



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2013-0140403
(43) 공개일자 2013년12월24일

(51) 국제특허분류(Int. Cl.)

G10L 19/12 (2006.01)

(21) 출원번호 10-2012-0063776

(22) 출원일자 2012년06월14일

심사청구일자 2012년06월14일

(71) 출원인

세종대학교산학협력단

서울특별시 광진구 능동로 209 (군자동, 세종대학교)

(72) 발명자

김무영

서울특별시 강남구 도곡2동 도곡테크아파트 210-402

노명훈

서울특별시 성동구 성수2가3동 성수현대아파트 107동 401호

이윤주

서울특별시 성북구 석관1동 두산아파트 121동 2102호

(74) 대리인

특허법인태백

전체 청구항 수 : 총 16 항

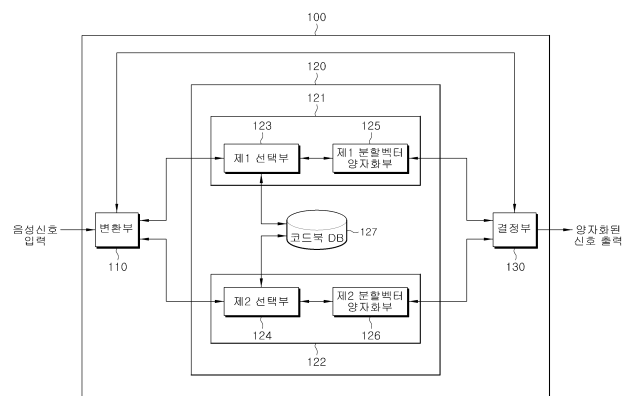
(54) 발명의 명칭 가우시안 혼합 모델을 기반으로 하는 스위치 분할 벡터 양자화 방법 및 그 장치

(57) 요약

본 발명은 가우시안 혼합 모델을 기반으로 하는 스위치 분할 벡터 양자화 방법 및 그 장치에 관한 것으로, 본 발명의 일 실시예에 따른 가우시안 혼합 모델을 기반으로 하는 스위치 분할 벡터 양자화 장치의 스위치 분할 벡터 양자화 방법은, 입력된 음성 신호의 각 프레임에 대한 선형 스펙트럼 주파수를 가우시안 혼합 모델을 기반으로 하는 입력 벡터로 변환하는 단계와, 상기 입력 벡터 또는 상기 입력 벡터의 예측 에러 벡터 각각을 스위치 분할 벡터 양자화하여 복수의 양자화 벡터를 생성하는 단계와, 상기 생성된 복수의 양자화 벡터를 상기 입력 벡터와 비교하여 기 설정된 유사도를 가지는 하나의 최종 양자화 벡터를 결정하는 단계를 포함한다.

이에 따라 음성 신호에 대한 프레임에 대해 인터 프레임 상관 관계와 인트라 프레임 상관 관계를 이용함으로써 채널 음성 신호의 화자 인식 성능을 향상시킬 수 있다.

대표도 - 도1



이 발명을 지원한 국가연구개발사업

과제고유번호	2011-0002502
부처명	교육과학기술부
연구사업명	기본연구지원사업
연구과제명	차세대 Seamless 사운드 코덱 표준화를 위한 핵심기초기술 연구 및 통합 소프트웨어 시뮬레이터 구현
기 여 율	1/1
주관기관	세종대학교 산학협력단
연구기간	2011.05.01 ~ 2012.04.30

특허청구의 범위

청구항 1

가우시안 혼합 모델을 기반으로 하는 스위치 분할 벡터 양자화 장치의 스위치 분할 벡터 양자화 방법에 있어서, 입력된 음성 신호의 각 프레임에 대한 선형 스펙트럼 주파수를 가우시안 혼합 모델을 기반으로 하는 입력 벡터로 변환하는 단계;

상기 입력 벡터 또는 상기 입력 벡터의 예측 에러 벡터 각각을 스위치 분할 벡터 양자화하여 복수의 양자화 벡터를 생성하는 단계; 및

상기 생성된 복수의 양자화 벡터를 상기 입력 벡터와 비교하여 기 설정된 유사도를 가지는 하나의 최종 양자화 벡터를 결정하는 단계를 포함하는 가우시안 혼합 모델을 기반으로 하는 스위치 분할 벡터 양자화 방법.

청구항 2

제1항에 있어서,

상기 복수의 양자화 벡터를 생성하는 단계는,

상기 입력 벡터의 예측 에러 벡터 각각을 KLT(Karhunen-Loeve Transform) 도메인으로 변환하여 분할 벡터 양자화하는 가우시안 혼합 모델을 기반으로 하는 스위치 분할 벡터 양자화 방법.

청구항 3

제1항에 있어서,

상기 복수의 양자화 벡터를 생성하는 단계는,

벡터 양자화 코드북으로부터 상기 입력 벡터와 기 설정된 유사도를 가지는 제1 코드 벡터를 선택하고, 상기 제1 코드 벡터를 복수의 서브 벡터로 분할하여 양자화한 값인 제1 양자화 벡터를 생성하는 단계; 및

상기 벡터 양자화 코드북으로부터 상기 예측 에러 벡터와 기 설정된 유사도를 가지는 제2 코드 벡터를 선택하고, 상기 제2 코드 벡터를 복수의 서브 벡터로 분할하여 양자화한 값인 제2 양자화 벡터를 생성하는 단계를 포함하는 가우시안 혼합 모델을 기반으로 하는 스위치 분할 벡터 양자화 방법.

