



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2021년01월29일

(11) 등록번호 10-2208855

(24) 등록일자 2021년01월22일

(51) 국제특허분류(Int. Cl.)
G10L 21/0232 (2013.01) G10L 25/18 (2013.01)

(52) CPC특허분류
G10L 21/0232 (2013.01)
G10L 25/18 (2013.01)

(21) 출원번호 10-2018-7013177

(22) 출원일자(국제) 2016년10월08일

심사청구일자 2019년12월27일

(85) 번역문제출일자 2018년05월09일

(65) 공개번호 10-2018-0067608

(43) 공개일자 2018년06월20일

(86) 국제출원번호 PCT/CN2016/101444

(87) 국제공개번호 WO 2017/063516

국제공개일자 2017년04월20일

(30) 우선권주장

201510670697.8 2015년10월13일 중국(CN)

(56) 선행기술조사문헌

JP평성08036400 A

EP2546831 A1

CN101968957 A

(73) 특허권자

어드밴스드 뉴 테크놀로지스 씨오., 엘티디.

케이만 군도, 그랜드 케이만 케이와이1-9008, 조지 타운, 27 하스피탈 로드, 케이만 코퍼레이트 센터

(72) 발명자

두 지준

중국 항저우 310099 완탕 로드 넘버 18 후양룽 타임즈 플라자 빌딩 비 17층 앤츠 패턴트 팀 내

(74) 대리인

김태홍, 김진희

전체 청구항 수 : 총 6 항

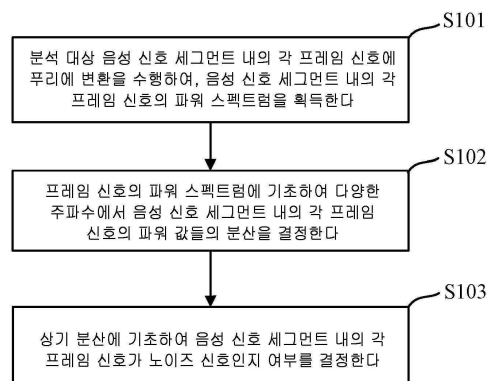
심사관 : 정성운

(54) 발명의 명칭 노이즈 신호 결정 방법과 장치, 및 음성 노이즈 제거 방법과 장치

(57) 요약

본 출원의 실시 예는 노이즈 신호 결정 방법과 장치, 및 음성 노이즈 제거 방법과 장치를 개시한다. 상기 노이즈 신호 결정 방법은 분석 대상 음성 신호 세그먼트의 각 프레임 신호에 푸리에 변환을 수행하여 상기 음성 신호 세그먼트 내의 각 프레임 신호의 파워 스펙트럼을 획득하는 단계; 상기 프레임 신호의 파워 스펙트럼에 기초하여 다양한 주파수에서 상기 음성 신호 세그먼트 내의 각 프레임 신호의 파워 값의 분산을 결정하는 단계; 및 상기 분산에 기초하여 상기 음성 신호 세그먼트 내의 각 프레임 신호가 노이즈 신호인지 여부를 결정하는 단계를 포함한다. 본 출원의 실시 예는 분석 대상 음성 신호 세그먼트 내에 포함된 여러 노이즈 프레임을 정확하게 얻을 수 있어 음성 노이즈 제거 효과가 향상된다.

대표도 - 도1



명세서

청구범위

청구항 1

음성 신호 세그먼트 내의 노이즈 신호를 결정하기 위한 방법에 있어서,

처리 대상(to-be-processed) 음성의 시간-영역 신호의 진폭 변화에 기초하여, 상기 처리 대상 음성 내에서 미리 설정된 문턱값보다 더 작은 진폭 변화를 갖는 음성 신호 세그먼트를 상기 음성 신호 세그먼트로서 결정하는 단계 - 상기 음성 신호 세그먼트는 복수의 프레임 신호를 포함함 -;

상기 음성 신호 세그먼트 내의 상기 복수의 프레임 신호 중 각 프레임 신호에 푸리에 변환을 수행하여 복수의 파워 스펙트럼을 획득하는 단계(S101) - 상기 복수의 파워 스펙트럼 중 각 파워 스펙트럼은 상기 음성 신호 세그먼트 내의 각 프레임 신호에 대응하고, 각 파워 스펙트럼은 상이한 주파수에 대응하는 다수의 파워 값을 포함함 -;

상기 프레임 신호의 파워 스펙트럼에 기초하여 다양한 주파수에서 상기 음성 신호 세그먼트 내의 각 프레임 신호의 파워 값의 분산을 결정하는 단계(S102); 및

상기 분산에 기초하여 상기 음성 신호 세그먼트 내의 각 프레임 신호가 노이즈 신호인지 여부를 결정하는 단계(S103)

를 포함하는, 음성 신호 세그먼트 내의 노이즈 신호를 결정하기 위한 방법.

청구항 2

제1항에 있어서, 상기 분산에 기초하여 상기 음성 신호 세그먼트 내의 각 프레임 신호가 노이즈 신호인지 여부를 결정하는 단계는:

상기 음성 신호 세그먼트 내의 각 프레임 신호에 대응하는 상기 분산이 제1 문턱값보다 더 큰지 여부를 결정하는 단계(S1031); 및

만일 그렇지 않다면, 상기 프레임 신호를 노이즈 신호로서 결정하는 단계(S1032)

를 포함하는 것인, 음성 신호 세그먼트 내의 노이즈 신호를 결정하기 위한 방법.

청구항 3

제2항에 있어서, 상기 파워 값의 분산을 결정하는 단계는:

다양한 주파수에서 상기 프레임 신호의 파워 값을 제1 주파수 구간(frequency interval)에 대응하는 제1 파워 값 세트로 분류하는 단계; 및

상기 제1 파워 값 세트 내에 포함된 파워 값의 제1 분산을 결정하는 단계(S1022)

를 포함하고,

상기 분산이 제1 문턱값보다 더 큰지 여부를 결정하는 단계는:

상기 제1 분산이 상기 제1 문턱값보다 더 큰지 여부를 결정하는 단계

를 포함하는 것인, 음성 신호 세그먼트 내의 노이즈 신호를 결정하기 위한 방법.

청구항 4

제1항에 있어서, 상기 파워 값의 분산을 결정하는 단계는:

다양한 주파수에서 각 프레임 신호의 파워 값을 제1 주파수 구간에 대응하는 제1 파워 값 세트 또는 제2 주파수 구간에 대응하는 제2 파워 값 세트로 적어도 분류하는 단계(S1021) - 상기 제1 주파수 구간의 최고 주파수는 상기 제2 주파수 구간의 최저 주파수보다 더 낮음 - ;

상기 제1 파워 값 세트 내에 포함된 파워 값의 제1 분산을 결정하는 단계(S1022); 및
상기 제2 파워 값 세트 내에 포함된 파워 값의 제2 분산을 결정하는 단계(S1023)
를 포함하고;

그 후, 상기 분산에 기초하여 상기 음성 신호 세그먼트 내의 각 프레임 신호가 노이즈 신호인지 여부를 결정하는 단계는:

각 프레임 신호에 대응하는 상기 제1 분산과 상기 제2 분산 사이의 차이가 제2 문턱값보다 더 큰지 여부를 결정하는 단계; 및

만일 그렇지 않다면, 상기 프레임 신호를 노이즈 신호로서 결정하는 단계

를 포함하는 것인, 음성 신호 세그먼트 내의 노이즈 신호를 결정하기 위한 방법.

청구항 5

제1항에 있어서, 상기 프레임 신호의 파워 스펙트럼에 기초하여 다양한 주파수에서 상기 음성 신호 세그먼트 내의 각 프레임 신호의 파워 값의 분산을 결정하는 단계 후에, 그리고 상기 분산에 기초하여 상기 음성 신호 세그먼트 내의 각 프레임 신호가 노이즈 신호인지 여부를 결정하는 단계 전에, 상기 방법은,

상기 분산의 크기에 따라 상기 음성 신호 세그먼트 내의 프레임 신호를 순위 매김(rank)하는 단계

를 더 포함하고,

그 후, 상기 분산에 기초하여 상기 음성 신호 세그먼트 내의 각 프레임 신호가 노이즈 신호인지 여부를 결정하는 단계는, 다양한 주파수에서 각 순위 매김된 프레임 신호의 파워 값의 분산에 기초하여 상기 음성 신호 세그먼트 내의 각 프레임 신호가 노이즈 신호인지 여부를 결정하는 단계를 포함하는 것인, 음성 신호 세그먼트 내의 노이즈 신호를 결정하기 위한 방법.

청구항 6

음성 신호 세그먼트 내의 노이즈 신호를 결정하기 위한 장치(100)에 있어서, 제1항 내지 제5항 중 어느 한 항의 방법을 수행하도록 구성된 다수의 유닛(101, 102, 103)을 포함하는, 음성 신호 세그먼트 내의 노이즈 신호를 결정하기 위한 장치(100).

