(19)대한민국특허청(KR)
(12) 공개특허공보(A)

(51) 。Int. Cl.
G10L 21/02 (2006.01)

(11) 공개번호 10-2006-0121778
(43) 공개일자 2006년11월29일

(21) 출원번호 10-2006-0103891
(22) 출원일자 2006년10월25일
심사청구일자 2006년10월25일

(71) 출원인 인하대학교 산학협력단
인천 남구 용현동 253 인하대학교

(72) 발명자 장준혁
서울 서초구 서초동 1660-31 B동 301호

(74) 대리인 김건우

전체 청구항 수 : 총 2 항

(54) 입력 음성을 수정함으로써 음성 품질을 향상시키는 방법 및이를 구현한 시스템

(57) 요약

본 발명은 입력 음성 신호를 수정함으로써 음성의 품질을 향상시키는 방법으로서, 보다 구체적으로는 Try 단계에서 실제 입력 신호를 사용하여 코덱 시스템을 먼저 평가한 후 Try 단계에서의 평가 결과에 기초하여 Catch 단계에서 입력 신호를 수정함으로써 코덱에서 출력되는 음성의 품질을 향상시키는 방법 및 이를 구현한 시스템에 관한 것이다.

본 발명은, 실제 입력 신호를 사용하여 코덱 시스템을 먼저 평가한 후 평가 결과에 기초하여 입력 신호를 수정함으로써, 종래의 방법들, 예컨대 RLS-GAbS 방법에 비해 코덱에서 출력되는 음성의 품질을 크게 향상시킬 수 있다.

대표도

도 2

특허청구의 범위

청구항 1.

입력-출력 특성을 모델링한 전달 함수가 벡터 **Q**인 코덱(CODEC)에 입력되는 입력 벡터를 **x**=[x(0), x(1), …, x(M-1)]^T, 상기 입력 벡터 **x**를 수정하여 얻어지는 수정 입력 벡터를 **y**=[y(0), y(1), …, y(M-1)]^T, 상기 수정 입력 벡터 **y**를 코더에 인가한 후 디코더에서 재합성되는 출력 벡터를 **z**=[z(0), z(1), …, z(M-1)]^{T}라고 하고, 상기 벡터 **x**, **y**, **z**의 변환 도메인 표현(transform domain representation)을 각각 **X**=[X(0), X(1), …, X(N-1)]^T, **Y**=[Y(0), Y(1), …, Y(N-1)]^T, **Z**=[Z(0), Z(1), …, Z(N-1)]^{T}라고 할 경우, 상기 입력 벡터 **x**를 상기 수정 입력 벡터 **y**로 수정함으로써 음성을 향상시키는 방법에 있어서,

t번째 입력 벡터 $\mathbf{X}_t$에 대하여, $\mathbf{X}_t$를 제1 수정 입력 벡터 $\mathbf{Y}_{t1}$에 직접 할당하는 단계;

상기 제1 수정 입력 벡터 $\mathbf{Y}_{t1}$를 상기 코덱에 인가함으로써 제1 출력 벡터 $\mathbf{Z}_{t1}$를 얻는 단계;

상기 제1 수정 입력 벡터 $\mathbf{Y}_{t1}$와 제1 출력 벡터 $\mathbf{Z}_{t1}$를 이용하여, 다음 식에 기초하여 상기 전달 함수 Q에 대한 최소 제곱 평가치(least square estimation) $\widehat{Q}_t$를 얻는 단계;

$$\widehat{Q}_t(k) = \frac{w_t Z_t(k) Y_t^*(k)}{w_t |Y_t(k)|^2}$$

(여기서, w_t는 t번째 프레임의 입력-출력 쌍($\mathbf{Y}_t$,$\mathbf{Z}_t$)에 할당되는 가중 팩터(weighting factor)를 나타냄)

상기 단계에서 얻어진 전달 함수 $\widehat{Q}_t$에 기초하여 입력 벡터 $\mathbf{X}_t$를 수정함으로써 제2 수정 입력 벡터 $\mathbf{Y}_{t2}$를 얻는 단계;

상기 제2 수정 입력 벡터 $\mathbf{Y}_{t2}$를 이용하여 향상된 제2 출력 벡터 $\mathbf{Z}_{t2}$를 실제로 얻는 단계; 및

양자화된 파라미터들을 디코더를 통과시키고, 그 결과를 이용하여 다음 번(t+ 1번째) 프레임에 대한 메모리를 갱신하는 단계

를 포함하는 방법.

청구항 2.

