(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2015-0022786
(43) 공개일자 2015년03월04일

(51) 국제특허분류(Int. Cl.)
G10L 15/22 (2006.01) *G10L 15/32* (2013.01)
G10L 15/02 (2006.01)
(21) 출원번호 10-2014-7033285
(22) 출원일자(국제) 2013년04월23일
심사청구일자 없음
(85) 번역문제출일자 2014년11월26일
(86) 국제출원번호 PCT/US2013/037679
(87) 국제공개번호 WO 2013/163113
국제공개일자 2013년10월31일
(30) 우선권주장
13/456,959 2012년04월26일 미국(US)

(71) 출원인
뉘앙스 커뮤니케이션즈, 인코포레이티드
미합중국, 매사추세츠 01803, 버링톤, 원 웨이사이드 로드
(72) 발명자
뉴만, 미첼 제이.
미국 메사추세스 02143, 12마운튼 아버뉴 서머빌
로스, 로버트
미국 메사추세스02460, 뉴턴 아파트9,73 엘름 로드
(뒷면에 계속)
(74) 대리인
특허법인가산

전체 청구항 수 : 총 24 항

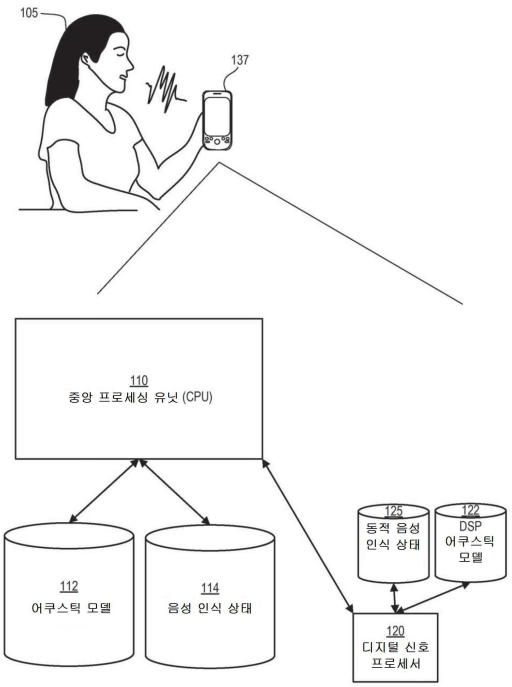
(54) 발명의 명칭 **사용자 정의 제약 조건으로 소형 풋프린트 음성 인식을 구성하는 임베디드 시스템**

(57) 요 약

본 명세서에 개시된 기술은, 음성 명령 기능을 수동적인 개시할 필요 없이 전자 디바이스를 웨이크업(wake-up)하는 음성 트리거(voice trigger)를 가능하게 하거나 상기 디바이스가 추가적인 음성 명령을 동작시키도록 하는 시스템 및 방법을 포함한다. 또한, 이러한 음성 트리거는 동적으로 프로그램 가능하거나 커스터마이징이 가능하다.

(뒷면에 계속)

대 표 도 - 도1



발성자(speaker)는 특정 구문을 음성 트리거로서 프로그램하거나 지정할 수 있다. 일반적으로, 본 명세서에 개시된 기술은, 중앙 프로세싱 유닛(central processing unit, CPU) 상에서 실행되는 대신, 전자 디바이스의 디지털 신호 프로세서(digital signal processor, DSP) 또는 기타 저전력, 보조 프로세싱 유닛 상에서 동작하는 음성 동작(voice-activated) 웨이크업 시스템을 실행한다. 음성 인식 관리부는 전자 디바이스 상에서 2 개의 음성 인식 시스템을 실행한다. CPU는 DSP 용 소형 음성 시스템을 동적으로 생성한다. 이러한 소형 시스템은 대기 모드 동안, 배터리 공급을 빨리 소진시키지 않고 연속적으로 실행될 수 있다.

(72) 발명자

알렉산더 윌리엄 디.

미국 메사추세스 02421 렉싱턴 리지 드라이브 렉싱턴

멀브레트, 폴 반

미국 메사추세스01778, 30 보로드 웨일런드

특허청구의 범위

청구항 1

컴퓨터로 구현된 음성 인식 관리 방법으로서,

제1 프로세서 및 제2 프로세서를 포함하는 전자 디바이스의 상기 제1 프로세서에서 수신되고, 트리거 구문(trigger phrase)를 포함하는 설정 입력(configuration input)을 상기 전자 디바이스의 음성 동작(voice-activated) 웨이크업(wake-up) 기능에서 수신하는 단계;

상기 트리거 구문에 대응하는 음성 인식 상태(speech recognition state)의 네트워크를, 상기 제1 프로세서가 실행시키는 제1 음성 인식 엔진을 이용하여 상기 제1 프로세서에서 생성하는 단계;

상기 음성 인식 상태의 네트워크를 상기 제1 프로세서로부터 상기 제2 프로세서로 전달하는 단계; 및

상기 트리거 구문에 대응하는 음성 인식 상태의 네트워크를 이용하여 상기 제2 프로세서 상에서 제2 음성 인식 엔진을 실행시키는 단계를 포함하고,

상기 제2 프로세서는 상기 제1 프로세서의 상기 제1 음성 인식 엔진이 비동작 상태(inactive state)인 동안 상기 제2 음성 인식 엔진을 실행시키는 컴퓨터로 구현된 음성 인식 관리 방법.

청구항 2

제1항에 있어서,

상기 제2 음성 인식 엔진을 실행시키는 단계는,

상기 제1 음성 인식 엔진이 비동작 상태인 동안 상기 제2 프로세서를 이용하여 발성음(spoken utterance)을 연속적으로 분석하는 단계; 및

특정 발성음이 상기 트리거 구문을 포함하는 것을 식별한 것에 대한 응답으로, 상기 제1 프로세서에 시그널링(signaling)을 하여 상기 제1 음성 인식 엔진을 동작 상태(active state)로 복귀시키는 단계를 포함하는 컴퓨터로 구현된 음성 인식 관리 방법.

청구항 3

제2항에 있어서,

상기 제1 음성 인식 엔진을 동작 상태로 복귀시키는 단계는,

상기 제1 프로세서가 후속의 음성 명령에 응답하는 단계를 포함하는 컴퓨터로 구현된 음성 인식 관리 방법.

청구항 4

제1항에 있어서,

상기 제1 프로세서의 상기 제1 음성 인식 엔진이 비동작 상태인 동안 상기 제2 음성 인식 엔진을 실행시키는 단계는,

상기 제1 프로세서가 비동작 음성 명령 모드(inactive voice command mode)인 단계; 및

특정 발성음(spoken utterance)이 상기 트리거 구문을 포함하는 것을 식별한 것에 대한 응답으로, 상기 전자 디바이스를 비동작 음성 명령 모드로부터 동작 음성 명령 모드(active voice command mode)로 변경하는 단계를 포함하는 컴퓨터로 구현된 음성 인식 관리 방법.

청구항 5

제1항에 있어서,

상기 제2 프로세서는, 상기 전자 디바이스가 대기 모드(standby mode)인 동안 상기 음성 동작 웨이크업 기능을 제공하는 상기 제2 음성 인식 엔진을 실행시키는 컴퓨터로 구현된 음성 인식 관리 방법.

청구항 6

제1항에 있어서,

상기 음성 인식 상태를 상기 제2 프로세서에 전달하는 단계는,

상기 음성 인식 상태를 디지털 신호 프로세서(digital signal processor, DSP)에 전달하는 단계를 포함하는 컴퓨터로 구현된 음성 인식 관리 방법.

청구항 7

제1항에 있어서,

상기 설정 입력을 수신하는 단계는,

상기 전자 디바이스의 사용자 인터페이스를 통해 텍스트 입력으로서 상기 트리거 구문을 수신하는 단계를 포함하는 컴퓨터로 구현된 음성 인식 관리 방법.

청구항 8

제7항에 있어서,

상기 설정 입력을 수신하는 단계는 상기 텍스트 입력을 확인하는 발성음(spoken utterance)을 수신하는 단계를 포함하는 컴퓨터로 구현된 음성 인식 관리 방법.

청구항 9

제1항에 있어서,

상기 제1 프로세서는 중앙 프로세싱 유닛(central processing unit)이고,

상기 제2 프로세서는 디지털 신호 프로세서(digital signal processor)인 컴퓨터로 구현된 음성 인식 관리 방법.

청구항 10

제9항에 있어서,

상기 제2 프로세서는, 상기 제1 프로세서가 동작 상태(active state)일 때, 상기 제1 프로세서보다 더 적은 전력을 사용하는 컴퓨터로 구현된 음성 인식 관리 방법.

청구항 11

제9항에 있어서,

상기 제1 프로세서 및 상기 제2 프로세서는 모바일 전화 내에 공동 배치된(co-located) 컴퓨터로 구현된 음성 인식 관리 방법.

청구항 12

음성 인식 관리 시스템으로서,

제1 프로세서;

제2 프로세서; 및

상기 제1 프로세서 및 상기 제2 프로세서에 연결된 메모리를 포함하고,

상기 메모리는, 상기 제1 프로세서 및 상기 제2 프로세서에 의해 실행되는 경우, 상기 음성 인식 관리 시스템으로 하여금:

상기 제1 프로세서 및 상기 제2 프로세서를 포함하는 전자 디바이스의 상기 제1 프로세서에서 수신되고, 트리거 구문(trigger phrase)를 포함하는 설정 입력(configuration input)을 상기 전자 디바이스의 음성 동작(voice-activated) 웨이크업(wake-up) 기능에서 수신하는 단계;

상기 트리거 구문에 대응하는 음성 인식 상태(speech recognition state)의 네트워크를, 상기 제1 프로세서가 실행시키는 제1 음성 인식 엔진을 이용하여 상기 제1 프로세서에서 생성하는 단계;

상기 음성 인식 상태의 네트워크를 상기 제1 프로세서로부터 상기 제2 프로세서로 전달하는 단계; 및

상기 트리거 구문에 대응하는 음성 인식 상태의 네트워크를 이용하여 상기 제2 프로세서 상에서 제2 음성 인식 엔진을 실행시키는 단계를 수행하도록 하고,

상기 제2 프로세서는 상기 제1 프로세서의 상기 제1 음성 인식 엔진이 비동작 상태(inactive state)인 동안 상기 제2 음성 인식 엔진을 실행시키는 음성 인식 관리 시스템.