청구항 7

삭제

청구항 8

삭제

청구항 9

삭제

청구항 10

삭제

청구항 11

삭제

청구항 12

삭제

청구항 13

삭제

청구항 14

삭제

청구항 15

삭제

청구항 16

삭제

청구항 17

삭제

청구항 18

삭제

발명의 설명

기술 분야

- [0001] 본 출원은 2015년 10월 13일에 출원되고 "노이즈 신호 결정 방법과 장치 및 음성 제거 방법과 장치"라는 명칭의 중국 특허 출원 번호 제201510670697.8호에 대한 우선권을 주장하며, 이는 전체가 본 명세서에 참고로 포함된다.
- [0002] 본 발명은 음성 노이즈 제거 기술 분야에 관한 것으로, 특히 노이즈 신호 결정 방법과 장치, 및 음성 노이즈 제거 방법과 장치에 관한 것이다.

배경 기술

- [0003] 음성 노이즈 제거 기술은 음성 신호로부터 환경 노이즈를 제거함으로써 음성 품질을 향상시킬 수 있다. 음성 신호에서의 노이즈 신호의 파워 스펙트럼은 먼저 음성 노이즈 제거 과정에서 결정될 필요가 있으며, 그 후 노이즈 신호의 결정된 파워 스펙트럼에 따라 음성 신호에서 노이즈가 제거될 수 있다.
- [0004] 종래 기술에서, 음성 신호 내의 노이즈 신호의 파워 스펙트럼은 일반적으로 다음과 같은 방식으로 결정될 수 있다: 음성 신호 내의 노이즈 신호의 파워 스펙트럼을 얻기 위해 처음 N개의 프레임 신호가 노이즈 신호라는 가정 하에 음성 신호 세그먼트 내의 처음 N개의 프레임 신호를 분석한다(즉, 인간의 음성 신호는 포함하지 않는다).
- [0005] 실제 응용 시나리오에서, 종래 기술에서 노이즈 신호로 가정되는 음성 신호 내의 처음 N개의 프레임 신호는 보통 실제 노이즈 신호와 일치하지 않으며, 따라서 얻어진 노이즈 신호 파워 스펙트럼의 정확도가 영향을 받는다.

발명의 내용

- [0006] 본 출원의 실시 예들의 목적은 노이즈 신호로 가정된 처음 N개의 프레임 신호가 실제 노이즈 신호와 일치하지 않기 때문에, 얻어진 노이즈 신호 파워 스펙트럼의 정확도가 영향을 받는다는 종래 기술의 문제점을 해결하기 위하여, 노이즈 신호 결정 방법과 장치, 및 음성 노이즈 제거 방법과 장치를 제공하는 것이다.
- [0007] 상기 기술적 문제점을 해결하기 위해, 본 출원의 실시 예들에 제공된 노이즈 신호 결정 방법과 장치, 및 음성 노이즈 제거 방법과 장치는 다음과 같이 구현된다:
- [0008] 노이즈 신호 결정 방법에 있어서,
- [0009] 분석 대상(to-be-analyzed) 음성 신호 세그먼트 내의 각 프레임 신호에 푸리에 변환을 수행하여 상기 음성 신호 세그먼트 내의 각 프레임 신호의 파워 스펙트럼을 획득하는 단계;
- [0010] 상기 프레임 신호의 파워 스펙트럼에 기초하여 다양한 주파수에서 상기 음성 신호 세그먼트 내의 각 프레임 신

호의 파워 값들의 분산을 결정하는 단계; 및

- [0011] 상기 분산에 기초하여 상기 음성 신호 세그먼트 내의 각 프레임 신호가 노이즈 신호인지 여부를 결정하는 단계를 포함하는 노이즈 신호 결정 방법.
- [0012] 음성 노이즈 제거 방법에 있어서,
- [0013] 처리 대상 음성 내에 포함된 분석 대상 음성 신호 세그먼트를 결정하는 단계;
- [0014] 상기 분석 대상 음성 신호 세그먼트 내의 각 프레임 신호에 푸리에 변환을 수행하여 상기 음성 신호 세그먼트 내의 각 프레임 신호의 파워 스펙트럼을 획득하는 단계;
- [0015] 상기 프레임 신호의 파워 스펙트럼에 기초하여 다양한 주파수에서 상기 음성 신호 세그먼트 내의 각 프레임 신호의 파워 값들의 분산을 결정하는 단계;
- [0016] 상기 음성 신호 세그먼트 내에 포함된 여러 노이즈 프레임을 얻기 위해 상기 분산에 기초하여 상기 음성 신호 세그먼트 내의 각 프레임 신호가 노이즈 신호인지 여부를 결정하는 단계; 및
- [0017] 상기 음성 신호 세그먼트 내에 포함된 여러 노이즈 프레임에 대응하는 평균 파워를 결정하고, 상기 노이즈 프레임의 평균 파워에 기초하여 상기 처리 대상 음성에서 노이즈를 제거하는 단계를 포함하는 것인 음성 노이즈 제거 방법.
- [0018] 노이즈 신호 결정 장치에 있어서,
- [0019] 분석 대상 음성 신호 세그먼트 내의 각 프레임 신호에 푸리에 변환을 수행하여 상기 음성 신호 세그먼트 내의 각 프레임 신호의 파워 스펙트럼을 획득하도록 구성된 파워 스펙트럼 획득 유닛;
- [0020] 상기 프레임 신호의 파워 스펙트럼에 기초하여 다양한 주파수에서 상기 음성 신호 세그먼트 내의 각 프레임 신호의 파워 값들의 분산을 결정하도록 구성된 분산 결정 유닛; 및
- [0021] 상기 분산에 기초하여 상기 음성 신호 세그먼트 내의 각 프레임 신호가 노이즈 신호인지 여부를 결정하도록 구성된 노이즈 결정 유닛을 포함하는 것인 노이즈 신호 결정 장치.
- [0022] 음성 노이즈 제거 장치에 있어서,
- [0023] 처리 대상 음성 내에 포함된 분석 대상 음성 신호 세그먼트를 결정하도록 구성된 세그먼트 결정 유닛;
- [0024] 상기 분석 대상 음성 신호 세그먼트 내의 각 프레임 신호에 푸리에 변환을 수행하여 상기 음성 신호 세그먼트 내의 각 프레임 신호의 파워 스펙트럼을 획득하도록 구성된 파워 스펙트럼 획득 유닛;
- [0025] 상기 프레임 신호의 파워 스펙트럼에 기초하여 다양한 주파수에서 상기 음성 신호 세그먼트 내의 각 프레임 신호의 파워 값들의 분산을 결정하도록 구성된 분산 결정 유닛;
- [0026] 상기 분산에 기초하여 상기 음성 신호 세그먼트 내의 각 프레임 신호가 노이즈 신호인지 여부를 결정하고, 상기 음성 신호 세그먼트 내에 포함된 여러 노이즈 프레임을 얻도록 구성된 노이즈 결정 유닛; 및
- [0027] 상기 음성 신호 세그먼트 내에 포함된 여러 노이즈 프레임에 대응하는 평균 파워를 결정하고, 상기 노이즈 프레임의 평균 파워에 기초하여 상기 처리 대상 음성에서 노이즈를 제거하도록 구성된 음성 노이즈 제거 유닛을 포함하는 것인 음성 노이즈 제거 장치.
- [0028] 본 출원의 실시 예들에서 제공된 기술적 솔루션으로부터 알 수 있는 바와 같이, 분석 대상 음성 신호 세그먼트에 대해 푸리에 변환을 수행하여 각 프레임 신호의 파워 스펙트럼을 획득하고, 다양한 주파수에서 상기 분석 대상 음성 신호 세그먼트 내의 각 프레임 신호의 파워 값들의 분산을 결정하고, 최종적으로 상기 분산에 기초하여 상기 프레임 신호가 노이즈 신호인지 여부를 결정함으로써, 본 출원의 실시 예들에서 제공된 음성 노이즈 제거 방법 및 장치 뿐만 아니라 노이즈 신호 결정 방법 및 장치가 상기 분석 대상 음성 신호 세그먼트 내에 포함된 여러 노이즈 프레임을 정확히 얻을 수 있다. 음성 노이즈 제거 프로세스에서 결정된 노이즈 프레임의 평균 파워에 기초하여 처리 대상 음성에서 노이즈가 제거될 수 있고, 따라서 음성 노이즈 제거 효과가 향상된다.