청구항 4

제3항에 있어서,

상기 제2 양자화 벡터를 생성하는 단계는,

상기 음성 신호의 현재 프레임의 예측 에러 벡터를 양자화한 값과, 이전 프레임에서 생성된 최종 양자화 벡터의 값을 합산하여 상기 제2 양자화 벡터를 생성하는 가우시안 혼합 모델을 기반으로 하는 스위치 분할 벡터 양자화 방법.

청구항 5

제1항에 있어서,

상기 예측 에러 벡터는,

상기 음성 신호의 현재 프레임의 입력 벡터와 이전 프레임의 입력 벡터를 이용하여 생성하는 가우시안 혼합 모델을 기반으로 하는 스위치 분할 벡터 양자화 방법.

청구항 6

제1항에 있어서,

상기 입력 벡터는,

상기 입력된 음성 신호의 각 프레임에 대한 선형 스펙트럼 주파수를 EM(Expectation Maximum) 알고리즘을 이용하여 생성된 가우시안 혼합 모델인 가우시안 혼합 모델을 기반으로 하는 스위치 분할 벡터 양자화 방법.

청구항 7

제3항에 있어서,

상기 제1 코드 벡터 또는 제2 코드 벡터는,

최대 우도 추정 알고리즘을 이용하여 선택되는 가우시안 혼합 모델을 기반으로 하는 스위치 분할 벡터 양자화 방법.

청구항 8

제1항에 있어서,

상기 하나의 최종 양자화 벡터를 결정하는 단계는,

상기 최종 양자화 벡터에 의해 할당되는 i 번째 클러스터의 n 차원을 가지는 비트($B_{i,n}$)는 다음의 수학적식을 이용하여 계산하는 가우시안 혼합 모델을 기반으로 하는 스위치 분할 벡터 양자화 방법:

$$B_{i,n} = \frac{k_n}{k} B_i + \frac{k_n}{2} \log_2 \frac{\prod_{p=1}^n \lambda_{i,l}^{1/k_n}}{\left[\prod_{j=1}^N \left(\frac{G_{i,k_j}}{k_j} \right)^{k_j} \right]^{1/k} \kappa_i}$$

여기서, B_i 는 i 번째 클러스터, k는 분할 벡터, $G_{i,k}$ 는 k-차원의 도형의 정규화된 2차 모멘트를 나타낸다.

청구항 9

입력된 음성 신호의 각 프레임에 대한 선형 스펙트럼 주파수를 가우시안 혼합 모델을 기반으로 하는 입력 벡터로 변환하는 변환부;

상기 입력 벡터 또는 상기 입력 벡터의 예측 에러 벡터 각각을 스위치 분할 벡터 양자화하여 복수의 양자화 벡터를 생성하는 양자화부; 및

상기 생성된 복수의 양자화 벡터를 상기 입력 벡터와 비교하여 기 설정된 유사도를 가지는 하나의 최종 양자화 벡터를 결정하는 결정부를 포함하는 가우시안 혼합 모델을 기반으로 하는 스위치 분할 벡터 양자화 장치.

청구항 10

제9항에 있어서,

상기 양자화부는,

상기 입력 벡터 또는 상기 입력 벡터의 예측 에러 벡터 각각을 KLT(Karhunen-Loeve Transform) 도메인으로 변환하여 분할 벡터 양자화하는 가우시안 혼합 모델을 기반으로 하는 스위치 분할 벡터 양자화 장치.

청구항 11

제9항에 있어서,

상기 양자화부는,

벡터 양자화 코드북으로부터 상기 입력 벡터와 기 설정된 유사도를 가지는 제1 코드 벡터를 선택하고, 상기 제1 코드 벡터를 복수의 서브 벡터로 분할하여 양자화한 값인 제1 양자화 벡터를 생성하는 제1 양자화부; 및

상기 벡터 양자화 코드북으로부터 상기 예측 에러 벡터와 기 설정된 유사도를 가지는 제2 코드 벡터를 선택하고, 상기 제2 코드 벡터를 복수의 서브 벡터로 분할하여 양자화한 값인 제2 양자화 벡터를 생성하는 제2 양자화부를 포함하는 가우시안 혼합 모델을 기반으로 하는 스위치 분할 벡터 양자화 장치.

청구항 12

제11항에 있어서,

상기 제2 양자화부는,

상기 음성 신호의 현재 프레임의 예측 에러 벡터를 양자화한 값과, 이전 프레임에서 생성된 최종 양자화 벡터의 값을 합산하여 상기 제2 양자화 벡터를 생성하는 가우시안 혼합 모델을 기반으로 하는 스위치 분할 벡터 양자화 장치.

청구항 13

제9항에 있어서,

상기 예측 에러 벡터는,

상기 음성 신호의 현재 프레임의 입력 벡터와 이전 프레임의 입력 벡터를 이용하여 생성하는 가우시안 혼합 모델을 기반으로 하는 스위치 분할 벡터 양자화 장치.

청구항 14

제9항에 있어서,

상기 입력 벡터는,

상기 입력된 음성 신호의 각 프레임에 대한 선형 스펙트럼 주파수를 EM(Expectation Maximum) 알고리즘을 이용하여 생성된 가우시안 혼합 모델인 가우시안 혼합 모델을 기반으로 하는 스위치 분할 벡터 양자화 장치.

청구항 15

제11항에 있어서,

상기 제1 코드 벡터 또는 제2 코드 벡터는,

최대 우도 추정 알고리즘을 이용하여 선택되는 가우시안 혼합 모델을 기반으로 하는 스위치 분할 벡터 양자화 장치.

