbf{L}}_M$ 은 상기 프레임에 대한 입력 벡터이고, $\{\hat{U}_M^n, \hat{U}_{M-1}^n, \dots, \hat{U}_{M-P}^n; n = 0, 1, \dots, N-1\}$ 은 상기 프레임 바로 이전에 처리된 사전결정된 수의 프레임의 선 스펙트럼 정보 파라미터에 대응하는 코드북 엔트리이고, $\{\alpha_0^n, \alpha_1^n, \dots, \alpha_P^n; n = 0, 1, \dots, N-1\}$ 은 $\{\alpha_0^n + \alpha_1^n + \dots + \alpha_P^n = 1; n = 0, 1, \dots, N-1\}$ 의 조건을 만족하는 개별 파라미터 가중치인 방법.

청구항 12.

제8항에 있어서, 상기 유도 단계는 다음 식에 따라 상기 양자화된 선 스펙트럼 정보 벡터를 유도하는 단계를 포함하며:

$$\hat{\mathbf{L}}_M \equiv \{\hat{L}_M^n = \alpha_0^n \hat{U}_M^n + \alpha_1^n \hat{U}_{M-1}^n + \dots + \alpha_P^n \hat{U}_{M-P}^n; n = 0, 1, \dots, N-1\}$$

여기서 $\{\hat{U}_M^n, \hat{U}_{M-1}^n, \dots, \hat{U}_{M-P}^n; n = 0, 1, \dots, N-1\}$ 은 상기 프레임 바로 이전에 처리된 사전결정된 수의 프레임의 선 스펙트럼 정보 파라미터에 대응하는 코드북 엔트리이고, $\{\alpha_0^n, \alpha_1^n, \dots, \alpha_P^n; n = 0, 1, \dots, N-1\}$ 은 $\{\alpha_0^n + \alpha_1^n + \dots + \alpha_P^n = 1; n = 0, 1, \dots, N-1\}$ 의 조건을 만족하는 개별 파라미터 가중치인 방법.

청구항 13.

제8항에 있어서, 상기 계산 단계는 다음 식에 따라 상기 등가 이동 평균 코드벡터를 계산하는 단계를 포함하며:

$$\tilde{\mathbf{U}}_{M-K} \equiv \left\{ \tilde{U}_{M-K}^n = \frac{(\hat{L}_{M-K}^n - \beta_1^n \hat{U}_{M-K-1}^n - \beta_2^n \hat{U}_{M-K-2}^n - \dots - \beta_K^n \hat{U}_{M-K-P}^n)}{\beta_0^n}; n = 0, 1, \dots, N-1 \right\}$$

여기서 $\hat{\mathbf{L}}_{M-P}$ 은 양자화된 선 스펙트럼 정보 파라미터들의 벡터이고, $\{\beta_1^n, \beta_2^n, \dots, \beta_K^n; n = 0, 1, \dots, N-1\}$ 은 $\{\beta_0^n + \beta_1^n + \dots + \beta_K^n = 1; n = 0, 1, \dots, N-1\}$ 의 조건을 만족하는 개별 등가 이동 평균 코드벡터 엘리먼트 가중치이고, $\{\tilde{\mathbf{U}}_{-1}, \tilde{\mathbf{U}}_{-2}, \dots, \tilde{\mathbf{U}}_{-P}\}$ 의 초기 조건이 설정되는 방법.

청구항 14.

음성 코더로서,

비이동-평균-예측-기반 벡터 양자화 방식을 사용하는 제 1 벡터 양자화 기술로 프레임의 선 스펙트럼 정보 벡터를 벡터 양자화하는 수단;

상기 제 1 벡터 양자화 기술에 대한 등가 이동 평균 코드벡터를 계산하는 수단;

상기 등가 이동 평균 코드벡터로 음성 코더에 의해 미리 처리된 사전결정된 수의 프레임에 대한 코드벡터의 이동 평균 코드북의 메모리를 갱신하는 수단;

상기 갱신된 이동 평균 코드북 메모리에 기초하여 이동-평균-예측-기반 벡터 양자화 방식을 사용하는 제 2 벡터 양자화 기술에 대한 목표 양자화 벡터를 계산하는 수단;

양자화된 목표 코드벡터를 생성하기 위해 상기 제 2 벡터 양자화 기술로 상기 목표 양자화 벡터를 양자화하는 수단;

상기 양자화된 목표 코드벡터로 상기 이동 평균 코드북의 메모리를 갱신하는 수단; 및

상기 양자화된 목표 코드벡터로부터 양자화된 선 스펙트럼 정보 벡터를 유도하는 수단을 포함하는 음성 코더.

청구항 15.

제14항에 있어서, 상기 프레임은 음성 프레임인 음성 코더.

청구항 16.

제14항에 있어서, 상기 프레임은 선형 예측 잔여 프레임인 음성 코더.

청구항 17.

제14항에 있어서, 상기 목표 양자화 벡터는 다음의 식으로 계산되며:

$$\mathbf{U}_M \equiv \left\{ U_M^n = \frac{(L_M^n - \alpha_1^n \hat{U}_{M-1}^n - \alpha_2^n \hat{U}_{M-2}^n - \dots - \alpha_p^n \hat{U}_{M-p}^n)}{\alpha_0^n}; \quad n = 0, 1, \dots, N-1 \right\}$$

여기서 L_M^n 은 상기 프레임에 대한 입력 벡터이고, $\{\hat{U}_{M-1}^n, \hat{U}_{M-2}^n, \dots, \hat{U}_{M-p}^n; \quad n = 0, 1, \dots, N-1\}$ 은 상기 프레임 이전에 바로 처리된 사전 결정된 수의 프레임의 선 스펙트럼 정보 파라미터에 대응하는 코드북 엔트리이고, $\{\alpha_1^n, \alpha_2^n, \dots, \alpha_p^n; \quad n = 0, 1, \dots, N-1\}$ 은 $\{\alpha_0^n + \alpha_1^n + \dots + \alpha_p^n = 1; \quad n = 0, 1, \dots, N-1\}$ 의 조건을 만족하는 개별 파라미터 가중치인 음성 코더.

청구항 18.

제14항에 있어서, 상기 양자화된 선 스펙트럼의 정보 벡터는 다음의 식으로 유도되며:

$$\hat{\mathbf{L}}_M \equiv \left\{ \hat{L}_M^n = \alpha_0^n \hat{U}_M^n + \alpha_1^n \hat{U}_{M-1}^n + \dots + \alpha_p^n \hat{U}_{M-p}^n; \quad n = 0, 1, \dots, N-1 \right\}$$

여기서 $\{\hat{U}_{M-1}^n, \hat{U}_{M-2}^n, \dots, \hat{U}_{M-p}^n; \quad n = 0, 1, \dots, N-1\}$ 은 상기 프레임 바로 이전에 처리된 사전리 결정된 수의 프레임의 선 스펙트럼 정보 파라미터에 대응하는 코드북 엔트리이고, $\{\alpha_1^n, \alpha_2^n, \dots, \alpha_p^n; \quad n = 0, 1, \dots, N-1\}$ 은 $\{\alpha_0^n + \alpha_1^n + \dots + \alpha_p^n = 1; \quad n = 0, 1, \dots, N-1\}$ 의 조건을 만족하는 개별 파라미터 가중치인 음성 코더.

청구항 19.

제14항에 있어서, 상기 등가 이동 평균 코드벡터는 다음의 식에 따라 계산되며:

$$\tilde{\mathbf{U}}_{M-K} \equiv \left\{ \tilde{U}_{M-K}^n = \frac{(\hat{L}_{M-K}^n - \beta_1^n \hat{U}_{M-K-1}^n - \beta_2^n \hat{U}_{M-K-2}^n - \dots - \beta_K^n \hat{U}_{M-K-P}^n)}{\beta_0^n}; \quad n = 0, 1, \dots, N-1 \right\}$$

여기서 \hat{L}_{M-R}^n 은 양자화된 선 스펙트럼 정보 파라미터들의 벡터이고, $\{\beta_1^n, \beta_2^n, \dots, \beta_p^n; n = 0, 1, \dots, N-1\}$ 는 $\{\beta_0^n + \beta_1^n + \dots + \beta_p^n = 1; n = 0, 1, \dots, N-1\}$ 의 조건을 만족하는 개별 등가 이동 평균 벡터 요소의 가중치이고, $\{\tilde{\mathbf{U}}_{-1}, \tilde{\mathbf{U}}_{-2}, \dots, \tilde{\mathbf{U}}_{-P}\}$ 의 초기 조건이 설정되는 음성 코더.

청구항 20.

제14항에 있어서, 상기 음성 코더는 무선 통신 시스템의 가입자 유닛에 존재하는 음성 코더.

명세서

기술분야

본 발명은 일반적으로 음성 처리 분야에 관한 것이며, 특히 음성 코더에서 선스펙트럼 정보를 양자화하는 방법과 장치에 관한 것이다.

