



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2015-0016824

(43) 공개일자 2015년02월13일

(51) 국제특허분류(Int. Cl.)

G10L 21/02 (2006.01)

(21) 출원번호 10-2013-0092752

(22) 출원일자 2013년08월05일

심사청구일자 2013년08월05일

(71) 출원인

세종대학교산학협력단

서울특별시 광진구 능동로 209 (군자동, 세종대학교)

(72) 발명자

김무영

서울 강남구 선릉로 221, 210동 402호 (도곡동, 도곡렉슬아파트)

임동석

서울 서대문구 홍지문길 13, 3층 (홍은동)

(74) 대리인

특허법인태백

전체 청구항 수 : 총 20 항

(54) 발명의 명칭 음성 향상 장치 및 그 방법

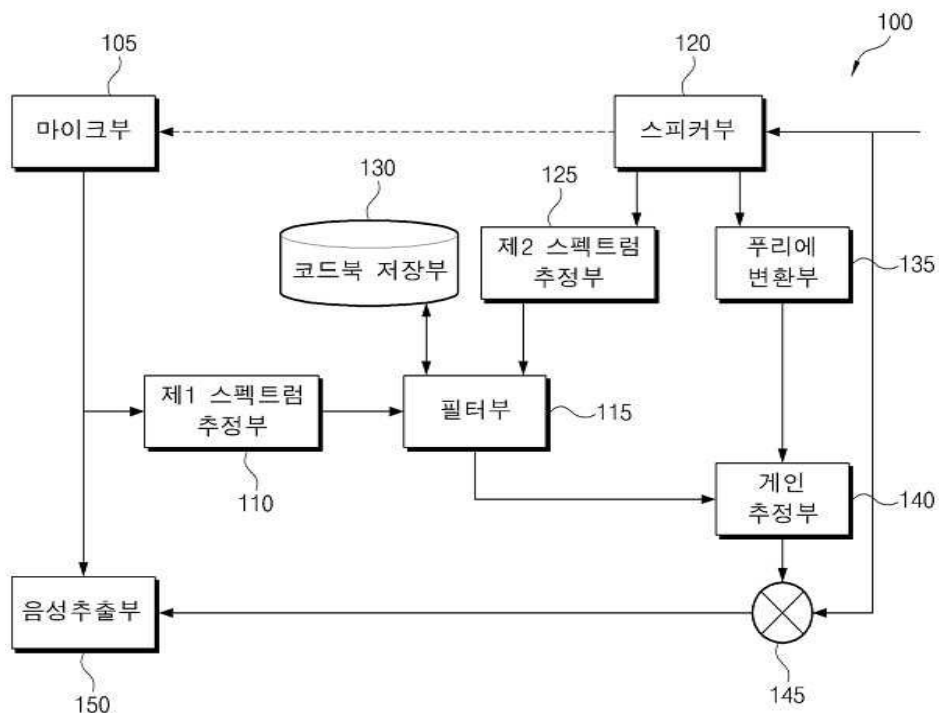
(57) 요약

본 발명은 음성 향상 장치 및 그 방법에 관한 것이다.

본 발명의 음성 향상 장치는, 마이크부로부터 입력된 노이즈 신호에 대한 LPC(Linear Prediction Coding Coefficient) 스펙트럼을 추정하는 제1 스펙트럼 추정부; 스피커부로부터 입력된 스피커 신호에 대한 LPC 스펙트럼

(뒷면에 계속)

대표도 - 도1



럼을 추정하는 제2 스펙트럼 추정부; 노이지 신호에 대한 LPC 스펙트럼과 입력된 스피커 신호에 대한 LPC 스펙트럼을 이용하여 위너 필터(Weiner filter)를 통해 스피커 신호 또는 노이지 신호에 포함된 클린 신호를 추정하는 필터부; 추정된 스피커 신호와 입력된 스피커 신호 간의 게인값(Gain)을 추정하는 게인 추정부; 및 기 저장된 지연 시간과 추정된 게인값, 입력된 스피커 신호를 이용하여 스피커 추정 신호를 생성하고, 스피커 추정 신호를 이용하여 노이지 신호로부터 스피커 신호를 제거하여 추출하는 음성 추출부를 포함한다.

이러한, 본 발명에 따르면, 종래 기기에서 이용하였던 양자화 처리 과정, 프라이어리 마이크와 RWM 등을 이용하지 않고도 종래기기보다 스피커 소음 등이 포함된 신호에서의 사용자의 음성인 노이지 신호에 대한 추출율을 상당히 향상 시킬 수 있는 효과를 기대할 수 있다.

이 발명을 지원한 국가연구개발사업

과제고유번호 711001261

부처명 미래창조과학부

연구관리전문기관 정보통신산업진흥원

연구사업명 정보통신기술인력양성

연구과제명 소환현실기반 콘텐츠 제작을 위한 실감형 인터랙션 NUI/NUX 플랫폼 기술 개발

기 여 율 1/1

주관기관 동국대학교 산학협력단

연구기간 2013.06.01 ~ 2016.12.31

특허청구의 범위

청구항 1

스피커 신호가 포함된 노이지 신호의 음성을 향상하기 위한 음성 향상 장치에 있어서,

마이크로부터 입력된 상기 노이지 신호에 대한 LPC(Linear Prediction Coding Coefficient) 스펙트럼을 추정하는 제1 스펙트럼 추정부;

스피커부로 입력되기 이전 신호에 대한 LPC 스펙트럼을 추정하는 제2 스펙트럼 추정부;

상기 노이지 신호에 대한 LPC 스펙트럼과 상기 스피커로 입력된 스피커 신호에 대한 LPC 스펙트럼을 이용하여 위너 필터(Weiner filter)를 통해 상기 스피커 신호 또는 상기 노이지 신호에 포함된 클린 신호를 추정하는 필터부;

상기 추정된 스피커 신호와 상기 입력된 스피커 신호 간의 게인값(Gain)을 추정하는 게인 추정부; 및

기 저장된 지연 시간과 상기 추정된 게인값, 상기 입력된 스피커 신호를 이용하여 스피커 추정 신호를 생성하고, 상기 노이지 신호로부터 상기 스피커 추정 신호를 제거하여 추출하는 음성 추출부

를 포함하는 음성 향상 장치.

청구항 2

제1항에 있어서,

복수의 음성 코드북을 저장하는 코드북 저장부를 더 포함하며,

상기 필터부는,

상기 노이지 신호의 LPC 스펙트럼을 기초로 상기 코드북 저장부에 포함된 음성 코드북을 선택하고,

선택된 음성 코드북에 포함된 코드벡터를 이용하여 상기 노이지 신호에 대한 LPC 스펙트럼을 추출하며,

상기 노이지 신호에 대한 LPC 스펙트럼과 상기 입력된 스피커 신호에 대한 LPC 스펙트럼을 이용하여 상기 노이지 신호를 추정하는 음성 향상 장치.

청구항 3

제2항에 있어서,

상기 필터부는,

상기 추정된 노이지 신호와 상기 입력된 노이지 신호에 최소 평균 제곱 오차(MMSE, Minimum Mean Square Error) 추정 방식을 적용하여 필터 파라미터를 산출하고, 상기 필터 파라미터를 상기 위너 필터에 적용하여 상기 스피커 신호를 추정하도록 하는 음성 향상 장치.

