



(19) 대한민국특허청(KR)
 (12) 등록특허공보(B1)

(45) 공고일자 2016년12월30일
 (11) 등록번호 10-1691549
 (24) 등록일자 2016년12월26일

- (51) 국제특허분류(Int. Cl.)
G10L 19/107 (2013.01)
- (52) CPC특허분류
G10L 19/107 (2013.01)
- (21) 출원번호 10-2015-7011110
- (22) 출원일자(국제) 2013년07월31일
 심사청구일자 2015년04월28일
- (85) 번역문제출일자 2015년04월28일
- (65) 공개번호 10-2015-0070200
- (43) 공개일자 2015년06월24일
- (86) 국제출원번호 PCT/EP2013/066074
- (87) 국제공개번호 WO 2014/053261
 국제공개일자 2014년04월10일
- (30) 우선권주장
 61/710,137 2012년10월05일 미국(US)
- (56) 선행기술조사문헌
 KR10200000074365 A*
- *는 심사관에 의하여 인용된 문헌
- (73) 특허권자
 프라운호퍼 케겔샵트 쭈르 퍼트레통 데어 안겐
 반텐 포르슝 에. 베.
 독일 80686 뮌헨 한자슈트라쎄 27 체
- (72) 발명자
 베스트룀, 탐
 독일 90443 뉘른베르크 바우에른가쎄 8-12
 몰트루스, 마르쿠스
 독일 90469 뉘른베르크 에츠라우프베그 7
 (뒷면에 계속)
- (74) 대리인
 윤의섭, 김수진

전체 청구항 수 : 총 22 항

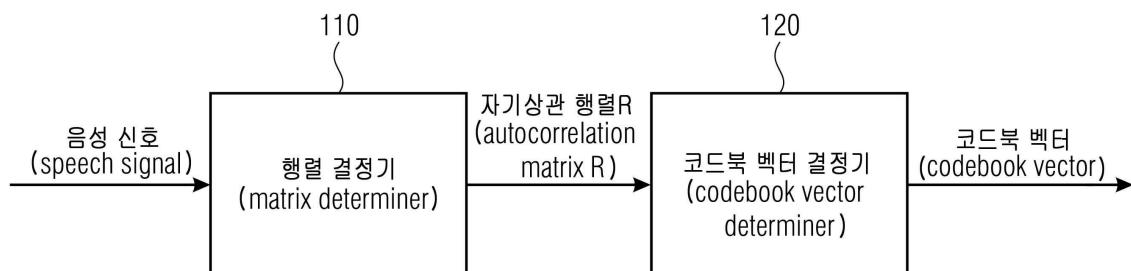
심사관 : 옥윤철

(54) 발명의 명칭 자기상관 영역에서 ACELP를 이용하는 음성 신호 인코딩 장치

(57) 요약

자기상관 행렬(autocorrelation matrix) R을 결정하는 행렬 결정기(matrix determiner)(110); 및 상기 자기상관 행렬 R에 의존하여 코드북 벡터를 결정하는 코드북 벡터 결정기(codebook vector determiner)(120);를 포함하는 음성 코딩 알고리즘(speech coding algorithm)의 코드북 벡터(codebook vector)를 결정(determining)함으로

(뒷면에 계속)

대 표 도 - 도1

로써, 음성 신호를 인코딩(encoding)하는 장치를 개시한다. 행렬 결정기(matrix determiner)(110)는 벡터 r 의 벡터 계수들(vector coefficients)을 결정함으로써 상기 자기상관 행렬 R 을 결정하도록 구성되고, 자기상관 행렬 R 은 복수개의 행(rows)과 열(columns)을 포함하는 벡터 r 은 자기상관 행렬 R 의 복수개의 열들(columns)중

하나 또는 복수개의 행들(rows) 중 하나를 나타내고, $R(i, j) = r(|i-j|)$ 에 있어서, $R(i, j)$ 는 자기상관 행렬 R 의 계수들(coefficients)을 나타내고, i 는 자기상관 행렬 R 의 복수개의 행(rows)들 중 하나인 첫번째 인덱스(first index)를 나타내고, j 는 복수개의 열들(columns)중 하나인 두번째 인덱스(second index)를 나타낸다.

(72) 발명자

디에츠, 마틴

폭스, 기욤

독일 90429 뉴른베르크 도이치헤른슈트라쎄 37

독일 91088 부벤라트 요셉-오토-콜프-슈트라쎄 31

헵리히, 크리스티안

독일 91054 에를랑겐 하우프트슈트라쎄 68

명세서

청구범위

청구항 1

음성 코딩 알고리즘(speech coding algorithm)의 코드북 벡터(codebook vector)를 결정(determining)함으로써, 음성 신호를 인코딩(encoding)하는 장치에 있어서,

행렬 결정기(matrix determiner)(110); 및

자기상관 행렬 R에 의존하여 코드북 벡터를 결정하는 코드북 벡터 결정기(codebook vector determiner)(120);를 포함하고,

상기 행렬 결정기(matrix determiner)(110)는 벡터 r의 벡터 계수들(vector coefficients)을 결정함으로써 자기상관 행렬 R을 결정하도록 구성되고,

상기 자기상관 행렬 R은 복수개의 행(rows)과 열(columns)을 포함하고,

벡터 r은 자기상관 행렬 R의 복수개의 열들(columns)중 하나 또는 복수개의 행들(rows) 중 하나를 나타내고,

$$R(i, j) = r(|i-j|)$$

에 있어서, 상기 $R(i, j)$ 는 자기상관 행렬 R의 계수들(coefficients)을 나타내고, 상기 i는 자기상관 행렬 R의 복수개의 행(rows)들 중 하나인 행 인덱스(row index)를 나타내고, j는 자기상관 행렬 R의 복수개의 열들(columns)중 하나인 열 인덱스(column index)를 나타내며,

상기 코드북 벡터 결정기(120)는

$$f(\hat{e}) = \frac{(d^T \hat{e})^2}{\hat{e}^T R \hat{e}}$$

수학식

의 적용에 의해 코드북 벡터를 결정하도록 구성되고,

상기 R은 자기상관 행렬(autocorrelation matrix)이고,

상기 \hat{e} 는 음성 코딩 알고리즘(speech coding algorithm)의 코드북 벡터들 중 하나이고, $f(\hat{e})$ 는 정규화된 상관(normalized correlation)이며,

$$\frac{(e^T R \hat{e})^2}{(\hat{e}^T R \hat{e})} = \frac{(d^T e)^2}{(\hat{e}^T R \hat{e})}.$$

에서 d 는 타겟 벡터이며,

e는 원래의, 양자화되지 않은 잔류 신호인 것을 특징으로 하는 음성신호 인코딩 장치.

청구항 2

제 1항에 있어서, 상기 코드북 벡터 결정기(120)는 정규화된 상관(normalized correlation)

$$f(\hat{e}) = \frac{(d^T \hat{e})^2}{\hat{e}^T R \hat{e}}$$

을 최대로 하는 음성 코딩 알고리즘의 코드북 벡터 \hat{e} 를 결정하도록 구성되는 것을 특징으로 하는 음성 신호 인코딩 장치.

청구항 3

제 1항에 있어서,

상기 행렬 결정기(110)는 공식:

$$r(k) = h(k) * h(-k) = \sum_l h(l)h(l-k)$$

을 적용하여 벡터 r 의 벡터 계수를 결정하도록 구성되며,

$h(k)$ 는 선형 예측 모델의 지각적으로 가중된 임펄스 응답을 표시하며, k 는 정수인 인덱스이고, l 은 정수인 인덱스인 것을 특징으로 하는 음성 신호 인코딩 장치.

청구항 4

제 1항에 있어서,

행렬 결정기(110)는 지각적으로 가중된 선형 예측기에 의존하여 자기상관을 결정하도록 구성되는 것을 특징으로 하는 음성 신호 인코딩 장치.

청구항 5

제 1항에 있어서,

상기 코드북 벡터 결정기(120)는 자기상관 행렬 R 을 행렬 분해(matrix decomposition)를 수행함으로써 분해하도록 구성되는 것을 특징으로 하는 음성 신호 인코딩 장치.

청구항 6

제 5항에 있어서, 상기 코드북 벡터 결정기(120)는

코드북 벡터 결정을 위한 대각 행렬 D (diagonal matrix D)을 결정하도록 행렬 분해(matrix decomposition)를 수행하도록 구성되는 것을 특징으로 하는 음성 신호 인코딩 장치.

청구항 7

제 6항에 있어서, 상기 코드북 벡터 결정기(120)는

$$\frac{(\hat{f}^H D \hat{f})^2}{\hat{f}^H D^H \hat{f}}.$$

를 사용함으로써 코드북 벡터를 결정하도록 구성되고,

상기 D 는 대각 행렬(diagonal matrix)이고 상기 f 는 첫 번째 벡터(first vector)이고, 상기 \hat{f} 는 두 번째 벡터인 것을 특징으로 하는 음성 신호 인코딩 장치.

청구항 8

제 6항에 있어서, 상기 코드북 벡터 결정기(120)는

코드북 벡터 결정을 위한 대각 행렬 D (diagonal matrix D)의 결정을 위한 행렬 분해(matrix decomposition)를 수행하는 자기상관 행렬 R 을 분해하도록 자기상관 행렬 R 에 대한 반데몬드 인수분해(Vandermonde factorization)를 수행하도록 구성되는 것을 특징으로 하는 음성 신호 인코딩 장치.

청구항 9

제 6항에 있어서, 상기 코드북 벡터 결정기(120)는 코드북 벡터 결정을 위해, 수학식 $\|Cx\|^2 = \|DVx\|^2$ 를 이용하도록 구성되고, 상기 C는 컨볼루션 행렬(convolution matrix)을 나타내고, 상기 V는 푸리에 트랜스폼(Fourier transform)을 나타내고, x는 음성 신호를 나타내는 것을 특징으로 하는 음성 신호 인코딩 장치.

청구항 10

제 6항에 있어서, 상기 코드북 벡터 결정기(120)는 코드북 벡터를 결정하기 위해 대각 행렬 D (diagonal matrix D)를 결정하도록 행렬 분해를 수행하기 위해 자기상관 행렬 R을 분해하는 자기상관 행렬 R에 대한 특이 값 분해(singular value decomposition)를 수행하도록 구성되는 것을 특징으로 하는 음성 신호 인코딩 장치.

청구항 11

제 6항에 있어서, 상기 코드북 벡터 결정기(120)는 코드북 벡터를 결정하기 위해 대각 행렬 D (diagonal matrix D)를 결정하도록 행렬 분해를 수행하기 위해 자기상관 행렬 R을 분해하는 자기상관 행렬 R에 대한 콜레스키 분해(Cholesky decomposition)를 수행하도록 구성되는 것을 특징으로 하는 음성 신호 인코딩 장치.

청구항 12

제 1항에 있어서, 상기 코드북 벡터 결정기(120)는 음성 신호의 제로 임펄스 응답(zero impulse response)에 의존하는 코드북 벡터를 결정하도록 구성되는 것을 특징으로 하는 음성 신호 인코딩 장치.

청구항 13

제 1항에 있어서, 상기 음성 신호 인코딩 장치는 선형 예측 음성 코딩(linear prediction speech coding)에 의해 여기된(excited) 대수 코드(algebraic code)를 이용함으로써 음성 신호(speech signal)를 인코딩(encoding)하는 인코더(encoder)이고,

상기 코드북 벡터 결정기(120)는 대수 코드북(algebraic codebook)의 코드북 벡터로서의 자기상관 행렬 R에 기반하여 코드북 벡터를 결정하도록 구성되는 것을 특징으로 하는 음성 신호 인코딩 장치.