배경기술

디지털 기술을 이용한 음성 전송은 확산되고 있으며 특히 장거리 및 디지털 무선 전화 분야에서는 더욱 그러하다. 이는 다시 재구성된 음성의 인지 품질을 유지하면서 하나의 채널을 통하여 전송되는 가장 적은 정보량을 결정하는데에 관심을 불러 일으켰다. 만약 음성이 단순한 샘플링과 디지타이징(digitizing)만으로 전송된다면 통상적인 아날로그 전화의 음질을 달성하기 위해 64kbps 정도의 초당 데이터 레이트가 요구된다. 하지만 음성 분석에 이은 적정한 코딩과 전송 그리고 수신기에서 재합성을 통해 데이터 레이트의 상당한 감소가 이루어질 수 있다.

음성을 압축하기 위한 장치는 여러 통신 분야에서 사용된다. 그 한 예가 무선 통신이다. 무선 통신 분야는 무선 전화, 무선 호출, 무선 가입자 회선, 핸드폰이나 개인휴대통신(PCS) 시스템과 같은 무선 통화, 이동 인터넷 프로토콜(IP) 통화, 그리고 위성 통신 시스템과 같은 많은 응용들이 있다. 특히 중요한 응용은 이동 전화 가입자를 위한 무선 통화이다.

작제

주파수 분할 다중 접속(FDMA), 시간 분할 다중 접속(TDMA), 그리고 코드 분할 다중 접속(CDMA)과 같이 무선 통신 시스템을 위한 여러 무선 인터페이스가 개발되고 있다. 이와 관련하여 Advanced Mobile Phone Service(AMPS), Global System for Mobile Communications(GSM), 그리고 Interim Standard 95(IS-95)와 같은 국내외 기준이 만들어졌다. 무선 전화 통신 시스템의 한 예가 코드 분할 다중 접속(CDMA) 시스템이다. IS-95 표준 그리고 이것에서 비롯한, IS-95A, ANSI J-STD-008, IS-95B, 그리고 3세대 표준인 IS-95C 와 IS-2000은 미국 전기 통신 공업 협회 (TIA)와 다른 저명한 표준화 기관에 의해 셀룰러나 PCS 전화 통신 시스템을 위한 CDMA 무선 인터페이스의 사용을 특정화하기 위해 발표되었다. IS-95 표준의 사용과 관련하여 만들어진 무선 통신 시스템의 예는 본 발명의 출원인에게 양도되고 이하 참고가 되는 미국 특허 번호 5,103,459와 4,901,307에 설명되어 있다.

인간 음성 발생 모델과 관련된 파라미터를 추출함으로써 음성을 압축하는 기술을 이용하는 장치가 음성 코더라고 불린다. 음성 코더는 입력 음성 신호를 시간 블럭, 또는 분석 프레임으로 나눈다. 음성 코더는 전형적으로 인코더와 디코더로 구성되어 있다. 인코더는 특정 관련 파라미터를 추출하기 위해 음성 프레임을 분석하고 그 파라미터를 비트 세트나 이진 데이터 패킷과 같이 이진 표현으로 양자화한다. 그 데이터 패킷은 통신 채널을 통하여 수신기와 디코더에 전달된다. 디코더는 데이터 패킷을 처리하고 파라미터를 생성하기 위해, 데이터 패킷을 비양자화하며, 비양자화된 파라미터를 이용하여 음성 프레임을 재합성한다.

음성 코더의 기능은 음성에서 본질적인 모든 자연(natural) 임여를 제거함으로써 디지털화된 음성 신호를 낮은 비트 레이트의 신호로 압축하는 것이다. 이러한 디지털 압축은 입력 음성 프레임을 파라미터 세트로 나타내고 상기 파라미터를 비트 세트로 표현하기 위해 양자화를 함으로 이루어진다. 만약 입력 음성 프레임이 N_i 개의 비트를 가지고 있고 음성 코더에 의해 생성된 데이터 패킷이 N_o 개의 비트를 가지고 있다면 음성 코더에 의해 이루어진 압축 인수는 $C_r = N_i/N_o$ 이다. 과제는 목표 압축 인수를 달성하면서 디코딩된 음성의 높은 음질을 유지하는 것이다. 음성 코더의 성능은 (1) 음성 모델 또는 위에서

설명한 분석과 합성 과정의 세트가 얼마나 잘 작동하는지와 (2)프레임 당 N_0 의 목표 비트 레이트로 파라미터 양자화 과정을 얼마나 잘 수행되는지에 달려있다. 따라서 음성 모델의 목표는 각 프레임 당 적은 세트의 파라미터로 음성 신호의 본질 또는 목표 음질을 획득하는 것이다.

아마도 음성 코더를 디자인하는데 가장 중요한 것은 음성 신호를 표현하기 위해 (벡터를 포함하는) 우수한 세트의 파라미터를 찾는 것이다. 우수한 세트의 파라미터는 인지적으로 정확한 음성 신호를 재구성하기 위해 낮은 시스템 대역폭을 요구한다. 간격, 신호 전력, 스펙트럼 포락선 (포르만트), 진폭, 위상 스펙트럼들이 음성 코딩 파라미터의 예이다.

음성 코더는 시간 영역 코더로 구성되는데 시간 영역 코더는 한번에 (전형적으로 5ms의 서브프레임과 같이) 작은 음성 단위를 인코딩하기 위해 높은 시간 분해 처리를 사용하여 시간 영역 음성 코더를 포획하고자 한다. 각각의 서브프레임에 대해, 코드북 공간으로부터 높은 정밀도 표현은 당업계에서 알려진 여러 탐색 알고리즘에 의해 찾을 수 있다. 선택적으로 음성 코더는 주파수 영역 코더로 구성될 수 있는데, 이는 파라미터 세트를 가진 입력 음성 프레임의 단기간 음성 스펙트럼을 포획하고 스펙트럼 파라미터에서 음파를 재형성하기 위해 상응하는 합성 과정을 사용한다. 파라미터 양자화기는 A.

Gersho & R. M. Gray의 "벡터 양자화와 신호 압축(1992)"이라는 책에서 설명하고 있는 공지의 양자화 기술과 일치하게 코드 벡터의 저장된 표현으로 파라미터를 표현함으로써 파라미터를 보존한다.

잘 알려진 시간 영역 음성 코더는 L. B. Rabiner & R. W. Schafer의 "음성 신호의 디지털 처리"라는 책 396-453(1978)에 설명된 Code Excited Linear Predictive(CELP) 코더이며 본 발명에서 참고로 하고 있다. CELP 코더에서 음성 신호에 있어서의 단기간 상관관계 또는 임여분은 선형예측(LP) 분석에 의해 제거되는데, 여기서 선형 예측 분석 방식은 단기간 포르만트 필터(short-term formant filter)의 계수를 판단한다. 입력 음성 프레임에 단기간 예측 필터를 적용하는 것은 LP 잔여 신호를 생성하는데, LP 잔여 신호는 장기 예측 필터 파라미터나 후속의 추계 코드 복을 가지고 좀 더 모델링되고 양자화된다. 그러므로 CELP 코딩은 시간 영역 음파의 인코딩 작업을 LP 단시간 필터 계수의 인코딩과 LP 잔여를 인코딩하는 작업으로 분리한다. 시간 영역 코딩은 정해진 레이트(즉, 각 프레임 당 같은 수의 비트 N_0 을 이용하는 것) 또는 가변 레이트(즉, 프레임 내용의 다른 종류에 대해 다른 비트 레이트)로 수행될 수 있다. 가변 레이트 코더는 목표 품질을 얻을 수 있을 정도의 코드 파라미터를 인코딩하기 위해 필요한 비트 양만을 사용하려 한다. 가변 레이트를 갖는 CELP 코더의 예는 본 발명의 출원인에게 양도되었고 본 발명에 참조된 미국 특허 번호 5,414,796에 설명되어 있다.

CELP 코더와 같은 시간 영역 코더는 전형적으로 시간 영역 음파의 정확성을 유지하기 위해 프레임 당 높은 비트 수 N_0 에 의존한다. 이러한 코더는 전형적으로 프레임 당 상대적으로 높은 비트 수 N_0 (8kbps 또는 그 이상)를 제공한다면 우수한 음질을 제공한다. 그러나 낮은 비트 레이트에서는(4kbps 또는 그 이하) 시간 영역 코더는 고품질이나 확고한 기능을 수행할 수 없는데, 이는 사용할 수 있는 비트의 수가 한정이 되어 있기 때문이다. 낮은 비트 레이트에서 제한된 코드북 공간은 종래의 시간 영역 코더의 과정 매칭 기능을 낳는데, 이는 보다 높은 레이트의 상업적 응용에는 성공적으로 전개되었다. 그러므로, 시간이 경과됨에 따른 개선에도 불구하고 낮은 비트 레이트에서 작동하는 여러 CELP 코딩 시스템은 전형적으로 잡음으로 표현되는 상당한 왜곡을 겪는다.