청구항 16

제9항에 있어서,

상기 결정부는,

상기 최종 양자화 벡터에 의해 할당되는 i 번째 클러스터의 n 차원을 가지는 비트($B_{i,n}$)는 다음의 수학적식을 이용하여 계산하는 가우시안 혼합 모델을 기반으로 하는 스위치 분할 벡터 양자화 장치:

$$B_{i,n} = \frac{k_n}{k} B_i + \frac{k_n}{2} \log_2 \frac{\frac{G_{i,k_n}}{k_n} \prod_{p=1}^n \lambda_{i,l}^{1/k_n}}{\left[\prod_{j=1}^N \left(\frac{G_{i,k_j}}{k_j} \right)^{k_j} \right]^{1/k} \kappa_i}$$

여기서, B_i 는 i 번째 클러스터, k 는 분할 벡터, $G_{i,k}$ 는 k -차원의 도형의 정규화된 2차 모멘트를 나타낸다.

명세서

기술분야

[0001] 본 발명은 가우시안 혼합 모델을 기반으로 하는 스위치 분할 벡터 양자화 방법 및 그 장치에 관한 것으로, 더욱 상세하게는 인터 프레임 상관관계와 인트라 프레임 상관관계를 이용하여 음성 신호를 양자화하는 기술이 개시된다.

배경기술

[0002] 음성 부호화기에서 고음질의 음성 부호화를 위해서 고음질의 음성 부호화를 위해서는 음성 신호의 단구간 상관도를 나타내는 선형 스펙트럼 주파수인 LSF(Line Spectral frequency) 계수를 효율적으로 양자화 하는 것이 매우 중요하다. LPC(Linear Predictive Coefficient) 필터의 최적 선형 예측 계수값은 입력 신호를 프레임 단위로 나누어 각 프레임별로 예측 오차의 에너지를 최소화시키는 개념으로 구해진다.

[0003] LPC에서 변환된 LSF 데이터는 코딩 효율성을 높이지만, LSF 데이터가 다차원으로 갈수록 복잡해진다는 문제점이 있다. VQ(Vector Quantization)의 차수가 커지면, 계산이 복잡해지고 메모리의 요구량도 증가하게 된다. 이러한 문제점을 해결하기 위해, SVQ(Split Vector Quantization)이 제안되었다.

[0004] LPC 필터의 계수를 직접 양자화할 경우, 필터의 특성이 계수의 양자화 오차에 매우 민감하고 계수 양자화 후의 LPC 필터의 안정성이 보장되지 않는 문제점이 있다. 따라서 LPC 계수를 양자화 성질이 좋은 다른 파라미터로 변환하여 양자화하여야 하며, 주로 반사 계수(reflection coefficient) 또는 LSF로 변환하여 양자화한다. 특히, LSF 값은 음성의 주파수 특성과 밀접하게 연관되는 성질이 있어 최근에 개발된 표준 음성 압축기들은 대부분 LSF 양자화 방법을 사용한다.

[0005] LSF 양자화 방법은 효율적인 양자화를 위하여 LSF 계수의 프레임간 상관관계를 이용한다. 즉, 현재 프레임의 LSF를 직접 양자화하지 않고 과거 프레임의 LSF 값 정보로부터 현재 프레임의 LSF를 예측하고 예측 오차를 양자화한다. LSF 값은 음성 신호의 주파수 특성과 밀접한 관계가 있으며, 따라서 시간적으로 예측이 가능하고 상당히 큰 예측 이득을 얻을 수 있다.

[0006] 이러한 방법은 재구성된 이전 프레임을 사용하기 때문에 채널 에러에 약한 문제가 있다. 만일 채널 내에서 오류가 발생하는 경우 이 오류로 인해 변형된 프레임은 다음에 오는 모든 데이터를 망칠 수 있게 된다. 제거를 위해 사용된 이전 프레임이 재구성된 프레임이기 때문이다. 따라서 채널에 민감한 실제 타임 시스템에서는 사용되기 어려운 문제점이 있다.

[0007] 본 발명의 배경이 되는 기술은 대한민국 공개특허공보 제10-2004-0078760호(2004. 09. 31)에 개시되어 있다.

발명의 내용

해결하려는 과제

[0008] 본 발명이 해결하고자 하는 기술적인 과제는 음성 신호에 대한 프레임에 대해 인터 프레임 상관 관계와 인트라 프레임 상관 관계를 이용하여 채널 음성 신호의 화자 인식 성능을 향상시키는 가우시안 혼합 모델을 기반으로 하는 스위치 분할 벡터 양자화 방법 및 그 장치를 제공하는 것을 목적으로 한다.

과제의 해결 수단

[0009] 본 발명의 일 실시예에 따른 가우시안 혼합 모델을 기반으로 하는 스위치 분할 벡터 양자화 장치의 스위치 분할 벡터 양자화 방법은, 입력된 음성 신호의 각 프레임에 대한 선형 스펙트럼 주파수를 가우시안 혼합 모델을 기반으로 하는 입력 벡터로 변환하는 단계와, 상기 입력 벡터 또는 상기 입력 벡터의 예측 에러 벡터 각각을 스위치 분할 벡터 양자화하여 복수의 양자화 벡터를 생성하는 단계와, 상기 생성된 복수의 양자화 벡터를 상기 입력 벡터와 비교하여 기 설정된 유사도를 가지는 하나의 최종 양자화 벡터를 결정하는 단계를 포함한다.