청구항 14

음성 코딩 알고리즘의 코드북 벡터를 결정함으로써 음성 신호를 인코딩하는 장치에 있어서, 상기 장치는:

행렬 결정기(110); 및

자기상관 행렬 R(autocorrelation matrix R)에 의존하는 코드북 벡터(codebook vector)를 결정하는 코드북 벡터 결정기(120);를 포함하고,

상기 행렬 결정기(110)는 벡터 r의 벡터 계수들(vector coefficients)을 결정하여 상기 자기상관 행렬 R(autocorrelation matrix R)을 결정하도록 구성되며,

상기 자기상관 행렬 R은 복수개의 행들(rows)과 복수개의 열들(columns)을 포함하고, 상기 벡터 r은 자기상관 행렬 R의 열들 중 하나 또는 행들 중 하나를 나타내고,

$$R(i, j) = r(|i-j|)$$

에서

상기 $R(i, j)$ 는 자기상관 행렬 R 의 계수들(coefficients)을 나타내고,

상기 i 는 자기상관 행렬 R 의 복수개의 행들(rows) 중 하나를 나타내는 행 인덱스(row index)이고, 상기 j 는 자기상관 행렬 R 의 복수개의 열들(columns) 중 하나를 나타내는 열 인덱스(column index)이며,

상기 코드북 벡터 결정기(120)는 자기상관 행렬 R 을 행렬 분해(matrix decomposition)를 수행함으로써 분해하도록 구성되며,

상기 코드북 벡터 결정기(120)는 코드북 벡터 결정을 위한 대각 행렬 D (diagonal matrix D)를 결정하기 위해 행렬 분해(matrix decomposition)를 수행하도록 구성되며,

상기 코드북 벡터 결정기(120)는

$$\frac{(f^H D \hat{f})^2}{\hat{f}^H D \hat{f}}.$$

를 사용함으로써 코드북 벡터를 결정하도록 구성되고,

상기 D 는 대각 행렬(diagonal matrix)이고 상기 f 는 첫 번째 벡터(first vector)이고, 상기 \hat{f} 는 두 번째 벡터인 것을 특징으로 하는 음성 신호 인코딩 장치.

청구항 15

음성 코딩 알고리즘의 코드북 벡터를 결정함으로써 음성 신호를 인코딩하는 장치에 있어서, 상기 장치는:

행렬 결정기(110); 및

자기상관 행렬 R (autocorrelation matrix R)에 의존하는 코드북 벡터(codebook vector)를 결정하는 코드북 벡터 결정기(120);를 포함하고,

상기 행렬 결정기(110)는 벡터 r 의 벡터 계수들(vector coefficients)을 결정하여 상기 자기상관 행렬 R (autocorrelation matrix R)을 결정하도록 구성되며,

상기 자기상관 행렬 R 은 복수개의 행들(rows)과 복수개의 열들(columns)을 포함하고, 상기 벡터 r 은 자기상관 행렬 R 의 열들 중 하나 또는 행들 중 하나를 나타내고,

$$R(i, j) = r(|i-j|) \quad \text{에서}$$

상기 $R(i, j)$ 는 자기상관 행렬 R 의 계수들(coefficients)을 나타내고,

상기 i 는 자기상관 행렬 R 의 복수개의 행들(rows) 중 하나를 나타내는 행 인덱스(row index)이고, 상기 j 는 자기상관 행렬 R 의 복수개의 열들(columns) 중 하나를 나타내는 열 인덱스(column index)이며,

상기 코드북 벡터 결정기(120)는 자기상관 행렬 R 을 행렬 분해(matrix decomposition)를 수행함으로써 분해하도록 구성되며,

상기 코드북 벡터 결정기(120)는 코드북 벡터 결정을 위한 대각 행렬 D (diagonal matrix D)를 결정하기 위해 행렬 분해(matrix decomposition)를 수행하도록 구성되며,

코드북 벡터 결정을 위한 대각 행렬 D (diagonal matrix D)의 결정을 위한 행렬 분해(matrix decomposition)를 수행하는 자기상관 행렬 R 을 분해하도록 자기상관 행렬 R 에 대한 반데몬드 인수분해(Vandermonde factorization)를 수행하도록 구성되는 것을 특징으로 하는 음성 신호 인코딩 장치.

청구항 16

음성 신호 입력(input speech signal)을 인코딩하여 인코딩된 음성 신호(encoded speech signal)를 획득하는 제1항 내지 제15항 중 어느 한 항에 따른 장치(210); 및

인코딩된 음성 신호(encoded speech signal)를 디코딩하여 디코딩된 음성 신호(decoded speech signal)를 획득하는 디코더(220);을 포함하며,

상기 디코더(220)는 인코딩된 음성 신호를 수신하도록 구성되며, 상기 인코딩된 음성 신호는 제1항 내지 제15항 중 어느 한 항에 따른 장치에 의해 결정되는, 코드북 벡터의 표시를 포함하며,

상기 디코더(220)는 코드북 벡터에 의존하여 디코딩된 음성 신호를 얻기 위해 인코딩된 음성 신호를 디코딩하도록 구성되는 것을 특징으로 하는, 시스템.

청구항 17

음성 코딩 알고리즘의 코드북 벡터를 결정함으로써 음성 신호를 인코딩 하는 방법에 있어서, 상기 방법은:

벡터 r 의 벡터 계수들(vector coefficients)을 결정하는 것을 이용하여 자기상관 행렬 R (autocorrelation matrix R)을 결정하는 단계;를 포함하며,

상기 자기상관 행렬 R 은 복수개의 행들(rows)과 복수개의 열들(columns)을 포함하고, 상기 벡터 r 은 자기상관 행렬 R 의 열들 중 하나 또는 행들 중 하나를 나타내고,

$$R(i, j) = r(|i-j|) \quad \text{에서}$$

상기 $R(i, j)$ 는 자기상관 행렬 R 의 계수들(coefficients)을 나타내고,

상기 i 는 자기상관 행렬 R 의 복수개의 행들(rows) 중 하나를 나타내는 행 인덱스(row index)이고, 상기 j 는 자기상관 행렬 R 의 복수개의 열들(columns) 중 하나를 나타내는 열 인덱스(column index)이며,

코드북 벡터를 결정하는 것은,

$$f(\hat{e}) = \frac{(d^T \hat{e})^2}{\hat{e}^T R \hat{e}}$$

수학식

의 적용에 의해 수행되고,

상기 R 은 자기상관 행렬(autocorrelation matrix)이고,

상기 \hat{e} 는 음성 코딩 알고리즘(speech coding algorithm)의 코드북 벡터들 중 하나이고, $f(\hat{e})$ 는 정규화된 상관(normalized correlation)이며,

$$\frac{(e^T R \hat{e})^2}{(\hat{e}^T R \hat{e})} = \frac{(d^T e)^2}{(\hat{e}^T R \hat{e})}.$$

에서 d 는 타겟 벡터이며,

e 는 원래의, 양자화되지 않은 잔류 신호인 것을 특징으로 하는 음성 신호 인코딩 방법.

청구항 18

음성 코딩 알고리즘의 코드북 벡터를 결정함으로써 음성 신호를 인코딩하는 방법에 있어서, 상기 방법은:

벡터 r 의 벡터 계수들(vector coefficients)을 결정하는 것을 이용하여 자기상관 행렬 R (autocorrelation matrix R)을 결정하는 단계; 및

상기 자기상관 행렬 R (autocorrelation matrix R)에 의존하는 코드북 벡터(codebook vector)를 결정하는 단계;를 포함하고,

상기 자기상관 행렬 R 은 복수개의 행들(rows)과 복수개의 열들(columns)을 포함하고, 상기 벡터 r 은 자기상관

행렬 R의 열들 중 하나 또는 행들 중 하나를 나타내고,

$$R(i, j) = r(|i-j|)$$
에서

상기 $R(i, j)$ 는 자기상관 행렬 R의 계수들(coefficients)을 나타내고,

상기 i는 자기상관 행렬 R의 복수개의 행들(rows) 중 하나를 나타내는 행 인덱스(row index)이고, 상기 j는 자기상관 행렬 R의 복수개의 열들(columns) 중 하나를 나타내는 열 인덱스(column index)이며,

상기 자기상관 행렬 R을 분해하는 단계는 행렬 분해(matrix decomposition) 실행에 의해 수행되고,

상기 행렬 분해를 수행하는 단계는 코드북 벡터 결정을 위한 대각 행렬 D(diagonal matrix D)을 결정하도록 수행되며,

코드북 벡터를 결정하는 것은,

$$\frac{(\mathbf{f}^H \mathbf{D} \hat{\mathbf{f}})^2}{\hat{\mathbf{f}}^H \mathbf{D} \hat{\mathbf{f}}}.$$

를 사용하여 수행되고,

상기 D는 대각 행렬(diagonal matrix)이고 상기 f는 첫 번째 벡터(first vector)이고, 상기 $\hat{\mathbf{f}}$ 는 두 번째 벡터인 것을 특징으로 하는 음성 신호 인코딩 방법.

청구항 19

음성 코딩 알고리즘의 코드북 벡터를 결정함으로써 음성 신호를 인코딩하는 방법에 있어서, 상기 방법은:

벡터 r의 벡터 계수들(vector coefficients)을 결정하는 것을 이용하여 자기상관 행렬 R(autocorrelation matrix R)을 결정하는 단계; 및

상기 자기상관 행렬 R(autocorrelation matrix R)에 의존하는 코드북 벡터(codebook vector)를 결정하는 단계;를 포함하고,

상기 자기상관 행렬 R은 복수개의 행들(rows)과 복수개의 열들(columns)을 포함하고, 상기 벡터 r은 자기상관 행렬 R의 열들 중 하나 또는 행들 중 하나를 나타내고,

$$R(i, j) = r(|i-j|)$$
에서

상기 $R(i, j)$ 는 자기상관 행렬 R의 계수들(coefficients)을 나타내고,

상기 i는 자기상관 행렬 R의 복수개의 행들(rows) 중 하나를 나타내는 행 인덱스(row index)이고, 상기 j는 자기상관 행렬 R의 복수개의 열들(columns) 중 하나를 나타내는 열 인덱스(column index)이며,

상기 자기상관 행렬 R을 분해하는 단계는 행렬 분해(matrix decomposition) 실행에 의해 수행되고,

상기 행렬 분해를 수행하는 단계는 코드북 벡터 결정을 위한 대각 행렬 D(diagonal matrix D)을 결정하도록 수행되며,

코드북 벡터 결정을 위해 대각 행렬 D(diagonal matrix D) 결정하는 행렬 분해(matrix decomposition)를 수행하는 단계는 자기상관 행렬 R을 분해하기 위해 자기상관 행렬 R에 대한 반데몬드 인수분해(Vandermonde factorization)가 수행되는 것을 특징으로 하는, 음성 신호 인코딩 방법.

청구항 20

제 17항 내지 제19항 중 어느 한 항에 따른 인코딩된 음성 신호(encoded speech signal)를 획득하기 위해 입력 음성 신호(input speech signal)를 인코딩하는 단계, - 상기 인코딩된 음성 신호는 코드북 벡터의 표시를 포함

-; 및

상기 코드북 벡터에 의존하여 디코딩된 음성 신호(decoded speech signal)를 획득하기 위해 인코딩된 음성 신호를 디코딩하는 단계;를 포함하는 방법.

청구항 21

컴퓨터 또는 신호 프로세서에서 실행될 때, 제17항 내지 제19항 중 어느 한 항의 방법을 실행하기 위한 컴퓨터 프로그램을 포함하는 컴퓨터-판독가능 매체.

청구항 22

컴퓨터 또는 신호 프로세서에서 실행될 때, 제20항의 방법을 실행하기 위한 컴퓨터 프로그램을 포함하는 컴퓨터-판독가능 매체.

발명의 설명

기술 분야

[0001] 본 발명은 오디오 신호 코딩(audio signal coding)에 관한 것이다. 특히, 자기상관 영역(autocorrelation domain)에서 ACELP를 이용하는 음성 신호 인코딩을 위한 장치에 관한 것이다.

배경 기술

[0002] CELP(Code-Excited Linear Prediction) 코드 선형 예측에 의한 음성 코딩에서, 음성 신호의 스펙트럼 포락선(spectral envelope) (또는 동일하게, 단시간의 시간 구조)은 선형예측(LP, linear predictive) 모델에 의해 설명된다. 그리고, 예측 잔류(prediction residual)은 LTP(long-time predictor, 적응적 코드북 adaptive codebook으로도 알려진 LTP)에 의해 모델링 된다. 그리고, 잔류 신호(residual signal)는 코드북(고정된 코드북으로도 알려진 codebook)에 의해 표시된다. 후자인 고정 코드북(fixed codebook)은 일반적으로 대수 코드북(algebraic codebook)으로 적용된다. 여기서, 코드북은 대수적인 수식 또는 알고리즘에 의해 표현된다. 이는, 빠른 검색 알고리즘(fast search algorithm)을 허용하는 동시에, 전체 코드북을 저장할 필요 없이, 알고리즘만 저장하도록 한다. 대수 코드북(algebraic codebook)에 적용하는 잔류(residual)에 대한 CELP 코덱은 대수 코드 선형 예측(ACELP, Algebraic Code-Excited Linear Prediction) 코덱으로 알려져 있다([1], [2], [3], 4 참조).