입력-출력 특성을 모델링한 전달 함수가 벡터 Q인 코덱(CODEC)에 입력되는 입력 벡터를 $\mathbf{x}$=[x(0), x(1), …, x(M-1)]^T, 상기 입력 벡터 $\mathbf{x}$를 수정하여 얻어지는 수정 입력 벡터를 $\mathbf{y}$=[y(0), y(1), …, y(M-1)]^T, 상기 수정 입력 벡터 $\mathbf{y}$를 코더에 인가한 후 디코더에서 재합성되는 출력 벡터를 $\mathbf{z}$=[z(0), z(1), …, z(M-1)]^{T}라고 하고, 상기 벡터 $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$의 변환 도메인 표현(transform domain representation)을 각각 $\mathbf{X}$=[X(0), X(1), …, X(N-1)]^T, $\mathbf{Y}$=[Y(0), Y(1), …, Y(N-1)]^T, $\mathbf{Z}$=[Z(0), Z(1), …, Z(N-1)]^{T}라고 할 경우, 상기 입력 벡터 $\mathbf{x}$를 상기 수정 입력 벡터 $\mathbf{y}$로 수정함으로써 음성을 향상시키는 시스템에 있어서,

입력 벡터 $\mathbf{x}$를 분기시키는 스위치부;

t번째 입력 벡터 $\mathbf{X}_t$에 대하여, 상기 스위치부로부터 입력된 상기 입력 벡터 $\mathbf{X}_t$를 제1 수정 입력 벡터 $\mathbf{Y}_{t1}$로 직접 할당한 후, 제1 출력 벡터 $\mathbf{Z}_{t1}$를 생성하는 제1 코덱부;

상기 제1 수정 입력 벡터 $\mathbf{Y}_{t1}$과 출력 벡터 $\mathbf{Z}_{t1}$를 이용하여, 다음 식에 기초하여 상기 전달 함수 Q에 대한 최소 제곱 평가치(least square estimation) $\widehat{Q}_t$를 구한 후, 얻어진 상기 전달 함수 $\widehat{Q}_t$에 기초하여 스위치부로부터 분기된 입력 벡터 $\mathbf{X}_t$를 수정함으로써 제2 수정 입력 벡터 $\mathbf{Y}_{t2}$를 얻는 입력 신호 수정부;

$$\widehat{Q}_t(k) = \frac{w_t Z_t(k) Y_t^*(k)}{w_t |Y_t(k)|^2}$$

(여기서, w_t는 t번째 프레임의 입력-출력 쌍($\mathbf{Y}_t$,$\mathbf{Z}_t$)에 할당되는 가중 팩터(weighting factor)를 나타냄)

상기 제1 코덱부와 동일한 코덱부로서, 상기 제2 수정 입력 벡터 $\mathbf{Y}_{t2}$를 입력받아 제2 출력 벡터 $\mathbf{Z}_{t2}$를 실제로 생성하는 제2 코덱부; 및

상기 제2 코덱부에 기초하여 상기 제1 코덱부의 다음 번(t+1번째) 입력 신호 수정에 사용될 메모리를 갱신하는 메모리 갱신부

를 포함하는 시스템.

명세서

발명의 상세한 설명

발명의 목적

발명이 속하는 기술 및 그 분야의 종래기술

본 발명은 입력 음성 신호를 수정함으로써 음성의 품질을 향상시키는 방법으로서, 보다 구체적으로는 실제 입력 신호를 사용하여 코덱 시스템을 먼저 평가한 후 평가 결과에 기초하여 입력 신호를 수정함으로써 코덱에서 출력되는 음성의 품질을 향상시키는 방법 및 이를 구현한 시스템에 관한 것이다.

저전송율 음성 코딩 기법은 오늘날의 이동 통신 시스템에서 중요한 부분이 되어 왔다. 저전송율 음성 코더에서, 음성 신호는 대개 벡터 양자화되는 파라미터들의 벡터들로서 표현된다. 양자화를 위한 효과적인 방법은, 파라미터 값을 원래 신호를 재합성하기 위해 사용하고, 가장 정확한 재조합의 결과를 생성하는 양자화된 값을 선택하는 것이다. 이 프로시저는 폐쇄-루프 양자화(closed-loop quantization) 및 분석합성(AbS; Analysis-by-Synthesis)으로 알려져 있다. 오늘날 사용되고 있는 거의 모든 저전송율 음성 코더들은 AbS에 기초하고 있다. 최근에, 저전송율 음성 코더에 입력으로서 인가되는 신호를 수정하는 전처리기를 도입하는 시스템이 제안된 바 있다. 특히, 양자화 오차와 수정으로 인해 발생하는 왜곡을 절충하기 위한 기준이 도입되기도 하였다. 또한, 코덱 특성들이 전달 매트릭스의 형태로 설명되고, 시스템 확인을 위한 재귀적 최소 제곱(RLS; recursive least squares) 기준에 따라 평가되기도 하였다.

상기와 같은 시스템에서 코덱 특성들을 가능한 한 정확하게 평가하는 것이 주관적 음성 품질을 향상시키는데 크게 도움이 된다는 점은 이미 널리 알려져 있다. 즉, 코덱 특성들을 미리 정확하게 평가할 수 있다면, 청취자가 느끼는 주관적 음성 품질을 훨씬 향상시키는 결과를 얻을 수 있다.