청구항 13

제12항에 있어서,

상기 제2 음성 인식 엔진을 실행시키는 단계는,

상기 제1 음성 인식 엔진이 비동작 상태인 동안 상기 제2 프로세서를 이용하여 발성음(spoken utterance)을 연속적으로 분석하는 단계; 및

특정 발성음이 상기 트리거 구문을 포함하는 것을 식별한 것에 대한 응답으로, 상기 제1 프로세서에 시그널링(signaling)을 하여 상기 제1 음성 인식 엔진을 동작 상태(active state)로 복귀시키는 단계를 포함하는 컴퓨터로 구현된 음성 인식 관리 시스템.

청구항 14

제13항에 있어서,

상기 제1 음성 인식 엔진을 동작 상태로 복귀시키는 단계는,

상기 제1 프로세서가 후속의 음성 명령에 응답하는 단계를 포함하는 컴퓨터로 구현된 음성 인식 관리 시스템.

청구항 15

제12항에 있어서,

상기 제1 프로세서의 상기 제1 음성 인식 엔진이 비동작 상태인 동안 상기 제2 음성 인식 엔진을 실행시키는 단계는,

상기 제1 프로세서가 비동작 음성 명령 모드(inactive voice command mode)인 단계; 및

특정 발성음(spoken utterance)이 상기 트리거 구문을 포함하는 것을 식별한 것에 대한 응답으로, 상기 전자 디바이스를 비동작 음성 명령 모드로부터 동작 음성 명령 모드(active voice command mode)로 변경하는 단계를 포함하는 컴퓨터로 구현된 음성 인식 관리 시스템.

청구항 16

제12항에 있어서,

상기 제2 프로세서는, 상기 전자 디바이스가 대기 모드(standby mode)인 동안 상기 음성 동작 웨이크업 기능을 제공하는 상기 제2 음성 인식 엔진을 실행시키는 컴퓨터로 구현된 음성 인식 관리 시스템.

청구항 17

제12항에 있어서,

상기 음성 인식 상태를 상기 제2 프로세서에 전달하는 단계는,

상기 음성 인식 상태를 디지털 신호 프로세서(digital signal processor, DSP)에 전달하는 단계를 포함하는 컴퓨터로 구현된 음성 인식 관리 방법.

청구항 18

제12항에 있어서,

상기 설정 입력을 수신하는 단계는,

상기 전자 디바이스의 사용자 인터페이스를 통해 텍스트 입력으로서 상기 트리거 구문을 수신하는 단계를 포함하는 컴퓨터로 구현된 음성 인식 관리 방법.

청구항 19

데이터 정보를 처리하기 위한 명령이 저장된 비일시적인(non-transitory) 컴퓨터 저장 매체를 포함하는 컴퓨터 프로그램 제품에 있어서,

상기 명령은, 프로세싱 디바이스에 의해 실행되는 경우, 상기 프로세싱 디바이스로 하여금:

제1 프로세서 및 제2 프로세서를 포함하는 전자 디바이스의 상기 제1 프로세서에서 수신되고, 트리거 구문(trigger phrase)를 포함하는 설정 입력(configuration input)을 상기 전자 디바이스의 음성 동작(voice-activated) 웨이크업(wake-up) 기능에서 수신하는 단계;

상기 트리거 구문에 대응하는 음성 인식 상태(speech recognition state)의 네트워크를, 상기 제1 프로세서가 실행시키는 제1 음성 인식 엔진을 이용하여 상기 제1 프로세서에서 생성하는 단계;

상기 음성 인식 상태의 네트워크를 상기 제1 프로세서로부터 상기 제2 프로세서로 전달하는 단계; 및

상기 트리거 구문에 대응하는 음성 인식 상태의 네트워크를 이용하여 상기 제2 프로세서 상에서 제2 음성 인식 엔진을 실행시키는 단계를 수행하도록 하고,

상기 제2 프로세서는 상기 제1 프로세서의 상기 제1 음성 인식 엔진이 비동작 상태(inactive state)인 동안 상기 제2 음성 인식 엔진을 실행시키는 컴퓨터 프로그램 제품.

청구항 20

제19항에 있어서,

상기 제2 음성 인식 엔진을 실행시키는 단계는,

상기 제1 음성 인식 엔진이 비동작 상태인 동안 상기 제2 프로세서를 이용하여 발성음(spoken utterance)을 연속적으로 분석하는 단계; 및

특정 발성음이 상기 트리거 구문을 포함하는 것을 식별한 것에 대한 응답으로, 상기 제1 프로세서에 시그널링(signaling)을 하여 상기 제1 음성 인식 엔진을 동작 상태(active state)로 복귀시키는 단계를 포함하는 컴퓨터로 구현된 컴퓨터 프로그램 제품.

청구항 21

컴퓨터로 구현된 음성 인식 관리 방법으로서,

트리거 구문(trigger phrase)를 포함하는 설정 입력(configuration input)을, 제1 음성 인식 엔진을 실행시키는 제1 프로세서 및 제2 프로세서를 포함하는 전자 디바이스의 음성 동작(voice-activated) 웨이크업(wake-up) 기능에서 수신하는 단계;

상기 설정 입력을 원격 서버 컴퓨터에 전송하는 단계;

상기 트리거 구문에 대응하는 음성 인식 상태(speech recognition state)의 네트워크를, 상기 원격 서버 컴퓨터에서 생성하는 단계;

상기 전자 디바이스에서 상기 트리거 구문에 대응하는 음성 인식 상태의 네트워크를 수신하는 단계;

상기 음성 인식 상태의 네트워크를 상기 제2 프로세서로 전달하는 단계; 및

상기 트리거 구문에 대응하는 음성 인식 상태의 네트워크를 이용하여 상기 제2 프로세서 상에서 제2 음성 인식 엔진을 실행시키는 단계를 포함하고,

상기 제2 프로세서는 상기 제1 프로세서의 상기 제1 음성 인식 엔진이 비동작 상태(inactive state)인 동안 상기 제2 음성 인식 엔진을 실행시키는 컴퓨터로 구현된 음성 인식 관리 방법.

청구항 22

제21항에 있어서,

상기 제2 음성 인식 엔진을 실행시키는 단계는,

상기 제1 음성 인식 엔진이 비동작 상태인 동안 상기 제2 프로세서를 이용하여 발성음(spoken utterance)을 연속적으로 분석하는 단계; 및

특정 발성음이 상기 트리거 구문을 포함하는 것을 식별한 것에 대한 응답으로, 상기 제1 프로세서에 시그널링(signaling)을 하여 상기 제1 음성 인식 엔진을 동작 상태(active state)로 복귀시키는 단계를 포함하는 컴퓨터로 구현된 음성 인식 관리 방법.

청구항 23

제22항에 있어서,

상기 제1 음성 인식 엔진을 동작 상태로 복귀시키는 단계는,

상기 제1 프로세서가 후속의 음성 명령에 응답하는 단계를 포함하는 컴퓨터로 구현된 음성 인식 관리 방법.

청구항 24

제21항에 있어서,

상기 제1 프로세서의 상기 제1 음성 인식 엔진이 비동작 상태인 동안 상기 제2 음성 인식 엔진을 실행시키는 단계는,

상기 제1 프로세서가 비동작 음성 명령 모드(inactive voice command mode)인 단계; 및

특정 발성음(spoken utterance)이 상기 트리거 구문을 포함하는 것을 식별한 것에 대한 응답으로, 상기 전자 디바이스를 비동작 음성 명령 모드로부터 동작 음성 명령 모드(active voice command mode)로 변경하는 단계를 포함하는 컴퓨터로 구현된 음성 인식 관리 방법.

명 세 서

기 술 분 야

[0001] 본 발명은 음성 인식, 특히 음성 동작(voice-activated) 또는 음성 명령 기능을 제공하는 음성 인식에 관한 것이다.

배 경 기 술

[0002] 음성 인식, 또는 자동 음성 인식은 발성된 단어를 식별하는 전산화된 프로세스를 포함한다. 음성 인식은 음성 전사(transcription), 음성 번역, 음성으로 디바이스 및 소프트웨어 어플리케이션을 제어, 라우팅 시스템을 호출, 인터넷에 대한 음성 검색 등을 포함하여 다양한 용도를 가진다. 음성 인식 시스템은 선택적으로 음성 언어 이해 시스템(spoken language understanding system)과 함께, 시스템과 상호 작용을 하면서 실행할 명령 및/또는 의미를 추출할 수 있다.

[0003] 음성 인식 시스템은 고도로 복잡하고, 발음(utterence)의 어쿠스틱 시그너처(acoustic signature)를 단어의 어쿠스틱 시그너처와 매칭함으로써 동작한다. 이러한 매칭은 통계적 언어 모델과 선택적으로 조합될 수 있다. 따라서, 어쿠스틱 모델링(acoustic modeling)과 언어 모델링(language modeling)은 모두 음성 인식 프로세스에서 사용된다. 어쿠스틱 모델은 발성음(spoken utterance)의 오디오 레코딩뿐 아니라 연관된 전사(transcription)로부터 생성될 수 있다. 이후 어쿠스틱 모델은 대응되는 단어에 대한 개별 소리의 통계적 표현을 정의한다. 음성 인식 시스템은 어쿠스틱 모델을 사용하여 일련의 소리를 식별하고, 음성 인식 시스템은 통계적 언어 모델을 사용하여 식별된 소리로부터 가능한 단어 시퀀스를 식별한다.