도면의 간단한 설명

- [0029] 본 출원 또는 종래 기술의 실시 예들의 기술적 솔루션을 보다 명확하게 설명하기 위해, 다음은 실시 예들 또는

종래 기술을 설명하기 위해 사용된 첨부 도면을 간단히 소개한다. 명백하게, 아래에서 설명되는 첨부된 도면은 본 출원에 언급된 단지 일부 실시 예에 불과하며, 당업자는 창조적인 노력 없이도 이들 도면으로부터 다른 도면을 여전히 유도할 수 있다.

도 1은 본 출원의 실시 예에 따른 노이즈 신호 결정 방법의 흐름도이다.

도 2는 본 출원의 실시 예에 따라 프레임 신호가 노이즈 신호인지 여부를 결정하는 단계들의 흐름도이다.

도 3은 본 출원의 실시 예에 따라 다양한 샘플링 포인트에서 프레임 신호의 파워 값들의 분산을 결정하는 단계들의 흐름도이다.

도 4는 본 출원의 실시 예에 따른 파워 값들의 분산의 곡선 그래프이다.

도 5는 본 출원의 실시 예에 따른 음성 노이즈 제거 방법의 흐름도이다.

도 6은 본 출원의 실시 예에 따른 노이즈 신호 결정 장치의 블록도이다.

도 7은 본 출원의 실시 예에 따른 음성 노이즈 제거 장치의 블록도이다.

도 8은 본 출원에 제공된 장치의 하드웨어 구현 예의 개략적인 구조도이다.

발명을 실시하기 위한 구체적인 내용

[0030] 당업자가 본 출원의 기술 솔루션을 더 잘 이해하게 하기 위해, 본 출원의 실시 예의 기술적 솔루션은 본 출원의 실시 예에서 첨부된 도면을 참조하여 아래에서 명확하고 완전하게 설명될 것이다. 기술된 실시 예들은 단지 본 출원의 실시 예 전부가 아닌 일부의 것임이 명백하다. 본 출원의 실시 예에 기초하여, 당업자는 창조적 노력 없이 다른 실시 예를 얻을 수 있으며, 이들은 모두 본 출원의 보호 범위 내에 있다.

[0031] 도 1은 본 출원의 실시 예에 따른 노이즈 신호 결정 방법의 흐름도이다. 분석 대상 음성 신호 세그먼트 내의 노이즈 신호를 결정하기 위해서, 본 실시 예의 노이즈 신호 결정 방법은 다음 단계들을 포함한다:

[0032] S101: 분석 대상 음성 신호 세그먼트 내의 각 프레임 신호에 대하여 푸리에 변환을 수행하여, 음성 신호 세그먼트 내의 각 프레임 신호의 파워 스펙트럼을 획득한다.

[0033] 분석 대상 음성 신호 세그먼트는 특정 규칙에 따라 처리 대상 음성으로부터 캡처될 수 있다. 분석 대상 음성 신호 세그먼트는 예비 결정에 기초하여 많은 노이즈 프레임을 포함할 수 있는 "의심되는 노이즈 프레임 세그먼트"일 수 있다. 바람직하게는, 단계 S101 전에, 상기 방법은,

[0034] 상기 처리 대상 음성의 시간-영역 신호의 진폭 변화에 기초하여, 상기 처리 대상 음성에서 미리 설정된 임계치보다 작은 진폭 변화를 갖는 음성 신호 세그먼트를 상기 분석 대상 음성 신호 세그먼트로서 결정하는 단계; 또는

[0035] 상기 처리 대상 음성에서 처음 N개의 프레임 음성 신호를 상기 분석 대상 음성 신호 세그먼트로서 캡처하는 단계를 더 포함한다.

[0036] 본 출원의 실시 예에서, 음성 신호의 시간 영역에서, 노이즈 신호는 일반적으로 진폭 변화가 작은 음성 신호 세그먼트 또는 일정한 진폭을 갖는 음성 신호 세그먼트인 반면, 인간 스피치 음성을 포함하는 음성 신호 세그먼트는 일반적으로 진폭 변화에 있어 크게 변동한다. 이러한 규칙에 기초하여, 처리 대상 음성(즉, 노이즈 제거 대상(to-be-denoised) 음성) 내에 포함된 "의심되는 노이즈 프레임 세그먼트"를 인식하기 위해 사용되는 미리 설정된 임계치가 미리 설정될 수 있다. 따라서, 처리 대상 음성 내의 미리 설정된 임계치보다 작은 진폭 변화를 갖는 음성 신호 세그먼트는 분석 대상 음성 신호 세그먼트로서 결정될 수 있다.

[0037] 본 출원의 실시 예에서, 프레임링(framing)은 음성 신호에 대해 먼저 수행될 수 있다. 프레임 신호는 단일 프레임 음성 신호를 지칭하며, 하나의 음성 신호 세그먼트는 여러 프레임 신호를 포함할 수 있다. 하나의 프레임 신호는 몇몇 샘플링 포인트, 예를 들어 1024 샘플링 포인트를 포함할 수 있다. 인접한 두 개의 프레임 신호는 서로 중첩될 수 있다(예를 들어, 중첩 비율은 50%일 수 있다). 본 실시 예에서는 시간 영역에서 음성 신호에 대해 단시간 푸리에 변환(short-time Fourier transform, STFT)을 수행하여 음성 신호의 파워 스펙트럼(주파수 영역)을 획득할 수 있다. 파워 스펙트럼은 상이한 주파수, 예를 들어 1024개의 파워 값에 대응하는 다수의 파워 값을 포함할 수 있다.

[0038] 본 출원의 실시 예에서, 일반적으로, 사람이 말하기 전의 시간(예를 들어, 1.5초)의 음성 신호는, 인간 음성을

포함하는 음성 신호 세그먼트 내의 노이즈 신호(환경 노이즈)이다. 따라서, 본 출원의 실시 예에서는 분석 대상 음성 신호가 음성 신호 세그먼트 내의 처음 N개의 프레임 신호라고 결정될 수 있다. 예를 들어 분석 대상 음성 신호는 처음 1.5초 내의 음성 신호: $\{f_1', f_2', \dots, f_n'\}$ 이며, f_1', f_2', \dots, f_n' 은 각각 음성 신호 내에 포함되는 프레임 신호를 나타낸다. 본 출원의 실시 예는 분석 대상 음성 신호에서 프레임 신호로부터 노이즈 신호를 결정하는 것을 목표로 한다.

