



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2014년08월08일
(11) 등록번호 10-1428938
(24) 등록일자 2014년08월04일

(51) 국제특허분류(Int. Cl.) <i>G10L 19/038</i> (2013.01) <i>G10L 19/16</i> (2013.01)	(73) 특허권자 세종대학교산학협력단
(21) 출원번호 10-2013-0098037	서울특별시 광진구 능동로 209 (군자동, 세종대학교)
(22) 출원일자 2013년08월19일 심사청구일자 2013년08월19일	(72) 발명자 김무영
(56) 선행기술조사문헌 KR100446630 B1 US20110135007 A1 US6389388 B1 US20110112841 A1	서울 강남구 선릉로 221, 210동 402호 (도곡동, 도곡텍슬아파트) 주현호 경기 용인시 기흥구 구성로 90, 207동 502호 (연남동, 삼성래미안2차아파트)
	(74) 대리인 특허법인태백

전체 청구항 수 : 총 12 항

심사관 : 이수철

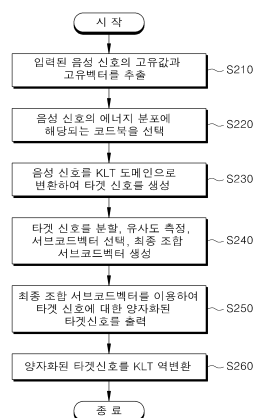
(54) 발명의 명칭 음성 신호의 벡터 양자화 장치 및 그 방법

(57) 요약

본 발명은 음성 신호 벡터 양자화 장치 및 그 방법에 관한 것이다. 본 발명의 음성 신호 벡터 양자화 장치는 입력된 상기 음성 신호의 고유값과 고유벡터를 추출하는 고유치 추출부; 상기 고유값을 이용하여 기 학습된 복수의 코드북 중에서 상기 음성 신호의 에너지 분포에 해당되는 코드북을 선택하는 코드북 선택부; 상기 고유벡터를 이용하여 상기 음성 신호를 KLT 도메인으로 변환하여 타겟 신호를 생성하는 KLT 변환부; 상기 타겟 신호를 복수의 타겟 서브벡터로 분할하고, 상기 코드북의 서브코드벡터와 상기 타겟 서브벡터 사이의 유사도를 측정하여, 상기 타겟 서브벡터 각각에 대하여 유사도가 큰 복수의 후보 서브코드벡터를 선택하고, 상기 선택된 후보 서브코드벡터를 조합하여 최종 조합 서브코드벡터를 생성하는 제어부; 및 상기 최종 조합 서브코드벡터를 이용하여 상기 타겟 신호에 대한 양자화된 타겟 신호를 출력하는 출력부를 포함한다.

이와 같이 본 발명에 따르면, 기존의 KLT-CVQ 음성 코더의 단점인 하나의 프레임에 사용 가능한 샘플 수가 5개로 한정되는 점과 피치(pitch) 정보를 효율적으로 이용하지 못하는 점을 보완하여, 하나의 프레임에 많은 샘플을 사용함으로써 피치 정보를 효율적으로 이용가능하고 코드벡터 서치 과정에서 계산량을 줄일 수 있으며 우수한 품질의 음성 서비스를 제공할 수 있다.

대 표 도 - 도2



이 발명을 지원한 국가연구개발사업

과제고유번호 1345192901

부처명 교육과학기술부

연구사업명 일반연구자지원사업

연구과제명 Interactive Music Application을 위한 객체기반 오디오 콘텐츠 코딩, 신호분리 및 잡음제거 기법에 관한 기초 기술 연구

기 여 율 1/1

주관기관 세종대학교 산학협력단

연구기간 2012.05.01 ~ 2015.04.30

특허청구의 범위

청구항 1

KLT-CVQ 코더를 이용한 음성 신호의 벡터 양자화 장치에 있어서,

입력된 상기 음성 신호의 고유값과 고유벡터를 추출하는 고유치 추출부;

상기 고유값을 이용하여 기 학습된 복수의 코드북 중에서 상기 음성 신호의 에너지 분포에 해당되는 코드북을 선택하는 코드북 선택부;

상기 고유벡터를 이용하여 상기 음성 신호를 KLT 도메인으로 변환하여 타겟 신호를 생성하는 KLT 변환부;

상기 타겟 신호를 복수의 타겟 서브벡터로 분할하고, 상기 코드북의 서브코드벡터와 상기 타겟 서브벡터 사이의 유사도를 측정하여, 상기 타겟 서브벡터 각각에 대하여 유사도가 큰 복수의 후보 서브코드벡터를 선택하고, 상기 선택된 후보 서브코드벡터를 조합하여 최종 조합 서브코드벡터를 생성하는 제어부; 및

상기 최종 조합 서브코드벡터를 이용하여 상기 타겟 신호에 대한 양자화된 타겟 신호를 출력하는 출력부를 포함하는 음성 신호 양자화 장치.

청구항 2

제 1항에 있어서,

상기 양자화된 타겟 신호에 대해 KLT 역변환하는 KLT 역변환부를 더 포함하는 음성 신호 양자화 장치.

청구항 3

제 1항 또는 제 2항에 있어서,

상기 제어부는,

상기 선택된 코드북의 서브코드벡터와 동일하게 상기 타겟 신호를 분할하여 상기 타겟 서브벡터를 생성하는 벡터 분할부;

상기 서브코드벡터와 상기 타겟 서브벡터 사이의 유사도를 측정하여, 상기 타겟 서브벡터 각각에 대하여 유사도가 큰 복수의 후보 서브코드벡터를 선택하는 후보 서브코드벡터 선택부;

상기 복수의 i번째 후보 서브코드벡터와 상기 선택된 복수의 i+1번째 후보 서브코드벡터를 모든 경우의 수만큼 조합하고, 상기 조합된 후보 서브코드벡터 중에서 상기 타겟 신호의 i번째와 i+1번째 타겟 서브벡터와 유사도가 큰 조합된 후보 서브코드벡터를 복수 개 선택하는 조합 서브코드벡터 선택부; 및

상기 조합 서브코드벡터 선택부에 의해 기 조합된 후보 서브코드벡터와 상기 후보 서브코드벡터 선택부에 의해 선택된 복수의 마지막 번째 후보 서브코드벡터를 모든 경우의 수만큼 조합하고, 상기 조합된 후보 서브코드벡터 중에서 상기 타겟 서브벡터와 유사도가 가장 큰 최종 조합 서브코드벡터를 생성하는 최종 조합 서브코드벡터 생성부를 포함하는 음성 신호 양자화 장치.