현재는 중간에서 낮은 비트 레이트(2.4kbps 내지 4kbps의 범위 및 그 이하)에서 동작하는 고품질 음성 코더의 개발에 많은 관심과 상업적 필요성이 고조되고 있다. 그리고 그 응용 범위는 무선 전화, 위성 통신, 인터넷 전화, 여러 멀티미디어, 그리고 음성 스트리밍 응용 분야, 음성 메일, 다른 음성 저장 시스템이다. 그리고 그러한 발전의 추진력은 패킷 손실 상황하에서도 높은 성능과 확고한 성능에 대한 요구이다. 최근 여러 음성 코딩 표준화 노력은 낮은 레이트의 음성 코딩 알고리즘에 대한 개발과 연구를 가속하는 다른 추진력이 되고 있다. 낮은 레이트의 음성 코더는 허용된 응용 대역폭 당 보다 많은 채널과 사용자를 생성하며, 적당한 채널 코딩의 추가 충과 결합한 낮은 레이트의 음성 코더는 코더 사양의 전체 비트-버짓(bit-budget)에 맞춰지고 채널 에러 상태에서도 확실한 기능을 수행한다.

낮은 비트 레이트에서 음성을 효과적으로 인코딩하는 효과적인 기술은 다중모드 코딩이다. 다중 모드 코딩 기술의 예는 1998년 12월 21일 출원된 "가변 레이트 음성 코딩"이라는 제하의 미국 특허 출원 제 09/217,341호에 기재되어 있으며 상기의 발명은 본 발명의 출원인에게 양도되었고 본 발명에 참조된다. 종래의 다중모드 코더는 다른 형태의 입력 음성 프레임에 다른 모드 또는 인코딩-디코딩 알고리즘을 적용한다. 각각의 모드 또는 인코딩-디코딩 처리는 유음 음성, 무음 음성, 전이 음성(유음과 무음 사이), 그리고 주위 잡음(무음성)과 같은 음성 부분의 특정 형태를 최적으로 표현하도록 만들어진다. 외부, 개방 루프 모드 결정 메커니즘은 입력 음성 프레임을 조사하여 어떤 모드를 프레임에 적용할 것인지를 결정한다. 상기 개방루프 모드 결정은 전형적으로 입력 프레임으로부터 수 개의 파라미터를 추출하고, 그 파라미터 특성의 일시적 스펙트럼 특성을 평가하고, 그 평가에 근거하여 모드 결정에 기초하여 수행된다.

종래의 많은 음성 코더에서, 선 스펙트럼 쌍 또는 선 스펙트럼 코사인과 같은 선 스펙트럼 정보는 코딩율을 충분히 감소시키지 않고 유음 음성 프레임을 인코딩함으로써 유음 음성의 정상상태 성질을 이용하지 않고 전송된다. 따라서 유용한 대역폭이 낭비된다. 다른 종래의 음성 코더, 다중 모드 음성 코더, 또는 낮은 비트 레이트 음성 코더에서는 유음 음성의 정상상태 성질은 매 프레임마다 이용이 된다. 따라서 비정상 상태 프레임은 저하되고 음질은 떨어진다.

본 발명은 주기적 및 비주기적 프레임 사이의 변화에 적응하는 다중 벡터 양자화 방법을 사용하는 음성 코더에 관한 것이다. 따라서, 본 발명의 일 특징에 따르면, 음성 코더는 바람직하게는 프레임을 분석하여 상기 분석에 기초한 선 스펙트럼 정보 코드벡터를 생성하도록 구성된 선형 예측 필터; 및 상기 선형 예측 필터에 연결되며 비이동-평균-예측-기반 양자화 방식을 사용하는 제 1 벡터 양자화 기술로 상기 선형 스펙트럼 정보 벡터를 벡터 양자화하도록 구성된 양자화기를 포함하며, 상기 양자화기는 제 1 기술에 대한 등가 이동 평균 코드벡터를 계산하며, 등가 이동 평균 코드벡터로 상기 음성 코더에 의해 미리 처리된 사전결정된 수의 프레임에 대해 코드벡터의 이동 평균 코드북의 메모리를 갱신하며, 상기 갱신된 이동 평균 코드북 메모리에 기초하여 제 2 기술에 대한 목표 양자화 벡터를 계산하며, 양자화된 목표 코드벡터를 생성하기 위해 제 2 벡터 양자화 기술로 상기 목표 양자화 벡터를 벡터 양자화하며 -상기 제 2 벡터 양자화 기술은 이동-평균-예측-기반 방식을 사용하며-, 상기 이동 평균 코드북의 메모리를 상기 양자화된 목표 코드벡터로 갱신하며, 및 상기 양자화된 목표 코드벡터로부터 양자화된 선 스펙트럼 정보 벡터를 계산하도록 구성된다.

각 프레임의 음성 콘텐츠의 성질에 따라 반응하는 적응성 코딩 방법을 제공하는 것이 바람직하다. 게다가 음성 신호는 보통 비정상 상태 또는 비고정 상태이므로, 음성 코딩에서 이용되는 선 스펙트럼 정보(LSI) 파라미터의 양자화의 효율은 음성의 각 프레임의 LSI파라미터가 이동-평균(MA)-예측-기반 벡터 양자화(VQ) 또는 다른 표준 VQ 방법을 이용하여 선택적으로 코딩되는 방식을 사용하여 향상시킬 수 있다. 그러한 방식은 상기 VQ의 두 가지 방법중 하나의 장점을 적절히 이용할 수 있다. 따라서 한 방법에서 다른 방법으로 바뀌는 경계에서 두 방식을 적당히 조합하여 VQ의 두 방법을 인터리빙하는 음성 코더를 제공함이 바람직하다. 따라서 주기적 프레임과 비주기적 프레임간의 변화에 적응하기 위해 다중 벡터 양자화 방법을 이용하는 음성 코더가 필요하다.

발명의 상세한 설명

본 발명은 주기적 프레임과 비주기적 프레임의 변화에 적응하기 위해 다중 벡터 양자화 방법을 사용하는 음성 코더에 관한 것이다. 따라서, 본 발명의 일 특징에 따르면, 음성 코더는 바람직하게는 프레임을 분석하여 상기 분석에 기초한 선 스펙트럼 정보 코드벡터를 생성하도록 구성된 선형 예측 필터; 및 상기 선형 예측 필터에 연결되며 비이동-평균-예측-기반 양자화 방식을 사용하는 제 1 벡터 양자화 기술로 상기 선형 스펙트럼 정보 벡터를 벡터 양자화하도록 구성된 양자화기를 포함하며, 상기 양자화기는 제 1 기술에 대한 등가 이동 평균 코드벡터를 계산하며, 등가 이동 평균 코드벡터로 상기 음성 코더에 의해 미리 처리된 사전결정된 수의 프레임에 대해 코드벡터의 이동 평균 코드북의 메모리를 갱신하며, 상기 갱신된 이동 평균 코드북 메모리에 기초하여 제 2 기술에 대한 목표 양자화 벡터를 계산하며, 양자화된 목표 코드벡터를 생성하기 위해 제 2 벡터 양자화 기술로 상기 목표 양자화 벡터를 벡터 양자화하며 -상기 제 2 벡터 양자화 기술은 이동-평균-예측-기반 방식을 사용하며-, 상기 이동 평균 코드북의 메모리를 상기 양자화된 목표 코드벡터로 갱신하며, 및 상기 양자화된 목표 코드벡터로부터 양자화된 선 스펙트럼 정보 벡터를 계산하도록 구성된다.

본 발명의 또 다른 특징에 따르면, 제 1 및 제 2 벡터 양자화 기술을 이용하여 프레임의 선 스펙트럼 정보 벡터를 벡터 양자화하는 방법 - 상기 제 1 기술은 비이동-평균-예측-기반 벡터 양자화 방식을 사용하고 제 2 기술은 이동-평균-예측-기반 벡터 양자화 방식을 사용함 -은 바람직하게는 상기 제 1 벡터 양자화 기술로 선 스펙트럼 정보 벡터를 벡터 양자화하는 단계; 상기 제 1 기술에 대한 등가 이동 평균 코드벡터를 계산하는 단계; 상기 등가 이동 평균 코드북 벡터로 음성 코더에 의해 미리 처리된 사전결정된 수의 프레임에 대한 코드 벡터의 이동 평균 코드북 메모리를 갱신하는 단계; 상기 갱신된 이동 평균 코드북 메모리에 기초하여 상기 제 2 기술에 대해 목표 양자화 벡터를 계산하는 단계; 양자화된 목표 코드벡터를 생성하기 위해 상기 제 2 벡터 양자화 기술로 상기 목표 양자화 벡터를 벡터 양자화하는 단계; 상기 양자화된 목표 코드벡터로 상기 이동 평균 코드북의 메모리를 갱신하는 단계; 및 상기 양자화된 목표 코드벡터로부터 양자화된 선 스펙트럼 정보 벡터를 유도하는 단계를 포함한다.