[0003] 음성 코딩에서, 대수적인 잔류 코드북(algebraic residual codebook)을 이용하는 것은 [17], [13], [18]과 같은 주류(main stream) 코덱(codecs)에 접근하는 것이다. ACELP는 선형 예측(LP, linear predictive)필터, LTP(long time predictor)에 의한 음성 음향(voiced sounds)의 기저 주파수 및 대수적인 코드북에 의한 예측 잔류(prediction residual)에 의한 스펙트럼 포락선(spectral envelope) 모델링을 기초로 한다. LTP와 대수 코드북(algebraic codebook) 파라미터들(parameters)은 지각영역(perceptual domain)의 최소 제곱 알고리즘(least squares algorithm)에 의해 최적화 된다.

[0004] 계산적으로 ACELP-타입 알고리즘의 가장 복잡한 부분(병목현상 부분)은 잔여 코드북의 최적화이다. 현재 까지 알려진 최적 알고리즘은, 모든 서브 - 프레임(sub-frame)에 대해 크기 N^P 공간을 철저하게 검색(exhaustive search)하는 것이다. 여기서, 모든 포인트마다 $\mathcal{O}(N^2)$ 복잡성의 평가가 요구된다. 전형적인 값들은 펄스 p = 8에서 서브-프레임 길이(sub-frame length) N = 64 (i.e. 5ms)이기 때문에, 이는 초당 10^{20} 번 이상의 연산 수행을 의미한다. 분명히 이는 실행 가능한 옵션이 아니다. 하드웨어 요구 사항(hardware requirements)에 의해 설정된 복잡성 한계(complexity limits) 내에 유지하기 위해, 코드북 최적화 방법은 비반복적인 최적화 알고리즘을 사용하는 경우가 많다.

리즘(non-optimal iterative algorithms)으로 동작해야 한다. 최적화 과정을 위한 이러한 수많은 알고리즘과 개선방안은 예컨대 [17], [19], [20], [21], [22]와 같이 과거에 제안되었다.

[0005] 명시적으로, ACELP 최적화는 측정된 음성 신호가 (1)과 같은 선형 예측 모델(linear predictive model)의 출력으로서의 음성 신호(speech signal) $x(n)$ 을 표현하는 것에 기반한다.

$$\hat{x}(n) = -\sum_{k=1}^m a(k) \hat{x}(n-k) + \hat{e}(k) \quad (1)$$

[0006] 여기서, $a(k)$ 는 LP 계수들(LP coefficients)이고, $\hat{e}(k)$ 는 잔류 신호(residual signal)이다. 벡터 형태에서, 이 수학식은 (2)와 같이 표현될 수 있다.

$$\hat{x} = H \hat{e} \quad (2)$$

[0007] 여기서, 행렬 H 는 대각항 $h(0)$ (diagonal $h(0)$)과 하부 대각항들 $h(1), \dots, h(39)$ (lower diagonals $h(1), \dots, h(39)$)의 하 삼각 플리츠 컨볼루션 행렬(lower triangular Toeplitz convolution matrix)로서 정의된다. 그리고, 벡터 $h(k)$ 는 LP 모델의 임펄스 응답이다. 이 표기에 인지모델(perceptual model, 일반적으로 가중된 LP 모델에 대응하는 인지모델)이 생략된 것은 주의해야 한다. 하지만, 인지모델은 임펄스 응답 $h(k)$ 에 포함되는 것으로 가정한다. 이러한 생략은 표기를 단순화 할뿐, 결과의 일반성에 영향을 미치지 않는다. 인지 모델 포함은 [1]에서와 같이 적용된다.

[0008] 모델의 적합성(fitness of the model)은 제곱 오차(squared error)에 의해 측정된다. 이는 (3)과 같다.

$$\epsilon^2 = \sum_{k=1}^N (x(k) - \hat{x}(k))^2 = (e - \hat{e})^H H^H H (e - \hat{e}). \quad (3)$$

[0009] 이는 제곱 오차(squared error)는 최적 모델 파라미터들(optimal model parameters)을 찾기 위해 사용된다. 여기에서, LTP와 펄스 코드북(pulse codebook)은 모두 벡터 e 를 모델링 하는데 사용되는 것으로 가정한다. 실제 응용 프로그램은 관련 간행물(relevant publications)에서 찾을 수 있다([1-4] 참조).

[0010] 실제로, 전술한 적합성의 측정은 아래와 같이 단순화 될 수 있다.

[0011] 행렬 $B = H^T H$ 는 $h(n)$ 의 상관(correlations)을 포함하도록 하고, c_k 는 k 번째 고정된 코드북 벡터(fixed codebook vector)로 하고, $\hat{e} = g c_k$ 로 설정한다. 여기서, g 는 게인 팩터(gain factor)이다.

[0012] g 가 최적으로 선택되는 것을 가정하면, 코드북은 검색 기준(search criterion)

$$\frac{C_k^2}{E_k} = \frac{(x^T H c_k)^2}{c_k^T B c_k} = \frac{(d^T c_k)^2}{c_k^T B c_k} \quad (4)$$

[0013] 의 최대화에 의해 검색된다.

[0014] 여기서, $d = H^T x$ 는 타겟 벡터와 임펄스 응답 $h(n)$ 사이의 상관관계(correlation)를 포함하는 벡터이다. 그리고 첨자 T는 전치(transpose)를 나타낸다.

[0015] 벡터 d 와 행렬 B 는 코드북 검색(codebook search) 전에 연산된다. 이 수식은 LTP와 펄스 코드북 모두의 최적화에 일반적으로 사용된다.

[0016] 수많은 연구들은 전술한 수식을 사용을 최적화 하는 것에 투자해 왔다.

[0017] 예컨대,

- [0022] 1) 상기 행렬 B의 성분들(elements of matrix B)만이 연산되고, 이들이 서치 알고리즘(search algorithm)에 의해 실제로 접근된다. 또는:
- [0023] 2) 펄스 서치(pulse search)의 시행착오 알고리즘(The trial-and-error algorithm)은 사전 선별(prior screening)(예컨대, [1,5] 참조)에 기초한 높은 성공 확률(high probability of success)의 코드북 벡터들만을 시도함으로써 감소된다.

[0024] ACELP 알고리즘의 실질적인 세부 사항은 제로 임펄스 응답(ZIR, zero impulse response) 개념에 관련된다. 이는 합성된 잔류(synthesised residual)와 비교되는 오리지널 도메인 합성 신호(original domain synthesis signal)를 고려할 때 나타나는 개념이다. 잔류는 프레임 또는 서브-프레임 크기(sub-frame size)에 대응하는 블록들(blocks)로 인코딩(부호화)된다. 그러나, 수학식 1의 LP 모델로 오리지널 도메인 신호를 합성할 때, 고정된 길이의 잔류는 LP 필터의 임펄스 응답에 대응하는 무한한 길이인 “테일(tail)”을 가질 것이다. 즉, 잔여 코드북 벡터(residual codebook vector)의 길이는 제한되어 있지만, 그것은 현재 프레임(current frame) 또는 서브-프레임(sub-frame) 이후의 합성 신호(synthesis signal)에 영향을 미칠 것이다. 이후(future)로의 프레임 효과는 0들로 코드북 벡터를 확장하고, 이 확장된 신호에 대해 수학식 1의 합성 출력(synthesis output)을 계산함으로써 연산 될 수 있다. 합성된 신호의 이러한 확장은 제로 임펄스 응답(ZIR,Zero Impulse Response)으로 알려져 있다. 그리고, 현재 프레임(current frame) 인코딩에서 이전 프레임(prior frame)의 영향을 고려하기 위해, 이전 프레임의 ZIR은 현재 프레임의 타겟으로부터 제외된다. 따라서, 현재 프레임 인코딩에 있어서, 이전 프레임에 의해 이미 모델링 되지 않은 신호의 일부만이 고려된다.

[0025] 실제로, ZIR은 후술하는 내용을 고려한다. (서브)프레임(sub)frame N-1 이 인코딩 되면, 양자화된 잔류(quantized residual)는 0들(zeros)로 다음 (서브) 프레임 N의 길이로 확장된다. 확장된 양자화된 잔류(extended quantized residual)는 LP에 의해 양자화된 신호(quantized signal)의 ZIR을 획득하기 위해 필터링된다. 그러면, 양자화된 신호의 ZIR은 원래의(양자화 되지 않은)신호로부터 감산되고, 이러한 수정 신호(modified signal)는 (서브) 프레임 N을 인코딩 할 때 타겟 신호(목표신호 target signal)를 형성한다. 이러한 방식으로, (서브)프레임 N-1에서 만들어진 모든 양자화 오차들(quantization errors)은 (서브)프레임 N을 양자화(quantizing)할 때, 고려될 것이다. 이는 출력 신호의 인지 품질(perceptual quality of the output signal)을 상당히 개선한다.

[0026] 그러나 이는 오디오 코딩에 대한 향상된 개념이 제공되는 경우에, 높게 평가될 것이다.

발명의 내용

해결하려는 과제

[0027] 본 발명의 목적은 오디오 객체 코딩(audio object coding)에 대한 향상된 개념을 제공하는 것이다.

과제의 해결 수단

[0028] 본 발명의 목적은 청구항 1항에 따른 장치, 청구항 15항에 따른 인코딩 방법, 청구항 16항에 따른 디코더, 청구항 17항에 따른 디코딩 방법, 청구항 18항에 따른 시스템, 청구항 19항에 따른 방법 및 청구항 20항에 따른 컴퓨터 프로그램에 의해 달성된다.

[0029] 음성 부호화 알고리즘(speech coding algorithm)의 코드북 벡터(codebook vector)를 결정(determining)함으로써, 음성 신호를 부호화(encoding) 하는 장치가 제공된다.

[0030] 상기 장치는 자기상관 행렬(autocorrelation matrix) R을 결정하는 행렬 결정기(matrix determiner); 및 자기상관 행렬 R에 의존하여 코드북 벡터를 결정하는 코드북 벡터 결정기(codebook vector determiner)를 포함한다. 행렬 결정기(matrix determiner)(110)는 벡터 r의 벡터 계수들(vector coefficients)을 결정함으로써 자기상관

행렬 R 결정하도록 구성되고, 상기 자기상관 행렬 R 은 복수개의 행(rows)과 열(columns)을 포함하고, 상기 벡터 r 은 자기상관 행렬 R 의 복수개의 열들(columns) 중 하나 또는 복수개의 행들(rows) 중 하나를 나타내고, $R(i, j) = r(|i-j|)$ 에 있어서, 상기 $R(i, j)$ 는 자기상관 행렬 R 의 계수들(coefficients)을 나타내고, 상기 i 는 자기상관 행렬 R 의 복수개의 행(rows)들 중 하나인 첫 번째 인덱스(first index)를 나타내고, 상기 j 는 복수개의 열들(columns) 중 하나인 두 번째 인덱스(second index)를 나타낸다.

[0031] 장치는 음성 신호를 인코딩 하는 코드북 벡터를 이용하도록 구성된다. 예컨대, 상기 장치는 인코딩 된 음성 신호(encoded speech signal)를 생성할 수 있다. 인코딩 된 음성 신호는 복수의 선형 예측 계수들(Linear Prediction coefficients), 음성 음향(voiced sounds)의 기저 주파수(fundamental frequency)표시(예컨대, 퍼치 파라미터, pitch parameters), 코드북 벡터 표시, 예컨대, 코드북 벡터 인덱스 등을 포함한다.

[0032] 또한, 전술한 실시예에 따른 장치에 의해, 인코딩 된 음성 신호(encoded speech signal)를 디코딩하여, 디코딩 된 음성신호(decoded speech signal)를 획득하는 디코더가 제공된다.

[0033] 또한, 시스템이 제공된다. 시스템은 인코딩 된 음성 신호(encoded speech signal)를 획득하기 위해 입력 음성 신호(input speech signal)를 인코딩 하는 전술한 실시예에 따른 장치를 포함한다. 또한, 시스템은 디코딩된 음성 신호(decoded speech signal)를 획득하기 위해 인코딩 된 음성 신호(encoded speech signal)를 디코드(복호화)하는 전술한 실시예에 따른 디코더(decoder)를 포함한다.

[0034] 음성 코딩 알고리즘 ACELP의 목적 함수(objective function)에 대한 향상된 개념이 제공된다. 이는 현재 프레임의 파라미터들을 최적화 할 때, 이전 프레임(previous frame)에서 현재 프레임(current frame)으로의 임펄스 응답(impulse response)의 영향 뿐만 아니라, 현재 프레임의 다음 프레임(next frame)으로의 임펄스 응답의 영향 또한 고려한다. 일부 실시예들은 상관 행렬(correlation matrix)을 변경 함으로써 이러한 개선을 구현한다. 이는 헤미시안 토플리츠(Hermitian Toeplitz)구조를 가지는 자기 상관 행렬(autocorrelation matrix)에 대한 종래 ACELP 최적화(optimisation)의 핵심이다.