청구항 4

제 3항에 있어서,

상기 후보 서브코드벡터 선택부와 상기 조합 서브코드벡터 선택부는,

아래 수학적식을 통하여 유사도를 판단하는 음성 신호 양자화 장치:

$$g_i = \frac{\phi_i^T c_i^n}{c_i^{nT} c_i^n}$$

$$n_i^* = \operatorname{argmin}(\phi_i - \overline{g_i^n} c_i^n)^2$$

여기서, ϕ_i 는 타겟 신호의 i번째 타겟 서브벡터이며, $\overline{g_i^n}$ 은 타겟 서브벡터와 서브코드벡터 사이의 계인이고, c_i^n 은 i번째 서브벡터의 n번째 코드벡터이고, n_i^* 는 타겟 서브벡터와 서브코드벡터 사이의 유사도이다.

청구항 5

제 3항에 있어서,

상기 최종 조합 서브코드벡터 생성부는,

아래 수학적식을 통하여 유사도를 판단하는 음성 신호 양자화 장치:

$$g_o = \frac{y_z^{kT} \overline{y^k}}{y^{kT} y^k}$$

$$L^* = \operatorname{argmin}(y_z - g_o^{L-L})$$

여기서, g_o 는 상기 타겟 서브벡터와 상기 조합된 후보 서브코드벡터의 계인이며, $\overline{y^k}$ 는 조합된 코드북이고, y_z^k 는 고유벡터 행렬을 음성 신호의 영상태응답에 곱해서 구한 KLT 변환된 타겟 신호이고, L^* 은 타겟 서브벡터 (y_z)와 상기 조합된 후보 서브코드벡터 사이의 유사도이다.

청구항 6

제 5항에 있어서,

상기 출력부는,

아래 수학적식과 같이 상기 최종 조합 서브코드벡터의 계인 및 형태를 이용하여 상기 타겟 신호로부터 상기 양자

화된 타겟 신호($\hat{y}_z$)를 생성하는 음성 신호 양자화 장치:

$$\hat{y}_z = g_o^{L^*} \overline{y}^{L^*}$$

여기서, g_o 는 최종 조합 서브코드벡터의 계인이며, $\overline{y}$ 는 최종 조합 서브코드벡터이다.

청구항 7

KLT-CVQ 코더를 이용한 음성 신호의 벡터 양자화 방법에 있어서,

입력된 상기 음성 신호의 고유값과 고유벡터를 추출하는 단계;

상기 고유값을 이용하여 기 학습된 복수의 코드북 중에서 상기 음성 신호의 에너지 분포에 해당되는 코드북을 선택하는 단계;

상기 고유벡터를 이용하여 상기 음성 신호를 KLT 도메인으로 변환하여 타겟 신호를 생성하는 단계;

상기 타겟 신호를 복수의 타겟 서브벡터로 분할하고, 상기 코드북의 서브코드벡터와 상기 타겟 서브벡터 사이의 유사도를 측정하여, 상기 타겟 서브벡터 각각에 대하여 유사도가 큰 복수의 후보 서브코드벡터를 선택하고, 상기 선택된 후보 서브코드벡터를 조합하여 최종 조합 서브코드벡터를 생성하는 단계; 및

상기 최종 조합 서브코드백터를 이용하여 상기 타겟 신호에 대한 양자화된 타겟 신호를 출력하는 단계를 포함하는 음성 신호 양자화 방법.

청구항 8

제 7항에 있어서,

상기 양자화된 타겟 신호에 대해 KLT 역변환하는 단계를 더 포함하는 음성 신호 양자화 방법.

청구항 9

제 7항 또는 제 8항에 있어서,

상기 조합 서브코드백터를 생성하는 단계는,

상기 선택된 코드북의 서브코드백터와 동일하게 상기 타겟 신호를 분할하여 상기 타겟 서브백터를 생성하는 단계;

상기 서브코드백터와 상기 타겟 서브백터 사이의 유사도를 측정하여, 상기 타겟 서브백터 각각에 대하여 유사도가 큰 복수의 후보 서브코드백터를 선택하는 단계;

상기 복수의 i번째 후보 서브코드백터와 상기 선택된 복수의 i+1번째 후보 서브코드백터를 모든 경우의 수만큼 조합하고, 상기 조합된 후보 서브코드백터 중에서 상기 타겟 신호의 i번째와 i+1번째 타겟 서브코드백터와 유사도가 큰 조합된 후보 서브코드백터를 복수 개 선택하는 단계; 및

상기 조합 서브코드백터 선택부에 의해 기 조합된 후보 서브코드백터와 상기 후보 서브코드백터 선택부에 의해 선택된 복수의 마지막번째 후보 서브코드백터를 모든 경우의 수만큼 조합하고, 상기 조합된 후보 서브코드백터 중에서 상기 타겟 신호와 유사도가 가장 큰 최종 조합 서브코드백터를 생성하는 단계를 포함하는 음성 신호 양자화 방법.

청구항 10

제 9항에 있어서,

상기 유사도가 큰 복수의 후보 서브코드백터를 선택하는 단계와 상기 유사도가 큰 조합된 후보 서브코드백터를 복수 개 선택하는 단계는,

아래 수학적식을 통하여 유사도를 판단하는 음성 신호 양자화 방법:

$$\bar{g}_i^n = \frac{\phi_i^T c_i^n}{c_i^{nT} c_i^n}$$

$$n_i^* = \operatorname{argmin}(\phi_i - \bar{g}_i^n c_i^n)^2$$

여기서, ϕ_i 는 타겟 신호의 i번째 타겟 서브백터이며, $\bar{g}_i^n$ 은 타겟 서브백터와 서브코드백터 사이의 계인이고, c_i^n 은 i번째 서브백터의 n번째 코드백터이고, n_i^* 는 타겟 서브백터와 서브코드백터 사이의 유사도이다.

청구항 11

제 9항에 있어서,

상기 최종 조합 서브코드백터를 생성하는 단계는,

아래 수학적식을 통하여 유사도를 판단하는 음성 신호 양자화 방법:

$$g_o = \frac{y_z^{kT} \bar{y}^k}{y^{kT} y^k}$$

$$L^* = \operatorname{argmin}(y_z - g_o^{L-L})$$

여기서, g_o 는 상기 타겟 서브벡터와 상기 조합된 후보 서브코드벡터의 계인이며, $\bar{y}^k$ 는 조합된 코드북이고, y_z^k 는 고유벡터 행렬을 음성 신호의 영상태응답에 곱해서 구한 KLT 변환된 타겟 신호이고, L^* 은 타겟 서브벡터 (y_z)와 상기 조합된 후보 서브코드벡터 사이의 유사도이다.