[0004] 음성 동작 또는 음성 명령 기능을 제공하는 음성 인식은 발성자(speaker)가 다양한 명령을 발성함에 의해 디바이스 및 시스템을 제어할 수 있도록 한다. 예를 들어, 발성자는 커맨드(command)를 발음하여 특정 태스크를 실행하거나, 쿼리(query)를 발음하여 특정 결과를 인출할 수 있다. 발성된 입력(spoken input)은 특정 태스크를

수행하는 구문의 엄격한 세트(rigid set)를 따를 수 있고, 또는 발성된 입력은, 음성 인식 시스템의 자연 언어 유닛(natural language unit)에 의해 해석되는 자연 언어(natural language)일 수 있다. 음성 명령 기능은 휴대용 디바이스, 특히, 휴대폰, 랩톱 및 태블릿 컴퓨터와 같이 배터리 전원으로 동작하는 휴대용 디바이스 상에서 점점 대중적이 되고 있다.

발명의 내용

해결하려는 과제

[0005] 소정의 디바이스의 음성 명령 모드는 능동적으로 발성된 명령을 청취하거나, 사용자가 수동적으로 음성 명령 모드를 동작시킬 때까지 꺼져 있을 수 있다. 예를 들어, 일부 모바일 폰은, 특정 개인에게 전화를 하거나, 이메일 메시지를 확인하거나, 특정 노래를 재생하는 등과 같이 발성된 사용자 입력에 대한 응답으로 태스크를 실행하는 기능을 포함한다. 통상적으로, 사용자는 폰 상의 버튼을 눌러 (또는 아이콘을 선택하여) 음성 명령 모드를 활성화한 후, "존 스미스에게 전화를 걸어(Call John Smith)"와 같은 명령을 발성한다. 이에 대한 응답으로, 폰 또는 디바이스는, 연락처 리스트 또는 기타 디렉터리로부터 검색된 전화 번호와 같이, 대응되는 전화 번호로 통화를 개시할 것이다.

[0006] 이러한 음성 명령 기능은 편리하지만, 그럼에도 불구하고 이러한 기능에 접근하는 것은 사용자로 하여금 버튼을 누르게 하거나, 아니면 선택하고자 하는 메뉴를 찾도록 한다. 따라서, 음성 명령 모드 또는 기능을 개시하는 것은 핸드-프리(hands-free) 프로세스가 아니다. 또한, 연속적으로 실행되는 음성 명령 음성 인식 엔진은 종래의 음성 인식 엔진의 상당한 전력 요구 때문에, 핸드-프리 대체 수단으로서 바람직하지 않다. 예를 들어, 종래의 음성 인식 프로그램을 음성 명령 모드의 일부로서 연속적으로 실행시키는 통상적인 모바일 폰은, (배터리 스펙에 의존적으로) 몇 시간 내에 배터리 전원을 모두 소진시킬 것이다. 종래의 음성 인식 시스템은 장치 효율 등급(device efficiency rating)에 부정적인 영향을 미치기 때문에, 전력 소비는 또한 유선 디바이스 또는 장치의 경우에도 문제가 된다. 이에 따라, 적은 전력을 사용하면서, 이에 따라 효율적인 핸드-프리 동작을 제공하는, 모바일 폰 또는 기타 디바이스의 음성 인식 프로그램을 개시하는 음성 동작 "웨이크업" 기능이 요구된다.

과제의 해결 수단

[0007] 따라서, 본 명세서에 개시된 기술은, 디바이스를 웨이크업하는 음성 트리거(voice trigger)를 가능하게 하거나, 아니면, (버튼을 누르는) 음성 명령 기능의 수동적 개시와 유사하게, 디바이스가 모든/나머지 음성 명령을 동작시키도록 하는 시스템 및 방법을 포함한다. 또한, 이러한 음성 트리거는 동적으로 프로그램 가능하거나 커스터마이징이 가능하다. 예를 들어, 발성자는 특정 구문(단어 또는 단어들)을 음성 트리거로서 프로그램하거나 지정할 수 있고, 이러한 음성 트리거를 원하는 대로 변경할 수 있다. 일반적으로, 본 명세서의 기술은, 전자 디바이스의 디지털 신호 프로세서(digital signal processor, DSP) 또는 기타 저전력, 보조 프로세싱 유닛 상에서 동작하는 음성 동작 웨이크업 시스템을 실행한다. 종래의 음성 인식 시스템은 상대적으로 크기 때문에 DSP 상에서 음성 인식 프로그램을 실행하는 것은 문제가 될 수 있다. DSP 상의 웨이크업 기능을 실행하는 것에 대한 다른 문제는 동적으로 프로그램 가능한 웨이크업 기능을 제공하는 것이다. 모바일 폰을 제조하는 동안과 같이, 미리 프로그램된 웨이크업 구문은 DSP 상에서 웨이크업 시스템을 실행하는 것을 용이하게 하지만, 사용자로 하여금 명령 단어를 설정하도록 허용하는 것은 어려운 문제이다.

[0008] 본 명세서의 기술은 전자 디바이스 상에서 2 개의 음성 인식 시스템을 실행하는 것을 포함한다. 상대적으로 큰 제1 음성 시스템은 디바이스의 중앙 프로세싱 유닛(central processing unit, CPU) 상에서 실행되고, 상대적으로 작은 제2 음성 시스템은 DSP 상에서 실행된다. 소정의 휴대용 전자 디바이스의 CPU는 통상적으로 전자 디바이스가 동작 상태에 있는 동안의 대부분의 프로세싱을 처리한다. DSP는 통상적으로 상대적으로 저전력으로 실행되고, 대기 프로세싱을 위해 사용된다. 예를 들어, 전자 디바이스가 모바일 폰으로서 구현되는 경우, DSP는 통상적으로 모바일 폰이 대기 모드(CPU는 비동작 상태(inactive state)에 있음)인 동안 수신되는 전화 호출을 청취하고 응답한다. 제1 음성 인식 시스템은 커스텀 트리거 구문(custom trigger phrase)을 취하여, 음성 모델(speech model) 및 코드를 상대적으로 작은 시스템으로 변환한다. 그 후 이러한 제2 음성 시스템은 DSP에 전달되어, 전자 디바이스가 대기 모드에 있는 경우와 같이 CPU가 비동작 상태인 동안 DSP는 제2 음성 시스템을 실행할 수 있다. 이러한 제2 음성 시스템을 사용하는 DSP는 커스텀 트리거 구문을 청취한다. 커스텀 트리거 구문이 발음된 것을 검출하면, DSP는 전자 디바이스에가 동작 상태, 또는 CPU가 이어지는 음성 커맨드에 응답할 수 있는 상태로 복귀하도록 신호를 보낸다. 더 작은 제2 음성 시스템을 DSP 상에서 실행시킴으로써, 전자 디바이스는

상대적으로 낮은 전력을 사용하는 핸드-프리 음성 동작 웨이크업 기능을 제공할 수 있다.

[0009] 본 발명의 일 실시예는 음성 인식 프로세스를 실행하는 음성 인식 관리부 또는 전자 디바이스 상에서 음성 인식을 관리하기 위한 시스템을 포함한다. 음성 인식 관리부는 전자 디바이스의 음성 동작 웨이크업 기능에서 설정 입력(configuration input), 즉, 웨이크업 기능을 설정하기 위한 인터페이스를 수신한다. 설정 입력은 트리거 구문을 포함한다. 설정 입력은 전자 디바이스의 제1 프로세서에서 수신된다. 이러한 전자 디바이스는 제1 프로세서뿐 아니라 제2 프로세서도 구비한다. 음성 인식 관리부는 트리거 구문에 대응하는 음성 인식 상태(speech recognition state)의 네트워크를 생성한다. 음성 인식 상태의 네트워크는 제1 프로세서가 실행하는 제1 음성 인식 엔진을 사용하여 제1 프로세서에서 생성된다. 음성 인식 관리부는 음성 인식 상태의 네트워크를, CPU에서 DSP로와 같이, 제1 프로세서로부터 제2 프로세서로 전달한다. 그 후 음성 인식 관리부는 트리거 구문에 대응하는 음성 인식 상태의 네트워크를 사용하여 제2 프로세서 상에서 제2 음성 인식 엔진을 실행한다. 제1 프로세서의 제1 음성 인식 엔진이 비동작 상태인 동안, 제2 프로세서는 제2 음성 인식 엔진을 실행한다. 이러한 제2 음성 인식 엔진을 실행하는 것은, 제1 음성 인식 엔진, 또는 제1 프로세서가 비동작 상태인 동안 발성음(spoken utterance)을 연속적으로 분석하는 것을 포함할 수 있다. 특정 발성음이 트리거 구문을 포함하는 것을 식별한 것에 대한 응답으로, DSP 또는 음성 인식 관리부는 제1 프로세서에 신호를 보내 제1 음성 인식 엔진이 동작 상태로 복귀하도록 하여, 전자 디바이스를 제어하기 위한 음성 명령 모드를 가능하게 한다.