[0010] 또한, 상기 복수의 양자화 벡터를 생성하는 단계는, 상기 입력 벡터 또는 상기 입력 벡터의 예측 에러 벡터 각각을 KLT 도메인으로 변환하여 분할 벡터 양자화할 수 있다.

[0011] 또한, 상기 복수의 양자화 벡터를 생성하는 단계는, 벡터 양자화 코드북으로부터 상기 입력 벡터와 기 설정된 유사도를 가지는 제1 코드 벡터를 선택하고, 상기 제1 코드 벡터를 복수의 서브 벡터로 분할하여 양자화한 값인 제1 양자화 벡터를 생성하는 단계와, 상기 벡터 양자화 코드북으로부터 상기 예측 에러 벡터와 기 설정된 유사도를 가지는 제2 코드 벡터를 선택하고, 상기 제2 코드 벡터를 복수의 서브 벡터로 분할하여 양자화한 값인 제2 양자화 벡터를 생성하는 단계를 포함할 수 있다.

[0012] 또한, 상기 제2 양자화 벡터를 생성하는 단계는, 상기 음성 신호의 현재 프레임의 예측 에러 벡터를 양자화한 값과, 이전 프레임에서 생성된 최종 양자화 벡터의 값을 합산하여 상기 제2 양자화 벡터를 생성할 수 있다.

[0013] 또한, 상기 예측 에러 벡터는, 상기 음성 신호의 현재 프레임의 입력 벡터와 이전 프레임의 입력 벡터를 이용하여 생성할 수 있다.

[0014] 또한, 상기 입력 벡터는, 상기 입력된 음성 신호의 각 프레임에 대한 선형 스펙트럼 주파수를 EM(Expectation Maximum) 알고리즘을 이용하여 생성된 가우시안 혼합 모델일 수 있다.

[0015] 또한, 상기 제1 코드 벡터 또는 제2 코드 벡터는, 최대 우도 추정 알고리즘을 이용하여 선택될 수 있다.

[0016] 또한, 상기 하나의 최종 양자화 벡터를 결정하는 단계는, 상기 최종 양자화 벡터에 의해 할당되는 i 번째 클러스터의 n 차원을 가지는 비트($B_{i,n}$)는 다음의 수학식을 이용하여 계산할 수 있다:

$$B_{i,n} = \frac{k_n}{k} B_i + \frac{k_n}{2} \log_2 \frac{\frac{G_{i,k_n}}{k_n} \prod_{p=1}^n \lambda_{i,l}^{1/k_n}}{\left[\prod_{j=1}^N \left(\frac{G_{i,k_j}}{k_j} \right)^{k_j} \right]^{1/k} \kappa_i}$$

[0017]

[0018] 여기서, B_i 는 i 번째 클러스터, k는 분할 벡터, $G_{i,k}$ 는 k-차원의 도형의 정규화된 2차 모멘트를 나타낸다.

[0019] 본 발명의 또 다른 실시예에 따른 가우시안 혼합 모델을 기반으로 하는 스위치 분할 벡터 양자화 장치는, 입력된 음성 신호의 각 프레임에 대한 선형 스펙트럼 주파수를 가우시안 혼합 모델을 기반으로 하는 입력 벡터로 변환하는 변환부와, 상기 입력 벡터 또는 상기 입력 벡터의 예측 에러 벡터 각각을 스위치 분할 벡터 양자화하여 복수의 양자화 벡터를 생성하는 양자화부와, 상기 생성된 복수의 양자화 벡터를 상기 입력 벡터와 비교하여 기 설정된 유사도를 가지는 하나의 최종 양자화 벡터를 결정하는 결정부를 포함한다.

발명의 효과

[0020] 이와 같이 본 발명의 실시예에 따르면 음성 신호에 대한 프레임에 대해 인터 프레임 상관 관계와 인트라 프레임 상관 관계를 이용함으로써 채널 음성 신호의 화자 인식 성능을 향상시킬 수 있다.

도면의 간단한 설명

[0021] 도 1은 본 발명의 일 실시예에 따른 가우시안 혼합 모델을 기반으로 하는 스위치 분할 벡터 양자화 장치의 구성도,

도 2는 도 1에 따른 가우시안 혼합 모델을 기반으로 하는 스위치 분할 벡터 양자화 방법의 흐름도,

도 3은 본 발명의 또 다른 실시예에 따른 가우시안 혼합 모델을 기반으로 하는 스위치 분할 벡터 양자화 장치의 구성도,

도 4는 도 1 및 도 3에 따른 가우시안 혼합 모델을 기반으로 하는 스위치 분할 벡터 양자화 장치의 입력 벡터 및 그에 대응하는 양자화 벡터를 설명하기 위한 예시도이다.

발명을 실시하기 위한 구체적인 내용

[0022] 이하, 첨부된 도면들을 참조하여 본 발명의 실시예를 상세하게 설명한다. 사용되는 용어들은 실시예에서의 기능을 고려하여 선택된 용어들로서, 그 용어의 의미는 사용자, 운용자의 의도 또는 관례 등에 따라 달라질 수 있다. 그러므로 후술하는 실시예들에서 사용된 용어의 의미는, 본 명세서에 구체적으로 정의된 경우에는 그 정의에 따르며, 구체적인 정의가 없는 경우는 당업자들이 일반적으로 인식하는 의미로 해석되어야 할 것이다.

[0023] 도 1은 본 발명의 일 실시예에 따른 가우시안 혼합 모델을 기반으로 하는 스위치 분할 벡터 양자화 장치의 구성도이고, 도 2는 도 1에 따른 가우시안 혼합 모델을 기반으로 하는 스위치 분할 벡터 양자화 방법의 흐름도이다.

