



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2016-0023767
(43) 공개일자 2016년03월03일

- | | |
|---|---|
| (51) 국제특허분류(Int. Cl.)
<i>G1OL 15/02</i> (2006.01) <i>G1OL 25/00</i> (2014.01)
<i>G1OL 25/69</i> (2013.01) | (71) 출원인
퀄컴 인코포레이티드
미국 92121-1714 캘리포니아주 샌디에고 모어하우스 드라이브 5775 |
| (52) CPC특허분류
<i>G1OL 15/02</i> (2013.01)
<i>G1OL 25/00</i> (2013.01) | (72) 발명자
센 디파운드
미국 92121 캘리포니아주 샌디에고 모어하우스 드라이브 5775 |
| (21) 출원번호 10-2016-7000189 | (74) 대리인
루 웬량
미국 92121 캘리포니아주 샌디에고 모어하우스 드라이브 5775 |
| (22) 출원일자(국제) 2014년06월25일 | 특허법인코리아나 |
| 심사청구일자 없음 | |
| (85) 번역문제출일자 2016년01월05일 | |
| (86) 국제출원번호 PCT/US2014/044163 | |
| (87) 국제공개번호 WO 2014/210204 | |
| 국제공개일자 2014년12월31일 | |
| (30) 우선권주장
61/839,807 2013년06월26일 미국(US)
(뒷면에 계속) | |

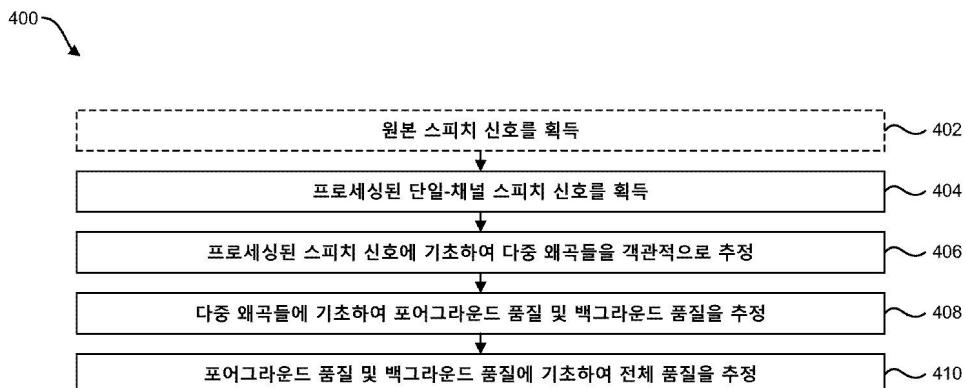
전체 청구항 수 : 총 30 항

(54) 발명의 명칭 스피치 신호 품질을 측정하기 위한 시스템들 및 방법들

(57) 요 약

전자 디바이스에 의해 스피치 신호 품질을 측정하기 위한 방법이 설명된다. 방법은 수정된 단일-채널 스피치 신호를 획득하는 단계를 포함한다. 방법은 또한, 수정된 단일-채널 스피치 신호에 기초하여 다중 객관적 왜곡들을 추정하는 단계를 포함한다. 다중 객관적 왜곡들은 적어도 하나의 포어그라운드 왜곡 및 적어도 하나의 백그라운드 왜곡을 포함한다. 방법은 다중 객관적 왜곡들에 기초하여 포어그라운드 품질 및 백그라운드 품질을 추정하는 단계를 더 포함한다. 방법은 추가적으로, 포어그라운드 품질 및 백그라운드 품질에 기초하여 전체 품질을 추정하는 단계를 포함한다.

대 표 도



(52) CPC특허분류
G10L 25/69 (2013.01)

(30) 우선권주장

61/839,796	2013년06월26일	미국(US)
61/839,800	2013년06월26일	미국(US)
61/876,177	2013년09월10일	미국(US)
61/888,945	2013년10월09일	미국(US)
14/314,019	2014년06월24일	미국(US)

명세서

청구범위

청구항 1

전자 디바이스에 의해 스피치 신호 품질을 측정하기 위한 방법으로서, 수정된 단일-채널 스피치 신호를 획득하는 단계; 상기 수정된 단일-채널 스피치 신호에 기초하여 다중 객관적 왜곡들을 추정하는 단계로서, 상기 다중 객관적 왜곡들은 적어도 하나의 포어그라운드 왜곡 및 적어도 하나의 백그라운드 왜곡을 포함하는, 상기 다중 객관적 왜곡들을 추정하는 단계; 상기 다중 객관적 왜곡들에 기초하여 포어그라운드 품질 및 백그라운드 품질을 추정하는 단계; 및 상기 포어그라운드 품질 및 상기 백그라운드 품질에 기초하여 전체 품질을 추정하는 단계를 포함하는, 스피치 신호 품질을 측정하기 위한 방법.

청구항 2

제 1 항에 있어서,

스피치 신호 품질을 측정하는 것은, 복수의 계층적 레벨들을 포함하는 계층적 구조에 따라 수행되고, 각각의 계층적 레벨은 임의의 그리고 각각의 상태적으로 더 낮은 계층적 레벨에 종속되는, 스피치 신호 품질을 측정하기 위한 방법.

청구항 3

제 2 항에 있어서,

상기 복수의 계층적 레벨들은 3 개의 레벨들을 포함하고, 제 1 계층적 레벨은 상기 다중 객관적 왜곡들을 추정하는 것을 포함하고, 제 2 계층적 레벨은 상기 포어그라운드 품질 및 상기 백그라운드 품질을 추정하는 단계를 포함하고, 제 3 계층적 레벨은 상기 전체 품질을 추정하는 단계를 포함하는, 스피치 신호 품질을 측정하기 위한 방법.

청구항 4

제 1 항에 있어서,

상기 전체 품질을 추정하는 단계는 추가로, 상기 다중 객관적 왜곡들에 직접적으로 기초하는, 스피치 신호 품질을 측정하기 위한 방법.

청구항 5

제 1 항에 있어서,

상기 다중 객관적 왜곡들의 각각은 스피치 품질의 독립적인 차원을 나타내는, 스피치 신호 품질을 측정하기 위한 방법.

청구항 6

제 1 항에 있어서,

상기 포어그라운드 왜곡은 러프니스 (roughness), 불연속성 (discontinuity), 둔탁성 (dullness), 및 빈약함 (thinness) 을 포함하는, 스피치 신호 품질을 측정하기 위한 방법.

청구항 7

제 1 항에 있어서,

상기 백그라운드 왜곡은 히싱 (hissing) 및 가변성 (variability) 을 포함하는, 스피치 신호 품질을 측정하기 위한 방법.

청구항 8

제 1 항에 있어서,

상기 수정된 단일-채널 스피치 신호는 원본 스피치 신호에 기초하는, 스피치 신호 품질을 측정하기 위한 방법.

청구항 9

제 1 항에 있어서,

상기 다중 객관적 왜곡들은 원본 스피치 신호 없이 추정되는, 스피치 신호 품질을 측정하기 위한 방법.

청구항 10

제 1 항에 있어서,

원본 스피치 신호를 획득하는 단계를 더 포함하고, 상기 다중 객관적 왜곡들을 추정하는 단계는 추가로, 상기 원본 스피치 신호에 기초하는, 스피치 신호 품질을 측정하기 위한 방법.

청구항 11

스피치 신호 품질을 측정하기 위한 전자 디바이스로서,

수정된 단일-채널 스피치 신호에 기초하여 다중 객관적 왜곡들을 추정하는 연산 왜곡 추정기로서, 상기 다중 객관적 왜곡들은 적어도 하나의 포어그라운드 왜곡 및 적어도 하나의 백그라운드 왜곡을 포함하는, 상기 연산 왜곡 추정기;

상기 연산 왜곡 추정기에 결합된 연산 포어그라운드 품질 추정기로서, 상기 연산 포어그라운드 품질 추정기는 상기 다중 객관적 왜곡들에 기초하여 포어그라운드 품질을 추정하는, 상기 연산 포어그라운드 품질 추정기;

상기 연산 왜곡 추정기에 결합된 연산 백그라운드 품질 추정기로서, 상기 연산 백그라운드 품질 추정기는 상기 다중 객관적 왜곡들에 기초하여 백그라운드 품질을 추정하는, 상기 연산 백그라운드 품질 추정기; 및

상기 연산 왜곡 추정기에 결합된 연산 전체 품질 추정기로서, 상기 연산 전체 품질 추정기는 상기 포어그라운드 품질 및 상기 백그라운드 품질에 기초하여 전체 품질을 추정하는, 상기 연산 전체 품질 추정기를 포함하는, 스피치 신호 품질을 측정하기 위한 전자 디바이스.

청구항 12

제 11 항에 있어서,

스피치 신호 품질을 측정하는 것은, 복수의 계층적 레벨들을 포함하는 계층적 구조에 따라 수행되고, 각각의 계층적 레벨은 임의의 그리고 각각의 상대적으로 더 낮은 계층적 레벨에 종속되는, 스피치 신호 품질을 측정하기 위한 전자 디바이스.

청구항 13

제 12 항에 있어서,

상기 복수의 계층적 레벨들은 3 개의 레벨들을 포함하고, 제 1 계층적 레벨은 상기 다중 객관적 왜곡들을 추정하는 것을 포함하고, 제 2 계층적 레벨은 상기 포어그라운드 품질 및 상기 백그라운드 품질을 추정하는 것을 포함하고, 제 3 계층적 레벨은 상기 전체 품질을 추정하는 것을 포함하는, 스피치 신호 품질을 측정하기 위한 전자 디바이스.

청구항 14

제 11 항에 있어서,

상기 다중 객관적 왜곡들은 원본 스피치 신호 없이 추정되는, 스피치 신호 품질을 측정하기 위한 전자

디바이스.

청구항 15

제 11 항에 있어서,

상기 연산 왜곡 추정기는 추가로 원본 스피치 신호에 기초하여 상기 다중 객관적 왜곡들을 추정하는, 스피치 신호 품질을 측정하기 위한 전자 디바이스.

청구항 16

명령들을 갖는 비-일시적인 유형의 컴퓨터-판독가능한 매체를 포함하는, 스피치 신호 품질을 측정하기 위한 컴퓨터-프로그램 제품으로서,

상기 명령들은,

전자 디바이스로 하여금 수정된 단일-채널 스피치 신호를 획득하게 하기 위한 코드;

상기 전자 디바이스로 하여금 상기 수정된 단일-채널 스피치 신호에 기초하여 다중 객관적 왜곡들을 추정하게 하기 위한 코드로서, 상기 다중 객관적 왜곡들은 적어도 하나의 포어그라운드 왜곡 및 적어도 하나의 백그라운드 왜곡을 포함하는, 상기 다중 객관적 왜곡들을 추정하게 하기 위한 코드;

상기 전자 디바이스로 하여금 상기 다중 객관적 왜곡들에 기초하여 포어그라운드 품질 및 백그라운드 품질을 추정하게 하기 위한 코드; 및

상기 전자 디바이스로 하여금 상기 포어그라운드 품질 및 상기 백그라운드 품질에 기초하여 전체 품질을 추정하게 하기 위한 코드를 포함하는, 컴퓨터-프로그램 제품.

청구항 17

제 16 항에 있어서,

스피치 신호 품질을 측정하는 것은, 복수의 계층적 레벨들을 포함하는 계층적 구조에 따라 수행되고, 각각의 계층적 레벨은 임의의 그리고 각각의 상대적으로 더 낮은 계층적 레벨에 종속되는, 컴퓨터-프로그램 제품.

청구항 18

제 17 항에 있어서,

상기 복수의 계층적 레벨들은 3 개의 레벨들을 포함하고, 제 1 계층적 레벨은 상기 다중 객관적 왜곡들을 추정하는 것을 포함하고, 제 2 계층적 레벨은 상기 포어그라운드 품질 및 상기 백그라운드 품질을 추정하는 것을 포함하고, 제 3 계층적 레벨은 상기 전체 품질을 추정하는 것을 포함하는, 컴퓨터-프로그램 제품.

청구항 19

제 16 항에 있어서,

상기 다중 객관적 왜곡들은 원본 스피치 신호 없이 추정되는, 컴퓨터-프로그램 제품.

청구항 20

제 16 항에 있어서,

상기 전자 디바이스로 하여금 원본 스피치 신호를 획득하게 하기 위한 코드를 더 포함하고, 상기 다중 객관적 왜곡들을 추정하는 것은 추가로, 상기 원본 스피치 신호에 기초하는, 컴퓨터-프로그램 제품.

청구항 21

스피치 신호 품질을 측정하기 위한 장치로서,

수정된 단일-채널 스피치 신호를 획득하기 위한 수단;

상기 수정된 단일-채널 스피치 신호에 기초하여 다중 객관적 왜곡들을 추정하기 위한 수단으로서, 상기 다중 객관적 왜곡들은 적어도 하나의 포어그라운드 왜곡 및 적어도 하나의 백그라운드 왜곡을 포함하는, 상기 다중 객

관적 왜곡들을 추정하기 위한 수단;

상기 다중 객관적 왜곡들에 기초하여 포어그라운드 품질 및 백그라운드 품질을 추정하기 위한 수단; 및

상기 포어그라운드 품질 및 상기 백그라운드 품질에 기초하여 전체 품질을 추정하기 위한 수단을 포함하는, 스피치 신호 품질을 측정하기 위한 장치.

청구항 22

제 21 항에 있어서,

스피치 신호 품질을 측정하는 것은, 복수의 계층적 레벨들을 포함하는 계층적 구조에 따라 수행되고, 각각의 계층적 레벨은 임의의 그리고 각각의 상대적으로 더 낮은 계층적 레벨에 종속되는, 스피치 신호 품질을 측정하기 위한 장치.

청구항 23

제 22 항에 있어서,

상기 복수의 계층적 레벨들은 3 개의 레벨들을 포함하고, 제 1 계층적 레벨은 상기 다중 객관적 왜곡들을 추정하는 것을 포함하고, 제 2 계층적 레벨은 상기 포어그라운드 품질 및 상기 백그라운드 품질을 추정하는 것을 포함하고, 제 3 계층적 레벨은 상기 전체 품질을 추정하는 것을 포함하는, 스피치 신호 품질을 측정하기 위한 장치.

청구항 24

제 21 항에 있어서,

상기 전체 품질을 추정하는 것은 추가로, 상기 다중 객관적 왜곡들에 직접적으로 기초하는, 스피치 신호 품질을 측정하기 위한 장치.

청구항 25

제 21 항에 있어서,

상기 다중 객관적 왜곡들의 각각은 스피치 품질의 독립적인 차원을 나타내는, 스피치 신호 품질을 측정하기 위한 장치.

청구항 26

제 21 항에 있어서,

상기 포어그라운드 왜곡은 러프니스, 불연속성, 둔탁성, 및 빈약함을 포함하는, 스피치 신호 품질을 측정하기 위한 장치.

청구항 27

제 21 항에 있어서,

상기 백그라운드 왜곡은 허성 및 가변성을 포함하는, 스피치 신호 품질을 측정하기 위한 장치.

청구항 28

제 21 항에 있어서,

상기 수정된 단일-채널 스피치 신호는 원본 스피치 신호에 기초하는, 스피치 신호 품질을 측정하기 위한 장치.

청구항 29

제 21 항에 있어서,

상기 다중 객관적 왜곡들은 원본 스피치 신호 없이 추정되는, 스피치 신호 품질을 측정하기 위한 장치.

청구항 30

제 21 항에 있어서,

원본 스피치 신호를 획득하기 위한 수단을 더 포함하고, 상기 다중 객관적 왜곡들을 추정하는 것은 추가로, 상기 원본 스피치 신호에 기초하는, 스피치 신호 품질을 측정하기 위한 장치.

발명의 설명

기술 분야

관련 출원들

이 출원은 "SYSTEMS AND METHODS FOR MEASURING SPEECH SIGNAL QUALITY (스피치 신호 품질을 측정하기 위한 시스템들 및 방법들)"에 대하여 2013년 06월 26일자로 출원된 미국 특허 출원 제 61/839,807 호, "SYSTEMS AND METHODS FOR MEASURING SPEECH SIGNAL QUALITY (스피치 신호 품질을 측정하기 위한 시스템들 및 방법들)"에 대하여 2013년 10월 09일자로 출원된 미국 특허 출원 제 61/888,945 호, "SYSTEMS AND METHODS FOR INTRUSIVE FEATURE EXTRACTION (침투적 특징 추출을 위한 시스템들 및 방법들)"에 대하여 2013년 06월 26일자로 출원된 미국 특허 출원 제 61/839,796 호, "SYSTEMS AND METHODS FOR NON-INTRUSIVE FEATURE EXTRACTION (비-침투적 특징 추출을 위한 시스템들 및 방법들)"에 대하여 2013년 06월 26일자로 출원된 미국 특허 출원 제 61/839,800 호, 및 "SYSTEMS AND METHODS FOR FEATURE EXTRACTION (특징 추출을 위한 시스템들 및 방법들)"에 대하여 2013년 09월 10일자로 출원된 미국 특허 출원 제 61/876,177 호에 관련되고 이에 대한 우선권을 주장한다.

기술 분야

본 개시물은 일반적으로 전자 디바이스들에 관한 것이다. 더욱 구체적으로, 본 개시물은 스피치 신호 품질 (speech signal quality)을 측정하기 위한 시스템들 및 방법들에 관한 것이다.

배경 기술

[0005] 최근 수십 년 내에, 전자 디바이스들의 이용이 보편적인 것으로 되었다. 특히, 전자 기술에서의 진보들은 점점 더 복잡하고 유용한 전자 디바이스들의 비용을 감소시켰다. 비용 감소 및 소비자 요구는, 전자 디바이스들이 실제적으로 현대 사회에서 유비쿼터스 (ubiquitous) 가 되도록, 전자 디바이스들의 이용을 급증시켰다. 전자 디바이스들의 이용이 확장함에 따라, 전자 디바이스들의 새롭고 개선된 특징들에 대한 요구도 확장하였다. 더욱 구체적으로, 새로운 기능들을 수행하고 및/또는 기능들을 더 빠르게, 더욱 효율적으로, 또는 더 높은 품질로 수행하는 전자 디바이스들이 종종 추구된다.

[0006] 일부의 전자 디바이스들 (예컨대, 셀룰러 전화들, 스마트폰들, 오디오 레코더들, 캠코더들, 컴퓨터들 등)은 오디오 신호들을 캡처하거나 사용한다. 예를 들어, 스마트폰은 스피치 신호를 캡처할 수도 있다.

[0007] 스피치 신호 품질은 수량화하기 어려울 수도 있다. 또한, 사람들이 스피치 신호들을 어떻게 지각하는지에 영향을 주는 스피치 신호들의 특성들은 식별하고 및/또는 수량화하기가 어려울 수도 있다. 스피치 신호 품질을 평가하는 것은 시간 소비적이고 마찬가지로 고비용일 수도 있다. 이 논의로부터 관찰될 수 있는 바와 같이, 스피치 신호들의 평가를 개선시키는 시스템들 및 방법들이 유익할 수도 있다.

발명의 내용

과제의 해결 수단

[0008] 전자 디바이스에 의해 스피치 신호 품질을 측정하기 위한 방법이 설명된다. 수정된 단일-채널 스피치 신호가 획득된다. 다중 객관적 왜곡 (multiple objective distortion) 들은 수정된 단일-채널 스피치 신호에 기초하여 추정된다. 다중 객관적 왜곡들은 적어도 하나의 포어그라운드 왜곡 (foreground distortion) 및 적어도 하나의 백그라운드 왜곡 (background distortion) 을 포함한다. 포어그라운드 품질 및 백그라운드 품질은 다중 객관적 왜곡들에 기초하여 추정된다. 전체 품질은 포어그라운드 품질 및 백그라운드 품질에 기초하여 추정된다.

[0009] 스피치 신호 품질을 측정하는 것은 복수의 계층적 레벨들을 포함하는 계층적 구조에 따라 수행될 수도 있다. 각각의 계층적 레벨은 임의의 그리고 각각의 상대적으로 더 낮은 계층적 레벨에 종속될 수도 있다. 복수의 계층적 레벨들은 3개의 레벨들을 포함할 수도 있다. 제 1 계층적 레벨은 다중 객관적 왜곡들을 추정하

는 것을 포함할 수도 있다. 제 2 계층적 레벨은 포어그라운드 품질 및 백그라운드 품질을 추정하는 것을 포함할 수도 있다. 제 3 계층적 레벨은 전체 품질을 추정하는 것을 포함할 수도 있다.

[0010] 전체 품질을 추정하는 것은 또한, 다중 객관적 왜곡들에 직접적으로 기초할 수도 있다. 다중 객관적 왜곡들의 각각은 스피치 품질의 독립적인 차원을 나타낼 수도 있다. 포어그라운드 왜곡들은 러프니스(roughness), 불연속성(discontinuity), 둔탁성(dullness) 및 빈약함(thinness)을 포함할 수도 있다. 백그라운드 왜곡들은 히싱(hissing) 및 가변성(variability)을 포함할 수도 있다.

[0011] 수정된 단일-채널 스피치 신호는 원본 스피치 신호(original speech signal)에 기초할 수도 있다. 다중 객관적 왜곡들은 원본 스피치 신호 없이 추정될 수도 있다. 원본 스피치 신호가 획득될 수도 있고, 다중 객관적 왜곡들은 원본 스피치 신호에 추가로 기초할 수도 있다.

[0012] 스피치 신호 품질을 측정하기 위한 전자 디바이스가 또한 설명된다. 전자 디바이스는 수정된 단일-채널 스피치 신호에 기초하여 다중 객관적 왜곡들을 추정하는 연산 왜곡 추정기(computational distortion estimator)를 포함한다. 다중 객관적 왜곡들은 적어도 하나의 포어그라운드 왜곡 및 적어도 하나의 백그라운드 왜곡을 포함한다. 전자 디바이스는 또한, 연산 왜곡 추정기에 결합된 연산 포어그라운드 품질 추정기(computational foreground quality estimator)를 포함한다. 연산 포어그라운드 품질 추정기는 다중 객관적 왜곡들에 기초하여 포어그라운드 품질을 추정한다. 전자 디바이스는 연산 왜곡 추정기에 결합된 연산 백그라운드 품질 추정기(computational background quality estimator)를 더 포함한다. 연산 백그라운드 품질 추정기는 다중 객관적 왜곡들에 기초하여 백그라운드 품질을 추정한다. 전자 디바이스는 연산 왜곡 추정기에 결합된 연산 전체 품질 추정기(computational overall quality estimator)를 더 포함한다. 연산 전체 품질 추정기는 포어그라운드 품질 및 백그라운드 품질에 기초하여 전체 품질을 추정된다.

[0013] 스피치 신호 품질을 측정하기 위한 컴퓨터-프로그램 제품이 또한 설명된다. 컴퓨터-프로그램 제품은 명령들을 갖는 비-일시적인 유형의 컴퓨터-판독 가능한 매체(non-transitory tangible computer-readable medium)를 포함한다. 명령들은 전자 디바이스로 하여금 수정된 단일-채널 스피치 신호를 획득하게 하기 위한 코드를 포함한다. 명령들은 또한, 전자 디바이스로 하여금 수정된 단일-채널 스피치 신호에 기초하여 다중 객관적 왜곡들을 추정하게 하기 위한 코드를 포함한다. 다중 객관적 왜곡들은 적어도 하나의 포어그라운드 왜곡 및 적어도 하나의 백그라운드 왜곡을 포함한다. 명령들은 또한, 전자 디바이스로 하여금 다중 객관적 왜곡들에 기초하여 포어그라운드 품질 및 백그라운드 품질을 추정하게 하기 위한 코드를 포함한다. 명령들은 전자 디바이스로 하여금 포어그라운드 품질 및 백그라운드 품질에 기초하여 전체 품질을 추정하게 하기 위한 코드를 더 포함한다.

[0014] 스피치 신호 품질을 측정하기 위한 장치가 또한 설명된다. 장치는 수정된 단일-채널 스피치 신호를 획득하기 위한 수단을 포함한다. 장치는 또한, 수정된 단일-채널 스피치 신호에 기초하여 다중 객관적 왜곡들을 추정하기 위한 수단을 포함한다. 다중 객관적 왜곡들은 적어도 하나의 포어그라운드 왜곡 및 적어도 하나의 백그라운드 왜곡을 포함한다. 장치는 다중 객관적 왜곡들에 기초하여 포어그라운드 품질 및 백그라운드 품질을 추정하기 위한 수단을 더 포함한다. 장치는 추가적으로, 포어그라운드 품질 및 백그라운드 품질에 기초하여 전체 품질을 추정하기 위한 수단을 포함한다.

도면의 간단한 설명

[0015] 도 1 은 주관적 방법을 이용한 품질 분석을 예시하고;

도 2 는 스피치 신호의 품질의 객관적 결정을 위한 2 개의 방법들을 예시하는 블록도이고;

도 3 은 지각적인 객관적 청취 품질 평가(perceptual objective listening quality assessment; P.O.LQA)의 현재의 ITU 표준을 이용하여 MOS를 예측하는 정확도를 예시하는 그래프이고;

도 4 는 스피치 신호 품질을 측정하기 위한 방법의 하나의 구성을 예시하는 흐름도이고;

도 5 는 스피치 신호 품질을 측정하기 위해 구성된 전자 디바이스의 하나의 구성을 예시하는 블록도이고;

도 6 은 본원에서 개시된 시스템들 및 방법들에 따라 구현될 수도 있는 계층적 구조의 하나의 예를 예시하는 블록도이고;

도 7 은 본원에서 개시된 시스템들 및 방법들에 따라 구현될 수도 있는 계층적 구조의 또 다른 예를 예시하는 블록도이고;

도 8 은 추출된 특징들 및 트레이닝 스피치 샘플들에 기초한 주관적 품질 측정 점수들의 하나의 단일 차원의 예측을 예시하는 블록도이고;

도 9 는 도 9a 내지 도 9c 에 대응하는 예의 다수의 부분들을 예시하는 블록도이고;

도 9a 는 특징 행렬 (feature matrix) 을 예시하는 블록도이고;

도 9b 는 추가적인 특징들을 예시하는 블록도이고;

도 9c 는 가중화 적용을 예시하는 블록도이고;

도 10 은 스피치 신호 품질을 측정하기 위해 구현될 수도 있는 계층적 구조의 하나의 예를 예시하는 흐름도이고;

도 11 은 스피치 신호 품질을 측정하기 위해 구성된 네트워크 디바이스 및 2 개의 전자 디바이스들의 하나의 구성을 예시하는 블록도이고;

도 12 는 스피치 신호 품질을 측정하기 위해 구성된 전자 디바이스의 하나의 구성을 예시하는 블록도이고;

도 13 은 스피치 신호 품질을 측정하기 위하여 전자 디바이스에 의해 구현된 방법을 예시하는 흐름도이고;

도 14 는 스피치 신호 품질을 측정하기 위하여 네트워크 디바이스에 의해 구현된 방법을 예시하는 흐름도이고;

도 15 는 특징 추출을 위한 전자 디바이스의 하나의 구성을 예시하는 블록도이고;

도 16 은 특징 추출을 위한 방법의 하나의 구성을 예시하는 흐름도이고;

도 17 은 생리학적 달팽이관 모델 출력 (physiological cochlear model output) 의 하나의 예를 예시하는 그림이고;

도 18 은 침투적 특징 추출 모듈의 하나의 구성을 예시하는 블록도이고;

도 19 는 비-침투적 특징 추출 모듈의 하나의 구성을 예시하는 블록도이고;

도 20 은 섹션들 A-D 로 분할되는 달팽이관 모델의 출력의 하나의 예를 예시하는 블록도이고;

도 21 은 달팽이관 모델 출력들의 섹션들로부터 장소-기반 분석 벡터를 추출하는 하나의 예를 예시하는 블록도이고;

도 22 는 달팽이관 모델 출력의 섹션 A로부터 장소-기반 분석 벡터를 추출하는 또 다른 예를 예시하는 블록도이고;

도 23 은 달팽이관 모델 출력들의 섹션들로부터 시간-기반 분석 벡터를 추출하는 하나의 예를 예시하는 블록도이고;

도 24 는 달팽이관 모델 출력의 섹션 A로부터 시간-기반 분석 벡터를 추출하는 또 다른 예를 예시하는 블록도이고;

도 25 는 특징 결정 모듈의 구성들을 예시하는 블록도를 포함하고;

도 26 은 특징 결정의 예를 예시하고;

도 27 은 특징 결정의 또 다른 예를 예시하고;

도 28 은 본원에서 개시된 시스템들 및 방법들에 따라 객관적 예측의 하나의 예를 예시하고;

도 29 는 본원에서 개시된 시스템들 및 방법들에 따라 객관적 예측의 또 다른 예를 예시하고;

도 30 은 본원에서 개시된 시스템들 및 방법들에 따라 객관적 예측의 또 다른 예를 예시하고;

도 31 은 도 31a 내지 도 31d 에 대응하는 예의 다수의 부분들을 예시하는 블록도이고;

도 31a 는 침투적 특징 추출을 이용하여 러프니스 왜곡을 추정하거나 예측하는 하나의 예의 부분을 예시하는 블록도이고;

도 31b 는 침투적 특징 추출을 이용하여 러프니스 왜곡을 추정하거나 예측하는 하나의 예의 또 다른 부분을 예시하는 블록도이고;

도 31c 는 침투적 특징 추출을 이용하여 러프니스 왜곡을 추정하거나 예측하는 하나의 예의 또 다른 부분을 예시하는 블록도이고;

도 31d 는 침투적 특징 추출을 이용하여 러프니스 왜곡을 추정하거나 예측하는 하나의 예의 또 다른 부분을 예시하는 블록도이고;

도 32 는 도 32a 내지 도 32d 에 대응하는 예의 다수의 부분들을 예시하는 블록도이고;

도 32a 는 비-침투적 특징 추출을 이용하여 러프니스 왜곡을 추정하거나 예측하는 하나의 예의 부분을 예시하는 블록도이고;

도 32b 는 비-침투적 특징 추출을 이용하여 러프니스 왜곡을 추정하거나 예측하는 하나의 예의 또 다른 부분을 예시하는 블록도이고;

도 32c 는 비-침투적 특징 추출을 이용하여 러프니스 왜곡을 추정하거나 예측하는 하나의 예의 또 다른 부분을 예시하는 블록도이고;

도 32d 는 비-침투적 특징 추출을 이용하여 러프니스 왜곡을 추정하거나 예측하는 하나의 예의 또 다른 부분을 예시하는 블록도이고;

도 33 은 특징 추출을 위한 방법의 더욱 구체적인 구성을 예시하는 흐름도이고;

도 34 는 침투적 특징 추출을 위한 방법의 구성을 예시하는 흐름도이고;

도 35 는 비-침투적 특징 추출을 위한 방법의 구성을 예시하는 흐름도이고;

도 36 은 침투적 특징 추출의 더욱 구체적인 구성을 예시하는 블록도이고;

도 37 은 스피치 신호 품질을 측정하고 및/또는 특정 추출을 위한 시스템들 및 방법들이 구현될 수도 있는 무선 통신 디바이스의 하나의 구성을 예시하는 블록도이고; 그리고

도 38 은 전자 디바이스/무선 디바이스 내에 포함될 수도 있는 어떤 컴포넌트들을 예시한다.

발명을 실시하기 위한 구체적인 내용

[0016] 본원에서 개시된 시스템들 및 방법들의 일부의 구성들은 스피치 왜곡들 및 전체 품질을 진단하는 계층적 시스템을 사용한다. 스피치 신호들의 품질은 인간 청자 (human listener) 들에 의해 가장 정확하게 측정된다.

그러나, 다수의 청자들로 이러한 평가를 착수하는 것은 시간 및 돈의 양자의 측면에서 고비용이다. 이에 따라, (스피치 품질에 관한) 인간 청자 응답들을 정확하게 복제할 수 있는 시스템들 및 방법들 (예컨대, 알고리즘들, 연산 시스템들, 등) 이 유의할 것이다. 예를 들어, 이 시스템들 및 방법들은 다수의 인간 청자들에 대한 요건을 대체할 수 있다.

[0017] 본원에서 개시된 시스템들 및 방법들은, 인간 달팽이관의 유체-역학적 기능, 청각 신경으로의 그 기계적 신경지배, 및 내부/외부 모세포 (hair cell) 들의 기능을 에뮬레이팅하는 연산 엔진 (computational engine) 을 제공한다. (연산 엔진으로부터의) 내유모세포 (inner hair cell) 들의 출력은 인간 스피치 지각의 다양한 양태들을 추론하기 위해 이용될 수 있다. 인간 청취 프로세스는 고도로 복잡하여, 달팽이관에서의 유체역학들을, 상부 청각 경로 및 중추 신경계 (Central Nervous System; CNS) 에서 전기적으로 변환된 신호들을 신경 프로세싱하는 것 (달팽이관-이후 기능) 에 관련시킨다. 전체 품질을 (단일-차원 수량으로서) 예측하는 것을 시도하였던 알려진 접근법들은 광범위한 왜곡들을 커버하거나 높은 정확도를 달성하는 것에 실패하였다. 예를 들어, 현재의 국제 전기통신 연합 (International Telecommunications Union; ITU) 표준의 지각적인 객관적 청취 품질 평가 (P.OLQA) (평균 의견 점수 (mean opinion score; MOS) 을 예측하는 P.863) 는, P.OLQA 가 설계되지 않았던 시스템들에 의해 열화된 스피치의 품질을 추정할 때에 정확하지 않다.

[0018] 구체적으로, 본 시스템들 및 방법들은 인간 달팽이관의 연산 모델의 출력으로부터 특징들을 추출할 수도 있다.

다음으로, 특징들은 스피치 품질을 예측하기 위하여 추가로 분석될 수도 있다. 본원에서 설명된 특징들은 스피치 품질을 예측하는 상황에서 논의되지만, 특징들은 임의의 적당한 애플리케이션에서 이용될 수도 있고 스피치 품질 예측으로 제한되지 않는다. 이용된 달팽이관의 연산 모델은 스피치 신호들을 내유모세포 응답들로 정확하게 변환하는 유체-역학적인 생리학적 모델이다. 구체적으로, 기능적 달팽이관 모델은 달팽이관의 입력 및 출력에서 보일 수도 있고, 달팽이관이 어떻게 작동하는지를 이해하지 않고도 달팽이관의 기능을 에뮬레이팅하는 것을 시도할 수도 있다. 대조적으로, 본 시스템들 및 방법들의 생리학적 달팽이관 모델은 달

팽이관 (예컨대, 고막 (eardrum), 외이도 (ear canal), 달팽이관 내부의 유체, 다양한 막들 상의 힘들/변위/속도, 내유모세포들 - 그리고 질량, 탄성 및/또는 마찰 특성들과 같은 이 구조들의 생리학적 성질들을 감안함)의 구체적인 생리학적 응답들을 모방하는 것을 시도하고, 여기서, 모델링은 기본적인 물리학을 통해 수행되고, 출력을 생성하기 위하여 Heinz, Baumgarte, Meddis 또는 Zwicker 의 저작물에서의 같은 전자 회로들을 이용하여 모델링하는 것과 같은 근사화들을 이용하는 것으로 제한되지 않는다. 따라서, 스피치 품질을 예측하기 위하여 생리학적 달팽이관 모델을 적용하는 것은 본원에서 개시된 시스템들 및 방법들의 하나의 고유한 특징이다.

[0019] 본원에서 개시된 시스템들 및 방법들의 일부의 구성들은 "분할 및 정복 (divide and conquer" 전략을 사용한다. 예를 들어, 본원에서 개시된 시스템들 및 방법들은 먼저, 스피치 품질의 개별적인 속성들/차원들 (예컨대, 왜곡들)을 진단하는 것을 시도하고, 다음으로, 신호의 전체 품질을 예측하기 위하여 이 개별적인 속성들을 조합한다. 예를 들어, 계층적 접근법이 사용될 수도 있고, 여기서, 다중 왜곡 진단들이 먼저 수행되고, 그 다음으로, 포어그라운드 및 백그라운드 왜곡 진단들이 수행되고, 마지막으로, 전체 품질이 연산 모델로 추정된다.

[0020] 본원에서 개시된 시스템들 및 방법들은 스피치 품질의 객관적 측정을 제공한다. 전통적으로, 주관적 청취 테스트들은 평균 의견 점수 (MOS)를 결정하기 위하여 수행된다. 이 주관적 청취 테스트들은 비용 소비적이고, 시간 소비적일 수도 있으며, 반복가능하지 않을 수도 있다.

[0021] 스피치 품질의 객관적 측정은 주관적 청취 테스트들의 단점들을 극복할 수도 있다. 주관적 척도 (measure)들이 아마도 가장 정확하지만, 객관적 척도들은 주관적 척도들의 출력을 예측하기 위해 사용될 수도 있다.

[0022] 스피치 품질 테스팅의 객관적 측정 동안, 원본 스피치는 인코더 (예컨대, 강화된 가변 레이트 코덱 (Enhanced Variable Rate Codec; EVRC) 코덱)에 제공될 수도 있다. 음향심리학적 분석 (psychoacoustic analysis)은 원본 스피치 및 인코딩된 스피치에 대해 수행될 수도 있다. 결과들은 MOS를 추정하기 위하여 비교될 수도 있다. 이 접근법의 예들은 P.861, P.862 및 P.863와 같은 국제 전기통신 연합 (ITU) 표준들을 포함한다. 이를 중의 3개 모두는 ITU 표준 P.800에서 설명된 바와 같은 주관적 MOS 테스팅의 결과들을 예측하는 것을 시도한다.

[0023] 잡음은 원본 스피치 내로 주입될 수도 있다. 잡음-피해받은 스피치는 강화 알고리즘 (예컨대, 잡음 억압기)에 제공될 수도 있다. 음향심리학적 분석은 원본 스피치 및 강화된 스피치 (예컨대, 비-억압된 스피치)에 대해 수행될 수도 있다. 결과들은 MOS를 추정하기 위하여 비교될 수도 있다. 이 접근법의 하나의 예는 유럽 전기통신 표준 협회 (European Telecommunications Standards Institute; ETSI)에 따라 제공되고, 이는 ITU 표준 P.835에서 설명된 바와 같은 주관적 테스팅의 결과들을 예측하는 것을 시도한다.

[0024] 스피치 품질의 객관적 측정에 대한 알려진 접근법들은 1 차원일 수도 있다. 그러나, 스피치 품질은 다차원이다. 하나의 차원은 적합하거나 정확하지 않을 수도 있다. 예를 들어, 현재의 ITU 표준 P.OLQA (MOS를 예측하는 P.863)는 완전히 정확하지 않다. 그것은 EVRC 코덱들에 대하여 MOS를 부정확하게 예측할 수도 있다. 이것의 하나의 예는 ITU Q9 WP2 Contribution COM12-C270rev1-E (201110)이다. 본원에서 개시된 시스템들 및 방법들은 알려진 접근법들에 비해 MOS를 예측함에 있어서 더 높은 정확도를 제공할 수도 있다. 일부의 구성들에서, 본원에서 개시된 시스템들 및 방법들은 예상된 ITU 표준들 (예컨대, P.TCA, P.ONRA 및 P.AMD) 내로 편입될 수도 있다. 표 1은 스피치 품질을 측정하기 위한 일부의 접근법들 (예컨대, 표준화되고 진행 중임)을 예시한다. 본원에서 개시된 시스템들 및 방법들은 표준화가 진행 중인 표준들에 적용될 수도 있다.

표 1

주관적	객관적	주석
MOS	PSQM (지각적 스피치 품질 척도) PESQ (스피치 품질의 지각적 평가) P.OLOA	1 차원만: 전체 품질
P.835	P.ONRA (지각적인 객관적 청취 품질 평가 (진행 중))	3 차원들: 신호/백그라운드/전체 품질들
P.MULTI	P.AMD (다차원 분석을 위한 지각적 접근법들) (진행 중)	4 포어그라운드 / 2 백그라운드 품질들

[0025]

[0026] 본원에서 개시된 시스템들 및 방법들은 1 차원 스피치 신호 품질 테스팅을, 다중 독립적인 왜곡들을 검출하기 위한 능력으로 대체하기 위하여 적용될 수도 있다. 일단 다차원 '검출가능성 (detectability)'이 달성되면, 전체 품질은 이에 따라 예측될 수도 있다. 본원에서 개시된 시스템들 및 방법들은 "분할 및 정복" 전략을 채용할 수도 있다. 예를 들어, 신규의 프레임워크 및 특징들 (예컨대, 진보된 지각적 특징들)이 제공된다.

[0027]

계층적 시스템 또는 구조는 스피치 왜곡들 및 전체 품질을 진단하기 위해 사용될 수도 있다. 이것은 전체 품질에 대한 예측의 개선된 정확도로 귀착될 수도 있다. 추가적인 장점들은 스피치 신호들에 대한 진단 정보를 제공하는 것, 네트워크 불안들에 대한 진단 정보를 잠재적으로 제공하는 것, 코딩/강화 알고리즘들을 개발하기 위한 가능한 툴을 제공하는 것, 및 품질에 대한 실시간 피드백을 제공하는 것을 포함한다.

[0028]

다양한 구성들은 도면들을 참조하여 지금부터 설명되며, 여기서, 유사한 참조 번호들은 기능적으로 유사한 엘리먼트들을 표시할 수도 있다. 본원의 도면들에서 일반적으로 설명되고 예시된 바와 같은 시스템들 및 방법들은 폭넓게 다양한 상이한 구성들로 배치되고 설계될 수 있다. 이에 따라, 도면들에서 나타낸 바와 같이, 몇몇 구성들의 다음의 더욱 상세한 설명은 청구된 바와 같이 범위를 제한하도록 의도된 것이 아니라, 단지 시스템들 및 방법들을 대표한다.

[0029]

도 1은 주관적 방법을 이용한 스피치 신호 품질 분석을 예시한다. 주관적 방법은 인간 청자들 (104a 내지 104c)의 지각들에 기초하여 하나 이상의 스피치 신호들의 품질을 측정하기 위한 접근법이다. 주관적 모델은 하나 이상의 스피치 신호들을 하나 이상의 인간 청자들 (104a 내지 104c)에게 제공하는 스피치 신호 소스 (speech signal source; 102)를 포함할 수도 있다. 하나 이상의 인간 청자들 (104a 내지 104c)은 하나 이상의 스피치 신호들을 청취할 수도 있고, 인간 청자 (104a 내지 104c)의 품질의 지각들에 기초하여 하나 이상의 스피치 신호들의 품질을 등급부여할 수도 있다. 예를 들어, 인간 청자 (104a)는 스피치 신호의 품질이 양호한 것 (106a)으로 판단할 수도 있고; 또 다른 인간 청자 (104b)는 동일한 스피치 신호의 품질이 열악한 것 (106b)으로 판단할 수도 있고; 제 3 인간 청자 (104c)는 동일한 스피치 신호의 품질이 타당한 것 (106c)으로 판단할 수도 있다. 주관적 테스팅에서는, 3 개의 등급들이 평균화 프로그램 (108)에 의해 평균화될 것이고, 타당함의 평균 의견 점수 (MOS) (110)를 부여받는다.

[0030]

주관적 방법은 이 방법이 하나 이상의 인간 청자들 (104a 내지 104c) 및 테스터 (tester)의 양자에 대해 시간 소비적인 것으로 인해 일관적으로 이용하기가 어렵다. 이 주관적 방법을 이용하는 것에 대한 또 다른 잠재적인 부정적인 면은 주관적 방법이 고비용일 수도 있다는 것이다. 이 주관적 방법은 인간 청자들 (104a 내지 104c)이 그 시간을 쓰게 할 수도 있고; 그것은 또한, 결과들이 더욱 일관적이도록 하기 위하여 전용 구역을 필요로 할 수도 있다. 주관적 방법은 또한, 반복가능한 결과들을 획득하기가 어려울 수도 있다. 인간 청자들 (104a 내지 104c)이 스피치 신호의 품질을 결정할 때에 급격한 차이들을 가질 수도 있기 때문이다.

예를 들어, 동일한 스피치 신호는 2 개의 상이한 주관적 방법 테스트들에서 이용될 수도 있다. 제 1 주관적 방법 테스트는 스피치 신호가 타당함의 스피치 신호 품질 등급을 가지는 것으로 귀착될 수도 있고, 제 2 주관적 방법 테스트는 동일한 스피치 신호가 양호함의 스피치 신호 품질 등급을 가지는 것으로 귀착될 수도 있다.

[0031] 알고리즘들 또는 연산 시스템들은 인간 청자 (104a 내지 104c) 의 응답을 정확하게 복제할 수 있을 수도 있다.

이것은 테스팅에 소비된 돈을 감소시킬 수도 있다. 예를 들어, 인간 청자 (104a 내지 104c) 의 스피치의 지각을 시뮬레이팅할 수 있는 연산 엔진은 생리학적 달팽이관 모델일 수 있다. 생리학적 달팽이관 모델은 본원에서 설명된 방법들을 이용하여 스피치 신호의 전체 스피치 품질을 예측할 수 있다. 생리학적 달팽이관 모델은 달팽이관에서의 유체역학들과, 인간 귀의 상부 청각 경로 및 중추 신경계에서 행해지는 신경 프로세싱을 복제할 수 있다. 생리학적 달팽이관 모델은 스피치 품질의 개별적인 차원들을 진단할 수 있고, 그 다음으로, 스피치 신호의 전체 스피치 품질을 더욱 정확하게 예측하기 위하여 이 개별적인 속성들을 조합할 수 있다. 스피치 신호 품질 분석을 위한 생리학적 달팽이관 모델의 이용은 전체 품질의 예측을 개선시킬 수도 있고, 스피치 신호에 대한 진단 정보를 제공할 수도 있고, 실시간 품질 피드백을 이용하여 강화 알고리즘들을 위한 가능한 툴을 제공할 수도 있다.

[0032] 도 2 는 스피치 신호의 품질의 주관적 결정을 연산적으로 예측함으로써, 객관적 점수를 생성하기 위한 2 개의 방법들을 예시하는 블록도이다. 원본 스피치 신호 (212a 내지 212b) 는 예를 들어, 코더 (214) 및/또는 강화 (222) 를 통과한 후에 왜곡된 스피치 신호가 될 수도 있다. P.835 는 구체적으로, 왜곡되었고 그 다음으로, 다양한 알고리즘들을 이용하여 추후에 강화되었던 신호들에 대한 ITU 주관적 테스팅 표준이다. P.800 은, 범위에 있어서 P.835 보다 더욱 일반적이고 임의의 왜곡된 신호들에 대해 유효한 ITU 주관적 테스팅 표준이다. 도 2 에서 도시된 2 개의 방법들은 인간의 청각 지각을 시뮬레이팅하는 것에 기초하여 (코딩 및/또는 강화 (214, 222) 후의) 왜곡된 스피치 신호의 품질의 객관적 점수를 제공할 수도 있다. 도시된 2 개의 방법들은 MOS (P.800) 및 P.835 주관적 테스팅 방법론들을 이용한 주관적 평가들의 연산 예측들 (객관적 점수들) 을 산출한다. 방법 P.800 은 전체 점수화된 MOS 를 산출하고, 방법 P.835 는 3 개의 점수들을 산출한다: S-MOS, N-MOS 및 G-MOS. S-MOS 는 스피치 신호의 포어그라운드 품질에 대한 점수이고, N-MOS 는 스피치 신호의 백그라운드 잡음에 대한 점수이고, G-MOS 는 스피치 신호의 전체 품질에 대한 점수이다. 양자의 테스팅 방법들은 침투적 또는 비-침투적 방법을 이용할 수도 있다.

[0033] 명확함을 위하여, 주관적 스피치 품질의 ITU 표준에 대한 일부의 설명은 다음과 같이 주어진다: 스피치 신호에 대하여, P.800 은 MOS 로 칭해지는 하나의 OVERALL 점수만을 산출한다. P.835 는 3 개의 점수들을 산출한다: S-MOS, N-MOS 및 G-MOS. S-MOS 는 포어그라운드/신호만에 대한 점수이고 N-MOS 는 백그라운드/잡음 품질만에 대한 점수이다. G-MOS 는, P.800 에서의 MOS 와 실제적으로 동일한 전체 품질에 대한 점수이다. (2014 년 초기에 ITU 표준으로 완결되었던) P.806 은 8 개의 점수들을 산출한다.

[0034] 도 2 에서의 방법들은 (예를 들어, 코딩 및/또는 강화 (214, 222) 후의) 단지 왜곡된 스피치 신호, 또는 (코더 (214) 또는 강화 모듈 (222) 에 의해 출력된) 왜곡된/수정된 스피치 신호 및 (비교하기 위한 기준으로서 이용된) 원본 스피치 신호 (212a 내지 212b) 의 조합을 이용하여 행해질 수도 있다. (원본 스피치 신호 (212a 내지 212b) 가 존재하지 않을 때의) 전자는 비-침투적 방법인 반면, (수정된/왜곡된 신호 및 원본 스피치 신호 (212a 내지 212b) 의 양자가 존재할 때의) 후자는 침투적 방법이다. 수정된 스피치 신호는 프로세싱을 포함하는, 원본 스피치 (212a 내지 212b) 신호에 대한 변경들을 포함할 수도 있다. 프로세싱된 스피치 신호는 수정된 스피치 신호의 하나의 예일 수도 있다.

[0035] MOS 점수를 예측하는 방법은 원본 스피치 (212a) 신호를 획득할 수도 있다. 원본 스피치 (212a) 신호는 인간의 지각을 시뮬레이팅하는 것을 시도하였던 음향심리학적 분석 (216a) 을 위한 입력으로서 이용될 수도 있다.

원본 스피치 (212a) 신호는 또한, 무선 네트워크에 의해 행해질 수도 있는 신호 압축 및 다른 타입들의 신호 프로세싱을 시뮬레이팅하기 위하여 코더 (214) 에 대한 입력으로서 이용될 수도 있다. 코더 (214) 는 수정된 스피치 신호를 출력할 수도 있고, MOS 점수를 예측하는 방법이 테스팅의 비-침투적 방법을 포함하도록 한다.

수정된 스피치 신호는 인간의 지각을 시뮬레이팅하는 것을 시도하는 음향심리학적 분석 (216b) 을 위한 입력으로서 이용될 수도 있다.

[0036] 다음으로, 원본 스피치 신호 및 수정된 스피치 신호의 음향심리학적 분석 (216a 내지 216b) 으로부터의 결과들이 비교될 수도 있다 (218a). 원본 스피치 신호 및 수정된 스피치 신호의 비교는 전체 MOS (220a) 를 제공할 수도 있다. MOS 점수화는 1 로부터 5 까지의 범위일 수도 있다. 1 의 점수는 "나쁨 (bad)" 의 품질을 가질 수도 있고; 2 는 "열악함 (poor)" 의 품질을 가질 수도 있고; 3 은 "타당함 (fair)" 의 품질을 가질 수

도 있고; 4 는 "양호함 (good)" 의 품질을 가질 수도 있고; 5 는 "우수함 (excellent)" 의 품질을 가질 수도 있다.

[0037] 잡음 억압 알고리즘들을 포함하는 스피치 신호들을 평가하기 위한 객관적 테스트 방법론은 또한, 분석을 위하여 원본 스피치 (212b) 신호 및 잡음 (224) 을 획득할 수도 있다. 이 방법은 S-MOS, N-MOS 및 G-MOS 점수들을 포함할 수도 있는 P.835 점수들을 산출할 수도 있다. 원본 스피치 (212b) 신호 및 잡음 (224) 은 인간의 청각 지각을 시뮬레이팅하는 것을 시도하는 음향심리학적 분석 (216a) 을 위한 입력으로서 이용될 수도 있다. 원본 스피치 (212b) 신호 및 잡음 (224) 은 강화 모듈 (222) 에 대한 입력으로서 이용될 수도 있다. 예를 들어, 강화 모듈 (222) 은 원본 스피치 (212b) 신호에서 있을 수도 있는 잡음 또는 에코 (echo) 를 감소시킬 수도 있다. 강화 모듈 (222) 은 수정된 스피치 신호를 출력할 수도 있고, P.835 예측 방법이 테스팅의 비-침투적 방법을 포함하도록 한다. 수정된 스피치 신호는 인간의 청각 지각을 시뮬레이팅하는 것을 시도하는 음향심리학적 분석 (216d) 을 위한 입력으로서 이용될 수도 있다.