청구항 12

제 11항에 있어서,

상기 타겟 신호에 대한 양자화된 타겟 신호를 출력하는 단계는,

아래 수학적식과 같이 상기 최종 조합 서브코드벡터의 계인 및 형태를 이용하여 상기 타겟 신호로부터 상기 양자

화된 타겟 신호($\hat{y}_z$)를 생성하는 음성 신호 양자화 방법:

$$\hat{y}_z = g_o^{L-L} \bar{y}^{L^*}$$

여기서, g_o 는 최종 조합 서브코드벡터의 계인이며, $\bar{y}$ 는 최종 조합 서브코드벡터이다.

명세서

기술분야

[0001] 본 발명은 음성을 합성하는 음성 코더에 관한 것으로서, 더욱 상세하게는 높은 차원이 가능한 KLT-CVQ(Karhunen Loeve Transform based adaptive Classified Vector Quantization) 음성 코더를 이용하는 음성 신호의 벡터 양자화 장치 및 그 방법에 관한 것이다.

배경기술

[0002] 음성 신호의 벡터 양자화 장치 및 그 방법은 음성 신호를 부호화할 때, KLT를 기반으로 분할된 코드북과 음성 신호의 고유값 및 고유벡터를 이용하여 입력되는 음성 신호를 양자화 함으로써, 벡터 양자화의 이점을 극대화시킬 수 있다.

[0003] 음성 신호를 양자화 하는 장치 또는 그 방법에 있어서 핵심이 되는 것이 음성 코더인데, 기존의 기술에서는 음성 코더로서 보통 Code-Excited Linear Predictive (CELP) 코더를 사용한다. 상기 코더는 벡터 양자화의 장점을 모두 이용하지 못하기 때문에 성능이 감소하는 단점이 있다. 이러한 단점을 보완하기 위해서 Karhunen-Loeve transform (KLT)-based adaptive classified Vector Quantization(CVQ)이 제안되었고, CELP의 단점을 해결할 수 있었다. 그러나 KLT-CVQ는 하나의 frame 마다 음성 샘플을 5개만을 이용한다는 제한과 적은 sample만을 사용해서 pitch 정보를 효율적으로 이용하지 못한다는 단점을 갖고 있다.

[0004] 본 발명의 배경이 되는 기술은 대한민국 공개특허공보 제 2003-0087373호(2003.1.14)에 기재되어 있다.

발명의 내용

해결하려는 과제

[0005] 본 발명이 해결하고자 하는 과제는, 기존의 5개의 sample만 이용하던 KLT-CVQ 음성 코더에 더 많은 sample을 이용할 수 있게 함으로써 pitch 정보를 이용할 수 있고, 코드북 search 알고리즘에 있어서 shape 코드백터를 더욱 효율적으로 선택함으로써 계산량을 크게 줄일 수 있는 알고리즘에 관한 기술을 제공하는 것이다.

과제의 해결 수단

[0006] 상기한 바와 같은 목적을 달성하기 위한 본 발명의 하나의 실시예에 따른 음성 신호의 벡터 양자화 장치는, KLT-CVQ 코더를 이용한 음성 신호의 벡터 양자화 장치에 있어서, 입력된 상기 음성 신호의 고유값과 고유벡터를 추출하는 고유치 추출부; 상기 고유값을 이용하여 기 학습된 복수의 코드북 중에서 상기 음성 신호의 에너지 분포에 해당되는 코드북을 선택하는 코드북 선택부; 상기 고유벡터를 이용하여 상기 음성 신호를 KLT 도메인으로 변환하여 타겟 신호를 생성하는 KLT 변환부; 상기 타겟 신호를 복수의 타겟 서브벡터로 분할하고, 상기 코드북의 서브코드백터와 상기 타겟 서브벡터 사이의 유사도를 측정하여, 상기 타겟 서브벡터 각각에 대하여 유사도가 큰 복수의 후보 서브코드백터를 선택하고, 상기 선택된 후보 서브코드백터를 조합하여 최종 조합 서브코드백터를 생성하는 제어부; 및 상기 최종 조합 서브코드백터를 이용하여 상기 타겟 신호에 대한 양자화된 타겟 신호를 출력하는 출력부를 포함한다.

[0007] 또한, 음성 신호의 벡터 양자화 장치는 상기 양자화된 타겟 신호에 대해 KLT 역변환하는 KLT 역변환부를 더 포함할 수 있다.

[0008] 또한, 상기 제어부는, 상기 선택된 코드북의 서브코드백터와 동일하게 상기 타겟 신호를 분할하여 상기 타겟 서브벡터를 생성하는 서브벡터 분할부; 상기 서브코드백터와 상기 타겟 서브벡터 사이의 유사도를 측정하여, 상기 타겟 서브벡터 각각에 대하여 유사도가 큰 복수의 후보 서브코드백터를 선택하는 후보 서브코드백터 선택부; 상기 복수의 i번째 후보 서브코드백터와 상기 선택된 복수의 i+1번째 후보 서브코드백터를 모든 경우의 수만큼 조합하고, 상기 조합된 후보 서브코드백터 중에서 상기 타겟 신호의 i번째와 i+1번째 타겟 서브벡터와 유사도가 큰 조합된 후보 서브코드백터를 복수 개 선택하는 조합 서브코드백터 선택부; 및 상기 조합 서브코드백터 선택부에 의해 기 조합된 후보 서브코드백터와 상기 후보 서브코드백터 선택부에 의해 선택된 복수의 마지막번째 후보 서브코드백터를 모든 경우의 수만큼 조합하고, 상기 조합된 후보 서브코드백터 중에서 상기 타겟 신호와 유사도가 가장 큰 최종 조합 서브코드백터를 생성하는 최종 조합 서브코드백터 생성부를 포함할 수 있다.

[0009] 또한, 상기 후보 서브코드백터 선택부와 상기 조합 서브코드백터 선택부는, 아래 수학적식을 통하여 유사도를 판단할 수 있다.

$$g_i^n = \frac{\phi_i^T c_i^n}{c_i^{nT} c_i^n}$$

[0010]

$$n_i^* = \operatorname{argmin}(\phi_i - \bar{g}_i^n c_i^n)^2$$

[0011]

[0012] 여기서, ϕ_i 는 타겟 신호의 i번째 타겟 서브벡터이며, $\bar{g}_i^n$ 은 타겟 서브벡터와 서브코드백터 사이의 계인이고, c_i^n 은 i번째 서브벡터의 n번째 코드백터이고, n_i^* 는 타겟 서브벡터와 서브코드백터 사이의 유사도이다.