발명의 효과

[0035] 이러한 구조를 이용함으로써, 메모리 요구사항 뿐만 아니라 연산적인 복잡성의 두 관점 모두에서 보다 효율적인 ACELP최적화를 가능하게 한다.

[0036] 또한, 이와 동시에 적용된 지각 모델(perceptual model applied)은 보다 일관적이 되고, 프레임 간 종속성(interframe dependencies)은 패킷 손실(packet-loss)의 영향에서 성능을 개선 시키기 위해 방지 될 수 있다.

도면의 간단한 설명

[0037] 도 1은 본 발명의 실시예에 따른 음성 코딩 알고리즘의 코드북 벡터를 결정함으로써 음성 신호를 인코딩 하는 장치를 나타낸 도면이다.

도 2는 본 발명의 실시예에 따른 디코더 및 디코더를 나타낸 도면이다.

도 3은 본 발명의 실시예에 따라 음성 신호를 인코딩 하는 장치 및 디코더를 포함하는 시스템을 나타낸 도면이다.

발명을 실시하기 위한 구체적인 내용

[0038] ACELP 패러다임(paradigm)에 있어서 음성 코딩은 인지 영역(perceptual domain)에서 최소 제곱 알고리즘(least squares algorithm)에 기반한다. 여기서, 지각 영역은 필터에 의해 지정된다. 실시 예에 따라, 최소 제곱 문제에 대한 종래 정의(conventional definition)에 대한 연산 복잡성(computational complexity)은 다음 프레임(next frame)으로의 제로 임펄스 응답(zero impulse response)의 영향을 고려함으로써 감소 될 수 있다. 제공

된 변형들(provided modifications)은 목적 함수(objective function)에서 나타나는 상관 행렬(correlation matrix)에 폴리츠 구조(Toeplitz structure)를 도입하여, 구조를 단순화 하고, 연산을 감소시킨다. 제시된 개념은 인지되는 품질 저하 없이 연산적인 복잡성을 최대 17%까지 감소 시킨다.

[0039] 실시 예들은 목적 함수(objective function)의 약간의 수정(slight modification)에 의해, 잔류 코드북의 최적화에 대한 복잡성이 더 감소될 수 있다는 발견에 기초한다. 이러한 복잡성 감소는 인지적인 품질(perceptual quality) 감소 없이 제공된다. 대안으로서, ACELP 잔류 최적화(residual optimization)는 제시된 수정(presented modification)에 대한 반복적인 검색 알고리즘(iterative search algorithms)에 기반하기 때문에, 복잡성 증가 없이 반복 횟수를 늘릴 수 있고, 이로써, 개선된 인지 품질을 획득한다.

[0040] 수정된 목적 함수들 뿐만 아니라 종래의 목적함수들 모두 인지(perception)를 모델링하고, 지각적 왜곡(perceptual distortion)을 최소화하기 위해 노력한다. 그러나, 종래의 접근 방법에 대한 최적 방안은 수정된(변형된) 목적함수에 대해서는 반드시 최적이 아니고, 그 반대의 경우도 마찬가지이다. 이는 하나의 접근 방법이 다른 것보다 더 나은 것을 의미하지 않지만, 분석적인 논의(analytic arguments)는 수정된 목적 함수가 보다 일관성 있다는 것을 보여준다. 구체적으로, 종래의 목적 함수와는 대조적으로, 제공된 개념은 일관성, 명확한 인지(well-defined perceptual) 및 신호 모델들로 서브-프레임 내에서 모든 샘플들을 동일하게 처리한다.

[0041] 실시 예에서, 제안 된 수정(modifications)은 단지 잔류 코드북의 최적화를 변경하도록 적용될 수 있다. 따라서, 비트-스트림(bit-stream)구조를 변경하지 않고, 기존 ACELP 코덱(codecs)에 백 워드 호환 방식(backward compatible manner)으로 적용될 수 있다.

[0042] 또한, 음성 코딩 알고리즘의 코드북 벡터 결정에 의한 음성 신호 인코딩 방법이 제공된다.

[0043] 이 방법은

[0044] - 자기상관 행렬 (autocorrelation matrix) R 을 결정하는 단계; 및:

[0045] - 상기 자기상관 행렬 R 에 의존하는 코드북 벡터를 결정하는 단계;

[0046] 를 포함한다.

[0047] 자기상관 행렬 (autocorrelation matrix) R 을 결정하는 단계는 벡터 r 의 벡터 계수들을 결정하는 단계를 포함한다. 자기상관 행렬 (autocorrelation matrix) R 은 복수개의 행들(rows)과 복수개의 열들(columns)을 포함한다.

$$R(i, j) = r(|i-j|)$$

[0048]

[0049] 예 있어서, 벡터 r 은 자기상관 행렬 R 의 열들 중 하나 또는 행들 중 하나를 나타낸다. $R(i, j)$ 는 자기상관 행렬 R 의 계수들(coefficients)을 나타내고, 상기 i 는 자기상관 행렬의 복수개의 행들(rows) 중 하나를 나타내는 첫 번째 인덱스(first index)이고, 상기 j 는 자기상관 행렬의 복수개의 열들(columns) 중 하나를 나타내는 두 번째 인덱스(first index)이다.

[0050] 또한, 디코딩된 음성 신호를 획득하기 위한 전술한 실시예에 따른 음성 신호 인코딩 방법에 따라 인코딩 된 인코딩 음성 신호를 디코딩하는 방법이 제공된다.

[0051] 또한, 방법이 제공된다. 상기 방법은,

[0052] - 인코딩된 음성 신호(encoded speech signal)를 획득하기 위해 전술한 음성 신호 인코딩 방법에 따라 입력 음

성 신호(input speech signal)를 인코딩하는 단계; 및:

[0053] - 디코딩된 음성 신호를 획득하기 위해, 전술한 음성 신호 디코딩 방법에 따라 인코딩된 음성 신호를 디코딩 하는 단계; 를 포함한다.

[0054] 또한, 컴퓨터 또는 신호 프로세서에서 실행되는 경우 전술한 방법들을 구현하기 위한 컴퓨터 프로그램이 제공된다.

[0055] 바람직한 실시 예들은 종속 항들에서 제공 될 것이다.

[0056] 이하에서, 본 발명의 실시 예들은 도면을 참조하여 더욱 상세히 설명된다:

[0057] 도 1은 본 발명의 실시예에 따른 음성 코딩 알고리즘의 코드북 벡터를 결정함으로써 음성 신호를 인코딩 하는 장치를 나타낸 도면이다.

[0058] 상기 장치는 자기상관 행렬(autocorrelation matrix) R을 결정하는 행렬 결정기(matrix determiner)(110); 및 자기상관 행렬 R에 의존하여 코드북 벡터를 결정하는 코드북 벡터 결정기(codebook vector determiner)(120);를 포함한다.

[0059] 행렬 결정기(matrix determiner)(110)는 벡터 r의 벡터 계수들(vector coefficients)을 결정함으로써 자기상관 행렬 R 결정하도록 구성된다.

$$R(i, j) = r(|i-j|)$$

[0060] 자기상관 행렬 R은 복수개의 행(rows)과 열(columns)을 포함하고, 상기 벡터 r은 자기상관 행렬 R의 복수개의 열 들(columns)중 하나 또는 복수개의 행들(rows) 중 하나를 나타낸다.

[0061] $R(i, j)$ 는 자기상관 행렬 R의 계수들(coefficients)을 나타내고, 상기 i는 자기상관 행렬 R의 복수개의 행 (rows)들 중 하나인 첫 번째 인덱스(first index)를 나타내고, 상기 j는 복수개의 열 들(columns)중 하나인 두 번째 인덱스(second index)를 나타낸다.

[0062] 장치는 음성 신호를 인코딩 하는 코드북 벡터를 이용하도록 구성된다. 예컨대, 상기 장치는 인코딩 된 음성 신호(encoded speech signal)를 생성할 수 있다. 인코딩 된 음성 신호는 복수의 선형 예측 계수들(Linear Prediction coefficients), 음성 음향(voiced sounds)의 기저 주파수(fundamental frequency)표시(예컨대, 피치 파라미터, pitch parameters), 코드북 벡터 표시, 예컨대, 코드북 벡터 인덱스 등을 포함한다.

[0063] 예컨대, 특정 실시예에 따른 음성 신호 인코딩에 있어서, 장치는 음성 신호에 의존하는 복수개의 선형 예측 계수들 (linear predictive coefficients)(a(k))을 결정하도록 구성될 수 있다. 또한, 장치는 복수개의 선형 예측 계수들(a(k))에 의존하는 잔류 신호(residual signal)를 결정하도록 구성된다.

[0064] 또한, 행렬 결정기 110은 잔류 신호에 의존하는 자기상관 행렬 R을 결정하도록 구성될 수 있다.

[0065] 이하에서, 본 발명에 따른 몇몇 추가 실시 예들이 설명된다.

[0066] 수학식 3과 4로 돌아오면, 상기 수학식 3은 인지적 모델(perceptual model)의 적합성을 나타내는 제곱 오차(squared error)를 정의하고, 다음과 같다:

$$\epsilon^2 = \sum_{k=1}^N (x(k) - \hat{x}(k))^2 = (e - \hat{e})^H H^H H (e - \hat{e}). \quad (3)$$

[0068] 그리고 상기 수학식 4는

$$\frac{C_k^2}{E_k} = \frac{(x^T H c_k)^2}{c_k^T B c_k} = \frac{(d^T c_k)^2}{c_k^T B c_k} \quad (4)$$

[0069]

[0070] 최대화된 검색 기준(search criterion)을 나타낸다.

[0071]

ACELP 알고리즘은 수학식 4를 중심으로 한다. 이는 다시 수학식 3을 기반으로 한다.

[0072]

실시예들은 양자화된 잔류 값(quantized residual values) $e(k)$ 가 인덱스 k 에 의존하는 오차 에너지 (error energy) ϵ^2 에 매우 다른 영향을 미치는 것을 보여주는 이러한 수학식의 분석을 발견하는 것에 기반한다. 예컨대, 인덱스들(indices) $k=1$ 및 $k=N$ 을 고려할 때, 만약 잔류 코드북의 유일한 0이 아닌 값(only non-zero value)이 $k=1$ 로 나타난다면, 오차 에너지(error energy) ϵ^2 의 결과는 다음과 같다.

[0073]

$$\epsilon_1^2 = \sum_{k=1}^N (x(k) - e(1)h(k))^2 \quad (5)$$

[0074]

$k=N$ 동안의 오차 에너지 (error energy) ϵ^2 결과는 (6)이다.

[0075]

$$\epsilon_N^2 = (x(N) - e(N)h(1))^2 + \sum_{k=1}^{N-1} (x(k))^2. \quad (6)$$

[0076]

다시 말해, $e(1)$ 은 1부터 N 범위에서 임펄스 응답(impulse response) $h(k)$ 로 가중(weighted)되고, 반면 $e(N)$ 은 오직 $h(1)$ 로 가중된다. 가중 스펙트럼(spectral weighting)의 관점에서 이는, $e(k)$ 각각이 다른 스펙트럼 가중 함수(spectral weighting function)로 가중된 것을 의미한다. 극단적인 예로서, $e(N)$ 은 선형적으로 가중 (linearly-weighted)된다. 지각적 모델링(perceptual modelling) 관점에서는, 프레임 내의 모든 샘플들에 대해 동일한 지각적인 가중(same perceptual weight)을 적용하는 것이 알맞다. 따라서 수학식 3은 다음 프레임으로 ZIR을 고려하는 것과 같이 확장되어야 한다. 특히 주목 해야 할 종래 기술과의 차이는, 이전 프레임(previous frame)으로부터의 ZIR과 다음 프레임(next frame)으로의 ZIR이 모두 고려된다는 점이다.