[0010] 본 발명의 다른 실시예는 위에서 요약되고 아래에서 상세하게 개시되는 단계 및 동작을 수행하는 소프트웨어 프로그램을 포함한다. 이러한 일 실시예는, 컴퓨터 저장 매체(예컨대, 비일시적인(non-transitory), 유형의(tangible), 컴퓨터로 판독 가능한 매체, 분리되어 위치하거나 공통으로 위치하는 스토리지 매체, 컴퓨터 스토리지 매체 또는 매체들 등)를 구비한 컴퓨터 프로그램 제품을 포함하고, 컴퓨터 저장 매체는 컴퓨터 저장 매체 상에 인코딩된 컴퓨터 프로그램 로직을 포함한다. 컴퓨터 프로그램 로직은, 프로세서 및 대응되는 메모리를 구비하는 컴퓨팅 디바이스에서 실행되는 경우, 프로세서가 본 명세서에서 개시된 동작들을 수행하도록(또는 프로세서가 본 명세서에서 개시된 동작들을 수행하도록 만들도록) 프로그램한다. 이러한 구성은 통상적으로 소프트웨어, 펌웨어, 마이크로코드, 코드 데이터(예컨대, 데이터 구조) 등으로서 제공되고, 광학 매체(예컨대, CD-ROM), 플로피 디스크, 하드 디스크, 하나 이상의 ROM 또는 RAM 또는 PROM 칩, ASIC(Application Specific Integrated Circuit), FPGA(field-programmable gate array) 등의 컴퓨터로 판독 가능한 스토리지 매체 상에 배치되고 인코딩된다. 소프트웨어 또는 펌웨어 또는 기타 이러한 구성들은 컴퓨팅 디바이스 상에 설치되어, 컴퓨팅 디바이스로 하여금 본 명세서에서 설명된 기술을 수행하도록 할 수 있다.

[0011] 따라서, 본 발명의 일 실시예는 하나 이상의 비일시적 컴퓨터 스토리지 매체를 포함하는 컴퓨터 프로그램 제품에 관한 것이다. 하나 이상의 비일시적 컴퓨터 스토리지 매체는 다음과 같은 동작들을 지원하기 위해 하나 이상의 비일시적 컴퓨터 스토리지 매체 상에 저장된 명령을 포함한다: 전자 디바이스의 음성 동작 웨이크업 기능에서 설정 입력을 수신하고, 설정 입력은 트리거 구문을 포함하고, 설정 입력은 전자 디바이스의 제1 프로세서에서 수신되고, 전자 디바이스는 제1 프로세서뿐 아니라 제2 프로세서를 포함하고; 트리거 구문에 대응하는 음성 인식 상태의 네트워크를 생성하고, 음성 인식 상태의 네트워크는 제1 프로세서가 실행하는 제1 음성 인식 엔진을 사용하여 제1 프로세서에서 생성되고; 음성 인식 상태의 네트워크를 제1 프로세서로부터 제2 프로세서로 전달하고; 트리거 구문에 대응하는 음성 인식 상태의 네트워크를 사용하여 제2 프로세서 상에서 제2 음성 인식 엔진을 실행하고, 제1 프로세서의 제1 음성 인식 엔진이 비동작 상태인 동안 제2 프로세서는 제2 음성 인식 엔진을 실행한다. 본 명세서에서 설명된 명령 및 방법은, 각각의 컴퓨팅 디바이스의 프로세서에 의해 수행되는 경우, 프로세서로 하여금 본 명세서에 개시된 방법을 수행하도록 한다.

[0012] 본 발명의 다른 실시예는 위에서 요약되고 아래에서 상세하게 설명되는 방법 실시예의 임의의 단계 및 동작을 수행하는 소프트웨어 프로그램을 포함한다.

[0013] 물론, 본 명세서에서 설명된 바와 같은 여러 단계들에 대한 논의 순서는 명확성을 위해 제시된 것이다. 일반적으로 이러한 단계들은 임의의 적절한 순서에 따라 수행될 수 있다.

[0014] 또한, 본 명세서의 각각의 시스템, 방법, 장치 등은, 프로세서 내에서, 운영 체제 내에서 또는 소프트웨어 어플리케이션 내에서와 같이, 소프트웨어 프로그램으로서, 소프트웨어 및 하드웨어의 하이브리드(hybrid)로서, 또는 하드웨어 단독적으로 엄격하게 구현될 수 있음을 이해해야 한다. 또는, 본 명세서의 각각의 시스템, 방법, 장치 등은 동작의 일부 또는 전부를 수행하는 사람과 같은 비-소프트웨어 어플리케이션(non-software application)을 통해 구현될 수 있음을 이해해야 한다.

[0015] 상술한 바와 같이, 본 명세서의 기술은 음성 명령 음성 인식을 지원하는 소프트웨어 어플리케이션에 사용되는

것이 바람직하다. 그러나, 본 명세서의 실시예들은 이러한 어플리케이션의 사용에 한정되지 않으며, 본 명세서의 기술은 다른 어플리케이션에서도 또한 사용될 수 있다.

[0016] 또한, 본 명세서의 각각의 여러가지 특징, 기술, 설정 등은 본 개시의 다른 위치에서 논의될 수 있지만, 각각의 개념들은 서로 독립적으로 실행되거나, 서로 조합되어 실행될 수 있다. 따라서 본 발명은 다수의 상이한 방식으로 관찰되고 구현될 수 있다.

[0017] 본 명세서의 과제의 해결 수단은 본 발명 또는 청구항 발명의 모든 실시예 및/또는 점점 신규한 모든 측면들을 기재하고 있지 않다. 대신에, 과제의 해결 수단은 단지 종래 기술에 비해 신규한 여러 실시예들 및 이에 대응하는 논점들에 대한 예비적인 논의만을 제공할 뿐이다. 본 발명 및 실시예에 대한 추가적인 설명 및/또는 가능한 관점에 대해, 이어지는 발명을 실시하기 위한 구체적인 내용 및 도면을 참조할 것이다.

도면의 간단한 설명

[0018] 본 발명의 상술한 객체, 특징 및 이점 및 기타 객체, 특징 및 이점은 첨부된 도면에 도시된 바와 같은 본 명세서에서 이어지는 바람직한 실시예의 더욱 상세한 설명으로부터 명백해질 것이다. 유사한 참조 부호는 상이한 측면에 따른 동일한 요소를 참조할 수 있다. 도면은 반드시 스케일링되어 있지 않으며, 본 발명의 실시예, 원리 및 개념을 설명하기 위해 강조될 수 있다.

도 1은 본 발명에 따른 음성 트리거 웨이크업 기능을 지원하는 디바이스에 대한 시스템의 블록도이다.

도 2는 본 발명에 따른 음성 트리거 웨이크업을 지원하는 프로세스의 예를 설명하는 순서도이다.

도 3 및 도 4는 본 발명에 따른 음성 트리거 웨이크업을 지원하는 프로세스의 예를 설명하는 순서도이다.

도 5는 본 발명에 따른 컴퓨터/네트워크 환경에서 동작하는 음성 인식 관리부의 예시적인 블록도이다.

발명을 실시하기 위한 구체적인 내용

[0019] 본 명세서에 개시된 기술은, 디바이스를 웨이크업(wake-up)하는 음성 트리거(voice trigger)를 가능하게 하거나, (버튼을 누르는) 음성 명령 기능의 수동적 개시와 유사하게, 디바이스가 모든/나머지 음성 명령을 동작시키도록 하는 시스템 및 방법을 포함한다. 또한, 이러한 음성 트리거는 동적으로 프로그램 가능하거나 커스터마이징이 가능하다. 예를 들어, 발성자는 특정 구문(단어 또는 단어들)을 음성 트리거로서 프로그램하거나 지정할 수 있고, 이러한 음성 트리거 구문은 사용자가 원하는대로 수정되거나 변경될 수 있다. 일반적으로, 본 명세서의 기술은, 중앙 프로세싱 유닛(central processing unit, CPU) 상에서 실행하는 대신, 전자 디바이스의 디지털 신호 프로세서(digital signal processor, DSP) 또는 기타 저전력, 보조 프로세싱 유닛 상에서 동작하는 음성 동작(voice-activated) 웨이크업 시스템을 실행한다.

[0020] 본 명세서에 개시된 기술은 2 개의 프로세서를 사용하는 다수의 여러 전자 디바이스에 구현될 수 있지만, 실시예들에 대한 설명의 편의를 위해, 본 개시는 배터리 전원으로 동작하는 셀룰러 폰과 같은 모바일 폰을 주로 참조한다. 종래의 모바일 폰에서는 통상적으로 2 개의 프로세서가 있다. 종래의 모바일 폰에는 상대적으로 고성능인 메인 또는 주 CPU가 있다. 이러한 CPU는, 전화 통화, 유틸리티 어플리케이션, 이메일, 게임 등과 같이 폰이 능동적으로 사용되는 경우, 대부분의 프로세싱을 처리한다. 모바일 폰은 또한 통상적으로 2차 프로세서, 즉, 디지털 신호 프로세서(digital signal processor, DSP)를 포함한다. DSP는 매우 낮은 전력으로 동작할 수 있다. DSP는 또한 (그 자신에 비해) 고전력 모드로 동작할 수도 있다. 매우 긴 시간동안 실행되는 모바일 폰 기능에 대해, 이러한 기능들은 통상적으로 DSP 상에서 실행된다. DSP는 동작 상태로, 폰이 대기 모드 또는 능동적으로 사용되지 않는 동안 수신되는 전화 호출를 청취하는 모바일 폰의 구성 요소이다.

[0021] 본 명세서의 기술은 전자 디바이스 상에서 2 개의 음성 인식 시스템을 실행하는 것을 포함한다. 상대적으로 큰 제1 음성 시스템은 디바이스의 중앙 프로세싱 유닛(central processing unit, CPU) 상에서 실행되고, 상대적으로 작은 제2 음성 시스템은 DSP 상에서 실행된다. CPU 음성 시스템은, 연락처, 이메일, 노래를 접근하거나, 태스크를 수행하거나, 어플리케이션을 실행시키기 위한 커맨드를 인식하기 위한 것과 같은 특정 모바일 어플리케이션에 대한 음성 인식 모델의 풀 세트(full set)를 포함할 수 있다. 이러한 모델의 풀 세트는 사용자로 하여금 발성된 웨이크업 트리거로서 사용하기 위한 커스텀 구문 또는 단어를 선택할 수 있도록 한다.