[0038] 다음으로, 원본 스피치 (212b) 신호 및 수정된 스피치 신호의 음향심리학적 분석 (216c 내지 216d) 으로부터의 결과들이 비교될 수도 있다 (218b). 원본 스피치 (212b) 신호 및 수정된 스피치 신호의 비교는 P.835 (220b) 테스트의 품질 등급들 중의 하나 이상을 제공할 수도 있다. P.835 점수화는 3 개의 등급들을 가질 수도 있다; 포어그라운드 스피치 품질에 대한 S-MOS, 백그라운드 잡음 품질에 대한 N-MOS, 및 전체 스피치 품질에 대한 G-MOS. 모두 3 개의 등급들은 1 부터 5 까지의 범위일 수도 있다. S-MOS 에 대한 1 의 점수는 "매우 왜곡됨 (very distorted)" 의 등급을 가질 수도 있고; 2 는 "상당히 왜곡됨 (fairly distorted)" 의 등급을 가질 수도 있고; 3 은 "다소 왜곡됨 (somewhat distorted)" 의 등급을 가질 수도 있고; 4 는 "약간 왜곡됨 (slightly distorted)" 의 등급을 가질 수도 있고; 5 는 "왜곡되지 않음 (not distorted)" 의 등급을 가질 수도 있다. N-MOS 에 대한 1 의 점수는 "매우 침투적 (very intrusive)" 의 등급을 가질 수도 있고; 2 는 "다소 침투적 (somewhat intrusive)" 의 등급을 가질 수도 있고; 3 은 "현저하지만 침투적이지 않음 (noticeable but not intrusive)" 의 등급을 가질 수도 있고; 4 는 "약간 현저함 (slightly noticeable)" 의 등급을 가질 수도 있고; 5 는 "현저하지 않음 (not noticeable)" 의 등급을 가질 수도 있다. G-MOS 에 대한 1 의 점수는 "나쁨" 의 등급을 가질 수도 있고; 2 는 "열악함" 의 등급을 가질 수도 있고; 3 은 "타당함" 의 등급을 가질 수도 있고; 4 는 "양호함" 의 등급을 가질 수도 있고; 5 는 "우수함" 의 등급을 가질 수도 있다.

[0039] 도 3 은 지각적인 객관적 청취 품질 평가 (P.OLQA) 의 현재의 ITU 표준을 이용하여 MOS 를 예측하는 정확도를 예시하는 그래프이다. 그래프의 X-축은 스피치 신호의 MOS (319) 등급에 대응하고, Y 축은 스피치 신호의 P.OLQA (317) 등급에 대응한다. P.OLQA 는 디지털 분석을 통해 스피치 신호의 품질을 예측하기 위해 이용된 표준이다.

[0040] 도 3 에서의 그래프는 스피치 신호의 P.OLQA 점수를 P.800 점수 (MOS) 와 비교한다. P.OLQA 분석이 점수를 적절하게 예측할 경우, 그 점수는 스피치 신호의 MOS 와 일치해야 하거나 스피치 신호의 MOS 에 근접해야 한다. 그래프에서의 대각선은 P.OLQA 점수 및 MOS 가 근접할 경우에 포인트 (point) 가 어디에 있어야 하는지를 표시한다.

[0041] 도 3 에서의 그래프는 현재의 P.OLQA 분석이 스피치 신호들의 MOS 등급과 항상 비슷하지는 않다는 것을 표시한다. 전형적으로, 강화된 가변 레이트 CODEC (EVRC) 설정을 거친 스피치 신호의 P.OLQA 등급은 그래프에서 도시된 바와 같이 예측된 것의 아래에 있을 수도 있다. EVRC 설정된 스피치 신호 결과들은 마름모들 및 정사각형들로서 도시된다.

[0042] 도 4 는 스피치 신호 품질을 측정하기 위한 방법 (400) 의 하나의 구성을 예시하는 흐름도이다. 방법 (400) 은 전자 디바이스에 의해 수행될 수도 있다. 전자 디바이스의 예들은 이동 디바이스들 (예컨대, 스마트폰들, 셀룰러 전화들, 태블릿 디바이스들, 웹톱 컴퓨터들 등), 테스크톱 컴퓨터들, 개인 정보 단말들 (Personal Digital Assistants; PDA), 기기들, 텔레비전들, 게임용 시스템들 및 서버들 (예컨대, 네트워크 디바이스 내에 포함된 서버) 을 포함한다. 전자 디바이스는 원본 스피치 신호를 선택적으로 획득할 수도 있다 (402). 예를 들어, 전자 디바이스는 하나 이상의 마이크로폰들로 스피치 신호를 캡처할 수도 있거나, 또 다른 디바이스 (예컨대, 저장 디바이스, 컴퓨터, 전화, 헤드셋 등) 로부터 스피치 신호를 수신할 수도 있다. 원본 스피치 신호는 원시 (raw) 또는 수정되지 않은 (unmodified) 신호일 수도 있다. 예를 들어, 원본 스피치 신호는 인코딩 및/또는 압축되지 않은 스피치 신호의 전자적으로 샘플링된 버전일 수도 있다. 원본 스피치 신호가 획득되는 접근법들은 침투적 접근법들 또는 방법들로서 지칭될 수도 있다. 예를 들어, 침투적 접근법들은 원본 스피치 신호 및 수정된 스피치 신호 (예컨대, 왜곡된 스피치 신호) 사이의 명시적 감산

(explicit subtraction) 을 포함할 수도 있다. 다른 접근법들에서는, 원본 스피치 신호가 획득되지 않을 수도 있다. 이 접근법들은 비-침투적 접근법들 또는 방법들로서 지칭될 수도 있다. 비-침투적 접근법들에서는, 예를 들어, 원본 스피치 신호가 이용가능하지 않을 수도 있다 (예컨대, 수정된 스피치 신호 및 원본 스피치 신호 사이에 명시적 감산이 없을 수도 있음). 예를 들어, 방법 (400)은 일부의 구현예들 및/또는 시나리오들에서 원본 스피치 신호를 획득하지 않으면서 수행될 수도 있다. 구체적으로, '원본' 또는 '클린 (clean)' 신호가 비교를 위해 이용가능하지 않을 때, 특징들은 특히 품질을 예측하는 것에 순응할 수도 있다. 원본/클린 신호를 요구하지 않는 이러한 시스템들은 비-침투적 시스템들로서 알려져 있다. 위에서 설명된 바와 같이, 달팽이관의 연산 모델은 스피치 신호들을 내유모세포 응답들로 정확하게 변환하는 유체-역학적 모델일 수도 있다. 따라서, 비-침투적 특징들은 지각적 도메인에서의 스피치 및 오디오 신호 분석을 위한 달팽이관 모델로부터 추출될 수도 있다. 다른 유사한 모델들이 또한 이용될 수도 있다.

[0043] 일부의 구성들에서, 전자 디바이스는 수정된 스피치 신호를 획득하기 위하여 원본 스피치 신호를 프로세싱할 수도 있다. 예를 들어, 전자 디바이스는 수정된 스피치 신호를 획득하기 위하여 원본 스피치 신호를 인코딩할 수도 있다 (및/또는 예를 들어, 결과적인 인코딩된 스피치 신호를 디코딩함). 추가적으로 또는 대안적으로, 전자 디바이스는 수정된 스피치 신호를 획득하기 위하여 원본 스피치 신호를 강화할 수도 있다. 예를 들어, 전자 디바이스는 원본 스피치 신호의 잡음-주입된 버전에 대해 잡음 억압을 수행할 수도 있다. 전자 디바이스는 수정된 단일-채널 스피치 신호를 획득할 수도 있다 (404). "단일-채널" 스피치 신호는 스피치 신호를 임의의 하나의 시간에서의 (예를 들어, 압력의) 단일 샘플로서 나타낼 수도 있다. 예를 들어, 수정된 단일-채널 스피치 신호는 공간적으로 구별되는 정보를 제공하는 다중 신호들을 포함하지 않을 수도 있다. 일부의 구성들에서는, 예를 들어, 수정된 단일-채널 스피치 신호가 단일 마이크로폰으로 캡처되었던 원본 스피치 신호에 기초할 수도 있다.

[0044] 다른 구성들에서, 전자 디바이스는 또 다른 디바이스 (예컨대, 저장 디바이스, 컴퓨터, 전화, 헤드셋, 인코더 등)로부터 수정된 스피치 신호를 획득할 수도 있다. 예를 들어, 전자 디바이스는 네트워크화된 전자 디바이스로부터 수정된 스피치 신호를 수신할 수도 있다.

[0045] 수정된 스피치 신호는 원본 스피치 신호에 기초할 수도 있다. 예를 들어, 수정된 스피치 신호는 원본 스피치 신호의 왜곡된 버전일 수도 있다. 원본 스피치 신호의 강화, 인코딩, 디코딩, 잡음 억압, 신장 (expansion) 및 압축 (compression)은, 수정된 스피치 신호의 하나 이상의 특성들이 원본 스피치 신호의 특성들과는 상이하도록, 원본 스피치 신호를 왜곡시킬 수도 있다.

[0046] 일부의 구성들에서, 전자 디바이스는 본원에서 설명된 바와 같은 생리학적 모델 (예컨대, 생리학적 달팽이관 모델)을 이용하여 수정된 스피치 신호를 프로세싱할 수도 있다. 생리학적 모델을 이용하여 수정된 스피치 신호를 프로세싱하는 것은 (예를 들어, 기능적 모델보다) 인간 청각 기관에 의해 수행된 프로세싱을 더욱 정확하게 근사화할 수도 있다. 이것은 인간 청자에 의해 지각된 바와 같은 스피치 신호 품질을 더욱 정확하게 추정하는 것을 도울 수도 있다. 예를 들어, 생리학적 모델을 이용하여 수정된 스피치 신호를 프로세싱하는 것은 내유모세포 데이터를 제공할 수도 있다. 내유모세포 데이터 및 기저막 (basilar membrane) 데이터 사이의 차이는, 내유모세포 데이터가 기저막 데이터보다 인간 청각 기관에서의 더 이후의 데이터 수집 포인트로부터 수집되고, 그러므로, 기저막에서 발견되는 것을 초월하여 추가의 프로세싱을 거치게 된다는 것이다. 가장 기능적인 모델들은 기저막 응답을 주로 근사화하고, 약간의 다른 기능적인 모델들은 내유모세포 응답을 근사화하는 것을 계속하지만, 그러나 직렬 방식에서는, 이 기능적 모델들 중의 어느 것도 생리학적 구성요소들이 피드-포워드 (feed-forward) 및 피드-백 (feed-back) 기법에서 서로에 대해 종속적인 전체론적 방식으로 완전한 달팽이관 (생리학적 구조들, 유체 등) 기관을 취급하지는 않는다. 본원에서 개시된 시스템들 및 방법들은 내유모세포 데이터, 기저막 데이터 또는 양자의 데이터 타입들을 이용하여 적용될 수도 있다.

[0047] 전자 디바이스는 수정된 스피치 신호에 기초하여 다중 객관적 왜곡들을 추정할 수도 있다 (406). 예를 들어, 객관적 왜곡은 수정된 스피치 신호에서 왜곡들을 나타내는 하나 이상의 메트릭 (metric) 들 (예컨대, 원본 스피치 신호 및 수정된 스피치 신호 사이의 차이들) 일 수도 있다. 일부의 구현예들에서, 전자 디바이스는 추가적으로 원본 스피치 신호에 기초하여 다중 객관적 왜곡들을 추정할 수도 있다. 예를 들어, 다중 객관적 왜곡들 중의 하나 이상은 수정된 스피치 신호, 원본 스피치 신호, 또는 수정된 스피치 신호 및 원본 스피치 신호의 양자에 기초할 수도 있다. 다중 객관적 왜곡들은 예를 들어, 일부의 접근법들에서 별별로 추정될 수도 있다.

[0048] 일부의 구성들에서, 다중 객관적 왜곡들의 각각은 스피치 품질의 독립적인 차원을 나타낼 수도 있다. 예를

들어, 왜곡들의 각각은 인간 청자 피드백에 따라 서로 대체로 상관되지 않을 수도 있다. 예를 들어, 인간 청자들은 스피치 신호들에서 별도의 왜곡들을 검출할 수도 있다. 다시 말해서, 왜곡들을 설명하기 위한 많은 디스크립터 (descriptor) 들이 주어지면, 인간 청자들은 유사한 점수들을 스피치 품질의 상관된 차원들에 배정하는 경향이 있다. 하나의 예에서, 인간 청자들은 왜곡들을 설명하기 위하여 디스크립터들 "러프 (rough)", "하쉬 (harsh)", "플러터링 (fluttering)", "불연속적", "둔탁한 (dull)", "머플드 (muffled)", "떨어진 (distant)", "빈약한 (thin)", "히싱 (hissing)", "러싱 (rushing)", "로링 (roaring)", "버블링 (bubbling)", "간헐적 (intermittent)" 및 "가변적 (variable)" 을 부여받는다. 이 예에서, 인간 청자들은 유사한 점수들을 "러프" 및 "하쉬"; "플러터링" 및 "불연속적"; "둔탁한" 및 "머플드"; "떨어진" 및 "빈약한"; "히싱", "러싱" 및 "로링"; 그리고 "버블링", "간헐적" 및 "가변적"에 배정하는 경향이 있다. 따라서, 이 디스크립터들은 스피치 품질의 상관되지 않거나 독립적인 차원들로 그룹화될 수 있다.

[0049] 일부의 구성들에서, "러프니스" 왜곡은 "러프" 및 "하쉬" 디스크립터들에 대응할 수도 있다. 스피치 신호의 러프니스는 스피치 신호의 빠른 시간적으로 국소화된 열화 (temporally localized degradation) 일 수도 있다.

러프니스는 시간적 포어그라운드 왜곡일 수도 있다. "불연속성" 왜곡은 "플러터링" 및 "불연속적" 디스크립터들에 대응할 수도 있다. 스피치 신호의 불연속성은 스피치 신호의 느린 시간적으로 국소화된 열화일 수도 있다. 불연속성은 시간적 포어그라운드 왜곡일 수도 있다. "둔탁성" 왜곡은 "둔탁한" 및 "머플드" 디스크립터들에 대응할 수도 있다. 스피치 신호의 둔탁성은 스피치 신호의 고주파수 대역에서의 열화일 수도 있다. 둔탁성은 스펙트럼 (예컨대, 주파수) 포어그라운드 왜곡일 수도 있다. "빈약함" 왜곡은 "떨어진" 및 "빈약한" 디스크립터들에 대응할 수도 있다. 스피치 신호의 빈약함은 스피치 신호의 저주파수 대역에서의 열화일 수도 있다. 빈약함은 스펙트럼 포어그라운드 왜곡일 수도 있다. "히싱" 왜곡은 "히싱", "러싱" 및 "로링" 디스크립터들에 대응할 수도 있다. 스피치 신호의 히싱은 스피치 신호에서의 백그라운드 잡음의 레벨로 인한 열화일 수도 있다. "가변성" 왜곡은 "버블링", "간헐적" 및 "가변적" 디스크립터들에 대응할 수도 있다. 스피치 신호의 가변성은 스피치 신호에서의 잡음의 가변성으로 인한 열화일 수도 있다.

이 왜곡들의 각각은 스피치 품질의 독립적인 차원들을 나타낼 수도 있다. 다중 객관적 왜곡들의 각각은 등급부여될 수도 있고 점수를 제공받을 수도 있다. 점수는 다중 객관적 왜곡들의 각각이 스피치 신호에 야기시킬 수도 있는 왜곡의 레벨을 표시할 수도 있다. 다른 용어들은 스피치 품질의 독립적인 차원들을 위해 이용될 수도 있다는 것에 주목해야 한다. 예를 들어, ITU 표준 P.806은 다중 객관적 왜곡들에 밀접하게 대응하는 지각적 품질 (Perceptual Quality; PQ) 스케일을 제공한다. PQ 스케일은 S-FLT, S-RUF, S-LFC, S-HFC, B-LVL 및 B-VAR을 포함한다. S-FLT은 스피치 신호에서의 느리게 변동되는 열화 (slow-varying degradation)에 관련된다 (불연속성). S-RUF은 스피치 신호에서의 빠르게 변동되는 열화 (fast varying degradation)에 관련된다 (러프니스). S-LFC는 스피치 신호에서의 저주파수 컬러레이션 (coloration)의 열화에 관련된다 (둔탁성). S-HFC는 스피치 신호에서의 저주파수 컬러레이션 (coloration)의 열화에 관련된다 (빈약함). B-LVL은 스피치 신호에서의 백그라운드 잡음의 레벨로 인한 열화에 관련된다 (히싱). B-VAR은 스피치 신호에서의 백그라운드 잡음의 가변성으로 인한 열화에 관련된다 (가변성). 또한, P.806은 라우드니스 (loudness)에 대한 LOUD 점수를 포함하고 OVRL은 전체 품질에 대한 점수인 것이 주목될 수도 있다. OVRL은 P.800에서의 MOS 및 P.835에서의 G-MOS와 동일할 수도 있다.

[0050] 일부의 구성들에서, 다중 객관적 왜곡들은 적어도 하나의 포어그라운드 왜곡 및/또는 적어도 하나의 백그라운드 왜곡을 포함할 수도 있다. 예를 들어, 다중 객관적 왜곡들은 4 개의 포어그라운드 왜곡들 및 2 개의 백그라운드 왜곡들을 포함할 수도 있다. 4 개의 포어그라운드 왜곡들은 "러프니스", "불연속성", "둔탁성" 및 "빈약함"을 포함할 수도 있다. 2 개의 백그라운드 왜곡들은 스피치 신호에서의 스피치의 열화로 인한 스피치 신호의 왜곡들일 수도 있다. 백그라운드 왜곡들은 스피치 신호에서 발견된 잡음으로 인한 스피치 신호의 왜곡들일 수도 있다. 따라서, 본원에서 개시된 시스템들 및 방법들은 시간적 및 스펙트럼 왜곡들의 양자를 사용할 수도 있다.

[0051] 전자 디바이스는 다중 객관적 왜곡들에 기초하여 포어그라운드 품질 및 백그라운드 품질을 추정할 수도 있다 (408). 포어그라운드 왜곡들은 연산 포어그라운드 품질 추정기에 의해 수신될 수도 있고, 그 다음으로, 연산 포어그라운드 품질 추정기는 본원에서 설명된 방법들 및 절차들을 이용하여 스피치 신호의 포어그라운드 품질을 계산할 수도 있다. 포어그라운드 품질은, ITU 표준 P.835의 S-MOS 점수에 대응할 수도 있는, 포어그라운드 왜곡들에 대응하는 수정된 스피치 신호의 품질을 표시하는 메트릭이다. 백그라운드 왜곡들은 연산 백그라운드 품질 추정기에 의해 수신될 수도 있고, 그 다음으로, 연산 백그라운드 품질 추정기는 본원에서 설명된 방법들 및 절차들을 이용하여 스피치 신호의 백그라운드 품질을 계산할 수도 있다. 백그라운드 품질은, ITU P.835의 N-MOS 점수에 대응할 수도 있는, 백그라운드 왜곡들에 대응하는 수정된 스피치 신호의 품질을 표

시하는 메트릭이다.

[0052] 전자 디바이스는 포어그라운드 품질 및 백그라운드 품질에 기초하여 전체 품질을 추정할 수도 있다 (410). 예를 들어, 전자 디바이스는 본원에서 개시된 방법들 및 절차들을 이용하여 수정된 스피치 신호의 전체 품질을 추정하는 메트릭을 생성할 수도 있다. 이 전체 품질은 포어그라운드 품질 및 백그라운드 품질에 기초한 객관적 척도일 수도 있다. 이 전체 품질은 또한, 스피치 품질의 주관적 척도들을 근사화할 수도 있다. 예를 들어, 전체 품질은 주관적 청자들의 평균 의견 점수 (MOS)를 근사화할 수도 있다. 예를 들어, 전체 품질은 국제 전기통신 연합 (ITU) 표준들 P.800, 및/또는 P.835 및/또는 P.806에 따라 취해진 주관적 척도들을 근사화할 수도 있다.

[0053] 일부의 구성들에서, 전체 품질을 추정하는 것은 또한, 다중 객관적 왜곡들에 직접적으로 기초할 수도 있다. 예를 들어, 전자 디바이스는 포어그라운드 품질 및 백그라운드 품질에 추가하여, (예를 들어, 포어그라운드 품질 및 백그라운드 품질을 통해 다중 객관적 왜곡들에 오직 간접적으로 기초하는 것이 아니라) 다중 객관적 왜곡들에 직접적으로 기초하여 전체 품질을 추정할 수도 있다. 다시 말해서, 전체 품질은 포어그라운드 품질, 백그라운드 품질, 러프니스 점수, 불연속성 점수, 둔탁성 점수, 빈약함 점수, 허싱 점수 및 가변성 점수에 기초할 수도 있다. 다시 말해서, 전체 품질이 포어그라운드 품질, 백그라운드 품질, 러프니스 점수, 불연속성 점수, 둔탁성 점수, 빈약함 점수, 허싱 점수 및 가변성 점수에 기초할 경우, 전체 품질은 이 상이한 인자들에 대해 불확정적일 수도 있다.

[0054] 전체 품질은 하나 이상의 방법들로 적용될 수도 있다. 예를 들어, 전체 품질은 인코더, 인코더/디코더 (예컨대, 코덱 (codec)) 및/또는 잡음 억압기 (예컨대, 잡음 억압 알고리즘)의 품질을 평가하기 위해 사용될 수도 있다. 일부의 구성들에서, 전체 품질은 음성 품질을 개선시키기 위하여 통신 시스템들 (예컨대, 셀룰러 전화 네트워크들)에 대한 네트워크 및/또는 디바이스 세팅들을 자동으로 조절하기 위해 사용될 수도 있다. 전체 품질은 인코더들, 인코더들/디코더들 (예컨대, 코덱들) 및/또는 작업 억압기들 (예컨대, 잡음 억압 알고리즘들)을 설계하기 위해 사용될 수도 있다. 예를 들어, 전체 품질은 디코딩, 인코딩 및 잡음 억압을 비교하기 위한 어떤 동작 파라미터들을 테스트하기 위해 이용될 수도 있다.

[0055] 본원에서 개시된 시스템들 및 방법들은 주관적 척도들을 근사화하는 전체 품질의 객관적 척도를 제공할 수도 있다. 본원에서 개시된 시스템들 및 방법들의 이익들의 일부는 비용 감소를 포함한다. 예를 들어, 전체 품질은 더욱 신속하게, 그리고 인간 청자들 없이 추정될 수도 있다. 추가적으로, 본원에서의 전체 품질 추정들은 반복가능할 수도 있는 반면, 주관적 방법들은 일관적인 결과들을 제공하지 않을 수도 있다. 본원에서 개시된 시스템들 및 방법들에 따라 결정된 전체 품질 추정들은 다중 객관적 왜곡들 (예컨대, 스피치 품질의 다수의 차원들)에 기초할 수도 있는 반면, 다른 객관적 접근법들은 단일-차원이다. 따라서, 본원에서 개시된 시스템들 및 방법들에 따른 전체 품질 추정들은 다른 객관적 접근법들 (예컨대, ITU 표준 P.863 (P.OLQA)) 보다 더욱 정확하게 주관적 척도들 (예컨대, MOS)을 근사화할 수도 있다.

[0056] (도 4 와 관련하여 설명된 바와 같이) 스피치 신호 품질을 측정하는 것은 계층적 구조에 따라 수행될 수도 있다. 계층적 구조는 복수의 계층적 레벨들을 포함할 수도 있고, 여기서, 각각의 계층적 레벨은 임의의 그리고 각각의 상대적으로 더 낮은 계층적 레벨에 종속된다. 예를 들어, 복수의 계층적 레벨들은 3 개의 레벨들을 포함할 수도 있다. 제 1 계층적 레벨 (예컨대, 최저 계층적 레벨)은 다중 객관적 왜곡들을 추정하는 것을 포함할 수도 있다. 제 2 계층적 레벨은 포어그라운드 품질 및 백그라운드 품질을 추정하는 것을 포함할 수도 있다. 제 3 계층적 레벨 (예컨대, 최고 계층적 레벨)은 전체 품질을 추정하는 것을 포함할 수도 있다. 따라서, 제 2 계층적 레벨은 제 1 계층적 레벨에 기초할 수도 있고, 제 3 계층적 레벨은 제 2 계층적 레벨 및 (적어도 간접적으로) 제 1 계층적 레벨에 기초할 수도 있다. 이에 따라, 포어그라운드 품질 및 백그라운드 품질은 다중 객관적 왜곡들에 기초하여 추정되고, 전체 품질은 포어그라운드 품질 및 백그라운드 품질에 기초하여 추정된다.

[0057] 본원에서 개시된 시스템들 및 방법들에 따라 다수의 계층적 레벨들을 이용하는 것은 스피치 신호의 송신에 관한 문제들을 진단 및 해결하기 위한 능력에 있어서 개선사항들을 생성할 수도 있다. 예를 들어, 스피치 신호 품질이 수용가능할 경우, 전체 품질은 제공되고 검증하기 위한 데이터의 단일 포인트 (single point)이다. 그러나, 스피치 신호 품질이 수용가능하지 않을 경우, 스피치 신호는 포어그라운드 품질 데이터 및 백그라운드 품질 데이터를 이용하여 추가로 분석될 수 있다. 백그라운드 품질 데이터가 백그라운드 품질이 수용가능하지 않는다는 것을 표시할 경우, 즉시, 포어그라운드 품질에 대응하는 4 개의 다중 객관적 왜곡들이 배제된다. 다음으로, 스피치 신호는 백그라운드 품질에 대응하는 다중 객관적 왜곡들을 이용하여 추가로 분석될 수 있다.

[0058] 본원에서 개시된 시스템들 및 방법들에 의해 제공된 바와 같은 스피치 신호 품질을 측정하는 것 (및/또는 그 안에 포함된 하나 이상의 절차들)은 사실상 예측적일 수도 있다는 것에 주목해야 한다. 예를 들어, 전체 품질을 "추정하는 것"은 주관적 청취 테스트가 실제적으로 수행되었을 경우에 주관적 품질 척도 (예컨대, MOS)가 무엇이었는지를 예측하는 것을 포함할 수도 있다. 그러나, 청취 테스트들은 본원에서 개시된 시스템들 및 방법들에 따라 필요하지 않다는 것에 또한 주목해야 한다. 다시 말해서, 본원에서 개시된 시스템들 및 방법들은 객관적 결과들을 비교하기 위한 주관적 청취 테스트 없이 수행될 수도 있다.

[0059] 도 5 는 스피치 신호 품질을 측정하기 위해 구성된 전자 디바이스 (556)의 하나의 구성을 예시하는 블록도이다. 전자 디바이스는 도 4 와 관련하여 설명된 방법 (400)의 절차들 중의 하나 이상을 수행할 수 있다. 전자 디바이스 (556)는 선택적인 연산 포어그라운드 품질 추정기 (540) 및 선택적인 연산 백그라운드 품질 추정기 (542) 와 결합된 연산 왜곡 추정기 (526) 및/또는 연산 전체 품질 추정기 (552)를 포함한다. 하드웨어가 결합될 때, 컴포넌트들은 직접적으로 또는 간접적으로 중의 어느 하나로 접속된다. 예를 들어, 제 1 컴포넌트가 제 2 컴포넌트에 결합될 경우, 컴포넌트들은 직접 링크를 가질 수 있거나, 링크에서의 매개 컴포넌트들이 존재할 수도 있다. 전자 디바이스 (556) 및/또는 그 하나 이상의 컴포넌트들은 하드웨어 (예컨대, 회로부), 소프트웨어 또는 양자의 조합으로 구현될 수도 있다. 전자 디바이스 (556)의 예들은 이동 디바이스들 (예컨대, 스마트폰들, 셀룰러 전화들, 태블릿 디바이스들, 랩톱 컴퓨터들 등), 서버들 (예컨대, 네트워크 디바이스 내에 포함된 서버) 및 소프트-전화 디바이스들 (예컨대, 스카이프 (Skype), 페이스타임 (FaceTime) 등)을 포함한다.

[0060] 전자 디바이스 (556) (및/또는 예를 들어, 연산 왜곡 추정기 (526))는 수정된 단일-채널 스피치 신호 (524)를 획득할 수도 있다. 일부의 구현예들 및/또는 시나리오들에서, 전자 디바이스 (556)는 추가적으로 원본 스피치 신호 (512)를 획득할 수도 있다. 다른 구현예들 및/또는 시나리오들에서, 전자 디바이스 (556)는 원본 스피치 신호 (512)를 획득하지 않을 수도 있다. 이것은 도 4 와 관련하여 위에서 설명된 바와 같이 달성될 수도 있다.

[0061] 연산 왜곡 추정기 (526)는 수정된 단일-채널 스피치 신호 (524)에 기초하여 (및/또는 선택적으로 원본 스피치 신호 (512)에 기초하여) 다중 객관적 왜곡들을 추정할 수도 있다. 청자가 왜곡을 주관적으로 평가하지 않으므로, 왜곡들은 객관적이다. 일부의 구현예들 및/또는 시나리오들에서 (예컨대, 비-침투적 접근법들에서), 연산 왜곡 추정기 (526)는 원본 스피치 신호 (512) 없이 다중 객관적 왜곡들을 추정할 수도 있다. 이것은 도 4 와 관련하여 위에서 설명된 바와 같이 달성될 수도 있다. 예를 들어, 연산 왜곡 추정기 (526)는 수정된 단일-채널 스피치 신호 (524) 및/또는 원본 스피치 신호 (512)에 기초하여 러프니스 (528), 불연속성 (530), 둔탁성 (532), 빈약함 (534), 허싱 (536) 및/또는 가변성 (538)을 추정할 수도 있다. 연산 왜곡 추정기 (526)는 객관적 왜곡 추정들을 수행하기 위한 연산 알고리즘을 사용할 수도 있다. 주관적 인간 청자 (들)가 왜곡 추정들을 결정하지 않으므로, 이것은 "객관적" 일 수도 있다.

[0062] 일부의 구성들에서, 다중 객관적 왜곡들 중의 하나 이상은 연산 포어그라운드 품질 추정기 (540)에 제공될 수도 있다. 이것은 도 4 와 관련하여 위에서 설명된 바와 같이 달성될 수도 있다. 예를 들어, 러프니스 (528), 불연속성 (530), 둔탁성 (532) 및 빈약함 (534) 왜곡들은 연산 포어그라운드 품질 추정기 (540)에 제공될 수도 있다. 연산 포어그라운드 품질 추정기 (540)는 다중 객관적 왜곡들 (예컨대, 러프니스 (528), 불연속성 (530), 둔탁성 (532) 및 빈약함 (534) 왜곡들)에 기초하여 포어그라운드 품질 (550)을 추정할 수도 있다. 포어그라운드 품질 (550)은 연산 전체 품질 추정기 (552)에 제공될 수도 있다.

[0063] 일부의 구성들에서, 다중 객관적 왜곡들 중의 하나 이상은 연산 백그라운드 품질 추정기 (542)에 제공될 수도 있다. 이것은 도 4 와 관련하여 위에서 설명된 바와 같이 달성될 수도 있다. 예를 들어, 허싱 (536) 및 가변성 (538) 왜곡들은 연산 백그라운드 품질 추정기 (542)에 제공될 수도 있다. 연산 백그라운드 품질 추정기 (542)는 다중 객관적 왜곡들 (예컨대, 허싱 (536) 및 가변성 (538) 왜곡들)에 기초하여 백그라운드 품질 (548)을 추정할 수도 있다. 백그라운드 품질 (548)은 연산 전체 품질 추정기 (552)에 제공될 수도 있다.

[0064] 연산 전체 품질 추정기 (552)는 포어그라운드 품질 (550) 및 백그라운드 품질 (548)에 기초하여 전체 품질 (554)을 추정할 수도 있다. 이것은 도 4 와 관련하여 위에서 설명된 바와 같이 달성될 수도 있다. 예를 들어, 연산 전체 품질 추정기 (552)는 포어그라운드 품질 (550) 및 백그라운드 품질 (548)에 기초하여 전체 품질 (554)을 추정할 수도 있다. 또 다른 예에서, 연산 전체 품질 추정기 (552)는 포어그라운드 품질 (550) 및 백그라운드 품질 (548)에 추가하여, 러프니스 (528), 불연속성 (530), 둔탁성 (532), 빈약함 (534), 허싱 (536) 및/또는 가변성 (538) 왜곡들에 직접적으로 기초하여 전체 품질 (554)을 추정할 수도 있다. 전

자 디바이스 (556) 는 도 4 와 관련하여 설명된 바와 같은 계층적 구조에 따라 스피치 신호 품질을 측정할 수도 있다는 것에 주목해야 한다.

[0065] 도 6 은 본원에서 개시된 시스템들 및 방법들에 따라 구현될 수도 있는 계층적 구조 (600) 의 하나의 예를 예시하는 블록도이다. 도 6 에서 예시된 계층적 구조 (600) 는 침투적 접근법의 하나의 예이다. 일부의 구성들에서, 특징들은 원본 스피치 신호 (612) 및 수정된 스피치 신호 (624) 로부터 추출될 수도 있다 (656). 예를 들어, 원본 스피치 신호 (612) 및 수정된 스피치 신호 (624) 는 인간의 달팽이관을 모델링하는 달팽이관 모델러 (cochlear modeler) 에 제공될 수도 있다. 달팽이관 모델은 청각 프로세싱에서 인간의 달팽이관의 효과들을 근사화하기 위하여 원본 스피치 신호 (612) 및 수정된 스피치 신호 (624) 를 수정한다.

[0066] 계층적 구조 (600) 에서, 다중 객관적 왜곡들은 원본 스피치 신호 (612) (예컨대, 달팽이관 모델링된 원본 스피치 신호) 및 수정된 스피치 신호 (624) (예컨대, 달팽이관 모델링된 수정된 스피치 신호) 에 기초하여 추정될 수도 있다. 하나의 예에서, 왜곡들은 러프니스 (628), 불연속성 (630), 둔탁성 (632), 빈약함 (634), 허싱 (636) 및/또는 가변성 (638) 왜곡들을 포함할 수도 있다. 위에서 설명된 바와 같이, 러프니스 (628) 왜곡은 "러프" 및 "하쉬" 디스크립터들에 대응할 수도 있다. 러프니스 (628) 왜곡은 러프니스 모듈 (658) 에 의해 결정될 수도 있다. 불연속성 (630) 왜곡은 "플러터링" 및 "불연속적" 디스크립터들에 대응할 수도 있다. 불연속성 (630) 왜곡은 불연속성 모듈 (660) 에 의해 결정될 수도 있다. 둔탁성 (632) 왜곡은 "둔탁한" 및 "머플드" 디스크립터들에 대응할 수도 있다. 둔탁성 (632) 왜곡은 둔탁성 모듈 (662) 에 의해 결정될 수도 있다. 빈약함 (634) 왜곡은 "떨어진" 및 "빈약한" 디스크립터들에 대응할 수도 있다. 빈약함 (634) 왜곡은 빈약함 모듈 (664) 에 의해 결정될 수도 있다. 허싱 (636) 왜곡은 "허싱", "러싱" 및 "로링" 디스크립터들에 대응할 수도 있다. 허싱 (636) 왜곡은 허싱 모듈 (666) 에 의해 결정될 수도 있다. 가변성 (638) 왜곡은 "버블링", "간헐적" 및 "가변적" 디스크립터들에 대응할 수도 있다. 가변성 (638) 왜곡은 가변성 모듈 (668) 에 의해 결정될 수도 있다.

[0067] 러프니스 (628), 불연속성 (630), 둔탁성 (632) 및 빈약함 (634) 왜곡들은 포어그라운드 왜곡들로서 범주화될 수도 있다. 허싱 (636) 및 가변성 (638) 왜곡들은 백그라운드 왜곡들로서 범주화될 수도 있다. 위에서 설명된 바와 같이, 포어그라운드 품질 (650) 은 선택적으로, 러프니스 (628), 불연속성 (630), 둔탁성 (632) 및 빈약함 (634) 왜곡들에 기초하여 추정될 수도 있다. 또한, 백그라운드 품질 (648) 은 선택적으로, 허싱 (636) 및 가변성 (638) 왜곡들에 기초하여 추정될 수도 있다.

[0068] 구조에서 예시된 바와 같이, 전체 품질 (654) 은 포어그라운드 품질 (650) 및 백그라운드 품질 (648) 에 직접적으로 기초할 수도 있다. 선택적으로, 전체 품질 (654) 은 포어그라운드 품질 (650) 및 백그라운드 품질 (648) 에 추가하여, 다중 객관적 왜곡들에 직접적으로 기초할 수도 있다.

[0069] 개별적인 왜곡들의 객관적 점수들은 2 개의 전체 점수들로 합성될 수도 있다: 포어그라운드 품질 (650) 에 대한 하나 및 백그라운드 품질 (648) 에 대한 다른 하나. 예를 들어, 포어그라운드 품질 (650) 은 신호 품질 (SIG, SGN) 로 나타내어질 수도 있고, 백그라운드 품질 (648) 은 백그라운드 품질 (436) (BGN) 로서 나타내어질 수도 있다.

[0070] 포어그라운드 품질 (650) 및 백그라운드 품질 (648) 점수들은 하나의 최종적인 전체 품질 (654) 점수로 합성될 수도 있다. 전체 품질 (654) 은 전체 품질 (654) (MOS) 로서 나타내어질 수도 있다. 일부의 가능한 합성 접근법들은 다음과 같이 주어진다 (그러나 이것으로 제한되지 않음): 선형 회귀 (예컨대, $MOS = b_2 * SIG + b_1 * BGN + b_0$) 또는 비선형 회귀 (예컨대, $MOS = b_4 * SGN^2 + b_3 * SGN + b_2 * BGN^2 + b_1 * BGN + b_0$).

[0071] 본원에서 개시된 시스템들 및 방법들은, 고주파수 왜곡들 및 백그라운드 잡음의 레벨과 같이, 스피치 품질 왜곡들에 대한 더 많은 세부사항들을 제공할 수도 있는, 테스트 중인 각각의 스피치 신호에 대한 객관적 점수들의 3 개의 레벨들을 제공할 수도 있다. 추가적으로, 본원에서 개시된 시스템들 및 방법들은 알고리즘들의 개발을 더욱 용이하게 할 수도 있다. 예를 들어, 개발자들은 속성들/왜곡들의 어떤 타입들에 집중할 수 있고, 그러므로, 탐구하기 위한 인자들의 범위를 감소시킬 수 있다. 본원에서 개시된 시스템들 및 방법들은 또한, 전체 품질 (654) 에 대한 예측의 개선된 정확도를 제공할 수도 있다. 몇몇 개별적인 점수들로부터 합성될 때, 예를 들어, 전체 품질 (654) 의 예측은 단일 점수들에 기초하여 직접적으로 그것을 예측하는 것보다 더욱 정확 할 수 있다.

[0072] 계층적 구조 (600) 는 이동 디바이스, 이동 스테이션 (mobile station), 가입자 스테이션 (subscriber station), 클라이언트, 클라이언트 스테이션 (client station), 사용자 장비 (user equipment; UE), 원격 스테이션 (remote station), 액세스 단말, 이동 단말, 단말, 사용자 단말, 가입자 유닛 등으로서 또한 알려진 무선 통신 디바이스와 같은 전자 디바이스 (556)에 의해 구현될 수도 있다. 전자 디바이스들 (556)의 다른 예들은 랙톱들 또는 데스크톱 컴퓨터들, 셀룰러 전화들, 스마트폰들, 무선 모뎀들, e-리더들, 태블릿 디바이스들, 게임용 시스템들 등을 포함한다. 또한, 본 시스템들 및 방법들은 기지국, 적응적 잡음 상쇄를 수행하는 전자 디바이스 (556) 등에서 이용될 수도 있다.

[0073] 계층적 구조 (600)에 의해 결정되는 전체 품질 (654)은 스피치의 주어진 부분에 대한 인간 주관적 점수들을 시뮬레이팅할 수도 있다. 다시 말해서, 계층적 구조 (600)는 인간 청자들 (140a 내지 140c)이 주관적 점수들을 실시간으로 제공할 것을 요구하는 대신에, 트레이닝된 데이터에 기초하여 전체 품질 (654)을 결정할 수도 있다. 이것을 행하기 위하여, 계층적 구조 (600)는 상이한 왜곡들을 분리하기 위하여 수정된 스피치 신호 또는 원본 스피치 신호 (612)로부터의 특징들의 추출 (656)을 이용할 수도 있다. 특징들은 다중 객관적 왜곡 차원들의 각각에 대한 예측 점수를 결정하기 위해 이용될 수도 있다.

[0074] 도 7은 본원에서 개시된 시스템들 및 방법들에 따라 구현될 수도 있는 계층적 구조 (700)의 또 다른 예를 예시하는 블록도이다. 도 7에서 예시된 계층적 구조 (700)는 비-침투적 접근법의 하나의 예이다. 이 예에서 설명된 엘리먼트들은 도 6과 관련하여 설명된 바와 같은 대응하는 엘리먼트들과 유사할 수도 있다. 그러나, 이 예에서는, 특징들이 추출될 수도 있고 왜곡들이 (예컨대, 원본 스피치 신호 (612) 없이) 수정된 스피치 신호 (724)에 기초하여 결정될 수도 있다.

[0075] 이 계층적 구조 (700)에서, 다중 객관적 왜곡들은 수정된 스피치 신호 (724) (예컨대, 달팽이판 모델링된 수정된 스피치 신호)에 기초하여 추정될 수도 있다. 하나의 예에서, 왜곡들은 러프니스 (728), 불연속성 (730), 둔탁성 (732), 빈약함 (734), 허싱 (736) 및/또는 가변성 (738) 왜곡들을 포함할 수도 있다. 위에서 설명된 바와 같이, 러프니스 (728) 왜곡은 "러프" 및 "하쉬" 디스크립터들에 대응할 수도 있다. 러프니스 (728) 왜곡은 러프니스 모듈 (758)에 의해 결정될 수도 있다. 불연속성 (730) 왜곡은 "플러터링" 및 "불연속적" 디스크립터들에 대응할 수도 있다. 불연속성 (730) 왜곡은 불연속성 모듈 (760)에 의해 결정될 수도 있다. 둔탁성 (732) 왜곡은 "둔탁한" 및 "머플드" 디스크립터들에 대응할 수도 있다. 둔탁성 (732) 왜곡은 둔탁성 모듈 (762)에 의해 결정될 수도 있다. 빈약함 (734) 왜곡은 "떨어진" 및 "빈약한" 디스크립터들에 대응할 수도 있다. 빈약함 (734) 왜곡은 빈약함 모듈 (764)에 의해 결정될 수도 있다. 허싱 (736) 왜곡은 허싱 모듈 (766)에 의해 결정될 수도 있다. 가변성 (738) 왜곡은 "버블링", "간헐적" 및 "가변적" 디스크립터들에 대응할 수도 있다. 가변성 (738) 왜곡은 가변성 모듈 (768)에 의해 결정될 수도 있다.

[0076] 러프니스 (728), 불연속성 (730), 둔탁성 (732) 및 빈약함 (734) 왜곡들은 포어그라운드 왜곡들로서 범주화될 수도 있다. 허싱 (736) 및 가변성 (738) 왜곡들은 백그라운드 왜곡들로서 범주화될 수도 있다. 위에서 설명된 바와 같이, 포어그라운드 품질 (750)은 선택적으로, 러프니스 (728), 불연속성 (730), 둔탁성 (732) 및 빈약함 (734) 왜곡들에 기초하여 추정될 수도 있다. 또한, 백그라운드 품질 (748)은 선택적으로, 허싱 (736) 및 가변성 (738) 왜곡들에 기초하여 추정될 수도 있다.

[0077] 구조에서 예시된 바와 같이, 전체 품질 (754)은 포어그라운드 품질 (750) 및 백그라운드 품질 (748)에 직접적으로 기초할 수도 있다. 선택적으로, 전체 품질 (754)은 포어그라운드 품질 (750) 및 백그라운드 품질 (748)에 추가하여, 다중 객관적 왜곡들에 직접적으로 기초할 수도 있다.

[0078] 개별적인 왜곡들의 객관적 점수들은 2개의 전체 점수들로 합성될 수도 있다: 포어그라운드 품질 (750)에 대한 하나 및 백그라운드 품질 (748)에 대한 다른 하나. 예를 들어, 포어그라운드 품질 (750)은 신호 품질 (SIG, SGN)로 나타내어질 수도 있고, 백그라운드 품질 (748)은 백그라운드 품질 (436) (BGN)로서 나타내어질 수도 있다.

[0079] 포어그라운드 품질 (750) 및 백그라운드 품질 (748) 점수들은 하나의 최종적인 전체 품질 (754) 점수로 합성될 수도 있다. 전체 품질 (754)은 전체 품질 (754) (MOS)로서 나타내어질 수도 있다. 일부의 가능한 합성 접근법들은 다음과 같이 주어진다 (그러나 이것으로 제한되지 않음): 선형 회귀 (예컨대, $MOS = b_2 * SIG + b_1 * BGN + b_0$) 또는 비선형 회귀 (예컨대,

$$MOS = b_4 * SGN^2 + b_3 * SGN + b_2 * BGN^2 + b_1 * BGN + b_0) .$$

[0080] 본원에서 개시된 시스템들 및 방법들은, 고주파수 왜곡들 및 백그라운드 잡음의 레벨과 같이, 스피치 품질 왜곡 들에 대한 더 많은 세부사항들을 제공할 수도 있는, 테스트 중인 각각의 스피치 신호에 대한 객관적 점수들의 3 개의 레벨들을 제공할 수도 있다. 추가적으로, 본원에서 개시된 시스템들 및 방법들은 알고리즘들의 개발을 더욱 용이하게 할 수도 있다. 예를 들어, 개발자들은 속성들/왜곡들의 어떤 타입들에 집중할 수 있고, 그러므로, 탐구하기 위한 인자들의 범위를 감소시킬 수 있다. 본원에서 개시된 시스템들 및 방법들은 또한, 전체 품질 (754)에 대한 예측의 개선된 정확도를 제공할 수도 있다. 몇몇 개별적인 점수들로부터 합성될 때, 예를 들어, 전체 품질 (754)의 예측은 단일 점수들에 기초하여 직접적으로 그것을 예측하는 것보다 더욱 정확 할 수 있다.

[0081] 계층적 구조 (700)는 이동 디바이스, 이동 스테이션, 가입자 스테이션, 클라이언트, 클라이언트 스테이션, 사용자 장비 (UE), 원격 스테이션, 액세스 단말, 이동 단말, 단말, 사용자 단말, 가입자 유닛 등으로서 또한 알려진 무선 통신 디바이스와 같은 전자 디바이스 (556)에 의해 구현될 수도 있다. 전자 디바이스들 (556)의 다른 예들은 랙톱들 또는 데스크톱 컴퓨터들, 셀룰러 전화들, 스마트폰들, 무선 모뎀들, e-리더들, 태블릿 디바이스들, 게임용 시스템들 등을 포함한다. 또한, 본 시스템들 및 방법들은 기지국, 적응적 잡음 상쇄를 수행하는 전자 디바이스 (556) 등에서 이용될 수도 있다.

[0082] 계층적 구조 (700)에 의해 결정되는 전체 품질 (754)은 스피치의 주어진 부분에 대한 인간 주관적 점수들을 시뮬레이팅할 수도 있다. 다시 말해서, 계층적 구조 (700)는 인간 청자들 (140a 내지 140c)이 주관적 점수들을 실시간으로 제공할 것을 요구하는 대신에, 트레이닝된 데이터에 기초하여 전체 품질 (754)을 결정할 수도 있다. 이것을 행하기 위하여, 계층적 구조 (700)는 상이한 왜곡들을 분리하기 위하여 수정된 스피치 신호 또는 원본 스피치 신호 (712)로부터의 특징들의 추출 (756)을 이용할 수도 있다. 특징들은 다중 객관적 왜곡 차원들의 각각에 대한 예측 점수를 결정하기 위해 이용될 수도 있다.

[0083] 도 8은 추출된 특징들 (856a) 및 하나 이상의 트레이닝 스피치 샘플들 (868)에 기초한 주관적 품질 측정 점수들의 하나의 단일 차원의 예측을 위한 방법 (800)을 예시하는 블록도이다. 트레이닝 데이터베이스 (882)는 특징들이 추출 (856a)되는 하나 이상의 트레이닝 스피치 샘플들 (868)을 포함할 수도 있다. 트레이닝 데이터베이스 (882)는 또한, 하나 이상의 트레이닝 스피치 샘플들 (868)에 대한 대응하는 주관적 점수들 (872)을 포함할 수도 있다. 이것들은 주관적 인간 청자 방법을 이용하여 수집될 수도 있고, 즉, 여기서, 많은 인간 청자들 (104a 내지 104c)은 하나 이상의 트레이닝 스피치 샘플들 (868)을 청취하도록 각각 요청받고, 다음으로, 하나 이상의 범주들에서 하나 이상의 트레이닝 스피치 샘플들 (868)을 등급부여하도록 요청받는다. 하나의 구성에서, 트레이닝 데이터베이스 (882)는 도 5에서 예시된 왜곡 차원들, 즉, 러프니스 (528), 불연속성 (530), 둔탁성 (532), 빈약함 (534), 허성 (536) 및 가변성 (538)의 각각에서 하나 이상의 트레이닝 스피치 샘플들 (868)에 대한 점수들을 포함할 수도 있다. 또한, 트레이닝 데이터베이스 (882)는 대응하는 주관적 점수들 (872)과 함께, 많은 상이한 조건들 (예컨대, 상이한 코텍들, 상이한 네트워크 기술들, 상이한 변조 방식들 등)에 대한 하나 이상의 트레이닝 스피치 샘플들 (868)을 포함할 수도 있다.

[0084] 다음으로, 추정 가중치들 (870)은 하나 이상의 트레이닝 스피치 샘플들 (868)로부터의 추출된 특징들 (856a) 및 대응하는 주관적 점수들 (872)에 기초하여 추정될 수도 있다. 다시 말해서, 하나 이상의 트레이닝 스피치 샘플들 (868)로부터의 추출된 특징들 (856a)로 하여금, 하나 이상의 트레이닝 스피치 샘플들 (868)에 대응하는 추정 가중치들 (870)을 생성하게 할 추정 가중치들 (870)이 결정될 수도 있다. 이 트레이닝은 (예를 들어, 도 5에서 예시된 바와 같은) 연산 전체 품질 추정기 (540)가 원본 스피치 신호 (512)의 부분에 대한 전체 품질을 결정하기 전에 오프라인 (offline)으로 수행될 수도 있다.

[0085] 다음으로, 특징들에 대한 가중치들 (874)은 예컨대, 선형 회귀 (878) 알고리즘을 이용하여 하나 이상의 테스팅 스피치 샘플들 (876) (즉, 전체 품질 (554)이 희망되는 원본 스피치 (512) 또는 열화된 스피치)로부터의 추출된 특징들 (856b)에 적용될 수도 있다. 하나 이상의 테스팅 스피치 샘플들 (876)은 테스팅 데이터베이스 (884)에서 위치될 수도 있다. 하나의 구성에서, 특징들에 대한 가중치들 (874)의 세트는 왜곡 차원들, 즉, 러프니스 (528), 불연속성 (530), 둔탁성 (532), 빈약함 (534), 허성 (536) 및 가변성 (538)의 각각에 대한 트레이닝 데이터로부터 결정될 수도 있다. 그러므로, 특정 차원에 대한 주관적 점수들의 예측 (880)은 특정 왜곡 차원에 대한 특징들에 대한 가중치들 (874)을 테스팅 스피치 샘플들 (876)의 추출된 특징들 (856b)에 적용함으로써 결정될 수도 있다.

[0086] 도 9 는 도 9a 내지 도 9c 에 대응하는 예의 다수의 부분들을 예시하는 블록도이다. (도 9c 에서 벡터 b 로 서 나타낸) 가중치들 (996) 은 트레이닝 주기 동안에 (예컨대, 전체 품질 (554) 이 결정되기 전에 오프라인에서) 결정될 수도 있다. 가중치들 (996) 은 스피치의 전체 품질 (554) 을 결정하기 위한 실행시간 (runtime) 동안에 도 9b 에서 도시된 스피치 특징들 (992a 내지 992n) (침투적 또는 비-침투적) 에 적용될 수도 있다. 구체적으로, 가중치들 (996) 의 세트, b 는 각각의 왜곡 차원에 대해 결정될 수도 있다.

[0087] 트레이닝 데이터베이스 (882) 는 도 9a 에서 도시된 바와 같은 N 개의 조건들 (988a 내지 988n) 에 대한 트레이닝 스피치 샘플들 (868) 을 포함할 수도 있고, 각각의 조건 (988a 내지 988n) 은 스피치가 수신되었던 상황들의 상이한 세트, 예컨대, 상이한 코덱들, 상이한 네트워크 기술들, 상이한 변조 방식들 등에 대응할 수도 있다.

트레이닝 데이터베이스 (882) 는 또한, N 개의 조건들의 각각에 대하여 각각의 왜곡 차원에 대한 주관적 점수 (986a 내지 986n) 를 포함할 수도 있다. 다시 말해서, 각각의 조건에 대하여, 트레이닝 데이터베이스는 6 개의 주관적 점수들 (872), 각각의 왜곡 차원에 대해 하나를 가질 수도 있다. 집합적으로, 특정 왜곡 차원 (도 9c 에서의 러프니스) 에 대한 모든 N 개의 조건들에 대한 주관적 점수들은 S 로서 지칭될 수도 있다.