[0077]

$e(k)$ 는 원래의, 양자화 되지 않은 잔류(unquantized residual)로, $\hat{e}(n)$ 는 양자화된 잔류(quantised residual)로 한다. 또한, 두 잔류 모두 1부터 N 까지의 범위에서는 0이 아니도록(non-zero) 하고, 이외의 범위에서는 0(zero)으로 한다. 그러면,

[0078]

$$\begin{aligned} x(n) &= -\sum_{k=1}^m a(k)x(n-k) + e(n) = \sum_{k=1}^{\infty} e(n-k)h(k) \\ \hat{x}(n) &= -\sum_{k=1}^m a(k)\hat{x}(n-k) + \hat{e}(n) = \sum_{k=1}^{\infty} \hat{e}(n-k)h(k) \end{aligned} \quad (7)$$

[0079]

동등하게, 행렬 형태(matrix form)에서의 동일한 관계는 다음과 같이 표현 될 수 있다:

[0080]

$$\begin{aligned} x &= \tilde{H}e \\ \hat{x} &= \tilde{H}\hat{e} \end{aligned} \quad (8)$$

[0081]

여기에서 \tilde{H} 는 임펄스 응답(impulse response) $h(k)$ 에 대응하는 무한 차원 컨볼루션 행렬(infinite dimensional convolution matrix)이다. 수학식 3의 수율(yields)을 대입한다.

[0082]

$$\epsilon^2 = \|\tilde{H}e - \tilde{H}\hat{e}\|^2 = (e - \hat{e})^T \tilde{H}^T \tilde{H} (e - \hat{e}) = (e - \hat{e})^T R (e - \hat{e}) \quad (9)$$

[0083] 여기에서 $R = \tilde{H}^T \tilde{H}$ 는 $h(n)$ 의 자기상관(autocorrelation)에 대응하는 토리츠 행렬(Toeplitz matrix)로 유한한(finite) 크기이다. 수학식 4와 유사한 유도(similar derivation)에 의해, 목적함수(objective function)가 획득된다.

$$\frac{(e^T R \hat{e})^2}{(\hat{e}^T R \hat{e})} = \frac{(d^T e)^2}{(\hat{e}^T R \hat{e})}. \quad (10)$$

[0085] 이 목적함수는 수학식 4와 매우 유사하다. 상관 행렬(correlation matrix) B 대신에 허미시안 토리츠 행렬(Hermitian Toeplitz matrix) R 이 분모(denominator)에 있다는 것이 주요한 차이이다.

[0086] 상술한 바와 같이, 이러한 새로운 수식으로 프레임 내의 모든 잔류 샘플들(samples of the residual)이 동일한 인지적인 가중(perceptual weighting)을 받게 되는 이점을 갖는다. 하지만, 중요한 것은, 이 수식은 연산 복잡성(computational complexity) 및 메모리 요구사항(memory requirements)에서도 고려할 만한 이익을 도입하는 것이다.

[0087] R 은 허미시안 토리츠 행렬(Hermitian Toeplitz matrix) 이기 때문에, 첫 번째 열(column) $r(0) \dots r(N-1)$ 은 행렬을 완전하게 정의한다. 다시 말해, 완전한 $N \times N$ 행렬을 저장(storing)하는 대신, $N \times 1$ 벡터 $r(k)$ 만 저장한다. 이로써, 메모리 할당(memory allocation)에 대해 상당한 절약을 얻는다. 또한, 연산적인 복잡성은 모든 $N \times N$ 성분들(elements)을 결정할 필요 없이, 오직 첫 번째 $N \times 1$ 열(column)만 결정하기 때문에 감소된다. 또한, 성분

(i, j) 는 $R(i, j) = r(|i-j|)$ 에 의해 발견되기 때문에, 행렬 내의 표기(indexing)가 간단하다.

[0088] 수식 10의 목적 함수는 수식 4와 매우 비슷하기 때문에, 일반적인 ACELP의 구조가 유지될 수 있다. 구체적으로, 다음 동작들(operations)은 간단한 알고리즘 변형으로 목적함수(objective function) 중 어느 하나로 수행될 수 있다.

[0089] 1. LTP 지연의 최적화 (적응적 코드북)

[0090] Optimisation of the LTP lag (adaptive codebook)

[0091] 2. 잔류 모델링에 대한 펄스 코드북의 최적화(고정 코드북)

[0092] Optimisation of the pulse codebook for modelling the residual (fixed codebook)

[0093] 3. LTP 및 펄스의 이득 최적화, 개별적이거나 공동으로 (Optimisation of the gains of LTP and pulses, either separately or jointly)

[0094] 4. 수학식 3의 제곱 오차에 의해 성능이 측정 될 수 있는 임의의 다른 파라미터의 최적화. (Optimisation of any other parameters whose performance can be measured by the squared error of Equation 3)

[0095] 종래의 ACELP 응용(applications)에서 수정되어야 하는 유일한 부분은 상관 행렬(correlation matrix) B 의 처리(the handling)이다. 다음 프레임(following frame)으로의 ZIR 을 포함해야 하는 타겟(target)뿐만 아니라 행렬 R 로 대체된다.

[0096] 일부 실시예들은 상관 행렬(correlation matrix) B 가 나타나는 ACELP 알고리즘 어디에서든 자기 상관 행렬

(autocorrelation matrix) R에 의해 대체되는 본 발명의 개념을 채용한다. 만약 행렬 B의 모든 인스턴스들 (instances)이 생략된 경우에는, 그것의 값 연산을 회피할 수 있다.

[0097] 예컨대, 자기상관 행렬 R은 자기상관 행렬 R의 첫 번째 열(column) $r(0), \dots, r(N-1)$ 의 계수들(coefficients)을 결정함으로써 결정된다.

[0098] $R = H^T H$ 에 의해 행렬 R이 정의되고, 행렬 R의 성분들(elements) $R_{ij} = r(i-j)$ 은 (9a)를 통해 연산 될 수 있다.

$$r(k) = h(k) * h(-k) = \sum_l h(l)h(l-k) \quad (9a)$$

[0099] 즉, $r(k)$ 배열(sequence)은 $h(k)$ 의 자기상관(autocorrelation)이다.

[0100] 그러나, 종종, $r(k)$ 는 보다 더 효율적인 수단에 의해 획득 될 수 있다. 구체적으로, AMR 및 G.718과 같은 음성 코딩 표준(speech coding standards)에서, $h(k)$ 배열(sequence)은 프리-эм파시스(pre-emphasis)를 포함하도록 하는 인지적인 가중 함수(perceptual weighting function) $W(z)$ 에 의해 필터링 되는 선형 예측 필터(linear predictive filter) $A(z)$ 의 임펄스 응답(impulse response)이다. 다시 말해, $h(k)$ 는 선형 예측 모델의 인지적 으로 가중된 임펄스 응답(perceptually weighted impulse response)을 나타낸다.

[0101] 필터 $A(z)$ 는 일반적으로 음성 신호 $rX(k)$ 의 자기상관으로부터 측정된다. 즉, $rX(k)$ 는 이미 알려져 있다. $H(z) = A^{-1}(u)W(z)$ 이기 때문에, 자기상관 배열(autocorrelation sequence) $r(k)$ 는 9b로 $w(k)$ 의 자기상관(autocorrelation)을 연산함으로써 결정될 수 있다.

$$r_w(k) = w(k) * w(-k) = \sum_l w(l)w(l-k) \quad (9b)$$

[0102] 이로써 $h(k)$ 의 자기상관은 (9c)이다.

$$r(k) = r_x(k) * r_w(k) = \sum_l r_w(l)r_x(l-k). \quad (9c)$$

[0103] 전체 시스템의 디자인에 의존하여, 이러한 수학식들은, 몇몇 실시예에 따라, 적절하게 수정(변형) 될 수 있다.

[0104] 그리고 예컨대, 코드북의 코드북 벡터는 자기상관 행렬(autocorrelation matrix) R에 기초하여 결정될 수 있다. 특히, 수학식 10은 몇몇 실시예에 따라 코드북의 코드북 벡터를 결정하는데 이용 될 수 있다.

$$f(\hat{e}) = \frac{(d^T \hat{e})^2}{\hat{e}^T R \hat{e}}$$

[0105] 이러한 맥락에서, 수학식 10은 형태 $f(\hat{e})$ 에서 목적 함수를 정의한다. 이는, 음성 코딩 표준 (speech coding standards)인 AMR 및 G.718과 같은 형태이지만, 그러나 행렬 R은 현재 대칭 폴리즈(symmetric Toeplitz) 구조를 가진다. 목적 함수는 기본적으로 타겟 벡터(target vector) d와 코드북 벡터(codebook vector) \hat{e} 사이의 균등화된 상관관계 (코렐레이션)(normalized correlation)이다. 그리고, 최적의 코드북 벡터는 정규화된 상관관계 $f(\hat{e})$ (코렐레이션)에 대해 최고 값(highest value)을 제공하는 예컨대, 정규화된 상관관계(correlation) $f(\hat{e})$ 를 최대화 하는 것이다.

[0106] 따라서, 코드북 벡터들은 전술한 표준과 같은 동일한 접근 방법으로 최적화 될 수 있다. 구체적으로, 예컨대, 잔류에 대한 최상의 대수 코드북(예컨대, 고정 코드북) 벡터 \hat{e} 를 발견하기 위한 매우 간단한 알고리즘은 후술하는 바와 같이 적용될 수 있다. 그러나, 상당한 노력이 효율적인 서치 알고리즘(search algorithms) 디자인에 투자되었다는 것에 유념해야 한다.

[0110] 그리고, 이러한 서치 알고리즘은 단지 어플리케이션의 예시이다.

[0111] 1. 초기 코드북 벡터(initial codebook vector) $\hat{e}_0 = [0, 0 \dots 0]^T$ 를 정의하고, $p = 0$ 으로 펄스의 수를 설정한다.

[0112] 2. 초기 코드북 품질 측정(codebook quality measure) 을 $f_0 = 0$ 로 설정한다.

[0113] 3. 임시의 코드북 품질 측정(temporary codebook quality)을 $\hat{f}_p = f_{p-1}$ 로 설정한다.

[0114] 4. 코드북 벡터에서의 각 위치(position) k는

[0115] (i) p가 1씩 증가함

[0116] (ii) k위치가 이미 음의 펄스(negative pulse)를 포함하는 경우 vii 단계로 진행함

[0117] (iii) 임시 코드북 벡터 $\varepsilon_p^+ = \hat{e}_{p-1}$ 를 생성하고, k 위치에 양의 펄스(positive pulse)를 더함

[0118] (iv) $f(\varepsilon_p^+)$ 에 의한 임시 코드북 벡터의 품질을 평가

[0119] (v) 만약 임시 코드북 벡터가 이전 어떤 값들보다 양호하다면, $f(\varepsilon_p^+) > \hat{f}_p$ 이고, 이러한 코드북 벡터를 저장(save)하고, $\hat{f}_p = f(\varepsilon_p^+)$ 로 설정하고, 다음 반복(next iteration)을 지속함

[0120] (vi) 만약, k 위치가 이미 양의 펄스(positive pulse)를 포함하고 있다면, 다음 반복을 지속함

[0121] (vii) 임시 코드북 벡터(temporary codebook vector) $\varepsilon_p^- = \hat{e}_{p-1}$ 를 생성하고, k 위치에서 음의 펄스(negative pulse)를 추가함

[0122] (viii) $f(\varepsilon_p^-)$ 에 의한 임시 코드북 벡터(temporary codebook vector)의 품질을 평가함

[0123] (ix) 만약 임시 코드북 벡터가 이전의 어떤 값들보다 양호 하다면, $f(\varepsilon_p^-) > \hat{f}_p$ 이고, 이 코드북 벡터를 저장(save)하고, $\hat{f}_p = f(\varepsilon_p^-)$ 로 설정하고, 다음 반복을 지속함

[0124] 5. 저장된 코드북 벡터들 중 마지막 값(즉, 최상의 값)으로 코드북 벡터 \hat{e}_p 를 정의함

[0125] 6. 만약, 펄스 p의 수가 이미 펄스의 원하는(바람직한) 수(desired number)에 도달한 경우, 출력 벡터(output vector) 를 $\hat{e} = \hat{e}_p$ 로 정의하고, 정지함. 그렇지 않으면, step 4를 지속함

[0126] 이미 지적한 바와 같이, 종래의 ACELP와 비교하면, 몇몇 실시예에 있어서, 타겟(target)은 다음 프레임(following frame)으로의 ZIR을 포함하는 것과 같이 수정된다.