청구항 4

제3항에 있어서,

상기 필터부는,

상기 노이지 신호에 대한 LPC 스펙트럼과 상기 입력된 스피커 신호에 대한 LPC 스펙트럼에 대한 게인을 각각 추정하고, 다음의 수학적식을 이용하여 상기 추정된 노이지 신호($\hat{Y}^i(k)$)를 생성하는 음성 향상 장치:

$$|\hat{Y}^i(k)|^2 = g_x^i C_x^i(k) + g_d^i C_d(k)$$

여기서, g_x^i, g_d^i 는 각각 상기 노이지 신호에 대한 LPC 스펙트럼(C_x)의 게인 값, 상기 입력된 스피커 신호에 대한 LPC 스펙트럼(C_d)의 게인 값이다.

청구항 5

제4항에 있어서,

상기 필터부는,

다음의 수학적식을 이용하여 상기 노이즈 신호로부터 상기 스피커 신호를 추정하는 음성 향상 장치:

$$|\tilde{D}(k)|^2 = |Y(k)|^2 H(k) \text{ 이고,}$$

$$H(k) = \frac{\hat{g}_d C_d(k)}{\hat{g}_x \hat{C}_x(k) + \hat{g}_d C_d(k)}$$

여기서, $\tilde{D}(k)$ 는 상기 추정된 스피커 신호이고, $Y(k)$ 는 상기 노이즈 신호이다.

청구항 6

제1항 내지 제5항 중 어느 한 항에 있어서,

상기 게인 추정부는,

상기 추정된 스피커 신호와 상기 입력된 스피커 신호 사이의 상기 게인값을 다음의 수학적식을 이용하여 계산하는 음성 향상 장치:

$$\tilde{g}_* = \underset{\tilde{g}}{\operatorname{argmin}} \sum_{k=1}^{K-1} (|\tilde{D}(k)| - \tilde{g}|D(k)|)^2$$

여기서, $\tilde{D}(k)$ 는 상기 추정된 스피커 신호이고, $D(k)$ 는 상기 입력된 스피커 신호이다.

청구항 7

제6항에 있어서,

상기 지연 시간은 상기 스피커부로부터의 상기 스피커 신호 송출 시간과 상기 마이크부로부터의 상기 스피커 신호 입력 시간에 대한 차이에 대응하는 시간인 것을 특징으로 하는 음성 향상 장치.

청구항 8

제1항에 있어서,

상기 위너 필터는,

다음의 수학적식을 이용하여 주파수 도메인에서 상기 클린 신호를 추정하며,

$$H(k) = \frac{\hat{g}_x \hat{C}_x(k)}{\hat{g}_x \hat{C}_x(k) + \hat{g}_d C_d(k)}$$

상기 음성 추출부는, 다음의 수학적식을 통해 상기 클린 신호를 추출하는 음성 향상 장치:

$$|\tilde{X}(k)|^2 = |Y(k)|^2 H(k)$$

여기서, $Y(k)$ 는 상기 입력된 노이즈 신호이다.

청구항 9

제8항에 있어서,

복수의 음성 코드북을 저장하는 코드북 저장부를 더 포함하며,
 상기 필터부는,
 상기 노이지 신호의 LPC 스펙트럼을 기초로 상기 코드북 저장부에 포함된 음성 코드북을 선택하고,
 선택된 음성 코드북에 포함된 코드벡터를 이용하여 상기 노이지 신호에 대한 LPC 스펙트럼을 추출하며,
 상기 노이지 신호에 대한 LPC 스펙트럼과 상기 입력된 스피커 신호에 대한 LPC 스펙트럼을 이용하여 상기 노이지 신호를 추정하는 음성 향상 장치.

청구항 10

제9항에 있어서,
 상기 필터부는,
 상기 추정된 노이지 신호와 상기 입력된 노이지 신호에 최소 평균 제곱 오차 (MMSE, Minimum Mean Square Error) 추정 방식을 적용하여 필터 파라미터를 산출하고, 상기 필터 파라미터를 상기 위너 필터에 적용하여 상기 노이지 신호를 추정하도록 하는 음성 향상 장치.

청구항 11

제1항 내지 제5항 및 제7항 내지 제10항 중 어느 한 항에 있어서,
 상기 스피커부로 입력되기 이전 신호에 대하여 기 저장된 상기 스피커부와 상기 마이크부에 대한 잡음 정보를 기초로 상기 이전 신호에 스피커와 마이크에 대한 잡음을 콘볼루션하여 상기 제2 스펙트럼 추정부로 제공하는 잡음 삽입부를 더 포함하는 음성 향상 장치.

청구항 12

스피커 신호가 포함된 노이지 신호의 음성을 향상하기 위한 장치의 음성 향상 방법에 있어서,
 마이크로부터 입력된 상기 노이지(Noisy) 신호에 대한 LPC(Linear Prediction Coding Coefficient) 스펙트럼을 추정하는 단계;
 스피커로부터 입력된 스피커 신호에 대한 LPC 스펙트럼을 추정하는 단계;
 상기 노이지 신호에 대한 LPC 스펙트럼과 상기 입력된 스피커 신호에 대한 LPC 스펙트럼을 이용하여 위너 필터(Weiner filter)를 통해 상기 스피커 신호 또는 상기 노이지 신호에 포함된 클린 신호를 추정하는 단계;
 상기 추정된 스피커 신호와 상기 입력된 스피커 신호 간의 게인값(Gain)을 추정하는 단계;
 기 저장된 지연 시간과 상기 추정된 게인값, 상기 입력된 스피커 신호를 이용하여 스피커 추정 신호를 생성하는 단계; 및
 상기 노이지 신호로부터 상기 스피커 추정 신호를 제거하여 추출하는 단계를 포함하는 음성 향상 방법.

청구항 13

제12항에 있어서,
 상기 스피커 신호 또는 상기 클린 신호를 추정하는 단계는,
 상기 노이지 신호의 LPC 스펙트럼을 기초로 코드북 저장부에 기 저장된 음성 코드북을 선택하는 단계;
 상기 선택된 음성 코드북에 포함된 코드벡터를 이용하여 상기 노이지 신호에 대한 LPC 스펙트럼을 추출하는 단계; 및
 상기 노이지 신호에 대한 LPC 스펙트럼과 상기 입력된 스피커 신호에 대한 LPC 스펙트럼을 이용하여 상기 노이지 신호를 추정하는 단계를 포함하는 음성 향상 방법.

청구항 14

제13항에 있어서,

상기 스피커 신호 또는 상기 클린 신호를 추정하는 단계는,

상기 추정된 노이지 신호와 상기 입력된 노이지 신호에 최소 평균 제곱 오차 (MMSE, Minimum Mean Square Error) 추정 방식을 적용하여 필터 파라미터를 산출하는 단계; 및

상기 필터 파라미터를 상기 위너 필터에 적용하여 상기 스피커 신호를 추정하도록 하는 단계를 포함하는 음성 향상 방법.