발명이 이루고자 하는 기술적 과제

본 발명은 상기와 같은 인식에서 비롯된 것으로서, 코덱 시스템이 입력 신호에 어떻게 응답하는지를 실제 전송 전에 알 수 있게 함으로써, 종래의 시스템(예컨대, RLS-GAbS 시스템)에 비해 코덱 전달 매트릭스의 보다 정확한 평가가 가능하며, 그 결과 종래의 시스템에 비해 주관적 음성 품질을 훨씬 향상시킬 수 있는 입력 신호 수정에 의한 새로운 음성 품질 향상 방법 및 이를 구현한 시스템을 제공하는 것을 그 목적으로 한다. 본 발명에 의해 제안된 이러한 구조는 GAbS 메커니즘을 TC 구조(Try-and-Catch structure)와 결합시킨 것으로서, 이하에서 TC-GAbS라고 지칭하기로 한다.

발명의 구성

상기한 목적을 달성하기 위한 본 발명의 특징에 따른, 입력 음성 신호를 수정함으로써 음성 품질을 향상시키는 방법은 다음과 같다.

먼저, 입력-출력 특성을 모델링한 전달 함수가 벡터 **Q**인 코덱(CODEC)에 입력되는 입력 벡터를 $\mathbf{x}=[x(0), x(1), \cdots, x(M-1)]^T$, 상기 입력 벡터 **x**를 수정하여 얻어지는 수정 입력 벡터를 $\mathbf{y}=[y(0), y(1), \cdots, y(M-1)]^T$, 상기 수정 입력 벡터 **y**를 코더에 인가한 후 디코더에서 재합성되는 출력 벡터를 $\mathbf{z}=[z(0), z(1), \cdots, z(M-1)]^T$라고 하고, 상기 벡터 **x**, **y**, **z**의 변환 도메인 표현(transform domain representation)을 각각 $\mathbf{X}=[X(0), X(1), \cdots, X(N-1)]^T$, $\mathbf{Y}=[Y(0), Y(1), \cdots, Y(N-1)]^T$, $\mathbf{Z}=[Z(0), Z(1), \cdots, Z(N-1)]^T$라고 하자.

본 발명의 특징에 따른 음성 품질 향상 방법은,

(1) t번째 입력 벡터 $\mathbf{X}_t$에 대하여, $\mathbf{X}_t$를 제1 수정 입력 벡터 $\mathbf{Y}_{t1}$에 직접 할당하는 단계;

(2) 제1 수정 입력 벡터 $\mathbf{Y}_{t1}$를 코덱에 인가함으로써 제1 출력 벡터 $\mathbf{Z}_{t1}$를 얻는 단계;

(3) 제1 수정 입력 벡터 $\mathbf{Y}_{t1}$와 제1 출력 벡터 $\mathbf{Z}_{t1}$를 이용하여, 다음 수학식 1에 기초하여 전달 함수 Q에 대한 최소 제곱 평가치(least square estimation) $\widehat{Q}_t$를 구하는 단계;

$$\widehat{Q}_t(k) = \frac{w_t Z_t(k)\, Y_t^*(k)}{w_t |Y_t(k)|^2}$$

(여기서, w_t는 t번째 프레임의 입력-출력 쌍($\mathbf{Y}_t$,$\mathbf{Z}_t$)에 할당되는 가중 팩터(weighting factor)를 나타냄)

(4) 단계 (3)에서 얻어진 전달 함수 $\widehat{Q}_t$에 기초하여 입력 벡터 $\mathbf{X}_t$를 수정함으로써 제2 수정 입력 벡터 $\mathbf{Y}_{t2}$를 얻는 단계;

(5) 제2 수정 입력 벡터를 이용하여 향상된 제2 출력 벡터 $\mathbf{Z}_{t2}$를 실제로 얻는 단계; 및

(6) 양자화된 파라미터들을 디코더를 통과시키고, 그 결과를 이용하여 다음 번(t+ 1번째) 프레임에 대한 메모리를 갱신하는 단계

를 포함하는 것을 그 특징으로 한다.