[0022] 제1 시스템/주 시스템은 커스텀 트리거 구문을 수신한 후, 이러한 트리거 구문을 사용하여 음성 모델 및 코드를 상대적으로 작은 시스템으로 변환한다. 따라서, CPU는 더 작은 음성 시스템을 미리 연산한다. 이러한 시스템은 2 개의 부분을 가질 수 있다: (1) 인식 상태(recognition state)의 네트워크 및 (2) 어쿠스틱 모델(acoustic

model) 정보. 인식 상태의 네트워크는, 상태의 시퀀스를 통해 가장 좋은 경로를 검색하는 디코더에 의해 사용될 수 있다. 어쿠스틱 모델 정보는 음성 레코딩으로부터 측정된 것을 포함할 수 있다. 실시예들은 DSP 웨이크업 기능을 위해 사용되는 어쿠스틱 모델 정보의 매우 작은 서브셋(subset)들로 매우 양호하게 기능할 수 있다. 이러한 2 부분은 서로 결합되어, CPU 음성 인식 시스템에 비해 작은, 작은 시스템을 이룰 수 있다. 한정되지 않는 실시예들에 의해, 인식 상태의 네트워크(CPU에 의해 컴파일된 데이터)는 약 5 킬로바이트(kilobyte) 또는 10 킬로바이트의 사이즈를 가질 수 있는 반면, 주 시스템으로부터의 네트워크는 약 300 킬로바이트 내지 약 2 메가바이트(megabyte)가 될 수 있다. 따라서, 이러한 기술은 현저한 사이즈의 감소를 제공한다.

[0023] 커스텀 입력(커스텀 트리거 구문)은 텍스트 입력으로서 명세될 수 있다. 이와 다른 실시예에서, 입력은 발성되거나 및/또는 텍스트로서 입력될 수 있다. 가청 입력(audible input)은 텍스트 입력을 검증함에 의해 인식 정확도를 향상시키는 데 도움을 줄 수 있다. 그러나 트리거 구문에 대한 가청 입력은 필요하지 않음을 주목해야 한다. 시스템은 초기에 오로지 텍스트로서만 입력된 트리거 구문에 기초하여 웨이크업 기능을 실행할 수 있다. 그 후 미리 연산된 시스템은 DSP로 전달될 수 있다. 본 기술은 상대적으로 작은 음성 디코더를 DSP 상에서 실행시키는 것을 포함한다. 이러한 DSP 디코더는, 예컨대, 약 30 킬로바이트의 코드 사이지를 가질 수 있다. 이와 대조적으로, CPU 상에서 실행되는 음성 디코더는 약 1 내지 2 메가바이트의 사이즈를 가질 수 있다. 예를 들어, DSP 인식 코드는 DSP 상에서 실행되는 몇 백 라인의 코드만을 가질 수 있다. 종래의 인식기에서는, 통상적으로 각각이 상대적으로 큰 다수의 모듈이 존재한다. 이러한 큰 모듈의 전부는 DSP 상에서 실행되는 단일 모듈로 교체될 수 있다. 다시 말해서, 일반적인 인식 엔진을 실행하는 것 대신, 매우 세부적인 인식 엔진이 DSP 상에서 사용되기 위해 생성될 수 있다.

[0024] 다음으로, DSP 상의 인식기는 웨이크업 트리거로서 기능한다. 예를 들어, 인식기는 트리거 구문이 발성되었는지 아닌지의 여부를 필수적으로 결정한다. 만일 DSP가 웨이크업 구문을 수신된 발성음(spoken utterance)으로부터 인식하지 못했다면, 아무 일도 일어나지 않는다. 만일 DSP 인식기가 발음된 웨이크업 구문을 결정한다면, CPU 인식기는 활성화되고, 발성된 명령 또는 쿼리에 대해 응답하는 것을 시작하거나 계속할 수 있다.

[0025] 발성된 구문 또는 단어로 웨이크업이 가능한 디바이스들이 존재하지만, 이러한 트리거 단어들은 고정되어 있다. 즉, 소정의 트리거 명령이 전자 디바이스에 고정되어, 변경될 수 없다. 이러한 고정된 트리거 단어는 통상적으로 제조 공정 중에 설정된다. 다시 말해서, 트리거 단어는 디바이스의 일생 동안 영속적이다. 이와 대조적으로, 본 명세서에 개시된 기술은, 완전히 커스터마이징될 수 있거나 사용자에게 맞춤형인 설정 가능하고/변경 가능한 트리거 단어를 제공한다. 본 명세서에서 제공되는 이러한 커스터마이징 기능은 추가적인 코드를 온라인으로 다운로드거나, 트리거 명령을 재설정하기 위해 디바이스를 공장으로 되돌려 보낼 것을 요구하지 않는다.

[0026] 일 실시예에서, 트레이닝 모드 또는 커스터마이징 인터페이스는 발성음을 필요로 하지는 않지만, 웨이크업 시스템을 생성하여 DSP 상에서 실행하기 위한 텍스트 입력을 사용할 수 있다. 몇몇의 실시예들은 또한 텍스트 입력을 확인하기 위한 음성 입력을 수신할 수 있지만, 이것을 요구도지 않는다. 텍스트 입력만으로부터 웨이크업 시스템을 생성하는 것은, 미리 정해진 어쿠스틱 모델이 실질적으로 더 작은 모델로 감소될 수 있기 때문에 유용하지만, 소정의 트리거 단어에 대해 유사한 인식 성능/정확도를 갖는다. 따라서, 몇몇의 실시예에서, DSP 시스템으로 송신된 어쿠스틱 모델 데이터는 사전에 존재하는 발성자에 대해 독립적인(pre-existing speaker-independent) 어쿠스틱 모델로부터 추출될(abstracted) 수 있다.

[0027] 일 특징은, DSP 어쿠스틱 모델이 특정 사용자의 목소리에 적응할 필요가 없다는 것이다. 대신에 시스템은 더 큰 모델로부터 특정 트리거 구문에 대해 필요한 상태들을 선택할 수 있다. 소정의 트리거 구문에 대해, 대응되는 개수의 필요한 상태 ID(state ID)가 존재할 수 있다. 제한되지 않는 예시적인 방식으로, 만일 소정의 트리거 구문이 2, 7, 48, 138, 455, ...의 상태 ID를 필요로 한다면, 음성 인식 관리부 또는 웨이크업 관리부는 그 후 이러한 상태 ID를 당기고(pull), 더 작은 모델에서 사용하기 위해 리넘버링을 할 것이다. 이것은 더 작고 더욱 효율적인 모델을 야기한다. 따라서, 소정의 단어의 시퀀스에 대해, 주 시스템은 이러한 단어의 시퀀스를 음소(phoneme)의 시퀀스 및 기타 종래의 음성 표현으로 변환한다. 소정의 구문에 대해, 유지될 필요가 없는 적은 개수의 모델 상태만이 존재한다. 따라서 대부분의 모델 데이터는 세부적인 웨이크업 트리거 단어/구문에 대해 사용될 필요가 없다. 이것은 또한 음성 인식에 사용되는 코드의 대부분이 DSP와 같은 2차 프로세싱 유닛 상에서 실행되는 더 작은 시스템에서 사용될 필요가 없음을 의미한다.

[0028] 이제 도 1을 참조하면, 사용자(105)는 전자 디바이스(137)를 동작시킨다. 전자 디바이스(137)는 음성 명령 기능을 포함하고, 음성 명령 기능은 커스터마이징 가능한 웨이크업 기능을 포함한다. 사용자(105)는 특정 트리거 구문에 대한 텍스트 입력을 할 수 있다. 예를 들어, 사용자(105)는 "지니어스 버튼(genius button)", "존의 폰을

활성화(activate John's phone)", "시퀘스터(sequester)", "퍼플 멍키 식기세척기(purple monkey dishwasher)" 등의 임의의 구문을 타이핑할 수 있다. 본 명세서에서 사용된 "구문" 또는 "트리거 구문"이라는 용어는 하나의 단어 또는 복수의 단어를 가리킬 수 있음을 주목해야 한다. (초기 셋업(set-up)을 위해) 사용자(105)가 트리거 구문을 입력한 후, 중앙 프로세싱 유닛(110)은 디지털 신호 프로세서(120)에 의해 사용되는 음성 인식 시스템 또는 상태를 생성한다. CPU(110)는 웨이크업 기능을 위한 데이터를 생성하는 동안 어쿠스틱 모델(112) 및 음성 인식 상태(114)를 접근할 수 있다. 음성 인식 상태(114)을 생성하는 것에 추가적으로, CPU(110)는 어쿠스틱 모델(122)로서 사용되기 위한 어쿠스틱 모델(112)(발성자로부터 독립적인 어쿠스틱 모델)로부터 어쿠스틱 모델 데이터를 추출 또는 수집할 수 있다. 이후 DSP(120)는 이러한 데이터를 수신하여 동적 음성 인식 상태(125)로서 저장할 수 있다. 다음으로, 음성 인식 동안, DSP(12)는 음성 인식 상태(125)와 어쿠스틱 모델(122)을 모두 접근할 수 있다. 초기의 또는 새로운 트리거 구문 및 이에 수반된 데이터가 생성된 후, DSP(120)는 트리거 단어가 발음되었는지 식별하기 위해 음성을 모니터링하는 인식 프로세스를 실행할 수 있다. 이러한 웨이크업 특징은 CPU(110)가 비동작(inactive), 대기 중, 또는 적어도 CPU(110) 상에서 실행되는 음성 명령 기능이 비동작인 동안 실행될 수 있다. CPU(110)는 트리거 구문이 검출될 때까지 비동작으로 유지될 수 있다. DSP(120)가, 트리거 구문이 발성되었음을 검출한 후, DSP는 CPU(120)가 동작 상태 또는 음성 엔진을 실행하고 음성 입력에 대한 응답으로 태스크를 실행할 수 있는 상태로 복귀하도록 신호를 전송한다. 소정의 트리거 단어 및 이에 대응하는 데이터 네트워크는, 사용자가 새로운 트리거 구문을 생성하기를 원하는 때까지 DSP 상에 유지된다. 새로운 트리거 구문을 수신하면, 시스템은 DSP(120)에 대해 네트워크 상태를 생성하는 프로세스를 반복할 수 있다.