[0088] 특징 행렬 (994) 에서의 각각의 특징 벡터 (예컨대, FM 에서의 각각의 열 (column)) 는 특정 조건 (988a 내지 988n) 에 대하여, 즉, 달팽이판 모델 출력 (990a 내지 990n) 의 선택된 섹션들의 분석 시리얼 (analysis serial) 들을 통한 침투적 또는 비-침투적 특징 추출을 이용하여 결정될 수도 있다. 특징 벡터들은 특징 행렬 (994) FM 에서 배치된다. 그러므로, N 개의 조건들 (988a 내지 988n) 이 이용될 경우, 특징 벡터는 N 개의 열들을 가질 수도 있다. 구체적으로, FM (994) 은 이 예에서 $54 \times N$ 행렬이지만, 데이터의 구체적인 사이즈들은 변동될 수도 있다.

[0089] 다음으로, 가중치들 (996) 은 특징 행렬 (994) FM 및 알려진 주관적 점수들 (986a 내지 986n) S 에 기초하여 추정될 수도 있다. 다시 말해서, 특징 행렬 (986), FM 으로 하여금, N 개의 조건들 (988a 내지 988n) 에 대응하는 주관적 점수들 (986a 내지 986n), S 를 생성하게 할 가중치들 (996), b 가 결정될 수도 있다. 그러므로, 도 9c 에서 도시된 바와 같은 가중치들 (996), b 는 수학식 (1) 을 충족시키도록 계산된다:

$$FM * b = S \quad (1)$$

[0091] 여기서, FM 은 트레이닝 스피치의 N 개의 조건들 (988a 내지 988n) 에 대해 결정된 특징 행렬 (994) 이고, b 는 특정 왜곡 차원에 대한 희망하는 가중치들 (996) 이고, S 는 특정 왜곡 차원에 대한 주관적 점수 (986a 내지 986n) 벡터이다. 그러므로, 가중치들 (996) 은 수학식 (2) 에 따라 계산될 수도 있다:

$$b = FM^{-1} * S \quad (2)$$

[0093] 여기서, FM^{-1} 은 특징 역행렬 (994) 이다. 가중치 세트 (996), b 는 각각의 왜곡 차원에 대해 결정될 수도 있고, 각각의 왜곡 차원, 즉, 러프니스 (528), 불연속성 (530), 둔탁성 (532), 빈약함 (534), 허성 (536) 및 가변성 (538) 에 대한 예측 점수들의 추후의 예측을 위하여 저장될 수도 있다. 수학식 (2) 는 이론적 솔루션이라는 것에 주목해야 한다. 실제로는, $FM * b$ 를 S 와 최상으로 일치하게 하는 "b", 예컨대, 다중 선형 회귀를 구하기 위한 다른 방법들이 있을 수도 있다.

[0094] 비록 이 트레이닝이 (예를 들어, 도 5 에서 예시된 바와 같은) 연산 전체 품질 추정기 (552) 가 원본 스피치 신호 (512) 의 부분에 대한 전체 품질 (554) 을 결정하기 전에 오프라인으로 수행될 수도 있더라도, 가중치들 (996) 은 다음으로, 전체 품질 (554) 이 희망되는 테스팅 스피치 샘플들 (876) 로부터 추출된 특징들에 적용될 수도 있다.

[0095] 도 10 은 스피치 신호 품질을 측정하기 위해 구현될 수도 있는 계층적 구조 (1000) 의 하나의 예를 예시하는 흐름도이다. 계층적 구조 (1000) 는 전자 디바이스 (556) 에 의해 수행될 수도 있다. 전자 디바이스 (556) 는 수정된 스피치 신호 (524) 의 다중 객관적 왜곡들을 추정함으로써, 또는 원본 스피치 신호 (512) 에 기초하여 제 1 계층적 구조 레벨을 수행할 수도 있다 (1002). 예를 들어, 객관적 왜곡은 수정된 스피치 신호 (524) 에서 왜곡들을 표현하는 하나 이상의 메트릭들일 수도 있다. 다중 객관적 왜곡들은 스피치 품질의 독립적인 차원을 나타낼 수도 있다. 예를 들어, 다중 객관적 스피치 왜곡들은 러프니스 (528), 불연속성 (530), 둔탁성 (532), 빈약함 (534), 허성 (536) 및/또는 가변성 (538) 일 수도 있다.

[0096] 다음으로, 전자 디바이스 (556) 는 포어그라운드 품질 (550) 및 백그라운드 품질 (548) 을 추정함으로써 제 2 계층적 구조 레벨을 수행할 수도 있다 (1004). 포어그라운드 품질 (550) 및 백그라운드 품질 (548) 은 제 1 계층적 구조 레벨에서 추정된 다중 객관적 왜곡들에 기초할 수도 있다. 다시 말해서, 제 2 계층적 구조 레

벨은 제 1 계층적 구조 레벨 전에 수행되지 않을 수도 있다. 포어그라운드 품질 (550) 은 포어그라운드 품질 추정기 (540) 에 의해 추정될 수도 있고, 백그라운드 품질 (548) 은 백그라운드 품질 추정기 (542) 에 의해 추정될 수도 있다.

[0097] 포어그라운드 품질 (550) 은 객관적 왜곡들 중의 하나 이상에 기초할 수 있다. 예를 들어, 러프니스 (528), 불연속성 (530), 둔탁성 (532) 및 빈약함 (534) 은 포어그라운드 객관적 왜곡들일 수도 있다. 포어그라운드 품질 (550) 은 오직 둔탁성 (532) 및 빈약함 (534), 또는 포어그라운드 객관적 왜곡들의 임의의 다른 가능한 조합을 이용하여 결정될 수도 있다. 백그라운드 품질 (548) 은 객관적 왜곡들 중의 하나 이상에 기초할 수 있다. 백그라운드 품질 (548) 은 허싱 (536) 및 가변성 (538) 일 수도 있다. 백그라운드 품질 (548) 은 허싱 (536) 및 가변성 (538) 의 양자, 또는 단지 허싱 (536) 또는 가변성 (538) 을 이용하여 구해질 수도 있다.

본원에서 개시된 시스템들 및 방법들은 객관적 왜곡들의 임의의 조합을 이용할 수도 있다. 다음으로, 객관적 왜곡들은 단지 러프니스 (528), 불연속성 (530), 둔탁성 (532), 빈약함 (534), 허싱 (536) 또는 가변성 (538) 보다 더 많은 것을 포함할 수도 있다.

[0098] 다음으로, 전자 디바이스 (556) 는 스피치 신호의 전체 품질 (554) 을 추정함으로써 제 3 계층적 구조 레벨을 수행할 수도 있다 (1006). 스피치 신호의 전체 품질 (554) 은 포어그라운드 품질 (550), 백그라운드 품질 (548) 에 기초할 수도 있고, 그리고 선택적으로, 다중 객관적 왜곡들에 직접적으로 기초할 수도 있다. 다시 말해서, 제 3 계층적 구조 레벨은 제 1 계층적 구조 레벨 또는 제 2 계층적 구조 레벨 전에 수행되지 않을 수도 있다. 객관적 왜곡들은 전체 품질 (554) 을 결정하기 위하여 포어그라운드 품질 (550) 및 백그라운드 품질 (548) 을 통해 간접적으로 이용될 수도 있다. 추가적으로, 객관적 왜곡들은 포어그라운드 품질 (550) 및 백그라운드 품질 (548) 에 추가하여, 전체 품질 (554) 을 결정하기 위해 직접적으로 이용될 수도 있다. 전체 품질 (554) 은 스피치 품질의 주관적 척도들을 근사화할 수도 있다.

[0099] 도 11 은 스피치 신호 품질을 측정하기 위해 구성된 네트워크 디바이스 (1101) 및 2 개의 전자 디바이스들 (1156a 내지 1156b) 의 하나의 구성을 예시하는 블록도이다. 네트워크 디바이스 (1101) 는 스피치 평가기 (1198b), 적응 모듈 (1103b), 및 피드백 모듈 (1105b) 을 포함할 수도 있다. 네트워크 디바이스 (1101) 는 무선 라운터, 서버, 기지국, 셀 전화 타워 또는 컴퓨터 시스템일 수도 있다. 스피치 평가기 (1198b) 는 본원에서 개시된 방법들을 수행하기 위해 이용될 수도 있다. 스피치 평가기 (1198b) 는 연산 왜곡 추정기 (526), 연산 포어그라운드 품질 추정기 (540), 연산 백그라운드 품질 추정기 (542) 및 연산 전체 품질 추정기 (552) 를 포함할 수도 있다. 네트워크 디바이스 (1101) 는 하나 이상의 전자 디바이스들 (1156a 내지 1156b) 로부터 하나 이상의 스피치 신호들 (1107a 내지 1107b) 을 획득할 수도 있다. 스피치 신호들 (1107a 내지 1107b) 은 수정된 스피치 신호 (524), 원본 스피치 신호 (512), 또는 수정된 스피치 신호 (524) 및 원본 스피치 신호 (512) 의 양자일 수도 있다. 네트워크 디바이스 (1101) 는 스피치 신호들 (1107a 내지 1107b) 에서 발견된 다중 객관적 왜곡들을 결정하기 위해 스피치 평가기 (1198b) 를 이용할 수도 있다. 다중 객관적 왜곡들은 스피치 신호들 (1107a 내지 1107b) 의 포어그라운드 품질 (550) 및 백그라운드 품질 (548) 을 결정하기 위해 이용될 수도 있다. 다음으로, 스피치 신호들 (1107a 내지 1107b) 의 전체 품질 (554) 은 포어그라운드 품질 (550) 및 백그라운드 품질 (548) 을 이용하여 결정될 수도 있다.

[0100] 스피치 신호들 (1107a 내지 1107b) 의 전체 품질 (554) 에 따라서는, 네트워크 디바이스 (1101) 가 프로세싱 구조에 대해 변경들이 행해질 필요가 있는 것으로 결정할 수도 있다. 변경들은 적응 모듈 (1103b) 에 의해 행해질 수도 있다. 예를 들어, 적응 모듈 (1103b) 은 네트워크 디바이스 (1101) 에 의해 행해진 인코딩, 디코딩 또는 트랜스코딩 (transcoding) 을 수정할 수 있을 수도 있다. 적응 모듈 (1103b) 은 또한, 스피치 신호들 (1107a 내지 1107b) 에 대해 할당된 대역폭을 변경시키거나, 네트워크 디바이스 (1101) 의 비트 레인트를 변경시킬 수 있을 수도 있다. 또 다른 예에 대하여, 전자 디바이스 (1156a 내지 1156b) 는 스피치 신호 (1107a 내지 1107b) 를 네트워크 디바이스 (1101) 로 전송할 수도 있다. 네트워크 디바이스 (1101) 는 동일한 스피치 신호 (1107a 내지 1107b) 를 또 다른 전자 디바이스 (1156a 내지 1156b) 로 전송할 수도 있고, 그 후에, 스피치 평가기 (1198b) 는 네트워크 디바이스 (1101) 가 스피치 신호 (1107a 내지 1107b) 를 수신하였을 때에 스피치 신호 (1107a 내지 1107b) 의 전체 품질 (554) 을 결정할 수도 있고, 네트워크 디바이스 (1101) 가 그 것을 다른 전자 디바이스 (1156a 내지 1156b) 로 전송할 때에 스피치 신호 (1107a 내지 1107b) 를 결정할 수도 있다. 전송된 스피치 신호 (1107a 내지 1107b) 의 전체 품질 (554) 이 너무 낮을 경우, 네트워크 디바이스 (1101) 는 열화를 야기하였을 수도 있는 네트워크 디바이스 (1101) 에 의해 수행된 인코딩을 결정하기 위해 스피치 평가기 (1198b) 를 이용할 수도 있다. 다음으로, 네트워크 디바이스 (1101) 는 인코딩 방법을, 스피치 신호 (1107a 내지 1107b) 로 더욱 양호하게 수행하는 것으로 변경시키기 위해 적응 모듈 (1103b) 을 이용할 수

도 있다. 적응 모듈 (1103b) 은 이 예에서 단지 이 변경들을 행할 수 있을 수도 있지만, 적응 모듈 (1103b) 이 행할 수도 있는 특정 변경들은 다른 구성들에서 변동될 수도 있다.

[0101] 네트워크 디바이스 (1101) 는 네트워크 디바이스 (1101) 가 하나 이상의 전자 디바이스들 (1156a 내지 1156b) 에 접속되는 동안에 변경들을 행할 수도 있다. 네트워크 디바이스 (1101) 는 또한, 더욱 대규모의 변경들이 행해질 필요가 있을 수도 있는 것으로 결정할 수도 있고, 전자 디바이스 (1156a 내지 1156b) 가 네트워크 디바이스 (1101) 에 접속되지 않는 동안에 이 변경들을 행할 수도 있다. 네트워크 디바이스 (1101) 는 또한, 피드백 모듈 (1105b) 에서 스피치 신호들 (1107a 내지 1107b) 의 전체 품질 (554) 의 점수를 저장할 수도 있다.

정비 및 유지가 네트워크 디바이스 (1101) 에 대해 행해지고 있을 때, 피드백 모듈 (1105b) 은 스피치 신호들 (1107a 내지 1107b) 의 전체 품질 (554) 의 점수를 제공할 수도 있다. 전체 품질 (554) 의 저장된 점수를 이용하여, 정비 및 유지 동안에 하드웨어에 대해 어떤 변경들이 행해질 수도 있다. 예를 들어, 스피치 신호들 (1107a 내지 1107b) 의 전체 품질 (554) 이 셀 전화 타워에서 너무 낮은 것으로 일관적으로 결정될 경우, 셀 전화 타워의 하드웨어는 업데이트될 수도 있거나 더욱 새로운 하드웨어로 대체될 수도 있다.

[0102] 피드백 모듈 (1105b) 은 또한, 피드백을 네트워크 디바이스 (1101) 에 접속되는 하나 이상의 전자 디바이스들 (1156a 내지 1156b) 에 제공할 수도 있다. 피드백은 네트워크 디바이스 (1101) 가 스피치 신호들 (1107a 내지 1107b) 을 수신하였을 때에 스피치 신호들 (1107a 내지 1107b) 에 대한 전체 품질 (554) 의 점수를 포함할 수도 있고, 또한, 네트워크 디바이스 (1101) 가 스피치 신호들 (1107a 내지 1107b) 을 또 다른 전자 디바이스 (1156a 내지 1156b) 로 전송하였을 때에 스피치 신호들 (1107a 내지 1107b) 에 대한 전체 품질 (554) 의 점수를 포함할 수도 있다. 피드백은 네트워크 디바이스 (1101) 가 스피치 신호들 (1107a 내지 1107b) 의 스피치 신호 열화의 원인이 아닐 수도 있다는 것을 표시할 수도 있다. 하나 이상의 전자 디바이스들 (1156a 내지 1156b) 에 제공된 피드백은 또한, 전자 디바이스들 (1156a 내지 1156b) 로부터 원래 송신되었을 때의 스피치 신호의 전체 품질 (554) 이 낮았다는 것을 보여줄 수도 있고, 이것은 아마도 네트워크 디바이스 (1101) 가 신호 열화의 원인이 아닐 수도 있다는 것을 표시할 수도 있다. 피드백은 전자 디바이스들 (1156a 내지 1156b) 이 송신된 스피치 신호들 (1107a 내지 1107b) 의 전체 품질 (554) 을 개선시킬 수 있는 방법들을 표시할 수도 있다. 예를 들어, 피드백은 전자 디바이스들 (1156a 내지 1156b) 에 의해 수행된 스피치 신호들 (1107a 내지 1107b) 의 압축이 적당하게 기능하고 있지 않다는 것을 표시할 수도 있다.

[0103] 전자 디바이스 (1156a) 는 스피치 평가기 (1198a), 적응 모듈 (1103a), 및 피드백 모듈 (1105a) 을 포함할 수도 있다. 스피치 평가기 (1198a) 는 본원에서 개시된 방법들을 수행하기 위해 이용될 수도 있다. 전자 디바이스 (1156a) 는 스피치 신호 (1107a) 를 획득할 수도 있거나, 스피치 신호 (1107a) 를 네트워크 디바이스 (1101) 로 송신할 수도 있다. 스피치 신호 (1107a) 는 수정된 스피치 신호 (524), 원본 스피치 신호 (512), 또는 수정된 스피치 신호 (524) 및 원본 스피치 신호 (512) 의 양자일 수도 있다. 전자 디바이스 (1156a) 는 스피치 신호 (1107a) 의 다중 객관적 왜곡들 및 전체 품질 (554) 을 결정하기 위해 스피치 평가기 (1198a) 를 이용할 수도 있다. 적응 모듈 (1103a) 은 스피치 신호 (1107a) 의 전체 품질 (554) 에 기초하여 전자 디바이스 (1156a) 의 성능을 변경시킬 수도 있다. 피드백 (1105a) 은 전체 품질 (554) 및 전자 디바이스 (1156a) 에 의해 수행되고 있는 프로세싱의 타입들에 대하여 캐리어 피드백을 네트워크 디바이스 (1101) 에 제공할 수도 있다. 따라서, 본원에서 개시된 시스템들 및 방법들은 (사용자 및/또는 네트워크 제공자가 예를 들어, 그 음성 대화들의 품질의 '메트릭' 을 얻을 수 있을 수도 있도록) 스마트폰들 및/또는 다른 디바이스들에서의 객관적 품질의 척도들 (및 스피치 품질 장애들의 관련된 진단들) 을 제공할 수도 있다. 상기한 것과 유사하게, 품질의 이 메트릭들은 또한 Skype 등과 같은 소프트-폰 (soft-phone) 들 애플리케이션들 내에 포함될 수도 있다.

[0104] 도 12 는 스피치 신호 품질을 측정하기 위해 구성된 전자 디바이스 (1256) 의 하나의 구성을 예시하는 블록도이다. 전자 디바이스 (1256) 는 컴퓨터 시스템, 게임용 시스템, 서버 또는 이동 디바이스일 수도 있다. 전자 디바이스 (1256) 는 또한, 함께 작동하는 하나 이상의 전자 디바이스들 (1256) 일 수도 있다. 즉, 블루투스 헤드폰들, 잡음 상쇄 헤드폰들, 이동 디바이스 또는 스피커들.

[0105] 전자 디바이스 (1256) 는 스피치 평가기 (1298), 적응 모듈 (1203), 및 디스플레이 (1205) 를 포함할 수도 있다. 스피치 평가기 (1298) 는 연산 왜곡 추정기 (526), 연산 포어그라운드 품질 추정기 (540), 연산 백그라운드 품질 추정기 (542) 및 연산 전체 품질 추정기 (552) 를 포함할 수도 있다. 스피치 평가기 (1298) 는 전자 디바이스 (1256) 가 전송 및 수신하고 있는 하나 이상의 스피치 신호들의 다중 객관적 왜곡들, 포어그라운드 품질 (550), 백그라운드 품질 (548) 및 전체 품질 (554) 을 결정하기 위해 이용될 수도 있다. 예를 들어, 전자 디바이스 (1256) 는 상이한 무선 통신 서비스 제공자로부터 발신되었던 스피치 신호를 수신하고 있

는 이동 디바이스일 수도 있다. 스피치 평가기 (1298) 는 전자 디바이스 (1256) 에 의해 수신될 때의 스피치 신호의 전체 품질 (554) 을 결정할 수도 있다. 다음으로, 스피치 평가기 (1298) 는 이동 디바이스에 의해 수신된 스피치 신호의 전체 품질 (554) 과, 스피치 신호가 무선 통신 서비스 제공자의 네트워크 내의 네트워크 디바이스 (1101) 에 의해 최초에 수신되었을 때의 스피치 신호의 전체 품질 (554) 을 비교하기 위하여, 피드백을 네트워크 디바이스 (1101) 로 전송할 수도 있다.

[0106] 전자 디바이스 (1256) 는 또한, 적응 모듈 (1203) 을 이용하여 그 성능 및 프로세싱 파라미터들을 적응시킬 수 있을 수도 있다. 적응 모듈 (1203) 은 전자 디바이스 (1256) 에 의해 행해진 인코딩, 디코딩 또는 트랜스코딩을 수정할 수 있을 수도 있다. 적응 모듈 (1203) 은 또한, 하나 이상의 스피치 신호들에 대해 할당된 대역폭을 변경시키거나, 전자 디바이스 (1256) 의 비트 레인트를 변경시킬 수 있을 수도 있다. 예를 들어, 스피치 신호의 전체 품질 (554) 은 너무 낮을 수도 있고, 적응 모듈 (1203) 은 전자 디바이스 (1256) 가 안테나 전력을 증가시켜야 하는 것으로 결정할 수도 있다. 안테나 전력을 증가시키는 것은 셀 타워 및 전자 디바이스 (1256) 사이의 접속을 개선시킬 수도 있다. 스피치 평가기 (1298) 는 전체 품질 (554) 의 새로운 점수가 수용가능한 것으로 결정할 수도 있고, 적응 모듈 (1203) 은 전자 디바이스 (1256) 가 증가된 안테나 전력으로 계속할 것을 명령할 수도 있다. 또 다른 예에 대하여, 전자 디바이스 (1256) 는 잡음 상쇄 헤드폰들의 세트 일 수도 있다. 잡음 상쇄 헤드폰들은 능동 잡음 상쇄를 수행할 수도 있고, 여기서, 헤드폰들은 어떤 잡음이 억압되고 있는지와, 어떤 잡음이 다중 객관적 왜곡들을 이용하여 허용되는지를 결정할 수도 있다. 왜곡들 중의 하나 이상이 스피치 신호를 열화시키고 있을 경우, 잡음 상쇄 헤드폰들은 상쇄되고 있는 잡음들과, 어떤 잡음들이 허용되는지를 적응시키기 위하여, 적응 모듈 (1203) 내에 위치된 능동 잡음 억압을 이용할 수도 있다.

[0107] 전자 디바이스 (1256) 는 전자 디바이스 (1256) 에서 전체 품질 (554) 의 점수를 보여주기 위해 디스플레이 (1205) 를 이용할 수도 있다. 디스플레이 (1205) 는 스피치 신호의 다중 객관적 왜곡 점수들, 포어그라운드 품질 (550), 백그라운드 품질 (548) 또는 전체 품질 (554) 을 보여줄 수도 있다. 이 정보는 전자 디바이스 (1256) 의 조작자에 의해, 또는 전자 디바이스 (1256) 의 하드웨어 또는 프로세싱 파라미터들에 대한 수정들 또는 업그레이드를 행하기 위한 정비 동안에 이용될 수도 있다. 디스플레이 (1205) 상에 제공된 정보는 또한, 스피치 신호가 네트워크 디바이스 (1101) 에 의해 수신되었을 때의 스피치 신호의 전체 품질 (554) 을 보여주기 위해 이용될 수도 있다. 이 정보는 전자 디바이스 (1256) 의 조작자가, 스피치 신호의 열화가 전자 디바이스 (1256) 상에서 발생하는지, 또는 스피치 신호의 열화가 네트워크 디바이스 (1101) 상에서 발생하는지, 또는 스피치 신호가 네트워크 디바이스 (1101) 에 의해 수신되었을 때에 스피치 신호가 이미 열화된 것을 인지하도록 할 수도 있다.

[0108] 도 13 은 스피치 신호 품질을 측정하기 위하여 전자 디바이스 (1256) 에 의해 구현된 방법 (1300) 을 예시하는 흐름도이다. 방법 (1300) 은 전자 디바이스 (1256) (예컨대, 도 11 및 도 12 와 관련하여 설명된 전자 디바이스) 에 의해 수행될 수도 있다. 전자 디바이스 (1256) 의 예들은 이동 디바이스들 (예컨대, 스마트폰들, 셀룰러 전화들, 태블릿 디바이스들, 랩톱 컴퓨터들 등), 데스크톱 컴퓨터들, 개인 정보 단말들 (PDA), 기기들, 텔레비전들, 게임용 시스템들 및 서버들 (예컨대, 네트워크 디바이스 내에 포함된 서버) 을 포함한다. 전자 디바이스 (1256) 는 스피치 신호를 획득할 수도 있다 (1302). 스피치 신호는 수정된 스피치 신호 (524), 원본 스피치 신호 (512), 또는 수정된 스피치 신호 (524) 및 원본 스피치 신호 (512) 의 양자일 수도 있다. 전자 디바이스 (1256) 는 스피치 신호에 기초하여 다중 객관적 왜곡들을 결정하기 위해 (1304) 스피치 평가기 (1298) 를 이용할 수도 있다. 즉, 러프니스 (528), 불연속성 (530), 둔탁성 (532), 빈약함 (534), 허성 (536) 및 가변성 (538).

[0109] 다음으로, 전자 디바이스 (1256) 는 선택적으로, 캐리어 피드백을 네트워크 디바이스 (1101) 로 전송할 수도 있다 (1306). 캐리어 피드백은 다중 객관적 왜곡들 분석을 포함할 수도 있거나, 단지 다중 객관적 왜곡 점수들일 수도 있다. 캐리어 피드백은 전자 디바이스 (1256), 네트워크 디바이스 (1101), 또는 전자 디바이스 (1256) 및 네트워크 디바이스 (1101) 의 양자 상에서의 어느 하나의 적응 프로세스들에 의해 스피치 신호의 품질을 개선시키기 위해 이용될 수도 있다.

[0110] 도 14 는 스피치 신호 품질을 측정하기 위하여 네트워크 디바이스 (1101) 에 의해 구현된 방법 (1400) 을 예시하는 흐름도이다. 방법은 네트워크 디바이스 (1101) (예컨대, 도 11 과 관련하여 설명된 네트워크 디바이스) 에 의해 수행될 수도 있다. 네트워크 디바이스 (1101) 의 예들은 데스크톱 컴퓨터들, 서버들 및 셀 타워들을 포함한다. 네트워크 디바이스 (1101) 는 스피치 신호를 획득할 수도 있다 (1402). 스피치 신호는 수정된 스피치 신호 (524), 원본 스피치 신호 (512), 또는 수정된 스피치 신호 (524) 및 원본 스피치 신호 (512) 의 양자일 수도 있다. 네트워크 디바이스 (1101) 는 스피치 신호에 기초하여 다중 객관적 왜곡들

을 결정하기 위해 (1404) 스피치 평가기 (1198b) 를 이용할 수도 있다. 즉, 러프니스 (528), 불연속성 (530), 둔탁성 (532), 빈약함 (534), 허싱 (536) 및 가변성 (538).

[0111] 다음으로, 네트워크 디바이스 (1101) 는 선택적으로, 다중 객관적 왜곡들에 기초하여 스피치 신호 프로세싱의 하나 이상의 양태들을 적용시키는 것으로 결정할 수도 있다 (1406). 예를 들어, 네트워크 디바이스 (1101) 는 스피치 신호가 최초에 획득될 때에 디코딩이 네트워크 디바이스 (1101) 에 의해 수행되고 있는 것이 적합하지 않은 것으로 결정할 수도 있다. 다음으로, 네트워크 디바이스 (1101) 는 선택적으로, 피드백을 네트워크 디바이스 (1101) 에 접속되는 전자 디바이스 (1156a 내지 1156b) 에 제공할 수도 있다 (1408). 피드백은 다중 객관적 왜곡들 중의 하나 이상을 개선시키기 위하여 네트워크 디바이스 (1101) 가 행하고 있는 적응들을 표시할 수도 있다. 다음으로, 전자 디바이스 (1156a 내지 1156b) 는 통신이 네트워크 디바이스 (1101) 및 전자 디바이스 (1156a 내지 1156b) 사이에서 계속되도록 하기 위하여 이에 따라, 적응들을 행할 수도 있다.

[0112] 도 15 는 특징 추출을 위한 전자 디바이스 (1507) 의 하나의 구성을 예시하는 블록도이다. 전자 디바이스 (1507) 는 특징 추출 모듈 (1529) 을 포함할 수도 있다. 전자 디바이스 (1507) 및/또는 그 하나 이상의 컴포넌트들은 하드웨어 (예컨대, 회로부) 로, 또는 하드웨어 및 소프트웨어의 조합으로 구현될 수도 있다. 추가적으로 또는 대안적으로, 용어 "모듈" 은 컴포넌트가 하드웨어 (예컨대, 회로부) 로, 또는 하드웨어 및 소프트웨어의 조합으로 구현되는 것을 표시할 수도 있다. 예를 들어, 특징 추출 모듈 (1529) 은 하드웨어 (예컨대, 회로부) 로, 또는 하드웨어 및 소프트웨어의 조합 (예컨대, 실행 가능한 명령들을 갖는 프로세서) 로 구현될 수도 있다. 도면들 중의 하나 이상에서 도시된 라인들 또는 화살표들은 컴포넌트들 및/또는 모듈들 사이의 결합들을 나타낼 수도 있다. "결합" 은 직접적 또는 간접적일 수도 있다. 예를 들어, 하나의 모듈은 또 다른 모듈에 (임의의 매개 컴포넌트 없이) 직접적으로 또는 (하나 이상의 매개 컴포넌트들과 함께) 간접적으로 결합될 수도 있다.

[0113] 전자 디바이스 (1507) 의 예들은 이동 디바이스들 (예컨대, 스마트폰들, 셀룰러 전화들, 태블릿 디바이스들, 랩톱 컴퓨터들 등), 컴퓨터들 (예컨대, 테스크톱 컴퓨터들), 네트워크 디바이스들 (예컨대, 기지국들, 라우터들, 스위치들, 게이트웨이들, 서버들 등), 텔레비전들, 자동차 전자기기들 (예컨대, 자동차의 계기판 내로 통합된 전자 디바이스들), 게임용 시스템들, 전자 기기들 등을 포함한다. 일부의 구성들에서, 전자 디바이스 (1507) 는 도 5 와 관련하여 설명된 전자 디바이스 (556) 의 컴포넌트들 중의 하나 이상을 포함할 수도 있고, 및/또는 도 5 와 관련하여 설명된 전자 디바이스 (556) 의 기능들 중의 하나 이상을 수행할 수도 있다.

[0114] 특징 추출 모듈 (1529) 은 수정된 스피치 신호 (1511) 에 기초하여 하나 이상의 특징들 (1527) 을 결정할 수도 있다. 일부의 구성들에서, 하나 이상의 특징들 (1527) 을 결정하는 것은 수정된 스피치 신호 (1511) 및 원본 스피치 신호 (1509) 에 기초할 수도 있다.

[0115] 전자 디바이스 (1507) 는 선택적으로, 원본 스피치 신호 (1509) 를 획득할 수도 있다. 예를 들어, 전자 디바이스 (1507) 는 마이크로폰으로 스피치 신호를 캡처할 수도 있거나, 또 다른 디바이스 (예컨대, 저장 디바이스, 컴퓨터, 전화, 헤드셋 등)로부터 스피치 신호를 수신할 수도 있다. 원본 스피치 신호 (1509) 는 원시 또는 프로세싱되지 않은 신호일 수도 있다. 예를 들어, 원본 스피치 신호 (1509) 는 수정 (예컨대, 열화, 왜곡, 인코딩, 압축, 디코딩, 프로세싱 등) 되지 않은 스피치 신호의 전자적으로 샘플링된 버전일 수도 있다. 원본 스피치 신호 (1509) 가 획득되는 접근법들은 침투적 접근법들 또는 방법들로서 지칭될 수도 있다. 예를 들어, 침투적 접근법들은 원본 스피치 신호 (1509) 및 수정된 스피치 신호 (1511) (예컨대, 프로세싱, 열화, 왜곡, 강화 등이 행해진 스피치 신호) 사이의 명시적 감산 (explicit subtraction) 을 포함할 수도 있다. 다른 접근법들에서는, 원본 스피치 신호 (1509) 가 획득되지 않을 수도 있다. 이 접근법들은 비-침투적 접근법들 또는 방법들로서 지칭될 수도 있다. 비-침투적 접근법들에서는, 예를 들어, 원본 스피치 신호 (1509) 가 이용가능하지 않을 수도 있다 (예컨대, 수정된 스피치 신호 (1511) 및 원본 스피치 신호 (1509) 사이에 명시적 감산이 없을 수도 있음). 예를 들어, 본원에서 개시된 시스템들 및 방법들은 일부의 구현예들 및/또는 시나리오들에서 원본 스피치 신호 (1509) 를 획득하지 않으면서 사용될 수도 있다.

[0116] 침투적 및/또는 비-침투적 접근법들은 스피치 품질들 (예컨대, 왜곡-차원들) 을 객관적으로 측정하기 위해 사용될 수도 있다. 특징들은 주어진 애플리케이션에 따라 스피치의 유성음 및/또는 묵음 부분들에 대해 본원에서 설명된 바와 같이 추출될 수도 있다. 예를 들어, 특징들의 2 개의 세트들은 지각적 도메인 (perceptual domain) 에서 있을 수도 있고, 그러므로, 당연히 스피치 품질 측정을 위해 적당할 수도 있다. 모든 특징들이 본원에서 개시된 시스템들 및 방법들의 애플리케이션에 대해 필요하지는 않을 수도 있다는 것에 주목해야 한다. 간단한 합성 프로세스들이 트레이닝 및 예측을 위해 사용될 수도 있고 및/또는 복잡한 프로세스들이 개

선을 위해 사용될 수도 있다. 설명된 많은 특징들은 광범위한 스피치 왜곡들을 참작한다. 맵핑 모델은 추출된 특징들을 어떤 타입의 왜곡 점수들에 맵핑하기 위해 사용될 수도 있다. 하나의 간단한 접근법은 본원에서 설명된 바와 같이, 선형 회귀를 이용하고 있다. 더 높은 차수의 회귀들 또는 신경망들과 같은 더욱 복잡한 모델들이 추출된 특징들을 예측 점수들에 마찬가지로 맵핑하기 위해 채용될 수도 있다. 주어진 애플리케이션에 대하여, 트레이닝 프로세스는 맵핑 모델에 대한 적당한 파라미터들 및/또는 가중화를 세팅 (set up) 하기 위해 사용될 수도 있다. 예를 들어, 최적화된 가중화가 달성될 수도 있고, 이것은 주관적 점수들에 대하여 최소 에러를 갖는 예측들을 생성한다. 다음으로, 트레이닝된 모델은 (예를 들어, 트레이닝 폴 (training pool)에서가 아니라) 왜곡된 신호들에 직접적으로 적용될 수 있다. 예를 들어, 트레이닝된 맵핑 모델은 스피치 품질 점수들의 예측을 달성하기 위하여 테스팅되어야 할 스피치로부터 추출된 특징들과 함께 공급될 수도 있다.

[0117]

전자 디바이스 (1507) 는 수정된 스피치 신호 (1511) 를 획득할 수도 있다. 일부의 구성들에서, 전자 디바이스 (1507) 는 수정된 스피치 신호 (1511) 를 획득하기 위하여 원본 스피치 신호 (1509) 를 프로세싱할 수도 있다. 예를 들어, 전자 디바이스 (1507) 는 수정된 스피치 신호 (1511) 를 획득하기 위하여 원본 스피치 신호 (1509) 를 인코딩할 수도 있다 (및/또는 예를 들어, 결과적인 인코딩된 스피치 신호를 디코딩함). 추가적으로 또는 대안적으로, 전자 디바이스 (1507) 는 수정된 스피치 신호 (1511) 를 획득하기 위하여 원본 스피치 신호 (1509) 를 강화할 수도 있다. 예를 들어, 전자 디바이스 (1507) 는 원본 스피치 신호 (1509) 의 잡음-주입된 버전에 대해 잡음 억압을 수행할 수도 있다. "단일-채널" 스피치 신호는 스피치 신호를 임의의 하나의 시간 순간에서의 (예를 들어, 압력의) 단일 샘플로서 나타낼 수도 있다. 예를 들어, 단일-채널 스피치 신호는 공간적으로 구별되는 정보를 제공하는 다중 신호들을 포함하지 않을 수도 있다. 일부의 구성들에서 는, 예를 들어, 수정된 스피치 신호 (1511) 가 단일 마이크로폰으로 캡처되었던 원본 스피치 신호 (1509) 에 기초할 수도 있다.

[0118]

일부의 구성들에서, 전자 디바이스 (1507) 는 또 다른 디바이스 (예컨대, 저장 디바이스, 컴퓨터, 전화, 헤드셋, 인코더 등)로부터 수정된 스피치 신호 (1511) 를 획득할 수도 있다. 예를 들어, 전자 디바이스 (1507) 는 네트워크화된 전자 디바이스로부터 수정된 스피치 신호 (1511) 를 수신할 수도 있다.

[0119]

수정된 스피치 신호 (1511) 는 원본 스피치 신호 (1509) 에 기초할 수도 있다. 예를 들어, 수정된 스피치 신호 (1511) 는 원본 스피치 신호 (1509) 의 왜곡된 버전일 수도 있다. 원본 스피치 신호 (1509) 를 강화, 인코딩, 디코딩, 트랜스코딩, 송신, 수신 및/또는 에러-정정하는 것은, 수정된 스피치 신호 (1511) 의 하나 이상의 특성들이 원본 스피치 신호 (1509) 의 특성들과는 상이하도록, 원본 스피치 신호 (1509) 를 왜곡시킬 수도 있다.

[0120]

일부의 구성들에서, 수정된 스피치 신호 (1511) 및/또는 원본 스피치 신호 (1509) 는 주기들 (예컨대, "프레임들") 로 분할될 수도 있다. 예를 들어, 수정된 스피치 신호 (1511) 의 각각의 주기는 시간에 대한 다수의 샘플들을 포함할 수도 있다. 시간 주기들 또는 프레임들은 길이에 있어서 균일할 수도 있거나, 상이한 길이들일 수도 있다.

[0121]

특정 추출 모듈 (1529) 은 수정된 스피치 신호 (1511) 를 획득할 수도 있다. 선택적으로, 특정 추출 모듈 (1529) 은 추가적으로 원본 스피치 신호 (1509) 를 획득할 수도 있다. 특정 추출 모듈 (1529) 은 생리학적 달팽이관 모델 (1513), 섹션 분석 모듈 (1517), 벡터 추출 모듈 (1521), 및/또는 특정 결정 모듈 (1525) 을 포함할 수도 있다. 일부의 구성들에서, 특정 추출 모듈 (1529) 은 도 5 와 관련하여 설명된 연산 왜곡 추정기 (526) 내에 포함될 수도 있다. 다른 구성들에서, 특정 추출 모듈 (1529) 은 연산 왜곡 추정기 (526) 로부터 분리되어 있을 수도 있고, 연산 왜곡 추정기 (526) 를 포함할 수도 있다. 일부의 구성들에서, 특정 추출 모듈 (1529) 은 도 6 과 관련하여 설명된 특정 추출 모듈 (656) 의 예일 수도 있고, 및/또는 도 7 과 관련하여 설명된 특정 추출 모듈 (756) 의 예일 수도 있다.

[0122]

전자 디바이스 (1507) (예컨대, 특정 추출 모듈 (1529)) 는 하나 이상의 생리학적 달팽이관 모델들 (1513) 을 이용하여 스피치 (예컨대, 수정된 스피치 신호 (1511) 및/또는 원본 스피치 신호 (1509)) 를 프로세싱할 수도 있다. 생리학적 달팽이관 모델 (1513) 은 달팽이관의 하나 이상의 물리적 컴포넌트들의 응답을 모델링할 수도 있다. 예를 들어, 생리학적 달팽이관 모델 (1513) 은 내유모세포들 (IHC), 달팽이관 길이 및/또는 달팽이관의 유체역학들을 모델링할 수도 있다. 생리학적 달팽이관 모델 (1513) 을 이용하여 스피치 (예컨대, 수정된 스피치 신호 (1511)) 를 프로세싱하는 것은 (예를 들어, 기능적 모델과 비교하여) 인간 청각 기관에 의해 수행된 프로세싱을 더욱 정확하게 근사화할 수도 있다. 이것은 인간 청자에 의해 지각된 바와 같은 스피치

신호 품질을 더욱 정확하게 추정하는 것을 도울 수도 있다. 예를 들어, 생리학적 모델을 이용하여 수정된 스피치 신호 (1511) 를 프로세싱하는 것은 내유모세포 데이터를 제공할 수도 있다. 이것은 예를 들어, 기능적인 모델을 이용하여 근사화되는 기저막 데이터와는 상이하다.

[0123] 일부의 구성들에서, 생리학적 달팽이관 모델은 다음의 수학식들 및 설명들 중의 하나 이상에 따라 구현될 수도 있다. 그런 함수 적분 수학식 (Green's function integral equation) 은, 기저막 (BM) 속도에 대한 적분이 등자뼈 입자 속도 (stapes particle velocity) u_s X 스칼라 (scala) 높이 H (수직 속도는 상부 벽 상에서 제로 임) 와 동일하다는 것을 보장한다. 그런 함수 적분 수학식을 구하는 것은 본원에서 설명된 시스템들 및 방법들의 시작점이다. 그런 함수 적분 수학식은 다음에 의해 주어진다:

$$\phi(x, y) = \int_{x', y'} G(x, y | x', y') V_n(x', y') dx' dy' \quad (3)$$

[0124] 여기서, ϕ 는 전위의 속도이고, V_n 는 박스 내로 +로 정의된 스칼라 속도의 수직 성분이다.

[0125] 변수들의 다음의 리스트는 본원에서 개시된 수학식들에서 이용될 수도 있다:

[0126] $p(x, t); 2p(x, t)$ 달팽이구멍 (helicotrema) 에 관한 압력; BM 에 걸친 압력

[0127] x 등자뼈로부터 측정된, BM 을 따른 위치 변수

[0128] f 주파수 (Hz)

[0129] ω 라디안 주파수 = $2\pi f$

[0130] ρ 물의 밀도

[0131] n 물의 점성

[0132] $K_p(x)$ BM 파티션 스티프니스 파라미터 (partition stiffness parameter)

[0133] $R_p(x)$ BM 파티션 저항

[0134] m_p BM 파티션 질량

[0135] V_{ohc} OHC 전압

[0136] $T (V_{ohc})$ BM 장력

[0137] x_L BM 의 길이

[0138] y_H 달팽이관 스칼라의 높이

[0139] $X(f)$ 달팽이관 맵 함수 (BM 장소 대 주파수)

[0140] $Xz(f)$ 제 2 달팽이관 맵 함수

[0141] i $\sqrt{-1}$

[0142] $\xi(x, t), \dot{\xi}(x, t), \ddot{\xi}(x, t)$ BM 입자 변위, 속도, 가속도

[0143] $\zeta(x, t), \dot{\zeta}(x, t), \ddot{\zeta}(x, t)$ 섭모 (cilia) 변위, 속도, 가속도

[0144] $f(x, t)$ TM 흠

[0145] $us(t), u^s(t)$ 등자뼈 입자 속도 및 가속도

[0146] $(n+1, n, n-1)$ 이산 시간 (미래, 현재, 과거)

[0147] $*$ 공간적 컨볼루션

[0148] $L(t)$ "순간적" 라우드니스

[0150] Im, Ip 마스커 (masker) 및 프로브 (probe) 의 강도

[0151] 기본적인 탈팽이관 수학식은 다음에 의해 주어진다:

$$2\rho F(x) * \ddot{\xi}(x) + z_p(x,t)x\dot{\xi}(x) = 2\rho \dot{u}_s(t)(|x| - L))_L \quad (4)$$

[0153] 여기서, ρ 는 물의 밀도이고, $\ddot{\xi}(x)$ 는 BM 입자 가속도이고, *는 공간적 컨볼루션이고, \dot{u}_s 는 등자뼈 입자 가속도이다. 이 수학식은 Allen 및 Sondhi에 의해 작성된 수학식과 유사하지만, 덮개막 (Tectorial Membrane; TM) 힘을 포함하도록 수정된다. 덮개막 힘은 다음에 의해 정의된다:

$$f(x,t) \equiv g z_{tc}(x,t)x\dot{\xi}(x,t) \quad (5)$$

[0155] 여기서, $\dot{\xi}(x,t)$ 는 BM 입자 속도이다.

[0156] 수학식 (4)의 적분은 등자뼈 가속도 \dot{u}_s 가 주어질 경우에, 누군가가 BM 변위 $\xi(x,t)$ 를 구하도록 한다. 그러나, 최종 수학식들이 용이하게 풀어지지 않으므로, 이 엄격한 접근법으로 진행할 때에는 심각한 문제가 있다.

[0157] Allen 및 Sondhi의 수학식에서는, 임피던스 Z_t 가 존재하지 않았다 (즉, 무한대). 본 공식에서의 TM 임피

던스의 추가는 작은 선두 계수를 가지는, 수학식 (3)에서의 4 차 시간 항 $\frac{d^4\xi(x,t)}{dt^4}$ 을 초래한다. 이러한 수학식은 고전적으로, 특이 속성 (singular property)들을 가지는 스티프 차동 방정식 (stiff differential equation)으로서 알려져 있다.

[0158] p 및 ξ (즉, $z_p(t)$) 사이의 전달 함수의 차수를 확립하기 위한 방법은 라플라스 변환 (Laplace transforms)의 이용에 의한 것이다. 근사적인 파티션 임피던스 $Z_p^{(1)}$ 는 섬모 임피던스 $Z_c(x,f)$ 를 무시함으로써 획득된다. 위첨자 (1)은 수정된 (즉, 근사적인) Z_{tc} 를 표시한다. BM 변위 $\Xi(x,s)$ 에 대한 트랜스 (trans) BM 압력 $-2P(x,s)$ 의 측면에서 전체 파티션 임피던스를 작성하는 것은 $sZ_p(x,s)$ 를 제공한다:

$$-2\frac{P(s)}{\Xi(s)} = [m_b s^2 + R_O s + K_O] + \left[g^2 \frac{(m_c s^2 + r_c s + k_c)(m_t s^2 + r_t s + k_t)}{(m_c + m_t)s^2 + (r_c + r_t)s + (k_c + k_t)} \right] \quad (6)$$

[0160] 이것으로부터의 결론은, 섬모 임피던스 (cilia impedance)를 제거하는 것이 BM 임피던스에서의 5 퍼센트 변경으로부터 오직 설명된다는 것이다. 이에 따라, 그것이 수학식의 차수를 2로부터 4로 늘리지만, 수학식은 본질적으로 시간 수학식에서 2 차 (second order)로서 작용하고 있다. 이것은 매우 열악한 수치 속성을 초래하고, 이것은 수학식의 공식화에서 수반되어야 한다.

[0161] TM 및 섬모 임피던스 $Z_{tc}(x,t)$ 의 병렬은 프로세싱하기 위하여 이 시점에 근사화되어야 한다. 수학식 (4)

로부터 상대적으로 큰 (즉, 스티프) 임피던스 $Z_c \gg Z_t$ 를 낮춘 후에, 최종 수학식에 대한 솔루션이 획득될 수도 있다.

$$2\rho F(x) * \ddot{\xi}(x) + z_O(x,t)x\dot{\xi}(x) + g^2(x)z_t(x,t)x\dot{\xi}(x) = 2\rho \dot{u}_s(t)(|x| - L))_L \quad (7)$$

[0163] 수학식 (7)에서는, $Z_c \gg Z_t$ 라는 관찰로부터 뒤따르는 $Z_{tc} \equiv Z_t \parallel Z_c \approx Z_t$ 의 근사화가 있다.

[0164] 커널 함수 (kernel function) $F(x)$ 는 BM 가속도에 따라 동작한다. 또한, BM 가속도에 비례하는 BM 임피던스의 성분이 있다. 이 2 개의 가속도 항들은 수학식이 구해질 수도 있기 전에 그룹화되어야 한다. 임피던스가 2 차 형태일 때, 우리는 이 재그룹화가 용이하게 행해지는 것을 볼 수 있을 것이다. 일반적인 경우에 어떻게 진행할 것인지는 덜 분명하고, 그러므로, 2 차여서 임피던스에서 가속도 항을 노출하는 수학식 (7) 을 초래하는 BM 임피던스 함수에 대한 근사화를 행하는 것이 필요하였다.

[0165] 임피던스 함수들은 최소 위상이고, $M(s)$ 라는 매 최소 위상 함수는 다음의 형태로 기재될 수도 있다:

$$M(s) \equiv m_0 \frac{1 - R(s)}{1 + R(s)} \quad (8)$$

$$m_0 \equiv \int_0^t m(t) dt |_{t \rightarrow 0}$$

[0166] 여기서, $m(t)$ 이고, 여기서, $R(s)$ 는 $R(s)$ 에 대한 상기 수학식을 푸는 것에 의해 구해진 $M(s)/m_0$ 에 대응하는 반사율 (reflectance) 이다. 이 형태에서 임피던스 (즉, M) 를 기재하고, 그것을 시간 도메인에서 표현함으로써, 그런 함수 수학식 (3) 과 함께, 달팽이관 응답을 정의하는, $m(t)$ 에 대한 재귀적 시간 도메인 컨볼루션 수학식을 형성하는 것이 가능하다. 최종적인 달팽이관 수학식은 커널 함수 F 를 반전할 때에 임피던스에서의 가속도 성분을 참작해야 하므로, 이 외견상 복잡한 접근법이 필요하다.

[0167] 증강된 커널은 다음에 의해 정의된다:

$$Q(x) = \frac{1}{2} m_p ((\delta))_L(x) + 2\rho F(x) \quad (9)$$

$$m_p \approx m_0 + g^2 m_t \quad \text{이고, 그리고}$$

$$((\delta))_L(x) \equiv \sum_{k=-\infty}^{\infty} \delta(2Lk - x) \quad (10)$$

[0168] 질량 항은 BM 가속도 ξ 와의 공간에서의 컨볼루션으로서 재기재된다. 이것은 질량 항 및 커널이 그룹화되도록 한다. 이 표현은 컨볼루션 관계에 종속된다:

$$M_p \xi(x, t) = \frac{m_p}{2} ((\delta))_L(x) * \xi(x, t) \quad (11)$$

[0169] 이 수학식은 용이하게 검증된다.

[0170] 증강된 커널은 임피던스 질량 항 $m\xi$ 을 커널에 포함하기 위하여 정의될 수도 있는데, 이것은 양자가 BM 가속도 ξ 에 따라 동작하기 때문이다. 이 증강은, 실수인 홀수-고조파 대칭성 (odd-harmonic symmetry) 을 따르는, 전체 질량의 절반을 각각 갖는 2 개의 특이 포인트 (singular point) 들에서의 2 텔타 함수들에서 질량을 확산하는 것을 요구한다.

[0171] 증강된 커널의 측면에서, 모션 (motion) 의 BM 수학식은 다음과 같이 된다:

$$Q(x) * \xi + (R_O + g^2 r_I) \xi + (K_O + g^2 k_I) \xi = 2(|x| - L)_L \rho \dot{u}_s \quad (12)$$

[0172] 이것은 $Q(x)$ 를 반전함으로써 풀리며, 다음을 제공한다:

$$\xi = -Q^{-1} * [(R_O + g^2 r_I) \xi + (K_O + g^2 k_I) \xi - 2(|x| - L)_L \rho \dot{u}_s] \quad (13)$$

[0180] 일단 BM 변위가 수학식 (13) 의 적분에 의해 구해지면, 섬모 변위는, BM 대 섬모 변위 전달 함수에 의해 설명된 바와 같이, BM 변위 $\Xi(x, s) \leftrightarrow \xi(x, t)$ 가 주어질 경우에, 섬모 변위 $\Theta(x, s) \leftrightarrow \theta(x, t)$ 에 대하여 푸는 것에 의해 연산될 수도 있다:

$$\frac{\Theta(x, s)}{\Xi(x, s)} \equiv g(x) \frac{Z_t(x, s)}{Z_t(x|, s) + Z_c(x, s)} \quad (14)$$

[0181] 또는 시간 도메인에서:

$$[z_t(x, t) + z_c(x, t)]x\theta(x, t) = g(x)z_t(x, t)x\xi(x, t) \quad (15)$$

[0184] 우리는 간결함을 위하여, 수학식 (13) 에서 BM 변위에 대해 풀 때에 Z_t 가 무시되었지만, 이 수학식에서 그것을 무시하는 것이 필요하지도 않고 적당하지도 않다는 것을 반복한다. 수학식 (13) 의 경우에는, 그것을 무시하기 위하여 정당화되는 작은 항이다. 섬모 수학식에서는, 그것이 크고 임계적으로 중요한 임피던스이다.