본 발명의 또 다른 특징에 따르면, 음성 코더는 바람직하게는 비이동-평균-예측-기반 양자화 방식을 사용하는 제 1 벡터 양자화 기술로 프레임의 선 스펙트럼 정보 벡터를 벡터 양자화하는 수단; 상기 제 1 기술에 대해 등가 평균이동 코드벡터를 계산하는 수단; 상기 등가 이동 평균 코드 벡터로 음성 코더에 의해 미리 처리된 사전결정된 수의 프레임에 대한 코드 벡터의 이동 평균 코드북의 메모리를 갱신하는 수단; 상기 갱신된 이동 평균 코드북 메모리에 기초하여 제 2 기술에 대한 목표 양자화 벡터를 계산하는 수단; 양자화된 목표 코드벡터를 생성하기 위해 상기 제 2 벡터 양자화 기술로 상기 목표 양자화 벡터를 양자화하는 수단; 상기 양자화된 목표 코드벡터로 상기 이동 평균 코드북의 메모리를 갱신하는 수단; 및 상기 양자화된 목표 코드벡터로부터 양자화된 선 스펙트럼 정보 벡터를 유도하는 수단을 포함한다.

실시예

여기서 이하 설명하는 실시예는 무선 인터페이스에서 CDMA를 사용하기 위해 구성된 무선 전화 통신 시스템에 있다. 그럼에도 불구하고 당업자에게는 부표본화 방법과 현재의 발명의 특징을 드러내는 장치는 공지의 기술의 광 범위에 존재할 수 있는 것으로 이해가 될 것이다.

도면 1에서와 같이, CDMA 무선 통신 시스템은 보통 다수의 이동전화 가입자 유닛(10), 다수의 기지국(12), 기지국 제어기(BSC)(14), 및 이동전화 교환국(MSC)(16)를 포함한다. MSC(16)는 종래의 공중전화교환망(PSTN)(18)과 인터페이싱하도록 구성된다. 또한 MSC(16)는 BSC(14)와도 인터페이싱하도록 구성된다. BSC(14)는 기지국(12)과 백홀 라인(backhaul line)을 통해 연결된다. 백홀 라인은 예를 들면, E1/T1, ATM, IP, PPP, 프레임 릴레이, HDSL, ADSL 또는 xDSL을 포함하는 여러 공지된 인터페이스를 지원하도록 구성된다. 시스템내에 두 개 이상의 BSC(14)가 존재해야 한다는 것을 알 수 있을 것이다. 각각의 기지국(12)은 바람직하게 적어도 하나의 섹터(미도시)를 가지고 있어야 하고, 각각의 섹터는 무지향성 안테나 또는 기지국(12)으로부터 방사적으로 이격한 특정 방향의 안테나로 이루어져 있다. 선택적으로 각각의 섹터는 다이버시티 수신을 위해 2개의 안테나로 이루어질 수 있다. 각각의 기지국(12)은 바람직하게 다수의 주파수 할당을 지원할 수 있도록 설계된다. 섹터와 주파수 할당의 인터섹션은 CDMA 채널이라고도 한다. 기지국(12)은 또한 기지국 트랜시버 서브시스템(BTS)(12)이라고도 알려져 있다. 선택적으로 "기지국"은 산업상 BSC(14) 및 한 개 이상의 BTS(12)를 조합적으로 지칭하는데 사용된다. BTS(12)는 "셀 사이트(12)"로 표시될 수 있다. 선택적으로, 주어진 BTS(12)의 개별 섹터는 셀 사이트라고도 한다. 이동전화 가입자 유닛(10)은 전형적으로 셀룰러 또는 PCS 전화(10)이다. 시스템은 바람직하게 IS-95 표준에 따른 사용을 위해 구성된다.

셀룰러 전화 시스템의 전형적인 동작 동안, 기지국(12)은 이동전화 유닛(10) 세트로부터 역방향 링크 신호 세트를 수신한다. 이동전화 유닛(10)은 전화호출 또는 다른 통신을 수행한다. 주어진 기지국(12)에 의해 수신된 각각의 역방향 링크 신호는 그 기지국(12) 내에서 처리된다. 그로 인해 생성된 데이터는 BSC(14)로 전송된다. BSC(14)는 셀 자원 할당 및 기지국(12)들 간에 소프트 핸드오프를 조정하는 것을 포함하는 이동성 관리 기능을 제공한다. BSC(14)는 또한 수신된 데이터를 MSC(16)로 라우팅하는데, MSC(16)는 PSTN(18)과의 인터페이스를 위한 추가의 라우팅 서비스를 제공한다. 유사하게, PSTN(18)은 MSC(16)와 인터페이싱하고, MSC(16)는 BSC(14)와 인터페이싱하며, 이는 다시 순방향 링크 신호 세트를 이동전화 유닛(10) 세트로 전송하기 위해 기지국(12)을 제어한다.

도면 2에서, 제 1 인코더(100)는 디지털화된 음성 샘플 $s(n)$ 을 수신하고 전송 매체(102) 또는 통신 채널(102)를 통해 제 1 디코더(104)로 전송하기 위해 샘플 $s(n)$ 을 인코딩한다. 디코더(104)는 인코딩된 음성 샘플을 디코딩하고 출력 음성 신호 $s_{SYNTH}(n)$ 을 생성한다. 반대 방향으로 전송을 위해, 제 2 인코더(106)는 디지털화된 음성 샘플 $s(n)$ 을 인코딩하고, 이는 통신 채널(108)로 전달된다. 제 2 디코더(110)는 인코딩된 음성 샘플을 수신하여 디코딩하며, 합성된 출력 음성 신호 $s_{SYNTH}(n)$ 를 생성한다.

음성 샘플 $s(n)$ 은 예를 들면, 펠스 코드 변조(PCM), commanded μ -law, 또는 A-law를 포함하는 공지된 여러 방법에 따라 디지털화되고 양자화된 음성 신호를 나타낸다. 당업계에서 알려진 것과 같이, 음성 샘플 $s(n)$ 은 각각의 프레임이 사전결정된 수의 디지털화된 음성 샘플 $s(n)$ 로 구성된 입력 데이터의 프레임으로 구성된다. 일 실시예에서 8kHz의 샘플링 레이트가 사용되고, 각각의 20ms의 프레임은 160개 샘플을 포함한다. 후술될 실시예에서 데이터 전송 레이트는 바람직하게 13.2kbps(풀 레이트), 6.2kbps(1/2 레이트), 2.6kbps(1/4 레이트), 1kbps(1/8 레이트)까지 프레임별 기반으로 가변할 수 있다. 데이터 전송 레이트의 변화는 상대적으로 적은 음성 정보를 가지고 있는 프레임에서는 낮은 비트 레이트를 사용할 수 있기 때문이다. 당업자라면 알 수 있듯이, 다른 샘플링 레이트, 프레임 크기, 및 데이터 전송 레이트가 사용될 수 있다.

제 1 인코더(100)와 제 2 디코더(110)는 함께 제 1 음성 코더 또는 음성 코덱을 포함한다. 음성 코더는 도면 1을 참조하여 상술된 가입자 유닛, BTS 또는 BSC를 포함하는 음성 신호 전달을 임의의 통신 장비에서도 사용될 수 있다. 유사하게, 제 2 인코더(106)와 제 1 디코더(104)는 함께 제 2 음성 코더를 포함한다. 당업자라면 음성 코더가 디지털 신호 처리기(DSP), 응용주문형 직접회로(ASIC), 이산 게이트 로직, 펌웨어 또는 통상적인 프로그램 가능 소프트웨어 모듈 및 마이크로프로세서로 이루어질 수 있음을 알 수 있을 것이다. 소프트웨어 모듈은 RAM 메모리, 플래시 메모리, 레지스터, 또는 당업계에서 알려진 다른 형태의 기록 가능 메모리에 상주한다. 선택적으로 통상적인 프로세서, 제어기, 또는 상태 기계가 마이크로프로세서를 대신할 수 있다. 음성 코딩을 위해 특별히 제작된 ASIC의 예는 본 발명의 출원인에게 양도되고 본 출원에 전체적으로 참조되는 미국 특허 제5,727,123과 1994년 2월 16일에 출원되어 "보코더 ASIC"라는 제하의 미국 특허출원 제08/197,417에 설명되어 있다.

도면 3에서, 음성 코더에 사용되는 인코더(200)는 모드 결정 모듈(202), 피치 추정 모듈(204), LP 분석 모듈(206), LP 분석 필터(208), LP 양자화 모듈(210) 및 잔여 양자화 모듈(212)을 포함한다. 입력 음성 프레임 $s(n)$ 은 모드 결정 모듈(202), 피치 추정 모듈(204), LP 분석 모듈(206) 및 LP 분석 필터(208)에 제공된다. 모드 결정 모듈(202)은 각각의 입력 음성 프레임 $s(n)$ 의 다른 특징 중에도 주기성, 에너지, 신호-대-잡음 비(SNR) 부호 변환점 레이트에 기초하여 모드 인덱스 I_M 과 모드 M을 생성한다. 주기성에 따라 음성 프레임을 분류하는 여러 방법은 본 발명의 출원인에게 양도되어 있고 본 출원에 전체적으로 참조되는 미국 특허 제 5,911,128에 설명되어 있다. 그러한 방법은 또한 통신 산업 협회 산업 규정 표준 TIA/EIA IS-127과 TIA/EIA IS-733에 통합된다. 모드 결정 방식의 예는 언급된 미국 특허출원 제 09/217,341에 설명되어 있다.