또한, 본 발명의 특징에 따른 음성 품질 향상 시스템은,

(a) 입력 벡터 x를 분기시키는 스위치부;

(b) t번째 입력 벡터 $\mathbf{X}_t$에 대하여, 스위치부로부터 입력된 입력 벡터 $\mathbf{X}_t$를 제1 수정 입력 벡터 $\mathbf{Y}_{t1}$로 직접 할당한 후, 제1 출력 벡터 $\mathbf{Z}_{t1}$를 생성하는 제1 코덱부;

(c) 제1 수정 입력 벡터 $\mathbf{Y}_{t1}$과 제1 출력 벡터 $\mathbf{Z}_{t1}$를 이용하여, 전술한 수학식 1에 기초하여 전달 함수 Q에 대한 최소 제곱 평가치 $\widehat{Q}_t$를 구한 후, 구해진 전달 함수 $\widehat{Q}_t$에 기초하여 스위치부로부터 분기된 입력 벡터 $\mathbf{X}_t$를 수정함으로써 제2 수정 입력 벡터 $\mathbf{Y}_{t2}$를 얻는 입력 신호 수정부;

(d) 제1 코덱부와 동일한 코덱부로서, 제2 수정 입력 벡터 $\mathbf{Y}_{t2}$를 입력받아 제2 출력 벡터 $\mathbf{Z}_{t2}$를 생성하는 제2 코덱부; 및

(e) 제2 코덱부에 기초하여 제1 코덱부의 다음 번(t+ 1번째) 입력 신호 수정에 사용될 메모리를 갱신하는 메모리 갱신부

를 포함하는 것을 그 특징으로 한다.

이하에서 첨부된 도면을 참조하여, 입력 신호를 수정하여 음성 품질을 향상시키는 종래의 방법을 먼저 간단히 살펴본 후, 본 발명의 일 실시예에 따른 방법을 상세하게 설명하기로 한다.

<u>입력 신호를 수정함으로써 음성 품질을 향상시키는 방법</u>

도 1은 일반화된 분석합성(GAbS; Generalized Analysis-by-Synthesis)에 기초하여 신호를 수정하는 종래의 시스템(예컨대, RLS-GAbS 시스템)을 나타내는 도면이다. 도 1에 도시된 바와 같이, 종래의 시스템(100)은, 코더에 인가되기 전에, 입력 벡터 **x**(105)를 수정하여 수정된 입력 벡터 **y**(115)가 음성 코더에 보다 적합하도록 만드는 구조를 채택하고 있다. 여기서, **y**=[y(0), y(1), …, y(M-1)]^{T}는 입력 벡터 **x**=[x(0), x(1), …, x(M-1)]^{T}를 수정함으로써 얻어지는 수정 입력 벡터이다. **z**=[z(0), z(1), …, z(M-1)]^{T}를 **y**가 코더에서 적용되고 난 후 디코더에서 재합성되는 경우에 생성되는 코덱(110)의 출력 벡터라고 하고, 또한 **X**=[X(0), X(1), …, X(N-1)]^T, **Y**=[Y(0), Y(1), …, Y(N-1)]^T, **Z**=[Z(0), Z(1), …, Z(N-1)]^{T}를 각각 **x**(105), **y**(115), **z**(125)의 변환 도메인 표현(transform domain representation)이라고 하자.

수학적으로 다루기 쉬운 접근법을 개발하기 위해, 입력 및 출력 쌍을 갖는 코덱 시스템이 전달 매트릭스를 나타내는 **Q**에 의해 선형 시스템 모델로서 다음과 같이 근사화된다고 가정한다.

$$Z = Q\,Y$$

여기서, **Q**는 대각 매트릭스로 가정된다.

이와 같이 선형 시스템 모델로 근사화함으로써, 신호 수정 스킴에서 아주 잘 동작하는 최적해를 구하는 효과적인 방법을 찾을 수 있다. 주어진 **Q**에 대하여, 입력 벡터 **X**의 수정이 다음 기준에 따라 행해진다.

$$\hat{Y} = \arg\ \min_{Y}\ J(Y)$$

여기서, $\hat{Y}$ 은 원하는 수정 입력 벡터를 나타내며, 목적 함수 J(**Y**)는 다음 수학식 4에 의해 주어진다.

$$J(Y) = \| X - Y \|_{W}^{2} + K \| Y - QY \|_{W}^{2}$$

여기서, $\| a \|_{W}^{2} = a^{H} W a$ 이고, K는 음성 수정을 위한 양의 상수이며, W는 인지 가중 필터를 나타내고, H는 Hermitian 연산을 의미한다. 제안된 목적 함수의 정의, J(**Y**)는 다음과 같은 상당히 합리적인 가정들에 기초한다. 즉, 만약 입력 신호가 순수 음성 사운드로부터 추출된다면, 코덱 입력 및 출력 간의 차이는 작고 거의 수정이 요구되지 않을 것인 반면에, 입력 신호가 순수 음성 특성들로부터 크게 벗어난다면, 양자화 오차는 더욱 커질 것이며 입력 신호를 코덱 양자화 오차가 감소할 수 있도록 수정하는 것이 더 좋을 것이라는 가정들에 기초한다.