[0185] 유한 차분 이산-시간 달팽이관 수학식 (finite difference discrete-time cochlear equation) 은 다음과 같다:

$$\begin{aligned} \frac{Q * [\xi_{n+1} - 2\xi_n + \xi_{n-1}]}{T^2} + \frac{(R_O + g^2 r_t)(\xi_n - \xi_{n-1})}{T} + \\ (K_O + g^2 k_t)\xi_n = \frac{2\rho(|x| - L)u_n - u_{n-1}}{T} \end{aligned} \quad (16)$$

[0186] 푸리에 변환 (Fourier transforms) 에 의해 수치 공간적 컨볼루션 (numerical spatial convolution) 들을 행할 때, 길이 dx 의 엘리먼트는 $\Delta \equiv L/K$ 로서 정의되고, 여기서, 정수 K 는 2 의 제곱이고, L 은 달팽이관의 길이이다. 이에 따라, 기저막 좌표 $x=0\dots L$ 는 다음과 같다:

$$x_k = k\Delta, \quad k = 0\dots K \quad (17)$$

[0188] 길이 $4L$ 에 대한 "홀수-고조파" 원형 컨볼루션에 대한 공식, 즉,

$$Q((x))_L * \xi((x))_L \equiv \int_{x=0}^{4L} Q(x')\xi(x-x')dx \quad (18)$$

[0191] 커널 함수 샘플 값들 Q_k 및 테스트 함수 샘플 값 ξ_k 사이의 이산적인 경우 (discrete case) 에 대하여, 여기서, k 는 공간적 인덱스를 나타내고 (시간 인덱스 n 은 이 예에 대하여 억압됨), 다음에 의해 주어진다:

$$\Delta \sum_{k'=0}^{4K-1} Q_{k'} \xi_{k-k'} \quad (19)$$

[0192] 대역 제한된 함수들에 대하여, 샘플 값들은 샘플 시간들에서 평가된 함수의 간단한 값들이며, 즉, $\xi_k \equiv \xi(x_k)$ 이다. $k=0$ 일 때 (즉, $x = 0$), 커널 샘플 값 F_0 은 무한대이다. 불운하게도, 이 로그 특이성 (log singularity) 은 적분 가능하다. 이에 따라, 우리는 Δ 에 의해 나누어진 로그 특이성에 대해 적분함으로써 제로에서 샘플 값을 정의한다.

$$F_0 \equiv \frac{L}{2H} - \frac{1}{\Delta\pi} \int_{-\Delta/2}^{\Delta/2} \log\left(1 - e^{\pi|x|/H}\right) dx \quad (20)$$

[0194] 적분은 테일러 급수 (Taylor series) 에서 지수를 확장함으로써, 그리고 다음으로, 최저 차수 항을 적분함으로써 행해질 수도 있고, 이것은 다음을 제공한다:

[0196]
$$F_0 = \frac{L}{2H} - \frac{1}{\pi} \left[\log \left(\frac{\pi \Delta}{2H} \right) - 1 \right] \quad (21)$$

[0197] 유사한 방법으로, 증강된 커널에서 2 개의 특이 질량 항들의 샘플 값들은 다음으로서 유사하게 정의되어야 한다:

[0198]
$$m_{eff} \equiv \frac{1}{\Delta} \int_{-\Delta/2}^{\Delta/2} m_p \delta(x) dx = \frac{m_p}{\Delta} \quad (22)$$

[0199] 정의들로부터, 우리는 다음을 구한다:

[0200]
$$Q_0 = \frac{m_p}{2\Delta} + \frac{\rho L}{H} - \frac{2\rho}{\pi} \left[\ln \left(\frac{\pi \Delta}{2H} \right) - 1 \right] \quad (23)$$

[0201] 반면에, $1 \leq k \leq K-1$ 에 대해서는, 다음과 같다:

[0202]
$$Q_k = \frac{\rho L}{H} \left(1 - \frac{k}{K} \right) - \frac{2\rho}{\pi} \ln \left(1 - e^{-k\pi\Delta/H} \right) \quad (24)$$

[0203] 훌수-고조파 대칭성으로 인해, $Q_k = Q_{-k} = -Q_{2K-k}$ 이다. 최종적으로, 역 커널 Q^{-1} 은 다음으로서 연산된다:

[0204]
$$Q^{-1} \equiv F^{-1}[1/FQ] \quad (25)$$

[0205] 여기서, F 는 길이 4L 의 FFT 이다.

[0206] 유체 체적 (fluid volume) 의 보존이 지속되어야 하므로, BM 에 따른 체적 적분은 등자뼈 체적 속도 $H\dot{u}_s$ 와 동일해야 한다. 이 중요한 수치적인 제어는 등자뼈 입력을 제로로 설정함으로써 (즉, $\dot{u}_s = 0$), 그리고 t=0 에서의 체적 속도를 $x=L/2$ 에서 1 로 설정함으로써, 그리고 이 초기 조건을 전파함으로써 최종적인 프로그램에서 테스트될 수도 있다. 이 테스트에 대하여, BM 의 체적 속도는 펄스가 $x=L$ 에 도달할 때까지 1 로 유지되어야 한다.

[0207] ξ_{n+1} 에 대하여 수학식 (16) 을 푸는 것은 다음을 제공한다:

[0208]
$$\xi_{n+1} = 2\xi_n - \xi_{n-1} - T^2 Q^{-1} * b_n \quad (26)$$

[0209] 여기서:

[0210]
$$b_n \equiv \frac{(R_O + g^2 r_t) (\xi_n - \xi_{n-1})}{T} + (K_O + g^2 k_t) \xi_n - \frac{2(|x| - L) L (u_n - u_{n-1}) \rho}{T} \quad (27)$$

[0211] 수학식들 (26) 및 (27) 은 달팽이관 응답의 최종 수치 솔루션이고, 달팽이관 회귀 수학식 (BM 및 유체 사이의 피드백 루프) 을 나타낸다.

[0212] 섬모 변위는 최종적으로 수학식 (13) 으로부터 계산될 수도 있고, 이것은 다음과 같다:

[0213]
$$(m_c + m_t) \ddot{\theta} + (r_c + r_t) \dot{\theta} + (k_c + k_t) \theta = g [m_t \ddot{\xi} + r_t \dot{\xi} + k_t \xi] \quad (28)$$

[0214] 또는 이산 시간 형태에서는:

$$\frac{(m_c + m_t)}{T^2}(\theta_n + 2\theta_{n-1} + \theta_{n-2}) + \frac{(r_c + r_t)}{T}(\theta_{n-1} - \theta_{n-2}) + (k_c + k_t)\theta_{n-1} = g \left[\frac{(m_t)}{T^2}(\xi_n + 2\xi_{n-1} + \xi_{n-2}) + \frac{(r_t)}{T}(\xi_{n-1} - \xi_{n-2}) + (k_t)\xi_{n-1} \right] \quad (29)$$

[0215] [0216] X_n 에 대하여 수학식 (29)를 푸는 것은 다음과 같다:

$$\theta_n = 2\theta_{n-1} - \theta_{n-2} - T \frac{(r_c + r_t)}{(m_c + m_t)}(\theta_{n-1} - \theta_{n-2}) - T^2 \frac{(k_c + k_t)}{(m_c + m_t)}\theta_{n-1} - \frac{gm_t}{(m_c + m_t)} \left[(\xi_n + 2\xi_{n-1} + \xi_{n-2}) + T \frac{r_t}{m_t}(\xi_{n-1} - \xi_{n-2}) + T^2 \frac{k_t}{m_t}\xi_{n-1} \right] \quad (30)$$

[0217] [0218] 공통 항들에서 재 배열되고:

$$\theta_n = -a_1\theta_{n-1} - a_2\theta_{n-2} + b[\xi_n + b_1\xi_{n-1} + b_2\xi_{n-2}] \quad (31)$$

[0219] [0220] 이것은 검사에 의하여 계수들 $a \equiv [a_1, a_2]$ and $b \equiv [b, b_1, b_2]$ 을 정의한다:

$$a_1 = -2 + T \frac{(r_c + r_t)}{(m_c + m_t)} + T^2 \frac{(k_c + k_t)}{(m_c + m_t)} \quad (32)$$

$$a_2 = 1 - T \frac{(r_c + r_t)}{(m_c + m_t)} \quad (33)$$

$$b_1 = -2 + T \frac{r_t}{m_t} + T^2 \frac{k_t}{m_t} \xi_{n-1} \quad (34)$$

$$b_2 = 1 + T \frac{r_t}{m_t} \quad (35)$$

$$b = \frac{gm_t}{(m_c + m_t)} \times \frac{A_{fp}}{HW_{bm}} \quad (36)$$

[0221] [0222] 계수 벡터 b 는 수학식 (27)에 의해 정의된 b_n 에 상관되지 않는 것에 주목한다.

[0223] 마지막 단계는 Rabiner 및 Gold에서 설명된 바와 같이, 임펄스 불변 변환 (impulse invariance transformation)에 기초하여, s 장소로부터 z 평면까지의 제곱근 변환 (root transformation)으로 최상으로 취해진다. 이 변환은 디지털 및 아날로그 샘플 값의 불변성에 기초한다. 다시 말해서, $\xi_n \equiv \xi(nT)$ 는 s 내지 Z 도메인들 사이의 맵핑을 결정한다. 이 임펄스 불변 계수들은 더욱 정확하고, 디지털 솔루션을 더 높은 주파수들로 확장한다 (즉, 최대 주파수에 웨이브 근접할수록, 나이퀴스트 (Nyquist) 샘플링 레이트들은 $1/2$ 이다).

[0224] 2 차 디지털 공진기는 임펄스 불변 원뿔 형태 (impulse invariant conical form)에 의해 극 (pole)들 s_p 및 제로 (zero)들 s_z 의 측면에서 공통적으로 정의된다:

[0225] 이 시스템은 $s_p = \sigma_p \pm j\omega_p$ 및 $s_z = \sigma_z \pm j\omega_z$ 에 의해 결정된 한 쌍의 복소 공액 풀 및 제로 라디안 주파수들을

가지며, 댐핑 파라미터 (damping parameter)들은 $R_p = e^{\sigma p T}$ 및 $R_z = e^{\sigma z T}$ 에 의해 정의된다. 나이퀴스트

[0225] 트 주파수는 $f_{Nyquist}=1/T$ 에 의해 샘플 주기 T 에 관련된다.

$$\theta_n + 2R_p \cos(\omega_p T) \theta_{n-1} + R_p^2 \theta_{n-2} = H_0 [\xi_n + 2R_z \cos(\omega_z T) \xi_{n-1} + R_z^2 \xi_{n-2}] \quad (37)$$

[0226] [0227] 계수들 a, b 의 이 2 개의 세트들은 아날로그 도메인에서 정의된 보조 파라미터들을 통해 최상으로 정의된다:

$$\sigma_p = -0.5(r_c + r_t)/(m_c + m_t) \quad (38)$$

$$\sigma_z = -0.5r_t/m_t \quad (39)$$

$$\omega_p = \sqrt{(k_c + k_t)/(m_c + m_t)} \quad (40)$$

$$\omega_z = \sqrt{k_t/m_t} \quad (41)$$

$$g = 1 \quad (42)$$

[0228] [0229] 임펄스 불변성에 기초하여 디지털 공진기 계수 정의들을 초래하면:

$$R_p = e^{\sigma_p T} \quad (43)$$

$$R_z = e^{\sigma_z T} \quad (44)$$

$$a_1 = -2R_p * \cos\left(\sqrt{\omega_p^2 - \sigma_p^2} T\right) \quad (45)$$

$$a_2 = R_p^2 \quad (46)$$

$$b_1 = -2R_z * \cos\left(\sqrt{\omega_z^2 - \sigma_z^2} T\right) \quad (47)$$

$$b_2 = R_z^2 \quad (48)$$

[0230] [0231] 유한 차분 및 임펄스 불변성 계수들 사이에는 간단한 관계가 있다. 임펄스 불변성에 기초한 a_2 가 테일러 급수에서 T 에 있어서 차수 1 로 확장될 경우, 덜 정확한 유한 차분 a_2 는 다음의 결과로 된다:

$$a_2 = e^{-2\sigma_a T} = 1 - 2\sigma_a T + O(T^2) \quad (49)$$

[0232] [0233] [0234] 계수들 k_t , k_c , m_t 및 m_c 를 구하기 위하여, 우리는 문헌으로부터 알려져 있는, $w_p(x)$, $w_x(x)$ 및 $w_{cf}(x)$ 에 대한 정의들로부터 결정된 3 개의 다음의 수학식들을 풀다:

$$\omega_z^2(x) \equiv \frac{k_t}{m_t} = [0.17\omega_p^{1.11}]^2 \quad (50)$$

$$\omega_p^2(x) \equiv \frac{k_t + k_c}{m_t + m_c} \quad (51)$$

[0235]

[0236] 그리고, 후미 (tail) 에서 투닝 곡선 기울기 (tuning curve slope) 를 결정하는 수학식은 다음과 같다:

$$\frac{k_t}{k_c + k_t} = 0.01 \quad (52)$$

[0237] Poiseuille 공식으로부터의 섬모 파라미터들은 다음과 같다:

$$r_c = \frac{\mu W_c}{h W_{bm}} \quad (53)$$

$$m_c = \frac{\rho W_c h}{3 W_{bm}} \quad (54)$$

[0239] 그리고 최종적으로, TM 질량 수학식은 다음과 같다:

$$(m_c + m_t) = .02 \quad (55)$$

[0240] 일부의 구성들에서, 생리학적 달팽이관 모델 (1513) 은 달팽이관 길이에 대한 응답 데이터 (예컨대, 내유모세포 데이터) 제공할 수도 있다. 예를 들어, 다수의 장소 포인트들 (예컨대, 샘플들) 은 스피치의 각각의 샘플에 대하여 (예컨대, 수정된 스피치 신호 (1511) 의 각각의 샘플에 대하여) 달팽이관 길이에 대한 생리학적 달팽이관 모델 (1513) 의 응답을 모델링할 수도 있다. 장소 포인트들은 달팽이관의 길이에 따른 장소들에 대응할 수도 있다. 달팽이관의 길이에 따른 장소들은 특정 주파수들에서의 사운드 (sound) 들에 대응할 수도 있고 이 사운드들에 대해 응답할 수도 있다. 예를 들어, 장소 포인트들의 세트에서의 제 1 장소 포인트는 대략 20 퀄로헤르쯔 (kHz) 범위에서의 사운드들에 대응할 수도 있는 반면, 최후 장소 포인트는 매우 낮은 주파수들 (예컨대, 12 헤르쯔 (Hz)) 에서의 사운드들에 대응할 수도 있다. 따라서, 생리학적 달팽이관 모델 (1513) 은 그것이 각각의 스피치 샘플에 대한 다수의 장소 포인트들을 생성할 수도 있다는 점에서 "오버샘플링 (oversample)" 할 수도 있다. 일부의 구성들에서, 생리학적 달팽이관 모델 (1513) 은 수정된 스피치 신호 (1511) 의 각각의 샘플에 대한 장소 포인트들의 세트를 생성할 수도 있다.

[0241] [0242] 생리학적 달팽이관 모델 (들) (1513) 은 출력 (1515) 을 제공할 수도 있다. 출력 (1515) 은 시간의 길이에 대한 장소 포인트들을 포함할 수도 있다. 예를 들어, 출력 (1515) 은 시간의 길이에 대한 각각의 스피치 샘플에 대한 장소 포인트들의 세트를 포함할 수도 있다. 출력 (1515) 은 섹션 분석 모듈 (1517) 에 제공될 수도 있다. 섹션 분석 모듈 (1517) 은 생리학적 달팽이관 모델 (1513) 의 출력 (1515) 의 섹션 (section) 들을 분석할 수도 있다. 예를 들어, 섹션 분석 모듈 (1517) 은 출력 (1515) (예컨대, 출력의 부분들) 을 다수의 섹션들 (1519) 로 그룹화 (예컨대, 분할) 할 수도 있다. 섹션들 (1519) 의 각각은 달팽이관 길이에 따른 장소 범위에 대응할 수도 있다. 일부의 구성들에서, 섹션 분석 모듈 (1517) 은 출력 (1515) 을 4 개의 섹션들 (1519) 로 그룹화할 수도 있다. 예를 들어, 제 1 섹션은 1 부터 150 까지의 장소 포인트들을 포함할 수도 있고, 제 2 섹션은 151 부터 275 까지의 장소 포인트들을 포함할 수도 있고, 제 3 섹션은 276 부터 450 까지의 장소 포인트들을 포함할 수도 있고, 제 4 섹션은 451 부터 512 까지의 장소 포인트들을 포함할 수도 있다. 섹션들의 각각은 시간의 길이에 대한 장소 포인트들 (예컨대, N 샘플들) 을 포함할 수도 있다. 다른 섹션 (1519) 사이즈들이 사용될 수도 있다는 것에 주목해야 한다.

[0243] [0244] 섹션들 (1519) 은 벡터 추출 모듈 (1521) 에 제공될 수도 있다. 벡터 추출 모듈 (1521) 은 각각의 섹션 (1519) 에 대한 벡터들 (1523) 을 추출할 수도 있다. 특히, 벡터 추출 모듈 (1521) 은 각각의 섹션 (1519) 에 대한 장소-기반 분석 벡터 및 시간-기반 분석 벡터를 추출할 수도 있다. "장소-기반 분석 벡터" 는 장소에 대한 다수의 값들을 포함하는 벡터이다. 예를 들어, 벡터 추출 모듈 (1521) 은 섹션 (1519) 을 시간에 대하여 평균화함으로써 장소-기반 분석 벡터를 결정할 수도 있다 (이것은 예를 들어, 장소에 대한 다수의 값들을 갖는 벡터를 산출함). "시간-기반 분석 벡터" 는 시간에 대한 다수의 값들을 포함하는 벡터이다. 예를 들어, 벡터 추출 모듈 (1521) 은 섹션 (1519) 을 장소에 대하여 평균화함으로써 시간-기반 분석 벡터를 결정할 수도 있다 (이것은 예를 들어, 시간에 대한 다수의 값들을 갖는 벡터를 산출함).

[0245] 벡터들 (1523) (예컨대, 하나 이상의 시간-기반 분석 벡터들 및 하나 이상의 장소-기반 분석 벡터들) 은 특징 결정 모듈 (1525) 에 제공될 수도 있다. 특징 결정 모듈 (1525) 은 각각의 벡터 (1523) (예컨대, 분석 벡터)로부터 하나 이상의 특징들 (1527) 을 결정할 수도 있다. 특징 (1527) 은 벡터 (1523) 특성을 수량화 하는 메트릭일 수도 있다. 특징들 (1527) 의 예들은 평균들 (예컨대, 평균치들), 중앙값 (median) 들, 기하학적 시프트 (geometric shift) 들, 조화 평균 (harmonic mean) 들, 표준 편차 (standard deviation) 들, 비대칭도 (skewness) 들, 분산 (variance) 들 및 그 외의 것들을 포함한다. 특징 결정 모듈 (1525) 은 각각의 벡터 (1523) 로부터 이 종류의 특징들 (1527) 중의 하나 이상을 결정할 수도 있다. 일부의 구성들에서, 특징 결정 모듈 (1525) 은 각각의 벡터 (1523) 의 평균, 중앙값, 기하학적 시프트, 조화 평균, 표준 편차 및 비대칭도를 결정할 수도 있다.

[0246] 일부의 구성들에서, 전자 디바이스 (1507) 는 하나 이상의 특징들 (1527) 에 기초하여 왜곡을 추정할 수도 있다. 예를 들어, 전자 디바이스 (1507) 는 하나 이상의 특징들 (1527) 에 기초하여 하나 이상의 왜곡들을 추정하는 왜곡 추정 모듈 (도시되지 않음) 을 포함할 수도 있다. 예를 들어, 왜곡 추정 모듈은 왜곡을 추정하기 위하여 특징들 (1527) 중의 하나 이상 및 하나 이상의 가중치들에 기초하여 회귀 (예컨대, 선형 회귀, 다항식 회귀 (polynomial regression), 2 차 회귀 (second order regression), 비선형 회귀 등) 를 수행할 수도 있다. 일부의 구성들에서, 전자 디바이스 (1507) 는 본원에서 설명된 바와 같은 하나 이상의 왜곡들에 기초하여 하나 이상의 품질들 (예컨대, 포어그라운드 품질, 백그라운드 품질, 전체 품질 등) 을 추정할 수도 있다.

[0247] 도 16 은 특징 추출을 위한 방법 (1600) 의 하나의 구성을 예시하는 흐름도이다. 전자 디바이스 (1507) 는 방법 (1600) 의 하나 이상의 단계들, 기능들 및/또는 절차들을 수행할 수도 있다.

[0248] 전자 디바이스 (1507) 는 하나 이상의 생리학적 달팽이관 모델들 (1513) 을 이용하여 스피치 (예컨대, 수정된 스피치 신호 (1511) 및/또는 원본 스피치 신호 (1509)) 를 프로세싱할 수도 있다 (1602). 이것은 위에서 설명된 바와 같이 달성될 수도 있다. 예를 들어, 전자 디바이스 (1507) 는 수정된 스피치 신호 (1511) 및/또는 원본 스피치 신호 (1509) 에 기초하여 하나 이상의 생리학적 달팽이관 모델들 (1513) 의 응답을 결정할 수도 있다. 예를 들어, 전자 디바이스 (1507) 는 스피치 신호 (예컨대, 수정된 스피치 신호 (1511)) 의 각각의 샘플에 대한 장소 포인트들 (예컨대, 샘플들) 의 세트를 결정할 수도 있다. 생리학적 달팽이관 모델 (1513) 의 출력 (1515) 은 시간의 길이에 대한 장소 포인트들의 세트들 (예컨대, N 샘플들) 을 포함할 수도 있다.

[0249] 전자 디바이스 (1507) 는 생리학적 달팽이관 모델 (1513) 의 출력 (1515) 의 색션들을 분석할 수도 있다 (1604). 이것은 위에서 설명된 바와 같이 달성될 수도 있다. 예를 들어, 전자 디바이스 (1507) 는 출력 (1515) 의 부분들을 다수의 색션들 (1519) (예컨대, 4 개의 색션들 또는 또 다른 수의 색션들) 로 그룹화 (예컨대, 분할) 할 수도 있다. 색션들 (1519) 의 각각은 특정 사이즈 (예컨대, 장소 포인트들의 수 X N 샘플들의 수) 를 가질 수도 있다.

[0250] 전자 디바이스 (1507) 는 각각의 색션 (1519) 에 대한 벡터들 (1523) 을 추출할 수도 있다 (1606). 특히, 전자 디바이스 (1507) 는 각각의 색션 (1519) 에 대한 장소-기반 분석 벡터 및 시간-기반 분석 벡터를 추출할 수도 있다. 이것은 위에서 설명된 바와 같이 달성될 수도 있다. 예를 들어, 전자 디바이스 (1507) 는 장소-기반 분석 벡터 (1523) 를 생성하기 위하여 시간에 대하여 색션 (1519) 을 평균화할 수도 있고, 시간-기반 분석 벡터 (1523) 를 생성하기 위하여 장소에 대하여 색션 (1519) 을 평균화할 수도 있다.

[0251] 전자 디바이스 (1507) 는 각각의 벡터 (1523) (예컨대, 분석 벡터) 로부터 하나 이상의 특징들 (1527) 을 결정할 수도 있다 (1608). 이것은 위에서 설명된 바와 같이 달성될 수도 있다. 예를 들어, 전자 디바이스 (1507) 는 각각의 벡터 (1523) 의 평균, 중앙값, 기하학적 시프트, 조화 평균, 표준 편차 및 비대칭도를 결정할 수도 있다.

[0252] 도 17 은 생리학적 달팽이관 모델 출력의 하나의 예를 예시하는 그래프이다. 특히, 그래프는 스피치의 유성 음 부분에 대한 생리학적 달팽이관 모델 출력의 하나의 예를 예시한다. 그래프의 축들은 (샘플들에서의) 장소 (1735), (밀리초 (ms) 인) 시간 (1731) 및 내유모세포 (1733) (입력 스피치 신호의 진폭) 를 포함한다. 생리학적 달팽이관 모델 (CM) 은 알려진 음향심리학적 마스킹 모델들 (Psychoacoustic Masking Models; PMM) 보다 더욱 정확하다. 구체적으로, 생리학적 달팽이관 모델은 더 높은 시간-공간 분해능 (time-space resolution) 을 제공할 수도 있다. 생리학적 달팽이관 모델은 인간 지각을 근사화하는 사운드의 측정을 가능하게 한다. 이것은 스피치 왜곡들에 대한 인간의 지식을 더욱 양호하게 반영하는 스피치 품질 점수들의 결정을 가능하게 할 수도 있다. CM 출력의 추세는 그래프 내에 포함된 2 개의 라인들 (1737a 내지 1737b)

에 의해 표시된다.

[0253] 도 17에서 예시된 예에서, CM 출력은 3 개의 축들을 가진다. 시간 축은 간단하고, 여기서, 매 입력은 하나의 출력을 가진다. 도 17은 3900 내지 4150 밀리초 (ms) 사이에서 시간 (1731) 도표를 예시한다.

킬로헤르츠 (kHz) 샘플링 레이트를 갖는 입력 스피치에 대하여, 이것은 실제적으로 8 포인트들/ms를 산출한다. 장소 (1735) 축은 15 내지 20000 헤르츠 (Hz)에 (비선형적으로) 맵핑되는 512 포인트들을 함께 제공한다.

도 17은 더욱 양호한 예시를 위하여 장소 (1735) 축을 따라 300 내지 400 으로부터 도표화된다. IHC 축 (1733)은 입력 진폭이다.

[0254] 도 18은 침투적 특징 추출 모듈 (1829)의 하나의 구성을 예시하는 블록도이다. 침투적 특징 추출 모듈 (1829)은 도 15와 관련하여 설명된 특징 추출 모듈 (1529)의 하나의 예일 수도 있다. 침투적 특징 추출 모듈 (1829)은 지연 추정 모듈 (1837), 달팽이관 모델들 A 내지 B (1813a 내지 1813b), 부분 선택 모듈 (1843), 선택 분석 모듈 (1817), 벡터 추출 모듈 (1821) 및/또는 특징 결정 모듈 (1825)을 포함할 수도 있다.

원본 스피치 신호 (1809) 및 수정된 스피치 신호 (1811) (예컨대, 원본 스피치 신호 (1809)의 열화된 버전)는 특징 추출을 위한 입력으로서 이용될 수도 있다. 다시 말해서, 침투적 특징 추출 모듈 (1829)은 하나 이상의 특징들 (1827)을 결정할 수도 있다. 하나 이상의 특징들 (1827)은 수정된 스피치 신호 (1811)에 대한 하나 이상의 왜곡들 (1849)을 추정 (예컨대, 예측)하기 위해 이용될 수도 있다. 원본 스피치 신호 (1809) 및/또는 수정된 스피치 신호 (1811)의 길이는 하나의 프레임, 다수의 프레임들, 또는 임의의 적당한 시간 길이 (예컨대, 1, 2, 5, 10 초 등) 일 수도 있다. 추가적으로 또는 대안적으로, 원본 스피치 신호 (1809) 및/또는 수정된 스피치 신호 (1811)의 길이는 스피치 자체 (예컨대, 전체 문장)에 기초할 수도 있다.

예를 들어, 원본 스피치 신호 및/또는 수정된 스피치 신호 (1811)의 길이는 (예를 들어, 무선 네트워크의 조작자에 의해, 및/또는 사용자에 의해) 구성 가능할 수도 있다.

[0255] 원본 스피치 신호 (1809) 및 수정된 스피치 신호 (1811)는 선택적으로 지연 추정 모듈 (1837)에 제공될 수도 있다. 지연 추정 모듈 (1837)은 원본 스피치 신호 (1809) 및 수정된 스피치 신호 (1811) 사이의 지연을 추정할 수도 있다. 예를 들어, 지연 추정 모듈 (1837)은 (예를 들어, 지연이 있을 경우) 지연을 결정하기 위하여 원본 스피치 신호 (1809) 및 수정된 스피치 신호 (1811) 사이의 상관을 수행할 수도 있다. 지연 추정 모듈 (1837)은 수정된 스피치 신호 (1811) 및 원본 스피치 신호 (1809)를 정렬하기 위하여 수정된 스피치 신호 (1811), 원본 스피치 신호 (1809), 또는 양자를 지연시킬 수도 있다. 예를 들어, 지연 추정 모듈 (1837)이 수정된 스피치 신호 (1811)가 원본 스피치 신호 (1809)에 관련된 다수의 샘플들만큼 지연되는 것으로 추정할 경우, 지연 추정 모듈 (1837)은 원본 스피치 신호 (1809) 및 수정된 스피치 신호 (1811)를 정렬하기 위하여 원본 스피치 신호 (1809)를 지연시킬 수도 있다. 따라서, 지연 추정 모듈 (1837)은 정렬된 원본 스피치 신호 (1839) 및 정렬된 수정된 스피치 신호 (1841)를 제공할 수도 있고, 여기서, 정렬된 원본 스피치 신호 (1839) 및 정렬된 수정된 스피치 신호 (1841)는 시간 (예컨대, 샘플들)에 있어서 정렬된다.

[0256] 정렬된 원본 스피치 신호 (1839)는 달팽이관 모델 A (1813a)에 제공될 수도 있다. 정렬된 수정된 스피치 신호 (1841)는 달팽이관 모델 B (1813b)에 제공될 수도 있다. 달팽이관 모델들 A 내지 B (1813a 내지 1813b)는 도 15와 관련하여 설명된 생리학적 달팽이관 모델 (1513)의 예들일 수도 있다. 달팽이관 모델들 A 내지 B (1813a 내지 1813b)는 도 15와 관련하여 설명된 바와 같이 기능할 수도 있다. 따라서, 달팽이관 모델 A (1813a)는 출력 A (1815a)를 생성할 수도 있고, 달팽이관 모델 B (1813b)는 출력 B (1815b)를 생성할 수도 있다.

[0257] 평균 인간 달팽이관은 길이에 있어서 약 3.5 센티미터 (cm)이고, 여기서, 달팽이관 상의 각각의 포인트는 상이한 주파수에 대해 응답한다. 예를 들어, 최후 장소 포인트는 근사적인 최저 지각 가능한 주파수에 대응할 수도 있고, 제 1 장소 포인트는 근사적인 최고 지각 가능한 주파수에 대응할 수도 있다. 일부의 구성들에서, 달팽이관 모델들 A 내지 B (1813a 내지 1813b)는 달팽이관의 길이를 다수의 장소 포인트들 (예컨대, 512 개의 장소 포인트들 또는 샘플들)로 이산화 (discretize) 할 수도 있다. 예를 들어, 원본 스피치 신호 (1809)의 각각의 샘플에 대하여, 달팽이관 모델 A (1813a)는 512 개의 장소 포인트들을 생성할 수도 있고, 수정된 스피치 신호 (1811)의 각각의 샘플에 대하여, 달팽이관 모델 B (1813b)는 512 개의 장소 포인트들을 생성할 수도 있다. 출력들 A 내지 B (1815a 내지 1815b)는 샘플들의 범위에 대한 장소 포인트들의 세트를 포함할 수도 있다 (예컨대, N). 예를 들어, 출력 A (1815a)는 원본 스피치 신호 (1809) (예컨대, 정렬된 원본 스피치 신호 (1839))에 대응하는 장소 포인트들의 N 개의 세트들을 포함할 수도 있고, 출력 B (1815b)는 수정된 스피치 신호 (1811) (예컨대, 정렬된 수정된 스피치 신호 (1841))에 대응하는 장소 포인트들의 N 개의 세트들

을 포함할 수도 있다.

[0258] 출력들 A 내지 B (1815a 내지 1815b) 는 부분 선택 모듈 (1843) 에 제공될 수도 있다. 부분 선택 모듈 (1843) 은 달팽이관 모델들 A 내지 B (1813a 내지 1813b) 의 출력들 A 내지 B (1815a 내지 1815b) 로부터 부분들 A 내지 B (1845a 내지 1845b) (예컨대, "관심 부분들") 를 선택할 수도 있다. 예를 들어, 원본 스피치 신호 (1809) 및 수정된 스피치 신호 (1811) 의 유성음 부분들이 선택될 수도 있다. 예를 들어, 스피치 신호들의 유성음 부분들은 지각된 스피치 품질에 상당히 기여할 수도 있는 모음 (vowel sound) 들을 포함할 수도 있다. 일부의 구성들에서, 부분 선택 모듈 (1843) 은 출력 A (1815a) 및/또는 출력 B (1815b) 의 에너지를 결정할 수도 있다. 부분 A (1845a) 는 출력 A (1815a) 의 샘플들의 범위로서 선택될 수도 있고, 여기서, 출력 A (1815a) 의 에너지는 제 1 임계점 (first threshold) 을 초과한다. 부분 B (1845b) 는 출력 B (1815b) 의 샘플들의 범위로서 선택될 수도 있고, 여기서, 출력 B (1815b) 의 에너지는 제 2 임계점을 초과한다. 제 1 및 제 2 임계점들은 서로 동일하거나 상이할 수도 있다. 부분 선택 모듈 (1843) 은 부분들 A 내지 B (1845a 내지 1845b) 를 결정하기 위해 다른 음성 활성 검출 (voice activity detection; VAD) 접근법들 또는 모듈들을 사용할 수도 있다.

[0259] 출력들 A 내지 B (1815a 내지 1815b) 또는 부분들 A 내지 B (1845a 내지 1845b) 는 섹션 분석 모듈 (1817) 에 제공될 수도 있다. 섹션 분석 모듈 (1817) 은 도 15 와 관련하여 설명된 섹션 분석 모듈 (1517) 의 하나의 예일 수도 있다. 섹션 분석 모듈 (1817) 은 달팽이관 모델들 A 내지 B (1813a 내지 1813b) 의 출력들 A 내지 B (1815a 내지 1815b) 또는 부분들 A 내지 B (1845a 내지 1845b) 의 섹션들을 분석할 수도 있다. 예를 들어, 섹션 분석 모듈 (1817) 은 출력들 A 내지 B (1815a 내지 1815b) 또는 부분들 A 내지 B (1845a 내지 1845b) 를 다수의 섹션들 A 내지 B (1819a 내지 1819b) 로 그룹화 (예컨대, 분할) 할 수도 있다. 일부의 구성들에서, 섹션 분석 모듈 (1817) 은 출력들 A 내지 B (1815a 내지 1815b) 또는 부분들 A 내지 B (1845a 내지 1845b) 의 각각을 4 개의 섹션들 A 내지 B (1819a 내지 1819b) 각각으로 그룹화할 수도 있다. 예를 들어, 제 1 섹션은 1 부터 150 까지의 장소 포인트들을 포함할 수도 있고, 제 2 섹션은 151 부터 275 까지의 장소 포인트들을 포함할 수도 있고, 제 3 섹션은 276 부터 450 까지의 장소 포인트들을 포함할 수도 있고, 제 4 섹션은 451 부터 512 까지의 장소 포인트들을 포함할 수도 있다. 섹션들 A 내지 B (1819a 내지 1819b) 의 각각은 시간의 길이에 대한 장소 포인트들 (예컨대, N 샘플들) 을 포함할 수도 있다.

[0260] 섹션들 A 내지 B (1819a 내지 1819b) 는 벡터 추출 모듈 (1821) 에 제공될 수도 있다. 벡터 추출 모듈 (1821) 은 도 15 와 관련하여 설명된 벡터 추출 모듈 (1521) 의 하나의 예일 수도 있다. 벡터 추출 모듈 (1821) 은 섹션들 A 내지 B (1819a 내지 1819b) 에 대한 벡터들 (1823) 을 추출할 수도 있다. 특히, 벡터 추출 모듈 (1821) 은 섹션들 A 내지 B (1819a 내지 1819b) 에 대한 장소-기반 분석 벡터 및 시간-기반 분석 벡터들을 추출할 수도 있다.

[0261] 벡터들 (1823) (예컨대, 하나 이상의 시간-기반 분석 벡터들 및 하나 이상의 장소-기반 분석 벡터들) 은 특징 결정 모듈 (1825) 에 제공될 수도 있다. 특징 결정 모듈 (1825) 은 도 15 와 관련하여 설명된 특징 결정 모듈 (1525) 의 하나의 예일 수도 있다. 특징 결정 모듈 (1825) 은 각각의 벡터 (1823) (예컨대, 분석 벡터)로부터 하나 이상의 특징들 (1827) 을 결정할 수도 있다. 예를 들어, 특징 결정 모듈 (1825) 은 각각의 벡터 (1823) 의 평균, 중앙값, 기하학적 시프트, 조화 평균, 표준 편차 및 비대칭도를 결정할 수도 있다.

[0262] 일부의 구성들에서, 특징들 (1827) 은 왜곡 추정 모듈 (1847) 에 제공될 수도 있다. 왜곡 추정 모듈 (1847) 은 도 15 와 관련하여 설명된 왜곡 추정 모듈의 하나의 예일 수도 있다. 왜곡 추정 모듈 (1847) 은 하나 이상의 특징들 (1827) 에 기초하여 하나 이상의 왜곡들 (1849) 을 추정할 수도 있다. 예를 들어, 왜곡 추정 모듈 (1847) 은 하나 이상의 특징들 (1827) 에 기초하여 하나 이상의 왜곡들 (1849) 을 추정하는 선형 회귀 모듈 (1851) 을 포함할 수도 있다. 예를 들어, 선형 회귀 모듈 (1851) 은 하나 이상의 왜곡들 (1849) 을 추정하기 위하여, 특징들 (1827) 중의 하나 이상 및 가중화 (1853) (예컨대, 하나 이상의 가중치들) 에 기초하여 선형 회귀를 수행할 수도 있다. 가중화 (1853) 는 본원에서 설명된 바와 같은 (예컨대, 도 8 내지 도 9 중의 하나 이상과 관련하여 설명된 바와 같은) 트레이닝에 기초하여 결정될 수도 있다. 일부의 구성들에서, 왜곡 추정 모듈 (1847) 은 추가적으로 또는 대안적으로, 왜곡 (들) (1849) 을 추정하기 위하여, 다항식 회귀, 2 차 회귀, 비선형 회귀 등을 수행할 수도 있다. 일부의 구성들에서, 하나 이상의 품질들 (예컨대, 포어그라운드 품질, 백그라운드 품질, 전체 품질 등) 은 본원에서 설명된 바와 같은 하나 이상의 왜곡들 (1849) 에 기초하여 추정될 수도 있다.

[0263] 본원에서 개시된 시스템들 및 방법들과 관련하여 설명된 침투적 접근법의 일부의 장점들은 다음 중의 하나 이상

을 포함할 수도 있다. 접근법은 인간 지각 지향될 수도 있다. 그것은 스피치 품질 측정에서 높은 정확도를 제공할 수도 있다. 그것은 다양한 타입들의 스피치 왜곡들에 대한 통찰 (예컨대, 설명) 을 제공할 수도 있다. 접근법은 유체-역학적 달팽이관 모델 출력을 사용할 수도 있다 (그 반면에, 다른 알려진 솔루션들은 그려하지 않을 수도 있음).

[0264] 도 19 는 비-침투적 특징 추출 모듈 (1929) 의 하나의 구성을 예시하는 블록도이다. 예를 들어, 도 19 는 지각 도메인에서의 스피치 및 오디오 신호 분석을 위한 달팽이관 모델로부터의 비-침투적 특징 추출을 위한 접근법을 예시한다. 비-침투적 특징 추출 모듈 (1929) 은 도 15 와 관련하여 설명된 특징 추출 모듈 (1529) 의 하나의 예일 수도 있다. 비-침투적 특징 추출은 특징 추출의 침투적 모델과 유사할 수도 있지만, 원본 스피치 (왜곡되지 않음) 는 이용가능하지 않을 수도 있다. 비-침투적 특징 추출 모듈 (1929) 은 달팽이관 모델 (1913), 부분 선택 모듈 (1943), 선택 분석 모듈 (1917), 벡터 추출 모듈 (1921) 및/또는 특정 결정 모듈 (1925) 을 포함할 수도 있다. 수정된 스피치 신호 (1911) (예컨대, 원본 스피치 신호의 열화된 버전) 는 특징 추출을 위한 입력으로서 이용될 수도 있다. 다시 말해서, 비-침투적 특징 추출 모듈 (1929) 은 하나 이상의 특징들 (1927) 을 결정할 수도 있다. 하나 이상의 특징들 (1927) 은 수정된 스피치 신호 (1911) 에 대한 하나 이상의 왜곡들 (1949) 을 추정 (예컨대, 예측) 하기 위해 이용될 수도 있다. 원본 스피치 신호 (1909) 및/또는 수정된 스피치 신호 (1911) 의 길이는 하나의 프레임, 다수의 프레임들, 또는 임의의 적당한 시간 길이 (예컨대, 1, 2, 5, 10 초 등) 일 수도 있다. 추가적으로 또는 대안적으로, 수정된 스피치 신호 (1911) 의 길이는 스피치 자체 (예컨대, 전체 문장) 에 기초할 수도 있다. 예를 들어, 수정된 스피치 신호 (1911) 의 길이는 (예를 들어, 무선 네트워크의 조작자에 의해, 및/또는 사용자에 의해) 구성가능할 수도 있다.

[0265] 수정된 스피치 신호 (1911) 는 달팽이관 모델 (1913) 에 제공될 수도 있다. 달팽이관 모델 (1913) 은 도 15 와 관련하여 설명된 생리학적 달팽이관 모델 (1513) 의 예일 수도 있다. 달팽이관 모델 (1913) 은 도 15 와 관련하여 설명된 바와 같이 기능할 수도 있다. 따라서, 달팽이관 모델 (1913) 은 출력 (1915) 을 생성할 수도 있다.

[0266] 위에서 설명된 바와 같이, 평균 인간 달팽이관은 길이에 있어서 약 3.5 cm 이고, 여기서, 달팽이관 상의 각각의 포인트는 (예를 들어, 근사적인 최저 지각가능한 주파수로부터 근사적인 최고 지각가능한 주파수까지의 범위인) 상이한 주파수에 대해 응답한다. 일부의 구성들에서, 달팽이관 모델 (1913) 은 달팽이관의 길이를 다수의 장소 포인트들 (예컨대, 512 개의 장소 포인트들 또는 샘플들) 로 이산화할 수도 있다. 예를 들어, 수정된 스피치 신호 (1911) 의 각각의 샘플에 대하여, 달팽이관 모델 (1913) 은 512 개의 장소 포인트들을 생성할 수도 있다. 출력 (1915) 은 샘플들의 범위에 대한 장소 포인트들의 세트를 포함할 수도 있다 (예컨대, N). 예를 들어, 출력 (1915) 은 수정된 스피치 신호 (1911) 에 대응하는 장소 포인트들의 N 개의 세트들을 포함할 수도 있다.

[0267] 출력 (1915) 은 부분 선택 모듈 (1943) 에 제공될 수도 있다. 부분 선택 모듈 (1943) 은 달팽이관 모델 (1913) 의 출력 (1915) 로부터 부분 (1945) (예컨대, "관심 부분") 을 선택할 수도 있다. 예를 들어, 수정된 스피치 신호 (1911) 의 유성음 부분들이 선택될 수도 있다 (예컨대, 모음들을 포함하는 유성음 부분들). 일부의 구성들에서, 부분 선택 모듈 (1943) 은 출력 (1915) 의 에너지를 결정할 수도 있다. 부분 (1945) 은 출력 (1915) 의 샘플들의 범위로서 선택될 수도 있고, 여기서, 출력 (1915) 의 에너지는 임계점을 초과한다. 부분 선택 모듈 (1943) 은 부분 (1945) 을 결정하기 위해 다른 음성 활성 검출 (VAD) 접근법들 또는 모듈들을 사용할 수도 있다. 대안적으로, 목음 부분들이 검출 및/또는 선택될 수도 있다.

[0268] 부분 (1945) 의 출력 (1915) 은 섹션 분석 모듈 (1917) 에 제공될 수도 있다. 섹션 분석 모듈 (1917) 은 도 15 와 관련하여 설명된 섹션 분석 모듈 (1517) 의 하나의 예일 수도 있다. 섹션 분석 모듈 (1917) 은 생리학적 달팽이관 모델 (1913) 의 출력 (1915) 또는 부분 (1945) 의 섹션들을 분석할 수도 있다. 예를 들어, 섹션 분석 모듈 (1917) 은 출력 (1915) 또는 부분 (1945) 을 다수의 섹션들 (1919) 로 그룹화 (예컨대, 분할) 할 수도 있다. 일부의 구성들에서, 섹션 분석 모듈 (1917) 은 출력 (1915) 또는 부분 (1945) 을 4 개의 섹션들 (1919) 로 그룹화할 수도 있다. 예를 들어, 제 1 섹션은 1 부터 150 까지의 장소 포인트들을 포함할 수도 있고, 제 2 섹션은 151 부터 275 까지의 장소 포인트들을 포함할 수도 있고, 제 3 섹션은 276 부터 450 까지의 장소 포인트들을 포함할 수도 있고, 제 4 섹션은 451 부터 512 까지의 장소 포인트들을 포함할 수도 있다. 섹션들 (1919) 의 각각은 시간의 길이에 대한 장소 포인트들 (예컨대, N 샘플들) 을 포함할 수도 있다.

[0269] 섹션들 (1919) 의 각각은 벡터 추출 모듈 (1921) 에 제공될 수도 있다. 벡터 추출 모듈 (1921) 은 도 15 와 관련하여 설명된 벡터 추출 모듈 (1521) 의 하나의 예일 수도 있다. 벡터 추출 모듈 (1921) 은 섹션들

(1919) 의 각각에 대한 벡터들 (1923) 을 추출할 수도 있다. 특히, 벡터 추출 모듈 (1921) 은 각각의 섹션 (1919) 에 대한 장소-기반 분석 벡터 및 시간-기반 분석 벡터를 추출할 수도 있다.

[0270] 벡터들 (1923) (예컨대, 하나 이상의 시간-기반 분석 벡터들 및 하나 이상의 장소-기반 분석 벡터들) 은 특징 결정 모듈 (1925) 에 제공될 수도 있다. 특징 결정 모듈 (1925) 은 도 15 와 관련하여 설명된 특징 결정 모듈 (1525) 의 하나의 예일 수도 있다. 특징 결정 모듈 (1925) 은 각각의 벡터 (1923) (예컨대, 분석 벡터)로부터 하나 이상의 특징들 (1927) 을 결정할 수도 있다. 예를 들어, 특징 결정 모듈 (1925) 은 각각의 벡터 (1923) 의 평균, 중앙값, 기하학적 시프트, 조화 평균, 표준 편차 및 비대칭도를 결정할 수도 있다.

[0271] 일부의 구성들에서, 특징들 (1927) 은 왜곡 추정 모듈 (1947) 에 제공될 수도 있다. 왜곡 추정 모듈 (1947) 은 도 15 와 관련하여 설명된 왜곡 추정 모듈의 하나의 예일 수도 있다. 왜곡 추정 모듈 (1947) 은 하나 이상의 특징들 (1927) 에 기초하여 하나 이상의 왜곡들 (1949) 을 추정할 수도 있다. 예를 들어, 왜곡 추정 모듈 (1947) 은 하나 이상의 특징들 (1927) 에 기초하여 하나 이상의 왜곡들 (1949) 을 추정하는 선형 회귀 모듈 (1951) 을 포함할 수도 있다. 예를 들어, 선형 회귀 모듈 (1951) 은 하나 이상의 왜곡들 (1949) 을 추정하기 위하여, 특징들 (1927) 중의 하나 이상 및 가중화 (1953) (예컨대, 하나 이상의 가중치들) 에 기초하여 선형 회귀를 수행할 수도 있다. 가중화 (1953) 는 본원에서 설명된 바와 같은 (예컨대, 도 8 내지 도 9 중의 하나 이상과 관련하여 설명된 바와 같은) 트레이닝에 기초하여 결정될 수도 있다. 일부의 구성들에서, 왜곡 추정 모듈 (1947) 은 추가적으로 또는 대안적으로, 왜곡 (들) (1949) 을 추정하기 위하여, 다항식 회귀, 2 차 회귀, 비선형 회귀 등을 수행할 수도 있다. 일부의 구성들에서, 하나 이상의 품질들 (예컨대, 포어그라운드 품질, 백그라운드 품질, 전체 품질 등) 은 본원에서 설명된 바와 같은 하나 이상의 왜곡들 (1949) 에 기초하여 추정될 수도 있다.

[0272] 본원에서 개시된 시스템들 및 방법들과 관련하여 설명된 비-침투적 접근법의 일부의 장점들은 다음 중의 하나 이상을 포함할 수도 있다. 접근법은 인간 지각 지향될 수도 있다. 그것은 스피치 품질 측정에서 높은 정확도를 제공할 수도 있다. 그것은 다양한 타입들의 스피치 왜곡들에 대한 이해 (예컨대, 설명) 를 제공할 수도 있다. 접근법은 유체-역학적 달팽이관 모델 출력을 사용할 수도 있다 (그 반면에, 다른 알려진 솔루션들은 그러하지 않을 수도 있음). 비-침투적 접근법은 침투적 방법들만큼 많은 정보에 대한 액세스를 가지지 않을 수도 있다는 것에 주목해야 한다. 그러므로, 그것은 침투적 접근법보다 품질 측정에 있어서 덜 정확할 수도 있다.

[0273] 도 20 은 섹션들 A 내지 D (2019a 내지 2019d) 로 분할되고 있는 달팽이관 모델의 출력 (2015) 의 하나의 예를 예시하는 블록도이다. 이 예에서, 달팽이관 모델은 매 샘플 입력에 대한 512 개의 장소 포인트들 (예컨대, 샘플들) 을 출력할 수도 있고, 여기서, 512 개의 장소 포인트들의 각각은 인간 달팽이관 상의 포인트에 대응한다. 그러므로, N 개의 샘플들이 입력될 경우, 달팽이관 모델은 $512 \times N$ 샘플들을 출력할 수도 있다 (예컨대, 달팽이관 모델 출력 (2015) 은 $512 \times N$ 의 사이즈를 가짐). 도 20 에서 예시된 바와 같이, 수평 축은 시간 (2057) 으로 예시되어 있다. 도 20 에서 예시된 바와 같이, 수직 축은 장소 (2055) (예컨대, 주파수에 맵핑될 수 있는, 달팽이관의 길이에 따른 장소 또는 위치) 로 예시되어 있다. 이 예에서, 출력 (2015) 은 스피치 신호의 N 개의 샘플들의 각각에 대한 512 개의 장소 포인트들을 포함한다.

[0274] 위에서 설명된 바와 같이, 출력 (2015) 은 더 작은 섹션들로 그룹화되거나 분할될 수도 있다. 예를 들어, 섹션 분석 모듈 (예컨대, 섹션 분석 모듈 (1517, 1817, 1917)) 은 달팽이관 모델의 출력 (2015) 을 4 개의 더 작은 섹션들 A 내지 D (2019a 내지 2019d) 로 분할할 수도 있다. 섹션 A (2019a) 는 시간 축 (2057) 의 N 개의 샘플들에 걸쳐 장소 포인트들 1 내지 150 을 포함할 수도 있다 (이것은 예를 들어, 장소 포인트들의 범위 ($K1=150$) 로 귀착됨). 섹션 B (2019b) 는 시간 축 (2057) 의 N 개의 샘플들에 걸쳐 장소 포인트들 151 내지 275 를 포함할 수도 있다 (이것은 예를 들어, 장소 포인트들의 범위 ($K2=125$) 로 귀착됨). 섹션 C (2019c) 는 시간 축 (2057) 의 N 개의 샘플들에 걸쳐 장소 포인트들 276 내지 450 을 포함할 수도 있다 (이것은 예를 들어, 장소 포인트들의 범위 ($K3=175$) 로 귀착됨). 섹션 D (2019d) 는 시간 축 (2057) 의 N 개의 샘플들에 걸쳐 장소 포인트들 451 내지 512 를 포함할 수도 있다 (이것은 예를 들어, 장소 포인트들의 범위 ($K4=62$) 로 귀착됨). 섹션들 (2019) 은 특정 값들로 예시되지만, 임의의 적당한 묘사는 섹션들 (2019) 을 정의하기 위해 이용될 수도 있다. 또한, 용어들 "섹션" 및 "영역" 은 달팽이관 모델 출력의 부분들을 지칭하기 위해 상호 교환가능하게 이용될 수도 있다. 출력 (2015) 을 더 작은 섹션들로 분할하는 것은 다양한 데이터를 관리하고 및/또는 주파수-관련 분석을 위한 접근법들을 가능하게 할 수도 있다.

[0275] 도 21 은 달팽이관 모델 출력들 (2115) 의 섹션들 (2119) 로부터 장소-기반 분석 벡터 (2123) 를 추출하는 하나

의 예를 예시하는 블록도이다. 특히, 도 21 은 원본 스피치에 대한 달팽이관 모델 출력 A (2115a) 의 색션 A (2119a) 및 수정된 (예컨대, 열화된) 스피치에 대한 달팽이관 모델 출력 B (2115b) 의 색션 E (2119e) 에 기초하여 장소-기반 (예컨대, "타입 1") 분석 벡터를 추출하는 예를 예시한다. 도 21 과 관련하여 설명된 동작들 중의 하나 이상은 도 15 와 관련하여 설명된 전자 디바이스 (1507) (예컨대, 특정 추출 모듈 (1529)) 에 의해 수행될 수도 있다.

[0276] 구체적으로, 이 예는 원본 스피치 신호에 기초하는 달팽이관 모델의 출력 A (2115a) 를 예시한다. 출력 A (2115a) 는 색션들 A 내지 D (2119a 내지 2119d) 를 포함한다. 또한, 이 예는 수정된 스피치 신호에 기초하는 달팽이관 모델의 출력 B (2115b) 를 예시한다. 출력 B (2115b) 는 색션들 E 내지 H (2119e 내지 2119h) 를 포함한다.

[0277] 색션 분석 모듈 A (2117a) 는 출력 A (2115a) 를 색션들 A 내지 D (2119a 내지 2119d) 로 분할하고, 색션 A (2119a) 를 평균화 모듈 A (2159a) 에 제공한다. 다시 말해서, 색션 분석 모듈 A (2117a) 는 달팽이관 모델 출력 A (2115a) 의 K₁ x N 샘플들 (예컨대, 색션 A (2119a)) 을 평균화 모듈 A (2159a) 에 제공한다. 평균화 모듈 A (2159a) 는 시간에 대하여 색션 A (2119a) 를 평균화한다. 이 평균은 평균적으로 log₁₀ 연산을 수행하는 로그 모듈 A (2161a) 에 제공된다. 평균의 로그 (예컨대, 제 1 로그 평균) 는 합산기 (2163) 에 제공된다.