[0013] 또한, 상기 최종 조합 서브코드백터 생성부는, 아래 수학적식을 통하여 유사도를 판단할 수 있다.

$$g_o = \frac{y_z^k \bar{y}^k}{y^{kT} y^k}$$

[0014]

$$L^* = \operatorname{argmin}(y_z - g_o^{L-L} \bar{y})$$

[0015]

[0016] 여기서, g_o 는 상기 타겟 서브벡터와 상기 조합된 후보 서브코드벡터의 계인이며, $\overline{y}^k$ 는 조합된 코드북이고, y_z^k 는 고유벡터 행렬을 음성 신호의 영상태응답에 곱해서 구한 KLT 변환된 타겟 신호이고, L^* 은 타겟 서브벡터 (y_z)와 상기 조합된 후보 서브코드벡터 사이의 유사도이다.

[0017] 또한, 상기 출력부는, 아래 수식과 같이 상기 최종 조합 서브코드벡터의 계인 및 형태를 이용하여 상기 타겟 신호로부터 상기 양자화된 타겟 신호를 생성할 수 있다.

$$\hat{y}_z = g_o^{L^*} \overline{y}^{L^*}$$

[0018]

[0019] 여기서, g_o 는 최종 조합 서브코드벡터의 계인이며, $\overline{y}$ 는 최종 조합 서브코드벡터이다.

[0020] 본 발명의 하나의 실시예에 따른 음성 신호의 벡터 양자화 방법은, KLT-CVQ 코더를 이용한 음성 신호의 벡터 양자화 방법에 있어서, 입력된 상기 음성 신호의 고유값과 고유벡터를 추출하는 단계; 상기 고유값을 이용하여 기 학습된 복수의 코드북 중에서 상기 음성 신호의 에너지 분포에 해당되는 코드북을 선택하는 단계; 상기 고유벡터를 이용하여 상기 음성 신호를 KLT 도메인으로 변환하여 타겟 신호를 생성하는 단계; 상기 타겟 신호를 복수의 타겟 서브벡터로 분할하고, 상기 코드북의 서브코드벡터와 상기 타겟 서브벡터 사이의 유사도를 측정하여, 상기 타겟 서브벡터 각각에 대하여 유사도가 큰 복수의 후보 서브코드벡터를 선택하고, 상기 선택된 후보 서브코드벡터를 조합하여 최종 조합 서브코드벡터를 생성하는 단계; 및 상기 최종 조합 서브코드벡터를 이용하여 상기 타겟 신호에 대한 양자화된 타겟 신호를 출력하는 단계를 포함한다.

발명의 효과

[0021] 본 발명인 KLT-CVQ 음성 코더를 이용한 음성 신호의 벡터 양자화 장치 및 그 방법에 따르면, 기존의 KLT-CVQ 음성 코더의 단점인 하나의 프레임에 사용 가능한 샘플 수가 5개로 한정되는 점과 피치(pitch) 정보를 효율적으로 이용하지 못하는 점을 보완하여, 하나의 프레임에 많은 샘플을 사용함으로써 피치 정보를 효율적으로 이용가능하고 코드벡터 서치 과정에서 계산량을 줄일 수 있으며 우수한 품질의 음성 서비스를 제공할 수 있다.

도면의 간단한 설명

[0022] 도 1은 본 발명의 실시예에 따른 음성 신호의 벡터 양자화 장치의 구성도이다.

도 2는 본 발명의 실시예에 따른 음성 신호의 벡터 양자화 방법의 순서도이다.

도 3은 본 발명의 실시예에 따른 음성 신호의 벡터 양자화 방법의 구성도이다.

도 4는 본 발명의 실시예에 따른 최종 조합 서브코드벡터를 생성하는 과정을 설명하기 위한 순서도이다.

도 5a 내지 도 5c는 본 발명의 실시예에 따른 최종 조합 서브코드벡터를 생성하기 위한 과정을 설명하기 위한 예시도이다.

발명을 실시하기 위한 구체적인 내용

[0023] 이하에서는 첨부한 도면을 참조하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 본 발명의 실시예를 상세히 설명한다. 그러나 본 발명은 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시예에 한정되지 않는다. 그리고 도면에서 본 발명을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.

[0024] 먼저 본 발명의 음성 신호의 벡터 양자화 장치에 대하여 설명한다.

[0025] 도 1은 본 발명의 실시예에 따른 음성 신호의 벡터 양자화 장치의 구성도이다.

[0026] 도 1를 참조하면, 본 발명의 실시예에 따른 음성 신호의 벡터 양자화 장치(100)는 고유치 추출부(110), 코드북 선택부(120), KLT 변환부(130), 제어부(140), 출력부(150) 및 KLT 역변환부(160)를 포함한다.

[0027] 고유치 추출부(110)는 입력된 음성 신호를 분석하여 고유값과 고유벡터를 추출한다. 여기서 고유값은 입력된 음성 신호의 분산에 대응되는 것으로 추후 코드북을 검색하여 선택하는 과정에서 이용되며, 고유벡터는 추후 선택

된 코드북 내에서 유사한 코드벡터를 검색하여 선택하는데 이용된다.

- [0028] 코드북 선택부(120)는 고유치 추출부(110)로부터 추출된 고유값을 이용하여 기 학습된 복수의 코드북 중에서 입력된 음성 신호의 에너지 분포에 해당되는 코드북을 선택한다. 즉, 코드북 선택부(120)는 복수의 코드북 중에서 입력된 음성 신호의 분산과 유사한 분산을 가지는 코드북을 선택한다. 여기서 복수의 코드북은 에너지 분포를 나타내는 분산값에 따라서 여러 가지 에너지 분포의 Class로 나뉘어져 있다.
- [0029] KLT 변환부(130)는 고유치 추출부(110)에서 추출된 고유벡터를 이용하여 음성 신호를 KLT 도메인으로 변환하여 타겟 신호를 생성한다. KLT 변환부(130)는 KLT-CVQ 코더로 구현될 수 있으며, KLT 변환부(130)는 입력되는 음성 신호를 KLT 도메인으로 변환한다.
- [0030] 제어부(140)는 코드북 선택부(120)에 의해 선택된 코드북 중에서 코드벡터를 서치(search)하는 과정을 제어하는데, 제어부(140)는 벡터 분할부(141), 후보 서브코드벡터 선택부(142), 조합 서브코드벡터 선택부(143) 및 최종 조합 서브코드벡터 생성부(144)를 포함한다.
- [0031] 벡터 분할부(141)는 KLT 변환부(130)에서 출력되는 타겟 신호를 복수의 타겟 서브벡터로 각각 분할한다.
- [0032] 후보 서브코드벡터 선택부(142)는 코드북의 서브코드벡터와 타겟 서브벡터 사이의 유사도를 측정하여, 타겟 서브벡터 각각에 대하여 유사도가 큰 복수의 후보 서브코드벡터를 서브코드벡터에서 선택한다.
- [0033] 조합 서브코드벡터 선택부(143)는 복수의 i번째 후보 서브코드벡터와 상기 선택된 복수의 i+1번째 후보 서브코드벡터를 모든 경우의 수만큼 조합하고, 조합된 후보 서브코드벡터 중에서 타겟 신호의 i번째와 i+1번째 타겟 서브벡터와 유사도가 큰 조합된 후보 서브코드벡터를 복수 개 선택한다.
- [0034] 최종 조합 서브코드벡터 생성부(144)는 조합 서브코드벡터 선택부(143)에 의해 기 조합된 후보 서브코드벡터와 후보 서브코드벡터 선택부(142)에 의해 선택된 복수의 마지막 번째 후보 서브코드벡터를 모든 경우의 수만큼 조합하고, 조합된 후보 서브코드벡터 중에서 타겟 신호와 유사도가 가장 큰 최종 조합 서브코드벡터를 생성한다.
- [0035] 출력부(150)는 최종 조합 서브코드벡터를 이용하여 타겟 신호에 대한 양자화된 타겟 신호를 출력한다.
- [0036] KLT 역변환부(160)는 양자화된 타겟 신호를 KLT 역변환한다.
- [0037] 이하에서는 도 2 내지 도 5를 통하여 본 발명의 실시예에 따른 음성 신호의 벡터 양자화 방법에 대하여 더욱 자세히 설명한다.
- [0038] 도 2는 본 발명의 실시예에 따른 음성 신호의 벡터 양자화 방법의 순서도이다.
- [0039] 먼저 고유치 추출부(110)는 입력된 음성 신호의 고유값과 고유벡터를 추출한다(S210). 즉, 고유치 추출부(110)는 다음의 수학적 1을 이용하여 고유값 **D**와 고유벡터 **U**를 추출할 수 있다.