청구항 15

제14항에 있어서,

상기 노이지 신호를 추정하는 단계는,

상기 노이지 신호에 대한 LPC 스펙트럼과 상기 입력된 스피커 신호에 대한 LPC 스펙트럼에 대한 계인을 각각 추정하는 단계; 및

다음의 수학적식을 이용하여 상기 추정된 노이지 신호($\hat{Y}^i(k)$)를 생성하는 단계를 포함하는 음성 향상 방법:

$$|\hat{Y}^i(k)|^2 = g_x^i C_x^i(k) + g_d^i C_d(k)$$

여기서, g_x^i, g_d^i 는 각각 상기 노이지 신호에 대한 LPC 스펙트럼(C_x)의 계인 값, 상기 입력된 스피커 신호에 대한 LPC 스펙트럼(C_d)의 계인 값이다.

청구항 16

제15항에 있어서,

상기 스피커 신호를 추정하도록 하는 단계는,

다음의 수학적식을 이용하여 상기 위너 필터가 상기 노이지 신호로부터 상기 스피커 신호를 추정하도록 하는 음성 향상 방법:

$$|\tilde{D}(k)|^2 = |Y(k)|^2 H(k) \text{ 이고,}$$

$$H(k) = \frac{\hat{g}_d C_d(k)}{\hat{g}_x \hat{C}_x(k) + \hat{g}_d C_d(k)}$$

여기서, $\tilde{D}(k)$ 는 추정된 스피커 신호이고, $Y(k)$ 는 입력되는 상기 노이지 신호이다.

청구항 17

제12항 내지 제16항 중 어느 한 항에 있어서,

상기 계인값을 추정하는 단계는,

상기 추정된 스피커 신호와 상기 입력된 스피커 신호 사이의 상기 계인값을 다음의 수학적식을 이용하여 계산하는 단계를 포함하는 음성 향상 방법:

$$\tilde{g}_* = \underset{\tilde{g}}{\operatorname{argmin}} \sum_{k=1}^{K-1} (|\tilde{D}(k)| - \tilde{g}|D(k)|)^2$$

여기서, $\tilde{D}(k)$ 는 상기 추정된 스피커 신호이고, $D(k)$ 는 상기 입력된 스피커 신호이다.

청구항 18

제12항에 있어서,

상기 스피커 신호 또는 상기 클린 신호를 추정하는 단계는,

다음의 수학적식을 이용하여 주파수 도메인에서 상기 노이즈 신호를 추정하는 단계; 및

$$H(k) = \frac{\hat{g}_x \hat{C}_x(k)}{\hat{g}_x \hat{C}_x(k) + \hat{g}_d C_d(k)}$$

다음의 수학적식을 통해 상기 노이즈 신호를 추출하는 단계를 포함하는 음성 향상 방법;

$$|\tilde{X}(k)|^2 = |Y(k)|^2 H(k)$$

여기서, $Y(k)$ 는 상기 입력된 노이즈 신호이다.

청구항 19

제18항에 있어서,

상기 스피커 신호 또는 클린 신호를 추정하는 단계는,

상기 노이즈 신호의 LPC 스펙트럼을 기초로 상기 코드북 저장부에 기 저장된 음성 코드북을 선택하는 단계;

상기 선택된 음성 코드북에 포함된 코드백터를 이용하여 상기 노이즈 신호에 대한 LPC 스펙트럼을 추출하는 단계; 및

상기 노이즈 신호에 대한 LPC 스펙트럼과 상기 입력된 스피커 신호에 대한 LPC 스펙트럼을 이용하여 노이즈 신호를 추정하는 단계를 포함하는 음성 향상 방법.

청구항 20

제18항에 있어서,

상기 클린 신호를 추정하는 단계는,

상기 추정된 노이즈 신호와 상기 입력된 노이즈 신호에 최소 평균 제곱 오차 (MMSE, Minimum Mean Square Error) 추정 방식을 적용하여 필터 파라미터를 산출하는 단계; 및

상기 필터 파라미터를 상기 위너 필터에 적용하여 상기 노이즈 신호를 추정하도록 하는 음성 향상 방법.

명세서

기술분야

[0001]

본 발명은 음성 향상 장치 및 그 방법에 관한 것으로, 스마트 기기에서 입력되는 노이즈 신호에 포함된 스피커 신호를 효율적으로 제거하여 음성을 향상시킬 수 있는 음성 향상 장치 및 그 방법에 관한 것이다.

배경기술

[0002]

스마트 기기의 음성인식을 위한 전처리 기술로써 다양한 음성 향상 기법이 사용되어 왔다. 대표적인 방법으로 Remote wireless microphone(RWM)을 이용한 음성 향상 기법이 있다.

[0003]

종래 RWM을 이용한 잡음 제거 방법은 잡음 소스에 가깝게, RWM을 설치하고, 마이크로폰(Microphone)을 사람 근처에 설치하여, 잡음 소스의 근처에 있는 RWM에서 잡음에 대한 정보를 실시간으로 수신하여 잡음에 대한 정보를 마이크로폰에 전송하고, RWM의 잡음 정보를 이용하여 마이크로폰에서 수신된 신호에 대한 잡음을 제거하는 방법

이다.

[0004] 하지만 이 방법의 단점은 마이크가 두 개가 필요하다는 점과, RWM에서 마이크로폰으로 잡음의 정보를 보내는데 추가적으로 사용되는 양자화 기법이 필요하다는 점이다. 즉 음성 향상을 위해 필요한 기법과 장비들을 많이 사용하는 단점이 있었다.

[0005] 본 발명의 배경이 되는 기술은 대한민국 공개특허공보 제2000-0050430(2000.06.15.)에 개시되어 있다.

발명의 내용

해결하려는 과제

[0006] 따라서 본 발명이 이루고자 하는 기술적 과제는 스마트 기기에서 입력되는 노이지 신호에 포함된 스피커 신호를 효율적으로 제거하여 음성을 향상시킬 수 있는 음성 향상 장치 및 그 방법에 을 제공하는 것이다.

과제의 해결 수단

[0007] 이러한 기술적 과제를 이루기 위한 본 발명의 실시예에 따른 스피커 신호가 포함된 노이지 신호의 음성을 향상 하기 위한 음성 향상 장치는, 마이크부로부터 입력된 노이지 신호에 대한 LPC(Linear Prediction Coding Coefficient) 스펙트럼을 추정하는 제1 스펙트럼 추정부; 스피커부로부터 입력된 스피커 신호에 대한 LPC 스펙트럼을 추정하는 제2 스펙트럼 추정부; 노이지 신호에 대한 LPC 스펙트럼과 입력된 스피커 신호에 대한 LPC 스펙트럼을 이용하여 위너 필터(Weiner filter)를 통해 스피커 신호 또는 노이지 신호에 포함된 클린 신호를 추정 하는 필터부; 추정된 스피커 신호와 입력된 스피커 신호 간의 게인값(Gain)을 추정하는 게인 추정부; 및 기 저장된 지연 시간과 추정된 게인값, 입력된 스피커 신호를 이용하여 스피커 추정 신호를 생성하고, 스피커 추정 신호를 이용하여 노이지 신호로부터 스피커 신호를 제거하여 추출하는 음성 추출부를 포함한다.

[0008] 여기서, 음성 향상 장치는, 복수의 음성 코드북을 저장하는 코드북 저장부를 더 포함할 수 있다.