- [0039] STFT 이후에 얻어진 분석 대상 음성 신호: $\{f_1', f_2', \dots, f_n'\}$ 의 파워 스펙트럼에 기초하여 각 프레임 신호에 대응하는 다수의 파워 값이 계산될 수 있다. 주파수에서의 프레임 신호의 파워 스펙트럼이 $a + bi$ 이고, 실수부 a 는 진폭을 나타낼 수 있고, 허수부 b 는 위상을 나타낼 수 있다고 가정한다. 그러면 주파수에서 프레임 신호의 파워 값은 $a^2 + b^2$ 일 수 있다. 상기 처리에 기초하여 상이한 주파수에서 각 프레임 신호의 파워 값을 얻을 수 있다. 예를 들어, 프레임 신호 $\{f_1', f_2', \dots, f_n'\}$ 각각이 1024개의 샘플링 포인트를 포함하면, 파워 스펙트럼에 기초하여 상이한 주파수에서 각 프레임 신호의 1024개의 파워 값을 얻을 수 있다. 예를 들어, 프레임 신호 f_1' 에 대응하는 파워 값들은 $\{p_{1,1}^1, p_{1,2}^1, \dots, p_{1,1024}^1\}$ 이고, 프레임 신호 f_2' 에 대응하는 파워 값들은 $\{p_{2,1}^2, p_{2,2}^2, \dots, p_{2,1024}^2\}$ 이고, 프레임 신호 f_n' 에 대응하는 파워 값들은 $\{p_{n,1}^n, p_{n,2}^n, \dots, p_{n,1024}^n\}$ 이다.
- [0040] S102: 다양한 주파수에서 음성 신호 세그먼트 내의 각 프레임 신호의 파워 값들의 분산은 프레임 신호의 파워 스펙트럼에 기초하여 결정된다.
- [0041] 다양한 주파수에서 프레임 신호 $\{f_1', f_2', \dots, f_n'\}$ 의 파워 값에 기초하여, 프레임 신호 $\{f_1', f_2', \dots, f_n'\}$ 의 파워 값의 분산 $\{\text{Var}(f_1'), \text{Var}(f_2'), \dots, \text{Var}(f_n')\}$ 은 분산 계산 공식에 따라 계산될 수 있다. 예를 들어, 각 프레임 신호가 1024개의 샘플링 포인트를 포함하면, $\text{Var}(f_1')$ 은 $\{p_{1,1}^1, p_{1,2}^1, \dots, p_{1,1024}^1\}$ 의 분산이고, $\text{Var}(f_2')$ 는 $\{p_{2,1}^2, p_{2,2}^2, \dots, p_{2,1024}^2\}$ 의 분산이고, $\text{Var}(f_n')$ 는 $\{p_{n,1}^n, p_{n,2}^n, \dots, p_{n,1024}^n\}$ 의 분산이다.
- [0042] S103: 상기 분산에 기초하여 음성 신호 세그먼트 내의 각 프레임 신호가 노이즈 신호인지 여부를 결정한다.
- [0043] 본 출원의 실시 예에서, 스피치 세그먼트를 포함하는 프레임 신호의 에너지(즉, 파워 값)는 일반적으로 대역에 따라 크게 변화하는 반면, 스피치 세그먼트가 없는 프레임 신호(즉, 노이즈 신호)의 에너지는 대역에 따라 약간 변화하고, 균등하게 분포된다. 따라서, 프레임 신호의 파워 값들의 분산에 기초하여 각 프레임 신호가 노이즈 신호인지 여부를 결정할 수 있다.
- [0044] 도 2는 본 출원의 실시 예에 따라 프레임 신호가 노이즈 신호인지 여부를 결정하는 단계들의 흐름도를 도시한다. 본 출원의 실시 예에서, 상기 단계 S103는 다음 단계들을 포함할 수 있다:
- [0045] S1031: 프레임 신호의 파워 값들의 분산이 제1 임계치 T_1 보다 큰지 여부를 결정한다.
- [0046] S1032: "아니오"인 경우, 프레임 신호는 노이즈 신호로서 결정된다.
- [0047] 만일 프레임 신호의 파워 값들의 분산이 제1 임계치 T_1 을 초과하면, 대역을 갖는 프레임 신호의 에너지의 변화 진폭(즉, 파워 값)이 제1 임계치 T_1 을 초과한다고 표시된다. 따라서, 프레임 신호는 노이즈 신호가 아니라고 결정될 수 있다. 반대로, 프레임 신호의 파워 값들의 분산이 제1 임계치 T_1 을 초과하지 않으면, 대역을 갖는 프레임 신호의 에너지의 변화 진폭(즉, 파워 값)이 제1 임계치 T_1 을 초과하지 않는다고 표시된다. 따라서, 프레임 신호는 노이즈 신호라고 결정될 수 있다.
- [0048] 위의 과정을 바탕으로, 노이즈 프레임 신호 $\{f_1', f_2', \dots, f_m'\}$ 와 비-노이즈 프레임 신호 $\{f_{m+1}', f_{m+2}', \dots, f_n'\}$ 는 분석 대상 음성 신호 $\{f_1', f_2', \dots, f_n'\}$ 에서 순차적으로 결정될 수 있다. 따라서, 음성 신호 세그먼트 내에 포함된 노이즈 신호가 결정될 수 있고, 이들 노이즈 신호 $\{f_1', f_2', \dots, f_m'\}$ 에 따라 음성 노이즈 제거가 수행될 수 있다.
- [0049] 도 3을 참조하면, 본 출원의 실시 예에서, 상기 단계 S102는 구체적으로 다음의 단계들을 포함할 수 있다:
- [0050] S1021: 다양한 주파수에서 프레임 신호 $\{f_1', f_2', \dots, f_n'\}$ 각각의 파워 값은 프레임 신호의 파워 스펙트럼에

대응하는 주파수가 속하는 주파수 구간에 따라, 제1 주파수 구간에 대응하는 제1 파워 값 세트와 제2 주파수 구간에 대응하는 제2 파워 값 세트로 적어도 분류되고, 제1 주파수 구간은 제2 주파수 구간보다 더 낮다.

[0051] 특정 실시 예에서, 각 프레임 신호의 분산은 통계를 통해 주파수 도메인에서 획득될 수 있다. 비-노이즈 신호는 일반적으로 저 중음(low-mid) 주파수 대역에 집중되는 반면, 노이즈 신호는 일반적으로 모든 주파수 대역에서 균일하게 분포된다. 따라서, 적어도 두 개의 상이한 주파수 대역(즉, 상기 주파수 구간)의 통계치를 통해 다양한 주파수에서 각 프레임 신호의 파워 값들의 분산을 얻을 수 있다.

[0052] 예를 들어, 제1 주파수 구간은 0 ~ 2000Hz(저주파수 대역)일 수 있고, 제2 주파수 구간은 2000 ~ 4000Hz(고주파수 대역)일 수 있다. 각 프레임 신호가 1024 개의 샘플링 포인트를 포함하는 경우, 각 프레임 신호에 대응하는 1024개의 파워 값은 파워 값에 대응하는 주파수 구간에 따라 0 ~ 2000Hz에 대응하는 제1 파워 값 세트 A와 2000 ~ 4000Hz에 대응하는 제2 파워 값 세트 B로 분류된다. 프레임 신호 f_1' 을 예로 들면, 1024개의 대응하는 파워 값은 $\{p_1^1, p_2^1, \dots, p_{1024}^1\}$ 이다. 주파수 구간에 따라, 제1 파워 값 세트 A에 포함되는 파워 값은 예를 들어 $\{p_1^1, p_2^1, \dots, p_{126}^1\}$ 이고, 제1 파워 값 세트 A에 포함된 파워 값은 예를 들어 $\{p_{127}^1, p_{128}^1, \dots, p_{1024}^1\}$ 이고, 나머지는 유추로 추론할 수 있다.

[0053] 본 출원의 다른 실시 예에서는 신호 파워 값의 분산이 2개가 넘는 주파수 대역에서의 통계를 통해 획득될 수 있다는 것을 알아야 한다.

[0054] S1022: 제1 파워 값 세트 내에 포함된 파워 값들의 제1 분산이 결정된다.

[0055] 전술한 바와 같이, 프레임 신호 f_1' 을 예로 들면, 제1 파워 값 세트 A 내에 포함되는 파워 값은, 예를 들면 $\{p_{127}^1, p_{128}^1, \dots, p_{1024}^1\}$ 이다. 따라서, 파워 값 $p_{127}^1 \sim p_{1024}^1$ 의 제1 분산 $\text{Var}_{\text{high}}(f_1')$ 는 분산 공식에 따라 계산될 수 있다.

[0056] S1021: 제2 파워 값 세트 내에 포함된 파워 값들의 제2 분산이 결정된다.

[0057] 전술한 바와 같이, 프레임 신호 f_1' 을 예로 들면, 제2 파워 값 세트 B 내에 포함되는 파워 값은 예를 들면 $\{p_1^1, p_2^1, \dots, p_{126}^1\}$ 이다. 따라서, 파워 값 $p_1^1 \sim p_{126}^1$ 의 제2 분산 $\text{Var}_{\text{low}}(f_1')$ 는 분산 공식에 따라 계산될 수 있다.

[0058] 도 4는 본 출원의 실시 예에 따른 분산의 개략적인 곡선 그래프를 도시한다. 그래프에서, 가로축은 프레임 신호의 프레임 번호를 나타내고, 세로축은 분산의 크기를 나타낸다. 제1 분산 곡선은 각 프레임 신호의 제1 분산의 경향을 나타내고, 제1 분산 곡선은 각 프레임 신호의 제2 분산의 경향을 나타낸다. 그래프에서 볼 수 있듯이 2000 ~ 4000Hz의 고주파수 대역에서 분산이 약간 변동하며, 0 ~ 2000Hz의 저주파수 대역에서 분산이 크게 변동한다. 이것은 비-노이즈 신호가 주로 저주파수 대역에 집중되어 있음을 증명할 수 있다.

[0059] 전술한 바와 같이, 본 출원의 선호되는 실시 예에서, 단계 S1031은 구체적으로:

[0060] 상기 프레임 신호의 파워 값들의 제1 분산이 제1 임계치 T_1 보다 큰지 여부를 결정하는 단계; 만약 예라면, 프레임 신호를 노이즈 신호로서 결정하는 단계를 포함한다. 프레임 신호 f_1' 을 예로 들면, 제1 분산 $\text{Var}_{\text{high}}(f_1')$ 가 제1 임계치 T_1 보다 큰지 여부가 결정된다.

[0061] 본 출원의 실시 예에서, 상기 단계 S103은 구체적으로:

[0062] 상기 제1 분산과 상기 제2 분산 사이의 차이가 제2 임계치 T_2 보다 큰지 여부를 결정하는 단계; 및

[0063] 아니오이면, 프레임 신호를 노이즈 신호로서 결정하는 단계를 포함한다.

[0064] 프레임 신호 f_1' 을 예로 들면, 제1 분산과 제2 분산의 차이는 $|\text{Var}_{\text{high}}(f_1') - \text{Var}_{\text{low}}(f_1')|$ 이다. 만일 $|\text{Var}_{\text{high}}(f_1') - \text{Var}_{\text{low}}(f_1')| < T_2$ 인 경우, 프레임 신호 f_1' 은 노이즈 신호로서 결정된다. 이 단계에 따라, 분석 대상 음성 프레임 신호 $\{f_1', f_2', \dots, f_n'\}$ 로부터 노이즈 신호가 순차적으로 결정될 수 있다.