수학적 1

$$E[s_z^k s_z^{kT}] = U^T D U$$

- [0040]
- [0041] 여기서, s_z^k 는 입력신호의 k차원 영상태응답이다. **U**행렬은 고유벡터 행렬, **D**행렬은 고유값 행렬이다. 두 행렬 모두 k-by-k 행렬이다.
- [0042] 수학적 1과 같이, 입력되는 음성 신호의 공분산 행렬에 대하여 고유값 분해를 통하여 고유값을 추출할 수 있다.
- [0043] 도 3은 본 발명의 실시예에 따른 음성 신호의 벡터 양자화 방법의 구성도이다.
- [0044] 도 3에서, 고유값(Eigen value)을 이용하여 복수의 코드북에서 Class를 선택하는 과정이 도시되어 있다.
- [0045] 코드북 선택부(120)는 고유값을 이용하여 기 학습된 복수의 코드북 중에서 상기 음성 신호의 에너지 분포에 해당되는 코드북을 선택한다(S220).
- [0046] 여기서, 학습이란 학습 데이터라고 하는 복수의 음성 신호를 이용하여 복수의 코드북을 만드는 과정이다. 즉, 코드북 학습은 학습 데이터의 에너지 분포를 밝혀서 에너지 분포 별로 코드북을 만들어 Class를 정하는 과정이다. 코드북 학습과정에서 코드북을 만드는 경우, 복수의 음성 신호를 분할하여 분할된 신호에 대한 서브코드벡

터를 이용한다.

- [0047] 기 학습된 복수의 코드북(Karhunen-Love transform based Adaptive Classified Vector Quantization Code Book: KLT-CVQ 코드북)은 복수의 음성 신호인 학습 데이터의 분산에 따라서 Class가 나뉘어 진다. 즉, 기 학습된 복수의 코드북은 학습단계를 통해 에너지 분포에 따라서 Class가 나뉘어 지는 형태로 생성된다. 복수의 코드북을 만들기 위해서는 변형된 가우시안 모델인GMM-SVQ(Gaussian Mixture Model- Split Vector Quantization)가 이용된다.
- [0048] 도 3에서, 고유값과 고유벡터 U 를 이용한 코드북의 모양을 찾는 과정(Shape search)과 코드북의 크기를 찾는 과정(Gain search)이 도시되어 있다. 코드북 선택부(120)의 코드북 선택 과정은 코드북의 크기를 찾는 과정에 해당된다.
- [0049] 고유치 추출부(110)에 의해 추출된 고유값은 입력신호의 분산을 나타내므로, 코드북 선택부(120)는 입력 신호의 고유값을 통하여 복수의 코드북 중에서 음성 신호와 분산이 가장 유사한 코드북을 선택한다.
- [0050] KLT 변환부(130)는 음성 신호의 영상태응답 값을 입력받고, 고유값 추출부(110)에서 생성된 고유벡터를 이용하여 입력된 음성 신호를 KLT 도메인으로 변환하여 타겟 신호를 생성한다(S230).

수학식 2

$$Us_z^k = y_z^k$$

- [0051]
- [0052] 즉, 수학식 2와 같이 KLT 변환부(130)는 고유벡터 U 행렬과 입력신호의 k 차원 영상태응답(s_z^k)을 곱하여 타겟 신호(y_z^k)을 생성한다.
- [0053] 제어부(140)는 타겟 신호를 복수의 타겟 서브벡터로 분할한다. 그리고 제어부(140)는 코드북의 서브코드벡터와 타겟 서브벡터 사이의 유사도를 측정하여, 타겟 서브벡터 각각에 대하여 유사도가 큰 복수의 후보 서브코드벡터를 선택하고, 선택된 후보 서브코드벡터를 조합하여 최종 조합 서브코드벡터를 생성한다(S240).
- [0054] 다음으로, 출력부(150)는 최종 조합 서브코드벡터를 이용하여 상기 타겟 신호에 대한 양자화된 타겟 신호를 출력한다(S250).
- [0055] 그리고, KLT 역변환부(160)는 양자화된 타겟 신호를 KLT 역변환한다(S260).
- [0056] 이하에서는 도 4 내지 도 5c를 통하여 제어부(140)가 최종 조합 서브코드벡터를 생성하는 과정(S240)에 대하여 더욱 상세하게 설명한다.
- [0057] 도 4는 본 발명의 실시예에 따른 최종 조합 서브코드벡터를 생성하는 과정을 설명하기 위한 순서도이고, 도 5a 내지 도 5c는 본 발명의 실시예에 따른 최종 조합 서브코드벡터를 생성하기 위한 과정을 설명하기 위한 예시도이다.
- [0058] 벡터 분할부(141)는 상기 선택된 코드북의 서브코드벡터와 동일하게 상기 타겟 신호를 분할하여 타겟 서브벡터를 생성한다(S241).
- [0059] 본 발명의 실시예에 따르면 모든 서브코드벡터를 한꺼번에 조합하는 방식이 아닌, 낮은 차원의 서브코드벡터부터 복수 개, 이를 테면 2개씩 차례대로 조합을 하는 방식을 이용한다.
- [0060] 코드북 선택부(120)에 의해 선택된 코드북은 1024개의 서브코드벡터로 이루어져 있다고 가정하면, 설명의 편의상 도 5a 및 도 5b에서 하나의 코드벡터는 8개의 서브코드벡터($i=1, 2, \dots, 8$)로 나누는 것으로 예시한다.
- [0061] 벡터 분할부(141)는 타겟 신호를 구성하는 코드벡터를 8개의 타겟 서브벡터로 분할한다.
- [0062] 다음으로 후보 서브코드벡터 선택부(142)는 코드북의 서브코드벡터와 타겟 신호를 구성하는 타겟 서브벡터 사이의 유사도를 측정하여, 타겟 서브벡터 각각에 대하여 유사도가 큰 복수의 후보 서브코드벡터를 선택한다(S242).
- [0063] 즉, 후보 서브코드벡터 선택부(142)는 아래 수학식 3을 이용하여 코드북의 서브코드벡터와 타겟 서브벡터 사이의 split 타겟 신호의 계인(optimal gain)을 연산한다.