[0127] 수학식 1은 ACELP-타입 코덱 (ACELP-type codecs)에서 사용되는 선형 예측 모델(linear predictive model)을 설명한다. 제로 임펄스 응답(ZIR, Zero Impulse Response, 또한 몇몇의 경우 제로 입력 응답(Zero Input Response)으로도 알려진)은 현재 프레임(current frame, 그리고 모든 미래의 프레임들)의 잔류(residual)가 0 으로 설정될 때, 선형 예측 모델의 출력(output)을 추론한다. ZIR은 위치 (position) N으로부터 0인 잔류를 정의함으로써 다음과 같이 용이하게 계산될 수 있다.

$$e_K'(n) = \begin{cases} e(n) & \text{for } n < K \\ 0 & \text{for } n \geq K \end{cases} \quad (10a)$$

[0129] 여기서 ZIR은 (10b)로서 정의될 수 있다.

$$[0130] \quad ZIR_K(n) = \sum_{k=0}^N h(k) e_K'(n-k). \quad (10b)$$

[0131] 입력 신호(input signal)로부터 이러한 ZIR을 빼면, 오직 현재 프레임 앞으로부터의 잔류에 의존하는 신호가 획득된다.

[0132] 동일하게, ZIR은 과거 입력 신호(past input signal)를 10c로서 필터링 함으로써 결정될 수 있다.

$$[0133] \quad ZIR_K(n) = \begin{cases} x(n) & \text{for } n < K \\ -\sum_{k=1}^m a(k) ZIR_K(n-k) & \text{for } n \geq K. \end{cases} \quad (10c)$$

[0134] ZIR 이 제거 되는 곳의 입력 신호(input signal)는 종종 타겟(target)으로서 알려지고, $d(n) = x(n) - ZIR_K(n)$.로서의 위치 K에서 시작하는 프레임을 위해 정의 될 수 있다. 이러한 타겟은 원칙적으로 정확하게 AMR과 G.718 표준(G.718 standards)에서의 타겟과 동일하다. 신호를 양자화(quantizing) 할 때, 양자화된 신호 $\hat{d}(n)$ 는 프레임 $K \leq n < K + N$.의 지속(duration)동안 $d(n)$ 과 비교된다.

[0135] 반대로, 현재 프레임의 잔류는 다음 프레임들(following frames)에 영향을 미친다. 여기서, 신호를 양자화 할 때 그것의 영향을 고려하는 것이 유용하다, 이로써, 현재 프레임, $n > K + N$ 을 넘어선 차이 $\hat{d}(n) - d(n)$ 또한 평가할 수 있다. 그러나, 이를 수행하기 위해서, 오직 다음 프레임들(following frames)의 0으로의 잔류 설정에 의해서(만) 현재 프레임의 잔류 영향(influence of the residual)을 고려 할 수 있다. 그러므로, 다음 프레임으로의 ZIR은 비교 될 수 있다. 다시 말해, 수정된 타겟(modified target)이 획득된다.

$$[0136] \quad d'(n) = \begin{cases} 0 & n < K \\ d(n) & K \leq n < K + N \\ -\sum_{k=1}^m a(k) d'(n-k) & n > K + N. \end{cases} \quad (10d)$$

[0137] 동일하게, A(z)의 임펄스 응답(impulse response) $h(n)$ 을 이용한다. 그러면,

$$[0138] \quad d'(n) = \sum_{k=K}^{K+N-1} e(k) h(n-k). \quad (10e)$$

[0139] 이 수학식(formula)은 $\mathbf{d}' = \mathbf{H}\mathbf{e}$ 에 의한 편리한 행렬 형태(convenient matrix form)로 기재될 수 있고, 여기서 H 와 e 는 수학식 2에서 정의된다. 수정된 타겟(modified target)은 정확하게 수학식 2의 x인 것을 알 수 있다.

[0140] 행렬 R의 연산에서, 이론적으로 유념해야 하는 것은, 임펄스 응답(impulse response) $h(k)$ 는 특정 시스템(practical system)에서 실현되지 않는 무한 배열(infinite sequence)라는 것이다.

[0141] 하지만,

[0142] 1) 임펄스 응답을 유한한 길이로 절단(truncating) 또는 윈도윙(windowing) 하는 과정과 절단된 임펄스 응답의 자기상관(autocorrelation)을 결정하는 과정, 또는

[0143] 2) 연관된 LP와 인지 필터들(perceptual filters)의 푸리에 스펙트럼(Fourier spectra)을 이용하여 임펄스 응답의 파워 스펙트럼(power spectrum)을 계산하는 과정, 그리고

[0144] 역 푸리에 트랜스폼(inverse Fourier transform)에 의한 자기상관을 획득하는 것 중 하나는 가능하다.

- [0145] 이제, LTP 를 이용하는 확장(extension)이 설명된다.
- [0146] LTP(long-time predictor)는 또한 실제로 선형 예측(linear predictor)이다.
- [0147] 실시예에 따르면, 행렬 결정기(matrix determiner) 110는 지각적으로 가중된 선형 예측에 의존하여, 예컨대, LTP(long-time predictor)에 의존하여, 자기상관 행렬 (autocorrelation matrix)R을 결정하도록 구성된다.
- [0148] LP와 LTP는 하나의 공동 예측인자(one joint predictor)로 컨볼루션 될(be convolved) 수 있다. 이는 스펙트럼 포락선 형태(spectral envelope shape)와 고조파 구조(harmonic structure) 둘 다를 포함한다. 이러한 예측인자(predictor)의 임펄스 응답은 매우 길 것이다. 이로 인해, 종래기술에서 취급하기 매우 어려웠다. 그러나, 만약 선형 예측(linear predictor)의 자기상관이 이미 알려진 것이라면, 그러면 공동 예측인자(joint predictor)의 자기상관(autocorrelation)은 전후방 LTP로의 자기상관을 간단히 필터링 함으로써, 또는 주파수 영역(frequency domain)에서 유사한 과정에 의해 연산 될 수 있다.
- [0149] 이전에 LTP를 이용하는 방법은 LTP 지연(LTP lag)이 프레임 길이보다 짧으면, LTP가 프레임 내에서 피드백 루프(feedback loop)를 야기하기 때문에 문제가 있었다. 목적함수(objective function)에서 LTP를 포함하는 것의 이점은, LTP의 지연(lag of the LTP)이 프레임 길이(frame length)보다 짧을 때, 이러한 피드백이 최적화에 명시적으로 고려된다는 것이다.
- [0150] 이하에서, 비상관 영역(uncorrelated domain)에서의 빠른 최적화(optimisation)가 설명된다.
- [0151] ACELP 시스템 디자인의 주요 과제는 연산 복잡도(computational complexity)의 절감(reduction)이어야 왔다. ACELP 시스템은 LP가 야기하는 잔류 샘플들 사이의 복잡한 상관관계(correlations)에 의한 필터링 때문에 복잡하고, 이는 행렬 B 또는 행렬 R에 의한 현재 컨텍스트(current context)에서 설명된다. $e(n)$ 샘플들은 상관되어(correlated)있기 때문에, (단지) 원하는 정확도(accuracy)로 $e(n)$ 을 양자화(quantise)하는 것이 불가능 하지만, 시행 착오 접근(trial-and-error) 방식의 다른 양자화(quantisations)들의 많은 조합(combinations)은 수학식 3 또는 수학식 10 각각의 목적함수(objective function)에 대한 최적의 양자화(best quantisation)를 찾기 위해, 시도된다.
- [0152] 행렬(matrix) R의 도입에 의해, 이러한 상관관계(correlations)에 대한 새로운 관점이 획득된다. 즉, R은 허미시안 폴리츠 구조(Hermitian Toeplitz structure)를 가지고, 몇몇의 효율적인 행렬 분해(efficient matrix decompositions)는 특이 값 분해(singular value decomposition), 콜레스키분해(Cholesky decomposition) 또는 한켈 행렬(Hankel matrices)(한켈 행렬은 폴리츠 행렬의 위아래가 뒤바뀐 upside-down 행렬이다)의 반데몬드 분해(Vandermonde decomposition)에 적용될 수 있다. 이로써 같은 분해는 폴리츠 (Toeplitz) 및 한켈 행렬(Hankel matrices)에 적용될 수 있다([6] 및 [7] 참조). $R = E D E^H$ 를 D와 같은 R의 분해로 설정하자. D는 동일한 크기의 대각 행렬이고, R의 랭크(rank)이다. 그러면 수학식 9는 아래와 같이 수정된다.
- [0153]
$$\epsilon^2 = (e - \hat{e})^H R (e - \hat{e}) = (e - \hat{e})^H E D E^H (e - \hat{e}) = (f - \hat{f})^H D (f - \hat{f}) \quad (11)$$
- [0154] 여기서 $\hat{f} = E^H \hat{e}$ 이다.
- [0155] D는 다이어고널(diagonal) (대각행렬)이기 때문에, $f(k)$ 의 각 샘플에 대한 에러(오차)(error)는 다른 샘플 $f(i)$ 에 의존한다. 수학식 10에서, 코드북 벡터(codebook vector)는 최적 게인(optimal gain)에 의해 크기가 조정(스케일)(scaled)된다고 가정된다. 이로써, 새로운 목적 함수는 12이다.

$$\frac{(f^H D \hat{f})^2}{\hat{f}^H D \hat{f}}. \quad (12)$$

[0156]

[0157] 여기서, 샘플들은 다시 연관(correlated)된다(라인 하나의 양자화 변화는 모든 라인에 대한 최적 게인(optimal gain)을 변화시키기 때문에). 하지만, 수학식 10과 비교하면, 연관의 영향(effect of correlation)은 여기에서 제한된다. 그러나, 심지어 연관 (correlation) 이 고려되었다고 하더라도, 이러한 목적 함수의 최적화는 수학식 3 또는 10의 최적화 보다 더욱 간단하다.

[0158]

이러한 문제 접근 방법을 사용하면, 1 및 2가 가능하다.

[0159]

1. 원하는 정확도(desired accuracy)로 임의의 종래의 스칼라 또는 벡터 양자화 기법에 적용, 또는

[0160]

2. 임의의 종래의 ACELP 펄스 검색 알고리즘(pulse search algorithm)에서 목적 함수(objective function)로서 수학식 12를 이용.

[0161]

두 가지 접근법은 수학식 12에 대해 최적에 가까운 양자화를 제공한다. 종래의 양자화 기법은 일반적으로 어떠한 무작위 대입방법(brute-force methods)도 요구하지 않고(가능한 rate-loop는 예외임), 그리고 행렬 D는 행렬 B 또는 R 중 하나 보다 간단하고, 두 양자화 방법은 모두 종래의 ACELP 펄스 검색 알고리즘(pulse search algorithms) 보다 간단하기 때문이다. 연산 복잡성(computational complexity)의 주된 원인(main source)은 행렬분해(matrix decomposition) 연산 때문이다.

[0162]

몇몇 실시예에서 코드북의 코드북 벡터를 결정하기 위해 수학식 12를 이용한다.

[0163]

예를 들면, $R = E^H D E$ 형태의 R에 대한 몇몇의 행렬 인수들(matrix factorizations)이 존재한다.

[0164]

예컨대,

[0165]

(a) 고유값 분해(eigenvalue decomposition)는 GNU 과학 라이브러리 (http://www.gnu.org/software/gsl/manual/html_node/Real-Symmetric-Matrices.html)를 이용함으로써 연산될 수 있다. 행렬 R은 실제(실시간)(real)이고, 대칭형이다(플리츠(Toeplitz) 행렬 또한 마찬가지이다). 이로인해, “gsl_eigen_symm()” 기능이 행렬 E 및 D를 결정하는데 이용 될 수 있다. 동일한 고유 값 분해(eigenvalue decomposition)의 다른 구현 실시예들은 문헌 [6]에서 쉽게 사용 가능하다.

[0166]

(b) 플리츠 행렬(Toeplitz matrices)[7]의 반데몬드 분해(Vandermonde factorization)는 [8]에 기술된 알고리즘을 이용하여 사용될 수 있다. 이러한 알고리즘은 행렬 E 및 D를 반데몬드 행렬(Vandermonde matrix)인 E와 같이 반환하고(return), 이는 불-균일 주파수 분포(non-uniform frequency distribution)에서 불연속 푸리에 트랜스폼(discrete Fourier transform)과 동일하다.

[0167]

이러한 인수분해를 이용하면, 잔류 벡터(residual vector) e는 $f = E^H e$ 또는 $f' = D^{1/2} E^H e$ 에 의해 변환 영역(transform domain)으로 변환(transformed) 될 수 있다.

[0168]

임의의 일반적인 양자화 방법은 이러한 영역에 적용될 수 있다, 예컨대,

[0169]

1. 벡터 f 는 ACELP의 일반적인 구현으로서 정확하게 대수 코드북(algebraic codebook)에 의해 양자화 될 수 있다. 그러나, f 의 구성요소(elements)는 상관되어 있지 않기(uncorrelated) 때문에, ACELP에서 복잡한 검색 함수(search function)가 필요하지 않지만, 간단한 알고리즘은 다음과 같이 적용될 수 있다.