[0029] 이제 더욱 상세하게, 새로운 트리거 구문이 CPU(110)에서 수신된 경우 이 문자 스트링은 음소의 시퀀스로 변환될 수 있다. 이 시퀀스는 이미 존재할 수 있거나, 생성될 필요가 있을 수 있다. 각각의 음소에 대해, 관리부는 이웃 음소(neighboring phoneme)(각 측에 하나씩)를 식별하여 트라이폰(tri-phone)을 생성한다. 그 후 각각의 트라이폰은 상태의 시퀀스로 변환된다. 각각의 트라이폰은 어쿠스틱 상태의 시퀀스에 대한 모델을 갖는다. 일반적으로 소정의 트라이폰 모델은 2 또는 몇 개의 상태: 시작 및 종료, 또는 시작, 중간, 종료의 상태를 갖는다. 그 결과는 어쿠스틱 모델 상태의 세트이고, 스코어링(scoring)을 위한 어쿠스틱 모델에서 검색된(looked-up) 것이다. 따라서 트라이폰은 어쿠스틱 모델 또는 어쿠스틱 모델 상태에 매핑되어, 시퀀스를 생성한다.

[0030] 음성 인식 엔진에 의해 발성음을 스코어링하는 것은 통상적으로 상대적인 스코어링 프로세스이다. 음성 인식 관리부는 인식 문법을 사용할 수 있다. 이러한 인식 문법은 트리거 구문을 평가할 뿐 아니라 디코이(decoy) 단어 또는 구문의 세트를 통과하는 다른 경로를 평가하여, 음성 인식 관리부가 너무 자주 인식(거짓 인식(false recognition))하지 않도록 할 수 있다. 인식 문법은 디코이 단어 또는 어쿠스틱 모델 상태 전부를 통하는 경로를 포함한다. 이러한 설정으로, 관리부는 어쿠스틱 모델의 나머지 또는 어떤 단어 모델도 필요로 하지 않는다. 이러한 인식 문법 부분은 DSP에 의해 사용되는 상대적으로 소형인 데이터로 컴파일된 것이다.

[0031] 음성 인식 상태는 확률 분포, 가우시안(gaussian)의 시퀀스로서 모델링될 수 있다. 발성음이 검출된 경우, 발성음은 프레임으로 변환되고, 프레임은 확률 분포에 대해 비교되어 스코어를 획득한다. 디코이는 랜덤 단어의 세트로서 선정될 수 있고, 이것은 트리거 구문과 유사하거나 또는 완전히 다른 것일 수 있다. 다음으로 음성 모델은 발성음 및 하나 이상의 디코이를 평가하여 기준 비교 스코어(reference comparison score)를 구축한다. 만일 발성음의 스코어가 랜덤/디코이 단어의 스코어보다 (미리 정해진 양만큼) 좋은 경우, 관리부는 발성음이 인식되었다고 결정한다. 절대 스코어를 기대하는 모델을 사용하는 것이 선택적으로 사용될 수 있지만, 이러한 기술은 통상적으로 정확도가 높지 않다. 상대 스코어를 사용하는 것은, 배경 잡음 및 음성을 확인하는 동안 발성된 단어를 정확하게 인식할 수 있도록 한다.

[0032] 본 명세서의 DSP 인식 엔진으로 인한 이점은 DSP 인식 엔진이 (트리거 구문 인식 동안) 가설을 세우거나(hypothesize), 단어를 음소의 시퀀스로 변환하거나, 단어를 트라이폰으로 변환하거나, 트라이폰은 시퀀스 상태로 변환할 필요가 없다는 점이다. DSP 인식기는 CPU에 의해 생성된 상태의 시퀀스 상에서 기능할 수 있기 때문에 이러한 프로세스 단계들은 불필요하다. 이러한 디코더는 유한 상태 변환기(finite state transducer, FST)로서 참조될 수 있다. 따라서, FST 네트워크는 주 CPU 상에서 컴파일된 후, FST 디코더로 전달되어(passed down) DSP 상에서 실행된 것이고, 이것은 커스터마이징 가능한 트리거 구문에 대한 응답으로 동적으로 실행되어, 특정 구문에 커스터마이즈된 시스템을 제공하고, 상기 시스템은 CPU 상에서 실행되는 초기 시스템보다 실질적으로 더 작다. DSP 디코더에 대해, 몇몇의 실시예에서, 코드는 동일할 수 있지만, CPU가 컴파일하는 네트워크는 다르다. 데이터는 어쿠스틱 모델(122)를 형성하기 위해 초기에 수집될 수 있지만, 단일 어쿠스틱 모델을 형성한 후, 음성 인식 관리부는 원하는 만큼의 다수의 상이한 작은 모델들(네트워크 상태(125))을 생성할 수 있다.

[0033] 음성 인식 동안, DSP는 발성음을 수신하고, 이것은 네트워크를 통해 프로세싱하여 스코어를 획득한다. DSP 디코

더는 또한 "개(dog)" 또는 "새총(catapult)" 등과 같은 랜덤/디코이 단어를 프로세싱한다. 만일 DSP 디코더가 네트워크를 통해 발성음에 대한 경로를 식별할 수 없다면, 이러한 가설은 폐기된다. 만일 발성음 및 디코이 단어가 네트워크를 통해 모두 형성되면, (트리거 단어가 되는) 발성음은 디코이 단어보다 훨씬 높은 스코어를 획득해야 하며, 스코어 차이는 트리거 단어가 발성된 시스템을 지시하고, CPU가 웨이크업되거나 다시 활성화되도록 한다. 디코이 단어는 임의의 단어가 발성되는 어떤 시점에라도 실행될 수 있다. 이러한 웨이크업 모드에서, DSP는 그것이 청취하는 모든 단어를 분석할 수 있다. 약 100 개 정도의 디코이 단어와 같이, 더 적은 수의 디코이 단어를 갖는 것은 프로세스가 더 빠르게 실행되는 것을 돕는다. 이와 다르게, 디코이 구문은 일반 음성 모델을 사용함에 따라 폐기될 수 있고, 이것은 트리거 단어를 검출함에 있어 적당한 정확도로 기능할 것이다. 구문을 제거하는 것은 메모리 소비를 감소시키지만, 또한 정확도를 떨어뜨린다. DSP(또는 2차 프로세싱 유닛)를 프로그래밍하는 것은 소정의 전자 디바이스의 세부적인 하드웨어 및 설정 측면에 의존적일 수 있다. 예를 들어, 모바일 폰에서 동작하는 음성 인식 관리부는 태블릿 컴퓨터, 데스크톱 컴퓨터, 원격 제어, 장치, 차량 등의 실시예와 상이한 설정을 가질 수 있다.

[0034] 도 5는 본 발명에 따른 컴퓨터/네트워크 환경에서 동작하는 음성 인식 관리부(140)의 예시적인 블록도이다. 도 5에 도시된 컴퓨터 시스템 하드웨어는 후술하는 순서도에 대한 설명에서 더욱 상세하게 설명될 것이다.

[0035] 음성 인식 관리부(140)와 연관된 기능은 이제부터 도 2 내지 도 4의 순서도 및 도면을 통해 설명될 것이다. 이어지는 논의를 위해, 음성 인식 관리부(140) 또는 기타 적절한 주체가 순서도의 단계를 수행한다.

[0036] 이제부터 실시예들을 더욱 상세하게 설명하면, 도 2는 본 발명의 실시예들을 설명하기 위한 순서도이다. 단계 210에서, 음성 인식 관리부는 전자 디바이스의 음성 동작 웨이크업 기능에서 설정 입력을 수신한다. 설정 입력은 트리거 구문을 포함한다. 다시 말해서, 사용자는 웨이크업 커스터마이징 메뉴를 접근하여 구체적인 구문(단어 또는 단어의 그룹)을 설정하고, 그 구문을 타이핑하거나, 아니면 커스텀 구문을 선택한다. 설정 입력은 전자 디바이스의 제1 프로세서에서 수신(로 송신)된다. 전자 디바이스는 제1 프로세서에 더해 제2 프로세서도 포함한다.

[0037] 단계 220에서, 음성 인식 관리부는, 트리거 구분에 기초하여, 트리거 구문에 대응하는 음성 인식 상태의 네트워크를 생성한다. 음성 인식 상태의 네트워크는 제1 프로세서가 실행하는 제1 음성 인식 엔진을 사용하여 제1 프로세서에서 생성된다.

[0038] 단계 230에서, 음성 인식 관리부는 음성 인식 상태의 네트워크를 제1 프로세서로부터 제2 프로세서로 전달한다. 즉, 음성 인식 상태의 네트워크는 제2 프로세서에 있는 스토리지 또는 제2 프로세서에 대해 접근 가능한 스토리지로 전달된다.