[0009] 여기서, 필터부는, 노이지 신호의 LPC 스펙트럼을 기초로 코드북 저장부에 포함된 음성 코드북을 선택하고, 선택된 음성 코드북에 포함된 코드백터를 이용하여 노이지 신호에 대한 LPC 스펙트럼을 추출하며, 노이지 신호에 대한 LPC 스펙트럼과 입력된 스피커 신호에 대한 LPC 스펙트럼을 이용하여 노이지 신호를 추정할 수 있다,

[0010] 여기서, 필터부는, 추정된 노이지 신호와 입력된 노이지 신호에 최소 평균 제곱 오차 (MMSE, Minimum Mean Square Error) 추정 방식을 적용하여 필터 파라미터를 산출하고, 필터 파라미터를 위너 필터에 적용하여 스피커 신호를 추정할 수 있다.

[0011] 여기서, 필터부는, 노이지 신호에 대한 LPC 스펙트럼과 입력된 스피커 신호에 대한 LPC 스펙트럼에 대한 게인을 각각 추정하고, 다음의 수학적식을 이용하여 추정된 노이지 신호($\hat{Y}^i(k)$)를 생성할 수 있다.

$$|\hat{Y}^i(k)|^2 = g_x^i C_x^i(k) + g_d^i C_d(k)$$

[0013] 여기서, g_x^i, g_d^i 는 각각 노이지 신호에 대한 LPC 스펙트럼(C_x)의 게인 값, 입력된 스피커 신호에 대한 LPC 스펙트럼(C_d)의 게인 값이다.

[0014] 여기서, 필터부는, 다음의 수학적식을 이용하여 노이지 신호로부터 스피커 신호를 추정할 수 있다.

$$|\tilde{D}(k)|^2 = |Y(k)|^2 H(k) \text{ 이고,}$$

$$H(k) = \frac{\hat{g}_d C_d(k)}{\hat{g}_x \hat{C}_x(k) + \hat{g}_d C_d(k)}$$

[0017] 여기서, $\tilde{D}(k)$ 는 추정된 스피커 신호이고, $Y(k)$ 는 노이지 신호이다.

[0018] 여기서, 게인 추정부는, 추정된 스피커 신호와 입력된 스피커 신호 사이의 게인값을 다음의 수학적식을 이용하여

계산할 수 있다.

$$\tilde{g}_* = \underset{\tilde{g}}{\operatorname{argmin}} \sum_{k=1}^{K-1} (|\tilde{D}(k)| - \tilde{g}|D(k)|)^2$$

여기서, $\tilde{D}(k)$ 는 추정된 스피커 신호이고, $D(k)$ 는 입력된 스피커 신호이다.

여기서, 지연 시간은 스피커부로부터의 스피커 신호 출력 시간과 마이크부로부터의 스피커 신호 입력 시간에 대한 차이에 대응하는 시간인 것을 특징으로 할 수 있다.

여기서, 위너 필터는, 다음의 수학적식을 이용하여 주파수 도메인에서 노이지 신호를 추정할 수 있다.

$$H(k) = \frac{\hat{g}_x \hat{C}_x(k)}{\hat{g}_x \hat{C}_x(k) + \hat{g}_d C_d(k)}$$

여기서, 음성 추출부는, 다음의 수학적식을 통해 노이지 신호를 추출할 수 있다.

$$|\tilde{X}(k)|^2 = |Y(k)|^2 H(k)$$

여기서, $Y(k)$ 는 입력된 노이지 신호이다.

여기서, 음성 향상 장치는, 복수의 음성 코드북을 저장하는 코드북 저장부를 더 포함할 수 있다.

여기서, 필터부는, 노이지 신호의 LPC 스펙트럼을 기초로 코드북 저장부에 포함된 음성 코드북을 선택하고, 선택된 음성 코드북에 포함된 코드백터를 이용하여 노이지 신호에 대한 LPC 스펙트럼을 추출하며, 노이지 신호에 대한 LPC 스펙트럼과 입력된 스피커 신호에 대한 LPC 스펙트럼을 이용하여 노이지 신호를 추정할 수 있다.

여기서, 필터부는, 추정된 노이지 신호와 입력된 노이지 신호에 최소 평균 제곱 오차 (MMSE, Minimum Mean Square Error) 추정 방식을 적용하여 필터 파라미터를 산출하고, 필터 파라미터를 위너 필터에 적용하여 노이지 신호를 추정하도록 할 수 있다.

본 발명의 다른 실시예에 따른 음성 향상 방법은, 마이크부로부터 입력된 노이지(Noisy) 신호에 대한 LPC(Linear Prediction Coding Coefficient) 스펙트럼을 추정하는 단계; 스피커부로부터 입력된 스피커 신호에 대한 LPC 스펙트럼을 추정하는 단계; 노이지 신호에 대한 LPC 스펙트럼과 입력된 스피커 신호에 대한 LPC 스펙트럼을 이용하여 위너 필터(Weiner filter)를 통해 스피커 신호 또는 노이지 신호에 포함된 클린 신호를 추정하는 단계; 추정된 스피커 신호와 입력된 스피커 신호 간의 게인값(Gain)을 추정하는 단계; 기 저장된 지연 시간과 추정된 게인값, 입력된 스피커 신호를 이용하여 스피커 추정 신호를 생성하는 단계; 및 스피커 추정 신호를 이용하여 노이지 신호로부터 스피커 신호를 제거하여 추출하는 단계를 포함한다.

이와 같이 본 발명에 따르면, 종래 기기에서 이용하였던 양자화 처리 과정, 프라이어머리 마이크와 RWM 등을 이용하지 않고도 종래기기보다 스피커 소음 등이 포함된 신호에서의 사용자의 음성인 노이지 신호에 대한 추출율을 상당히 향상 시킬 수 있는 효과를 기대할 수 있다.

발명의 효과

이와 같이 본 발명에 따르면, 종래 기기에서 이용하였던 양자화 처리 과정, 프라이어머리 마이크와 RWM 등을 이용하지 않고도 종래기기보다 스피커 소음 등이 포함된 신호에서의 사용자의 음성인 클린 신호에 대한 추출율을 상당히 향상 시킬 수 있는 효과를 기대할 수 있다.

도면의 간단한 설명

도 1은 본 발명의 실시예에 따른 음성 향상 장치를 도시한 도면이다.

도 2는 본 발명의 실시예에 따른 음성 향상 장치의 음성 향상 방법을 도시한 순서도이다.

도 3a 및 도 3b는 본 발명의 실시예에 따른 음성 향상 방법과 종래의 방법간의 잡음 제거 차이를 시간 도메인과 주파수 도메인 상에 적용시킨 결과를 나타낸 표이다.