[0065] 본 출원의 실시 예에서, 단계 S102와 단계 S103 사이에서 상기 방법은:

- [0066] 상기 분산의 크기에 따라 상기 분석 대상 음성 신호 세그먼트의 프레임 신호를 순위 매김(rank)하는 단계를 더 포함할 수 있다.
- [0067] 그 후, 상기 분산에 기초하여 상기 음성 신호 세그먼트 내의 각 프레임 신호가 노이즈 신호인지 여부를 결정하는 단계는:
- [0068] 다양한 주파수에서 각각의 순위가 매겨진 프레임 신호의 파워 값들의 분산에 기초하여 상기 음성 신호 세그먼트 내의 각 프레임 신호가 노이즈 신호인지 여부를 결정하는 단계를 포함한다.
- [0069] 전술한 바와 같이, 프레임 신호 $\{f_1', f_2', \dots, f_n'\}$ 의 파워 값들의 분산 $\{\text{Var}(f_1'), \text{Var}(f_2'), \dots, \text{Var}(f_n')\}$ 이 이 실시 예에서 결정될 수 있다. 프레임 신호는 파워 값의 분산의 오름차순으로 순위를 매길 수 있다. 더 작은 분산을 가진 신호는 노이즈 신호일 가능성이 크다. 따라서, 분석 대상 음성 신호 내의 노이즈 프레임 신호를 맨 앞쪽까지 순위 매김할 수 있다. 본 출원의 실시 예에서, 저주파수 대역(예를 들어, 0 ~ 2000Hz) 및 고주파수 대역(예를 들어, 2000 ~ 4000Hz)의 통계치를 통해 분산을 각각 획득하면, 다양한 주파수에서 프레임 신호 $\{f_1', f_2', \dots, f_n'\}$ 각각의 파워 값들이 프레임 신호의 파워 스펙트럼에 대응하는 주파수가 속하는 주파수 구간에 따라 제1 주파수 구간(예컨대, 0 ~ 2000Hz)에 대응하는 제1 파워 값 세트 A와 제2 주파수 구간(예컨대, 2000 ~ 4000Hz)에 대응하는 제2 파워 값 세트 B로 분류될 수 있다. 그 다음, 프레임 신호 $\{f_1', f_2', \dots, f_n'\}$ 에 대응하는 제1 파워 값 세트 내에 포함된 파워 값들의 제1 분산 $\{\text{Var}_{\text{low}}(f_1'), \text{Var}_{\text{low}}(f_2'), \dots, \text{Var}_{\text{low}}(f_n')\}$ 이 각각 결정될 수 있고, 프레임 신호 $\{f_1', f_2', \dots, f_n'\}$ 에 대응하는 제2 파워 값 세트 내에 포함된 파워 값들의 제2 분산 $\{\text{Var}_{\text{high}}(f_1'), \text{Var}_{\text{high}}(f_2'), \dots, \text{Var}_{\text{high}}(f_n')\}$ 이 각각 결정될 수 있다. 상기 단계 S104에서, 고주파수 및 저주파수에서의 분산 통계에 기초하여, 분석 대상 음성 신호 내에 포함된 노이즈 신호(분산의 크기에 따라 등급이 매겨진 음성 신호일 수 있음)는 다음과 같은 방식으로 결정될 수 있다:
- [0070] $\text{Var}_{\text{low}}(f_i') > T_1$ (1);
- [0071] $|\text{Var}_{\text{high}}(f_i') - \text{Var}_{\text{low}}(f_i')| > T_2$ (2);
- [0072] $\text{Var}_{\text{high}}(f'_{i+1}) - \text{Var}_{\text{low}}(f'_{i-1}) > T_3$ (3);
- [0073] $\text{Var}_{\text{high}}(f'_{i+1}) - \text{Var}_{\text{low}}(f'_{i-1}) > T_4$ (4);
- [0074] $i \in (1, n)$ 이다. 식 (1)에 기초하여 각 프레임 신호 f_i' 의 파워 값들의 제1 분산이 제1 임계치 T_1 보다 큰지 여부를 결정할 수 있다. 아니오이면, 프레임 신호 f_i' 는 노이즈 프레임 신호로서 결정된다. 결정된 노이즈 프레임 신호들의 세트는 노이즈 신호로서 결정된다.
- [0075] 식 (2)에 기초하여 각 프레임 신호 f_i' 의 파워 값들의 제2 분산이 제2 임계치 T_2 보다 큰지 여부를 결정할 수 있다. 아니오이면, 프레임 신호 f_i' 는 노이즈 프레임 신호로서 결정된다. 결정된 노이즈 프레임 신호들의 세트는 노이즈 신호로서 결정된다.
- [0076] 식 (3)에 기초하여 프레임 신호 f_i' 전의 프레임 신호 f'_{i-1} 의 파워 값들의 제2 분산 $\text{Var}_{\text{high}}(f'_{i-1})$ 및 프레임 신호 f_i' 다음의 프레임 신호 f'_{i+1} 의 파워 값들의 제2 분산 $\text{Var}_{\text{high}}(f'_{i+1})$ 의 차이 $\text{Var}_{\text{high}}(f'_{i+1}) - \text{Var}_{\text{high}}(f'_{i-1})$ 가 제3 임계치 T_3 보다 큰지 여부를 결정할 수 있다. 아니오이면, 프레임 신호 f_i' 는 노이즈 프레임 신호로서 결정된다. 결정된 노이즈 프레임 신호들의 세트는 노이즈 신호로서 결정된다.
- [0077] 식 (4)에 기초하여 프레임 신호 f_i' 전의 프레임 신호 f'_{i-1} 의 파워 값들의 제1 분산 $\text{Var}_{\text{low}}(f'_{i-1})$ 및 프레임 신호 f_i' 다음의 프레임 신호 f'_{i+1} 의 파워 값들의 제1 분산 $\text{Var}_{\text{low}}(f'_{i+1})$ 의 차이 $\text{Var}_{\text{low}}(f'_{i+1}) - \text{Var}_{\text{low}}(f'_{i-1})$ 가 제4 임계치 T_4 보다 큰지 여부를 결정할 수 있다. 아니오이면, 프레임 신호 f_i' 는 노이즈 프레임 신호로서 결정된다. 결정된 노이즈 프레임 신호들의 세트는 노이즈 신호로서 결정된다.
- [0078] 본 출원의 실시 예에서는, 상기 식 (1) ~ (4)를 이용하여, 분석 대상 음성 신호에 포함되는 노이즈 프레임을 인식할 수 있다. 즉, 상기 식 1 내지 4 중 어느 하나를 만족하는 임의의 프레임 신호 f_i' 는 비-노이즈 신호(노이즈 종료(noise end) 프레임)로서 결정될 수 있다. 다시 말해서, 상기 식 1 내지 4 중 아무 것도 만족하지 않는

임의의 프레임 신호 f_i' 는 노이즈 신호로서 결정될 수 있다. 상기 과정에 기초하여 노이즈 종료(noise end) 프레임 f_m' 을 결정할 수 있고, 그 후 노이즈 프레임은 $\{f_1', f_2', \dots, f_{m-1}'\}$ 을 포함한다.