[0039] 단계 240에서, 음성 인식 관리부는, 트리거 구문에 대응하는 음성 인식 상태의 네트워크를 사용하여 제2 프로세서 상에서 제2 음성 인식 엔진을 실행한다. 제2 프로세서는, 제1 프로세서의 제1 음성 인식 엔진이 비동작 상태인 동안 제2 음성 인식 엔진을 실행한다. 제1 프로세서 또는 CPU는 동작 상태일 수 있지만, 그럼에도 불구하고 제1 음성 인식 엔진은 상대적으로 비동작이거나 음성 명령 태스크에 응답하지 않을 수 있음을 유의해야 한다. 따라서, 전자 디바이스가 대기 상태이거나 (내용을 보거나 메시지를 체크하는 등과 같이) 동작하여 사용되는 경우, 전자 디바이스의 음성 명령 모드는 실행할 태스크를 능동적으로 청취하지 않을 수 있다. 다른 실시예에서, CPU는 완전히 비동작 상태일 필요는 없지만, 전자 디바이스와 상호작용을 하는 사용자에 의해서와 같이, 동작 상태에서의 전력 소비량에 비해 낮은 전력 모드에서 동작할 수 있다.

[0040] 도 3 및 도 4는 본 발명의 음성 인식 관리부(140)의 추가적 및/또는 대체적인 실시예와 선택적인 기능을 설명하기 위한 순서도이다. 단계 210에서, 음성 인식 관리부는 전자 디바이스의 음성 동작 웨이크업 기능에서 설정 입력을 수신한다. 설정 입력은 트리거 구문을 포함한다. 설정 입력은 전자 디바이스의 제1 프로세서에서 수신(로 송신)된다. 전자 디바이스는 제1 프로세서에 더해 제2 프로세서도 포함한다.

[0041] 단계 212에서, 음성 인식 관리부는 전자 디바이스의 사용자 인터페이스를 통해 텍스트 입력으로서 트리거 구문을 수신한다. 예를 들어, 사용자는 전자 디바이스를 웨이크업하기 위해 발성하고자 하는 구문을 타이핑할 수 있다.

[0042] 단계 213에서, 음성 인식 관리부는 텍스트 입력을 확인하는 발성음을 수신한다. 텍스트 입력이 충분하면, 음성 인식 관리부는 또한 텍스트 입력의 발성음을 프로세싱하여 정확한 인식을 확인할 수 있다.

[0043] 단계 215에서, 제1 프로세서는 중앙 프로세싱 유닛이고, 제2 프로세서는 디지털 신호 프로세서이다. 단계 216에서, 제1 프로세서가 동작 상태에 있는 경우 제2 프로세서는 제1 프로세서에 비해 더 적은 전력을 사용한다. 저

전력 프로세서 상에서 웨이크업 기능을 동작시킴으로써, 전자 디바이스는 빠른 배터리 소모 없이 트리거 단어를 청취할 수 있다. 단계 217에서, 제1 프로세서 및 제2 프로세서는 모바일 전화 내에 공동 배치된다(co-located).

[0044] 단계 220에서, 음성 인식 관리부는 트리거 구문에 대응하는 음성 인식 상태의 네트워크를 생성한다. 음성 인식 상태의 네트워크는 제1 프로세서가 실행하는 제1 음성 인식 엔진을 사용하여 제1 프로세서에서 생성된다.

[0045] 단계 230에서, 음성 인식 관리부는 음성 인식 상태의 네트워크를 제1 프로세서로부터 제2 프로세서로 전달한다. 단계 232에서, 음성 인식 관리부는 음성 인식 상태를 디지털 신호 프로세서로 전달한다.

[0046] 단계 240에서, 음성 인식 관리부는, 트리거 구문에 대응하는 음성 인식 상태의 네트워크를 사용하여 제2 프로세서 상에서 제2 음성 인식 엔진을 실행한다. 제2 프로세서는, 제1 프로세서의 제1 음성 인식 엔진이 비동작 상태인 동안 제2 음성 인식 엔진을 실행한다.

[0047] 단계 242에서, 제1 음성 인식 엔진이 비동작 상태인 동안 제2 음성 인식 엔진은 제2 프로세서를 사용하여 발성음을 연속적으로 분석한다. 그 후 음성 인식 관리부는, 특정 발성음이 트리거 구문을 포함하는 것을 식별한 것에 대한 응답으로, 제1 프로세서에 신호를 전송하여 제1 음성 인식 엔진을 동작 상태로 복귀시키도록 한다.

[0048] 단계 243에서, 제1 프로세서는 후속의 음성 명령에 대해 응답한다.

[0049] 단계 246에서, 제1 프로세서는 비동작 음성 명령 모드에 있고, 특정 발성음이 트리거 구문을 포함하는 것을 식별한 것에 대한 응답으로, 전자 디바이스가 비동작 음성 명령 모드로부터 동작 음성 명령 모드로 스위칭하도록 한다.

[0050] 단계 248에서, 제2 프로세서는 전자 디바이스가 대기 모드인 동안 음성 동작 웨이크업 기능을 제공하는 제2 음성 인식 엔진을 실행한다.

[0051] 다른 실시예에서, 새로운 상태 시퀀스 및 더 작은 어쿠스틱 모델이, 전자 디바이스에서 생성되는 대신에, 원격 서버에서 생성될 수 있다. 이러한 실시예에서, 전자 디바이스는 새로운 트리거 구문을 서버 또는 클라우드(cloud)에 전송할 수 있다. 새로운 트리거 구문은 전자 디바이스를 통해 텍스트 입력으로서 입력될 수 있다. 그 후 원격 서버는 트리거 구문에 대응하는 음성 인식 상태의 네트워크를 생성한 후, 생성된 상태 시퀀스 및 어쿠스틱 모델을 전자 디바이스로 전송하고, 이들은 2차 프로세서 또는 DSP에 의해 사용될 수 있다.

[0052] 도 6을 계속하여 참조하여 이어지는 논의는, 상술한 바와 같은 음성 인식 관리부(140)와 연관된 기능을 수행하는 방법을 나타내는 기본적인 실시예를 제공한다. 그러나, 음성 인식 관리부(140)를 수행하는 실제 구성은 각각의 응용에 따라 달라질 수 있음을 유의해야 한다. 예를 들어, 컴퓨터 시스템(149)은 본 명세서에서 설명된 프로세싱을 수행하는 하나 또는 다중 컴퓨터를 포함할 수 있다.

[0053] 다른 실시예에서, 컴퓨터 시스템(149)은, 셀룰러 폰, 개인용 컴퓨터 시스템, 데스크톱 컴퓨터, 랩톱, 노트북 또는 노트북 컴퓨터, 메인프레임 컴퓨터 시스템, 핸드헬드 컴퓨터, 워크스테이션, 네크워크 컴퓨터, 라우터, 네트워크 스위치, 브리지, 어플리케이션 서버, 스토리지 디바이스, 카메라, 캠코더, 셋톱박스, 모바일 디바이스, 비디오 게임 콘솔, 핸드헬드 비디오 게임 디바이스와 같은 컨슈머 전자 디바이스 또는 일반적인 임의의 타입의 컴퓨팅 또는 전자 디바이스를 포함하지만 이에 한정되지 않는 임의의 다양한 타입의 디바이스일 수 있다.

[0054] 컴퓨터 시스템(149)는 사용자(136)가 입력 디바이스(135)를 사용하여 동작하도록 하기 위한 그래픽 사용자 인터페이스(133)를 표시하는 디스플레이 모니터(130)에 연결된 것으로 도시되었다. 리포지터리(repository)(138)는 데이터 파일 및 컨텐트를 프로세싱 전후에 모두 저장하기 위해 선택적으로 사용될 수 있다. 입력 디바이스(135)는 키보드, 컴퓨터 마우스, 마이크 등과 같은 하나 이상의 디바이스를 포함할 수 있다.

[0055] 도시된 바와 같이, 본 실시예의 컴퓨터 시스템(149)는 메모리 시스템(141), 프로세서(142), I/O 인터페이스(144) 및 통신 인터페이스(145)를 연결하는 인터커넥트(interconnect)(143)를 포함한다.

[0056] I/O 인터페이스(144)는 컴퓨터 마우스, 키보드, 커서를 이동시키기 위한 선택 툴, 디스플레이 스크린 등을 포함하는 입력 디바이스(135)와 같은 주변 디바이스에 대한 연결성(connectivity)을 제공한다.

[0057] 통신 인터페이스(145)는 컴퓨터 시스템(149)의 음성 인식 관리부(140)가 네트워크를 통해 통신할 수 있도록 하고, 필요한 경우, 본 발명의 다양한 실시예에 따라 뷰를 생성하거나, 컨텐트를 프로세싱하거나, 사용자와 통신하는 등에 필요한 임의의 데이터를 검색할 수 있도록 한다.

[0058] 도시된 바와 같이, 메모리 시스템(141)은 상술한 바 및 아래에서 더 설명되는 것과 같은 기능을 지원하는 음성

인식 관리부(140-1)로 인코딩된다. 음성 인식 관리부(140-1)(및/또는 본 명세서에서 설명된 기타 자원)은 본 명세서에서 설명된 여러 실시예들에 따른 기능의 프로세싱을 지원하는 데이터 및/또는 논리 명령과 같은 소프트웨어 코드로서 구현될 수 있다.

[0059] 일 실시예의 동작 중, 프로세서(142)는, 음성 인식 관리부(140-1)의 논리 명령을 런치(launch), 실행(run, execute), 해석(interpret) 또는 기타 수행을 하기 위해, 인터커넥트(143)의 사용을 통해 메모리 시스템(1410을 접근한다. 음성 인식 관리부(140-1)의 실행은 음성 인식 관리 프로세스(140-2)의 기능을 프로세싱하는 것을 생성한다. 다시 말해서, 음성 인식 관리 프로세스(140-2)는 컴퓨터 시스템(149)의 프로세서(142) 내 또는 프로세서(142) 상의 음성 인식 관리부(140)의 하나 이상의 부분을 나타낸다.