피치 추정 모듈(204)은 각각의 입력 음성 프레임 $s(n)$ 에 기초하여 피치 인덱스 I_p 와 래그(lag) 값 P_0 을 생성한다. LP 분석 모듈(206)은 LP 파라미터 a 를 생성하기 위해 각각의 입력 음성프레임 $s(n)$ 에 대해 선형 예측 분석을 수행한다. LP 파라미터 a 는 LP 양자화 모듈(210)에 제공된다. LP양자화 모듈 (210)은 모드 M을 수신하고 그것에 따라 모드-의존 방법으로 양자화 과정을 수행한다. LP 양자화 모듈(210)은 LP 인덱스 I_{LP} 와 양자화된 LP 파라미터 \hat{a} 을 생성한다. LP 분석필터 (208)는 입력 음성 프레임 $s(n)$ 에 부가하여 양자화된 LP 파라미터 \hat{a} 를 수신한다. LP 분석 필터(208)는 양자화된 선형 예측 파라미터 \hat{a} 에 근거하여 입력 음성 프레임 $s(n)$ 과 재구성된 음성 사이의 에러를 나타내는 LP 잔여 신호 $R[n]$ 를 생성한다. LP 잔여 신호 $R[n]$, 모드 M, 및 양자화된 LP 파라미터 \hat{a} 는 잔여 양자화 모듈(212)에 제공된다. 이러한 값에 의해 잔여 양자화 모듈(212)은 잔여 인덱스 I_R 과 양자화된 잔여 신호 $\hat{R}[n]$ 을 생성한다.

도면 4에서, 음성 코더에서 사용될 수 있는 디코더(300)는 LP 파라미터 디코딩 모듈(302), 잔여 디코딩 모듈(304), 모드 디코딩 모듈(306) 및 LP 합성 필터(308)를 포함한다. 모드 디코딩 모듈(306)은 모드 인덱스 I_M 을 수신하고 디코딩을 하여 그로부터 모드 M을 생성한다. LP 파라미터 디코딩 모듈(302)은 모드 M 및 LP 인덱스 I_{LP} 를 수신한다. LP 파라미터 디코딩 모듈(302)은 양자화된 LP 파라미터 \hat{a} 을 생성하기 위해 수신된 값을 디코딩한다. 잔여 디코딩 모듈(304)은 잔여 인덱스 I_R , 피치 인덱스 I_p , 및 모드 인덱스 I_M 을 수신한다. 잔여 디코딩 모듈(304)은 양자화된 잔여 신호 $\hat{R}[n]$ 을 생성하기 위해 수신된 값을 디코딩한다. 양자화된 잔여 신호 $\hat{R}[n]$ 와 양자화된 LP 파라미터 \hat{a} 는 디코딩된 출력 음성 신호 $s^*[n]$ 을 합성하는 LP 합성 필터(308)에 제공된다.

도면 3의 인코더(200)와 도면 4의 디코더(300)의 여러 가지 모듈의 작동과 구현은 언급된 미국 특허 제 5,414,796 와 L. B. 라비너와 R. W. 스카퍼의 "음성 신호의 디지털 처리(1978)"의 396-453에 설명되어 있다.

도면 5의 흐름도에서 설명된 것과 같이, 일 실시예에서의 음성 코더는 전송을 위해 음성 샘플을 처리함에 있어서의 일련의 단계를 따라간다. 단계(400)에서 음성 코더는 연속적인 프레임에서 음성 신호의 디지털 샘플을 수신한다. 주어진 샘플을 수신한다면, 음성 코더는 단계(402)로 진행한다. 단계(402)에서 음성 코더는 프레임의 에너지를 검출한다. 에너지는 프레임의 음성 활동성의 측정치이다. 음성 검출은 디지털화된 음성 샘플의 진폭의 제곱을 합하고, 그 결과로 생성된 에너지를 임계값과 비교함으로 수행된다. 일 실시예에서 임계값은 주위 잡음의 크기 변화에 따라 달라진다. 예시적인 가변 임계 음성 활동성 검출기는 언급된 미국 특허 제 5,414,796에 설명되어 있다. 어떤 무음 음성은 오류로 주위 잡음으로 인코딩되는 매우 낮은 에너지 샘플이 될 수 있다. 이것을 방지하기 위해, 언급된 미국 특허 제 5,414,796에 설명된 것과 같이 낮은 에너지 샘플의 스펙트럼 텔트를 무음 음성과 주위 잡음을 구별하기 위해 사용한다.

프레임의 에너지를 검출한 후에, 음성 코더는 단계(404)로 진행한다. 단계(404)에서, 음성 코더는 검출된 프레임 에너지가 음성 정보를 가지고 있는 프레임으로 분류하기에 충분한 에너지를 가지고 있는지 아닌지를 결정한다. 만약 검출된 프레임 에너지가 미리 결정된 임계 레벨 이하로 떨어지면, 음성 코더는 단계(406)로 진행한다. 단계(406)에서, 음성 코더는 주위 잡음(즉, 무음 또는 침묵)으로 프레임을 인코딩한다. 일 실시예에서 주위 잡음 프레임은 1/8 레이트 또는 1kbps로 인코딩된다. 만약 단계(404)에서 검출된 프레임 에너지가 미리 결정된 임계 레벨을 충족하거나 초과하는 경우에는 그 프레임은 음성으로 분류되고 음성 코더는 단계(408)로 진행한다.

단계(408)에서, 음성 코더는 프레임의 주기성을 조사하여 그 프레임이 무음 음성인지 아닌지를 판단한다. 공지의 여러 주기성 결정은 부호 변환점을 이용하거나 정규화된 자기 상관 함수(NACF)를 이용하는 것을 포함한다. 특히 주기성을 검출하기 위해 부호 변환점과 자기 상관함수를 이용하는 것은 언급된 미국 특허 제 5,911,128과 미국 출원 제 09/217,341에 설명되어 있다. 게다가 유음 음성과 무음 음성을 구별하기 위해 위에서 설명한 방법은 통신 산업 협회 산업 잠정 표준 TIA/EIA IS-127과 TIA/EIA IS-733에 포함되어 있다. 만약 단계(408)에서 프레임이 무음 음성으로 결정되면 음성 코더는 단계(410)로 진행한다. 단계(410)에서, 음성 코더는 무음 음성을 프레임으로 인코딩한다. 일 실시예에서 무음 음성 프레임은 1/4 레이트 또는 2.6kbps로 인코딩된다. 만약 단계(408)에서, 프레임이 무음 음성이 아닌 것으로 결정되면, 음성 코더는 단계(412)로 진행한다.

단계(412)에서, 음성 코더는 예를 들면, 언급된 미국 특허 제 5,911,128에서 설명하고 있는 바와 같이, 공지된 주기성 결정법을 이용하여 프레임이 전이 음성인지 아닌지를 판단한다. 만약 그 프레임이 전이 음성이라고 판단되면 음성 코더는 단계(414)로 진행한다. 단계(414)에서, 그 프레임은 전이 음성(무음 음성으로부터 유성 음성으로의 전이)으로 인코딩된다. 일 실시예에서, 전이 음성 프레임은 "전이 음성 프레임의 멀티펄스 보간 코딩"이라는 제하의 1999년 5월 7일에 출원된 미국 특허출원 제 09/307,294(미국 특허 제 6,260,017호)에 설명된 멀티펄스 보간 방법에 따라 인코딩되며 상기 발명은 본 발명의 출원인에게 양도되었고 본 발명에 전체적으로 참조된다. 다른 실시예에서, 전이 음성 프레임은 풀 레이트 또는 13.2 kbps로 인코딩된다.

만약 단계(412)에서 음성 코더가 프레임이 전이 음성이 아니라고 결정하면, 음성 코더는 단계(416)로 진행한다. 단계(416)에서, 음성 코더는 프레임을 유음 음성으로 인코딩한다. 일 실시예에서, 유성 음성 프레임은 1/2 레이트 또는 6.2kbps로 인코딩된다. 또한 유음 음성 프레임을 풀 레이트 또는 13.2kbps(또는 8k CELP인코더에서는 풀 레이트, 8kbps)로 인코딩할 수 있다. 그러나 당업자라면 1/2 레이트로 유음 음성 프레임을 인코딩하는 것은 코더가 유음 프레임의 정상상태의 특성을 이용함으로써 상당한 대역폭을 절약할 수 있도록 한다는 것을 알 수 있을 것이다. 게다가 유음 음성을 인코딩하는데 사용된 레이트에 관계없이, 유음 음성은 바람직하게 지난 프레임의 정보를 이용하여 인코딩되며 이에 따라 예측한 대로 인코딩된다고 말할 수 있다.