- [0079] 본 출원의 다른 실시 예에서, 노이즈 종료 프레임은 식(1) 및 (2), 또는 식 (2) 및 (3)과 같이, 식 (1) 내지 (4)의 일부에 기초하여 결정될 수 있다. 또한, 본 출원의 실시 예에서 노이즈 종료 프레임을 결정하기 위한 공식은 상기 나열된 식에 제한되지 않는다. 임계치 T_1 , T_2 , T_3 및 T_4 는 다량의 테스트 샘플에 대한 통계로부터 모두 얻어진다.
- [0080] 도 5는 본 출원의 실시 예에 따른 음성 노이즈 제거 방법의 흐름도로서, 다음 단계들을 포함한다:
- [0081] S201: 처리 대상 음성 내에 포함되는 분석 대상 음성 신호 세그먼트를 결정한다.
- [0082] S202: 분석 대상 음성 신호 세그먼트 내의 각 프레임 신호에 푸리에 변환을 수행하여 음성 신호 세그먼트 내의 각 프레임 신호의 파워 스펙트럼을 획득한다.
- [0083] S203: 다양한 주파수에서 음성 신호 세그먼트 내의 각 프레임 신호의 파워 값의 분산은 프레임 신호의 파워 스펙트럼에 기초하여 결정된다.
- [0084] S204: 상기 분산에 기초하여 음성 신호 세그먼트 내의 각 프레임 신호가 노이즈 신호인지 여부를 결정하고, 음성 신호 세그먼트 내에 포함된 몇 개의 노이즈 프레임이 얻어진다.
- [0085] S205: 음성 신호 세그먼트 내에 포함된 여러 노이즈 프레임에 대응하는 평균 파워를 결정하고, 노이즈 프레임의 평균 파워에 기초하여 처리 대상 음성에서 노이즈를 제거한다.
- [0086] 본 출원의 실시 예에서는, 상기 방법에 따라 분석 대상 음성 세그먼트 내에 포함된 노이즈 프레임 $\{f_1', f_2', \dots, f_{m-1}'\}$ 을 획득한 후, 노이즈 프레임에 대응하는 원래 신호(순위 매김 전)의 프레임 번호를 각각 결정할 수 있고, 이들 프레임 신호의 평균 파워를 통계를 통해 구하여, 노이즈 신호의 파워 스펙트럼 추정 값 P_{noise} 를 얻을 수 있다. 노이즈 신호의 파워 스펙트럼 추정 값 P_{noise} 가 얻어진 후에 음성에서 노이즈를 제거할 수 있다. 상기 노이즈 제거 방법은 당업자에게 잘 알려져 있으므로 여기서 구체적으로 설명하지 않는다.
- [0087] 명확하게, 본 출원의 다른 실현 가능한 실시 예들에서, 분산들에 따라 프레임 신호들을 순위 매김하는 단계는 생략될 수 있고, 노이즈 프레임들은 원래 신호들의 분산에 기초하여 직접 결정될 수 있다. 또한, 본 출원에서 노이즈 신호의 다수의 프레임이 결정된 후, 파워 스펙트럼 추정값 P_{noise} 는 일반적으로 과대 추정(over-estimation)을 피하기 위해 프레임들 중 일부를 사용하여 계산된다. 예를 들어, 결정된 노이즈 신호가 50 프레임을 포함하면, 파워 스펙트럼 추정값 P_{noise} 를 계산하기 위해 처음 30 프레임이 캡처될 수 있다. 이와 같이, 파워 스펙트럼 추정값의 정확도를 향상시킬 수 있다.
- [0088] 본 출원의 실시 예는 상기 처리 구현에 대응하는 노이즈 신호 결정 장치를 더 제공한다. 상기 장치는 소프트웨어를 통해 구현될 수 있으며, 또한 하드웨어 또는 소프트웨어와 하드웨어의 조합을 통해 구현될 수도 있다. 소프트웨어 구현 방식을 예로 들자면, 서버의 중앙 처리 장치(Central Process Unit, CPU)를 통해 대응하는 컴퓨터 프로그램을 메모리로 판독하고 상기 컴퓨터 프로그램을 실행함으로써 로직 의미에서의 장치가 형성될 수 있다. 장치의 하드웨어 구조를 위해서는 도 8을 참조하라.
- [0089] 도 6은 본 출원의 실시 예에 따른 노이즈 신호 검출 장치의 블록도이다. 본 실시 예에서, 장치 내의 유닛들의 기능들은 상기 노이즈 신호 결정 방법의 단계들의 기능들에 대응할 수 있다. 자세한 내용은 위의 방법 실시 예를 참조하라. 노이즈 신호 결정 장치(100)는:
- [0090] 분석 대상 음성 신호 세그먼트 내의 각 프레임 신호에 푸리에 변환을 수행하여 상기 음성 신호 세그먼트 내의 각 프레임 신호의 파워 스펙트럼을 획득하도록 구성된 파워 스펙트럼 획득 유닛(101);
- [0091] 상기 프레임 신호의 파워 스펙트럼에 기초하여 다양한 주파수에서 상기 음성 신호 세그먼트 내의 각 프레임 신호의 파워 값들의 분산을 결정하도록 구성된 분산 결정 유닛(102); 및
- [0092] 상기 분산에 기초하여 상기 음성 신호 세그먼트 내의 각 프레임 신호가 노이즈 신호인지 여부를 결정하도록 구성된 노이즈 결정 유닛(103) 포함한다.
- [0093] 바람직하게는, 상기 장치는, 세그먼트 획득 유닛을 더 포함하고, 상기 세그먼트 획득 유닛은:

- [0094] 처리 대상 음성의 시간-영역 신호의 진폭 변화에 기초하여, 상기 처리 대상 음성에서 미리 설정된 임계치보다 작은 진폭 변화를 갖는 음성 신호 세그먼트를 분석 대상 음성 신호 세그먼트로서 결정하거나;
- [0095] 처리 대상 음성에서 처음 N개의 프레임 음성 신호를 상기 분석 대상 음성 신호 세그먼트로서 캡처하도록 구성된다.
- [0096] 바람직하게는 상기 노이즈 결정 유닛(103)은,
- [0097] 상기 음성 신호 세그먼트 내의 각 프레임 신호에 대응하는 상기 분산이 제1 임계치보다 큰지 여부를 결정하고;
- [0098] 아니오이면, 상기 프레임 신호를 노이즈 신호로서 결정하도록 구성된다.
- [0099] 바람직하게는 상기 분산 결정 유닛(102)은,
- [0100] 상기 파워 스펙트럼에 대응하는 주파수가 속하는 주파수 구간에 따라, 다양한 주파수에서 상기 프레임 신호의 파워 값들을 제1 주파수 구간에 대응하는 제1 파워 값 세트에 적어도 분류하고;
- [0101] 상기 제1 파워 값 세트 내에 포함된 파워 값들의 제1 분산을 결정하도록 구성된다.
- [0102] 그 후, 상기 노이즈 결정 유닛(103)은,
- [0103] 상기 제1 분산이 상기 제1 임계치보다 큰지 여부를 결정하고;
- [0104] 아니오이면, 상기 프레임 신호를 노이즈 신호로서 결정하도록 구성된다.
- [0105] 바람직하게는 상기 분산 결정 유닛(102)은 구체적으로,
- [0106] 상기 프레임 신호의 파워 값들에 대응하는 주파수가 속하는 주파수 구간에 따라, 다양한 주파수에서 각 프레임 신호의 파워 값들을 제1 주파수 구간에 대응하는 제1 파워 값 세트 및 제2 주파수 구간에 대응하는 제2 파워 값 세트에 적어도 분류하고 - 상기 제1 주파수 구간은 상기 제2 주파수 구간보다 낮음 -;
- [0107] 상기 제1 파워 값 세트 내에 포함된 파워 값들의 제1 분산을 결정하며;
- [0108] 상기 제2 파워 값 세트 내에 포함된 파워 값들의 제2 분산을 결정하도록 구성된다.
- [0109] 그 후, 상기 노이즈 결정 유닛(103)은,
- [0110] 각 프레임 신호에 대응하는 상기 제1 분산 및 상기 제2 분산의 차이가 제2 임계치보다 큰지 여부를 결정하고;
- [0111] 아니오이면, 상기 프레임 신호를 노이즈 신호로서 결정하도록 구성된다.
- [0112] 본 출원의 실시 예는 또한 상기 처리 구현에 대응하는 음성 노이즈 제거 장치를 제공한다. 상기 장치는 소프트웨어를 통해 구현될 수 있으며, 하드웨어 또는 소프트웨어와 하드웨어의 조합을 통해 구현될 수도 있다. 소프트웨어 구현 방식을 예로 들자면, 서버의 중앙 처리 장치(CPU)를 통해 대응하는 컴퓨터 프로그램을 메모리로 판독하고 상기 컴퓨터 프로그램을 실행함으로써 로직 의미에서의 장치가 형성될 수 있다. 장치의 하드웨어 구조를 위해서는 도 8을 참조하라.
- [0113] 도 7은 본 출원의 실시 예에 따른 음성 노이즈 제거 장치의 블록도이다. 이 실시 예에서, 장치 내의 유닛들의 기능은 상기 음성 노이즈 제거 방법의 단계들의 기능들에 대응할 수 있다. 자세한 내용은 상기 방법 실시 예를 참조하라. 이 실시 예에서, 음성 노이즈 제거 장치(200)는:
- [0114] 처리 대상 음성 내에 포함된 분석 대상 음성 신호 세그먼트를 결정하도록 구성된 세그먼트 결정 유닛(201);
- [0115] 상기 분석 대상 음성 신호 세그먼트 내의 각 프레임 신호에 푸리에 변환을 수행하여 상기 음성 신호 세그먼트 내의 각 프레임 신호의 파워 스펙트럼을 획득하도록 구성된 파워 스펙트럼 획득 유닛(202);
- [0116] 상기 프레임 신호의 파워 스펙트럼에 기초하여 다양한 주파수에서 상기 음성 신호 세그먼트 내의 각 프레임 신호의 파워 값들의 분산을 결정하도록 구성된 분산 결정 유닛(203);
- [0117] 상기 분산에 기초하여 상기 음성 신호 세그먼트 내의 각 프레임 신호가 노이즈 신호인지 여부를 결정하고, 상기 음성 신호 세그먼트 내에 포함된 여러 노이즈 프레임을 획득하도록 구성된 노이즈 결정 유닛(205); 및
- [0118] 상기 음성 신호 세그먼트 내에 포함된 여러 노이즈 프레임에 대응하는 평균 파워를 결정하고, 상기 노이즈 프레임의 평균 파워에 기초하여 상기 처리 대상 음성에서 노이즈를 제거하도록 구성된 음성 노이즈 제거 유닛(10)을

포함한다.