수학식 3

$$\bar{g}_i^n = \frac{\phi_i^T c_i^n}{c_i^{nT} c_i^n}$$

[0064]

[0065] 그리고 후보 서브코드벡터 선택부(142)는 계인을 기초로 아래 수학식 4를 이용하여 타겟 서브벡터와의 distortion이 적은 서브코드벡터를 순서대로 정렬시킨다.

수학식 4

$$n_i^* = \operatorname{argmin}(\phi_i - \bar{g}_i^n c_i^n)^2$$

[0066]

[0067] 여기서, n은 코드북 인덱스이고, i는 서브벡터 인덱스이다. 즉 ϕ_i 는 타겟 신호의 i번째 타겟 서브벡터이며, $\bar{g}_i^n$ 은 타겟 서브벡터와 서브코드벡터 사이의 계인(optimal gain)이고, c_i^n 은 i번째 서브벡터의 n번째 코드벡터이다.

[0068] 그리고 정렬된 서브코드벡터 중에서 유사도가 크고 왜곡(distortion)이 적은 상위 M개의 서브코드벡터를 선택한다. 여기서 M은 유사도가 높은 서브코드벡터를 순서대로 정렬시켰을 때 상위에 있는 서브코드벡터로서 후보 서브코드벡터라고도 한다. 도 5b에서는 설명의 편의상 M=2개로 가정한다.

[0069] 다음으로 조합 서브코드벡터 선택부(143)는 복수의 i번째 후보 서브코드벡터와 상기 선택된 복수의 i+1번째 후보 서브코드벡터를 모든 경우의 수만큼 조합하고, 상기 조합된 후보 서브코드벡터 중에서 상기 타겟 신호의 i번째와 i+1번째 타겟 서브벡터와 유사도가 큰 조합된 후보 서브코드벡터를 복수 개 선택한다(S243).

[0070] 이하에서는 도 5b를 통해 조합된 후보 서브코드벡터 중에서 유사도가 큰 조합된 후보 서브코드벡터를 선택하는 과정에 대해 설명한다.

[0071] 도 5b에는, i=1 후보 서브코드벡터와 i=2 후보 서브코드벡터가 조합되는 과정을 도시한 것이다. 첫 번째 서브코드벡터(i=1)의 경우, 후보 서브코드벡터는 (1-1) 및 (1-2)이고, 두 번째 서브코드벡터(i=2)의 경우, 후보 서브코드벡터는 (2-1) 및 (2-2)라고 가정한다.

[0072] 그리고, 첫 번째 서브코드벡터의 후보 서브코드벡터와 두 번째 서브코드벡터의 후보 서브코드벡터들을 조합하면 총 4개의 경우의 수가 생성된다. 즉, (1-1, 2-1), (1-1, 2-2), (1-2, 2-1) 및 (1-2, 2-2)의 4개의 조합된 후보 서브코드벡터가 생성된다.

[0073] 조합 서브코드벡터 선택부(143)는 수학식 3 및 수학식 4를 이용하여 4개의 조합된 후보 서브코드벡터와 타겟 신호의 첫 번째 서브벡터와 두 번째 서브벡터가 조합된 서브벡터의 유사도를 계산하여, 유사도가 큰 2개의 조합된 후보 서브코드벡터를 선택한다.

[0074] 도 5b에서는 (1-1, 2-2) 및 (1-2, 2-2)의 조합된 후보 서브코드벡터가 선택된 것으로 가정하였다.

[0075] 다음 마찬가지로, 조합된 후보 서브코드벡터 (1-1, 2-2) 및 (1-2, 2-2)와 세 번째 서브코드벡터(i=3)의 후보 서브코드벡터 (3-1) 및 (3-2)에 대해서 조합하여 총 4개의 조합된 후보 서브코드벡터를 생성한다. 즉, (1-1, 2-2, 3-1), (1-1, 2-2, 3-2), (1-2, 2-2, 3-1) 및 (1-2, 2-2, 3-2)의 4개의 조합된 후보 서브코드벡터가 생성된다.

[0076] 마찬가지로 조합 서브코드벡터 선택부(143)는 수학식 3 및 수학식 4를 이용하여 4개의 조합된 후보 서브코드벡터와 타겟 신호의 첫 번째, 두 번째, 세 번째 서브벡터가 조합된 서브벡터의 유사도를 계산하여, 유사도가 큰 2개의 조합된 후보 서브코드벡터 (1-1, 2-2, 3-2) 및 (1-2, 2-2, 3-1)를 선택한다.

[0077] 상기와 같은 방법을 반복적으로 수행하여, 조합 서브코드벡터 선택부(143)는 타겟 신호와 유사도가 높은 복수의

조합된 후보 서브코드벡터들을 선택해나갈 수 있다.

[0078] 위와 같은 과정을 정리하면, $i=1$ 인 경우, 첫 번째 후보 서브코드벡터와 두 번째 서브코드벡터가 조합되어 조합 서브코드벡터가 생성된다. 조합 서브코드벡터 선택부(143)는 조합된 후보 서브코드벡터 중에서 타겟 서브코드벡터와 유사도가 큰 조합된 후보 서브코드벡터를 복수 개 선택하는데, 이 경우 상기 수학식 3 및 수학식 4를 이용한다. 도 4를 참조하면 조합된 후보 서브코드벡터에서 유사도가 큰 조합된 서브코드벡터는 다시 $i=3$ 인 서브코드벡터와 조합된다.

[0079] 상기 방법을 반복함으로써, 조합된 후보 서브코드벡터 중에서 선택된 복수의 서브코드벡터는 $i=8$ 의 후보 서브코드벡터와 조합된다.