당업자라면 음성 신호 또는 대응하는 LP 잔여가 도면 5에서 보여지는 단계들을 따라 인코딩된다는 것을 알 수 있을 것이다. 잡음, 무음 음성, 전이 음성, 및 유음 음성의 파형 특성은 도면 6A의 그래프에 시간의 함수로서 도시된다. 잡음, 무음 음성, 전이 음성 및 유음 LP 잔여의 파형 특성은 도면 6B의 그래프에 시간의 함수로 도시된다.

일 실시예에서, 음성 코더는 선 스펙트럼 정보(LSI) 벡터 양자화(VQ)의 두 방법들을 인터리빙하기 위해 도면 7의 흐름도에 도시된 알고리즘 단계들을 수행한다. 음성 코더는 바람직하게는 음성 코더가 LSI VQ의 두 방법을 인터리빙할 수 있도록 하는 비-이동-평균(non-MA(moving-average))-예측-기반 LSI VQ에 대한 등가 이동-평균(MA) 코드북 벡터를 계산한다. MA 예측-기반 방식에서, MA는 이전에 처리된 프레임 수 P에 대해 계산되는데, MA는 이하 설명하는 것과 같이 파라미터 가중치에 각각의 벡터 코드북 엔트리를 곱하여 계산한다. 또한 이하 설명하는 것과 같이, MA는 목표 양자화 벡터를 생성하기 위해 LSI 파라미터의 입력 벡터로부터 감산된다. 당업자라면 비MA 예측-기반 VQ 방법이 MA 예측-기반 VQ 방식을 사용하지 않는 VQ 방법이라는 것을 알 수 있을 것이다.

LSI 파라미터는 전형적으로 프레임간 MA 예측으로 VQ를 이용함으로써 또는 예를 들면, 분할 VQ, 멀티스테이지 VQ (MSVQ), 스위칭 예언 VQ(SPVQ) 또는 이들 중 몇몇 혹은 전체의 조합을 사용하는 것과 같은 임의의 다른 표준 비MA-예측 기반 VQ 방법을 사용함으로써 양자화된다. 도면 7을 참고로 하여 설명한 일 실시예에서, 그 구조는 위에서 언급한 VQ 방법과 MA 예측-기반 VQ 방법을 혼합한 방식이 사용된다. 이는 MA 예측-기반 VQ 방법이 자연 상태에서(도면 6A-6B에서 정적 음성에 대해 도시된 바와 같은 신호를 나타냄) 정상상태 또는 정적인 음성 프레임의 장점을 가장 잘 활용하는데 사용되는 반면 비MA 예측-기반 VQ 방법은 자연 상태에서(도면 6A-6B에서 무음 프레임과 전이 프레임에 대해 도시된 바와 같은 신호를 나타냄) 비정상상태 또는 비정적인 음성 프레임의 장점을 가장 잘 활용하는데 사용되기 때문에 바람직하다.

N차원의 LSI 파라미터를 양자화하기 위한 비MA 예측-기반 VQ 방식에서 M 번째 프레임에 대한 입력벡터

$\mathbf{L}_M \equiv \{L_M^n; n = 0, 1, \dots, N-1\}$ 은 직접적으로 양자화를 위한 목표로 이용되고 위에서 언급한 표준 VQ 기술을 이용하여 벡터 $\hat{\mathbf{L}}_M \equiv \{\hat{L}_M^n; n = 0, 1, \dots, N-1\}$ 로 양자화된다.

프레임간 MA 예측 방식에서, 양자화를 위한 목표는 다음과 같이 계산된다.

$$\mathbf{U}_M \equiv \left\{ U_M^n = \frac{(L_M^n - \alpha_1^n \hat{U}_{M-1}^n - \alpha_2^n \hat{U}_{M-2}^n - \dots - \alpha_P^n \hat{U}_{M-P}^n)}{\alpha_0^n}; \quad n = 0, 1, \dots, N-1 \right\} \quad (1)$$

여기서 $\{\hat{U}_{M-1}^n, \hat{U}_{M-2}^n, \dots, \hat{U}_{M-P}^n; \quad n = 0, 1, \dots, N-1\}$ 은 프레임 M 바로 이전에 있는 P개의 프레임의 LSI 파라미터에 대응하는 코드북 엔트리이고, $\{\alpha_1^n, \alpha_2^n, \dots, \alpha_P^n; \quad n = 0, 1, \dots, N-1\}$ 은 $\{\alpha_0^n + \alpha_1^n + \dots + \alpha_P^n = 1; \quad n = 0, 1, \dots, N-1\}$ 를 만족하는 개별 가중치이다. 목표 양자화 U_M 은 위에서 언급된 VQ 기술을 이용하여 $\hat{\mathbf{U}}_M$ 으로 양자화한다. 양자화된 LSI 벡터는 다음과 같이 계산된다.

$$\hat{\mathbf{L}}_M \equiv \{\hat{L}_M^n = \alpha_0^n \hat{U}_M^n + \alpha_1^n \hat{U}_{M-1}^n + \dots + \alpha_P^n \hat{U}_{M-P}^n; \quad n = 0, 1, \dots, N-1\} \quad (2)$$

MA 예측 방식은 이전 P개의 프레임의 코드북 엔트리 $\{\hat{U}_{M-1}, \hat{U}_{M-2}, \dots, \hat{U}_{M-P}\}$ 의 이전 값의 존재를 요구한다. 코드북 엔트리가 MA 방식을 사용하여 양자화된 (이전 P개의 프레임중) 이들 프레임에 자동적으로 사용가능한 반면, 이전 P 프레임의 나머지는 비MA 예측-기반 VQ 방법을 이용하여 양자화할 수 있었으며, 해당 코드북 엔트리 $(\hat{\mathbf{U}})$ 는 이들 프레임에 대해 직접적으로 이용할 수 없다. 이것이 VQ의 두 가지 방법들을 인터리빙하거나 혼합하기 힘들게 한다.

도면 7을 참고로 한 실시예에서, 다음 식은 $K \in \{1, 2, \dots, P\}$ 경우 코드북 엔트리 $\hat{\mathbf{U}}_{M-K}$ 가 명백하게 이용가능하지 않을 경우 코드북 엔트리 $\hat{\mathbf{U}}_{M-K}$ 의 추정치, $\tilde{\hat{\mathbf{U}}}_{M-K}$ 을 계산하는데 유용하다.

$$\tilde{\hat{\mathbf{U}}}_{M-K} \equiv \left\{ \tilde{\hat{U}}_{M-K}^n = \frac{(\hat{L}_{M-K}^n - \beta_1^n \hat{U}_{M-K-1}^n - \beta_2^n \hat{U}_{M-K-2}^n - \dots - \beta_R^n \hat{U}_{M-K-P}^n)}{\beta_0^n}; \quad n = 0, 1, \dots, N-1 \right\} \quad (3)$$

이 식에서 $\{\beta_1^n, \beta_2^n, \dots, \beta_P^n; \quad n = 0, 1, \dots, N-1\}$ 은 $\{\beta_0^n + \beta_1^n + \dots + \beta_P^n = 1; \quad n = 0, 1, \dots, N-1\}$ 의 조건을 만족하는 각각의 가중치이고 $\{\tilde{\hat{U}}_{-1}, \tilde{\hat{U}}_{-2}, \dots, \tilde{\hat{U}}_{-P}\}$ 은 초기 조건이다. 예시적인 초기 조건은 $\{\tilde{\hat{U}}_{-1} = \tilde{\hat{U}}_{-2} = \dots = \tilde{\hat{U}}_{-P} = L^B\}$ 와 같으며 여기서 L^B 는 LSI의 바이어스 값이다. 다음은 가중치 세트의 한 예이다.

$$\left. \begin{cases} \beta_0^n = 1; \\ \beta_1^n = \dots = \beta_P^n = 0; \end{cases} \right\} \quad n = 0, 1, \dots, N-1$$

도면 7의 흐름도의 단계(500)에서, 음성 코더는 MA 예측-기반 VQ 기술을 이용하여 입력 LSI 벡터 L_M 을 양자화할 것인지 결정한다. 이러한 결정은 바람직하게는 프레임의 음성 콘텐츠를 바탕으로 이루어진다. 예를 들어, 정적 유음 프레임에 대한 LSI 파라미터는 MA 예측-기반 VQ 방법을 가장 잘 활용하여 양자화되는 반면, 무음 프레임 또는 전이 프레임에 대한 LSI 파라미터는 비MA 예측-기반 VQ 방법을 가장 잘 활용하여 양자화된다. 만약 음성 코더가 MA 예측-기반 VQ 기술로 입력 LSI 벡터 L_M 을 양자화하기로 결정하면, 음성 코더는 단계(502)로 진행한다. 한편 반대로 음성 코더가 MA 예측-기반 VQ 기술로 입력 LSI 벡터 L_M 을 양자화하지 않기로 결정하면, 음성 코더는 단계(504)로 진행한다.