발명을 실시하기 위한 구체적인 내용

- [0034] 아래에서는 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.
- [0035] 명세서 전체에서, 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다.
- [0036] 그러면 첨부한 도면을 참고로 하여 본 발명의 실시예에 대하여 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있도록 상세히 설명한다.
- [0037] 먼저, 본 발명의 실시예에 따른 코드북 기반의 음성 향상 장치에 대하여 상세히 설명하기로 한다.
- [0038] 도 1은 본 발명의 실시예에 따른 음성 향상 장치를 도시한 도면이다.
- [0039] 도 1에 나타난 바와 같이, 본 발명의 실시예에 따른 음성 향상 장치(100)는 마이크부(105), 제1 스펙트럼 추정부(110), 필터부(115), 스피커부(120), 제2 스펙트럼 추정부(125), 코드북 저장부(130), 푸리에 변환부(135), 계인 추정부(140), 곱셈기(145) 및 음성 추출부(150)를 포함한다. 마이크부(105)는 노이즈 신호를 수신한다.
- [0040] 여기서, 노이즈 신호는 사람의 음성에 대한 신호인 클린 신호, 클린 신호에 영향을 주는 배경 잡음 신호 및 스피커 신호를 포함한다.
- [0041] 여기서, 본 발명의 실시예에 따른 음성 향상 장치(100)는 배경 잡음 신호를 제거하기 위한 후처리 수단을 더 포함할 수 있다. 이러한, 배경 잡음 신호를 제거하기 위한 후처리 수단은 당업자라면 충분히 추가할 수 있는 기술이므로, 상세한 설명은 생략하기로 한다.
- [0042] 이러한, 본 발명의 실시예에 따른 음성 향상 장치(100)는 스마트 폰, 스마트 TV와 같은 영상 통화를 가능하게 하는 기기 내에 포함될 수 있으며, 영상 통화 서비스를 제공하는 다양한 기기 내에 포함될 수도 있다.
- [0043] 제1 스펙트럼 추정부(110)는 마이크부(105)로부터 입력받은 노이즈 신호의 LPC(Linear Prediction Coding Coefficient; 이하, "LPC"라고 함) 스펙트럼(Spectrum)을 도출하여 필터부(115)로 제공한다. 구체적으로, 제1 스펙트럼 추정부(110)는 마이크부(105)로부터 노이즈 신호를 수신하면, 노이즈 신호의 LPC를 도출하고, 도출된 LPC를 이용하여 하기 수학식 2의 노이즈 신호에 대한 LPC 스펙트럼을 산출한다.
- [0044] 필터부(115)는 하기의 수학식 3 내지 6을 이용하여 제1 스펙트럼 추정부(110)에 의해 추정된 노이즈 LPC 스펙트럼과 미리 저장하고 있는 스피치 코드(Speech code)를 기초로 코드북 저장부(130)로부터 복수의 음성 코드북들 중 적어도 하나의 음성 코드북을 검출한다. 그리고, 추출한 음성 코드북을 이용하여 하기의 수학식 3을 통해 음성 코드북의 LPC 스펙트럼을 계산한다.
- [0045] 또한, 필터부(115)는 제2 스펙트럼 추정부(125)로부터 스피커부(120)가 출력하는 스피커 신호에 대한 LPC 스펙트럼을 수신하고, 검출한 음성 코드북의 LPC 스펙트럼과 스피커 신호의 LPC 스펙트럼에 대한 각각의 계인을 추정한다. 그리고, 추정된 계인을 이용하여 노이즈 스펙트럼을 추정한다.
- [0046] 또한, 필터부(115)는 추정된 노이즈 스펙트럼과 마이크부(105)로부터 출력된 노이즈 스펙트럼에 대해 하기의 수학식 8 내지 10의 최소 평균 제곱 오차(Maximum Likelihood estimation; 이하, "MMSE"라고도 함) 추정 방법을 적용하여 스피커 신호에 대한 LPC스펙트럼을 도출하고, 이를 통해 필터 파라미터를 추정한다.
- [0047] 이렇게, 얻은 필터 파라미터 $\hat{\lambda} = \{\hat{C}_x, \hat{g}_x, \hat{g}_d\}$ 를 이용하여, 필터부(115)는 하기의 수학식 10과 같이 스피커 신호를 추정하기 위한 잡음 추정식을 적용한 위너 필터를 구성하고, 위너 필터를 이용하여 마이크부(105)로부터 입력되는 노이즈 신호에 포함된 스피커 신호를 추정하여 계인 추정부(140)로 제공한다.
- [0048] 한편, 노이즈 신호의 LPC파워 스펙트럼과 스마트 기기 안에 미리 저장된 스피커 신호에 대한 LPC파워 스펙트럼을 기초로 위너 필터를 구성하여 주파수 도메인에서 스피커 신호를 제거할 수도 있다. 필터부(115)는 시간 도메인 상에서 위너 필터를 이용하여 스피커 신호의 스펙트럼을 추정하지만, 주파수 도메인에서 노이즈 신호를 추정하기 위해 하기의 수학식 13의 분자에 스피커 신호의 스펙트럼 대신, 음성 코드북의 스펙트럼으로 바꾸어 위너 필터를 구성할 수도 있다. 그리고, 하기의 수학식 14와 같이, 음성 추출부(150)는 노이즈 신호의 스펙트럼의 곱

으로 주파수 도메인에서 스피커 신호를 제거한다.

- [0049] 스피커부(120)는 스피커를 포함하며, 소리를 출력하기 위한 신호를 생성하는 신호 생성부(생략)로부터 수신되는 스피커 신호를 송출하는 기능을 한다.
- [0050] 제2 스펙트럼 추정부(125)는 스피커부(120)로부터 입력된 스피커 신호를 LPC 스펙트럼으로 변환하여 필터부(115)로 제공한다.
- [0051] 게인 추정부(140)는 필터부(115)로부터 추정된 스피커 신호를 수신하여 매그니튜드(Magnitude) 형식으로 변환한다.
- [0052] 또한, 게인 추정부(140)는 푸리에 변환부(135)로부터 스피커 신호의 스펙트럼을 수신한다. 이때, 게인 추정부(140)는 매그니튜드(Magnitude) 형식으로 스피커 신호의 스펙트럼을 수신한다.
- [0053] 그리고, 게인 추정부(140)는 추정된 스피커 신호의 스펙트럼과 푸리에 변환부(135)로부터 수신한 스피커 신호에 대한 스펙트럼 사이의 게인($\tilde{g}_*$)을 하기의 수학적 식 12에 적용하여 추정한 후 곱셈기(145)로 전송한다.
- [0054] 곱셈기(145)는 추정된 게인에 스피커부(120)에 의해 출력되는 스피커 신호를 곱한 스피커 추정 신호를 생성하여 음성 추출부(150)로 전송한다.
- [0055] 음성 추출부(150)는 곱셈기(145)로부터 수신한 스피커 추정 신호에 기 저장되어 있는 시간 딜레이(delay T_A)를 이용하여 시간축에서의 스피커 추정 신호를 생성하고, 마이크부(105)로부터 수신한 노이지 신호 $y(n)$ 에 시간 딜레이가 적용된 스피커 추정 신호 $\tilde{g}_* d(n-T)$ 를 제거하여 노이지 신호를 추출한다. 여기서, 시간 딜레이는 스피커부(120)로부터의 스피커 신호 송출 시간과 마이크부(105)로부터의 스피커 신호 입력 시간에 대한 차이에 대응하는 시간이다.
- [0056] 한편, 본 발명의 실시 예에 따른 음성 향상 장치는 추가적으로 미리 저장하고 있는 기기 내의 마이크와 스피커 모델에 대한 잡음 정보를 기초로 마이크와 스피커에 대한 잡음을 삽입하기 위한 잡음 삽입부를 구성할 수 있다. 이 경우에, 잡음 삽입부는 원 신호에 마이크와 스피커에 대한 잡음을 콘볼루션하여 제2 스펙트럼 추정부와 푸리에 변환부로 입력시킨다. 이를 통해, 보다 정밀하게 사용자의 음성을 추출할 수 있다.
- [0057] 이러한, 본 발명에 따르면, 종래의 장치들에서 이용하였던 기존의 양자화 과정, RWM(Remote Wireless Microphone)과 프라이머리 마이크로폰(Primary microphone)등을 이용하지 않고도, 스마트폰, 스마트 TV(Television) 등과 같은 스마트기기에서 사용자의 음성을 효율적으로 추출하여 제공할 수 있는 큰 장점이 있다.
- [0058] 다음은 도 2를 통해 음성 향상 장치의 음성 향상 방법에 대하여 상세히 설명한다.
- [0059] 도 2는 본 발명의 실시 예에 따른 음성 향상 장치의 음성 향상 방법을 도시한 순서도이다.
- [0060] 도 2에 나타난 바와 같이, 본 발명의 실시 예에 따른 음성 향상 장치(100)의 마이크부(105)는 스피커 신호가 포함된 노이지 신호를 수신한다(S200).
- [0061] 먼저, 음성 향상 장치(100)의 마이크부(105)가 수신하는 신호에서 배경잡음에 대한 부분을 제외하면, 수신하는 노이지 신호는 다음의 수학적 식 1과 같이 나타낼 수 있다.