- [0119] 바람직하게는 상기 장치는: 상기 분산의 크기에 따라 상기 분석 대상 음성 신호 세그먼트 내의 프레임 신호를 순위 매김하도록 구성된 순위 매김(ranking) 유닛(204)을 더 포함한다.
- [0120] 그 후 노이즈 결정 유닛(205)은 구체적으로:
- [0121] 다양한 주파수에서 각각의 순위 매김된 프레임 신호의 파워 값들의 분산에 기초하여 음성 신호 세그먼트 내의 각 프레임 신호가 노이즈 신호인지 여부를 결정하도록 구성된다.
- [0122] 분석 대상 음성 신호 세그먼트에 푸리에 변환을 수행하여 각 프레임 신호의 파워 스펙트럼을 획득하고, 다양한 주파수에서 상기 분석 대상 음성 신호 세그먼트 내의 각 프레임 신호의 파워 값들의 분산을 결정하고, 최종적으로 상기 분산에 기초하여 상기 프레임 신호가 노이즈 신호인지 여부를 결정함으로써, 본 출원의 실시 예들에 제공된 음성 노이즈 제거 방법 및 장치 뿐만 아니라 노이즈 신호 결정 방법 및 장치가 상기 분석 대상 음성 신호 세그먼트 내에 포함된 여러 노이즈 프레임을 정확히 결정할 수 있다. 음성 노이즈 제거 프로세스에서 결정된 여러 노이즈 프레임의 평균 파워에 기초하여 처리 대상 음성에서 노이즈가 제거될 수 있고, 따라서 음성 노이즈 제거 효과가 향상된다.
- [0123] 설명을 용이하게 하기 위해, 장치는 각각의 설명을 위한 기능적 관점에서 다양한 유닛으로 분할된다. 당연히, 본 출원이 구현될 때, 상기 유닛들의 기능들은 동일한 소프트웨어 및/또는 하드웨어 컴포넌트 또는 다수의 소프트웨어 및/또는 하드웨어 컴포넌트들로 구현될 수 있다.
- [0124] 당업자는 본 발명의 실시 예가 방법, 시스템 또는 컴퓨터 프로그램 제품으로서 제공될 수 있다는 것을 이해해야 한다. 따라서, 본 발명은 완전한 하드웨어 실시 예, 완전한 소프트웨어 실시 예 또는 소프트웨어와 하드웨어를 결합한 실시 예로서 구현될 수 있다. 또한, 본 발명은 컴퓨터 사용 가능한 프로그램 코드를 포함하는 (자기 디스크 메모리, CD-ROM, 광학 메모리 등을 포함하지만 이에 한정되지 않는) 하나 이상의 컴퓨터 사용 가능한 저장 매체 상에 구현된 컴퓨터 프로그램 제품의 형태일 수 있다.
- [0125] 본 발명은 본 발명의 실시 예에 따른 방법, 디바이스(시스템) 및 컴퓨터 프로그램 제품에 따른 흐름도 및/또는 블록도를 참조하여 설명된다. 흐름도 및/또는 블록도에서 각각의 프로세스 및/또는 블록 및 프로세스 및/또는 블록의 조합을 구현하기 위해 컴퓨터 프로그램 명령어가 사용될 수 있다는 것을 이해해야 한다. 컴퓨터 프로그램 명령어는 범용 컴퓨터, 특수 목적 컴퓨터, 내장형 프로세서 또는 다른 프로그램 가능 데이터 처리 디바이스에 제공되어 기계(machine)를 생성할 수 있어, 컴퓨터 또는 다른 프로그램 가능 데이터 처리 디바이스의 프로세서는 흐름도의 하나 이상의 프로세스 및/또는 블록도 내의 하나 이상의 블록에서 지정된 기능들을 구현하도록 구성된 장치를 생성하기 위한 명령어를 실행한다.
- [0126] 컴퓨터 프로그램 명령어들은 또한 컴퓨터 또는 다른 프로그램 가능한 데이터 처리 디바이스가 특정 방식으로 동작하도록 안내할 수 있는 컴퓨터 판독 가능 저장 장치에 저장될 수 있어서, 컴퓨터 판독 가능 저장 장치에 저장된 명령어는, 흐름도의 하나 이상의 프로세스 및/또는 블록도의 하나 이상의 블록에 의해 지정된 기능들을 구현하는 명령어 장치를 포함하는 제조물(manufacture)을 생성한다.
- [0127] 또한, 컴퓨터 프로그램 명령어들은 또한 컴퓨터 또는 다른 프로그램 가능한 데이터 처리 디바이스에 로딩되어 일련의 동작 단계가 컴퓨터 구현 프로세싱을 생성하기 위해 컴퓨터 또는 다른 프로그램 가능한 디바이스상에서 실행될 수 있다. 따라서, 컴퓨터 또는 다른 프로그램 가능한 디바이스에서 실행되는 명령어는 흐름도의 하나 이상의 프로세스 및/또는 블록도의 하나 이상의 블록에서 지정된 기능을 구현하기 위한 단계들을 제공한다.
- [0128] "포함하다(include)" 또는 "포함하다(comprise)"라는 용어 또는 이들의 다른 변형은 비배타적인 것을 커버하도록 의도되며, 따라서, 일련의 구성 요소(element)를 포함하는 프로세스, 방법, 상품 또는 디바이스는 그 구성 요소들을 포함할 뿐만 아니라 명백히 나열되지 않은 다른 구성 요소들도 포함하거나, 프로세스, 방법, 상품 또는 디바이스의 내재된 구성 요소들을 더 포함한다는 것을 또한 주의하여야 한다. 더 이상 제한이 없는 경우, "...을 포함하는(including a/an ...)"에 의해 정의된 구성 요소는 그 구성 요소를 포함하는 프로세스, 방법, 상품 또는 디바이스가 다른 동일한 구성 요소를 더 포함한다는 것을 배제하지 않는다.
- [0129] 당업자는 본 출원의 실시 예가 방법, 시스템 또는 컴퓨터 프로그램 제품으로서 제공될 수 있다는 것을 이해해야 한다. 따라서, 본 출원은 완전한 하드웨어 실시 예, 완전한 소프트웨어 실시 예, 또는 소프트웨어와 하드웨어를 결합한 실시 예의 형태로 구현될 수 있다. 또한, 본 출원은 컴퓨터 사용 가능한 프로그램 코드를 포함하는 (자기 디스크 메모리, CD-ROM, 광학 메모리 등을 포함하지만 이에 한정되지 않는) 하나 이상의 컴퓨터 사용 가능한

저장 매체 상에 구현된 컴퓨터 프로그램 제품의 형태일 수 있다.

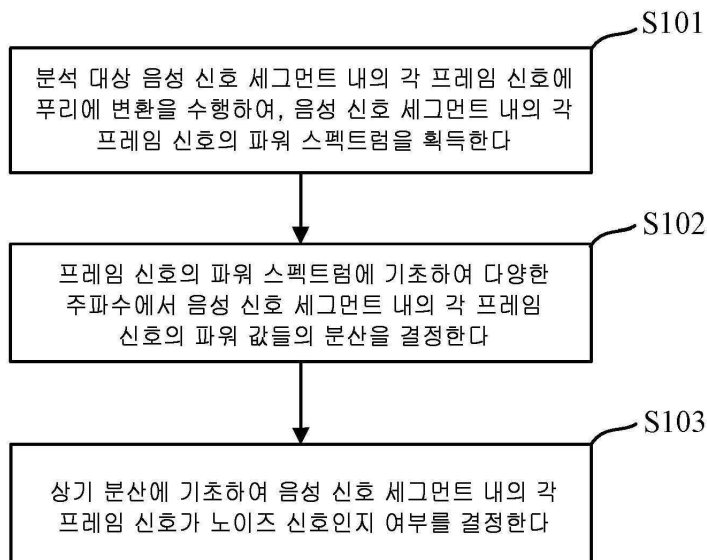
[0130] 본 출원은 컴퓨터에 의해 실행되는 컴퓨터 실행 가능 명령어, 예를 들어 프로그램 모듈의 공통 컨텍스트로 기술될 수 있다. 일반적으로, 프로그램 모듈은 특정 태스크를 실행하거나 특정 추상 데이터 타입을 구현하는데 사용되는 루틴, 프로그램, 객체, 어셈블리, 데이터 구조 등을 포함한다. 본 출원은 또한 통신 네트워크를 통해 접속된 원격 프로세싱 디바이스를 사용하여 태스크가 실행되는 분산 컴퓨팅 환경에서도 구현될 수 있다. 분산 컴퓨터 환경에서, 프로그램 모듈은 저장 디바이스를 포함하는 로컬 및 원격 컴퓨터 저장 매체에 위치할 수 있다.