J(**Y**)는 **Y**의 이차 함수이므로, **Y**에 대하여 편미분하고 0으로 두면 다음 수학식 5를 얻게 된다.

$$\frac{\partial J(Y)}{\partial Y} = -2W(X - Y) + 2K\tilde{Q}^{H} W \tilde{Q} Y = 0$$

여기서, $\tilde{Q} = I - Q$ 이고, **I**는 N x N 항등 매트릭스(identity matrix)를 의미한다. 상기 수학식 5를 풀면, 다음 수학식 6과 같은 결과를 얻는다.

$$\hat{Y} = [W + K\tilde{Q}^{H} W \tilde{Q}]^{-1} WX$$

시스템 전달 매트릭스 Q는 신호 수정에서 중요한 팩터로서, 종래에는 재귀적 최소 제곱법(RLS)을 사용하여 일반화된 AbS에 의해 평가되었다. T개의 입력 벡터, Y_1, Y_2, ..., Y_T를 음성 코덱에 인가하여, 대응 출력 벡터 Z_1, Z_2, ..., Z_T를 얻는다고 가정하자. 최소 제곱 평가(LSE) 기준에 따르면, $\hat{Q}$ 은 다음 수학식 7과 같이 얻어진다.

$$\begin{aligned}\hat{Q} &= \operatorname{argmin}_Q \sum_{t=1}^{T} w_t \| Z_t - QY_t \|^2 \\ &= \operatorname{argmin}_Q \sum_{t=1}^{T} w_t \left(\sum_{k=0}^{N-1} |Z_t(k) - Q(k)\, Y_t(k)|^2 \right)\end{aligned}$$

여기서, $\hat{Q}$ 은 Q에 대한 최소 제곱 평가치이며, w_t는 t번째 프레임 입력-출력 쌍, (Y_t, Z_t)에 할당되는 가중 팩터를 나타낸다.

오차 항을 나타내는 함수 E(Q)는 N개의 개별적인 함수들 $E_1(Q(1))$, $E_2(Q(2))$, ..., $E_{N-1}(Q(N-1))$로 분해되는데, 여기서 각각의 개별적인 함수들은 다음 수학식 8과 같이 주어진다.

$$E_k(Q(k)) = \sum_{t=1}^{T} w_t |Z_t(k) - Q(k)\, Y_t(k)|^2, k = 0,1,...,N-1$$

수학식 8의 E(Q)를 Q에 대하여 편미분하면, 다음 수학식 9를 얻는다.

$$\begin{aligned}\frac{\partial E(Q)}{\partial Q(k)} &= \frac{\partial E_k(Q(k))}{\partial Q(k)} \\ &= -\sum_{t=1}^{T} w_t (Z_t^*(k) - Y_t^*(k) Q_t^*(k))\, Y_t(k)\end{aligned}$$

여기서, k = 0, 1, ..., N-1이다.

각각의 미분식을 0으로 두고 풀면 다음 수학식 10과 같은 결과를 얻는다.

$$\hat{Q}(k) = \frac{\sum_{t=1}^{T} w_t Z_t(k)\, Y_t^*(k)}{\sum_{t=1}^{T} w_t |Y_t(k)|^2}$$

종래의 음성 품질 향상 방법에서는, 수학식 10에서 Q에 대한 평가를 RLS에 기초한 GAbS 방법을 사용하여 구현하였다.

GAbS에 의한 시스템 확인에서의 새롭게 제안하는 구조

전술한 바와 같이, 전달 매트릭스 Q는 입력 수정 전에 순차적인 방식으로 적절한 망각 팩터(forgetting factor)를 가지고 평가되는데, 이것은 코덱의 전달 특성들이 천천히 변화된다는 가정에 기초한 것이다. 그러나 코덱 입력-출력 특성들은 주

어진 음성 프레임에 따라 흔히 시변적이 되기도 하므로, 종종 RLS에 기초한 GAbS는 급격히 변화되는 음성 프레임들에 대해서는 제대로 동작하지 않는 경우가 관찰된다. 이러한 이유로, 코덱 시스템의 시변적 특성들을 고려하고 Q에 대한 평가의 정확도를 향상시키기 위해 시스템 확인에 대한 새로운 구조적 접근법이 본 발명에서 제안된다. 실제 코덱 입력-출력 특성들은 현재의 입력 신호가 코덱에 인가되기 전까지는 알려지지 않는다는 점에 주목하여, 본 발명에서는 이전 신호 대신에 현재 신호에 대한 실제 코덱 시스템의 응답을 분석하는 시스템에 대하여 제안하고 있다. 이러한 이유로, 우리는, 코덱 전달 매트릭스가 try 부분에서 평가된 후 catch 프로세스가 try 부분에서 평가된 Q를 사용하여 입력 신호를 수정하는 TC-GAbS를 채택한다. 여기서, 수정 프로세스 후에 catch 프로세스의 모든 메모리가 다음 프레임에 대한 try 프로세서로 복사되어야 한다는 점에 주의해야 한다. 코덱에서 메모리는, 메인 코더 또는 디코더 프로시저의 각각의 호 사이에서 유지되어야만 하는 모든 정적 변수들 및 어레이들을 지칭하며, 코더가 스위치 오프 될 때까지 다른 목적으로는 사용될 수 없다.