수학적 식 1

$$y(n) = g_x x(n) + g_d d(n-T)$$

- [0063] T는 delay값으로, 스마트 기기의 마이크부(105)와 스피커부(120)는 움직이지 않고 항상 같은 거리를 유지하며 고정적으로 위치해 있기 때문에, 신호의 전 구간에서 대하여 항상 일정한 값을 갖는다.
- [0064] 여기서, $g_x x(n)$ 은 노이지 신호이고, $g_d d(n-T)$ 는 스피커가 송출한 신호인 스피커 신호이다. 딜레이 T를 추정하는 알고리즘이 있는 스마트 기기의 경우, 시간 도메인에서 스피커 신호를 제거할 수 있다. 이때, 시간도메인에서 스피커 신호를 제거하는 방법은 다음과 같이 이루어진다.

[0065] 제1 스펙트럼 추정부(110)는 기 설정된 LPC 추정 방식을 통해 노이지 신호의 LPC 스펙트럼을 산출한다(S202).

[0066] 구체적으로, 제1 스펙트럼 추정부(110)는 수학적 식 1과 같은 형식의 노이지 신호가 마이크부(105)를 통해 입력으로 들어오면, 수신된 노이지 신호의 LPC를 도출하고, 도출한 LPC를 이용하여 LPC 스펙트럼을 산출한다. 노이지 신호의 LPC 스펙트럼은 다음의 수학적 식 2와 같이 나타낼 수 있다.

수학적 식 2

$$[0067] \quad |Y(k)|^2 = g_x |X(k)|^2 + g_d |D(k)|^2$$

[0068] 여기서, k 값은 $0 \leq k \leq K-1$ 의 범위를 갖으며, $Y(k)$ 는 $Y(n)$ 의 노이지 신호의 스펙트럼이다.

[0069] 필터부(115)는 제1 스펙트럼 추정부(110)에 의해 추정된 노이지 신호를 이용하여 코드북 저장부(130)로부터 복수의 음성 코드북들 중 적어도 하나의 음성 코드북을 검출하고(S204), 검출된 음성 코드북에 대한 LPC 스펙트럼을 산출한다(S206)

[0070] 그리고, 필터부(115)는 제2 스펙트럼 추정부(125)에 의해 계산된 스피커부(120)의 스피커 신호로부터 입력된 스피커 신호의 LPC 스펙트럼과 검출된 음성 코드북의 LPC 스펙트럼 각각에 대한 계인을 추정한다(S208). 이때, 음성 코드북과 스피커 신호의 정보에 대한 LPC 스펙트럼은 각각 $C_x^i(k)$ 와 $C_d(k)$ 로 나타낸다.

[0071] 여기서, 검출된 음성 코드북의 코드 벡터는 다음의 수학적 식 3과 같이 계산할 수 있다.

수학적 식 3

$$[0072] \quad C_x^i(k) = \frac{1}{\left| \sum_{m=0}^P a_{x_m}^i e^{-jkm} \right|^2}$$

[0073] 여기서, i 는 음성 코드 벡터의 인덱스이며, P 는 LPC 차수, a 는 LPC 계수이다.

[0074] 그리고, 필터부(115)는 추정된 계인을 이용하여 노이지 신호를 추정한다(S210). 이때, 음성 코드북의 LPC 스펙트럼과 스피커 신호의 LPC스펙트럼의 계인을 g_x^i, g_d^i 라고 한다면, 추정된 노이지 신호는 다음의 수학적 식 4와 같이 나타낼 수 있다.

수학적 식 4

$$[0075] \quad |\widehat{Y}^i(k)|^2 = g_x^i C_x^i(k) + g_d^i C_d(k)$$

[0076] 여기서, g_x^i, g_d^i 는 각각 상기 음성 코드북의 코드 벡터에 대한 노이지 LPC 스펙트럼, 스피커 신호의 LPC 스펙트럼에 대한 계인값이고, C_x, C_d 는 노이지 LPC 스펙트럼, 스피커 신호의 LPC 스펙트럼이다.

[0077] 이때, 각각의 계인은 다음의 수학적 식 5를 이용하여 구할 수 있다.

수학적 식 5

$$[0078] \quad \begin{bmatrix} g_x^i \\ g_d^i \end{bmatrix} = A^{-1} B$$

이때, 상기의 수학적 식 5의 행렬 A와 B는 다음의 수학적 식 6과 같이 나타낼 수 있다. 이러한, 행렬식을 이용하여 음성 코드북과 스피커 신호의 LPC 스펙트럼에 대한 계인을 추정할 수 있다.

수학적 식 6

$$A = \begin{bmatrix} \| (C_x^i(k))^2 / |Y(k)|^4 \| & \| (C_x^i(k)) C_d(k) / |Y(k)|^4 \| \\ \| (C_x^i(k)) C_d(k) / |Y(k)|^4 \| & \| (C_d(k))^2 / |Y(k)|^4 \| \end{bmatrix}$$

여기서, 수학적 식 6에서 사용된 $\|f(k)\| = \int |f(k)| dk$ 이다.

또한, 필터부(115)는 추정된 노이즈 스펙트럼 $|\hat{Y}^i(k)|$ 과 마이크부(105)로부터 입력된 노이즈 스펙트럼 $|Y(k)|$ 에 대하여 MMSE 추정 방법을 통해 스피커 신호에 대한 LPC스펙트럼을 추정한다(S212).

스피커 신호를 추출하기 위한 필터 파라미터 $\hat{\lambda}$ 을 $\hat{\lambda} = \{\hat{C}_x, \hat{g}_x, \hat{g}_d\}$ 라고 하면, 해당하는 필터 파라미터에 대한 MMSE 추정 방법은 다음의 수학적 식 7과 같이 나타낼 수 있다. 여기서, 필터 파라미터는 각각 노이즈 신호의 추정된 노이즈 신호, 추정된 노이즈 신호의 계인값, 추출된 스피커 신호의 계인값을 포함한다.