[0024] 도 1 및 도 2를 참조하면, 본 발명의 일 실시예에 따른 가우시안 혼합 모델을 기반으로 하는 스위치 분할 벡터 양자화 장치(100)는 변환부(110), 양자화부(120) 및 결정부(130)를 포함한다. 먼저 변환부(110)는 입력된 음성 신호의 각 프레임에 대한 선형 스펙트럼 주파수(LSF : Line Spectral frequency)를 가우시안 혼합 모델(GMM : Gaussian Mixture Model)을 기반으로 하는 입력 벡터로 변환한다(S210). 이 경우, 음성 신호는 복수의 프레임으로 구성될 수 있으며, 입력 벡터는 각각의 프레임에 선형 스펙트럼 주파수를 EM(Expectation Maximum) 알고리즘을 이용하여 생성된 가우시안 혼합 모델일 수 있다.

[0025] 또한, 변환부(110)는 입력 벡터의 예측 에러 벡터를 생성할 수 있다. 여기서, 예측 에러 벡터는 음성 신호의 현재 프레임의 입력 벡터와 이전 프레임의 입력 벡터를 이용하여 생성할 수 있다. 이는 LSF 계수의 프레임 간의 상관관계를 이용하는 방식인 인터 프레임 상관 관계를 통해 예측 에러 벡터를 구할 수 있다. 즉, 변환부(110)는 현재 프레임의 LSF를 직접 양자화하지 않고 과거 프레임의 LSF 값 정보로부터 현재 프레임의 LSF를 예측한 예측 에러를 구하게 된다. 여기서, LSF 값은 음성 신호의 주파수 특성과 밀접한 관계가 있으므로 시간적으로 예측이 가능하고 상당히 큰 예측 이득을 얻을 수 있다.

[0026] 다음으로, 양자화부(120)는 입력 벡터 또는 입력 벡터의 예측 에러 벡터 각각을 스위치 분할 벡터 양자화(SSVQ : Switched Split Vector Quantization)하여 복수의 양자화 벡터를 생성한다(S220). 즉, 양자화부(120)는 입력 벡터를 이용한 양자화 벡터와 예측 에러 벡터를 이용한 양자화 벡터를 생성하게 된다. 분할 벡터 양자화 알고리즘은 입력된 음성 신호의 프레임을 복수의 부차원(sub-dimension)으로 분할하여, 각각의 부차원에 대해 벡터 양자화를 하는 기술이다.

[0027] 구체적으로 양자화부(120)는 제1 양자화부(121), 제2 양자화부(122) 및 코드북 DB(127)를 포함하며, 본 명세서에서는 제1 양자화부(121)는 입력 벡터를 스위치 분할 벡터 양자화하며, 제2 양자화부(122)는 예측 에러 벡터를 스위치 분할 벡터 양자화하는 구성을 나타낸다. 코드북 DB(127)는 복수의 코드 벡터 $\{\mu_{m,X}\}_{m=1}^N$ 가 저장되어 있다.

예를 들어, 제1 양자화부(121)는 코드북 DB(127)로부터 입력 벡터 x^k 의 각각의 값에 대해 각 코드북의 k_1 내지 k_n 번째 코드 벡터들과 비교하여 가장 유사도가 높은 코드 벡터 i 를 선택하고, 제2 양자화부(122)는 예측 에러

벡터 e^k 의 각각의 값에 대해 각 코드북의 k_1 내지 k_n 번째 코드 벡터들과 비교하여 가장 유사도가 높은 코드 벡터 i 를 선택한다.

[0028] 제1 양자화부(121)는 제1 선택부(123) 및 제1 분할 벡터 양자화부(125)를 포함한다. 제1 선택부(123)는 벡터 양자화 코드북인 코드북 DB(127)로부터 입력 벡터와 기 설정된 유사도를 가지는 제1 코드 벡터를 선택한다. 여기서, 제1 코드 벡터는 최대 우도 추정(Maximum Likelihood Estimation) 알고리즘을 이용하여 선택된다. 제1 분할 벡터 양자화부(125)는 제1 선택부(123)에서 선택된 제1 코드 벡터를 복수의 서브 벡터로 분할하고, 각각의 서브 벡터를 양자화하여 제1 양자화 벡터를 생성한다.

[0029] 한편, 제2 양자화부(122)는 제2 선택부(124) 및 제2 분할 벡터 양자화부(126)를 포함한다. 제2 선택부(124)는 벡터 양자화 코드북인 코드북 DB(127)로부터 예측 에러 벡터와 기 설정된 유사도를 가지는 제2 코드 벡터를 선택한다. 여기서, 제2 코드 벡터는 최대 우도 추정(Maximum Likelihood Estimation) 알고리즘을 이용하여 선택된다. 제2 분할 벡터 양자화부(126)는 제2 선택부(124)에서 선택된 제2 코드 벡터를 복수의 서브 벡터로 분할하고, 각각의 서브 벡터를 양자화하여 제2 양자화 벡터를 생성한다.

[0030] 또한, 제2 양자화부(122)는 음성 신호의 현재 프레임의 예측 에러 벡터를 양자화한 값과, 이전 프레임에서 생성된 최종 양자화 벡터의 값을 합산하여 제2 양자화 벡터를 생성할 수 있다. 즉, 이전 프레임과 현재 프레임 간의 인터 프레임 상관 관계를 이용하여 제2 양자화 벡터를 생성한다.