본 발명에 따라 제안된 구조의 일 실시예가 도 2에 도시된다. 도 2는 본 발명에 따른 TC-GAbS(try-and-catch structure based on Generalized Analysis-by-Synthesis) 시스템의 일 실시예를 나타내는 도면이다. 도 2에 도시된 바와 같이, TC-GAbS 스킴을 사용하여 입력 신호를 수정함으로써 음성 품질을 향상시키는 본 발명의 일 실시예에 따른 음성 품질 향상 방법은 다음과 같이 Try 루틴과 Catch 루틴으로 구성될 수 있다.

Try 루틴

1) t번째 입력 벡터 $\mathbf{X}_t$(205)를 제1 수정 입력 벡터 $\mathbf{Y}_{t1}$(215)에 직접 할당하는 단계.

2) 제1 수정 입력 벡터 $\mathbf{Y}_{t1}$(215)를 제1 코덱부(210)에 인가함으로써 제1 출력 벡터 $\mathbf{Z}_{t1}$(225)를 얻는 단계.

3) 제1 수정 입력 벡터 $\mathbf{Y}_{t1}$(215)와 제1 출력 벡터 $\mathbf{Z}_{t1}$(225)를 이용하여 전술한 수학식 1(참고를 위해, 아래에 다시 한 번 기재함)에 기초하여 전달 매트릭스 $\widehat{Q}_t$ 를 구하는 단계.

$$\widehat{Q}_t(k) = \frac{w_t Z_t(k) Y_t^*(k)}{w_t |Y_t(k)|^2}$$

Catch 루틴

1) 상기 수학식 1에 따라 구해진 전달 매트릭스 $\widehat{Q}_t$ 에 기초하여 입력 벡터 $\mathbf{X}_t$(205)를 수정함으로써 제2 수정 입력 벡터 $\mathbf{Y}_{t2}$(235)를 얻는 단계.

2) 상기 제2 수정 입력 벡터 $\mathbf{Y}_{t2}$(235)를 이용하여 향상된 제2 출력 벡터 $\mathbf{Z}_{t2}$(245)를 실제로 얻는 단계.

3) 양자화된 파라미터들을 디코더를 통과시키고, 그 결과를 이용하여 다음 번(t+1번째) 프레임에 대한 메모리를 갱신하는 단계.

음성 품질 향상에서 종래의 방법에 비해 본 발명에 따른 방법이 우수한 결과를 낼 것이라는 것은 충분히 예상할 수 있지만, 그 대가로 종래의 방법과 비교하여 계산적인 복잡함 또한 예상된다. 종래의 시스템과 제안된 시스템은 인텔사의 펜티엄 IV 2.8GHz를 구비한 PC를 사용하여 실시간으로 구현되었다. 이러한 하드웨어에 기초하여, 마이크로소프트사의 Visual C++에서의 프로파일러(profiler)를 사용하여 전체 연산 시간을 계산하였다. 여기서, 레퍼런스 C 코드(부동 소수점)의 각각의 함수에 대한 개별적인 시간 및 전체 시간을 체크할 수 있었다. 전체 시스템에 대하여, 제안된 TC-GAbS 스킴의 연산 시간은, 평균적으로 RLS-GAbS의 연산 시간보다 39% 증가하는 것으로 확인되었다.

실험 및 결과

종래의 RLS-GAbS 방법 및 본 발명에 따른 TC-GAbS 방법으로 다수의 실험들을 수행하였다. 타깃 음성 코더로서는, G.729 CS-ACELP가 사용되었다. 음성 수정에 대해서는, 각각의 프레임의 데이터가 대응하는 이산 푸리에 변환(DFT) 계수들로 구성되는 벡터로 변환되었으며, 수정은 DFT 도메인에서 행해졌다. 모든 경우들에서, 음성 수정은 K=3의 조건 하에서 행해졌으며, 항등 매트릭스가 수학식 4에서 가중 매트릭스 W에 대해 사용되었다. 3명의 남성 화자들과 3명의 여성 화자들에 의해 말해진 24개의 문장이 실험 데이터로서 사용되었다. 음성 수정은 8kHz의 샘플링 레이트를 가진 10ms의 각각의 프레임에 대하여 행해졌다. 3가지 유형의 잡음 소스, 즉 NOISEX-92 데이터베이스로부터의 백색 잡음 및 배블 잡음과, Dynastat 데이터베이스로부터의 사무실 잡음이 0, 5, 10, 15 및 20dB의 SNR에서 깨끗한 음성 파형에 첨가되었다. 가우시안 잡음이 완전히 정상(stationary)인데 비하여, 배블 잡음은 비정상(nonstationary)이며, 사무실 잡음은 매우 비정상(highly nonstationary)이다.