단계(502)에서 음성 코더는 위의 식 (1)에 따른 양자화를 위하여 목표 U_M 을 계산한다. 그리고 나서 음성 코더는 단계(506)로 진행한다. 단계(506)에서 음성 코더는 당업계에서 알려진 여러 일반적인 VQ 기술중 하나를 이용하여 목표 U_M 을 양자화한다. 그 다음 음성 코더는 단계(508)로 진행한다. 단계(508)에서 음성 코더는 위의 식 (2)에 따라 양자화된 목표 \hat{U}_M 으로부터 양자화된 LSI 파라미터 벡터 \hat{L}_M 을 계산한다.

단계(504)에서 음성 코더는 당업계에서 잘 알려진 여러 비MA 예측-기반 VQ 기술중 하나에 따라 목표 L_M 을 양자화한다. (당업자가 이해할 수 있듯이 비MA 예측-기반 VQ 기술에서 양자화를 위한 벡터는 U_M 이 아니고 L_M 이다.) 그 다음 음성 코더는 단계(510)로 진행한다. 단계(510)에서 음성 코더는 위의 식 (3)에 따라 양자화된 LSI 벡터 \hat{L}_M 으로부터 등가 MA 코드벡터 \tilde{U}_M 을 계산한다.

단계(512)에서 음성 코더는 이전 P개의 프레임의 MA 코드북 벡터의 메모리를 갱신하기 위해 단계(506)에서 구한 양자화된 \hat{U}_M 과 단계(510)에서 구한 등가 MA 코드벡터 \tilde{U}_M 을 이용한다. 이전 P개 프레임의 MA 코드북 벡터의 갱신된 메모리는 단계(502)에서 다음 프레임에 대한 입력 LSI 벡터 L_{M+1} 에 대한 양자화를 위한 목표 U_M 을 계산한다.

따라서 음성 코더에서 선 스펙트럼 정보 양자화 방법들의 인터리빙을 위한 새로운 방법과 장치가 설명되었다. 당업자는 여러 로직 블럭과 알고리즘 과정으로 설명한 실시에는 디지털 신호 처리기(DSP), 응용주문형 직접회로(ASIC), 이산 게이트 또는 트랜지스터 로직, 저항 그리고 피포(FIFO)와 같은 이산 하드웨어 부품, 펌웨어 명령을 수행하는 처리기, 또는 종래의 프로그램할 수 있는 모듈과 처리기로 구성되어 기능을 수행한다. 처리기는 바람직하게는 마이크로프로세서이지만 선택적으로 종래의 처리기, 제어기, 마이크로콘트롤러 또는 상태기계도 가능하다. 소프트웨어 모듈은 RAM메모리, 플래시 메모리, 저항 또는 당업계에서 알려진 다른 저장성 매체에 존재할 수 있다. 당업자는 위의 설명에서 기술한 데이터, 지시, 명령, 정보, 신호, 비트, 심벌, 그리고 칩은 전압, 전류, 전자기파, 자장, 또는 자기 입자, 광학장 또는 입자, 또는 이들의 임의의 조합일 수 있을 것이다.

바람직한 본 발명의 실시예는 이상 설명되었다. 그러나 당업자의 기술 수준에서는 여기서 기술한 본 발명을 벗어나지 않는 범위에서 수많은 다른 실시가 가능하다는 것은 자명하다. 그로므로, 본 발명은 다음 청구항에 따른 것을 제외하고는 제한되지 않는다.

도면의 간단한 설명

도면 1은 무선 전화 시스템의 블록도이다.

도면 2는 음성 코더에 의해 각각의 말단에서 착신되는 통신 채널의 블록도이다.

도면 3은 인코더의 블록도이다.

도면 4는 디코더의 블록도이다.

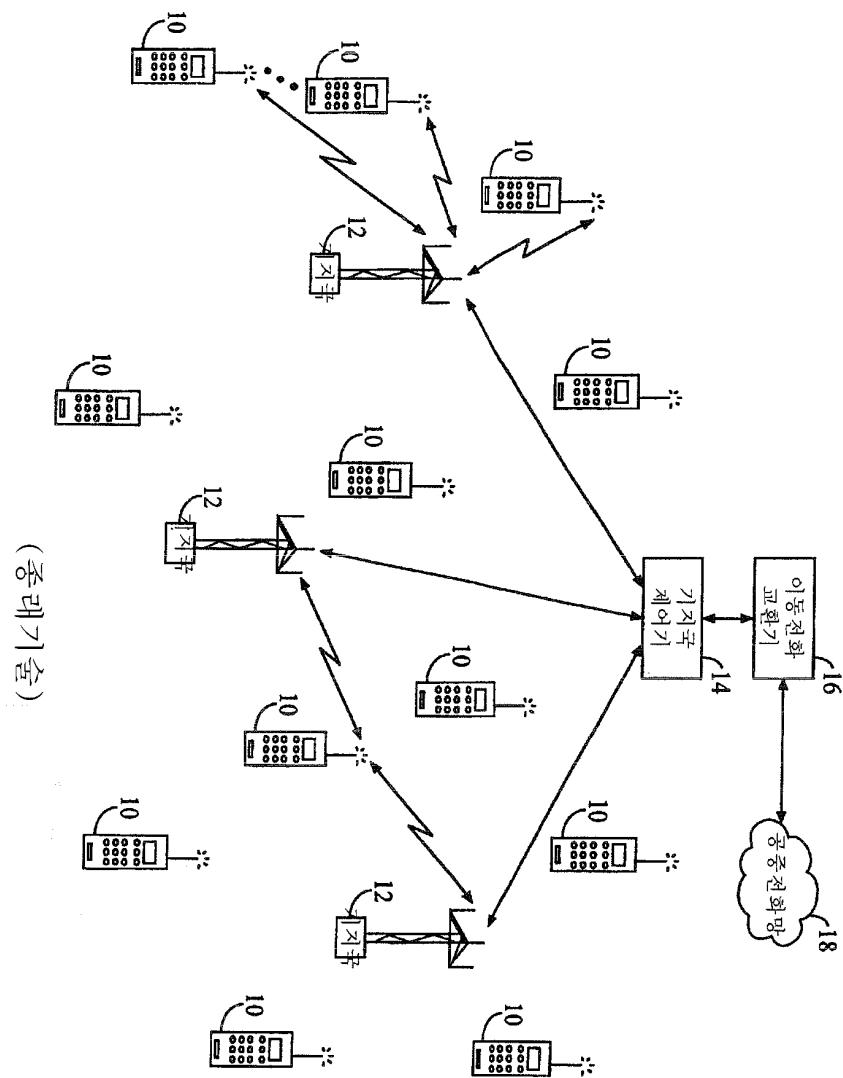
도면 5는 음성 코딩 결정 과정을 나타내고 있는 흐름도이다.

도면 6A는 시간에 대한 음성 신호 진폭의 그래프이고, 도면 6B는 시간에 대한 선형 예측(LP) 잔여 진폭의 그래프이다.

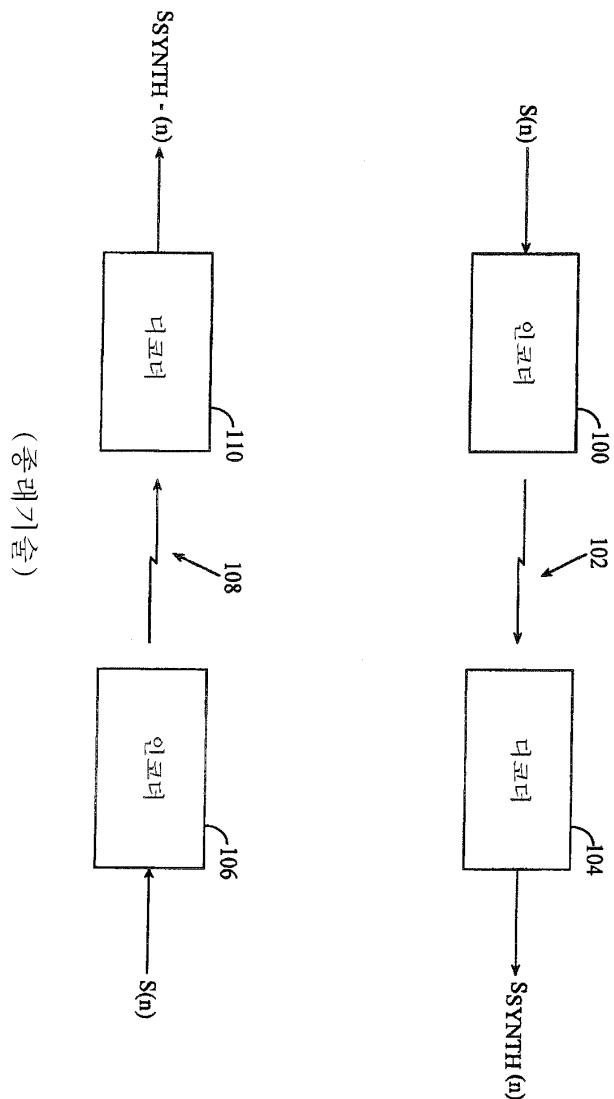
도면 7은 선 스펙트럼 정보(LSI) 벡터 양자화(VQ)의 두 방법을 인터리빙하기 위해 음성 코더에 의해 수행되는 방법 과정을 나타내고 있는 흐름도이다.