수학적 식 7

$$\hat{\lambda} = \frac{1}{N_x} \sum_{i=1}^{N_x} \lambda^i \frac{p(y|\hat{Y}^i)}{p(y)}$$

여기서, N_x 는 코드 벡터의 개수이며, $P(y)$ 와 $p(y|\hat{Y}^i)$ 는 각각 다음의 수학적 식 8과 수학적 식 9를 이용하여 계산된다.

수학적 식 8

$$p(y) = \frac{1}{N_x} \sum_{i=1}^{N_x} p(y|\hat{Y}^i)$$

수학적 식 9

$$p(y|\hat{Y}^i) = \left\| \left(\frac{|Y(k)|^2}{|\hat{Y}^i(k)|^2} \right) - \ln \left(\frac{|Y(k)|^2}{|\hat{Y}^i(k)|^2} \right) - 1 \right\|$$

이렇게, 필터부(115)는 도출된 필터 파라미터 $\hat{\lambda} = \{\hat{C}_x, \hat{g}_x, \hat{g}_d\}$ 를 이용하여, 입력되는 노이즈 신호에서 스피커 신호만을 추출할 수 있도록 필터부(115)에 포함된 위너 필터를 다음의 수학적 식 11과 같이 구성한다(S214).

필터부(115)는 노이즈 신호에서 스피커 신호만을 추출할 수 있도록, 수학적 식 10과 같은 스피커 신호를 검출하기 위한 위너 필터를 이용하여 마이크부(105)로부터 입력되는 노이즈 신호로부터 스피커 신호를 추정하여 계인 추정부(140)로 제공한다(S216).

수학식 10

$$|\tilde{D}(k)|^2 = |Y(k)|^2 H(k)$$

이때, H(K)는 다음의 수학식 11에 의해 계산된다.

수학식 11

$$H(k) = \frac{\hat{g}_d C_d(k)}{\hat{g}_x \hat{C}_x(k) + \hat{g}_d C_d(k)}$$

게인 추정부(140)는 필터부(115)에 의해 추정된 스피커 신호의 스펙트럼을 매그니튜드(Magnitude) 형식으로 변환하고, 푸리에 변환부(135)로부터 스피커부(120)에 의해 출력된 스피커 신호에 대한 매그니튜드(Magnitude)를 수신한다. 이때, 푸리에 변환부(135)로부터 수신된 스피커 신호에 대한 스펙트럼을 $|D(k)|$ 로 표현할 수 있다.

그리고, 게인 추정부(140)는 추정된 스피커 신호의 스펙트럼과 푸리에 변환된 스피커 신호의 스펙트럼 간의 게인($\tilde{g}^*$)을 다음의 수학식 12에 적용하여 추정한 후 곱셈기(145)로 전송한다.

수학식 12

$$\tilde{g}^* = \underset{\tilde{g}}{\operatorname{argmin}} \sum_{k=1}^{K-1} (|\tilde{D}(k)| - \tilde{g}|D(k)|)^2$$

곱셈기(145)는 추정된 게인에 스피커부(120)로부터 입력되는 스피커 신호를 곱한 스피커 추정 신호를 생성하여 음성 추출부(150)로 전송한다.

음성 추출부(150)는 곱셈기(145)로부터 수신한 스피커 추정 신호에 기 저장되어 있는 시간 딜레이(delay T_A)를 이용하여 시간축에서의 스피커 신호를 추정하고, 마이크부(105)로부터 수신한 노이지 신호 $y(n)$ 에 시간 딜레이가 적용된 스피커 신호 $\tilde{g}^* d(n-T)$ 를 제거하여 노이지 신호를 추출한다(S218).

한편, 스마트기기 내에 시간 딜레이 T를 추정하는 알고리즘이 없는 경우에는 주파수 도메인에서 스피커 신호를 제거할 수도 있다.

시간 도메인에서 스피커 신호를 제거하는 방법과 같은 방법으로 노이지 신호의 LPC파워스펙트럼과 스마트 기기 안에 미리 저장된 스피커 신호에 대한 LPC파워스펙트럼을 이용하여 필터부(115)의 위너 필터를 구성하고, 위너 필터를 이용하여 주파수 도메인에서 스피커 신호를 제거할 수도 있다.

필터부(115)는 시간 도메인 상에서 위너 필터를 이용하여 스피커 신호의 스펙트럼을 추정하지만, 주파수 도메인에서 음성을 추정하기 위해 수학식 13의 분자에 스피커 신호의 스펙트럼 대신, 음성 코드북의 스펙트럼으로 바꾸어 위너 필터를 구성할 수도 있다.

그리고, 하기의 수학식 14과같이, 음성 추출부(150)는 노이지 신호의 스펙트럼의 곱으로 주파수 도메인에서 스피커 신호를 제거하여 노이지 신호 $X(k)$ 를 추출한다.

수학식 13

$$H(k) = \frac{\hat{g}_x \hat{C}_x(k)}{\hat{g}_x \hat{C}_x(k) + \hat{g}_d C_d(k)}$$

수학식 14

$$|\tilde{X}(k)|^2 = |Y(k)|^2 H(k)$$

여기서, $|Y(k)|$ 는 입력되는 스피커가 포함된 노이지 신호이고, $X(k)$ 는 노이지 신호이다.

이처럼 스마트 기기 안에 딜레이를 추정하는 알고리즘이 없는 경우에는, 주파수 도메인에서 스피커 신호를 제거하여, 수신되는 신호에서 원 음성을 정확도 높게 추출할 수도 있다.

이하에서는 도 3a 및 도 3b를 통해 본 발명의 실시예에 따른 코드북 기반의 음성 향상 장치(100)와 종래의 장치간의 로그 스펙트럴 왜곡(Log Spectral Distortion, LSD) 성능을 설명한다.

도 3a 및 3b는 본 발명의 실시예에 따른 음성 향상 장치와 종래 장치간의 시간 도메인 상 로그 스펙트럴 왜곡 성능을 비교한 표이다.

도 3a 및 도 3b에 나타난 바와 같이, 본 발명의 실시예에 따른 음성 향상 장치(100)는 시간 도메인과 주파수 도메인상에서, 종래 ML 방식만 이용한 음성 향상 장치보다 스펙트럴 왜곡(LSD; Log Spectral Distortion) 성능을 크게 낮춘 것을 알 수 있다.

이를 통해, 본 발명의 음성 향상 장치(100)는 스마트 TV(Television)나 스마트 폰 등의 스마트 기기를 통해 프로그램을 시청하는 중에 마이크로 입력되는 신호에서의 노이지 신호 추출율을 종래의 장치보다 크게 향상시킨 것을 알 수가 있다.

이러한, 본 발명의 실시예에 따른 음성 향상 장치는 종래 기기에서 이용하였던 양자화 처리 과정, 프라이머리 마이크와 RWM 등을 이용하지 않고도 종래기기보다 노이지 신호에서의 사용자의 음성인 노이지 신호에 대한 추출율을 상당히 향상시킬 수 있는 큰 장점이 있다.