제안된 방법의 성능을 평가하기 위해, 우리는 먼저 ITU-T P.862 테스트들에 의해 생성된 음성 품질의 지각적 평가(PESQ; perceptual evaluation of speech quality)를 측정하였다. PESQ 결과는 표 1에 주어진다. 표 1은 TC-GAbS 접근법이 주어진 잡음 조건들 하에서 RLS-GAbS 방법보다 성능이 뛰어나는 것을 보여주는데, 이로부터 우리는 TC-GAbS의 결합이 PESQ 점수들에 대해 분명히 긍정적인 효과를 가진다는 것을 확인할 수 있다. TC-GAbS 접근법은 깨끗한 조건에서는 RLS-GAbS의 성능과 동일하다. 아래 표 1에서, G.729A는 입력 신호에 대한 수정이 없는 것을 나타내고, RLS-GAbS와 TC-GAbS는 각각 종래 및 본 발명에 따라 입력 신호를 수정한 결과를 나타낸다.

[표 1]

잡음	SNR(dB)	G.729A	RLS-GAbS	TC-GAbS
깨끗한 음성	∞	3.25	3.25	3.25
백색 잡음	0	1.48	1.58	1.61
	5	1.74	1.84	1.88
	10	2.15	2.24	2.27
	15	2.45	2.53	2.55
	20	2.78	2.85	2.87
배블 잡음	0	1.80	1.95	1.99
	5	2.14	2.31	2.36
	10	2.56	2.72	2.75
	15	2.81	2.93	2.96
	20	3.08	3.15	3.18
사무실 잡음	0	1.77	1.91	1.95
	5	2.09	2.26	2.30
	10	2.54	2.69	2.74
	15	2.78	2.90	2.94
	20	3.04	3.12	3.15

2번째 실험으로, 제안된 TC-GAbS 스킴의 주관적 품질을 평가하기 위해, 우리는 일련의 약식 듣기 테스트를 하였다. 주관적 의견 점수가 12명의 청취자에 의해 결정되었으며, 각각의 청취자는 각각의 테스트 문장에 대해 5(Excellent), 4(Good), 3(Fair), 2(Poor), 1(Bad)과 같은 점수를 평가하였다. 그 후, 점수들을 평균하여 평균 의견 점수(MOS; mean opinion score) 결과를 도출하였다. MOS 결과는 표 2에 도시된다. 표 2로부터, TC-GAbS 스킴을 채용하여 얻어지는 결과는 종래의 RLS-GAbS의 결과보다 항상 더 좋다는 것을 확인할 수 있었다. 성능 향상은 백색 잡음의 경우보다 배블 잡음 및 사무실 잡음 환경들에 대해 더 크게 나타났다. PESQ 점수를 사용하는 향상은 다소 작았으나, MOS 테스트가 다시 한 번 확인시켜주는 것처럼, 상당히 더 작은 양자화 잡음만이 가청적이었다.

[표 2]

잡음	SNR(dB)	G.729A	RLS-GAbS	TC-GAbS
깨끗한 음성	∞	4.30 ± 0.03	4.30 ± 0.03	4.30 ± 0.03

백색 잡음	0	1.37 ± 0.05	1.60 ± 0.05	1.66 ± 0.05
	5	1.66 ± 0.07	1.89 ± 0.07	1.95 ± 0.08
	10	1.91 ± 0.08	2.17 ± 0.09	2.26 ± 0.09
	15	2.10 ± 0.10	2.48 ± 0.11	2.53 ± 0.12
	20	2.33 ± 0.14	2.76 ± 0.14	2.85 ± 0.14
배블 잡음	0	1.85 ± 0.06	2.20 ± 0.06	2.30 ± 0.06
	5	2.20 ± 0.08	2.58 ± 0.08	2.66 ± 0.08
	10	2.49 ± 0.09	2.82 ± 0.10	2.86 ± 0.09
	15	2.79 ± 0.11	3.10 ± 0.11	3.18 ± 0.12
	20	2.98 ± 0.15	3.31 ± 0.15	3.43 ± 0.15
사무실 잡음	0	1.80 ± 0.06	2.15 ± 0.05	2.25 ± 0.06
	5	2.17 ± 0.07	2.55 ± 0.07	2.65 ± 0.07
	10	2.46 ± 0.08	2.81 ± 0.07	2.88 ± 0.07
	15	2.73 ± 0.09	3.07 ± 0.10	3.17 ± 0.10
	20	2.98 ± 0.13	3.28 ± 0.13	3.40 ± 0.13

전술한 실험들을 통해, TC-GAbS가 RLS-GAbS에 비해 더 높은 점수를 받았는데, 이는 충분히 예상되었던 결과이다. 한편, 상기 결과는 비교 카테고리 레이팅(CCR; comparison category rating) 테스트를 통해 다시 한 번 확인될 수 있었다. CCR 테스트는 방법 A에 의해 생성된 음성이 방법 B에 의해 생성된 음성보다 얼마나 잘 인식되는지를 알려준다. 3(Much better), 2(Better), 1(Slightly better), 0(About), -1(Slightly worse), -2(Worse), -3(Much worse)과 같은 7단계로 평가되었다. 깨끗한 조건과 잡음이 섞인 조건 모두에서 선호도 테스트 점수가 표 3에 도시된다. 역시, CCR 테스트의 결과로부터도, 제안된 TC-GAbS가 시스템 확인에 대한 새로운 구조 사양(new structure specification)에 의해 음성 품질 성능을 향상시키는 것으로 확인되었다.