[0278] 색션 분석 모듈 B (2117b) 는 출력 B (2115b) 를 색션들 E 내지 H (2119e 내지 2119h) 로 분할하고, 색션 E (2119e) 를 평균화 모듈 B (2159b) 에 제공한다. 다시 말해서, 색션 분석 모듈 B (2117b) 는 달팽이관 모델 출력 B (2115b) 의 K₁ x N 샘플들 (예컨대, 색션 E (2119e)) 을 평균화 모듈 B (2159b) 에 제공한다. 평균화 모듈 B (2159b) 는 시간에 대하여 색션 E (2119e) 를 평균화한다. 이 평균은 평균적으로 log₁₀ 연산을 수행하는 로그 모듈 B (2161b) 에 제공된다. 평균의 로그 (예컨대, 제 2 로그 평균) 는 합산기 (2163) 에 제공된다.

[0279] 합산기는 분석 벡터 (2123) (예컨대, "타입 1" 분석 벡터) 를 생성하기 위하여 제 1 로그 평균 및 제 2 로그 평균의 차이를 취한다. 이 분석 벡터 (2123) (예컨대, "타입 1" 분석 벡터) 는 장소-기반 분석 벡터 또는 침투적 장소-기반 분석 벡터로서 지칭될 수도 있다. 예를 들어, 장소-기반 분석 벡터 (2123) 는, 원본 스피치 신호 및 수정된 스피치 신호를 사용하는 침투적 접근법에서 결정될 수도 있다. 이 동작들은 출력들 A 내지 B (2115a 내지 2115b) 의 제 1 색션들 (색션 A (2119a) 및 색션 E (2119e)) 에 대해 예시되지만, 분석 벡터 (2123) 는 도 20 및 도 21 중의 하나 이상에서 예시된 4 개의 색션들 중의 임의의 것 및/또는 전부에 대해 결정될 수도 있다. 본원에서 이용된 바와 같이, 용어 "분석 시리얼 (analysis serial)" 및 "분석 벡터" 는 스피치의 특징들이 추출되는 중간 벡터를 지칭하기 위해 상호 교환가능하게 이용될 수도 있다.

[0280] 도 22 는 달팽이관 모델 출력 (2215) 의 색션 A (2219a) 로부터 장소-기반 분석 벡터 (2223) 를 추출하는 또 다른 예를 예시하는 블록도이다. 특히, 도 22 는 수정된 (예컨대, 열화된) 스피치에 대한 달팽이관 모델 출력 (2215) 의 색션 A (2219a) 에 기초하여 장소-기반 (예컨대, "타입 2") 분석 벡터를 추출하는 예를 예시한다. 유사한 접근법이 원본 스피치에 대한 장소-기반 분석 벡터를 추출하기 위해 사용될 수도 있다는 것에 주목해야 한다. 도 22 와 관련하여 설명된 동작들 중의 하나 이상은 도 15 와 관련하여 설명된 전자 디바이스 (1507) (예컨대, 특정 추출 모듈 (1529)) 에 의해 수행될 수도 있다.

[0281] 구체적으로, 이 예는 수정된 스피치 신호에 기초하는 달팽이관 모델의 출력 (2215) 을 예시한다. 출력 (2215) 은 색션들 A 내지 D (2219a 내지 2219d) 를 포함한다. 색션 분석 모듈 (2217) 은 출력 (2215) 을 색션들 A 내지 D (2219a 내지 2219d) 로 분할하고, 색션 A (2219a) 를 평균화 모듈 (2259) 에 제공한다. 다시 말해서, 색션 분석 모듈 (2217) 은 달팽이관 모델 출력 (2215) 의 K₁ x N 샘플들 (예컨대, 색션 A (2219a)) 을 평균화 모듈 (2259) 에 제공한다. 평균화 모듈 (2259) 은 시간에 대하여 색션 A (2219a) 를 평균화한다. 이 평균은 평균적으로 log₁₀ 연산을 수행하는 로그 모듈 (2261) 에 제공된다. 평균의 로그 (예컨대, 로그 평균) 는 분석 벡터 (2223) (예컨대, "타입 2" 분석 벡터) 이다. 이 분석 벡터 (2263) (예컨대, "타입 2" 분석 벡터) 는 장소-기반 분석 벡터 또는 비-침투적 장소-기반 분석 벡터로서 지칭될 수도 있다. 예를 들어, 장소-기반 분석 벡터 (2223) 는, 수정된 스피치 신호 (그리고 예를 들어, 원본 스피치 신호는 아님) 를 사용하는 비-침투적 접근법에서 결정될 수도 있다. 이 동작들은 출력 (2215) 의 제 1 색션 (색션 A (2219a)) 에 대해 예시되지만, 분석 벡터 (2223) 는 도 20 및 도 22 중의 하나 이상에서 예시된 4 개의 색션들 중의 임의의 것 및/또는 전부에 대해 결정될 수도 있다.

[0282] 도 23 은 달팽이관 모델 출력들 (2315) 의 색션들 (2319) 로부터 시간-기반 분석 벡터 (2323) 를 추출하는 하나의 예를 예시하는 블록도이다. 특히, 도 23 은 원본 스피치에 대한 달팽이관 모델 출력 A (2315a) 의 색션 A (2319a) 및 수정된 (예컨대, 열화된) 스피치에 대한 달팽이관 모델 출력 B (2315b) 의 색션 E (2319e) 에 기초하여 시간-기반 (예컨대, "타입 3") 분석 벡터를 추출하는 예를 예시한다. 도 23 과 관련하여 설명된 동작들 중의 하나 이상은 도 15 와 관련하여 설명된 전자 디바이스 (1507) (예컨대, 특정 추출 모듈 (1529)) 에 의해 수행될 수도 있다.

[0283] 구체적으로, 이 예는 원본 스피치 신호에 기초하는 달팽이관 모델의 출력 A (2315a) 를 예시한다. 출력 A (2315a) 는 색션들 A 내지 D (2319a 내지 2319d) 를 포함한다. 또한, 이 예는 수정된 스피치 신호에 기초하는 달팽이관 모델의 출력 B (2315b) 를 예시한다. 출력 B (2315b) 는 색션들 E 내지 H (2319e 내지 2319h) 를 포함한다.

[0284] 색션 분석 모듈 A (2317a) 는 출력 A (2315a) 를 색션들 A 내지 D (2319a 내지 2319d) 로 분할하고, 색션 A (2319a) 를 평균화 모듈 A (2359a) 에 제공한다. 다시 말해서, 색션 분석 모듈 A (2317a) 는 달팽이관 모델 출력 A (2315a) 의 K1 x N 샘플들 (예컨대, 색션 A (2319a)) 을 평균화 모듈 A (2359a) 에 제공한다. 평균화 모듈 A (2359a) 는 장소에 대하여 색션 A (2319a) 를 평균화한다. 이 평균은 평균적으로 \log_{10} 연산을 수행하는 로그 모듈 A (2361a) 에 제공된다. 평균의 로그 (예컨대, 제 1 로그 평균) 는 합산기 (2363) 에 제공된다.

[0285] 색션 분석 모듈 B (2317b) 는 출력 B (2315b) 를 색션들 E 내지 H (2319e 내지 2319h) 로 분할하고, 색션 E (2319e) 를 평균화 모듈 B (2359b) 에 제공한다. 다시 말해서, 색션 분석 모듈 B (2317b) 는 달팽이관 모델 출력 B (2315b) 의 K1 x N 샘플들 (예컨대, 색션 E (2319e)) 을 평균화 모듈 B (2359b) 에 제공한다. 평균화 모듈 B (2359b) 는 장소 (예컨대, 주파수에 대응할 수도 있는 달팽이관에 따른 장소) 에 대하여 색션 E (2319e) 를 평균화한다. 이 평균은 평균적으로 \log_{10} 연산을 수행하는 로그 모듈 B (2361b) 에 제공된다.

평균의 로그 (예컨대, 제 2 로그 평균) 는 합산기 (2363) 에 제공된다.

[0286] 합산기는 분석 벡터 (2323) (예컨대, "타입 3" 분석 벡터) 를 생성하기 위하여 제 1 로그 평균 및 제 2 로그 평균의 차이를 취한다. 이 분석 벡터 (2323) (예컨대, "타입 3" 분석 벡터) 는 시간-기반 분석 벡터 또는 침투적 시간-기반 분석 벡터로서 지정될 수도 있다. 예를 들어, 시간-기반 분석 벡터 (2323) 는, 원본 스피치 신호 및 수정된 스피치 신호를 사용하는 침투적 접근법에서 결정될 수도 있다. 이 동작들은 출력들 A 내지 B (2315a 내지 2315b) 의 제 1 색션들 (색션 A (2319a) 및 색션 E (2319e)) 에 대해 예시되지만, 분석 벡터 (2323) 는 도 20 및 도 23 중의 하나 이상에서 예시된 4 개의 색션들 중의 임의의 것 및/또는 전부에 대해 결정될 수도 있다.

[0287] 도 24 는 달팽이관 모델 출력 (2415) 의 색션 A (2419a) 로부터 시간-기반 분석 벡터 (2423) 를 추출하는 또 다른 예를 예시하는 블록도이다. 특히, 도 24 는 수정된 (예컨대, 열화된) 스피치에 대한 달팽이관 모델 출력 (2415) 의 색션 A (2419a) 에 기초하여 시간-기반 (예컨대, "타입 4") 분석 벡터를 추출하는 예를 예시한다. 유사한 접근법이 원본 스피치에 대한 시간-기반 분석 벡터를 추출하기 위해 사용될 수도 있다는 것에 주목해야 한다. 도 24 와 관련하여 설명된 동작들 중의 하나 이상은 도 15 와 관련하여 설명된 전자 디바이스 (1507) (예컨대, 특정 추출 모듈 (1529)) 에 의해 수행될 수도 있다.

[0288] 구체적으로, 이 예는 수정된 스피치 신호에 기초하는 달팽이관 모델의 출력 (2415) 을 예시한다. 출력 (2415) 은 색션들 A 내지 D (2419a 내지 2419d) 를 포함한다. 색션 분석 모듈 (2417) 은 출력 (2415) 을 색션들 A 내지 D (2419a 내지 2419d) 로 분할하고, 색션 A (2419a) 를 평균화 모듈 (2459) 에 제공한다. 다시 말해서, 색션 분석 모듈 (2417) 은 달팽이관 모델 출력 (2415) 의 K1 x N 샘플들 (예컨대, 색션 A (2419a)) 을 평균화 모듈 (2459) 에 제공한다. 평균화 모듈 (2459) 은 장소 (예컨대, 주파수에 대응할 수도 있는 달팽이관에 따른 장소) 에 대하여 색션 A (2419a) 를 평균화한다. 이 평균은 평균적으로 \log_{10} 연산을 수행하는 로그 모듈 (2461) 에 제공된다. 평균의 로그 (예컨대, 로그 평균) 는 분석 벡터 (2423) (예컨대, "타입 4" 분석 벡터) 이다. 이 분석 벡터 (2463) (예컨대, "타입 4" 분석 벡터) 는 시간-기반 분석 벡터 또는 비-침투적 시간-기반 분석 벡터로서 지정될 수도 있다. 예를 들어, 시간-기반 분석 벡터 (2423) 는, 수정된 스피치 신호 (그리고 예를 들어, 원본 스피치 신호는 아님) 를 사용하는 비-침투적 접근법에서 결정될 수도 있다. 이 동작들은 출력 (2415) 의 제 1 색션 (색션 A (2419a)) 에 대해 예시되지만, 분석 벡터 (2423) 는 도 20 및 도 24 중의 하나 이상에서 예시된 4 개의 색션들 중의 임의의 것 및/또는 전부에 대해 결정될 수도

있다.

[0289] 도 25 는 특정 결정 모듈 (2525) 의 구성들을 예시하는 블록도를 포함한다. 도 25 와 관련하여 설명된 특정 결정 모듈 (2525) 은 본원에서 설명된 특정 결정 모듈들 (1525, 1825, 1925) 중의 하나 이상의 특정 결정 모듈의 하나의 예일 수도 있다. 특정 결정 모듈 (2525) 은 특정 계산 모듈 (2563), 포지티브 모듈 (2565) 및/또는 네거티브 모듈 (2569) 을 포함할 수도 있다. 일부의 구성들에서, 포지티브 모듈 (2565) 및 네거티브 모듈 (2569) 은 침투적 접근법들에서 구현 및/또는 사용될 수도 있다. 일부의 구성들에서, 포지티브 모듈 (2565) 및 네거티브 모듈 (2569) 은 선택적이고, 및/또는 비-침투적 접근법들에서 구현 및/또는 사용되지 않을 수도 있다.

[0290] 벡터 (2523a) 는 특정 결정 모듈 (2525) 에 제공될 수도 있다. 특히, 벡터 (2523a) 는 특정 계산 모듈 (2563), 포지티브 모듈 (2565), 및/또는 네거티브 모듈 (2569) 에 제공될 수도 있다.

[0291] 포지티브 모듈 (2565) 은 벡터 (2523a) 의 포지티브 부분 (2567) 을 결정할 수도 있다. 예를 들어, 포지티브 모듈 (2565) 은 벡터 (2523a) 에서의 임의의 네거티브 값들을 제로로 변경시킬 수도 있다. 포지티브 모듈 (2565) 은 또한, 벡터 (2523a) 에서의 임의의 포지티브 값들을 변경되지 않은 채로 남겨둘 수도 있다.

[0292] 네거티브 모듈 (2569) 은 벡터 (2523a) 의 네거티브 부분 (2571) 을 결정할 수도 있다. 예를 들어, 네거티브 모듈 (2569) 은 벡터 (2523a) 에서의 임의의 포지티브 값들을 제로로 변경시킬 수도 있다. 네거티브 모듈 (2569) 은 또한, 벡터 (2523a) 에서의 임의의 네거티브 값들을 변경되지 않은 채로 남겨둘 수도 있다.

[0293] 벡터 (2523a), 포지티브 부분 (2567) 및/또는 네거티브 부분 (2571) 은 특정 계산 모듈 (2563) 에 제공될 수도 있다. 특정 계산 모듈 (2563) 은 벡터 (2523a), 포지티브 부분 (2567) 및/또는 네거티브 부분 (2571) 의 각각에 대한 하나 이상의 특징들을 결정 (예컨대, 계산) 할 수도 있다. 예를 들어, 특정 계산 모듈 (2563) 은 벡터 (2523a), 포지티브 부분 (2567) 및/또는 네거티브 파트 (2571) 의 각각에 대한 평균 (예컨대, 평균치), 중앙값, 기하학적 시프트, 조화 평균, 표준 편차, 비대칭도 및/또는 다른 특징을 계산할 수도 있다.

[0294] 벡터 (2523b) 의 포지티브 부분 및 네거티브 부분을 결정하기 위한 하나의 접근법이 도 25 에서 예시되어 있다. 특히, 도 25 는 포지티브 부분 및 네거티브 부분 결정 (2573) 의 하나의 예를 제공한다. 벡터 (2523b) (예컨대, 분석 벡터 또는 분석 시리얼) 는 n 개의 값들 또는 엔트리 (entry) 들을 가질 수도 있다: P1 내지 Pn. 포지티브 부분 (2575) (예컨대, 포지티브 벡터, 분석 시리얼 포지티브) 및 네거티브 부분 (2577) (예컨대, 네거티브 벡터, 분석 시리얼 네거티브) 이 생성될 수도 있다. 포지티브 부분 (2575) 에서의 각각의 값은 동일한 인덱스 (index) 를 갖는 벡터 (2523b) 에서의 대응하는 값에 기초하여 생성될 수도 있다 (예컨대, 포지티브 부분 (2575) 에서의 제 1 엔트리는 벡터 (2523b) 에서의 제 1 엔트리에 기초함). 예를 들어, 벡터 (2523b) 에서 P1 > 0 일 경우, 포지티브 부분 (2575) 에서의 P1 은 P1 이다. 그러나, 벡터 (2523b) 에서 P1 <= 0 일 경우, 포지티브 부분 (2575) 에서의 P1 은 0 이다. 반대로, 벡터 (2523b) 에서 P < 0 일 경우, 네거티브 부분 (2577) 에서의 P1 은 P1 이다. 그러나, 벡터 (2523b) 에서 P1 >= 0 일 경우, 네거티브 부분 (2577) 에서의 P1 은 0 이다. 이것은 특징들 (예컨대, 특징들 (2527)) 이 추출될 수도 있는 포지티브 부분 (2575) 및/또는 네거티브 부분 (2577) 을 패플레이팅 (populating) 하기 위하여 벡터 (2523b) 에서의 매 값 또는 엔트리에 대해 행해질 수도 있다.

[0295] 도 26 은 특정 결정의 예를 예시한다. 일부의 구성들에서, 도 26 에서 예시된 특정 결정은 본원에서 설명된 특정 결정 모듈들 (1525, 1825, 2525) 중의 하나 이상에 의해 수행될 수도 있다. 특히, 도 26 에서 예시된 특정 결정은 (예컨대, 침투적 장소-기반 벡터들 ("타입 1") 및 침투적 시간-기반 벡터들 ("타입 3") 에 대하여) 침투적 접근법들에서 수행될 수도 있다. 이 예에서, 단일 침투적 장소-기반 벡터 또는 단일 침투적 시간-기반 벡터는 18 개의 특징들 (예컨대, 특징 값들) 을 생성할 수도 있다: 벡터 (2623) 자체로부터 6 개, 분석 벡터 (2623) 의 포지티브 부분 (2667) 으로부터 6 개, 그리고 분석 벡터 (2623) 의 네거티브 부분 (2671) 으로부터 6 개. 일부의 구성들에서, 특징들 또는 특정 값들의 각각은 대응하는 모듈에 의해 결정 (예컨대, 계산) 될 수도 있다. 예를 들어, 각각의 모듈은 단일 특정 또는 특정 값을 신출할 수도 있다.

[0296] 일부의 구성들에서, 특정 결정 모듈 (예컨대, 도 25 와 관련하여 설명된 특정 결정 모듈 (2525)) 은 벡터 (2623) 의 포지티브 부분 (2667) 에 대한 평균 A (2679a), 중앙값 A (2681a), 기하학적 시프트 A (2683a), 조화 평균 A (2685a), 표준 편차 A (2687a) 및 비대칭도 A (2689a) 를 결정할 수도 있다. 추가적으로 또는 대안적으로, 특정 결정 모듈은 벡터 (2623) 자체에 대한 평균 B (2679b), 중앙값 B (2681b), 기하학적 시프트 B (2683b), 조화 평균 B (2685b), 표준 편차 B (2687b) 및 비대칭도 B (2689b) 를 결정할 수도 있다. 추가적

으로 또는 대안적으로, 특정 결정 모듈은 벡터 (2623) 의 네거티브 부분 (2671) 에 대한 평균 C (2679c), 중앙 값 C (2681c), 기하학적 시프트 C (2683c), 조화 평균 C (2685c), 표준 편차 C (2687c) 및 비대칭도 C (2689c) 를 결정할 수도 있다.

[0297] 포지티브 부분 (2667) 에 대하여, 특정 결정 모듈은 하나 이상의 다른 A (2691a) 특징들 또는 특정 값들을 결정 할 수도 있다. 벡터 (2623) 에 대하여, 특정 결정 모듈은 추가적으로 또는 대안적으로, 하나 이상의 다른 B (2691b) 특징들 또는 특정 값들을 결정할 수도 있다. 네거티브 부분 (2671) 에 대하여, 특정 결정 모듈은 추가적으로 또는 대안적으로, 하나 이상의 다른 C (2691c) 특징들 또는 특정 값들을 결정할 수도 있다. 하나 이상의 특징들 또는 특정 값들은 특정 세트에서 함께 그룹화될 수도 있다. 예를 들어, 평균 B (2679b), 중앙값 B (2681b), 기하학적 시프트 B (2683b), 조화 평균 B (2685b), 표준 편차 B (2687b) 및 비대칭도 B (2689b) 는 특정 세트로 그룹화될 수도 있다.

[0298] 도 27 은 특정 결정의 또 다른 예를 예시한다. 일부의 구성들에서, 도 27 에서 예시된 특정 결정은 본원에 서 설명된 특정 결정 모듈들 (1525, 1825, 1925, 2525) 중의 하나 이상에 의해 수행될 수도 있다. 특히, 도 27 에서 예시된 특정 결정은 (예컨대, 비-침투적 장소-기반 벡터들 ("타입 2") 및 비-침투적 시간-기반 벡터들 ("타입 4") 에 대하여) 비-침투적 접근법들에서, 및/또는 (예컨대, 침투적 장소-기반 벡터들 ("타입 1") 및 침 투적 시간-기반 벡터들 ("타입 3") 에 대하여) 침투적 접근법들에서 수행될 수도 있다. 이 예에서, 단일 침 투적 장소-기반 벡터 또는 단일 침투적 시간-기반 벡터는 벡터 (2723) 로부터 6 개의 특징들 (예컨대, 특정 값들) 을 생성할 수도 있다. 일부의 구성들에서, 특징들 또는 특정 값들의 각각은 대응하는 모듈에 의해 결정 (예컨대, 계산) 될 수도 있다. 예를 들어, 각각의 모듈은 단일 특징 또는 특징 값을 산출할 수도 있다.

[0299] 일부의 구성들에서, 특정 결정 모듈 (예컨대, 도 25 와 관련하여 설명된 특정 결정 모듈 (2525)) 은 벡터 (2723) 에 대한 평균 (2779), 중앙값 (2781), 기하학적 시프트 (2783), 조화 평균 (2785), 표준 편차 (2787) 및 비대칭도 (2789) 를 결정할 수도 있다. 벡터 (2723) 에 대하여, 특정 결정 모듈은 추가적으로 또는 대안 적으로, 하나 이상의 다른 (2791) 특징들 또는 특정 값들을 결정할 수도 있다. 하나 이상의 특징들 또는 특 징 값들은 특정 세트에서 함께 그룹화될 수도 있다. 예를 들어, 평균 (2779), 중앙값 (2781), 기하학적 시 프트 (2783), 조화 평균 (2785), 표준 편차 (2787) 및 비대칭도 (2789) 는 특정 세트로 그룹화될 수도 있다.

[0300] 도 28 은 본원에서 개시된 시스템들 및 방법들에 따라 객관적 예측의 하나의 예를 예시한다. 구체적으로, 도 28 은 그래프 A (2893a) 및 그래프 B (2893b) 를 포함한다. 그래프 A (2893a) 는, 1 내지 5 의 범위를 가지며 스피치 신호의 포어그라운드 품질에 대한 객관적 예측 점수들 또는 S-MOS 의 예측 (예컨대, SIG 는 포어 그라운드 점수의 예측을 의미함) 을 예시하는 S-MOS 예측 (2895) 점수들의 수직 축으로 예시된다. 그래프 A (2893a) 는 또한, 1 내지 5 의 범위를 또한 가지며 알려진 접근법을 이용한 포어그라운드 품질에 대한 척도의 예인, 주관적 S-MOS (2897) 점수들의 수평 축으로 예시된다. 그래프 A (2893a) 는 P.835 SIG 점수들 대 알 려진 접근법에서의 객관적 예측의 산포도 (scatter plot) 이다. 관찰될 수 있는 바와 같이, 알려진 접근법 은 2.5 보다 더 작은 주관적 점수들에 대해 양호하게 예측하지 못한다. 도 28 내지 도 30 에서, 포인트들이 대각선에 더 근접할수록, 이들이 나타내는 예측은 더욱 정확하다. 도 28 내지 도 30 에서의 A 그래프들의 전부는 하나의 데이터베이스 (예컨대, SIG (2893a), BAK (2993a) 및 OVR (3093a)) 에 대한 것이라는 것에 주목 해야 한다. 또한, 도 28 내지 도 30 에서의 B 그래프들의 전부는 또 다른 데이터베이스 (예컨대, SIG (2893b), BAK (2993b) 및 OVR (3093b)) 에 대한 것이라는 것에 주목해야 한다.

[0301] 그래프 B (2893b) 는, 1 내지 5 의 범위를 가지며 ITU 표준 P.835 를 이용한 스피치 신호의 포어그라운드 품질 에 대한 예측 점수인, 객관적 SIG (2899) 점수들의 수직 축으로 예시된다. 그래프 B (2893b) 는 또한, 1 내 지 5 의 범위를 또한 가지며 본원에서 개시된 시스템들 및 방법들에 따라 객관적 예측자 (objective predictor) 의 예인, 주관적 SIG (2802) 점수들의 수평 축으로 예시된다. 관찰될 수 있는 바와 같이, 본원에서 개시된 시스템들 및 방법들은 알려진 접근법보다 더 큰 정확도로 주관적 MOS 를 예측할 수도 있다.

[0302] 도 29 는 본원에서 개시된 시스템들 및 방법들에 따라 객관적 예측의 또 다른 예를 예시한다. 구체적으로, 도 29 는 그래프 A (2993a) 및 그래프 B (2993b) 를 포함한다. 그래프 A (2993a) 는, 1 내지 5 의 범위를 가지며 ITU 표준 P.835 를 이용한 스피치 신호의 백그라운드 잡음에 대한 예측 점수인, 객관적 P.835 BAK (2904) 점수들에서의 수직 축으로 예시된다. 그래프 A (2993a) 는 또한, 1 내지 5 의 범위를 또한 가지며 알려진 접근법을 이용한 백그라운드 잡음에 대한 척도의 예인, 주관적 N-MOS (2906) 점수들에서의 수평 축으로 예시된다. 그래프 A (2993a) 는 P.835 BAK 점수들 대 알려진 접근법에서의 객관적 예측의 산포도이다. 관찰될 수 있는 바와 같이, 알려진 접근법은 당면한 점수들을 상당히 근접하게 예측한다.

[0303] 그래프 B (2993b) 는, 1 내지 5 의 범위를 가지며 ITU 표준 P.835 를 이용한 스피치 신호의 백그라운드 잡음에 대한 예측 점수인, 객관적 P.385 NMOS (BAK) (2908) 점수들에서의 수직 축으로 예시된다. 그래프 B (2993b) 는 또한, 1 내지 5 의 범위를 또한 가지며 본원에서 개시된 시스템들 및 방법들에 따라 객관적 예측자의 예인, 주관적 NMOS (2910) 점수들의 수평 축으로 예시된다. 관찰될 수 있는 바와 같이, 본원에서 개시된 시스템들 및 방법들은 알려진 접근법보다 약간 더 작은 정확도로 점수들을 예측할 수도 있지만, 양자의 결과들은 주관적 점수들에 상당히 근접하다.

[0304] 도 30 은 본원에서 개시된 시스템들 및 방법들에 따라 객관적 예측의 또 다른 예를 예시한다. 구체적으로, 도 30 은 그래프 A (3093a) 및 그래프 B (3093b) 를 포함한다. 그래프 A (3093a) 는, 1 내지 5 의 범위를 가지며 ITU 표준 P.835 를 이용한 스피치 신호의 전체 품질에 대한 예측 점수인, 객관적 OVRL P.835 (3012) 점수들에서의 수직 축으로 예시된다. 그래프 A (3093a) 는 또한, 1 내지 5 의 범위를 또한 가지며 알려진 접근법을 이용한 전체 품질에 대한 척도의 예인, 주관적 G-MOS (3014) 점수들에서의 수평 축으로 예시된다. 그래프 A (3093a) 는 P.835 BAK 점수들 대 알려진 접근법에서의 객관적 예측의 산포도이다. 관찰될 수 있는 바와 같이, 알려진 접근법은 당면한 점수들을 상당히 근접하게 예측한다.

[0305] 그래프 B (3093b) 는, 1 내지 5 의 범위를 가지며 ITU 표준 P.835 를 이용한 스피치 신호의 전체 품질에 대한 예측 점수인, 객관적 GMOS (OVR) (3016) 점수들에서의 수직 축으로 예시된다. 그래프 B 는 또한, 1 내지 5 의 범위를 또한 가지며 본원에서 개시된 시스템들 및 방법들에 따라 객관적 예측자의 예인, 주관적 GMOS (OVR) (3018) 점수들에서의 수평 축으로 예시된다. 관찰될 수 있는 바와 같이, 본원에서 개시된 시스템들 및 방법들은 알려진 접근법보다 더 큰 정확도로 점수들을 예측할 수도 있다.

[0306] P.835 점수들을 예측하기 위하여 본원에서 개시된 시스템들 및 방법들의 능력의 일부의 예비적인 결과들이 표 2 에서 주어진다. 예를 들어, 본원에서 개시된 시스템들 및 방법들은 P.ONRA 에 대한 모델에 적용될 수도 있다. 본원에서 설명된 접근법들에서, 유체-역학적 달팽이관 모델은 스피치를 지각적 도메인으로 변환하기 위해 이용될 수도 있다. 달팽이관 모델 출력, 내유모세포 (IHC) 전압은 본질적으로 시간 및 공간 (예컨대, 기저막에 따른 거리) 축에서의 스피치 신호의 표현일 수도 있다. 전통적인 음향심리학적 모델들과 비교하면, 표현은 더 높은 시간적 분해능 및 더욱 정확한 공간적 정확도를 가진다. IHC 전압의 추가의 프로세싱은 예를 들어, SMOS, NMOS 및 GMOS 를 예측하기 위하여 간단한 선형 유럽 전기통신 표준 연구소 (European Telecommunications Standards Institute; ETSI) TS 103 106 회귀 모델들을 통해 추후에 공급될 수도 있는 '핵심 특징 세트 (salient feature set) 들' 을 산출한다.

[0307] 본원에서 제시된 예비적인 결과들에서, 본원에서 설명된 접근법들은 3 개의 주관적 데이터베이스들 (각각 60 개의 조건들을 가짐) 만으로 트레이닝된 반면; 비교예에서의 알려진 접근법은 7 개의 데이터베이스들로 트레이닝되었다. 예를 들어, 본원에서 개시된 시스템들 및 방법들은 확인 목적들을 위하여 일부의 P.835 데이터베이스들에 대해 테스트되었다. 4 개의 데이터베이스들이 이용되었다. 각각의 데이터베이스는 60 개의 조건들, 2 개의 화자 (speaker) 들 및 4 개의 문장들을 포함하였다. 3 개의 데이터베이스들이 트레이닝을 위해 이용되었고, 하나가 테스팅을 위해 이용되었다. 본원에서 설명되는 바와 같이, 트레이닝 및/또는 예측은 특징들의 선형 회귀로 수행될 수도 있다. 복잡한 트레이닝 모델 (예컨대, 신경망) 은 또한, 본원에서 개시된 시스템들 및 방법들에 따라 적용될 수 있다는 것에 주목해야 한다. 예를 들어, 트레이닝 프로세스가 개선될 수도 있다. 표 2 는 알려진 접근법 ("공지됨") 과 비교하여 본원에서 개시된 시스템들 및 방법들 ("신규") 의 일부의 예비적인 결과들의 예들을 제공한다. "RMSE" 는 "평균 제곱근 오차 (Root Mean Square Error)" 를 나타낸다.

표 2

	S-MOS		N-MOS		G-MOS	
	공지됨	신규	공지됨	신규	공지됨	신규
p	0.87	0.96	0.99	0.99	0.97	0.99
RMSE	0.45	0.24	0.13	0.22	0.36	0.16
RMSE*	0.33	0.13	0.04	0.12	0.23	0.09

[0308]

[0309] 표 2 는 하나의 주관적인 테스트 데이터베이스에 대한 예측 결과들을 예시한다. 동일한 데이터베이스 상에

서 (7 개의 주관적 데이터베이스들로 재트레이닝된) TS 103 106 에 의한 예측들은 비교를 위한 "공지된" 접근법으로서 여기에서 열거된다. 결과들은 본원에서 개시된 시스템들 및 방법들이 S-MOS 및 G-MOS 를 예측할 수 있는 것에 더욱 양호함을 보여준다. 상세한 비교들은 도 28 에서 도시된다. TS 103 106 의 성능은 3 보다 더 적은 주관적인 S-MOS 를 갖는 조건들에 대해 낮아진다. N-MOS 성능에 대한 상세한 비교는 도 29 에서 도시된다. 상관 계수들의 측면에서 차이가 없다 (양자 모두 0.99 를 나타냄). G-MOS 는 S-MOS 및 N-MOS 로부터 합성될 수도 있다. 본원에서 개시된 시스템들 및 방법들은 TS 103 106 보다 약간 더 양호한 예측을 제공한다. 주관적 및 객관적 점수들 사이의 상관 계수들은 0.99 대 0.97 이다. 현재의 모델은 또한, 더 작은 RMSE (0.16 대 0.36) 및 RMSE* (0.09 대 0.23) 를 나타낸다. G-MOS 의 비교는 도 30 에서 도시되어 있다.

[0310] 도 31 은 도 31a 내지 도 31d 에 대응하는 예의 다수의 부분들을 예시하는 블록도이다. 도 31a 내지 도 31d 는 침투적 특징 추출을 이용하여 러프니스 왜곡을 추정하거나 예측하는 하나의 예를 예시하는 블록도를 도시한다. 특정한 수들이 예시를 위해 이용될 수도 있지만, 본 시스템들 및 방법들의 임의의 포인트에서의 다양한 데이터의 실제적인 사이즈는 변동될 수도 있다.

[0311] 도 31a 에서 도시된 바와 같은 원본 스피치 신호 (3109) 및 수정된 (예컨대, 열화된) 스피치 신호 (3111) (각각은 1000 개의 샘플들의 길이로서 예시됨) 는 달팽이관 모델에 공급될 수도 있다. 달팽이관 모델은 매 샘플 입력에 대해 512 개의 샘플들을 출력할 수도 있다. 특히, 출력 A (3115a) 는 원본 스피치 신호 (3109) 에 대응할 수도 있고, 출력 B (3115b) 는 수정된 스피치 신호 (3111) 에 대응할 수도 있다.

[0312] 출력들 A 내지 B (3115a 내지 3115b) 로부터, 제 1 유성음 부분은 4 개의 섹션들 (3119a 내지 3119b) (예컨대, 원본 스피치 신호 (3109) 에 기초한 달팽이관 모델의 출력 A (3115a) 에 대한 4 개의 섹션들 (3119a) 과, 수정된 스피치 신호 (3111) 에 기초한 달팽이관 모델의 출력 B (3119b) 에 기초한 4 개의 섹션들 (3119b)) 로 분할될 수도 있다.

[0313] 하나 이상의 분석 벡터들 또는 분석 시리얼들 (analysis serials; AS) (3123) 은 섹션들 (3119a 내지 3119b)로부터 결정될 수도 있다. 예시된 바와 같이, R1 AS1 은 제 1 섹션들로부터 결정될 수도 있고, 여기서, R1 은 그것이 제 1 유성음 부분의 제 1 섹션들로부터 결정되었다는 사실을 지칭하고, AS1 은 그것이 타입 1 분석 벡터라는 사실을 지칭한다. R1 AS1 에 추가하여, 다음의 분석 벡터들 (3123) 이 결정될 수도 있다: R1 AS3, R2 AS1, R2 AS3, R3 AS1, R3 AS3, R4 AS1, R4 AS3, R3 AS2 및 R3 AS4. 이 벡터들은 도 15 내지 도 16, 도 18 및 도 21 내지 도 24 중의 하나 이상과 관련하여 설명된 바와 같이 결정될 수도 있다. 그러므로, 각각의 선택된 섹션에 대해 추출된 분석 벡터들 (3123) 의 타입들은, 스피치 품질의 침투적 측정을 위한 타입들을 예시한 표 3 에 따라 결정될 수도 있다:

표 3

섹션	입력 신호	추출된 분석 벡터
섹션 A (R1)	원본 및 수정	타입들 1 및 3
섹션 B (R2)	원본 및 수정	타입들 1 및 3
섹션 C (R3)	원본 및 수정	타입들 1 및 3
섹션 D (R4)	원본 및 수정	타입들 1 및 3
섹션 E (R5)	원본	타입들 2 및 4
SFP	원본 및 수정	타입 1

[0314]

[0315] 또한, 타입 1 분석 벡터는 원본 스피치 및 수정된 (예컨대, 열화된) 스피치로부터 결정되는 핵심 특징 포인트 (SFP) 벡터 (3124) 에 대해 추출될 수도 있다. SFP 벡터들은 시간적으로 국소화된 왜곡들에 관련된다. SFP 벡터를 획득하기 위한 프로세스는: (1) CM 출력을 획득하고 유성음 부분을 취하는 것; (2) 유성음 부분에서 하나 이상의 (예컨대, 전부) 트랙들을 찾는 것; (3) 진폭이 높은 하나 이상의 섹션들을 이용하는 것; 및 (4) 유성음 부분에서의 하나 이상의 트랙으로부터 SFP 벡터 (3124) 를 계산하는 것일 수도 있다. 그러므로, 선택된 섹션에 대한 특징들 (3127) (예컨대, 특징 벡터) 을 결정하기 위하여, 6 또는 18 개의 특징들을 각각 생성하는 11 개의 벡터들이 이용될 수도 있다 (10 개의 분석 벡터들 (3123) 및 SFP (3124)). SFP 는 "지각적 피치 (perceptual pitch)" 로 고려될 수도 있고, 전통적인 피치 개념과는 상이할 수도 있다. 이것은 SFP 가 3

차원일 수도 있기 때문이며, 이것은 SFP 에서의 각각의 포인트가 시간, 장소 및 진폭을 가진다는 것을 의미한다. 전통적인 퍼치 개념은 지각적 도메인에서 진폭을 가지지 않을 수도 있다.

[0316] 구체적으로, 각각의 타입 1 또는 타입 3 분석 벡터는 (예를 들어, 도 25 내지 도 26 과 관련하여 설명된 바와 같이) 18 개의 특징들 (3127) 을 생성할 수도 있다. 예시된 바와 같이, AS1 및 AS3 벡터들은 포지티브 부분들 (예컨대, RS1 AS1 포지티브) 및 네거티브 부분들 (예컨대, RS1 AS1 네거티브) 을 획득하기 위해 사용될 수도 있다.

[0317] 구체적으로, 각각의 타입 2 또는 타입 4 분석 벡터는 (예를 들어, 도 25 및 도 27 과 관련하여 설명된 바와 같이) 6 개의 특징들을 생성할 수도 있다. 특징들 (3127) 은 이 예에서, 174 개의 엔트리들을 가지는 특징 벡터에서 배치될 수도 있다. 특징 추출의 동일한 프로세스는 또 다른 174 개의 엔트리 특징 벡터를 생성하기 위하여 원본 스피치 및 수정된 스피치에 대한 달팽이관 출력의 제 2 부분 (예컨대, 유성음 부분) 에 대해 수행될 수도 있다. 동일한 프로세스는, 제 1 선택된 섹션의 특징 벡터와 유사한 특징 벡터를 각각 생성하는 임의의 적당한 수의 부분들 (예컨대, 유성음 부분들) 에 대해 반복될 수도 있다.

[0318] 일단 모든 선택된 섹션들에 대한 특징 벡터들이 결정되면, 특징들 (3127) 은 도 31b 및 도 31c 에서 예시된 바와 같이 평균화될 수도 있다. 다시 말해서, 특징 벡터들의 각각에서의 (평균들로서 도시된) 제 1 엔트리들이 평균화될 수도 있다. 유사하게, 매 인덱스에서의 엔트리들은, 이 예에서 174 개의 엔트리 벡터로서 도시된, 단일의 평균 특징 벡터 (3120) 를 생성하기 위하여 서로 평균화된다. 단일 평균 특징 벡터 (3120) 는 (예를 들어, 도 8 내지 도 9 중의 하나 이상과 관련하여 설명된 바와 같은) 트레이닝 동안에 미리 학습된 (예컨대, 가중치들을 이용하여 가중화될 수도 있는) 가중화 (3153) 와 승산되었다. 가중화 후에, 단일 평균 특징 벡터에서의 특징들의 각각은 도 31d 에서 도시된 바와 같은 러프니스 왜곡 (3149) (예컨대, 예측 점수) 을 생성하기 위하여 합산될 수도 있다. 유사한 절차들이 다른 왜곡들에 대해 후속될 수도 있다.

[0319] 스피치 품질의 객관적 측정을 위한 알려진 접근법은 ITU 표준들 PESQ/P.OLOQA 와 같이, 전체 품질을 직접적으로 올리는 것이다. 알려진 접근법들에서, 특징들은 인간의 각각의 러프 근사화들인 음향심리학적 마스킹 모델 또는 유사한 것으로부터 추출되었다. 지각적 포먼트-유사 (formant-like) 특징들은 유체-역학적 달팽이관 모델의 출력으로부터 추출되었고, 시간적으로 국소화된 왜곡들 측정을 위해 이용되었다.

[0320] 일부의 구성들에서, 시스템들 및 방법들의 하나 이상의 절차들은 다음과 같이 수행될 수도 있다. 하나의 접근법에서, 여기에서 도입된 새로운 특징 세트들은 (열화된 및 원본 스피치 신호의 양자를 이용하여) 침투적이다. 원본 및 대응하는 열화된 스피치 (S^{ori} 및 S^{dis} , 정렬되고 레벨링됨) 를 가정한다. S^{ori} 및 S^{dis} 는 (유체-역학적 달팽이관 모델들 또는 음향심리학적 모델과 같은) 청각의 연산 모델을 통과하게 될 수도 있다. 이러한 모델들은 CM 으로서 지칭될 수도 있고, 원본 및 왜곡된 신호들로 인한 출력들은 CM^{ori} 및 CM^{dis} 로서 지칭될 것이다.

[0321] 원본 및 왜곡된 스피치는 유체-역학적 달팽이관 모델 (CM) 을 통과하게 될 수도 있고, 출력은 시간 및 주파수 도메인들의 양자를 따라 높은 정밀도를 가진다. 출력은 지각적 도메인에서 스피치를 나타낸다.

[0322] 분석 시리얼들 (AS) 은 관심 있는 섹션들로부터 추출될 수도 있다. 특징 V SF144 는 분석의 목적에 따라서는, 본원에서 설명된 동일한 알고리즘들로 스피치의 유성음 섹션들 (VS) 및 묵음 (SIL) 섹션들로부터 추출될 수 있다. K 개의 관심 있는 섹션들이 있는 것으로 가정한다.

[0323] 주파수 관련 분석 시리얼들은 다음과 같이 설명될 수도 있다. k 번째 관심 있는 영역에 대하여, 열화된 CM

출력 $CM_k^{dis}(p, t)$ 및 대응하는 원본 CM 출력 $CM_k^{ori}(p, t)$ 은 각각 P*T 의 사이즈를 가진다.

$$CMOT_k^{dis}(p) = \log_{10} \frac{\sum_{t=1}^T CM_k^{dis}(p, t)}{T} \quad (56)$$

$$CMOT_k^{ori}(p) = \log_{10} \frac{\sum_{t=1}^T CM_k^{ori}(p, t)}{T} \quad (57)$$

$$CMOT_k^{diff}(p) = CMOT_k^{dis}(p) - CMOT_k^{ori}(p) \quad (58)$$

[0324]

[0325]

상이한 지각적 모델에 대하여, CM은 상이한 사이즈를 가질 수도 있다. 다음의 부분에서, 우리는 달팽이관 모델을 취한다. 숫자들은 어떤 출력에 대해 조절될 수도 있다는 것에 주목해야 한다. 이 예에서의 P는 범위 1 내지 512를 가지고, 4 개의 영역들로 절단된다: (A) 1 내지 150, (B) 151 내지 275, (C) 276 내지 450 및 (D) 451 내지 512. 따라서, 장소의 함수로서 4 개의 차분 함수 (difference function) 들이 있다:

$$\begin{aligned} P1_k(p) &= CMOT_k^{diff}(p), \quad p = 1 \dots 150 \\ P2_k(p) &= CMOT_k^{diff}(p), \quad p = 151 \dots 275 \\ P3_k(p) &= CMOT_k^{diff}(p), \quad p = 276 \dots 450 \\ P4_k(p) &= CMOT_k^{diff}(p), \quad p = 451 \dots 512 \end{aligned} \quad (59)$$

[0326]

[0327]

또 다른 것은 원본 스피치에 대하여, 그러나 핵심 영역 (영역 C)에서 설명된다:

$$P5_k(p - 275) = CMOT_k^{ori}(p), \quad p = 276 \dots 450 \quad (60)$$

(N의 길이를 갖는) $P1_k(p)$ 는 다음과 같이 분리될 수도 있다

$$\begin{aligned} P1_k^{positive}(p) &= \begin{cases} P1_k(p), & P1_k(p) > 0 \text{ 인 경우} \\ 0, & \text{그 외} \end{cases} \\ P1_k^{negative}(p) &= \begin{cases} P1_k(p), & P1_k(p) < 0 \text{ 인 경우} \\ 0, & \text{그 외} \end{cases} \end{aligned} \quad (61)$$

동일한 절차는 다음을 산출하기 위하여 $P2_k(p), P3_k(p), P4_k(p)$ 에 대하여 적용될 수도 있다

$$\begin{aligned} P1_k^{positive}(p), P1_k^{negative}(p), P1_k(p), \\ P2_k^{positive}(p), P2_k^{negative}(p), P2_k(p), \\ P3_k^{positive}(p), P3_k^{negative}(p), P3_k(p), \\ P4_k^{positive}(p), P4_k^{negative}(p), P4_k(p) \end{aligned} \quad (62)$$

및 $P5_k(p)$ 자체만.

[0333] 이것들은 분석 시리얼들로 칭해질 수도 있다.

[0334] 시간 관련 분석 시리얼들은 다음과 같이 획득될 수도 있다. 각각의 유성음 색션 (VS)에 대하여, 열화된 CM 출력 $CM_k^{dis}(p, t)$ 및 대응하는 원본 $CM_k^{ori}(p, t)$ 이 있으며, 여기서, 각각은 P*T의 사이즈를 가진다. 다음과 같이 계산될 수도 있다:

$$\begin{aligned}
T1_k^{dis}(t) &= \log_{10} \frac{\sum_{p=r1,1}^{r1,2} CM_k^{dis}(p, t)}{r_{1,2} - r_{1,1} + 1} \\
T1_k^{ori}(t) &= \log_{10} \frac{\sum_{p=r1,1}^{r1,2} CM_k^{ori}(p, t)}{r_{1,2} - r_{1,1} + 1} \\
T1_k^{diff}(t) &= T1_k^{dis}(t) - T1_k^{ori}(t) \\
T2_k^{dis}(t) &= \log_{10} \frac{\sum_{p=r2,1}^{r2,2} CM_k^{dis}(p, t)}{r_{2,2} - r_{2,1} + 1} \\
T2_k^{ori}(t) &= \log_{10} \frac{\sum_{p=r2,1}^{r2,2} CM_k^{ori}(p, t)}{r_{2,2} - r_{2,1} + 1} \\
T2_k^{diff}(t) &= T2_k^{dis}(t) - T2_k^{ori}(t) \\
T3_k^{dis}(t) &= \log_{10} \frac{\sum_{p=r3,1}^{r3,2} CM_k^{dis}(p, t)}{r_{3,2} - r_{3,1} + 1} \\
T3_k^{ori}(t) &= \log_{10} \frac{\sum_{p=r3,1}^{r3,2} CM_k^{ori}(p, t)}{r_{3,2} - r_{3,1} + 1} \\
T3_k^{diff}(t) &= T3_k^{dis}(t) - T3_k^{ori}(t) \\
T4_k^{dis}(t) &= \log_{10} \frac{\sum_{p=r4,1}^{r4,2} CM_k^{dis}(p, t)}{r_{4,2} - r_{4,1} + 1} \\
T4_k^{ori}(t) &= \log_{10} \frac{\sum_{p=r4,1}^{r4,2} CM_k^{ori}(p, t)}{r_{4,2} - r_{4,1} + 1} \\
T4_k^{diff}(t) &= T4_k^{dis}(t) - T4_k^{ori}(t) \\
&\begin{array}{ccc} r & 1 & 2 \\ 1 & 1 & 150 \\ 2 & 151 & 275 \\ 3 & 276 & 450 \\ 4 & 450 & 512 \end{array}
\end{aligned} \tag{63}$$

[0335]

[0336] 여기서, $r_{1,2}$ 는 "제 1 행, 제 2 열" (예컨대, 150) 을 의미한다. (예를 들어, 단지 영역 C 보다) 더 많은 영역들이 사용될 수도 있다는 것에 주목해야 한다. 따라서, 장소의 함수로서 4 개의 차분 함수들이 있다. 또 다른 것은 원본 스펙트럼에 대해, 그러나 핵심 영역 (영역 C) 에서만 설명된다.

$$T5_k(t) = T3_k^{ori}(t) \tag{64}$$

[0338] (T 의 길이를 갖는) $T1_k(t)$ 는 다음으로 분리될 수도 있다.

$$\begin{aligned}
T1_k^{positive}(t) &= \begin{cases} T1_k(t), & T1_k(t) > 0 \text{ 인 경우} \\ 0, & \text{그 외} \end{cases} \\
T1_k^{negative}(t) &= \begin{cases} T1_k(t), & T1_k(t) < 0 \text{ 인 경우} \\ 0, & \text{그 외} \end{cases}
\end{aligned} \tag{65}$$

[0339]

[0340] 동일한 절차는 다음을 산출하기 위하여 $T2_k(p)$, $T3_k(p)$, $T4_k(p)$ 에 대하여 적용될 수도 있다.

$$\begin{aligned} & T1_k^{positive}, T1_k^{negative}, T1_k, \\ & T2_k^{positive}, T2_k^{negative}, T2_k, \\ & T3_k^{positive}, T3_k^{negative}, T3_k, \\ & T4_k^{positive}, T4_k^{negative}, T4_k. \end{aligned} \quad (66)$$

[0341] [0342] 이것들은 또한, 분석 시리얼들로 칭해질 수도 있다.

[0343] SFP 가 또한 획득될 수도 있다. 관심 있는 섹션이 유성음 섹션일 경우, SFP_k 는 k 번째 유성음 섹션의 CM 출력으로부터 추출된다. 일부의 구성들에서, SFT 는 유럽 특허 출원 공개 제 EP2329399 A4 호에서 설명된 바와 같이 획득될 수도 있다. 3 개의 AS 가 여기에서 형성된다:

$$SFP_k^{positive}, SFP_k^{negative}, SFP_k, \quad (67)$$

[0344] [0345] AS로부터의 특징 추출은 다음과 같이 진행할 수도 있다. N 의 길이를 가지는 각각의 AS (예컨대, $P2_k^{negative}(p)$ 및 $T3_k$) 에 대하여, 특징들은 다음과 같이 추출된다:

$$\begin{aligned} FT1 &= mean(AS) \\ FT2 &= \frac{\sum_{p=1}^N AS(p)* p}{\sum_{p=1}^N AS(p)} - \bar{p}, \end{aligned} \quad (68)$$

[0346] 기하학적 시프트.

$$FT3 = median(AS),$$

[0348] 여기에서의 중앙값 함수는 $p(AS \geq F3) \geq \frac{1}{2}$ 을 의미하고, 그리고

$$p(AS \leq F3) \geq \frac{1}{2}.$$

$$FT4 = \frac{1}{\sum_{p=1}^N \frac{1}{AS(p)}},$$

조화 평균.

$$FT5 = std(AS),$$

표준 편차 (std).

$$FT6 = \frac{\frac{1}{N} \sum_{p=1}^N (AS(p) - \bar{AS})^3}{\left[\frac{1}{N} \sum_{p=1}^N (AS(p) - \bar{AS})^2 \right]^3}$$

[0349]

[0350] $P1_k^{positive}, P1_k^{negative}, P1_k$ 로부터 추출된 특징들은 F1 내지 F18 이다. $P2_k^{positive}, P2_k^{negative}, P2_k$ 로부터 추출된 특징들은 F19 내지 F36 이다. $P3_k^{positive}, P3_k^{negative}, P3_k$ 로부터 추출된 특징들은 F37 내지 F54 이다.

$P4_k^{positive}, P4_k^{negative}, P4_k$ 로부터 추출된 특징들은 F55 내지 F72 이다. 영역 C에서의 원본 값들인 $P5_k$ 에 대해서는, 6 개의 특징들만이 있다: F73 내지 F78.

[0351] $T1_k^{positive}, T1_k^{negative}, T1_k$ 로부터 추출된 특징들은 F79 내지 F96 이다. $T2_k^{positive}, T2_k^{negative}, T2_k$ 로부터 추출된 특징들은 F97 내지 F114 이다. $T3_k^{positive}, T3_k^{negative}, T3_k$ 로부터 추출된 특징들은 F115 내지 F132 이다.

$T4_k^{positive}, T4_k^{negative}, T4_k$ 로부터 추출된 특징들은 F133 내지 F150 이다. 영역 C에서 원본 값들인 $P5_k$ 에 대하여, 6 개의 특징들만이 있다: F151 내지 F156. $SFP_k^{positive}, SFP_k^{negative}, SFP_k$ 로부터 추출된 특징들은 F157 내지 F174 이다.