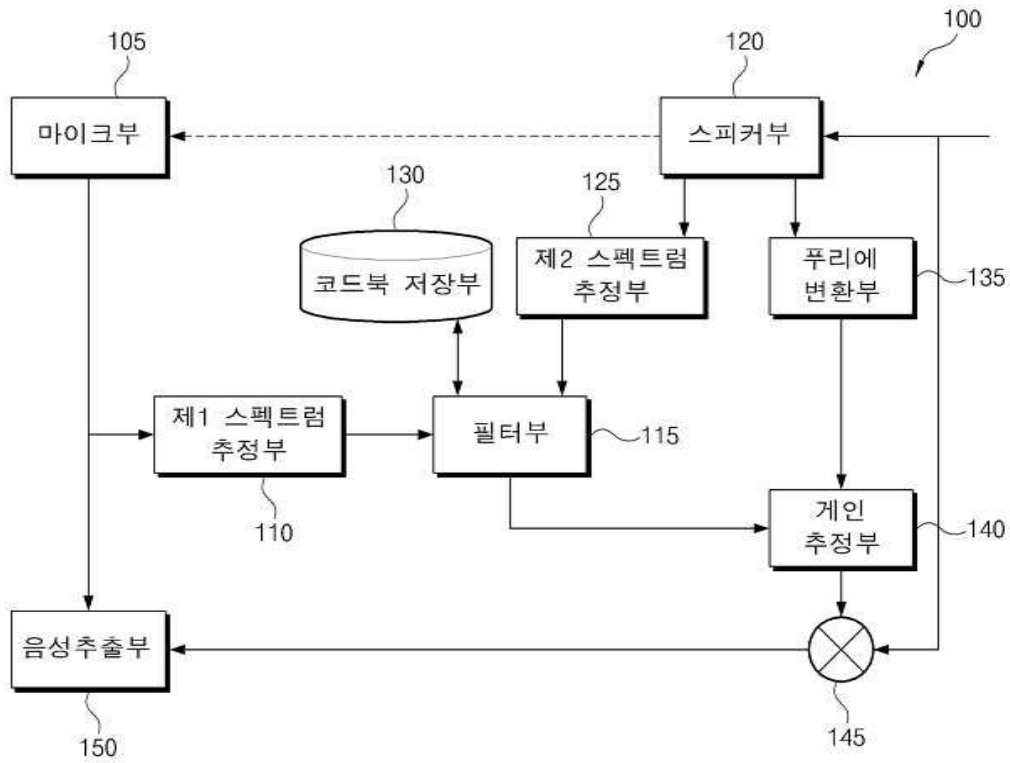
본 발명은 도면에 도시된 실시예를 참고로 설명되었으나 이는 예시적인 것에 불과하며, 본 기술 분야의 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 다른 실시예가 가능하다는 점을 이해할 것이다. 따라서, 본 발명의 진정한 기술적 보호 범위는 첨부된 특허청구범위의 기술적 사상에 의하여 정해져야 할 것이다.

부호의 설명

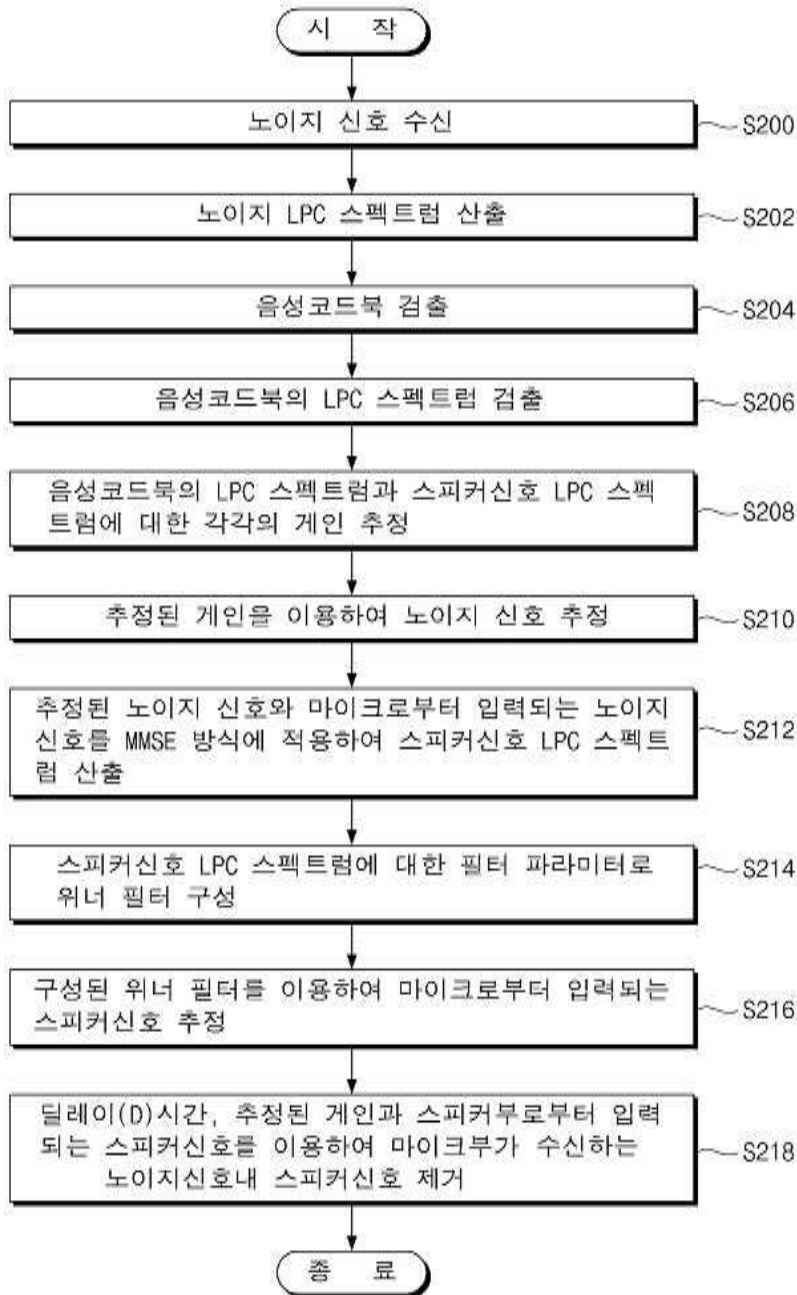
100: 음성 향상 장치	105: 마이크부
110: 제1 스펙트럼 추정부	115: 필터부
120: 스피커부	125: 제2 스펙트럼 추정부
130: 코드북 저장부	135: 푸리에 변환부
140: 게인 추정부	145: 곱셈기
150: 음성 추출부	

도면

도면1



도면2



도면3a

	종래 기기	본 발명
다큐멘터리	3.0052	1.2278
드라마	3.5561	1.2508
예능	3.9532	1.1670
뉴스	3.8162	1.1858
평균	3.5827	1.2079

<시간 도메인상 잡음 제거>

도면3b

	종래 기기	본 발명
다큐멘터리	3.0052	2.6398
드라마	3.5561	3.1719
예능	3.9532	3.4828
뉴스	3.8162	3.1166
평균	3.5827	3.1028

<주파수 도메인상 잡음 제거>