[0131] 본 명세서의 실시 예들은 계속해서 설명되고, 실시 예의 동일 또는 유사한 부분들이 서로 관련하여 얻어질 수 있으며, 각각의 실시 예는 다른 실시 예와 상이한 부분을 강조한다. 특히, 시스템 실시 예는 기본적으로 방법 실시 예와 유사하므로 간단히 설명되어 있다. 관련 부분에 대해서는, 방법 실시 예의 부분의 설명을 참조하라.

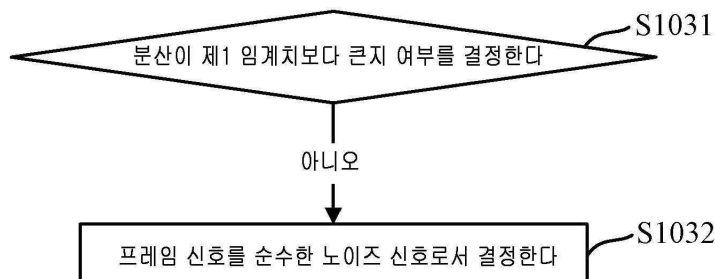
[0132] 상기 설명은 단지 본 출원의 실시 예에 불과하며, 본 출원을 제한하고자 하는 것은 아니다. 본 출원의 다양한 수정 및 변형이 당업자에게 가능하다. 본 출원의 정신 및 원리 내에서 이루어진 모든 수정, 균등한 대체, 개선 등은 모두 본 출원의 청구 범위 내에 속해야 한다.

도면

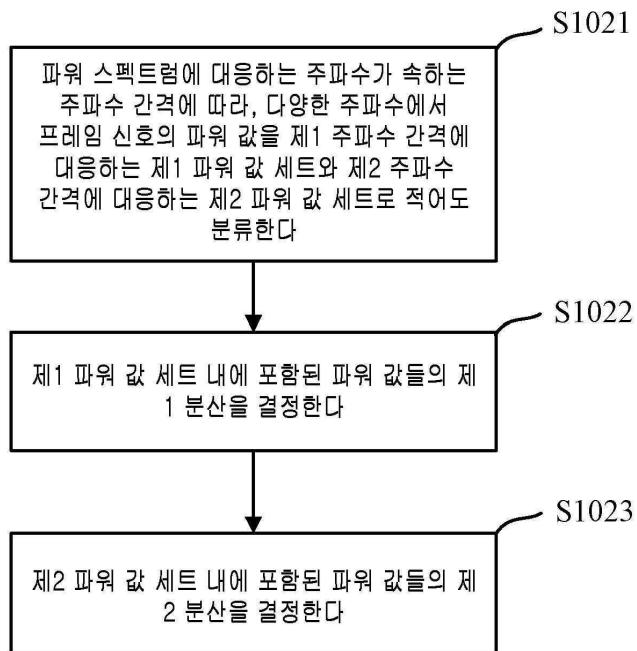
도면1



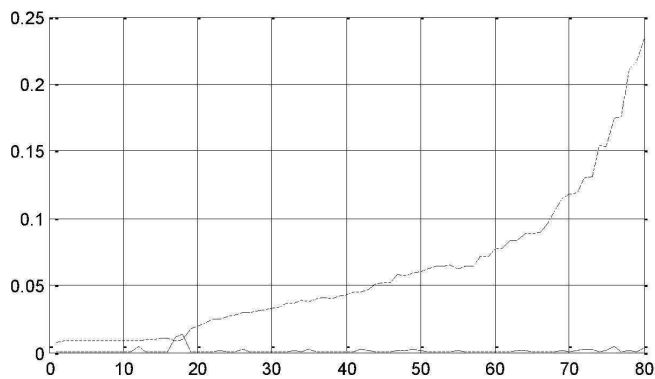
도면2



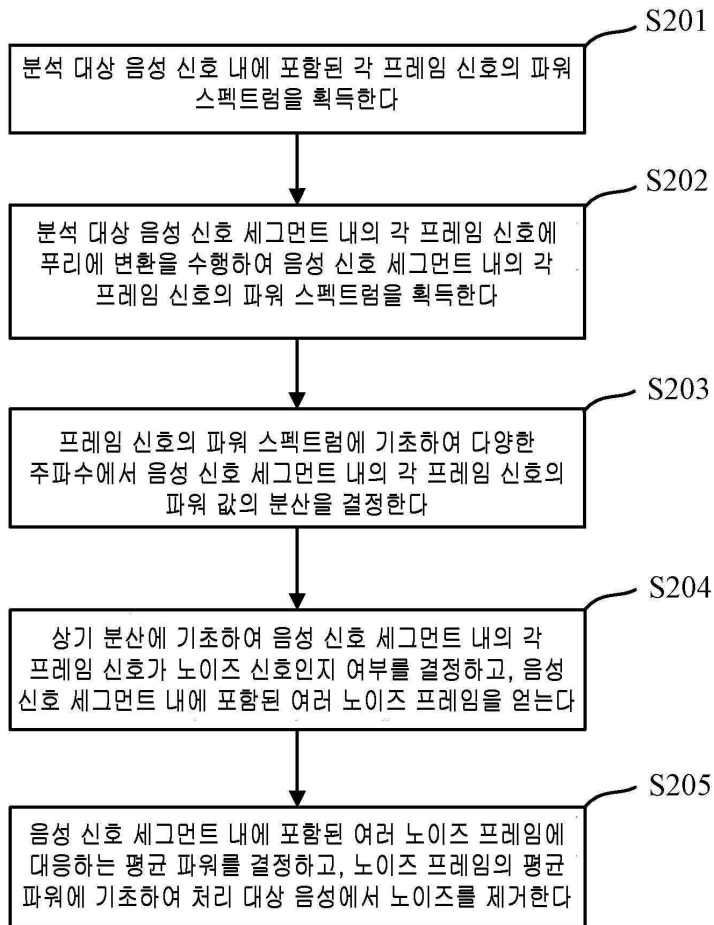
도면3



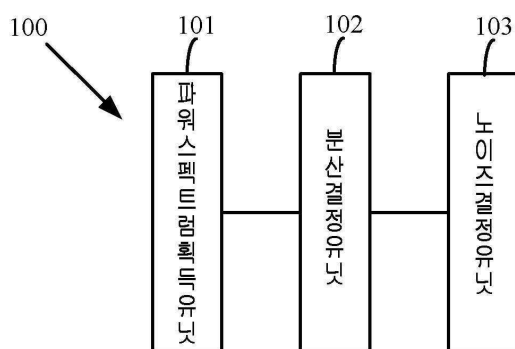
도면4



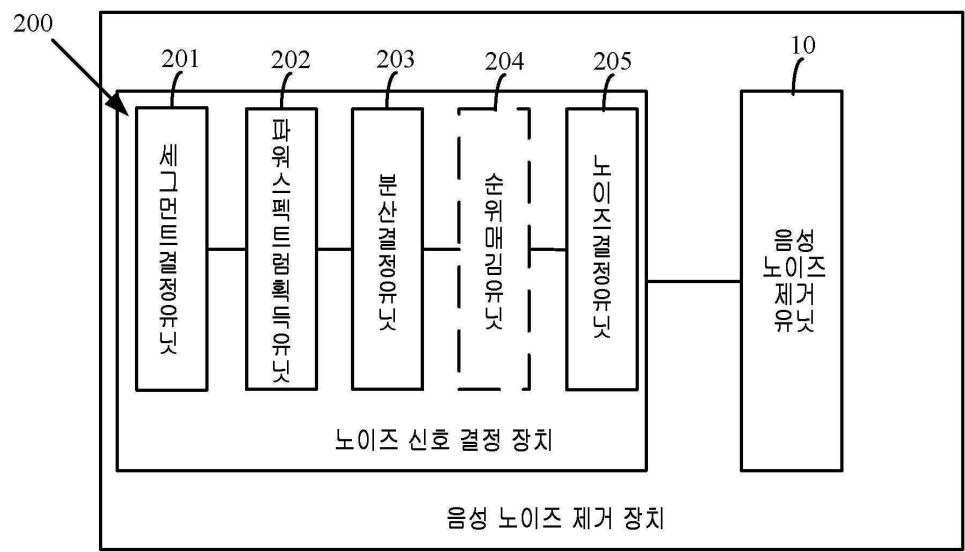
도면5



도면6



도면7



도면8