[0170]

(a) 초기 게인(이득)을 $g=1$ 로 설정함.

[0171]

(b) $\hat{f}' = \text{round}(gf')$. 예 의해 양자화 함.

[0172] (c) 만일, \hat{f} 에서 펠스의 수가 미리-정의된(pre-defined) 양 p 보다 크다면, $\|\hat{f}'\|_1 > p$, 계인(이득) g 를 증가시키고, step b로 돌아감.

[0173] (d) 반면, 만약 \hat{f} 에서 펠스의 수가 미리-정의된(pre-defined) 양 p 보다 작다면, $\|\hat{f}'\|_1 < p$, 계인(이득) g 를 감소시키고, step b로 돌아감.

[0174] (e) 그렇지 않으면, \hat{f}' 에서 펠스의 수가 미리-정의된(pre-defined) 양 p 와 같으면, $\|\hat{f}'\|_1 = p$, 처리과정이 중단됨.

[0175] 2. 산술 코더(arithmetic coder)는 표준 AMR-WB+ 또는 MPEG USAC에서 TCX의 스펙트럼 라인(spectral lines)의 양자화에서 사용되는 것과 유사하게 이용될 수 있다.

[0176] f' 의 구성요소는 직교(orthogonal)이기 때문에(수식 12형태에서 확인하는 바와 같이) 그리고, 그것들은 수식 12의 목적함수(objective function)에서 동일한 가중(weigh)을 가지기 때문에, 그것들은 동일한 양자화 스텝 사이즈로 개별적으로 양자화 될 수 있다. 양자화는 자동적으로 수학식 12에서 목적함수의 최적(가장 큰) 값을 찾는다(검출한다). 이는 특정(해당) 양자화 정확도(quantization accuracy)로 가능하다. 다시 말해, 양자화 알고리즘은 전술한 바와 같이, 모두 수학식 12에 대한 최적 양자화(optimal quantization)를 반환하는 것이다.

[0177] 최적화의 이점은 f' 의 구성요소들(elements)이 개별적으로 처리되는 것과 관련이 있다. 만약, 코드북 방법이 사용되는 경우, 여기에서 코드북 벡터 c_k 는 (적어도 하나 이상의 0이 아닌 구성요소를 가지는) non-trivial 이다. 그러면 이러한 코드북 벡터들은 더 이상 독립적인 구성요소(independent elements)를 가지지 않고, 행렬 분해(matrix factorization)의 이점도 순실한다.

[0178] 다시 말해, 반데몬드 행렬(Vandermonde matrix)은 주파수-변형 푸리에 트랜스폼(frequency-warped Fourier transform)에 대응한다. 이런 경우, 벡터 f 는 변형된 주파수 스케일(warped frequency scale)에서의 잔류 신호의 주파수 영역 표현(frequency domain representation)에 대응하는 것 이 된다 ([8] 의 “root-exchange property” 참조).

[0179] 이러한 결과가 잘 알려져 있지 않다는 점을 예고하는 것이 중요하다. 실제로, 이 결과는 만약 신호 x 가 컨볼루션 행렬(convolution matrix) C 로 필터링 되면 (13)의 결과를 나타낸다.

$$\|Cx\|^2 = \|DVx\|^2 \quad (13)$$

[0180] 여기에서 V 는 (예컨대 변형된) 푸리에 트랜스폼(단위 원(unit circle)의 구성요소의 반데몬드 행렬인) 및 대각 행렬(diagonal matrix) D 이다. 즉, 만약 필터링 된 신호(filtered signal)의 에너지를 측정하고자 하는 경우, 주파수-변형 신호(frequency-warped signal)의 에너지는 등가적으로 측정된다. 반대로, 변형 푸리에 도메인(warped Fourier domain)에서 수행되어야 하는 임의의 측정(evaluation)은 등가적으로 필터링 된 시간-영역(filtered time-domain)에서 수행될 수 있다. 시간과 주파수의 이중성(duality) 때문에, 시간-영역 윈도윙(time-domain windowing) 및 시간-변형(time-warping) 또한 존재한다. 그러나 현실적인 이슈는, 전술한 관계를 만족 시키는 컨볼루션 행렬(convolution matrix) C 를 찾는 것은 수치적으로 민감한 문제이기 때문에, 종종 근사화된 해(approximate solutions) \hat{C} 를 대신 찾는 것이 보다 용이하다.

$$\|Cx\|^2 = \|DVx\|^2 \quad \text{관계는 코드북의 코드북 벡터를 결정하는데 사용될 수 있다.}$$

[0183] 이를 위해, 여기에서 가장먼저, H 에 의해, 수학식 2에서와 같은 컨볼루션 행렬(convolution matrix) 은 C 대신에 표시 되는 것을 유념해야 한다. 만약, 그러면, 하나의 (해는) 양자화 잡음(quantization noise)

$e = Hx - H\hat{x}$ 을 최소화 하기를 원하고, 그것의 에너지는 아래와 같이 측정될 수 있다:

$$\begin{aligned} \varepsilon^2 &= \|Hx - H\hat{x}\|^2 = \|H(x - \hat{x})\|^2 = (x - \hat{x})^T H^T H(x - \hat{x}) = (x - \hat{x})^T R(x - \hat{x}) = (x - \hat{x})^T V^H D V(x - \hat{x}) = \|D^{1/2} V(x - \hat{x})\|^2 = \|D^{1/2} V(x - \hat{x})\|^2 = \\ [0184] \quad &\|D^{1/2}(f - \hat{f})\|^2 = \|f' - \hat{f}'\|^2. \end{aligned} \quad (13a)$$

[0185] 이제, 프레임-독립(frame-independence)에 대한 확장을 설명한다.

[0186] 인코딩 된 신호(encoded speech signal)가 라디오-웨이브(radio-waves)와 같은 불완전한 전송 라인(imperfect transmission lines)을 통해 전송 될 때, 예외 없이, 데이터 패킷이 손실되는 경우가 될 것이다. 만일 프레임들이 서로 의존해 있다면, 패킷 N을 N-1로 완벽하게 디코드 하는 것이 요구되는 것과 같이,(그리고) 패킷 N-1의 손실은 두 패킷 N-1 과 N의 합성을 손상시킬 것이다. 반면에, 만약 프레임들이 독립적인 경우, 그러면 패킷 N-1의 손실은 오직 패킷 N-1의 합성만을 손상시킬 것이다. 그러므로, 프레임간 의존(inter-frame dependencies)으로부터 자유로운 방법을 고안하는 것이 중요하다.

[0187] 종래의 ACEL시스템에서, 프레임간 의존의 주요 원인은 LTP와 또한 어느 정도의 LP이다. 구체적으로, 두 가지 모두 무한 임펄스 응답(IIR, infinite impulse response) 필터이기 때문에, 손상된 프레임은 손상된 샘플의 “무한” 테일(tail)을 야기한다. 실제로, 상기 테일(tail)은 몇몇 개의 프레임 만큼 길어 질 수 있고, 이는 인지적으로 거슬린다.

[0188] 본 발명의 프레임 워크(frame-work)를 사용하면, 프레임 간 종속성(inter-frame dependency)이 생성되는 경로는 현재 프레임으로부터 다음 프레임으로의 ZIR에 의해 정량화 될 수 있는 것이 실현 된다. 이러한 프레임간 종속성을 회피하기 위해, 종래의 ACELP에 대한 세 가지 수정을 구현할 필요가 있다.

[0189] 1. 이전 프레임으로부터 현재(서브)프레임으로 ZIR을 계산할 때, 양자화된 잔류가 아니라, 원래의(양자화 되지 않은) 0으로 확장된 잔류를 계산 해야만 한다. 이런 방식으로, 이전 (서브) 프레임으로부터의 양자화 오류(quantization errors)는 현재 (서브) 프레임으로 전파되지 않을 것이다.

[0190] 2. 현재 프레임을 양자화 할 때, 원신호(original)와 양자화된 신호들(quantized signals) 간 다음 프레임으로의 ZIR 오차(error)는 반드시 고려되어야 한다. 이는 상관 행렬(correlation matrix)B를 자기 상관 행렬(autocorrelation matrix) R로 대체 함으로써, 수행 될 수 있다. 이것은 다음 프레임으로의 ZIR에서의 에러(error)가 현재 프레임 내의 오류(error)와 함께 최소화되도록 보장한다.

[0191] 3. 에러 전파(error propagation)는 두 LP 와 LTP 로 기인되는 것이기 때문에, 두 구성요소는 ZIR에 포함되어야 한다. 여기에서 ZIR이 오직 LP에 대해서만 계산된다는 것이 종래의 접근 방법과의 차이이다.

[0192] 만약 현재 프레임을 양자화 할 때, 이전 프레임의 양자화 에러(quantization errors)가 고려되지 않는다면, 출력의 인지 품질 (perceptual quality)에 대한 효율성이 상실된다. 따라서, 에러 전파(error propagation)의 위험이 없을 때를 고려해야 이전 오류(previous errors)를 택할 수 있다. 예컨대, 종래의 ACELP 시스템은 매 20ms마다 프레임을 4 또는 5개의 서브 프레임들(subframes)로 서브-분할(sub-divided)하는 프레이밍(framing)을 적용한다. LTP와 잔류는 각 서브 프레임에 대하여 개별적으로 양자화 되고, 코드화되지만, 전체 프레임은 데이터의 한 블록으로서 전달된다. 따라서, 개별적인 서브프레임들은 손실될 수 있지만, 이는 오직 완전한 프레임에 한한다. 오직 프레임 경계에서만 프레임-독립 ZIR들을 사용하도록 요구되는 것이 뒤따르지만, ZIR들은 남아있는 서브프레임들 사이의 프레임간 의존으로 사용될 수 있다.

- [0193] 실시예는 종래의 ACELP알고리즘을 다음 프레임 으로의 현재 프레임의 임펄스 응답의 영향을 포함하는 것에 의해, 현재 프레임의 목적 함수(objective function)로 수정한다. 최적화 문제의 목적 함수에서, 이러한 수정은 허미시안 플리츠 구조(Hermitian Toeplitz structure)를 가지는 자기상관 행렬(autocorrelation matrix)로 상관 행렬(correlation matrix)을 대체하는 것에 대응한다. 이러한 수정은 아래와 같은 이익을 가진다.
- [0194] 1. 연산 복잡도와 메모리 요구사항이 자기상관 행렬의 허미시안 플리츠 구조 추가 때문에 감소된다.
- [0195] 2. 같은 인지 모델(perceptual model)이 모든 샘플에 적용될 것이다. 디자인과 인지 모델의 조정(tuning)을 보다 간단히 하고, 그것의 적용을 보다 효율적이고 지속적으로 한다.
- [0196] 3. 프레임간 상관(Inter-frame correlations)은 현재 프레임의 양자화에서, 이전 프레임(previous frame)으로 부터의 임펄스 응답과 다음 프레임으로의 양자화되지 않은 임펄스 응답(unquantized impulse response)만을 고려함으로써, 완전하게 회피 될 수 있다. 이는 패킷-순서가 예상되는 곳에서 시스템 안정성을 개선한다.
- [0197] 도 2는 전술한 실시예에 따라 디코딩된 음성 신호(speech signal)를 획득하기 위한 장치에 의해 인코딩 되는 인코딩 된 음성신호를 디코딩 하는 디코더 220을 나타낸 도면이다. 디코더 220은 인코딩 된 음성 신호(encoded speech signal)를 수신하도록 구성된다. 이때, 인코딩 된 음성 신호(encoded speech signal)는 전술한 실시예 중 하나에 따른 음성 신호를 인코딩 하는 장치에 의해 결정된, 코드북 벡터의 표시(indication of the codebook vector) 예컨대, 결정된 코드북 벡터의 인덱스(index)를 포함한다. 또한, 디코더 220은 코드북 벡터에 의존하는 디코딩된 음성 신호를 획득하기 위해, 인코딩 된 음성 신호(encoded speech signal)를 디코드 하도록 구성된다.
- [0198] 도 3은 실시예에 따른 시스템을 나타낸 도면이다. 상기 시스템은 전술한 실시예 중 어느 하나에 따라 인코딩 된 음성 신호(encoded speech signal)를 획득하기 위해 입력 음성신호(input speech signal)를 인코딩 하는 장치 210을 포함한다. 인코딩 된 음성 신호(encoded speech signal)는 음성신호 인코딩을 위한 장치 210에 의해 결정된 코드북 벡터의 표시(indication of the determined codebook vector)를 포함한다. 예컨대, 코드북 벡터의 인덱스를 포함한다. 또한, 시스템은 디코딩된 음성 신호를 획득하기 위해 인코딩 된 음성 신호를 디코딩하는 본 발명의 실시예에 따른 디코더 220를 포함한다. 디코더 220은 인코딩 된 음성 신호를 수신하도록 구성된다. 또한, 디코더 220은 결정된 코드북 벡터에 의존하는 디코딩된 음성 신호를 획득하기 위한 인코딩 된 음성 신호를 디코드 하도록 구성된다.
- [0199] 몇몇 측면은 장치의 맥락에서 설명되었지만, 이를 측면은 또한 디바이스가 블록 또는 방법 단계 또는 단계에 있어서의 기능에 대응하는 대응하는 방법의 설명을 나타낸다. 유사하게, 방법 단계들의 문맥으로 설명되는 측면은 해당 장치에 대응하는 블록 또는 항목 또는 기능에 대한 설명을 나타낸다.
- [0200] 본 발명의 분해된 신호(decomposed signal)는 디지털 저장 매체에 저장 될 수 있다. 또는 무선 전송 매체나 인터넷 등의 유선의 전송 매체 등의 전송 매체 상에 전송 될 수 있다.
- [0201] 특정 구현 요구 사항에 따라, 본 발명의 실시 예들은 하드웨어 또는 소프트웨어로 구현 될 수 있다. 전자적으로 판독 가능한 제어 신호가 저장을 갖는 (구현을 협력하는), 예컨대, 플로피 디스크, DVD, CD, ROM, PROM, EPROM, EEPROM 또는 FLASH 메모리를 디지털 저장 매체를 이용하여 수행 될 수 있다. 또는, 각각의 방법이 수행되도록 프로그래머블 컴퓨터 시스템과 협력 할 수 있다.
- [0202] 본 발명에 따른 일부 실시 예는 본원에 기재된 방법 중 하나가 수행되도록 프로그램 가능한 컴퓨터 시스템과 협력 할 수 있는 전자적으로 판독 가능한 제어 신호를 갖는 비 일시적인 데이터 캐리어를 포함한다.
- [0203] 일반적으로, 본 발명의 실시 예는 프로그램 코드를 가진 컴퓨터 프로그램 제품, 컴퓨터 프로그램 제품이 컴퓨터