[0352]

최종적인 특징 점수는 모든 유성음 색션들에 대한 그 특징의 평균들을 포함할 수도 있다는 것 (예컨대, $FS1 = \frac{1}{V} \sum_{v=1}^V F1_k$)에 주목해야 한다. 특징 F175는 화자 (speaker)가 남성 또는 여성인지 여부의 이진 표시자 (binary indicator) (예컨대, 남성에 대해 0, 그리고 여성에 대해 -1) 일 수도 있다. 특징 F176은 선형 회귀의 편리함을 위하여 상수 (예를 들어, 보통 1로 설정됨) 일 수도 있다.

[0353]

본원에서 개시된 시스템들 및 방법들은 스피치 품질 측정 출력을 제공할 수도 있다. 코딩 및/또는 송신 개발자를 위한 피드백 (예컨대, 실시간 피드백)이 추가적으로 또는 대안적으로 제공될 수도 있다. 스피치 왜곡들의 특성을 및/또는 통찰은 추가적으로 또는 대안적으로 제공될 수도 있다. 본원에서 개시된 시스템들 및 방법들은 지각적 도메인에서 주어진 스피치의 특성을 (예컨대, 왜곡들)을 설명하기 위하여 많은 파라미터들을 제공할 수도 있다. 이 특징들은 일반적인 툴박스 (toolbox) 내로 구축될 수도 있고 특정 목적들 (예컨대, 주파수-국소화된 왜곡들 측정)을 위해 조합될 수도 있다. 다시 말해서, 본원에서 설명된 시스템들 및 방법들은 스피치 품질 측정, 다양한 타입들의 스피치 왜곡들에 관한 설명에 대한 통찰, 인간-지각적으로 지

향된 접근법에서 높은 정확도를 제공할 수도 있다. 다른 알려진 접근법들은 유체-역학적 달팽이관 모델 출력을 이용하지 않을 수도 있다는 것에 주목해야 한다.

[0354] 도 32 는 도 32a 내지 도 32d 에 대응하는 예의 다수의 부분들을 예시하는 블록도이다. 도 32a 내지 도 32d 는 비-침투적 특징 추출을 이용하여 러프니스 왜곡을 추정하거나 예측하는 하나의 예를 예시하는 블록도이다.

특정한 수들이 예시를 위해 이용될 수도 있지만, 본 시스템들 및 방법들의 임의의 포인트에서의 다양한 데이터의 실제적인 사이즈는 변동될 수도 있다.

[0355] (1000 개의 샘플들의 길이로서 예시된) 도 32a 에서 도시된 바와 같은 수정된 (예컨대, 열화된) 스피치 신호 (3211) 는 달팽이관 모델에 공급될 수도 있다. 달팽이관 모델은 매 샘플 입력에 대해 512 개의 샘플들을 출력할 수도 있다. 특히, 출력 (3215) 은 수정된 스피치 신호 (3211) 에 대응할 수도 있다. 출력 (3215) 으로부터, 제 1 유성음 부분은 4 개의 섹션들 (3219) (예컨대, 수정된 스피치 신호 (3211) 에 기초한 달팽이관 모델의 출력 (3215) 에 기초한 4 개의 섹션들 (3219)) 로 분할될 수도 있다.

[0356] 하나 이상의 분석 벡터들 또는 분석 시리얼들 (AS) (3223) 은 섹션들 (3219) 로부터 결정될 수도 있다. 예시된 바와 같이, R1 AS2 는 제 1 섹션으로부터 결정될 수도 있고, 여기서, R1 은 그것이 제 1 유성음 부분의 제 1 섹션으로부터 결정되었다는 사실을 지칭하고, AS2 는 그것이 타입 2 분석 벡터라는 사실을 지칭한다. R1 AS2 에 추가하여, 다음의 분석 벡터들 (3223) 이 결정될 수도 있다: R1 AS4, R2 AS2, R2 AS4, R3 AS2, R3 AS4, R4 AS2 및 R4 AS4. 이 벡터들은 도 15 내지 도 16, 도 19, 도 22 및 도 24 중의 하나 이상과 관련하여 설명된 바와 같이 결정될 수도 있다. 그러므로, 각각의 선택된 섹션에 대해 추출된 분석 벡터들의 타입들은, 스피치 품질의 비-침투적 측정을 위한 타입들을 예시한 표 4 에 따라 결정될 수도 있다:

표 4

섹션	입력 신호	추출된 분석 벡터
섹션 A (R1)	수정	타입들 2 및 4
섹션 B (R2)	수정	타입들 2 및 4
섹션 C (R3)	수정	타입들 2 및 4
섹션 D (R4)	수정	타입들 2 및 4
SFP	수정	타입 2

[0357]

[0358] 또한, 타입 2 분석 벡터는 수정된 스피치로부터 결정되는 핵심 특징 포인트 (SFP) 벡터 (3224) 에 대해 추출될 수도 있다. SFP 벡터들은 시간적으로 국소화된 왜곡들에 관련된다. 그러므로, 선택된 섹션에 대한 특징 벡터를 결정하기 위하여, 6 개의 특징들을 각각 생성하는 9 개의 벡터들이 이용될 수도 있다 (8 개의 분석 벡터들 (3223) 및 SFP (3224)).

[0359] 구체적으로, 각각의 타입 2 또는 타입 4 분석 벡터는 (예를 들어, 도 25 및 도 27 과 관련하여 설명된 바와 같이) 6 개의 특징들 (3227) 을 생성할 수도 있다. 특징들 (3227) 은 이 예에서, 54 개의 엔트리들을 가지는 특징 벡터에서 배치될 수도 있다. 특징 추출의 동일한 프로세스는 또 다른 54 개의 엔트리 특징 벡터를 생성하기 위하여 달팽이관 출력 (3215) 의 제 2 부분 (예컨대, 유성음 부분)에 대해 수행될 수도 있다. 동일한 프로세스는, 제 1 선택된 섹션의 특징 벡터와 유사한 특징 벡터를 각각 생성하는 임의의 적당한 수의 부분들 (예컨대, 유성음 부분들)에 대해 반복될 수도 있다.

[0360] 일단 모든 선택된 섹션들에 대한 특징 벡터들이 결정되면, 특징들 (3227) 은 도 32b 및 도 32c 에서 예시된 바와 같이 평균화될 수도 있다. 다시 말해서, 특징 벡터들의 각각에서의 (평균들로서 도시된) 제 1 엔트리들이 평균화될 수도 있다. 유사하게, 매 인덱스에서의 엔트리들은, 이 예에서 54 개의 엔트리 벡터로서 도시된, 단일의 평균 특징 벡터 (3220) 를 생성하기 위하여 서로 평균화된다. 단일 평균 특징 벡터 (3220) 는 (예를 들어, 도 8 내지 도 9 중의 하나 이상과 관련하여 설명된 바와 같은) 트레이닝 동안에 미리 학습된 (예컨대, 가중치들을 이용하여 가중화될 수도 있는) 가중화 (3253) 와 승산될 수도 있다. 가중화 후에, 단일 평균 특징 벡터에서의 특징들의 각각은 도 32d 에서 도시된 바와 같은 러프니스 왜곡 (3249) (예컨대, 예측 점수) 을 생성하기 위하여 합산될 수도 있다. 유사한 절차들이 다른 왜곡들에 대해 후속될 수도 있다.

[0361] 도 31 내지 도 32 는 러프니스 왜곡 (3149, 3249) 에 대한 예측 점수를 예시하지만, 그러나, 예측 점수는 왜곡

차원들 중의 임의의 것에 대한 것일 수도 있다. 예를 들어, 둔탁성/머플드니스 (muffledness) 왜곡 차원에 대한 예측 점수를 결정할 때, 이용된 가중치들은 러프니스 왜곡 차원 대신에, 둔탁성/머플드니스 왜곡 차원에 대한 주관적 점수들에 대응할 수도 있다. 각각의 왜곡 차원에 대한 예측 점수들은 포어그라운드 품질 및 백그라운드 품질을 결정하기 위하여 (예를 들어, 도 6 내지 도 7 과 관련하여 설명된 바와 같은) 객관적 스피치 품질 모듈에 의해 이용될 수도 있다. 다음으로, 전체 품질이 결정될 수도 있다. 전체 품질은 스피치의 부분을 주관적으로 점수화하는 인간 청자들로부터 결정되는 평균 의견 점수 (MOS) 대신에, 또는 평균 의견 점수 (MOS)에 추가하여 이용될 수도 있다.

[0362] 일부의 구성들에서, 시스템들 및 방법들의 하나 이상의 절차들은 다음과 같이 수행될 수도 있다. 하나의 접근법에서, 여기에서 도입된 새로운 특징 세트들은 (열화된 스피치 신호를 이용하여) 비-침투적이다. 열화된 스피치 (S^{dis})를 가정한다. S^{dis} 는 (유체-역학적 달팽이관 모델 또는 음향심리학적 모델과 같은) 청취의 연산 모델을 통과하게 될 수도 있다. 이러한 모델들은 CM으로서 지칭될 수도 있고, 왜곡된 신호로 인한 출력은 CM^{dis}로서 지칭될 것이다. 출력은 시간 및 주파수 도메인들의 양자를 따라 높은 정밀도를 가진다. 출력은 지각적 도메인에서의 스피치를 나타낸다.

[0363] 분석 시리얼들 (AS)은 분석의 목적에 따라서는, 본원에서 설명된 동일한 알고리즘들로 스피치의 유성음 색션들 (VS) 및 목음 (SIL) 색션들로부터 추출될 수도 있다. K 개의 관심 있는 색션이 있는 것으로 가정한다.

[0364] 주파수 관련 분석 시리얼들은 다음과 같이 설명될 수도 있다. k 번째 관심 있는 영역에 대하여, 열화된 CM

출력 $CM_k^{dis}(p, t)$ 은 P*T 의 사이즈를 가진다.

$$CMOT_k^{dis}(p) = \log_{10} \frac{\sum_{t=1}^T CM_k^{dis}(p, t)}{T} \quad (69)$$

[0365] [0366] 상이한 지각적 모델에 대하여, CM은 상이한 사이즈를 가질 수도 있다. 다음의 부분에서, 우리는 달팽이관 모델을 취한다. 숫자들은 어떤 출력에 대해 조절될 수도 있다는 것에 주목해야 한다. 이 예에서의 P는 범위 1 내지 512를 가지고, 4 개의 영역들로 절단된다: (A) 1 내지 150, (B) 151 내지 275, (C) 276 내지 450 및 (D) 451 내지 512. 따라서, 장소의 함수로서 4 개의 차분 함수들이 있다:

$$\begin{aligned} P1_k(p) &= CMOT_k^{dis}(p), \quad p = 1 \dots 150 \\ P2_k(p) &= CMOT_k^{dis}(p), \quad p = 151 \dots 275 \\ P3_k(p) &= CMOT_k^{dis}(p), \quad p = 276 \dots 450 \\ P4_k(p) &= CMOT_k^{dis}(p), \quad p = 451 \dots 512 \end{aligned} \quad (70)$$

[0367] 이것들은 분석 시리얼들로 칭해질 수도 있다.

[0368] [0369] 시간 관련 분석 시리얼들은 다음과 같이 획득될 수도 있다. 각각의 유성음 색션 (VS)에 대하여, 열화된 CM 출력 $CM_k^{dis}(p, t)$ 은 P*T 의 사이즈를 가진다. 다음과 같이 계산될 수도 있다:

$$\begin{aligned}
 T1_k^{dis}(t) &= \log_{10} \frac{\sum_{p=r1,1}^{r1,2} CM_k^{dis}(p,t)}{r1,2 - r1,1 + 1} \\
 T2_k^{dis}(t) &= \log_{10} \frac{\sum_{p=r2,1}^{r2,2} CM_k^{dis}(p,t)}{r2,2 - r2,1 + 1} \\
 T3_k^{dis}(t) &= \log_{10} \frac{\sum_{p=r3,1}^{r3,2} CM_k^{dis}(p,t)}{r3,2 - r3,1 + 1} \\
 T4_k^{dis}(t) &= \log_{10} \frac{\sum_{p=r4,1}^{r4,2} CM_k^{dis}(p,t)}{r4,2 - r4,1 + 1}
 \end{aligned} \tag{71}$$

r	1	2
1	1	150
2	151	275
3	276	450
4	450	512

[0370]

[0371] 여기서, $r_{1,2}$ 는 "제 1 행, 제 2 열" (예컨대, 150) 을 의미한다. 이것들은 또한 분석 시리얼들로 칭해질 수도 있다.

[0372] SFP 가 또한 획득될 수도 있다. 관심 있는 섹션이 유성음 섹션일 경우, SFP_k 는 k 번째 유성음 섹션의 CM 출력으로부터 추출된다. AS로부터의 특징 추출은 다음과 같이 진행할 수도 있다. N 의 길이를 가지는 각각의 AS (예컨대, $T3_k$) 에 대하여, 특징들은 다음과 같이 추출된다:

$$\begin{aligned}
 FT1 &= mean(AS) \\
 FT2 &= \frac{\sum_{p=1}^N AS(p)*p}{\sum_{p=1}^N AS(p)} - \bar{p},
 \end{aligned}$$

[0373]

기하학적 시프트.

$$FT3 = median(AS),$$

[0375] 여기에서의 중앙값 함수는 $p(AS \geq F3) \geq \frac{1}{2}$ 을 의미하고, 그리고

$$p(AS \leq F3) \geq \frac{1}{2}. \quad (72)$$

$$FT4 = \frac{1}{\sum_{p=1}^N \frac{1}{AS(p)}},$$

조화 평균.

$$FT5 = std(AS),$$

표준 편차 (std).

$$FT6 = \frac{\frac{1}{N} \sum_{p=1}^N (AS(p) - \bar{AS})^3}{\left[\frac{1}{N} \sum_{p=1}^N (AS(p) - \bar{AS})^2 \right]^3}$$

[0377] $P1_k$ 로부터 추출된 특징들은 F1 내지 F6 이다.

$P3_k$ 로부터 추출된 특징들은 F13 내지 F18 이다.

$P2_k$ 로부터 추출된 특징들은 F7 내지 F12 이다.

$P4_k$ 로부터 추출된 특징들은 F19 내지 F24 이다.

[0378] $T1_k$ 로부터 추출된 특징들은 F25 내지 F30 이다.

$T2_k$ 로부터 추출된 특징들은 F31 내지 F36 이다.

$T3_k$ 로부터 추출된 특징들은 F37 내지 F42 이다.

$T4_k$ 로부터 추출된 특징들은 F43 내지 F48 이다.

SFP_k 로부터 추출된 특징들은 F49 내지 F54 이다.

[0379] 최종적인 특징 점수는 모든 유성음 색션들에 대한 그 특징의 평균들을 포함할 수도 있다는 것 (예컨대,

$FS1 = \frac{1}{V} \sum_{v=1}^V F1_k$)에 주목해야 한다. 특징 F55 는 화자가 남성 또는 여성인지 여부의 이진 표시자 (예컨대, 남성에 대해 0, 그리고 여성에 대해 -1) 일 수도 있다. 특징 F56 은 선형 회귀의 편리함을 위하여 상수 (예를 들어, 보통 1로 설정됨) 일 수도 있다.

[0380] 본원에서 개시된 시스템들 및 방법들은 스피치 품질 측정 출력을 제공할 수도 있다. 코딩 및/또는 송신 개발자를 위한 피드백 (예컨대, 실시간 피드백) 이 추가적으로 또는 대안적으로 제공될 수도 있다. 스피치 왜곡들의 특성을 및/또는 통찰은 추가적으로 또는 대안적으로 제공될 수도 있다. 본원에서 개시된 시스템들 및 방법들은 지각적 도메인에서 주어진 스피치의 특성을 (예컨대, 왜곡들) 을 설명하기 위하여 많은 파라미터들을 제공할 수도 있다. 이 특징들은 일반적인 툴박스 (toolbox) 내로 구축될 수도 있고 특정 목적들 (예컨대, 주파수-국소화된 왜곡들 측정) 을 위해 조합될 수도 있다. 다시 말해서, 본원에서 설명된 시스템들 및 방법들은 스피치 품질 측정, 다양한 타입들의 스피치 왜곡들에 관한 설명에 대한 통찰, 인간-지각적으로 지향된 접근법에서 높은 정확도를 제공할 수도 있다. 다른 알려진 접근법들은 유체-역학적 달팽이관 모델 출력을 이용하지 않을 수도 있다는 것에 주목해야 한다.

[0381] 도 33 은 특징 추출을 위한 방법 (3300) 의 더욱 구체적인 구성을 예시하는 흐름도이다. 전자 디바이스 (1507) (예컨대, 스피치 평가기 및/또는 특징 추출 모듈) 는 방법 (3300) 의 하나 이상의 단계들, 기능들 및/또는 절차들을 수행할 수도 있다.

[0382] 전자 디바이스는 원본 스피치 신호 및 수정된 (예컨대, 열화된) 스피치 신호를 획득할 수도 있다. 원본 스피치 신호 및 수정된 스피치 신호는 침투적 특징 추출을 위하여 동일한 주기 동안에 레코딩될 수도 있다. 다시 말해서, 수정된 스피치 신호는 원본 스피치 신호의 프로세싱된 버전일 수도 있다. 대안적으로, 수정된 스피치 신호만이 비-침투적 특징 추출을 위하여 획득될 수도 있다.

[0383] 전자 디바이스는 생리학적 달팽이관 모델을 이용하여 스피치를 프로세싱할 수도 있다 (3302). 이것은 예를 들어, 도 15 내지 도 19, 및 도 31 내지 도 32 중의 하나 이상과 관련하여 위에서 설명된 바와 같이 달성될 수도 있다.

[0384] 전자 디바이스는 달팽이관 모델의 출력으로부터 부분을 선택할 수도 있다 (3304). 이것은 예를 들어, 도 18 내지 도 19, 및 도 31 내지 도 32 중의 하나 이상과 관련하여 위에서 설명된 바와 같이 달성될 수도 있다. 부분은 원본 스피치 신호에 대한 생리학적 달팽이관 모델 출력의 부분, 및/또는 스피치 (예컨대, 유성음 부분들, 모음들 등)의 지각적 품질에 상당히 영향을 주는 수정된 스피치 신호에 대한 달팽이관 모델 출력의 부분일 수도 있다. 예를 들어, 부분을 선택하는 것 (3304)은 하나 이상의 신호들의 에너지를 연산하는 것, 및 에너지가 임계점을 초과하는 부분을 선택하는 것을 포함할 수도 있다.

[0385] 전자 디바이스는 생리학적 달팽이관 모델의 출력의 섹션들을 분석할 수도 있다 (3306). 이것은 예를 들어, 도 15 내지 도 16, 도 18 내지 도 24, 및 도 31 내지 도 32 중의 하나 이상과 관련하여 위에서 설명된 바와 같이 달성될 수도 있다. 예를 들어, 전자 디바이스는 출력의 부분들을 다수의 섹션들 (예컨대, 4 개의 섹션들 또는 또 다른 수의 섹션들)로 그룹화 (예컨대, 분할) 할 수도 있다. 섹션들의 각각은 특정 사이즈 (예컨대, 장소 포인트들의 수 X N 샘플들의 수)를 가질 수도 있다.

[0386] 전자 디바이스는 각각의 섹션에 대한 벡터들을 추출할 수도 있다 (3308). 특히, 전자 디바이스는 각각의 섹션에 대한 장소-기반 분석 벡터 및 시간-기반 분석 벡터를 추출할 수도 있다. 이것은 예를 들어, 도 15 내지 도 16, 도 18 내지 도 19, 도 21 내지 도 24 및 도 31 내지 도 32 중의 하나 이상과 관련하여 위에서 설명된 바와 같이 달성될 수도 있다. 예를 들어, 전자 디바이스는 장소-기반 분석 벡터를 생성하기 위하여 시간에 대하여 섹션을 평균화할 수도 있고, 시간-기반 분석 벡터를 생성하기 위하여 장소에 대하여 섹션을 평균화할 수도 있다. 일부의 구성들에서, 분석 벡터들은 상기 표 3에서 도시된 바와 같은 섹션 번호마다 추출될 수도 있다 (3308). 다시 말해서, 적어도 하나의 장소-기반 분석 벡터 및 하나의 시간-기반 분석 벡터는 각각의 섹션에 대하여 추출될 수도 있다 (3308).

[0387] 전자 디바이스는 각각의 벡터 (예컨대, 분석 벡터)로부터 하나 이상의 특징들을 결정할 수도 있다 (3310). 이것은 예를 들어, 도 15 내지 도 16, 도 18 내지 도 19, 도 25 내지 도 27, 및 도 31 내지 도 32 중의 하나 이상과 관련하여 설명된 바와 같이 달성될 수도 있다.

[0388] 전자 디바이스는 하나 이상의 특징들과, 왜곡 차원에 대응하는 가중화 (예컨대, 가중치 세트)에 기초하여 왜곡을 추정 (예컨대, 예측) 할 수도 있다 (3312). 이것은 도 15, 도 18 내지 도 19, 및 도 31 내지 도 32 중의 하나 이상과 관련하여 위에서 설명된 바와 같이 달성될 수도 있다. 예를 들어, 전자 디바이스는 특징들을 가중화하여 러프니스 왜곡 (예컨대, 러프니스 차원에 대한 예측 점수)을 생성하기 위하여 러프니스 왜곡에 대한 가중화 (예컨대, 오프라인으로 이전에 결정되었던 가중치 세트)를 이용할 수도 있다. 하나 이상의 왜곡들이 추정될 수도 있다 (3312). 예를 들어, 전자 디바이스는 위에서 설명된 바와 같은 러프니스, 불연속성, 둔탁성, 빈약함, 허성 및 가변성 중의 하나 이상을 추정할 수도 있다 (3312). 전자 디바이스는 추가적인 및/또는 대안적인 왜곡들을 추정할 수도 있다 (3312). 일부의 구성들에서, 왜곡들 (예컨대, 예측 점수들)은 도 4 내지 도 7과 관련하여 위에서 설명된 전부 4 개의 포어그라운드 왜곡 차원들 및 2 개의 백그라운드 왜곡 차원들에 대해 결정될 수도 있다. 일부의 구성들에서, 전자 디바이스는 왜곡들 (예컨대, 객관적 왜곡들)에 기초하여 포어그라운드 품질 및/또는 백그라운드 품질을 추정할 수도 있다. 이것은 도 4 내지 도 7 중의 하나 이상과 관련하여 위에서 설명된 바와 같이 달성될 수도 있다. 추가적으로 또는 대안적으로, 전체 품질은 포어그라운드 품질 및 백그라운드 품질로부터 추정될 수도 있다. 이것은 도 4 내지 도 7 중의 하나 이상과 관련하여 위에서 설명된 바와 같이 달성될 수도 있다. 전체 품질은, 주관적인 평균 의견 점수 (MOS) 대신에 사용될 수도 있는 수정된 스피치 신호의 전체적인 객관적 품질 점수로서 작용할 수도 있다.

[0389] 도 34는 침투적 특징 추출을 위한 방법 (3400)의 구성을 예시하는 흐름도이다. 전자 디바이스 (1507) (예컨대, 스피치 평가기 및/또는 특징 추출 모듈)는 방법 (3400)의 하나 이상의 단계들, 기능들 및/또는 절차들을 수행할 수도 있다.

[0390] 전자 디바이스는 원본 스피치 신호 및 수정된 (예컨대, 열화된) 스피치 신호를 획득할 수도 있다 (3402). 이것은 도 15와 관련하여 위에서 설명된 바와 같이 달성될 수도 있다. 예를 들어, 전자 디바이스는 원본 스피치 신호 및 수정된 스피치 신호의 하나 이상의 프레임들 또는 시간 주기들을 획득할 수도 있다. 원본 스피치 신호 및 수정된 스피치 신호는 침투적 특징 추출을 위하여 동일한 시간 주기 동안에 레코딩될 수도

있다. 다시 말해서, 수정된 스피치 신호는 원본 스피치 신호의 프로세싱된 버전일 수도 있다.

[0391] 전자 디바이스는 생리학적 달팽이관 모델을 이용하여 스피치를 프로세싱할 수도 있다 (3404). 이것은 예를 들어, 도 15 내지 도 19, 및 도 31 내지 도 32 중의 하나 이상과 관련하여 위에서 설명된 바와 같이 달성될 수도 있다. 특히, 전자 디바이스는 하나 이상의 달팽이관 모델들을 이용하여 원본 스피치 신호 및 수정된 스피치 신호를 프로세싱할 수도 있다 (3404).

[0392] 전자 디바이스는 생리학적 달팽이관 모델의 출력으로부터 부분을 선택할 수도 있다 (3306). 이것은 예를 들어, 도 18 내지 도 19, 및 도 31 내지 도 32 중의 하나 이상과 관련하여 위에서 설명된 바와 같이 달성될 수도 있다. 부분은 원본 스피치 신호에 대한 생리학적 달팽이관 모델 출력의 부분, 및/또는 스피치 (예컨대, 유성음 부분들, 모음들 등)의 지각적 품질에 상당히 영향을 주는 수정된 스피치 신호에 대한 달팽이관 모델 출력의 부분일 수도 있다. 예를 들어, 부분을 선택하는 것 (3406)은 하나 이상의 신호들의 에너지를 연산하는 것, 및 에너지가 임계점을 초과하는 부분을 선택하는 것을 포함할 수도 있다.

[0393] 전자 디바이스는 생리학적 달팽이관 모델의 출력의 섹션들을 분석할 수도 있다 (3408). 이것은 예를 들어, 도 15 내지 도 16, 도 18 내지 도 24, 및 도 31 내지 도 32 중의 하나 이상과 관련하여 위에서 설명된 바와 같이 달성될 수도 있다. 예를 들어, 전자 디바이스는 출력의 부분들을 다수의 섹션들 (예컨대, 4 개의 섹션들 또는 또 다른 수의 섹션들)로 그룹화 (예컨대, 분할) 할 수도 있다. 섹션들의 각각은 특정 사이즈 (예컨대, 장소 포인트들의 수 X N 샘플들의 수)를 가질 수도 있다.

[0394] 전자 디바이스는 각각의 섹션에 대한 벡터들을 추출할 수도 있다 (3410). 특히, 전자 디바이스는 각각의 섹션에 대한 장소-기반 분석 벡터 및 시간-기반 분석 벡터를 추출할 수도 있다 (3410). 이것은 예를 들어, 도 15 내지 도 16, 도 18 내지 도 19, 도 21 내지 도 24 및 도 31 내지 도 32 중의 하나 이상과 관련하여 위에서 설명된 바와 같이 달성될 수도 있다. 예를 들어, 전자 디바이스는 장소-기반 분석 벡터를 생성하기 위하여 시간에 대하여 섹션을 평균화할 수도 있고, 시간-기반 분석 벡터를 생성하기 위하여 장소에 대하여 섹션을 평균화할 수도 있다. 일부의 구성들에서, 분석 벡터들은 상기 표 3에서 도시된 바와 같은 섹션 번호마다 추출될 수도 있다 (3410). 다시 말해서, 적어도 하나의 침투적 장소-기반 분석 벡터 (예컨대, 타입 1) 및 하나의 시간-기반 분석 벡터 (예컨대, 타입 3)는 각각의 섹션에 대하여 추출될 수도 있다 (3410). 표 3에서 예시된 바와 같이, 전자 디바이스는 선택적으로 그리고 추가적으로, 하나 이상의 섹션들에 대하여 비-침투적 장소-기반 벡터 (예컨대, 타입 2) 및/또는 비-침투적 시간-기반 벡터 (예컨대, 타입 4)를 추출할 수도 있다.

[0395] 전자 디바이스는 각각의 벡터 (예컨대, 분석 벡터)로부터 하나 이상의 특징들을 결정할 수도 있다 (3412). 이것은 예를 들어, 도 15 내지 도 16, 도 18 내지 도 19, 도 25 내지 도 27, 및 도 31 내지 도 32 중의 하나 이상과 관련하여 설명된 바와 같이 달성될 수도 있다. 일부의 구성들에서, 6 및/또는 18 개의 특징들은 각각의 벡터에 기초하여 결정될 수도 있다 (3412).

[0396] 전자 디바이스는 하나 이상의 특징들과, 왜곡 차원에 대응하는 가중화 (예컨대, 가중치 세트)에 기초하여 왜곡을 추정 (예컨대, 예측) 할 수도 있다 (3414). 이것은 도 15, 도 18 내지 도 19, 및 도 31 내지 도 32 중의 하나 이상과 관련하여 위에서 설명된 바와 같이 달성될 수도 있다. 예를 들어, 전자 디바이스는 특징들을 가중화하여 러프니스 왜곡 (예컨대, 러프니스 차원에 대한 예측 점수)을 생성하기 위하여 러프니스 왜곡에 대한 가중화 (예컨대, 오프라인으로 이전에 결정되었던 가중치 세트)를 이용할 수도 있다. 하나 이상의 왜곡들이 추정될 수도 있다 (3414). 예를 들어, 전자 디바이스는 위에서 설명된 바와 같은 러프니스, 불연속성, 둔탁성, 빈약함, 허싱 및 가변성 중의 하나 이상을 추정할 수도 있다 (3414). 전자 디바이스는 추가적인 및/또는 대안적인 왜곡들을 추정할 수도 있다 (3414). 일부의 구성들에서, 왜곡들 (예컨대, 예측 점수들)은 도 4 내지 도 7과 관련하여 위에서 설명된 전부 4 개의 포어그라운드 왜곡 차원들 및 2 개의 백그라운드 왜곡 차원들에 대해 결정될 수도 있다. 일부의 구성들에서, 전자 디바이스는 왜곡들 (예컨대, 객관적 왜곡들)에 기초하여 포어그라운드 품질 및/또는 백그라운드 품질을 추정할 수도 있다. 이것은 도 4 내지 도 7 중의 하나 이상과 관련하여 위에서 설명된 바와 같이 달성될 수도 있다. 추가적으로 또는 대안적으로, 전체 품질은 포어그라운드 품질 및 백그라운드 품질로부터 추정될 수도 있다. 이것은 도 4 내지 도 7 중의 하나 이상과 관련하여 위에서 설명된 바와 같이 달성될 수도 있다. 전체 품질은, 주관적인 평균 의견 점수 (MOS) 대신에 사용될 수도 있는 수정된 스피치 신호의 전체적인 객관적 품질 점수로서 작용할 수도 있다.

[0397] 도 35는 비-침투적 특징 추출을 위한 방법 (3500)의 구성을 예시하는 흐름도이다. 전자 디바이스 (1507) (예컨대, 스피치 평가기 및/또는 특징 추출 모듈)는 방법 (3500)의 하나 이상의 단계들, 기능들 및/또는 절차들을 수행할 수도 있다.

[0398] 전자 디바이스는 수정된 (예컨대, 열화된) 스피치 신호를 획득할 수도 있다 (3502). 이것은 도 15 와 관련하여 위에서 설명된 바와 같이 달성될 수도 있다. 예를 들어, 전자 디바이스는 수정된 스피치 신호의 하나 이상의 프레임들 또는 시간 주기들을 획득할 수도 있다. 수정된 스피치 신호는 원본 스피치 신호의 프로세싱된 (예컨대, 열화된) 버전일 수도 있다. 비-침투적 접근법에서, 전자 디바이스는 원본 스피치 신호를 획득하지 않을 수도 있다. 이것은 원본 스피치 신호를 획득, 송신, 또는 수신할 필요 없이 다양한 디바이스들에서 스피치 품질 평가를 허용할 수도 있다.

[0399] 전자 디바이스는 생리학적 달팽이관 모델을 이용하여 수정된 스피치를 프로세싱할 수도 있다 (3504). 이것은 예를 들어, 도 15 내지 도 17, 도 19 및 도 32 중의 하나 이상과 관련하여 위에서 설명된 바와 같이 달성될 수도 있다. 특히, 전자 디바이스는 달팽이관 모델을 이용하여 수정된 스피치 신호 (그리고 원본 스피치 신호는 아님) 만을 프로세싱할 수도 있다 (3504).

[0400] 전자 디바이스는 생리학적 달팽이관 모델의 출력으로부터 부분을 선택할 수도 있다 (3506). 이것은 예를 들어, 도 19 및 도 32 중의 하나 이상과 관련하여 위에서 설명된 바와 같이 달성될 수도 있다. 부분은 스피치 (예컨대, 유성음 부분들, 모음들 등) 의 지각적 품질에 상당히 영향을 주는 수정된 스피치 신호에 대한 생리학적 달팽이관 모델 출력의 부분일 수도 있다. 예를 들어, 부분을 선택하는 것 (3506) 은 수정된 스피치 신호의 에너지를 연산하는 것, 및 에너지가 임계점을 초과하는 부분을 선택하는 것을 포함할 수도 있다.

[0401] 전자 디바이스는 생리학적 달팽이관 모델의 출력의 섹션들을 분석할 수도 있다 (3508). 이것은 예를 들어, 도 15 내지 도 16, 도 19 내지 도 20, 도 22, 도 24 및 도 32 중의 하나 이상과 관련하여 위에서 설명된 바와 같이 달성될 수도 있다. 예를 들어, 전자 디바이스는 출력의 부분들을 다수의 섹션들 (예컨대, 4 개의 섹션들 또는 또 다른 수의 섹션들) 로 그룹화 (예컨대, 분할) 할 수도 있다. 섹션들의 각각은 특정 사이즈 (예컨대, 장소 포인트들의 수 X N 샘플들의 수) 를 가질 수도 있다.

[0402] 전자 디바이스는 각각의 섹션에 대한 벡터들을 추출할 수도 있다 (3510). 특히, 전자 디바이스는 각각의 섹션에 대한 비-침투적 장소-기반 분석 벡터 및 비-침투적 시간-기반 분석 벡터를 추출할 수도 있다 (3510). 이것은 예를 들어, 도 15 내지 도 16, 도 19, 도 22, 도 24 및 도 32 중의 하나 이상과 관련하여 위에서 설명된 바와 같이 달성될 수도 있다. 예를 들어, 전자 디바이스는 장소-기반 분석 벡터를 생성하기 위하여 시간에 대하여 섹션을 평균화할 수도 있고, 시간-기반 분석 벡터를 생성하기 위하여 장소에 대하여 섹션을 평균화할 수도 있다. 일부의 구성들에서, 분석 벡터들은 상기 표 4 에서 도시된 바와 같은 섹션 번호마다 추출될 수도 있다 (3510). 다시 말해서, 적어도 하나의 비-침투적 장소-기반 분석 벡터 (예컨대, 타입 2) 및 하나의 비-침투적 시간-기반 분석 벡터 (예컨대, 타입 4) 는 각각의 섹션에 대하여 추출될 수도 있다 (3510).

[0403] 전자 디바이스는 각각의 벡터 (예컨대, 분석 벡터)로부터 하나 이상의 특징들을 결정할 수도 있다 (3512). 이것은 예를 들어, 도 15 내지 도 16, 도 19, 도 25, 도 27 및 도 32 중의 하나 이상과 관련하여 설명된 바와 같이 달성될 수도 있다. 일부의 구성들에서, 6 개의 특징들은 각각의 벡터에 기초하여 결정될 수도 있다 (3512).

[0404] 전자 디바이스는 하나 이상의 특징들과, 왜곡 차원에 대응하는 가중화 (예컨대, 가중치 세트) 에 기초하여 왜곡을 추정 (예컨대, 예측) 할 수도 있다 (3514). 이것은 도 15, 도 19 및 도 32 중의 하나 이상과 관련하여 위에서 설명된 바와 같이 달성될 수도 있다. 예를 들어, 전자 디바이스는 특징들을 가중화하여 러프니스 왜곡 (예컨대, 러프니스 차원에 대한 예측 점수) 을 생성하기 위하여 러프니스 왜곡에 대한 가중화 (예컨대, 오프라인으로 이전에 결정되었던 가중치 세트) 를 이용할 수도 있다. 하나 이상의 왜곡들이 추정될 수도 있다 (3514). 예를 들어, 전자 디바이스는 위에서 설명된 바와 같은 러프니스, 불연속성, 둔탁성, 빈약함, 허싱 및 가변성 중의 하나 이상을 추정할 수도 있다 (3514). 전자 디바이스는 추가적인 및/또는 대안적인 왜곡들을 추정할 수도 있다 (3514). 일부의 구성들에서, 왜곡들 (예컨대, 예측 점수들) 은 도 4 내지 도 7 과 관련하여 위에서 설명된 전부 4 개의 포어그라운드 왜곡 차원들 및 2 개의 백그라운드 왜곡 차원들에 대해 결정될 수도 있다. 일부의 구성들에서, 전자 디바이스는 왜곡들 (예컨대, 객관적 왜곡들) 에 기초하여 포어그라운드 품질 및/또는 백그라운드 품질을 추정할 수도 있다. 이것은 도 4 내지 도 7 중의 하나 이상과 관련하여 위에서 설명된 바와 같이 달성될 수도 있다. 추가적으로 또는 대안적으로, 전체 품질은 포어그라운드 품질 및 백그라운드 품질로부터 추정될 수도 있다. 이것은 도 4 내지 도 7 중의 하나 이상과 관련하여 위에서 설명된 바와 같이 달성될 수도 있다. 전체 품질은, 주관적인 평균 의견 점수 (MOS) 대신에 사용될 수도 있는 수정된 스피치 신호의 전체적인 객관적 품질 점수로서 작용할 수도 있다.

[0405] 도 36 은 침투적 특징 추출의 더욱 구체적인 구성을 예시하는 블록도이다. 예를 들어, 도 36 은 IHC 전압으

로부터 핵심 특징들을 추출하는 예를 제공한다. 신경망들과 같은 백엔드 (backend) 에서의 더욱 복잡한 통계적 모델들은 성능을 개선시킬 수도 있고 추가적으로 대안적으로 사용될 수도 있다는 것에 주목해야 한다. 도 26 과 관련하여 설명된 컴포넌트들 중의 하나 이상은 컴포넌트들 중의 하나 이상의 컴포넌트의 예들을 수도 있고, 및/또는 도 15 내지 도 27 및 도 31 내지 도 35 중의 하나 이상과 관련하여 설명된 절차들 중의 하나 이상을 수행할 수도 있다. 특히, 도 36 은 지연 추정 모듈 (3637), 달팽이관 모델들 A 내지 B (3613a 내지 3613b), 감산기 (3620), 부분 선택 모듈 (3643), 주파수에 대한 평균 모듈 (3622), 시간에 대한 평균 모듈 (3624), 특징 결정 모듈 (3625) 및/또는 회귀 모듈 (3651) 의 예들을 예시한다.

[0406] 원본 스피치 신호 (3609) 및 수정된 스피치 신호 (3611) (예컨대, 원본 스피치 신호 (3609) 의 열화된 버전) 는 지연 추정 모듈 (3637) 에 제공될 수도 있다. 지연 추정 모듈 (3637) 은, 달팽이관 모델 A (3613a) 및 달팽이관 모델 B (3613b) 에 제공될 수도 있는 원본 스피치 신호 (3609) 및 수정된 스피치 신호 (3611) 를 정렬할 수도 있다. 달팽이관 모델 A (3613a) 의 출력은 감산기 (3620) 에 의해 달팽이관 모델 B (3613b) 의 출력으로부터 감산될 수도 있다. 이 차이는 부분 선택 모듈 (3643) 에 제공될 수도 있다. 부분 선택 모듈 (3643) 은 차이로부터 부분들을 선택할 수도 있다. 예를 들어, 차이의 유성음 부분들이 선택될 수도 있다.

[0407] 유성음 부분들은 주파수에 대한 평균 모듈 (3622), 그리고 시간에 대한 평균 모듈 (3624) 에 제공될 수도 있다. 주파수에 대한 평균 모듈 (3622) 및 시간에 대한 평균 모듈 (3624) 은 주파수 및 시간에 대해 유성음 부분들을 각각 평균화할 수도 있다. 주파수에 대한 평균 모듈 (3622) 은 하나 이상의 시간 영역들을 생성할 수도 있다. 시간에 대한 평균 모듈 (3624) 은 하나 이상의 주파수 영역들 (3626) 을 생성할 수도 있다. 하나 이상의 주파수 영역들 (3626) 은 위에서 설명된 장소-기반 벡터들의 또 다른 표현일 수도 있다. 주파수에 대한 평균 모듈 (3622) 은 하나 이상의 시간 영역들 (3628) 을 생성할 수도 있다. 하나 이상의 시간 영역들 (3628) 은 위에서 설명된 시간-기반 벡터들에 대응할 수도 있다.

[0408] 주파수 영역 (들) (3626) 및/또는 시간 영역 (들) (3628) 은 특징 결정 모듈 (3625) 에 제공될 수도 있다. 특징 결정 모듈 (3625) 은 각각의 영역으로부터 특징들 (3627) (예컨대, 통계적 특징들) 을 결정 (예컨대, 추출) 할 수도 있다. 특징들 (3627) 은 특징 세트들로 그룹화될 수도 있다.

[0409] 일부의 구성들에서, 특징들 (3627) 은 회귀 모듈 (3651) 에 제공될 수도 있다. 회귀 모듈 (3651) 은 하나 이상의 왜곡들 (3649) 을 추정하기 위하여, 특징들 (3627) 중의 하나 이상 및 가중화 (3653) (예컨대, 하나 이상의 가중치들) 에 기초하여 선형 회귀를 수행할 수도 있다. 일부의 구성들에서, 회귀 모듈 (3651) 은 대안적으로, 왜곡 (들) (3649) 을 추정하기 위하여, 다향식 회귀, 2 차 회귀, 비선형 회귀 등을 수행할 수도 있다. 일부의 구성들에서, 하나 이상의 품질들 (예컨대, 포어그라운드 품질, 백그라운드 품질, 전체 품질 등) 은 본원에서 설명된 바와 같은 하나 이상의 왜곡들 (3649) 에 기초하여 추정될 수도 있다.

[0410] 도 37 은 스피치 신호 품질을 측정하고 및/또는 특징 추출을 위한 시스템들 및 방법들이 구현될 수도 있는 무선 통신 디바이스 (3730) 의 하나의 구성을 예시하는 블록도이다. 도 37 에서 예시된 무선 통신 디바이스 (3730) 는 본원에서 설명된 전자 디바이스들 (556, 1507) 중의 하나 이상의 전자 디바이스의 예일 수도 있다. 무선 통신 디바이스 (3730) 는 애플리케이션 프로세서 (3742) 를 포함할 수도 있다. 애플리케이션 프로세서 (3742) 는 일반적으로 무선 통신 디바이스 (3730) 상에서 기능들을 수행하기 위하여 명령들을 프로세싱한다 (예컨대, 프로그램들을 실행함). 애플리케이션 프로세서 (3742) 는 오디오 코더/디코더 (코덱) (3740) 에 결합될 수도 있다.

[0411] 오디오 코덱 (3740) 은 오디오 신호들을 코딩하고 및/또는 디코딩하기 위해 이용될 수도 있다. 오디오 코덱 (3740) 은 적어도 하나의 스피커 (3732), 이어폰 (earpiece; 3734), 출력 잭 (3736) 및/또는 적어도 하나의 마이크로폰 (3738) 에 결합될 수도 있다. 스피커들 (3732) 은 전기 또는 전자 신호들을 음향 신호들로 변환하는 하나 이상의 전기-음향 트랜스듀서들을 포함할 수도 있다. 예를 들어, 스피커들 (3732) 은 음악을 연주하거나 또는 스피커폰 대화를 출력하는 등을 하기 위해 이용될 수도 있다. 이어폰 (3734) 은 음향 신호들 (예컨대, 음성 신호들) 을 사용자에게 출력하기 위해 이용될 수 있는 또 다른 스피커 또는 전기-음향 트랜스듀서일 수도 있다. 예를 들어, 이어폰 (3734) 은 사용자만이 음향 신호를 신뢰성있게 청취할 수 있도록 이용될 수도 있다. 출력 잭 (3736) 은 다른 디바이스들을, 헤드폰들과 같이, 오디오를 출력하기 위한 무선 통신 디바이스 (3730) 에 결합하기 위해 이용될 수도 있다. 스피커들 (3732), 이어폰 (3734) 및/또는 출력 잭 (3736) 은 일반적으로 오디오 코덱 (3740) 으로부터 오디오 신호를 출력하기 위해 이용될 수도 있다. 적어도 하나의 마이크로폰 (3738) 은 (사용자의 음성과 같은) 음향 신호를, 오디오 코덱 (3740) 에 제공되는 전기 또는 전자 신호들로 변환하는 음향-전기 트랜스듀서일 수도 있다.

[0412]

일부의 구성들에서, 오디오 코덱 (3740) 은 스피치 평가기 A (3768a) 를 포함할 수도 있다. 스피치 평가기 A (3768a) 는 절차들 및/또는 기능들 중의 하나 이상을 수행할 수도 있고, 및/또는 도 4 내지 도 27 및 도 31 내지 도 36 중의 하나 이상과 관련하여 설명된 모듈들 및/또는 컴포넌트들 중의 하나 이상을 포함할 수도 있다. 특히, 스피치 평가기 A (3768a) 는 하나 이상의 특징들을 추출할 수도 있고, 및/또는 스피치 신호 품질을 측정할 수도 있다. 추가적으로 또는 대안적으로, 애플리케이션 프로세서 (3742) 는 스피치 평가기 B (3768b) 를 포함할 수도 있다. 스피치 평가기 B (3768b) 는 절차들 및/또는 기능들 중의 하나 이상을 수행할 수도 있고, 및/또는 스피치 평가기 A (3768a) 와 관련하여 설명된 모듈들 및/또는 컴포넌트들 중의 하나 이상을 포함할 수도 있다.

[0413]

애플리케이션 프로세서 (3742) 는 또한 전력 관리 회로 (3752) 에 결합될 수도 있다. 전력 관리 회로 (3752) 의 하나의 예는, 무선 통신 디바이스 (3730) 의 전기 전력 소비를 관리하기 위해 이용될 수도 있는 전력 관리 집적 회로 (power management integrated circuit; PMIC) 이다. 전력 관리 회로 (3752) 는 배터리 (3754) 에 결합될 수도 있다. 배터리 (3754) 는 일반적으로 전력을 무선 통신 디바이스 (3730) 에 제공할 수도 있다. 예를 들어, 배터리 (3754) 및/또는 전력 관리 회로 (3752) 는 무선 통신 디바이스 (3730) 내에 포함된 엘리먼트들 중의 적어도 하나에 결합될 수도 있다.

[0414]

애플리케이션 프로세서 (3742) 는 입력을 수신하기 위한 적어도 하나의 입력 디바이스 (3756) 에 결합될 수도 있다. 입력 디바이스들 (3756) 의 예들은 적외선 센서들, 이미지 센서들, 가속도계들, 터치 센서들, 키패드들 등을 포함한다. 입력 디바이스들 (3756) 은 무선 통신 디바이스 (3730) 와의 사용자 상호작용을 허용할 수도 있다. 애플리케이션 프로세서 (3742) 는 또한 하나 이상의 출력 디바이스들 (3758) 에 결합될 수도 있다. 출력 디바이스들 (3758) 의 예들은 프린터들, 프로젝터들, 스크린들, 햅틱 디바이스들 등을 포함한다. 출력 디바이스들 (3758) 은 무선 통신 디바이스 (3730) 가 사용자에 의해 경험될 수도 있는 출력을 생성하도록 할 수도 있다.

[0415]

애플리케이션 프로세서 (3742) 는 애플리케이션 메모리 (3760) 에 결합될 수도 있다. 애플리케이션 메모리 (3760) 는 전자 정보를 저장할 수 있는 임의의 전자 디바이스일 수도 있다. 애플리케이션 메모리 (3760) 의 예들은 이중 데이터 레이트 동기식 동적 랜덤 액세스 메모리 (double data rate synchronous dynamic random access memory; DDR SDRAM), 동기식 동적 랜덤 액세스 메모리 (synchronous dynamic random access memory; SDRAM), 플래시 메모리 등을 포함한다. 애플리케이션 메모리 (3760) 는 애플리케이션 프로세서 (3742) 를 위한 저장을 제공할 수도 있다. 예를 들어, 애플리케이션 메모리 (3760) 는 애플리케이션 프로세서 (3742) 상에서 실행되는 프로그램들의 기능을 위한 데이터 및/또는 명령들을 저장할 수도 있다.

[0416]

애플리케이션 프로세서 (3742) 는 디스플레이 제어기 (3762) 에 결합될 수도 있고, 이 디스플레이 제어기 (3762) 는 궁극적으로 디스플레이 (3764) 에 결합될 수도 있다. 디스플레이 제어기 (3762) 는 디스플레이 (3764) 상에서 이미지들을 생성하기 위해 이용되는 하드웨어 블록일 수도 있다. 예를 들어, 디스플레이 제어기 (3762) 는 애플리케이션 프로세서 (3742) 로부터의 명령들 및/또는 데이터를, 디스플레이 (3764) 상에서 제시될 수 있는 이미지들로 변환할 수도 있다. 디스플레이 (3764) 의 예들은 액정 디스플레이 (liquid crystal display; LCD) 패널들, 발광 다이오드 (light emitting diode; LED) 패널들, 음극선관 (cathode ray tube; CRT) 디스플레이들, 플라즈마 디스플레이들 등을 포함한다.

[0417]

애플리케이션 프로세서 (3742) 는 기저대역 프로세서 (3744) 에 결합될 수도 있다. 기저대역 프로세서 (3744) 는 일반적으로 통신 신호들을 프로세싱한다. 예를 들어, 기저대역 프로세서 (3744) 는 수신된 신호들을 복조하고 및/또는 디코딩할 수도 있다. 추가적으로 또는 대안적으로, 기저대역 프로세서 (3744) 는 송신을 준비하여 신호들을 인코딩하고 및/또는 변조할 수도 있다.

[0418]

기저대역 프로세서 (3744) 는 기저대역 메모리 (3766) 에 결합될 수도 있다. 기저대역 메모리 (3766) 는 SDRAM, DDRAM, 플래시 메모리 등과 같이, 전자 정보를 저장할 수 있는 임의의 전자 디바이스일 수도 있다. 기저대역 프로세서 (3744) 는 기저대역 메모리 (3766) 로부터 정보 (예컨대, 명령들 및/또는 데이터) 를 판독하고 및/또는 정보를 기저대역 메모리 (3766) 에 기록할 수도 있다. 추가적으로 또는 대안적으로, 기저대역 프로세서 (3744) 는 통신 동작들을 수행하기 위하여 기저대역 메모리 (3766) 내에 저장된 명령들 및/또는 데이터를 이용할 수도 있다.

[0419]

기저대역 프로세서 (3744) 는 라디오 주파수 (radio frequency; RF) 트랜시버 (3746) 에 결합될 수도 있다. RF 트랜시버 (3746) 는 전력 증폭기 (3748) 및 하나 이상의 안테나들 (3750) 에 결합될 수도 있다. RF 트랜시버 (3746) 는 라디오 주파수 신호들을 송신 및/또는 수신할 수도 있다. 예를 들어, RF 트랜시버 (3746)

는 전력 증폭기 (3748) 및 적어도 하나의 안테나 (3750) 를 이용하여 RF 신호를 송신할 수도 있다. RF 트랜서버 (3746) 는 또한 하나 이상의 안테나들 (3750) 을 이용하여 RF 신호들을 수신할 수도 있다.

[0420] 도 38 은 전자 디바이스/무선 디바이스 (3868) 내에 포함될 수도 있는 어떤 컴포넌트들을 예시한다. 전자 디바이스/무선 디바이스 (3868) 는 액세스 단말, 이동국, 사용자 장비 (user equipment; UE), 기지국, 액세스 포인트, 브로드캐스트 송신기, 노드 B, 진화형 노드 B, 서버, 컴퓨터, 라우터, 스위치 등일 수도 있다. 전자 디바이스/무선 디바이스 (3868) 는 본원에서 설명된 전자 디바이스들 (556, 1507) 중의 하나 이상에 따라 구현될 수도 있다. 전자 디바이스/무선 디바이스 (3868) 는 프로세서 (3884) 를 포함한다. 프로세서 (3884) 는 범용 단일-칩 또는 멀티-칩 마이크로프로세서 (예컨대, ARM), 특수 목적 마이크로프로세서 (예컨대, 디지털 신호 프로세서 (digital signal processor; DSP)), 마이크로제어기, 필드 프로그래밍 가능한 게이트 어레이 (field programmable gate array; FPGA) 등일 수도 있다. 프로세서 (3884) 는 중앙 프로세싱 유닛 (central processing unit; CPU) 이라고 지칭될 수도 있다. 단지 단일 프로세서 (3884) 가 전자 디바이스/무선 디바이스 (3868) 에서 도시되어 있지만, 대안적인 구성에서는, 프로세서들 (예컨대, ARM, DSP 및 FPGA) 의 조합이 이용될 수 있다.

[0421] 전자 디바이스/무선 디바이스 (3868) 는 또한, 메모리 (3870) 를 포함한다. 메모리 (3870) 는 전자 정보를 저장할 수 있는 임의의 전자 컴포넌트일 수도 있다. 메모리 (3870) 는 랜덤 액세스 메모리 (RAM), 판독 전용 메모리 (ROM), 자기디스크 저장 매체들, 광학 저장 매체들, RAM 에서의 플래시 메모리 디바이스들, 프로세서 (3884) 와 함께 포함된 온-보드 메모리, EPROM 메모리, EEPROM 메모리, 레지스터들, 및 그 조합들을 포함하는 기타 등등으로서 구체화될 수도 있다.