[0031] 마지막으로, 결정부(130)는 양자화부로부터 생성된 복수의 양자화 벡터를 입력 벡터와 비교하여 기 설정된 유사도를 가지는 하나의 최종 양자화 벡터를 결정한다(S230). 즉, 음성 신호에 대한 동일한 프레임 간의 상관 관계를 구하는 인트라 프레임 상관 관계(intra frame correlation) 알고리즘을 이용하게 된다. 따라서, 결정부(130)는 양자화부로부터 생성된 제1 양자화 벡터와 제2 양자화 벡터 중 입력 벡터와 가장 근접한 유사도를 가지는 양자화 벡터를 최종 양자화 벡터로 결정하며, 최종 양자화 벡터를 이용하여 음성 신호에 대한 비트를 할당하게 된다.

[0032] 또한, 결정부(130)는 전체 비트(B_{tot})를 각각의 클러스터(cluster) 및 하위 차원(subdimension)에 할당할 수 있다. 여기서, 원본 음성 데이터는 복수의 코드 벡터인 클러스터로 나뉘지며, 각 클러스터는 복수의 하위 차원으로 분할된다. 결정부(130)는 i 번째 클러스터에 할당되는 비트(B_i)를 다음의 수학적 식 1을 이용하여 계산할 수 있다.

수학적 식 1

$$B_i = B_{tot} + \log_2 \frac{(\alpha_i \beta_i \kappa_i)^{k/(k+2)}}{\sum_{j=1}^I (\alpha_j \beta_j \kappa_j)^{k/(k+2)}}$$

[0033]

[0034] 수학적 식 1에서, $\beta_i = k \left[\prod_{n=1}^N (G_{i,kn}/k_n)^{kn} \right]^{1/k}$, $\kappa_i = \left(\prod_{n=1}^N (\lambda_{i,1})^{1/k} \right)^{1/k}$, $G_{i,k}$ 는 k 차원의 도형의 정규화된 2차 모멘트를 나타낸다.

[0035] 또한, 결정부(130)는 최종 양자화 벡터에 의해 할당되는 i 번째 클러스터의 n 차원을 가지는 비트($B_{i,n}$)는 다음의 수학적 식 2를 이용하여 계산할 수 있다.

수학식 2

$$B_{i,n} = \frac{k_n}{k} B_i + \frac{k_n}{2} \log_2 \frac{\frac{G_{i,k_n}}{k_n} \prod_{p=1}^n \lambda_{i,l}^{1/k_n}}{\left[\prod_{j=1}^N \left(\frac{G_{i,k_j}}{k_j} \right)^{k_j} \right]^{1/k} \kappa_i}$$

[0036]

[0037]

수학식 2에서, B_i 는 i 번째 클러스터, k 는 분할 벡터, $G_{i,k}$ 는 k -차원의 도형의 정규화된 2차 모멘트를 나타낸다.

[0038]

도 3은 본 발명의 또 다른 실시예에 따른 가우시안 혼합 모델을 기반으로 하는 스위치 분할 벡터 양자화 장치의 구성도이다.

[0039]

도 3을 참조하면, 본 발명의 또 다른 실시예에 따른 가우시안 혼합 모델을 기반으로 하는 스위치 분할 벡터 양자화 장치(300)는 변환부(310), 양자화부(320) 및 결정부(330)를 포함한다. 이 경우, 변환부(310) 및 결정부(330)의 기능은 앞서 설명한 도 1의 스위치 분할 벡터 양자화 장치의 변환부(110) 및 결정부(130)와 동일하므로 이에 대한 설명은 생략하기로 한다. 이하, 도 3의 스위치 분할 벡터 양자화 장치의 양자화부(320)에 대해서 도 1의 스위치 분할 벡터 양자화 장치(100)의 양자화부(120)와의 차이점을 중심으로 설명하도록 한다.

[0040]

양자화부(320)는 입력 벡터 또는 입력 벡터의 예측 에러 벡터 각각을 KLT(Karhunen-Loeve Transform) 도메인으로 변환하여 분할 벡터 양자화할 수 있다. KLT는 원 음성 신호를 작은 크기의 서브 블록으로 분할하여 변환하는 방법 중 하나이다. 보다 구체적으로 양자화부(320)는 제1 양자화부(321), 제2 양자화부(322) 및 코드북 DB(329)를 포함한다. 제1 양자화부(321)에서는 입력 벡터에 대한 양자화 벡터를 구하고, 제2 양자화부(322)에서는 입력 벡터의 예측 에러 벡터에 대한 양자화 벡터를 구한다.

[0041]

제1 양자화부(321)는 제1 선택부(323), 제1 KLT 변환부(325-1), 제1 분할 벡터 양자화부(327) 및 제1 KLT 역변환부(325-2)를 포함한다. 제1 선택부(323)는 벡터 양자화 코드북인 코드북 DB(329)로부터 입력 벡터와 기 설정된 유사도를 가지는 제1 코드 벡터를 선택한다. 여기서, 제1 코드 벡터는 최대 우도 추정(Maximum Likelihood Estimation) 알고리즘을 이용하여 선택된다. 제1 KLT 변환부(325-1)는 제1 코드 벡터를 KLT 도메인으로 변환한다. 제1 분할 벡터 양자화부(327)는 KLT 변환된 제1 코드 벡터를 복수의 서브 벡터로 분할하고, 각각의 서브 벡터를 양자화한다. 제1 KLT 역변환부(325-2)는 제1 분할 벡터 양자화부(327)에서 양자화된 벡터를 KLT 역변환하여 원래의 도메인으로 변환시켜 제1 양자화 벡터를 생성한다. 따라서, 제1 양자화 벡터는 KLT 도메인에서 분할 벡터 양자화 처리를 거친 값이 된다.