에서 실행될 때 방법 중 하나를 수행하기 위한 프로그램 코드 인 동작으로 구현 될 수 있다. 프로그램 코드는, 예컨대, 머신 판독 가능 캐리어(machine readable carrier)상에 저장 될 수 있다.

[0204] 다른 실시 예는 기계 판독 가능 캐리어 상에 저장된 본 명세서에 기술 된 방법 중 하나를 수행하기 위한 컴퓨터 프로그램을 포함한다.

[0205] 즉, 본 발명의 방법 실시 예는, 따라서, 컴퓨터 프로그램이 컴퓨터에서 실행될 때, 본 명세서에 설명 된 방법 중 하나를 수행하기 위한 프로그램 코드를 갖는 컴퓨터 프로그램으로 구현 될 수 있다.

[0206] 본 발명의 방법에 따른 또 다른 실시 예는, 데이터 캐리어 (또는 디지털 저장 매체, 또는 컴퓨터 판독 가능 매체)를 포함하는, 본원에 기재된 방법 중 하나를 수행하기 위한, 컴퓨터 프로그램을 기록한 것이다.

[0207] 따라서, 본 발명의 방법의 또 다른 실시 예는, 데이터 스트림 또는 본원에 기재된 방법 중 하나를 수행하기 위한 컴퓨터 프로그램을 나타내는 신호들의 시퀀스(sequence)이다. 데이터 스트림 또는 신호들의 시퀀스는, 예컨대, 인터넷을 통해, 예컨대, 데이터 통신 접속을 통해 전송하도록 구성 될 수 있다.

[0208] 또 다른 실시 예는, 예를 들면, 컴퓨터, 또는 프로그래머블 로직 디바이스(programmable logic device)로 구성 되거나 본원에 기재된 방법 중 하나를 수행하도록 구성된 처리 수단을 포함한다.

[0209] 또 다른 실시 예는, 여기에 설명 된 방법 중 하나를 수행하기 위한 인스톨된 컴퓨터 프로그램을 포함한다.

[0210] 일부 실시 예에서, (예컨대, 필드 프로그래머블 게이트 어레이,a field programmable gate array) 프로그래머블 논리 디바이스는 여기에 설명 된 방법의 기능의 일부 또는 모두를 수행하는데 사용될 수 있다. 일부 실시 예에서, 필드 프로그래머블 게이트 어레이는 본원에 기재된 방법 중 하나를 수행하기 위해 마이크로 프로세서와 협력 할 수 있다. 일반적으로, 방법은 바람직하게는 임의의 하드웨어 장치에 의해 수행된다.

[0211] 상술 한 실시 예들은 본 발명의 원리에 대한 예시이다. 이는 본원에 기재된 변형 및 구성의 변형(modifications and variations of the arrangements)과 세부사항은 당업자에게 명백 할 것으로 이해된다. 따라서, 단지 기재된 특허 청구 범위(impending patent claims)에 의해서만 아니라 본원의 설명 및 실시 예의 설명에 의해 제시된 특정 세부 사항에 의해 제한되는 의도이다.

[0212] References

[0213] [1] Salami, R. and Laflamme, C. and Bessette, B. and Adoul, J.P., "ITU-T G. 729 Annex A: reduced complexity 8 kb/s CS-ACELP codec for digital simultaneous voice and data ", Communications Magazine, IEEE, vol 35, no 9, pp 56-63, 1997.

[0214] [2] 3GPP TS 26.190 V7.0.0 , "Adaptive Multi-Rate (AMR-WB) speech codec ", 2007.

[0215] [3] ITU-T G.718, "Frame error robust narrow-band and wideband embedded variable bit-rate coding of speech and audio from 8-32 kbit/s ", 2008.

[0216] [4] Schroeder, M. and Atal, B., "Code-excited linear prediction (CELP): High-quality speech at very low bit rates ", Acoustics, Speech, and Signal Processing, IEEE Int Conf, pp 937-940, 1985.

[0217] [5] Byun, K.J. and Jung, H.B. and Hahn, M. and Kim, K.S., "A fast ACELP codebook search method ",

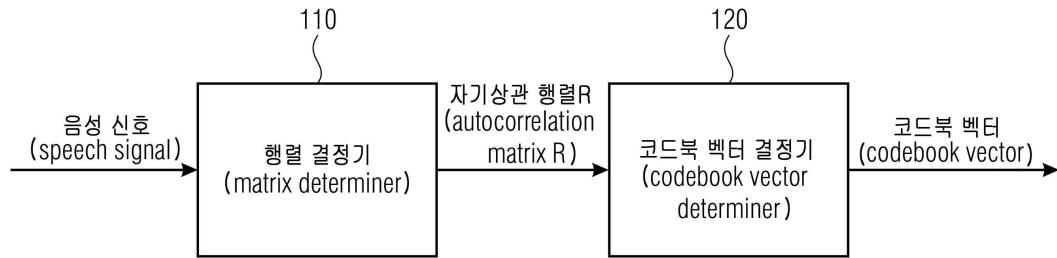
Signal Processing, 2002 6th International Conference on, vol 1, pp 422-425, 2002.

- [0218] [6] G. H. Golub and C. F. van Loan, "Matrix Computations ", 3rd Edition, John Hopkins University Press, 1996.
- [0219] [7] Boley, D.L. and Luk, F.T. and Vandevoorde, D., "Vandermonde factorization of a Hankel matrix ", Scientific computing, pp 27-39, 1997.
- [0220] [8] Backstrom, T. and Magi, C., "Properties of line spectrum pair polynomials - A review" , Signal processing, vol. 86, no. 11, pp. 3286-3298, 2006.
- [0221] [9] A. Harma, M. Karjalainen, L. Savioja, V. Valimaki, U. Laine, and J. Huopaniemi, "Frequencywarped signal processing for audio applications," J. Audio Eng. Soc, vol. 48, no. 11, pp. 1011?1031, 2000.
- [0222] [10] T. Laakso, V. Valimaki, M. Karjalainen, and U. Laine, "Splitting the unit delay [FIR/all pass filters design]," IEEE Signal Process. Mag., vol. 13, no. 1, pp. 30?60, 1996.
- [0223] [11] J. Smith III and J. Abel, "Bark and ERB bilinear transforms," IEEE Trans. Speech Audio Process., vol. 7, no. 6, pp. 697?708, 1999.
- [0224] [12] R. Schappelle, "The inverse of the confluent Vandermonde matrix," IEEE Trans. Autom. Control, vol. 17, no. 5, pp. 724~725, 1972.
- [0225] [13] B. Bessette, R. Salami, R. Lefebvre, M. Jelinek, J. Rotola-Pukkila, J. Vainio, H. Mikkola, and K. Jarvinen, "The adaptive multirate wideband speech codec (AMR-WB)," Speech and Audio Processing, IEEE Transactions on, vol. 10, no. 8, pp. 620?636, 2002.
- [0226] [14] M. Bosi and R. E. Goldberg, Introduction to Digital Audio Coding and Standards. Dordrecht, The Netherlands: Kluwer Academic Publishers, 2003.
- [0227] [15] B. Edler, S. Disch, S. Bayer, G. Fuchs, and R. Geiger, "A time-warped MDCT approach to speech transform coding," in Proc 126th AES Convention, Munich, Germany, May 2009.
- [0228] [16] J. Makhoul, "Linear prediction: A tutorial review," Proc. IEEE, vol. 63, no. 4, pp. 561?580, April 1975.
- [0229] [17] J.-P. Adoul, P. Mabilieu, M. Delprat, and S. Morissette, "Fast CELP coding based on algebraic codes," in Acoustics, Speech, and Signal Processing, IEEE Int Conf (ICASSP' 87), April 1987, pp. 1957?1960.
- [0230] [18] ISO/IEC 23003-3:2012, "MPEG-D (MPEG audio technologies), Part 3: Unified speech and audio coding," 2012.
- [0231] [19] F.-K. Chen and J.-F. Yang, "Maximum-take-precedence ACELP: a low complexity search method," in Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP' 01). 2001 IEEE International Conference on, vol. 2. IEEE, 2001, pp. 693~696.
- [0232] [20] R. P. Kumar, "High computational performance in code excited linear prediction speech model using faster codebook search techniques," in Proceedings of the International Conference on Computing: Theory and Applications. IEEE Computer Society, 2007, pp. 458~462.
- [0233] [21] N. K. Ha, "A fast search method of algebraic codebook by reordering search sequence," in Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on, vol. 1. IEEE, 1999, pp. 21~24.
- [0234] [22] M. A. Ramirez and M. Gerken, "Efficient algebraic multipulse search," in Telecommunications Symposium, 1998. ITS' 98 Proceedings. SBT/IEEE International. IEEE, 1998, pp. 231~236.
- [0235] [23] ITU-T Recommendation G.191, "Software tool library 2009 user's manual," 2009.
- [0236] [24] ITU-T Recommendation P.863, "Perceptual objective listening quality assessment," 2011.

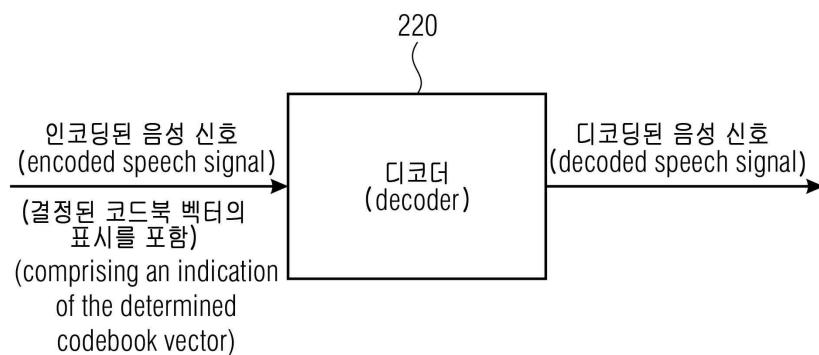
- [0237] [25] T. Thiede, W. Treurniet, R. Bitto, C. Schmidmer, T. Sporer, J. Beerends, C. Colomes, M. Keyhl, G. Stoll, K. Brandenburg et al., "PEAQ – the ITU standard for objective measurement of perceived audio quality," Journal of the Audio Engineering Society, vol. 48, 2012.
- [0238] [26] ITU-R Recommendation BS.1534-1, "Method for the subjective assessment of intermediate quality level of coding systems," 2003.

도면

도면1



도면2



도면3