[0422] 데이터 (3874a) 및 명령들 (3872a) 은 메모리 (3870) 내에 저장될 수도 있다. 명령들 (3872a) 은 본원에서 설명된 방법들 중의 하나 이상을 구현하기 위하여 프로세서 (3884) 에 의해 실행가능할 수도 있다. 명령들 (3872a) 을 실행하는 것은 메모리 (3870) 에서 저장되는 데이터 (3874a) 의 이용을 수반할 수도 있다. 프로세서 (3884) 가 명령들 (3872a) 을 실행할 때, 명령들 (3872b) 의 다양한 부분들은 프로세서 (3884) 상으로 로딩될 수도 있고, 데이터 (3874b) 의 다양한 피스 (piece) 들은 프로세서 (3884) 상으로 로딩될 수도 있다.

[0423] 전자 디바이스/무선 디바이스 (3868) 는 또한, 전자 디바이스/무선 디바이스 (3868) 로의 신호의 송신 및 전자 디바이스/무선 디바이스 (3868)로부터의 신호의 수신을 허용하기 위하여 송신기 (3876) 및 수신기 (3878) 를 포함할 수도 있다. 송신기 (3876) 및 수신기 (3878) 는 트랜시버 (3888) 로서 집합적으로 지칭될 수도 있다. 하나 이상의 안테나들 (3868a 내지 3868n) 은 트랜시버 (3888) 에 전기적으로 결합될 수도 있다. 전자 디바이스/무선 디바이스 (3868) 는 또한, (도시되지 않은) 다수의 송신기들, 다수의 수신기들, 다수의 트랜시버들, 및/또는 다수의 안테나들을 포함할 수도 있다.

[0424] 전자 디바이스/무선 디바이스 (3868) 는 디지털 신호 프로세서 (DSP) (3880) 를 포함할 수도 있다. 전자 디바이스/무선 디바이스 (3868) 는 또한, 통신 인터페이스 (3882) 를 포함할 수도 있다. 통신 인터페이스 (3882) 는 사용자가 전자 디바이스/무선 디바이스 (3868) 와 상호작용하도록 할 수도 있다.

[0425] 전자 디바이스/무선 디바이스 (3868) 의 다양한 컴포넌트들은 하나 이상의 버스들에 의해 함께 결합될 수도 있으며, 버스들은 전력 버스, 제어 신호 버스, 상태 신호 버스, 데이터 (3874) 버스 등을 포함할 수도 있다. 간결성을 위하여, 다양한 버스들은 도 38 에서 버스 시스템 (3890) 으로서 예시되어 있다.

[0426] 상기 설명에서, 참조 번호들은 때때로 다양한 용어들과 관련하여 이용되었다. 용어가 참조 번호와 관련하여 이용될 경우, 이것은 도면들 중의 하나 이상에서 도시되는 특정한 엘리먼트를 지칭하도록 의도될 수도 있다. 용어가 참조 번호 없이 이용될 경우, 이것은 임의의 특정한 도면에 대한 제한없이 일반적으로 용어를 지칭하도록 의도될 수도 있다.

[0427] 용어 "결정하는 것" 은 광범위한 액션들을 포함하고, 그러므로, "결정하는 것" 은 계산하는 것, 컴퓨팅하는 것, 프로세싱하는 것, 유도하는 것, 조사하는 것, 툭업하는 것 (예컨대, 테이블, 데이터베이스 또는 또 다른 데이터 구조에서 툭업하는 것), 확인하는 것 및 기타 등등을 포함할 수 있다. 또한, "결정하는 것" 은 수신하는 것 (예컨대, 정보를 수신하는 것), 액세스하는 것 (예컨대, 메모리 내의 데이터를 액세스하는 것) 및 기타 등등을 포함할 수 있다. 또한, "결정하는 것" 은 분석하는 것, 선택하는 것, 선정하는 것, 설정하는 것 등을 포함할 수 있다.

[0428] 어구 "~ 에 기초하여 (based on)" 는 명백히 이와 달리 규정되지 않는 한, "~ 에 오직 기초하여" 를 의미하지 않는다. 다시 말해서, 어구 "~ 에 기초하여" 는 "~ 에 오직 기초하여" 및 "~ 에 적어도 기초하여" 의 양자

를 설명한다.

[0429] 본원에서 설명된 구성들 중의 임의의 하나와 관련하여 설명된, 특징들, 기능들, 절차들, 구성요소들, 엘리먼트들, 구조들 등 중의 하나 이상은, 호환가능한 경우, 본원에서 설명된 다른 구성들 중의 임의의 것과 관련하여 설명된, 기능들, 절차들, 구성요소들, 엘리먼트들, 구조들 등 중의 하나 이상과 조합될 수도 있다는 것에 주목해야 한다. 다시 말해서, 본원에서 설명된, 기능들, 절차들, 구성요소들, 엘리먼트들 등의 임의의 호환가능한 조합은 본원에서 개시된 시스템들 및 방법들에 따라 구현될 수도 있다.

[0430] 본원에서 설명된 기능들은 프로세서-판독가능한 또는 컴퓨터-판독가능한 매체 상에 하나 이상의 명령들로서 저장될 수도 있다. 용어 "컴퓨터-판독가능한 매체"는 컴퓨터 또는 프로세서에 의해 액세스될 수 있는 임의의 이용가능한 매체를 지칭한다. 제한이 아닌 예로서, 이러한 매체는 RAM, ROM, EEPROM, 플래시 메모리, CD-ROM 또는 다른 광학 디스크 저장, 자기 디스크 저장 또는 다른 자기 저장 디바이스들, 또는 희망하는 프로그램 코드를 명령들 또는 데이터 구조들의 형태로 저장하기 위해 이용될 수 있으며 컴퓨터에 의해 액세스될 수 있는 임의의 다른 매체를 포함할 수도 있다. 디스크 (disk) 및 디스크 (disc) 는, 본원에서 이용될 때, 컴팩트 디스크 (compact disc; CD), 레이저 디스크 (laser disc), 광학 디스크 (optical disc), 디지털 다기능 디스크 (digital versatile disc; DVD), 플로피 디스크 (floppy disk) 및 Blu-ray® 디스크 (disc) 를 포함하며, 여기서, 디스크 (disk) 들은 통상적으로 데이터를 자기적으로 재생하는 반면, 디스크 (disc) 들은 레이저로 데이터를 광학적으로 재생한다. 컴퓨터-판독가능한 매체는 유형적이고 비일시적일 수도 있다는 것에 주목해야 한다. 용어 "컴퓨터-프로그램 제품"은 컴퓨팅 디바이스 또는 프로세서에 의해 실행되거나, 프로세싱되거나 또는 컴퓨팅될 수도 있는 코드 또는 명령들 (예컨대, "프로그램") 과 조합하여 컴퓨팅 디바이스 또는 프로세서를 지칭한다. 본원에서 이용될 때, 용어 "코드"는 컴퓨팅 디바이스 또는 프로세서에 의해 실행가능한 소프트웨어, 명령들, 코드 또는 데이터를 지칭할 수도 있다.

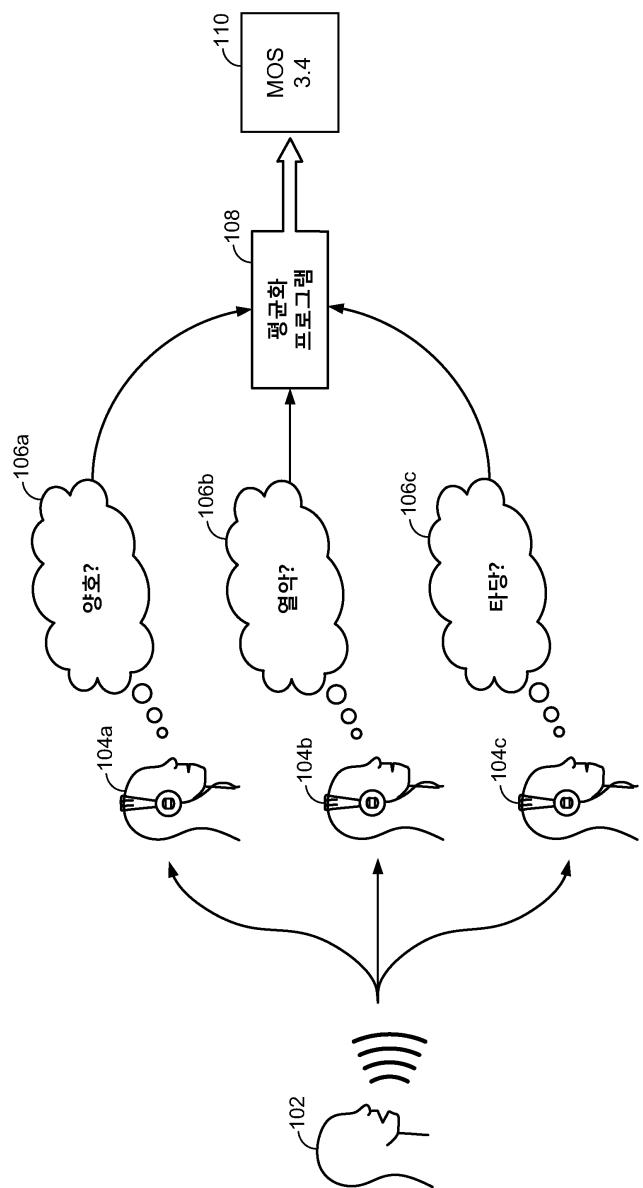
[0431] 소프트웨어 또는 명령들은 또한 송신 매체를 통해서 송신될 수도 있다. 예를 들어, 동축 케이블, 광섬유 케이블, 트위스트 페어 (twisted pair), 디지털 가입자 회선 (digital subscriber line; DSL), 또는 무선 기술들 예컨대, 적외선, 라디오, 및 마이크로파를 이용하여, 웹사이트, 서버, 또는 다른 원격 소스로부터 소프트웨어가 송신될 경우, 동축 케이블, 광섬유 케이블, 트위스트 페어, DSL, 또는 무선 기술들 예컨대, 적외선, 라디오, 및 마이크로파는 매체의 정의 내에 포함된다.

[0432] 본원에서 개시된 방법들은 설명된 방법을 달성하기 위한 하나 이상의 단계들 또는 액션들을 포함한다. 방법 단계들 및/또는 액션들은 청구의 범위로부터 이탈하지 않으면서 서로 상호 교환될 수도 있다. 다시 말해서, 설명되고 있는 방법의 적합한 동작을 위하여 단계들 또는 액션들의 특정한 순서가 요구되지 않는 한, 특정한 단계들 및/또는 액션들의 순서 및/또는 이용은 청구의 범위로부터 이탈하지 않으면서 수정될 수도 있다.

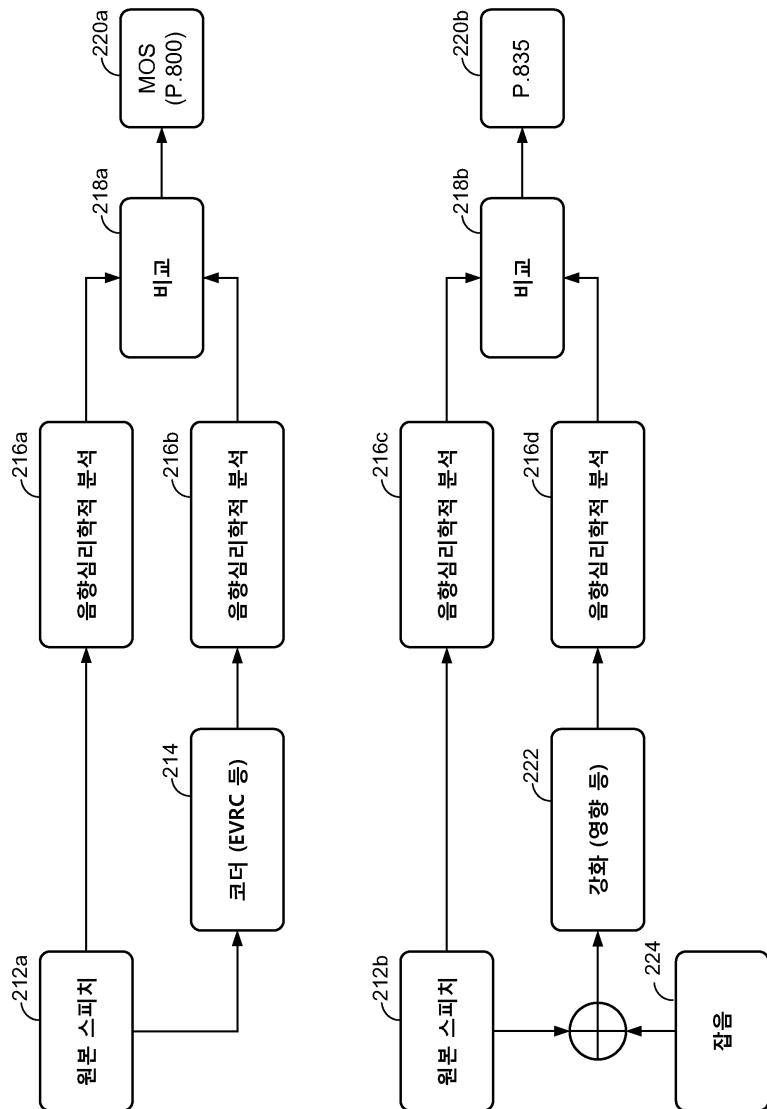
[0433] 청구항들은 위에서 예시된 정밀한 구성 및 컴포넌트들에 제한되지 않는 것으로 이해되어야 한다. 청구의 범위로부터 이탈하지 않고, 다양한 수정들, 변경들 및 변동들이 본원에서 설명된 시스템들, 방법들, 및 장치의 배치, 동작 및 세부사항들에서 행해질 수도 있다.

도면

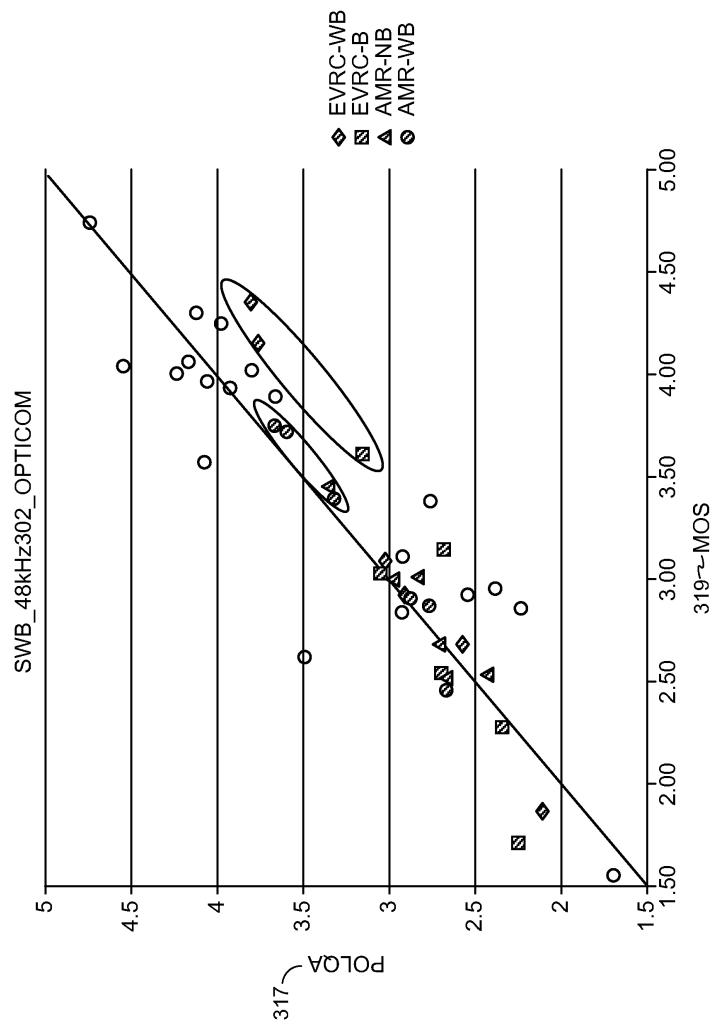
도면1



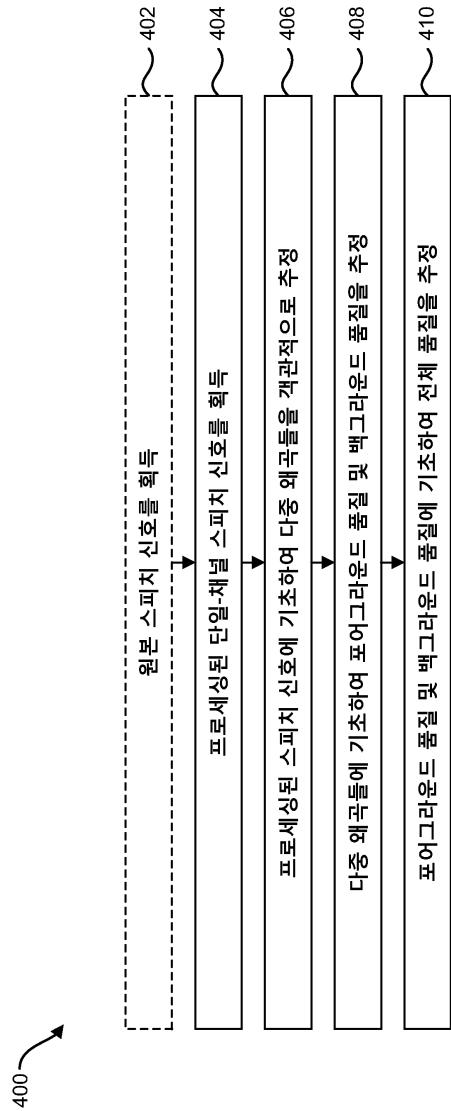
도면2



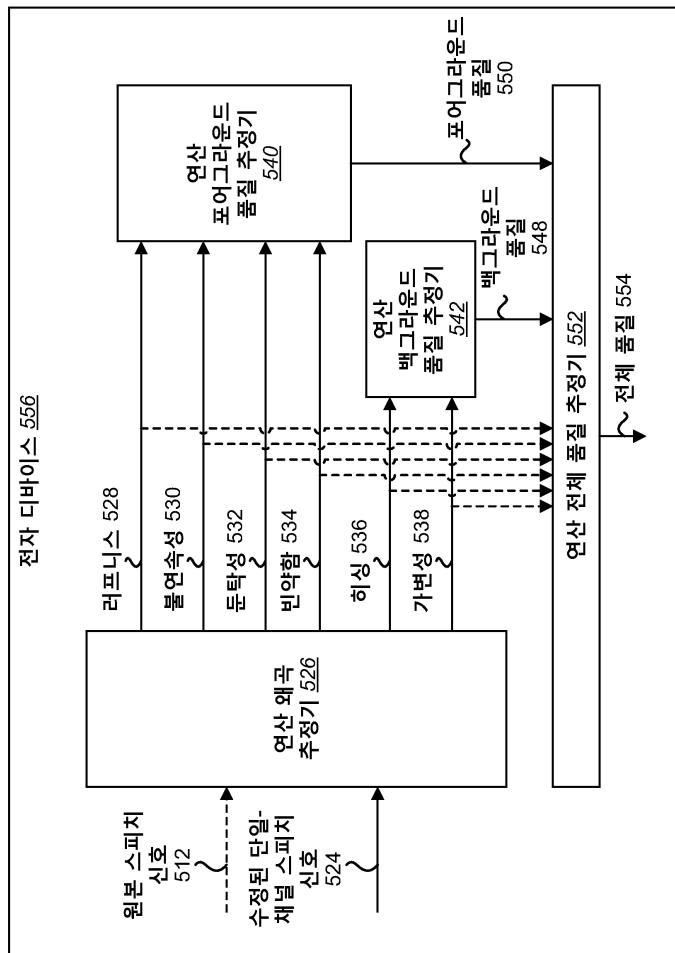
도면3



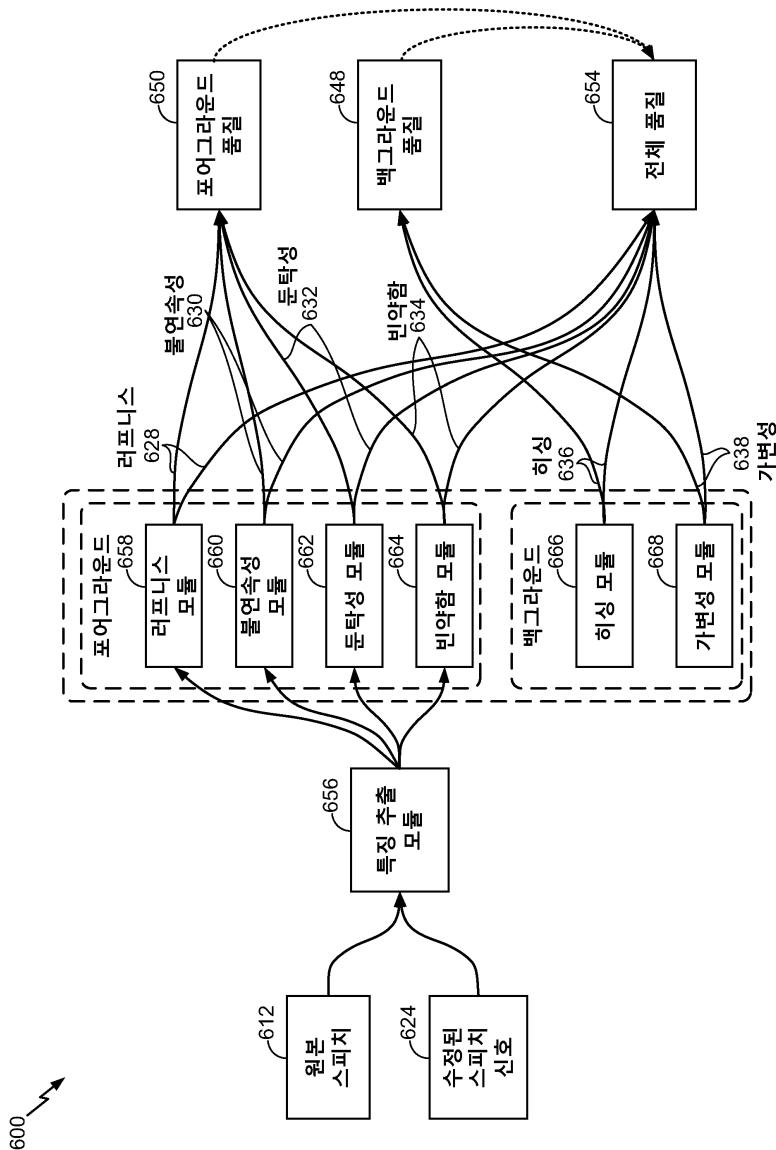
도면4



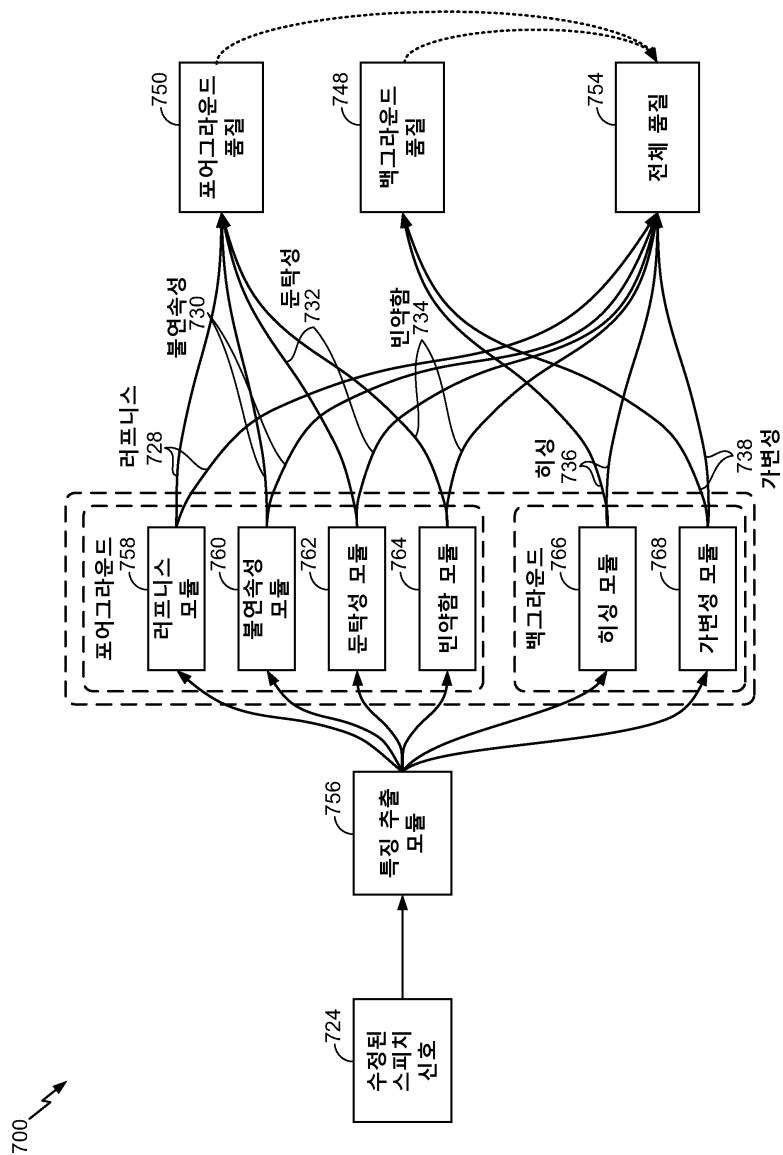
도면5



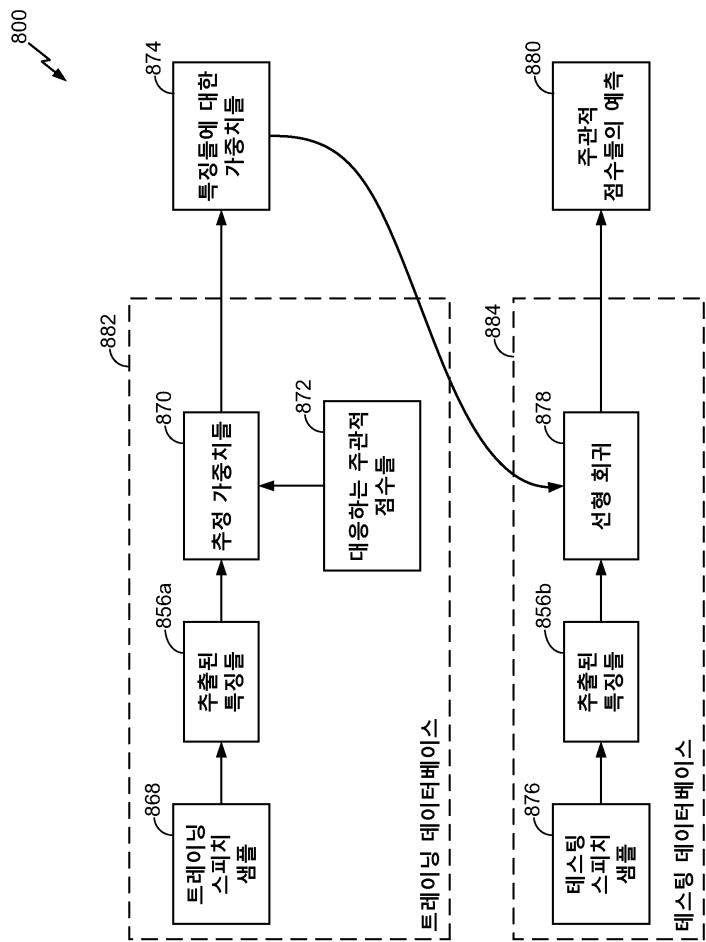
도면6



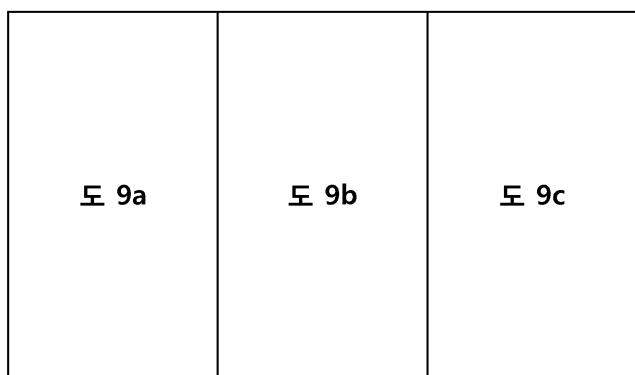
도면7



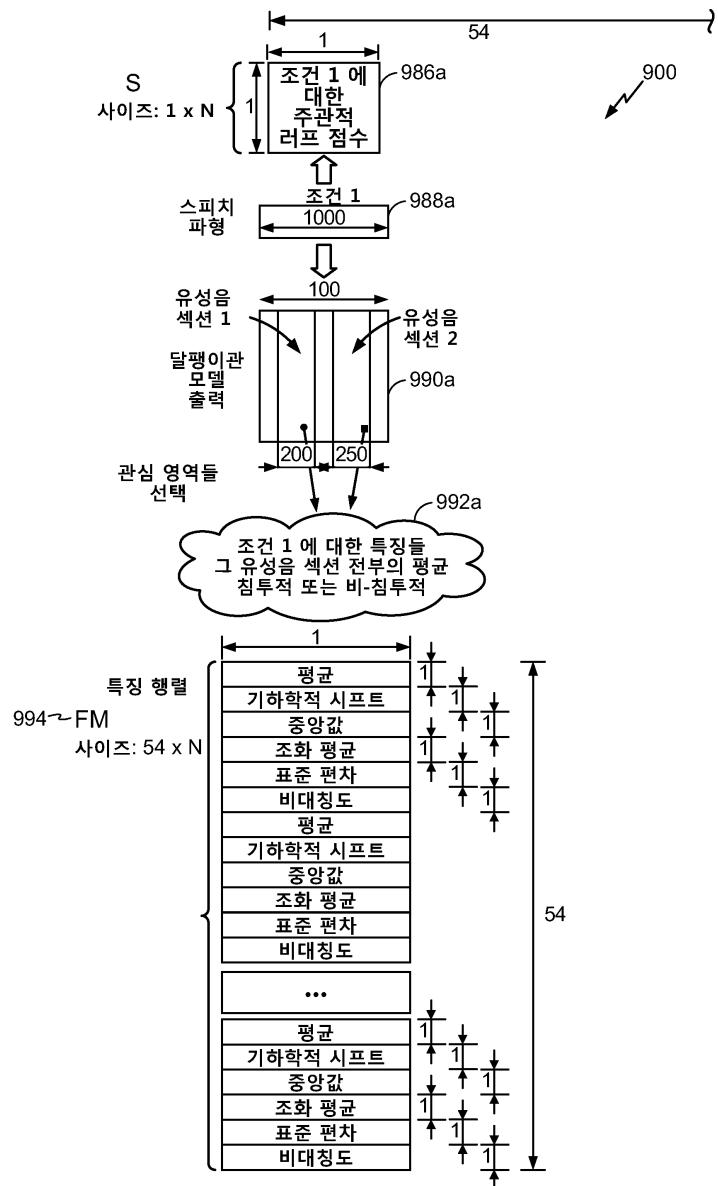
도면8



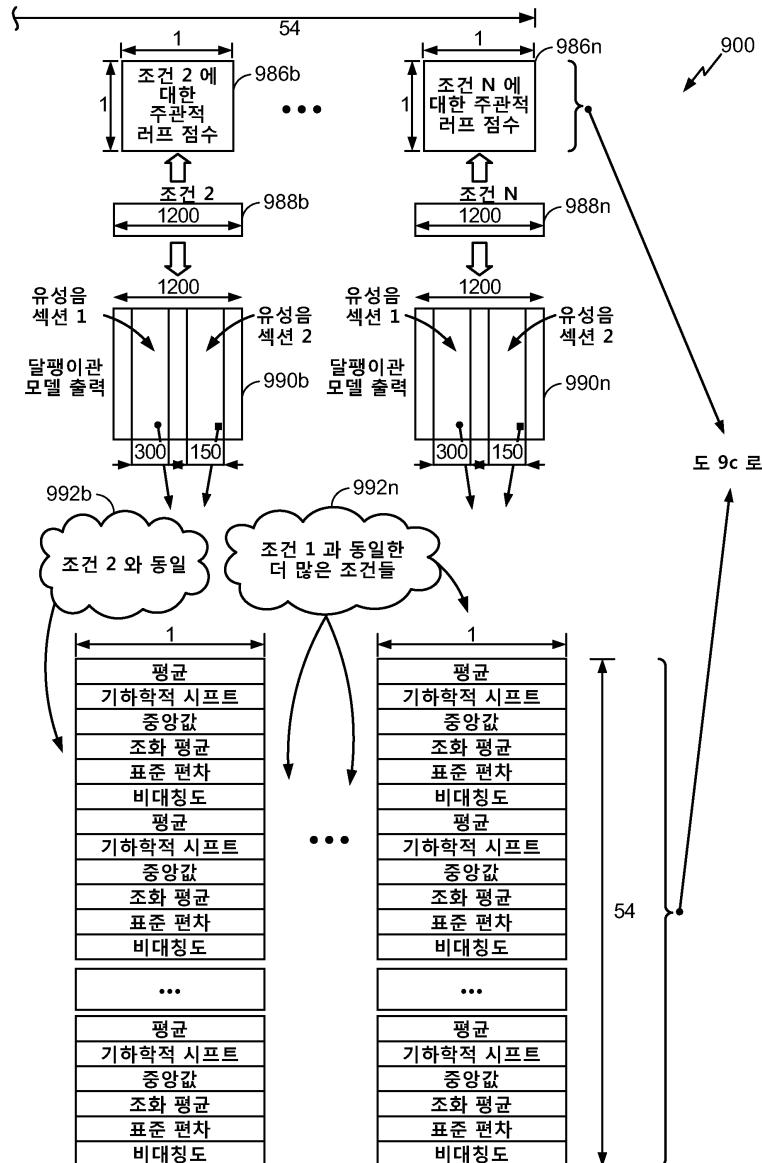
도면9



도면9a



도면9b



도면9c

900

러프 차원에 대한 트레이닝
가중화의 도면

침투적 및 비-침투적에 대해 동일
여기서는 비-침투적을 예로서 이용

여기서는 N 개의 조건들이
이용되는 것에 주목

도 9b
로부터

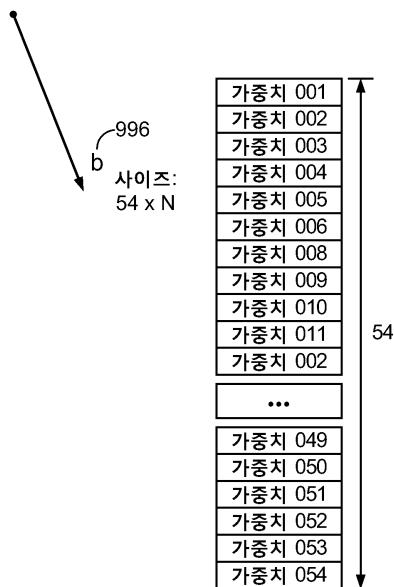
지금부터:

- 우리는 N 개의 조건들에 대한 주관적 "러프" 점수들을 가진다.
- 우리는 N 개의 조건들에 대해 추출된 특징들을 가진다 (각각의 조건에 대해 54 개의 특징들).

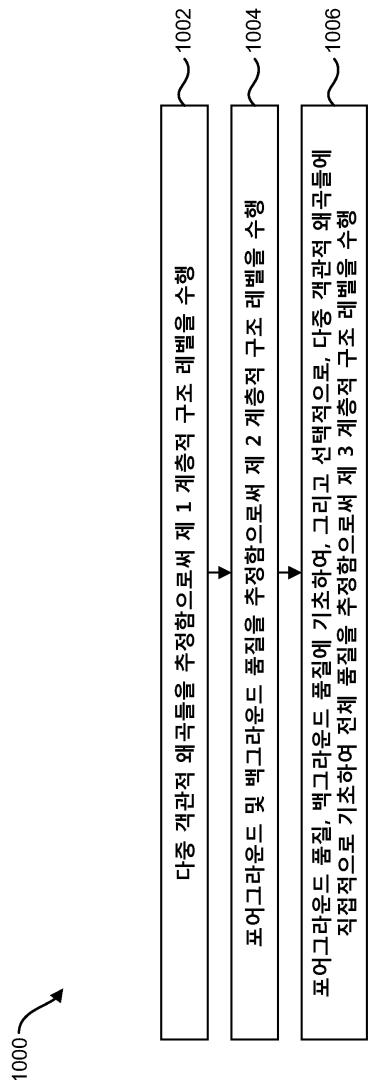
질문:
최소의 에러로 아래의 공식을 만드는 "b" 를 구함
 $FM^* b = S$

출력:

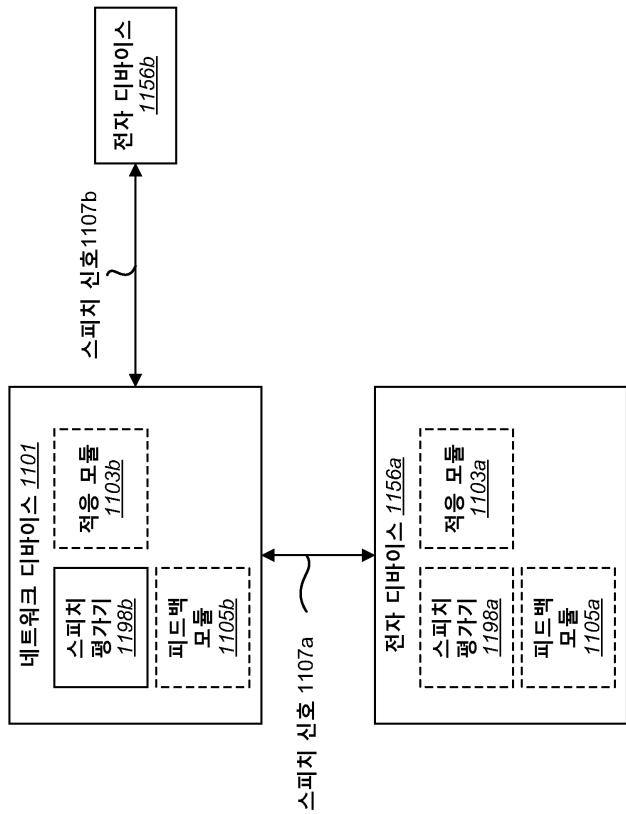
B 는 가중화. "러프" 의 추후의 테스팅에 대해 그것을 저장.