[0042]

한편, 제2 양자화부(322)는 제2 선택부(324), 제2 KLT 변환부(326-1), 제2 분할 벡터 양자화부(328) 및 제2 KLT 역변환부(326-2)를 포함한다. 제2 선택부(324)는 벡터 양자화 코드북인 코드북 DB(329)로부터 예측 에러 벡터와 기 설정된 유사도를 가지는 제2 코드 벡터를 선택한다. 여기서, 제2 코드 벡터는 최대 우도 추정(Maximum Likelihood Estimation) 알고리즘을 이용하여 선택된다. 제2 KLT 변환부(326-1)는 제2 코드 벡터를 KLT 도메인으로 변환한다. 제2 분할 벡터 양자화부(328)는 KLT 변환된 제2 코드 벡터를 복수의 서브 벡터로 분할하고, 각각의 서브 벡터를 양자화한다. 제2 KLT 역변환부(326-2)는 제2 분할 벡터 양자화부(328)에서 양자화된 벡터를 KLT 역변환하여 원래의 도메인으로 변환시켜 제2 양자화 벡터를 생성한다. 따라서, 제2 양자화 벡터는 KLT 도메인에서 분할 벡터 양자화 처리를 거친 값이 된다.

[0043]

또한, 제2 양자화부(322)는 음성 신호의 현재 프레임의 예측 에러 벡터를 양자화한 값과, 이전 프레임에서 생성

된 최종 양자화 벡터의 값을 합산하여 제2 양자화 벡터를 생성할 수 있다. 즉, 이전 프레임과 현재 프레임 간의 인터 프레임 상관 관계를 이용하여 제2 양자화 벡터를 생성한다.

[0044] 한편, 결정부(330)는 양자화부(320)로부터 생성된 복수의 양자화 벡터를 입력 벡터와 비교하여 기 설정된 유사도를 가지는 하나의 최종 양자화 벡터를 결정하는 것으로, 구체적인 설명은 도 1의 스위치 분할 벡터 양자화 장치와 동일하므로 생략하기로 한다.

[0045] 도 4는 도 1 및 도 3에 따른 가우시안 혼합 모델을 기반으로 하는 스위치 분할 벡터 양자화 장치의 입력 벡터 및 그에 대응하는 양자화 벡터를 설명하기 위한 예시도이다.

[0046] 도 4를 참조하면, 음성 신호에 대하여 SVQ(Split Vector Quantization) 알고리즘을 적용하여 양자화한 경우(a)와, 도 1에 따른 가우시안 혼합 모델을 기반으로 하는 스위치 스플릿 벡터 양자화 알고리즘인 GMM-SSVQ(Gaussian Mixture Model-Switched Split Vector Quantization)을 적용하여 양자화한 경우(b) 및 도 3에 따른 가우시안 혼합 모델 KLT 도메인 기반의 스위치 스플릿 벡터 양자화 알고리즘인 GMM-KLT-SSVQ(Gaussian Mixture Model-Karhunen Loeve Transform-Switched Split Vector Quantization)을 적용하여 양자화한 경우(c)를 나타낸다. 음성 신호에 대해 SVQ를 적용하여 양자화한 경우에 비해 GMM-SSVQ, GMM-KLT-SSVQ를 적용한 경우 원본 음성 데이터에 더욱 근접한 양자화된 데이터에 비트를 할당할 수 있다.

[0047] 이와 같이 본 발명의 실시예에 따르면 음성 신호에 대한 프레임에 대해 인터 프레임 상관 관계와 인트라 프레임 상관 관계를 이용함으로써 채널 음성 신호의 화자 인식 성능을 향상시킬 수 있다.

[0048] 이상에서 본 발명은 도면을 참조하면서 기술되는 바람직한 실시예를 중심으로 설명되었지만 이에 한정되는 것은 아니다. 따라서 본 발명은 기재된 실시예로부터 도출 가능한 자명한 변형예를 포괄하도록 의도된 특허청구범위의 기재에 의해 해석되어야 한다.