도면

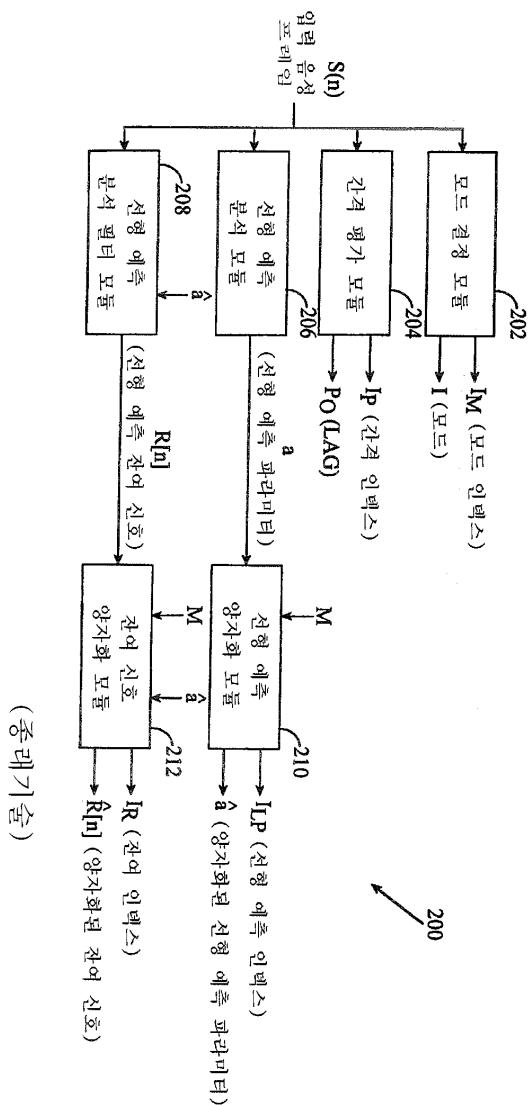
도면1



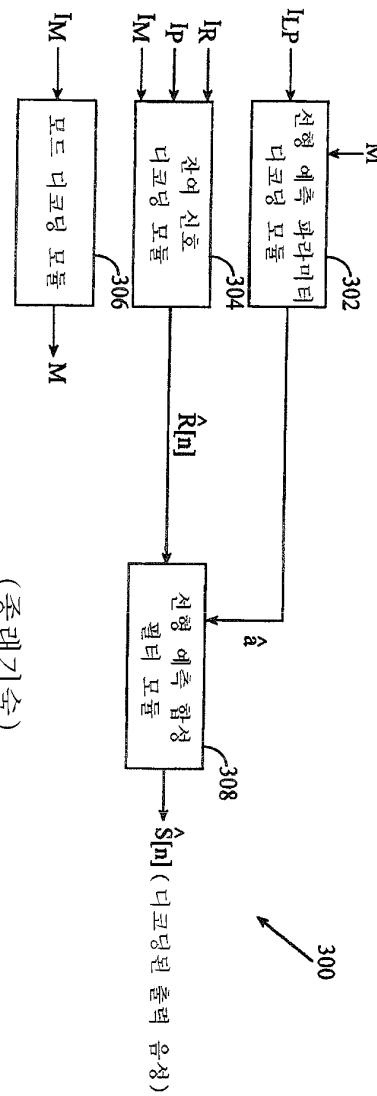
도면2



도면3

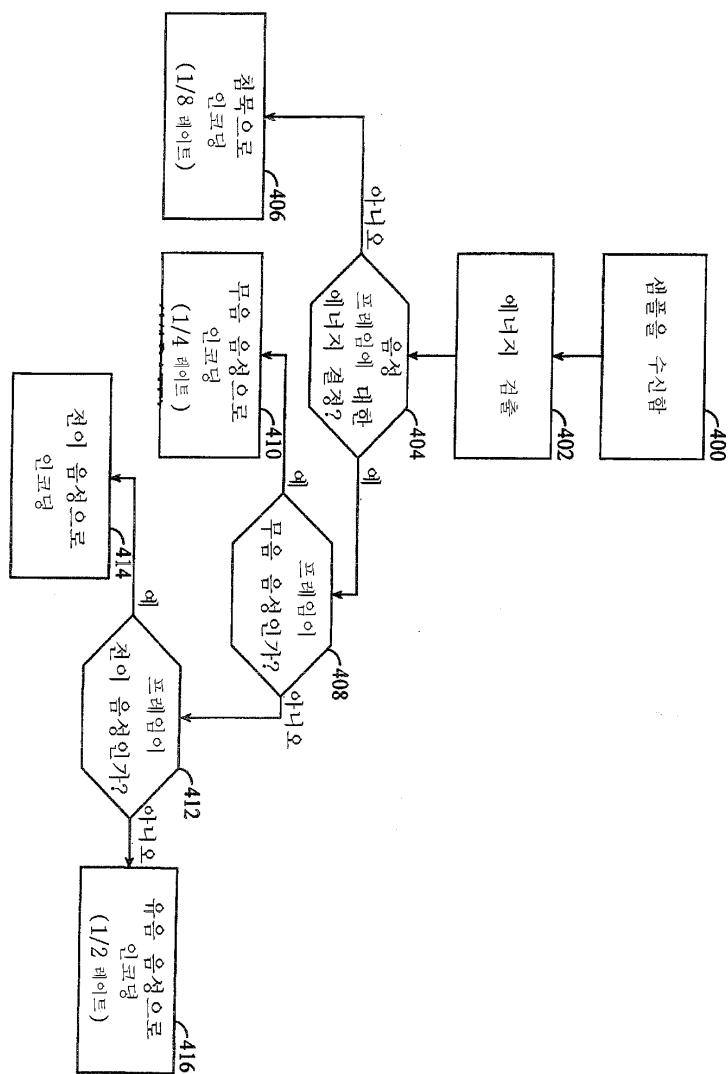


도면4

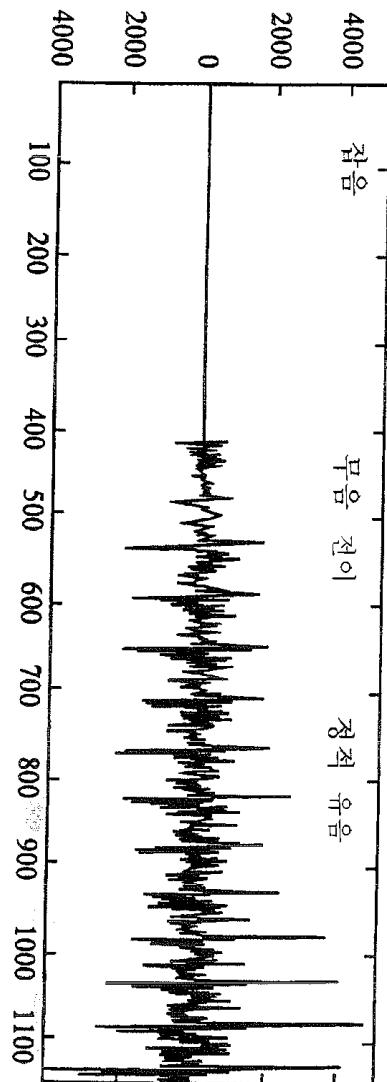


(종래기술)

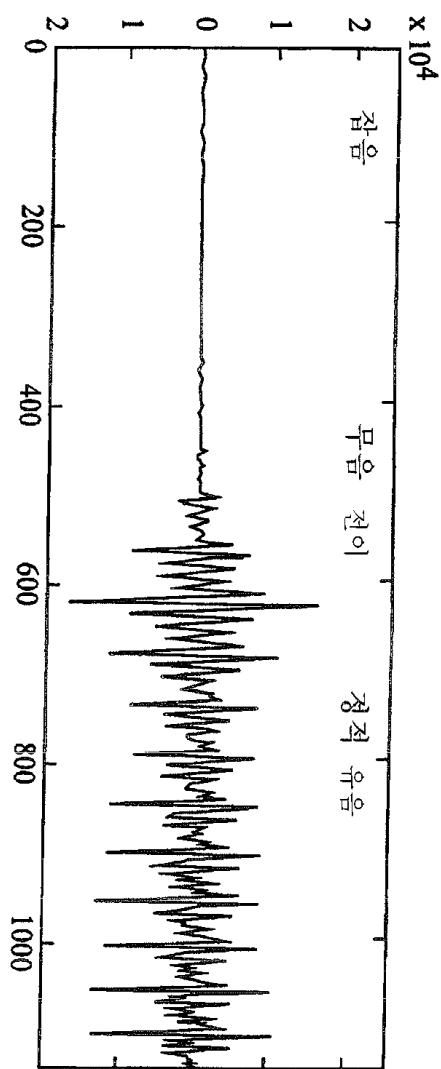
도면5



도면6a



도면6b



도면7