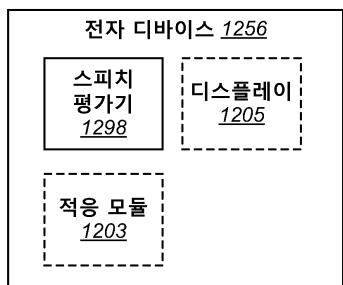
도면10



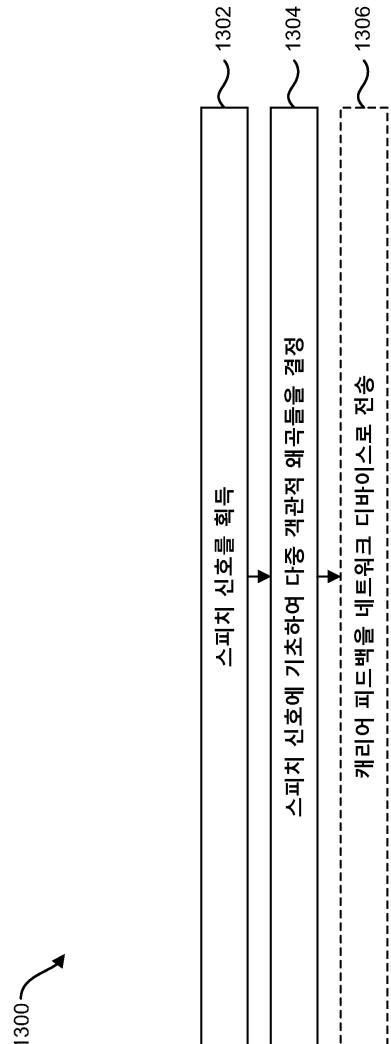
도면11



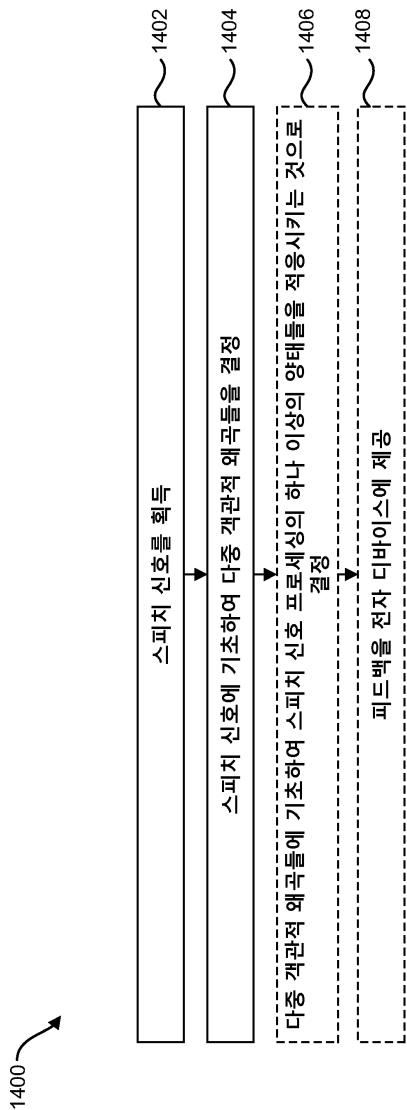
도면12



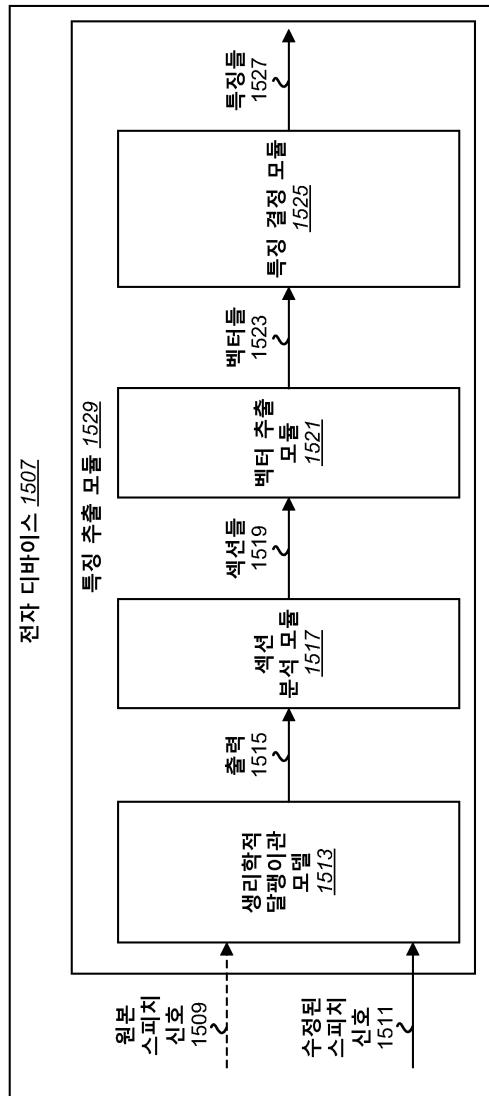
도면13



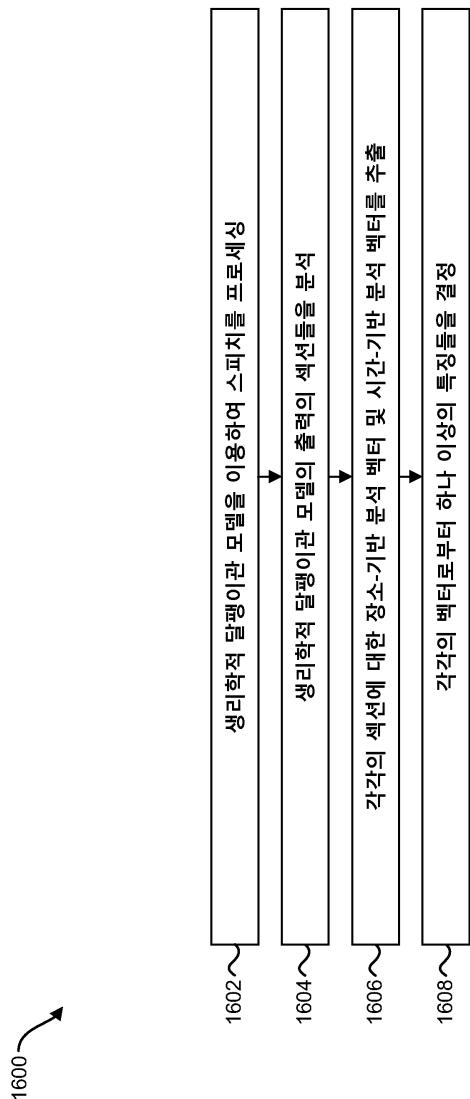
도면14



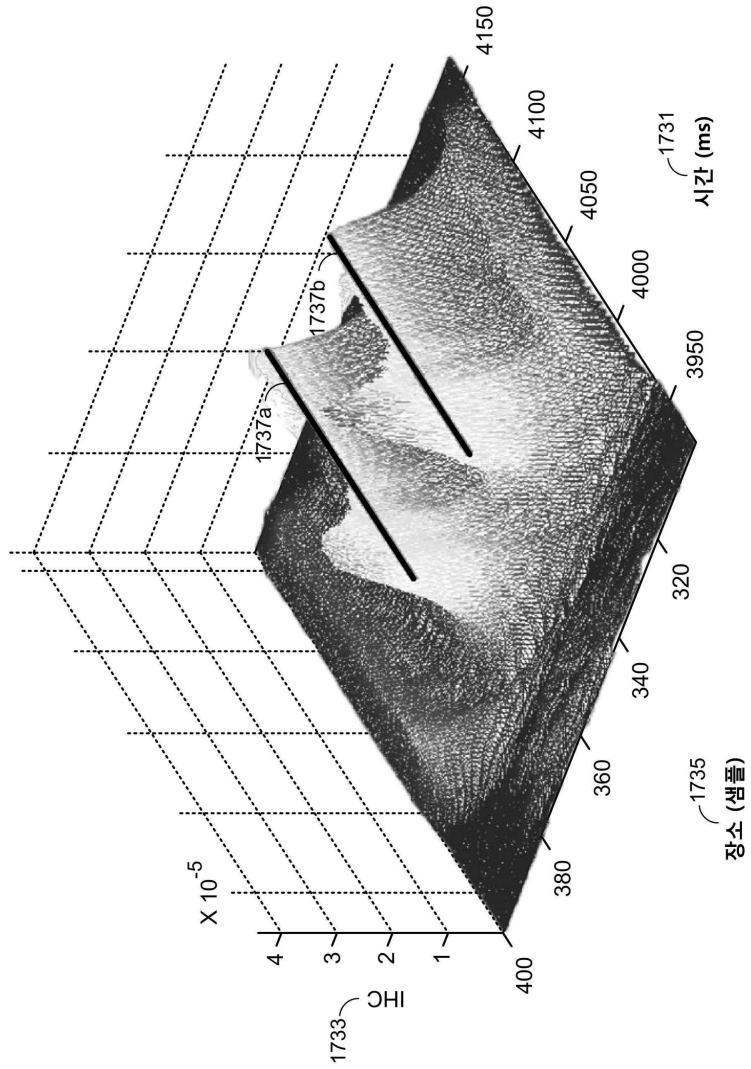
도면15



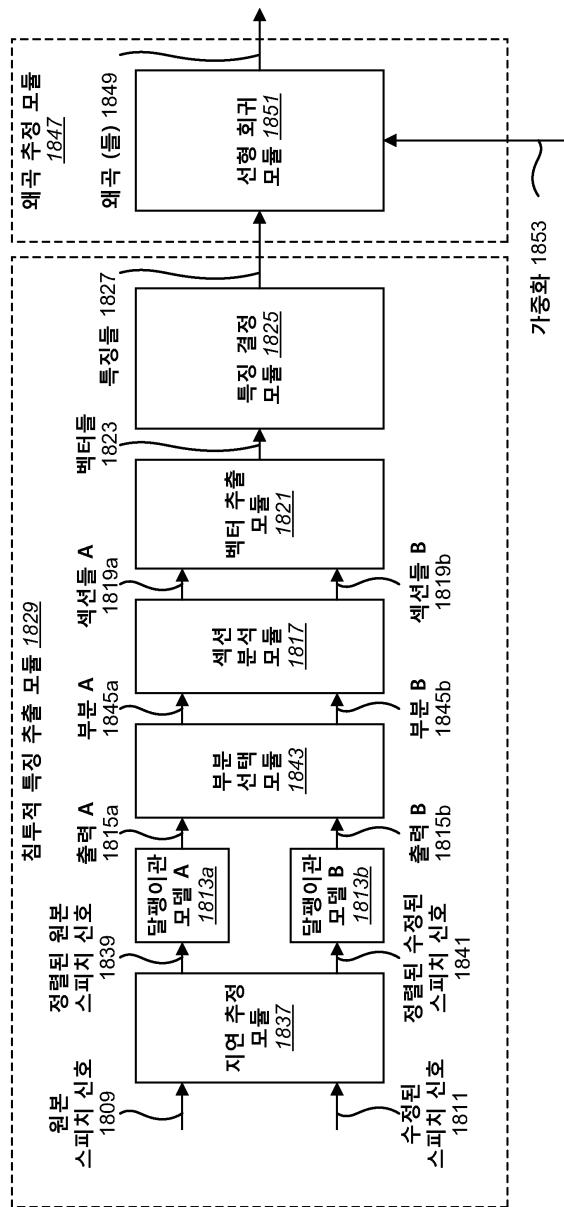
도면16



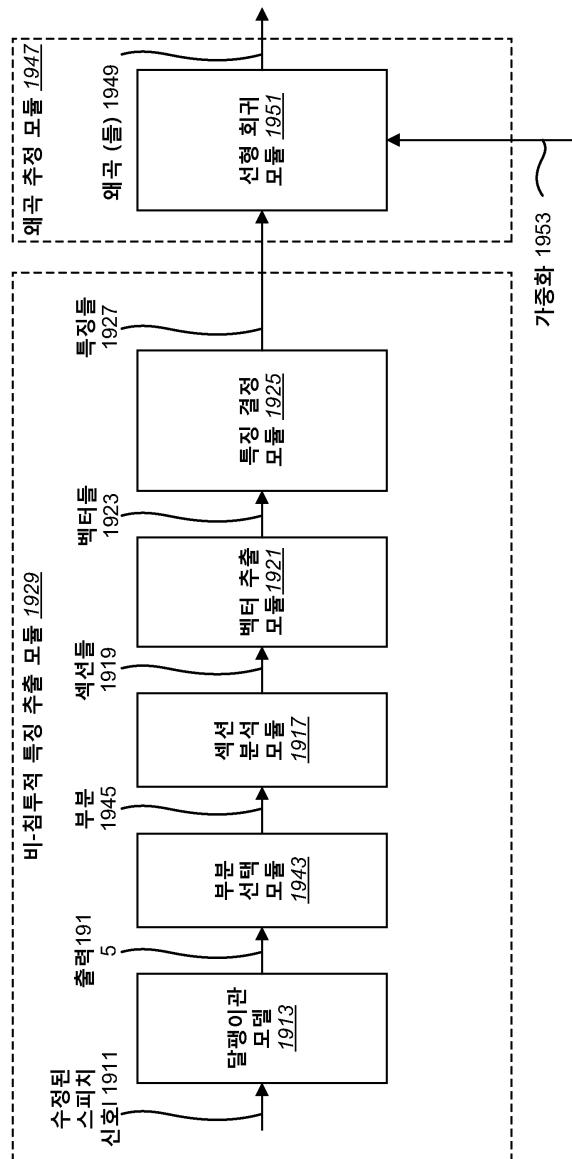
도면17



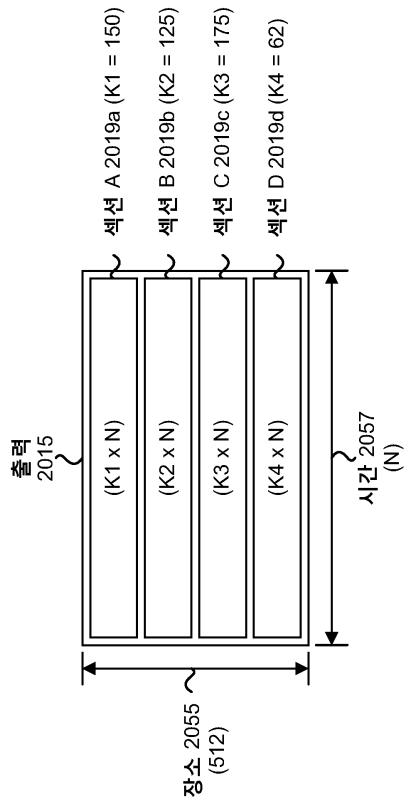
도면18



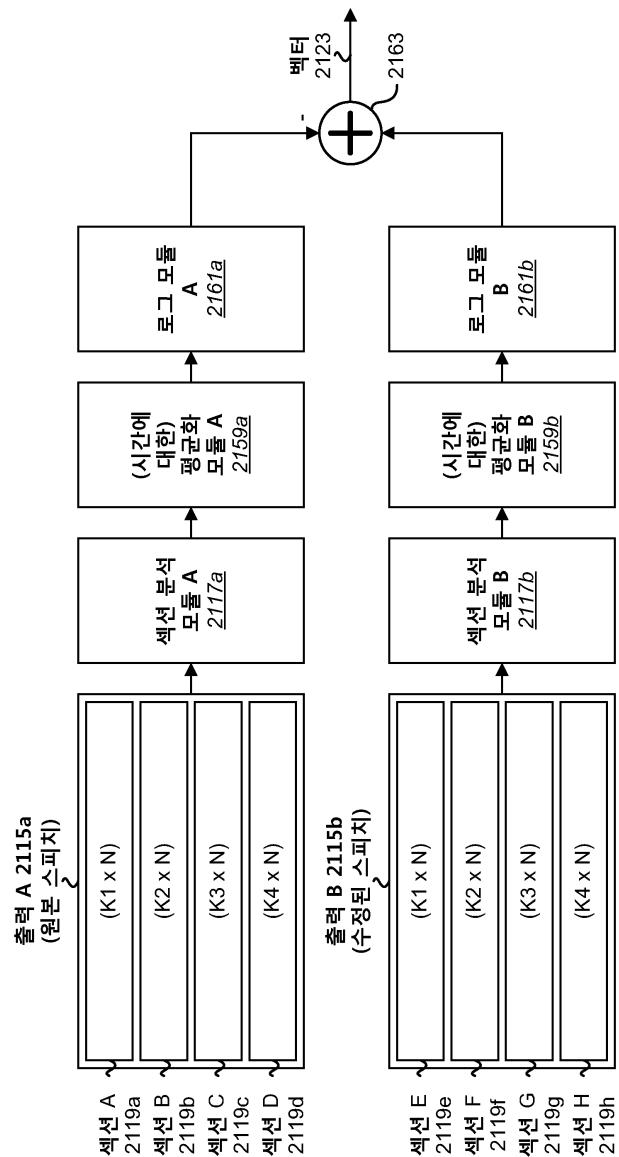
도면19



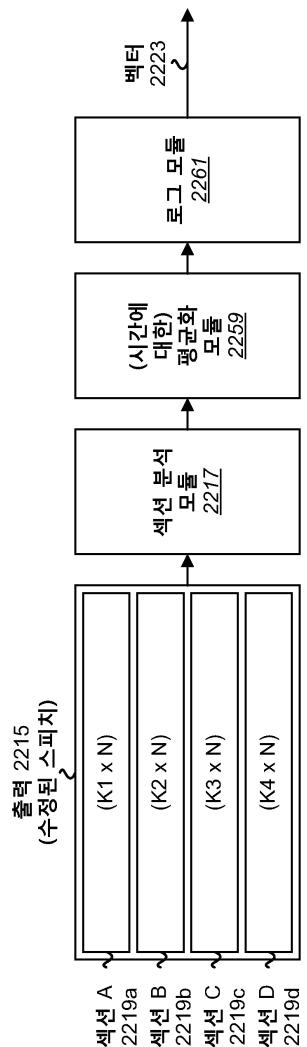
도면20



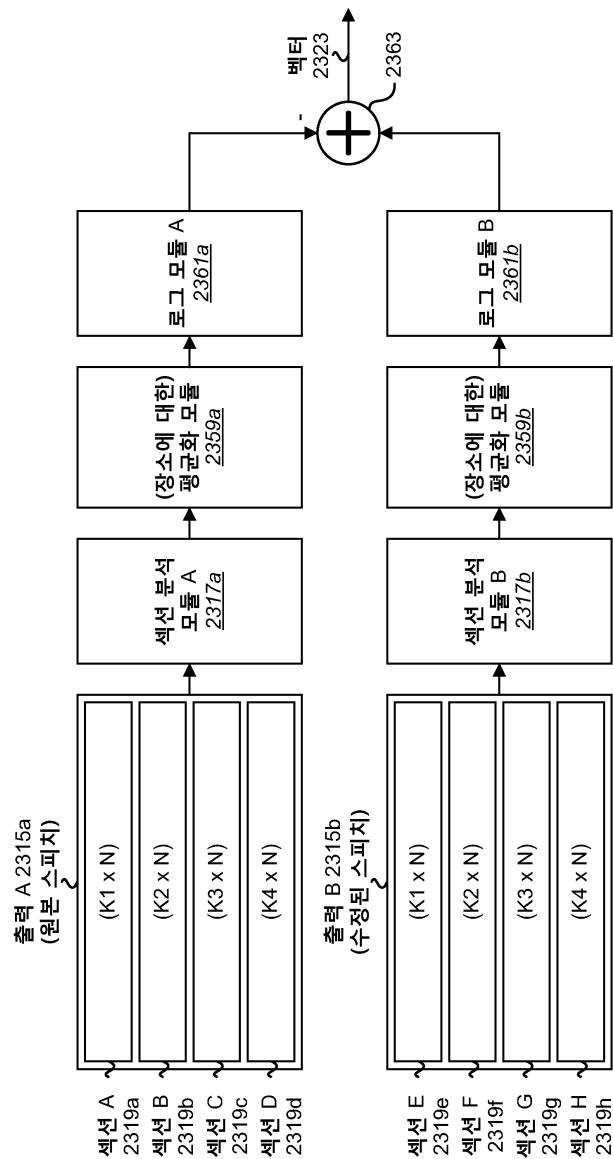
도면21



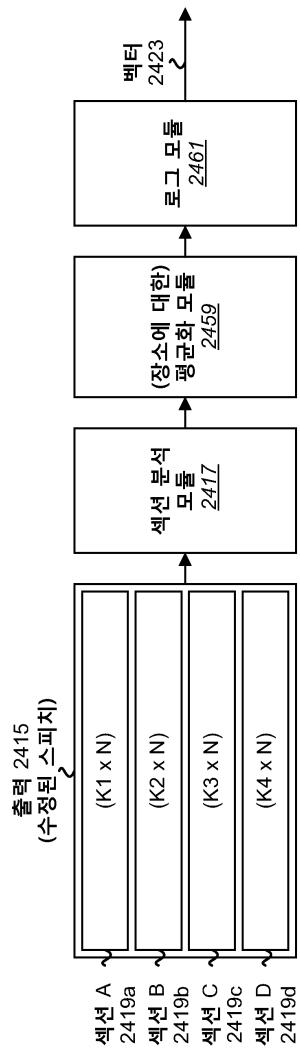
도면22



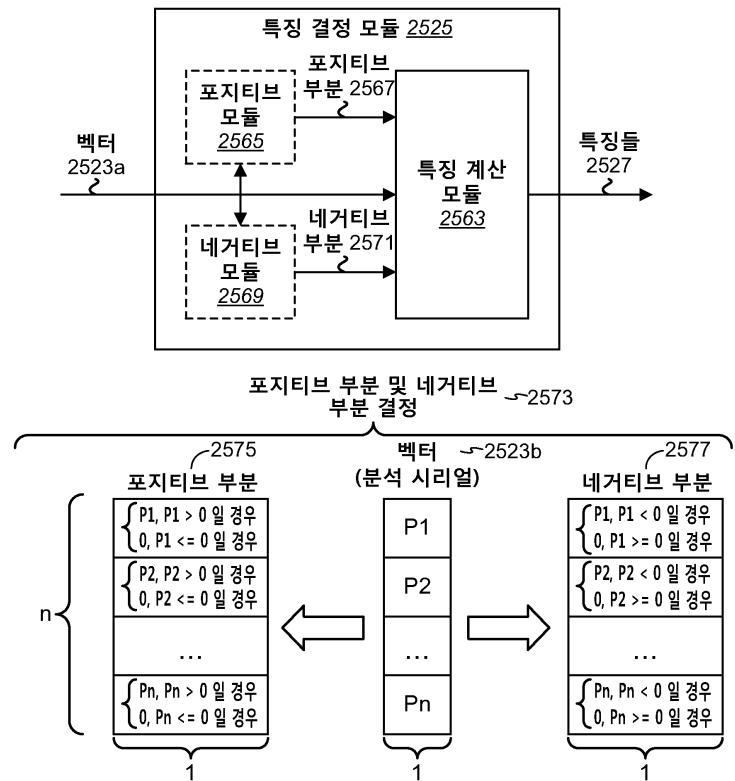
도면23



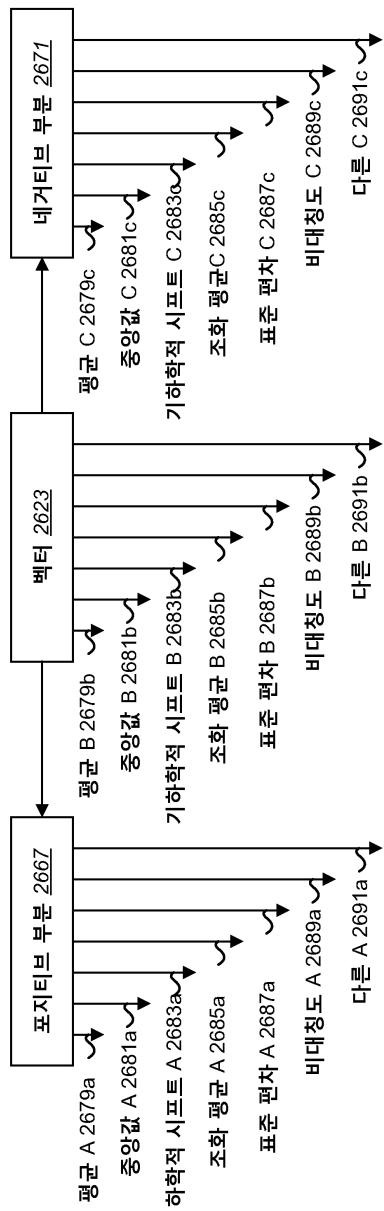
도면24



도면25



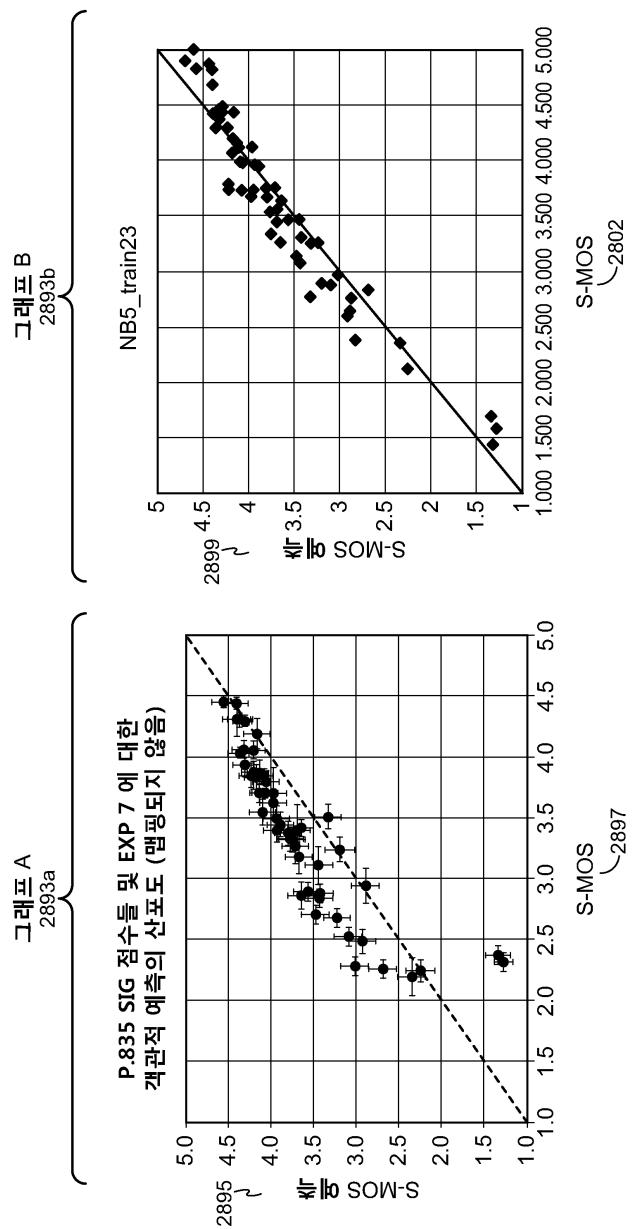
도면26



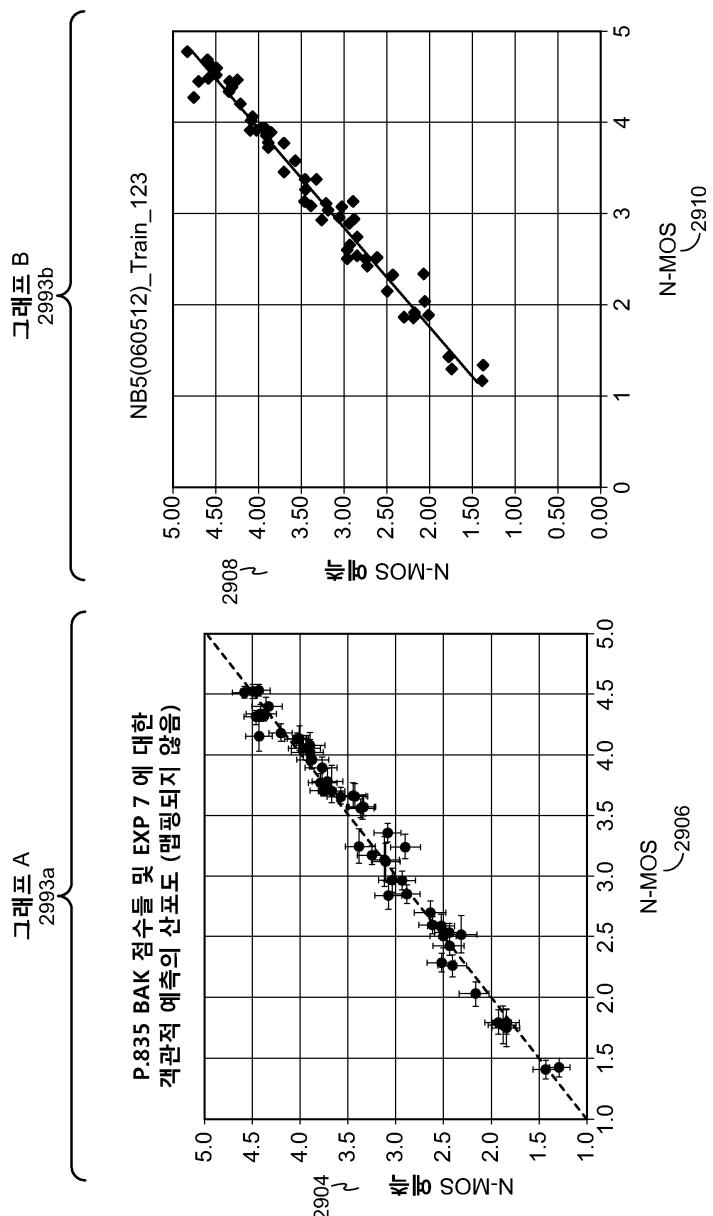
도면27



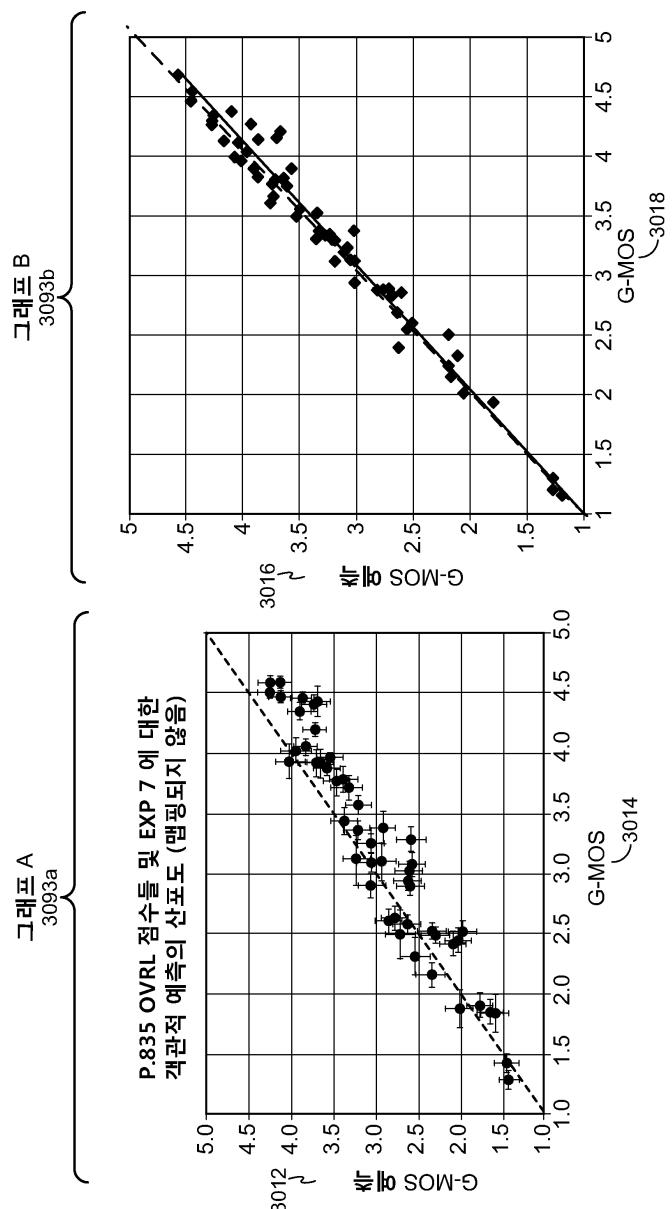
도면28



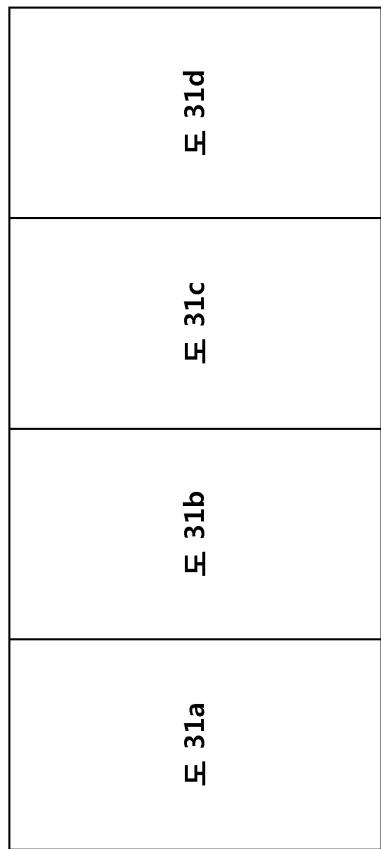
도면29



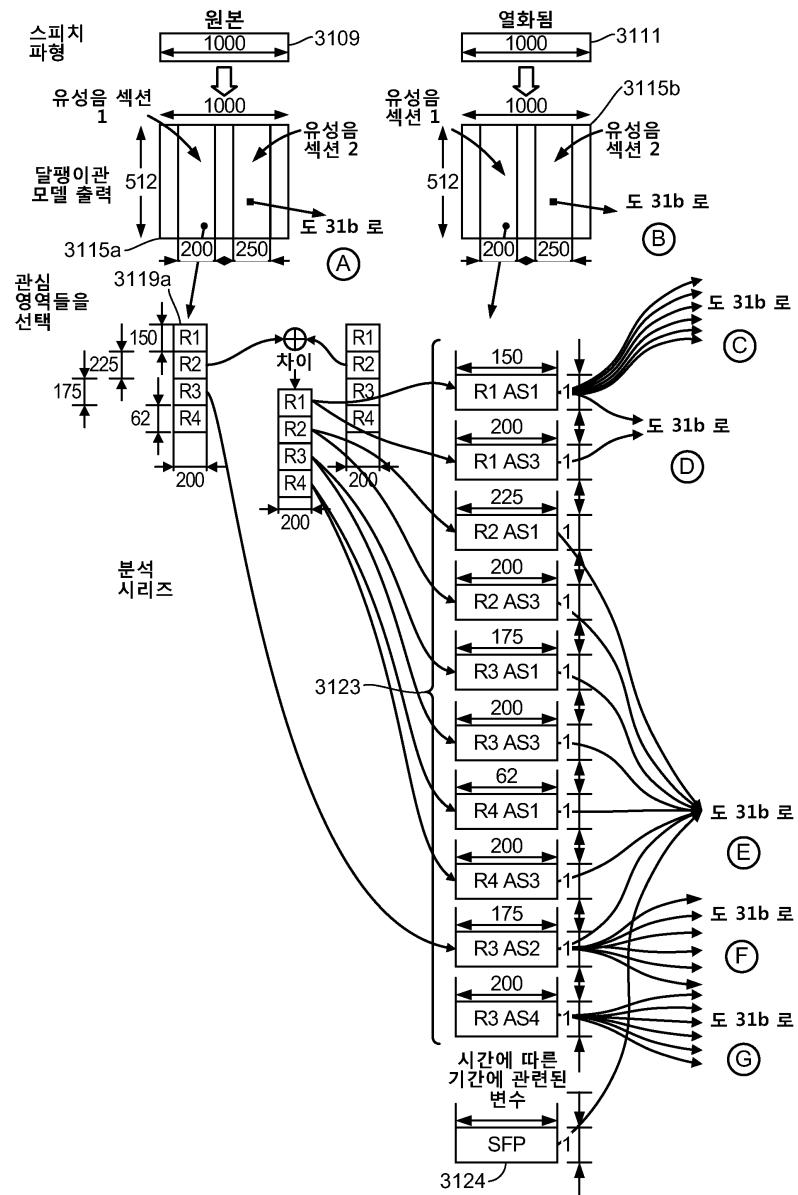
도면30



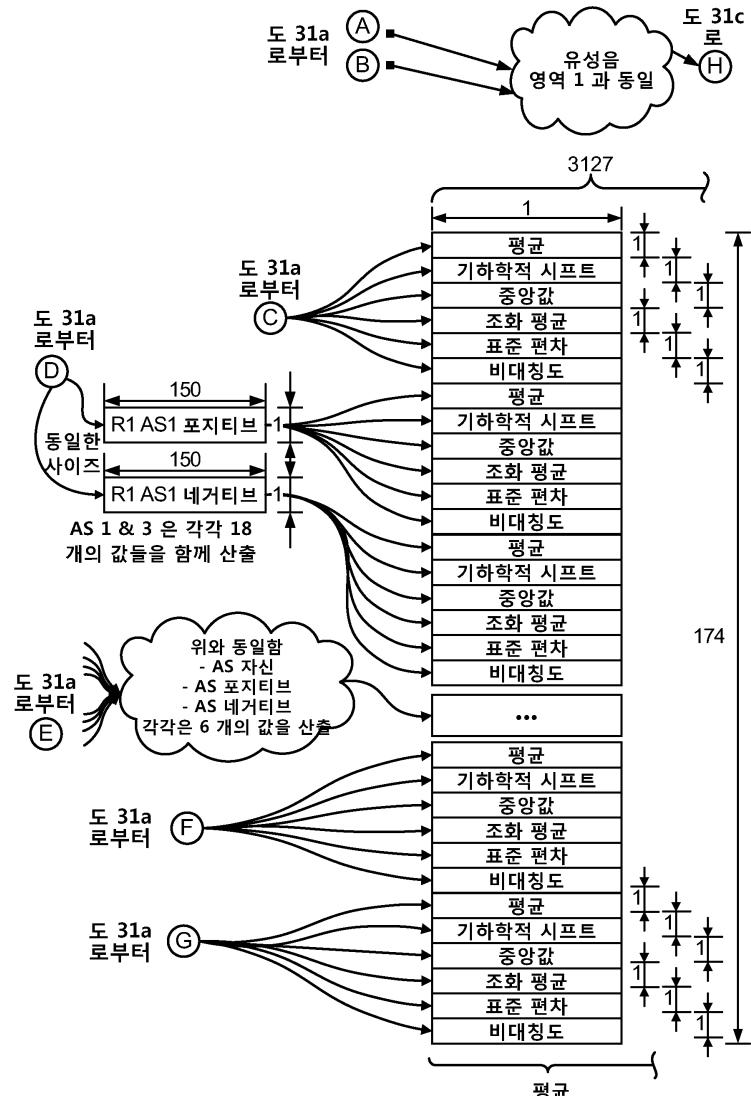
도면31



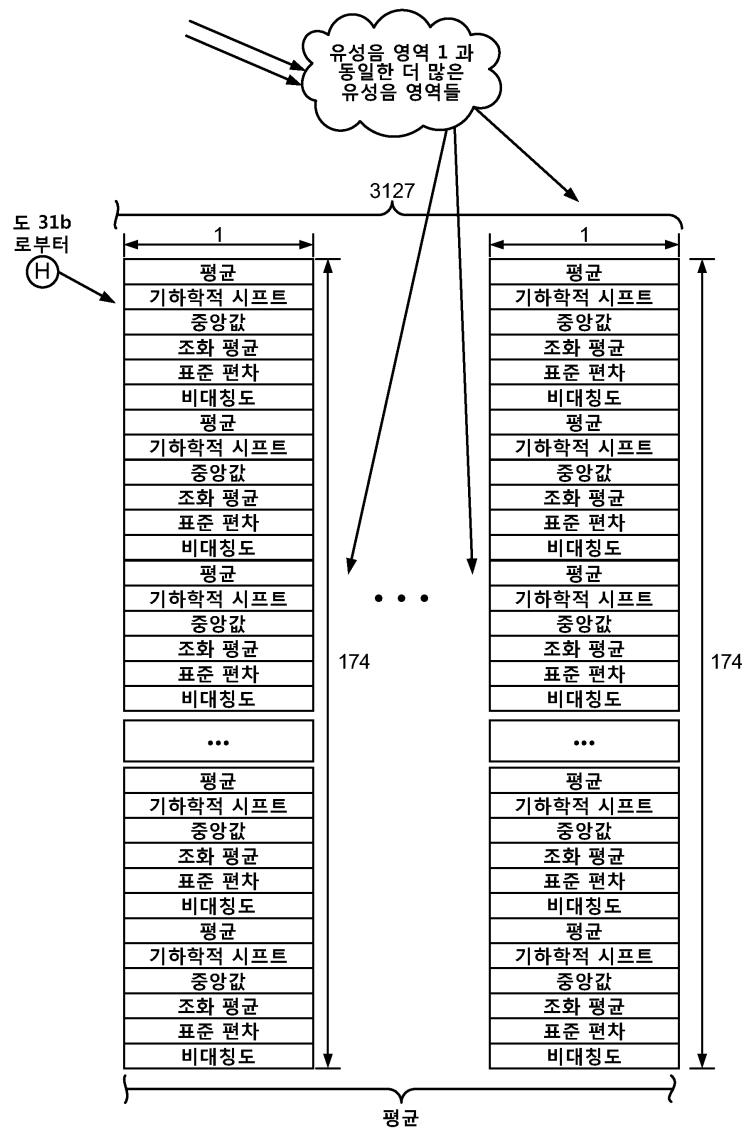
도면31a



도면31b



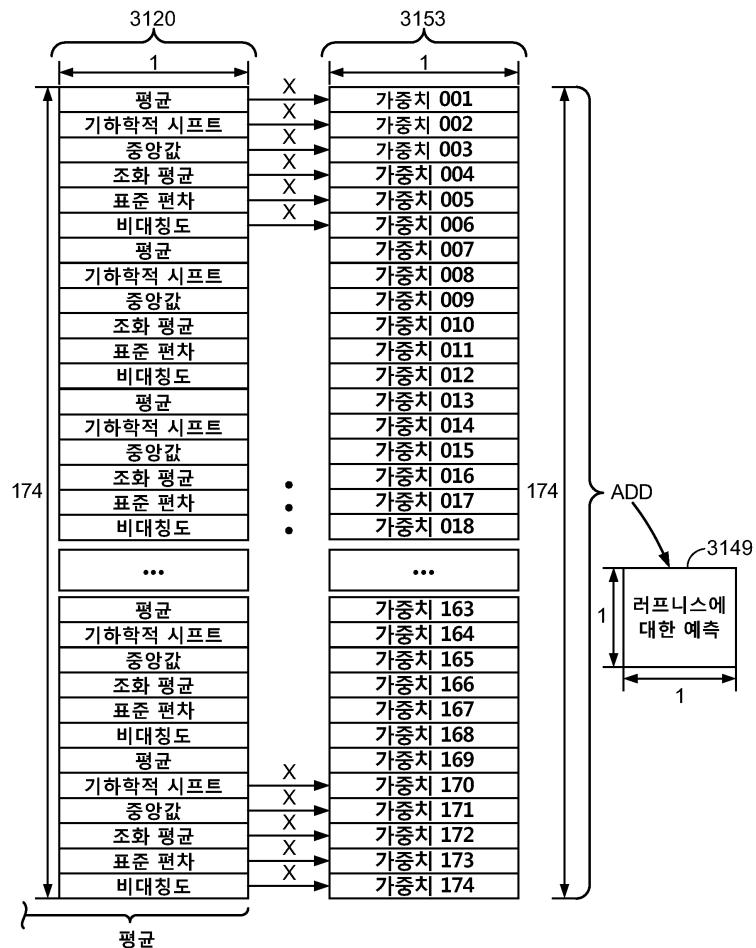
도면31c



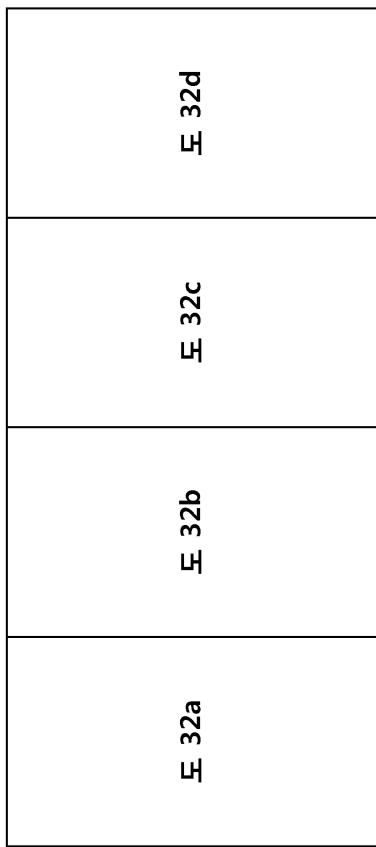
도면31d

러프 차원에 대한 침투적 측정의 도면

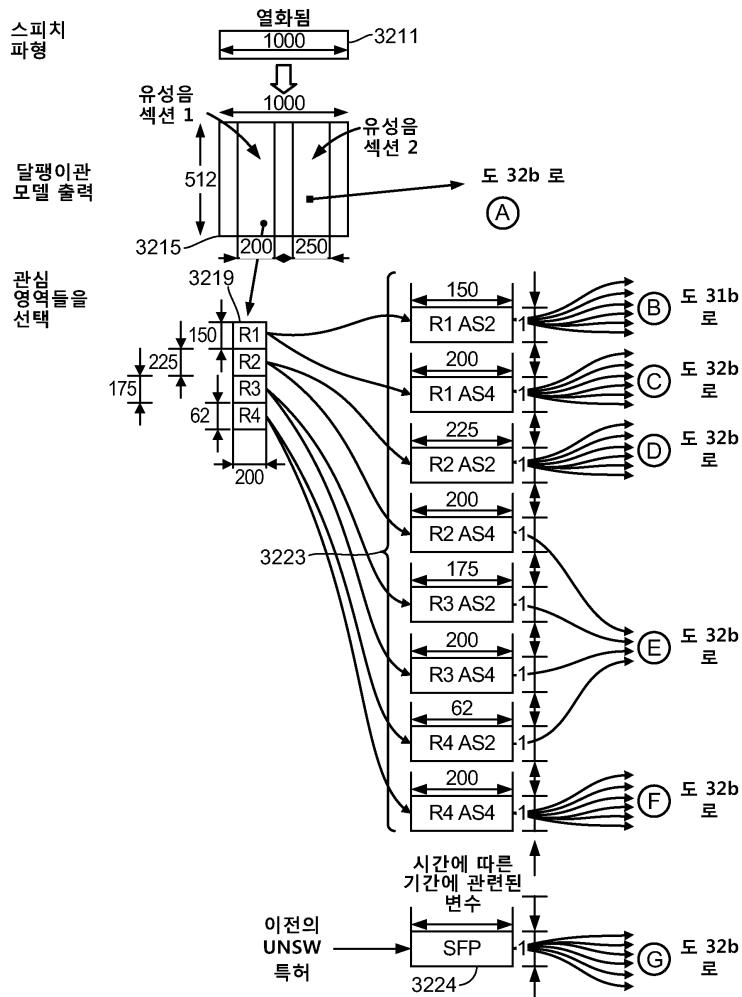
러프 차원에 대한 가중화
가중화의 값은 트레이닝으로부터 얻어짐.
이들은 테스트 동안에 일정한 것으로 고려될 수
있음.



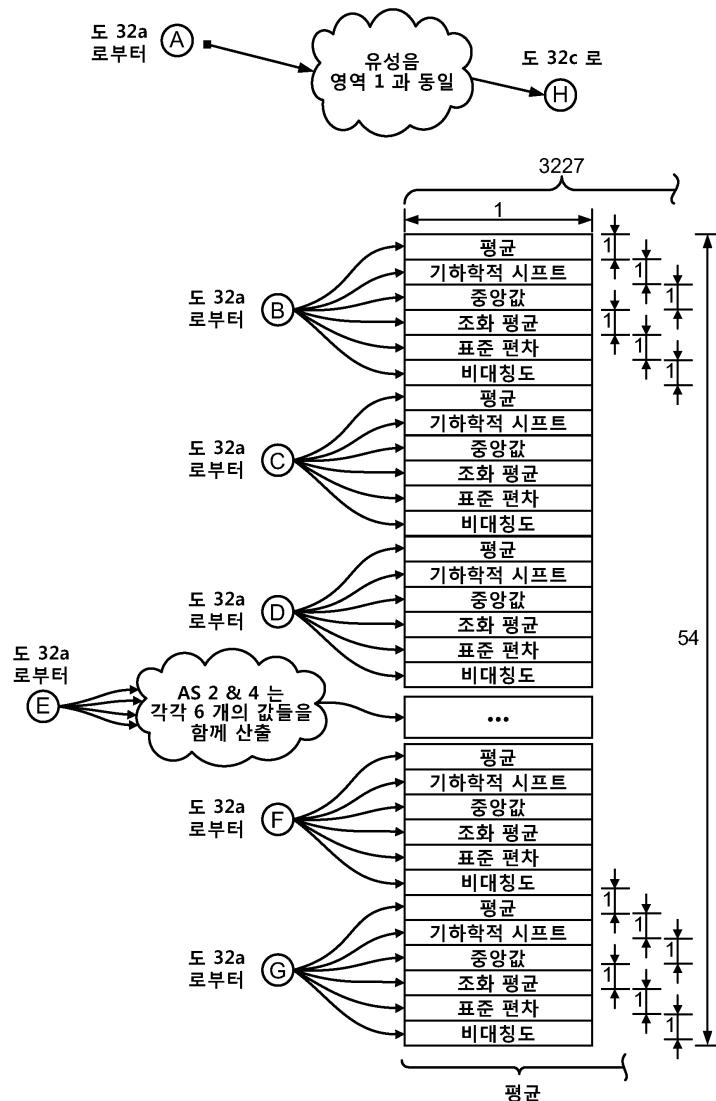
도면32



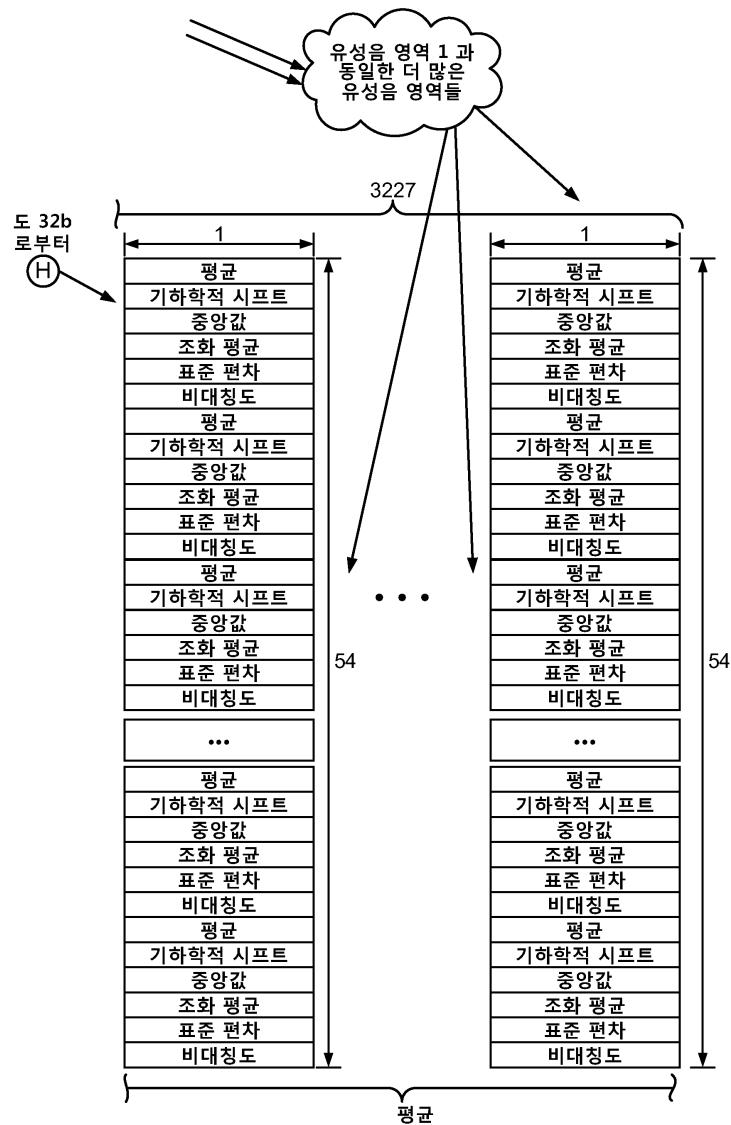
도면32a



도면32b



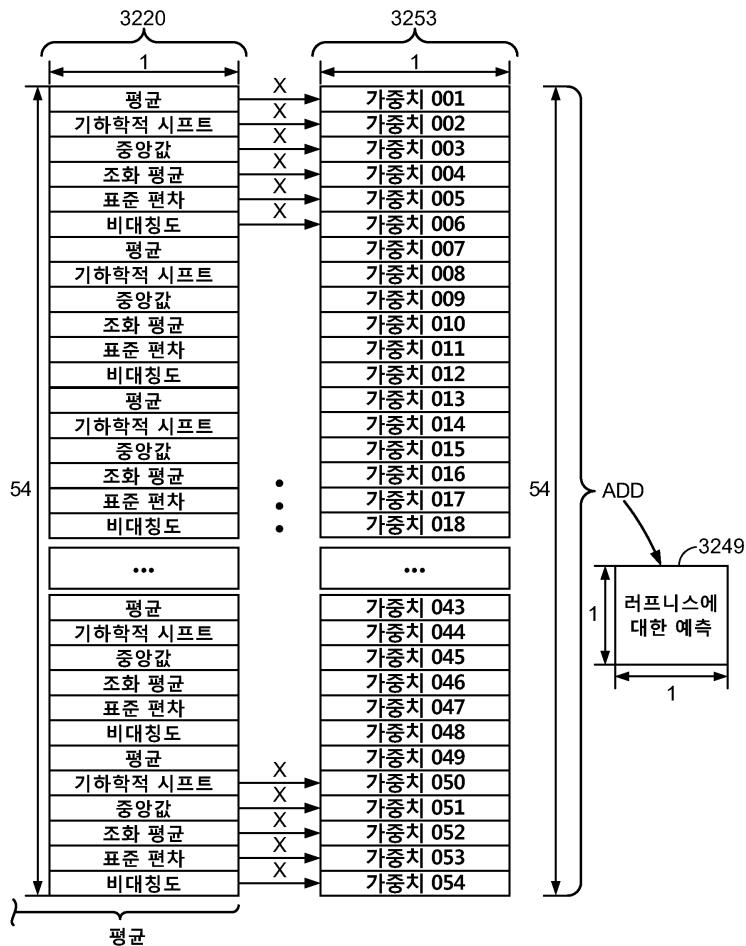
도면32c



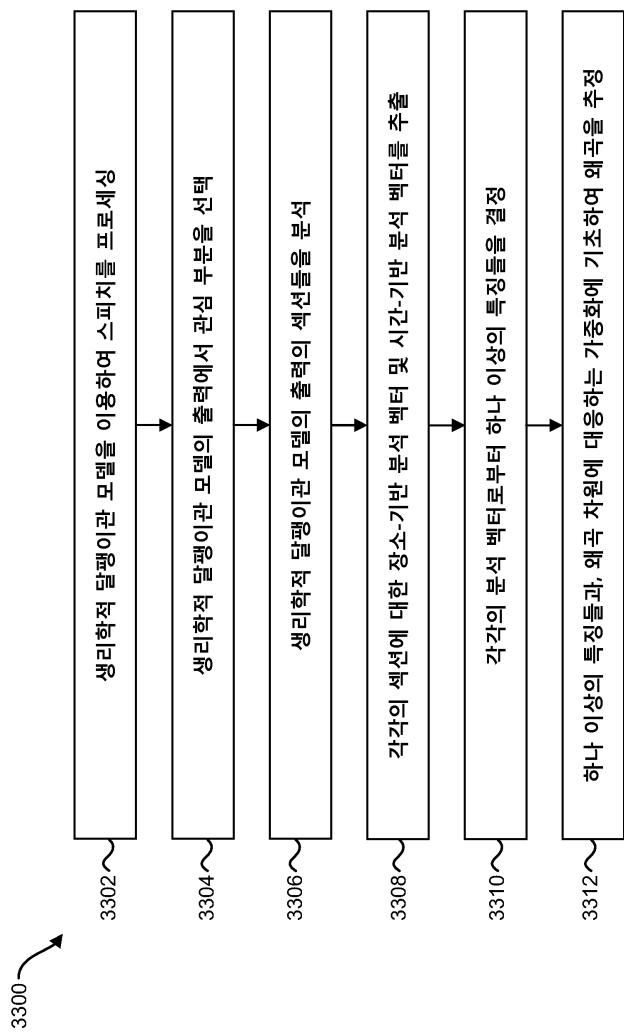
도면32d

러프 차원에 대한 침투적 측정의 도면

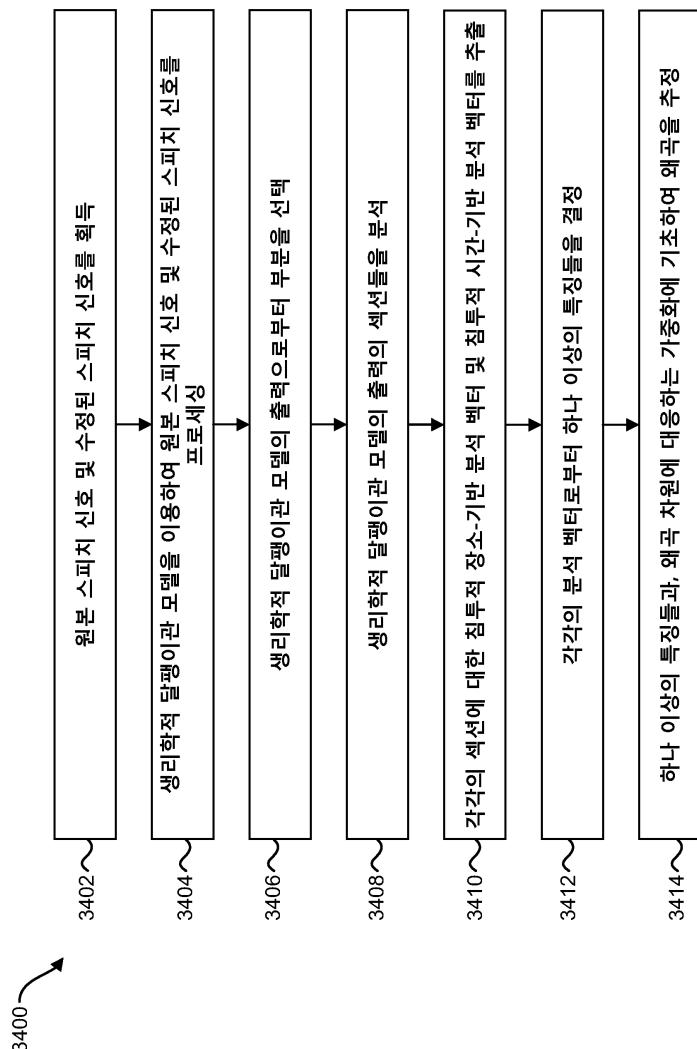
러프 차원에 대한 가중화
가중화의 값은 트레이닝으로부터 얻어짐.
이들은 테스트 동안에 일정한 것으로 고려될
수 있음.



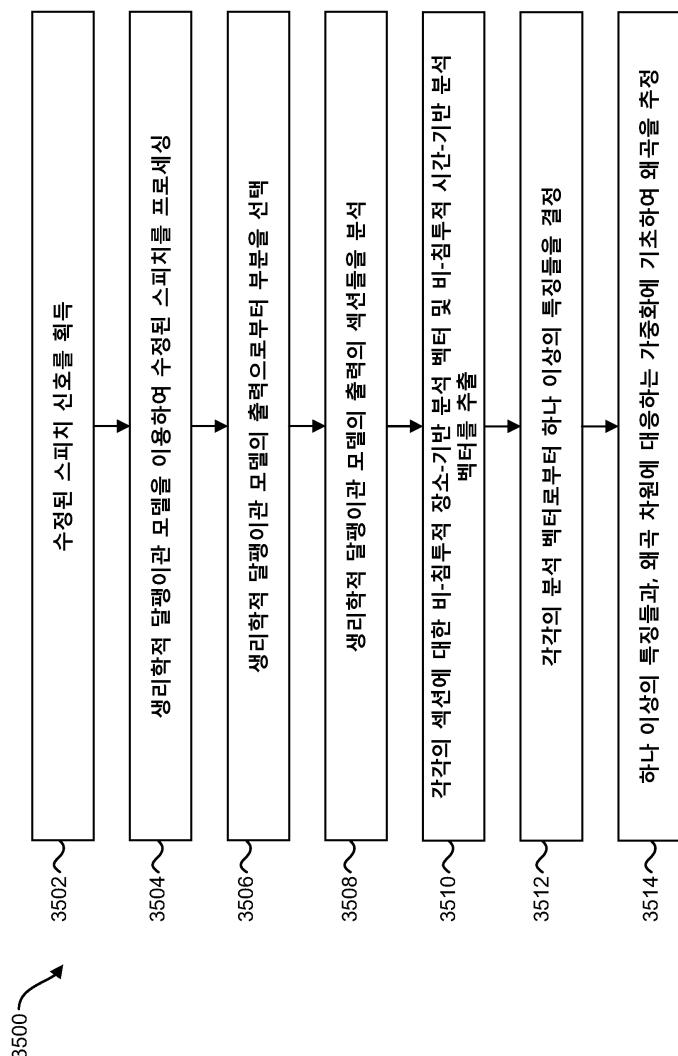
도면33



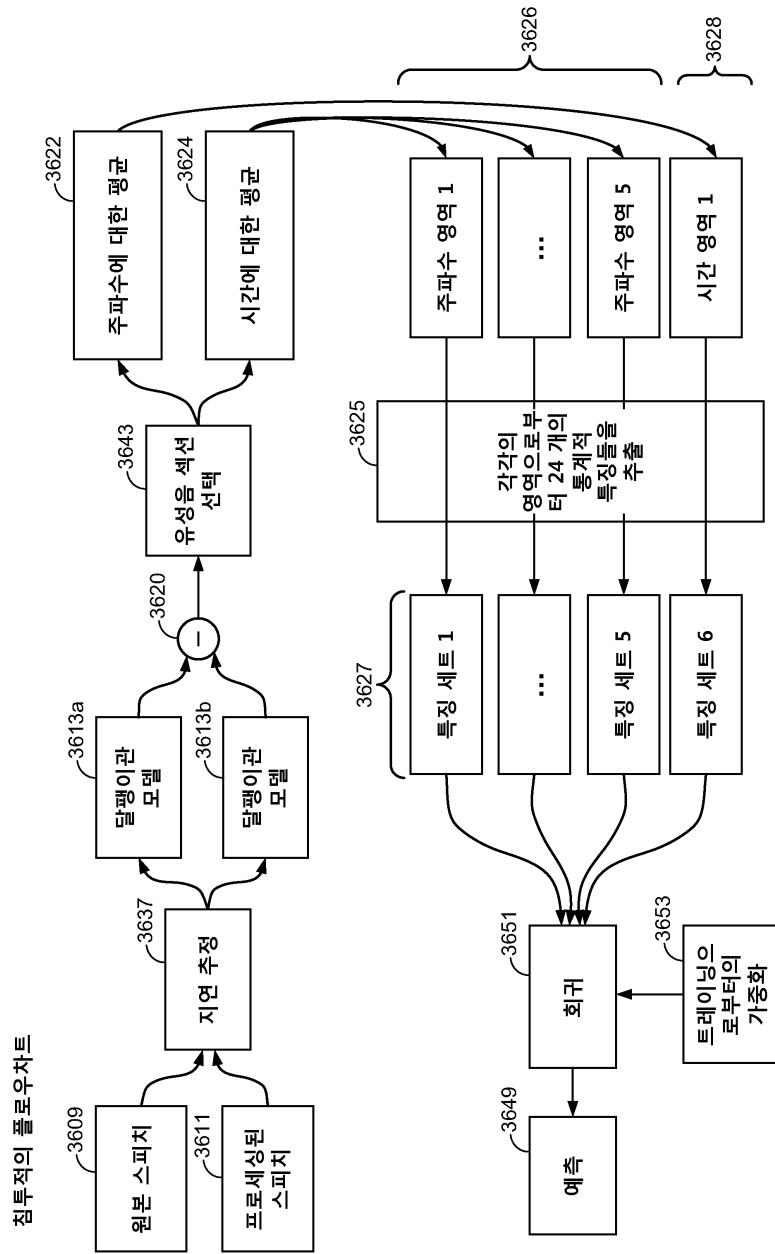
도면34



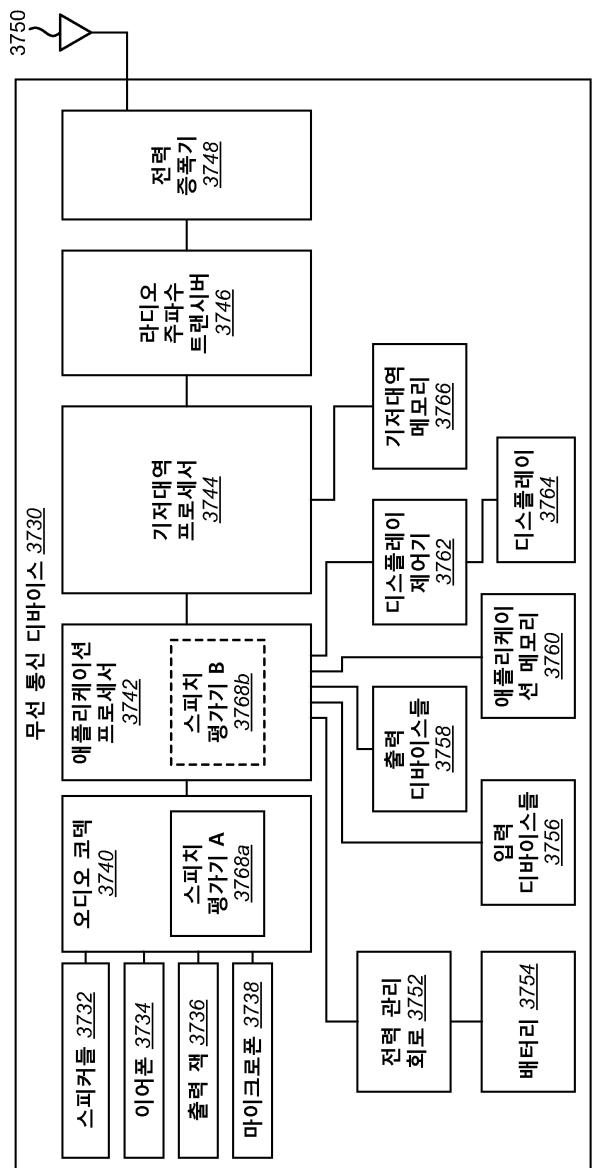
도면35



도면36



도면37



도면38