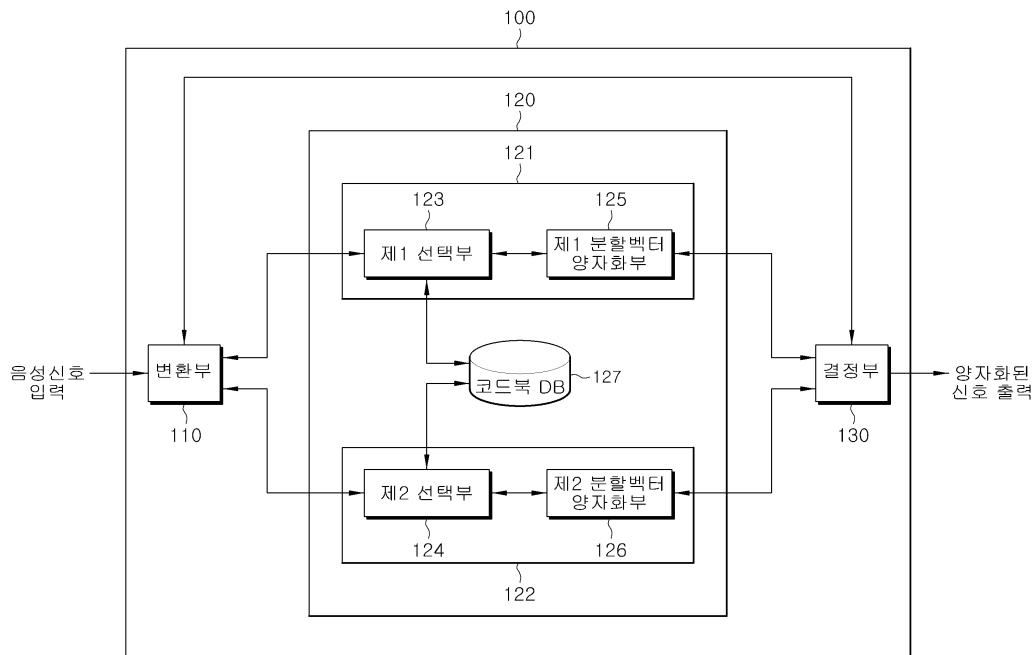
부호의 설명

[0049] 100 : 스위치 분할 벡터 양자화 장치
 110 : 변환부
 120 : 양자화부
 121 : 제1 양자화부
 122 : 제2 양자화부
 123 : 제1 선택부
 124 : 제2 선택부
 125 : 제1 분할 벡터 양자화부
 126 : 제2 분할 벡터 양자화부
 127 : 코드북 DB
 130 : 결정부
 300 : 스위치 분할 벡터 양자화 장치
 310 : 변환부
 320 : 양자화부
 321 : 제1 양자화부
 322 : 제2 양자화부

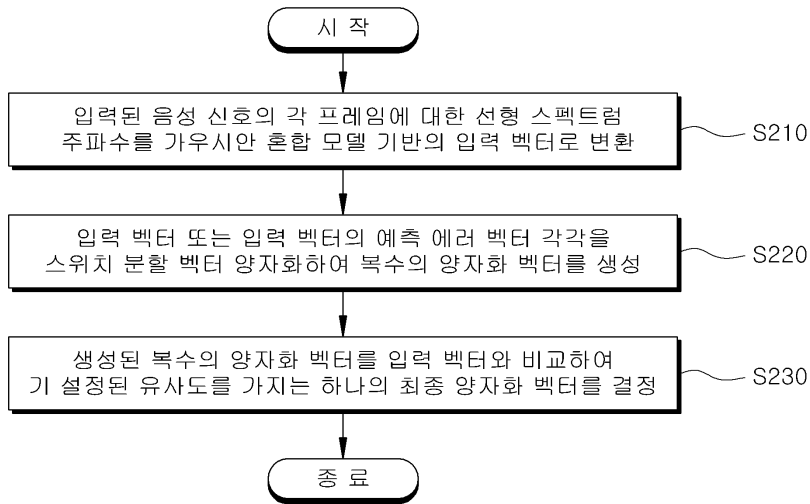
- 323 : 제1 선택부
- 324 : 제2 선택부
- 325-1 : 제1 KLT 변환부
- 325-2 : 제1 KLT 역변환부
- 326-1 : 제2 KLT 변환부
- 326-2 : 제2 KLT 역변환부
- 327 : 제1 분할 벡터 양자화부
- 328 : 제2 분할 벡터 양자화부
- 329 : 코드북 DB
- 330 : 결정부

도면

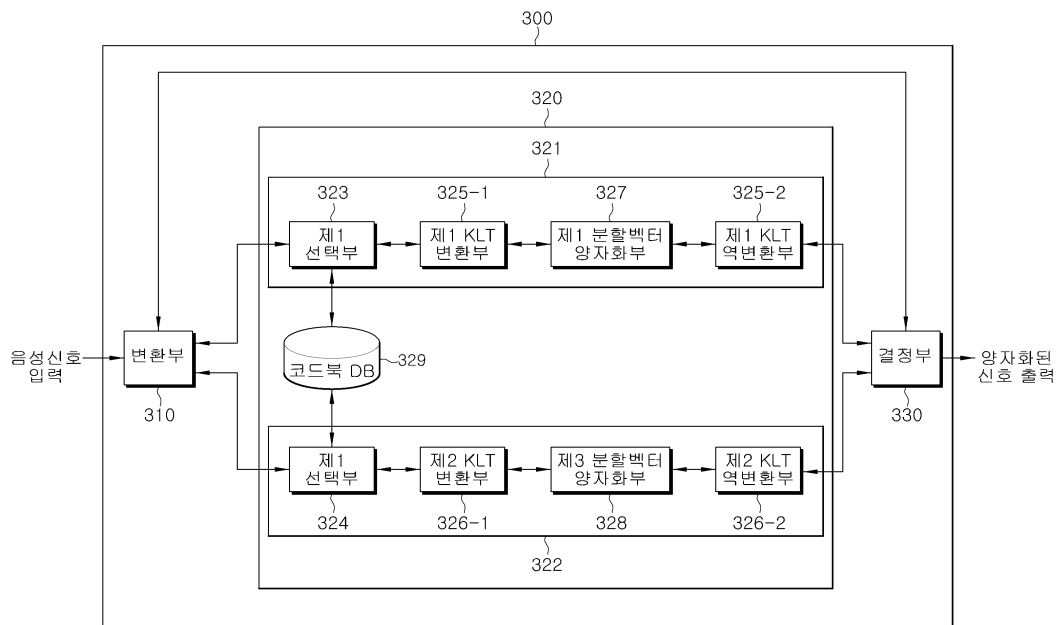
도면1



도면2



도면3



도면4