[0080] 이하에서도 도 5c를 이용하여 최종 조합 서브코드벡터를 선택하는 과정에 대해 설명한다.

[0081] 최종 조합 서브코드벡터 생성부(144)는 도 5c와 같이 상기 조합 서브코드벡터 선택부에 의해 기 조합된 후보 서브코드벡터와 후보 서브코드벡터 선택부(142)에 의해 선택된 복수의 마지막 번째 후보 서브코드벡터($(8-1)$, $(8-2)$)를 모든 경우의 수만큼 조합하고, 4개의 조합된 후보 서브코드벡터 중에서 타겟 신호(첫 번째부터 8번째 서브코드벡터가 모두 조합된 서브코드벡터)와 유사도가 가장 큰 최종 조합 서브코드벡터를 선택한다(S244).

[0082] 여기서 최종 조합 서브코드벡터 생성부(144)는 최종 조합 서브코드벡터를 선택하기 위하여 아래 수학식 5 및 수학식 6을 이용한다.

수학식 5

$$g_o = \frac{y_z^k T \overline{y^k}}{y^k T y^k}$$

[0083]

수학식 6

$$L^* = \operatorname{argmin}(y_z - g_o^L \overline{y^L})$$

[0084]

[0085] 여기서, g_o 는 상기 타겟 서브벡터와 상기 조합된 후보 서브코드벡터의 계인이며, $\overline{y^k}$ 는 조합된 코드북이고, y_z^k 는 고유벡터 행렬을 음성 신호의 영상태응답에 곱해서 구한 KLT 변환된 타겟 신호이고, L^* 은 타겟 서브벡터(y_z)와 상기 조합된 후보 서브코드벡터 사이의 유사도이다.

[0086] 다음으로 S250단계에서 출력부(160)는, 아래 수학식 7과 같이 상기 최종 조합 서브코드벡터의 계인 및 형태를 이용하여 상기 타겟 신호로부터 상기 양자화된 타겟 신호를 생성한다.

수학식 7

$$\hat{y}_z = g_o^{L^*} \overline{y^{L^*}}$$

[0087]

[0088] 여기서, g_o 는 최종 조합 서브코드벡터의 계인이며, $\overline{y}$ 는 최종 조합 서브코드벡터이다.

[0089] 한편, 도 5a 내지 도 5c에서는 각각의 서브코드벡터가 1차원으로 이루어진 것으로 도시하였으나, 5차원 이상의 고차원으로 이루어질 수 있다.

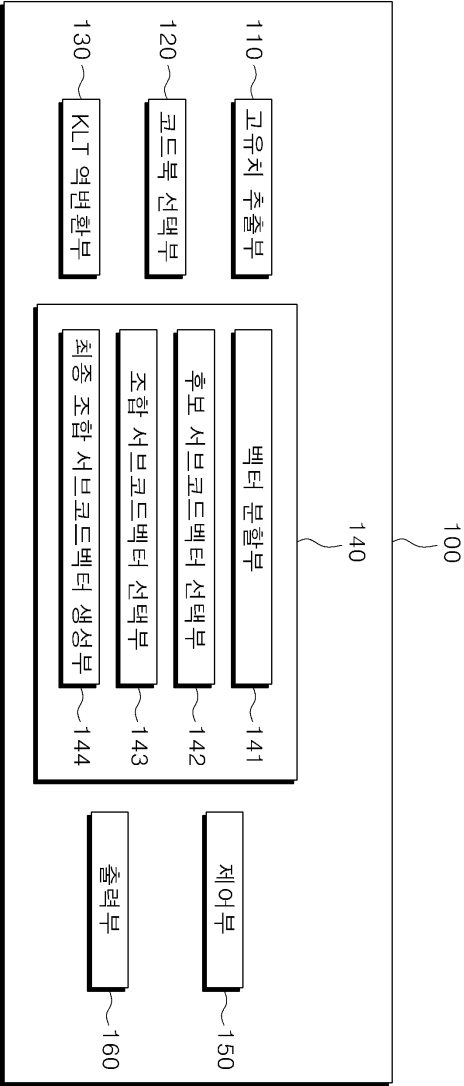
- [0090] 이상 음성 신호의 벡터 양자화 방법에 대하여 설명하였으며,
- [0091] 종래 기술에 따른 Full search와 M search의 경우 계산량의 결과 값에 서브코드벡터의 수를 나타내는 i 가 지수에 포함되어 있다. 즉, Full search와 M search에서 벡터 차원의 증가에 따라 서브코드벡터의 개수가 늘어날수록 계산량이 지수적으로 증가함을 알 수 있다. 이에 반하여 본 발명의 실시예에 따르면 I 가 지수에 포함되지 않기 때문에 상기 두 방법에 비해서 계산량이 줄어든 것을 알 수 있다.
- [0092] 따라서 본 발명의 실시예에 따르면 모든 서브코드벡터를 한꺼번에 조합하는 방식이 아니라, 낮은 차원의 서브코드벡터부터 복수 개, 이를 테면 2개씩 차례대로 조합을 하는 방식을 사용함으로써, 기존의 알고리즘에 비하여 계산량을 줄일 수 있다.
- [0093] 이와 같이 본 발명의 실시예에 따른 음성 신호의 벡터 양자화 장치 및 그 방법에 따르면, 기존의 KLT-CVQ 음성 코더의 단점인 하나의 프레임에 사용 가능한 샘플 수가 5개로 한정되는 점과 피치(pitch) 정보를 효율적으로 이용하지 못하는 점을 보완하기 위한 것으로, 하나의 프레임에 많은 샘플을 사용함으로써 피치 정보를 효율적으로 이용가능하고 코드벡터 서치 과정에서 계산량을 줄일 수 있으며 우수한 품질의 음성 서비스를 제공할 수 있다.
- [0094] 이제까지 본 발명에 대하여 실시예들을 중심으로 살펴보았다. 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자는 본 발명의 본질적인 특성에서 벗어나지 않는 범위에서 변형된 형태로 구현될 수 있음을 이해할 수 있을 것이다. 그러므로 개시된 실시예들은 한정적인 관점이 아니라 설명적인 관점에서 고려되어야 한다. 따라서 본 발명의 범위는 전술한 실시예에 한정되지 않고 특허청구범위에 기재된 내용 및 그와 동등한 범위 내에 있는 다양한 실시 형태가 포함되도록 해석되어야 할 것이다.