[표 3]

잡음	SNR(dB)	점수
깨끗한 음성	∞	0.01 ± 0.02
백색 잡음	0	0.07 ± 0.17
	5	0.07 ± 0.18
	10	0.06 ± 0.17
	15	0.06 ± 0.17
	20	0.06 ± 0.17
배블 잡음	0	0.11 ± 0.18
	5	0.10 ± 0.18
	10	0.11 ± 0.19
	15	0.10 ± 0.18
	20	0.10 ± 0.19
사무실 잡음	0	0.13 ± 0.20
	5	0.13 ± 0.20
	10	0.13 ± 0.20
	15	0.11 ± 0.21
	20	0.11 ± 0.22

이상 설명한 본 발명은 본 발명이 속한 기술분야에서 통상의 지식을 가진 자에 의하여 다양한 변형이나 응용이 가능하며, 본 발명에 따른 기술적 사상의 범위는 아래의 특허청구범위에 의하여 정해져야 할 것이다.

발명의 효과

본 발명은, 실제 입력 신호를 사용하여 코덱 시스템을 먼저 평가한 후 평가 결과에 기초하여 입력 신호를 수정함으로써, 종래의 방법들, 예컨대 RLS-GAbS 방법에 비해 코덱에서 출력되는 음성의 품질을 향상시킬 수 있다.

도면의 간단한 설명

도 1은 일반화된 분석합성(GAbS; Generalized Analysis-by-Synthesie)에 기초하여 신호를 수정하는 종래의 시스템(RLS-GAbS 시스템)을 나타내는 도면.

도 2는 본 발명에 따른 TC-GAbS(try-and-catch structure based on Generalized Analysis-by-Synthesis) 시스템의 일 실시예를 나타내는 도면.

<도면 중 주요 부분에 대한 부호의 설명>

100 : 종래의 RLS-GAbS 시스템

110 : 코덱부

120 : 입력 신호 수정부

105 : 입력 벡터

115 : 수정 입력 벡터

125 : 출력 벡터

200 : 본 발명에 따른 TC-GAbS 시스템

210 : 제1 코덱부

220 : 입력 신호 수정부

230 : 제2 코덱부

240 : 메모리 갱신부

205 : 입력 벡터

215 : 제1 수정 입력 벡터

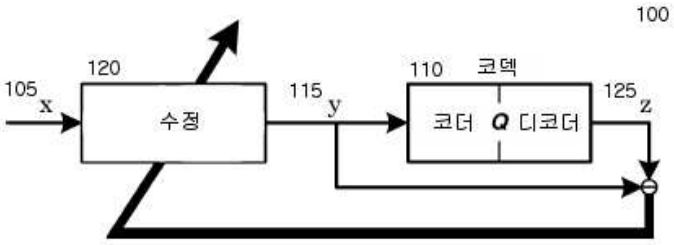
225 : 제1 출력 벡터

235 : 제2 수정 입력 벡터

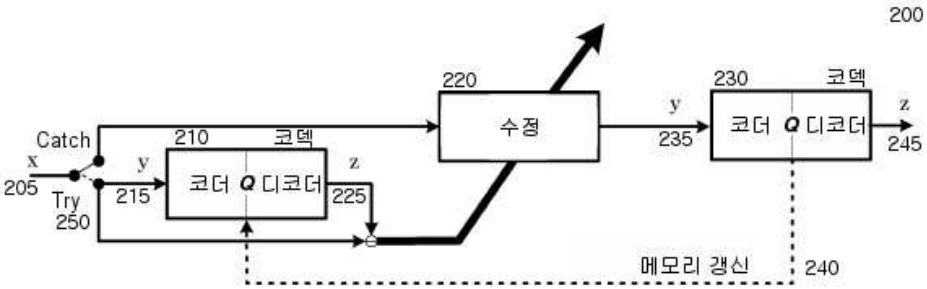
245 : 제2 출력 벡터

도면

도면1


100
120
105
x
수정
115
y
110
코덱
코더 Q 디코더
125
z

도면2


200
220
수정
230
코덱
y
z
235
코더 Q 디코더
245
Catch
x
205
Try
250
210
코덱
y
215
코더 Q 디코더
z
225
메모리 갱신
240