부호의 설명

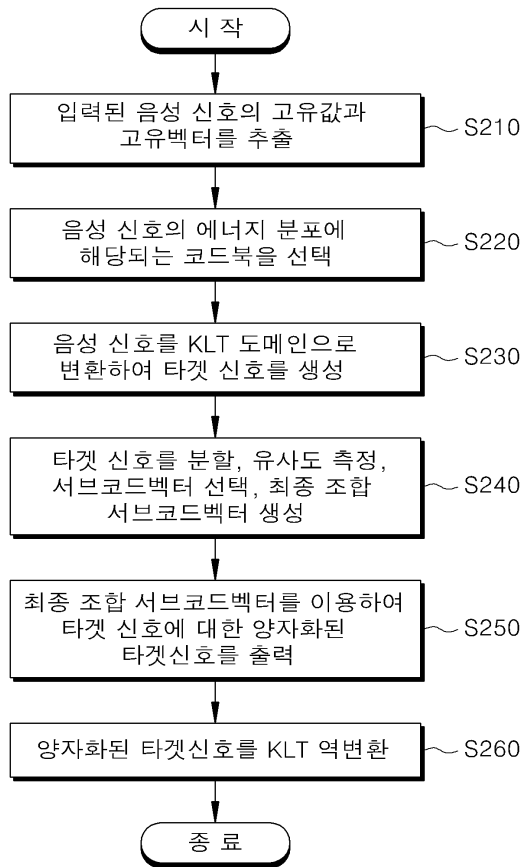
- [0095]
- | | |
|-----------------------|--------------------|
| 100: 음성 신호의 벡터 양자화 장치 | 110: 고유치 추출부 |
| 120: 코드북 선택부 | 130: KLT 변환부 |
| 140: 제어부 | 141: 벡터 분할부 |
| 142: 후보 서브코드벡터 선택부 | 143: 조합 서브코드벡터 선택부 |
| 144: 최종 조합 서브코드벡터 생성부 | 150: 출력부 |
| 160: KLT 역변환부 | |

도면

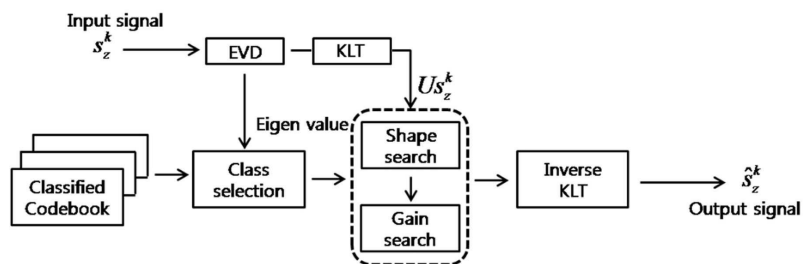
도면1



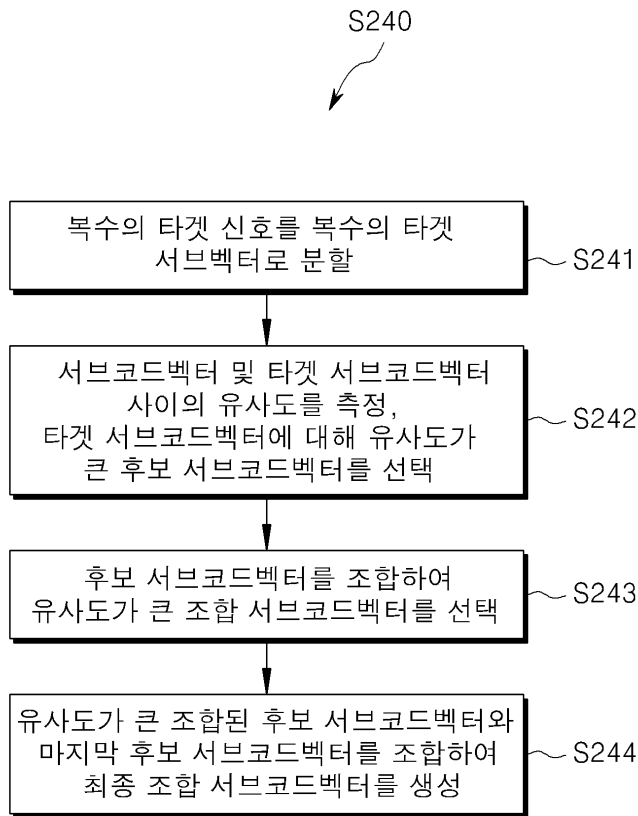
도면2



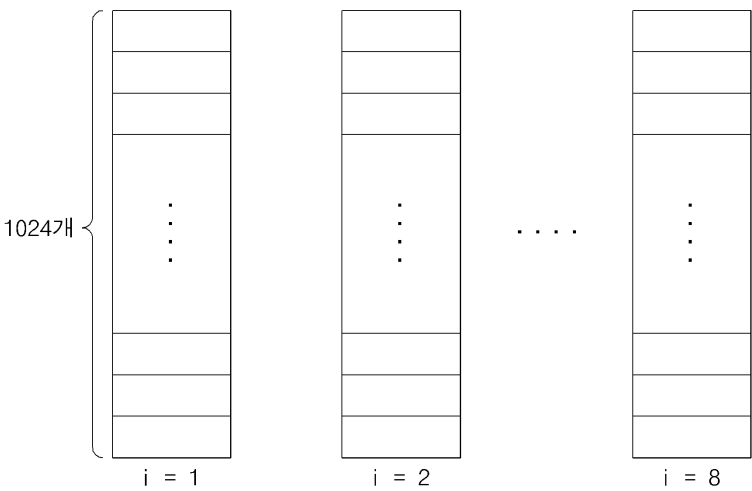
도면3



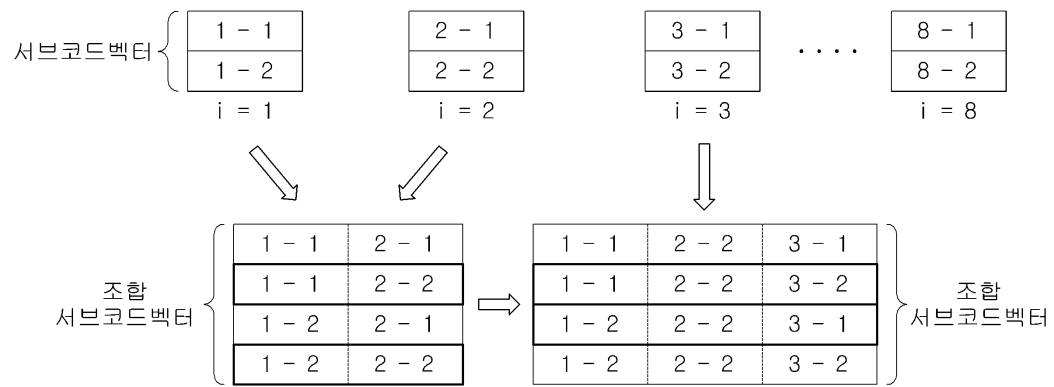
도면4



도면5a



도면5b



도면5c