[0060] 본 명세서에서 설명된 바와 같은 방법 동작을 수행하는 음성 인식 관리 프로세스(140-2)에 더해, 본 명세서의 다른 실시예는 음성 인식 관리부(140-1) 자신(즉, 실행되지 않은(un-executed) 또는 수행되지 않는(non-performing) 논리 명령 및/또는 데이터)을 포함함을 유의하여야 한다. 음성 인식 관리부(140-1)은, 플로피 디스크, 하드 디스크, 광학 매체 등과 같은 컴퓨터로 판독 가능한 스토리지 매체를 포함하는 비일시적인(non-transitory), 유형의(tangible) 컴퓨터로 판독 가능한 스토리지 매체 상에 저장될 수 있다. 다른 실시예에 따르면, 음성 인식 관리부(140-1)는 또한, 펌웨어, ROM(read only memory)과 같은 메모리 타입 시스템에, 또는, 본 실시예에서와 같이, 메모리 시스템(141) 내의 실행 가능한 코드로서 저장될 수도 있다.

[0061] 이러한 실시예들에 추가적으로, 본 명세서의 다른 실시예들은 프로세서(142)에서 음성 인식 관리 프로세스(140-2)로서 음성 인식 관리부(140-1)의 실행을 포함함에 유의하여야 한다. 따라서, 해당 기술 분야의 통상의 기술자는 컴퓨터 시스템(149)이, 할당을 제어하고 하드웨어 자원을 사용하는 운영체제 또는 다중 프로세서와 같은, 다른 프로세스 및/또는 소프트웨어 및 하드웨어 부품을 포함할 수 있음을 이해할 것이다.

[0062] 이상 첨부된 도면을 참조하여 본 발명의 실시예를 설명하였지만, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자는 본 발명이 그 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 실시될 수 있다는 것을 이해할 수 있을 것이다. 그러므로 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다.

부호의 설명

[0063] 105: 사용자

110: 중앙 프로세싱 유닛

112: 어쿠스틱 모델

114: 음성 인식 상태

120: 디지털 신호 프로세서

122: DSP 어쿠스틱 모델

125: 동적 음성 인식 상태

130: 디스플레이 모니터

133: 그래픽 사용자 인터페이스

135: 입력 디바이스

136: 사용자

137: 전자 디바이스

138: 리포지터리

140: 음성 인식 관리부

140-1: 음성 인식 관리부 어플리케이션

140-2: 음성 인식 관리부 프로세스

141: 메모리 시스템

142: 프로세서

143: 인터커넥트

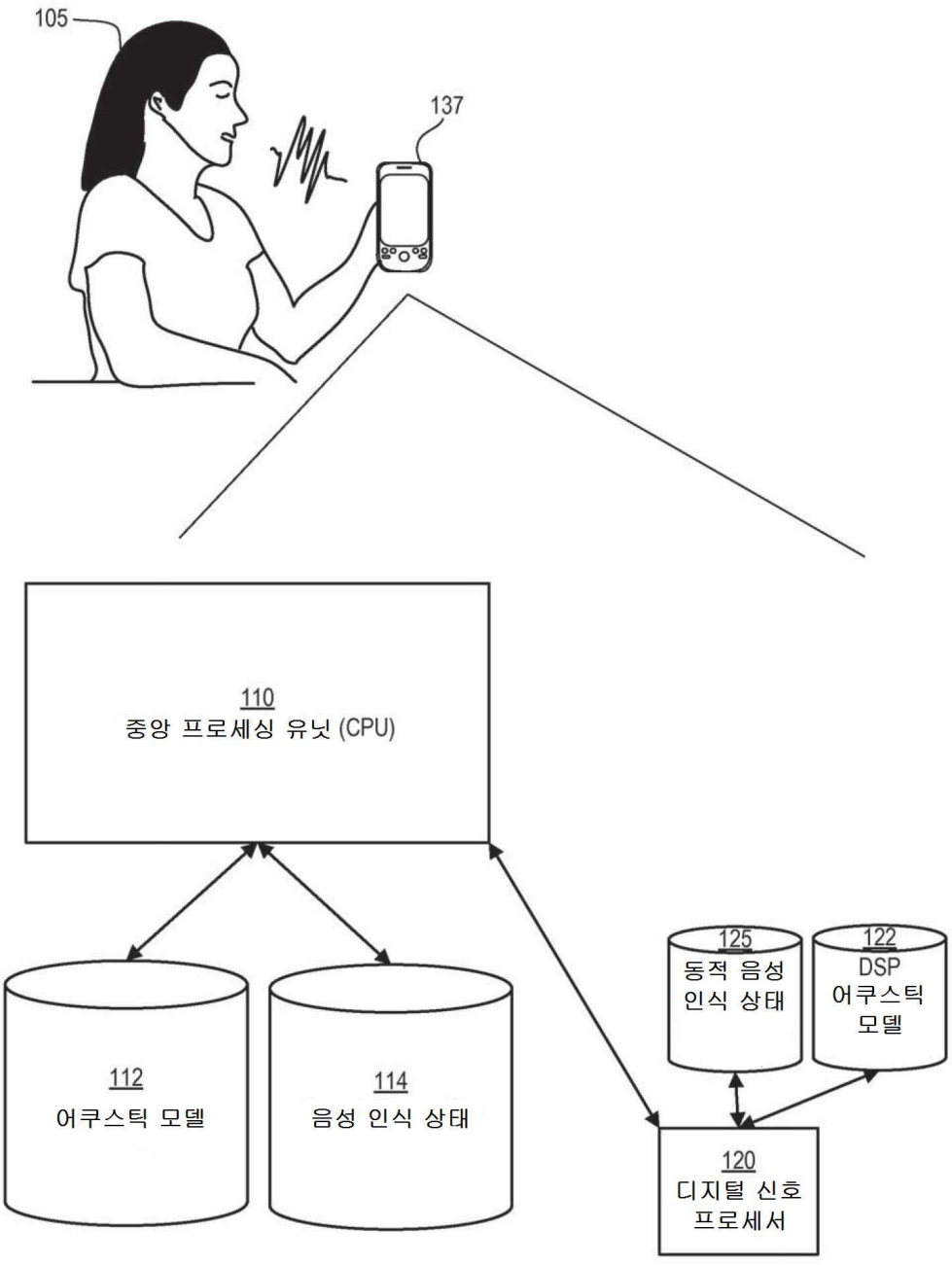
144: I/O 인터페이스

145: 통신 인터페이스

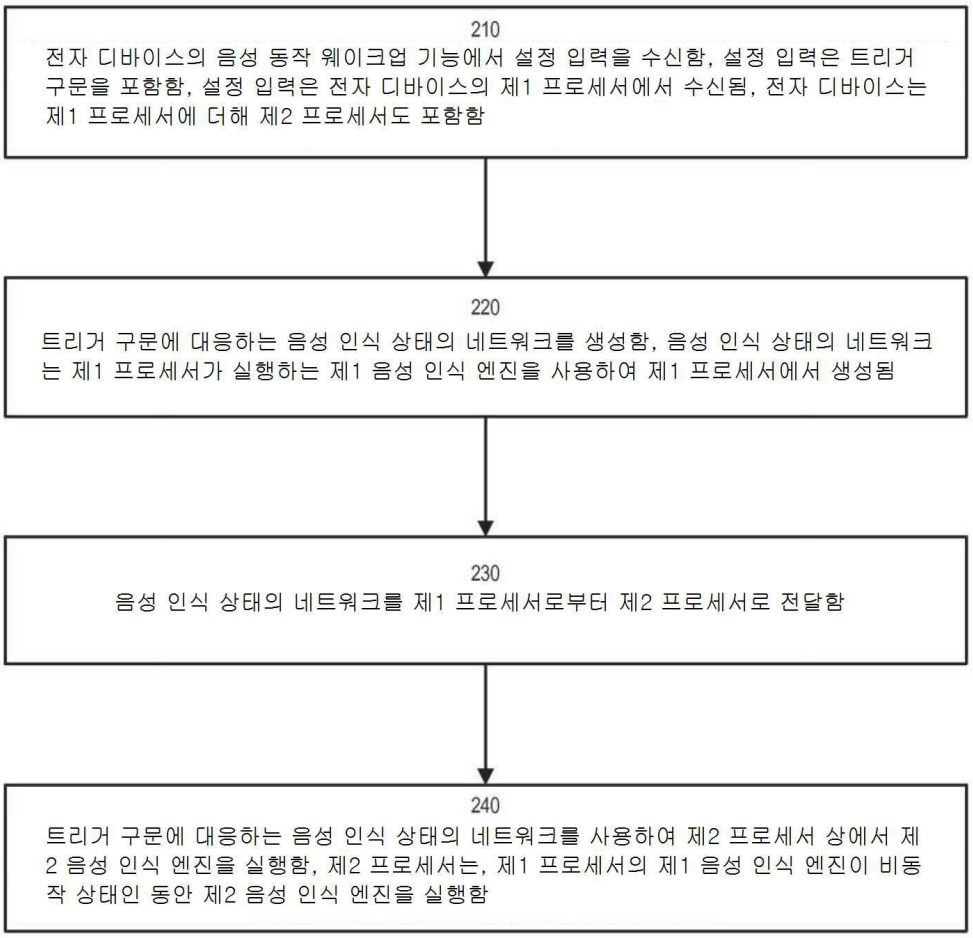
149: 컴퓨터 시스템

도면

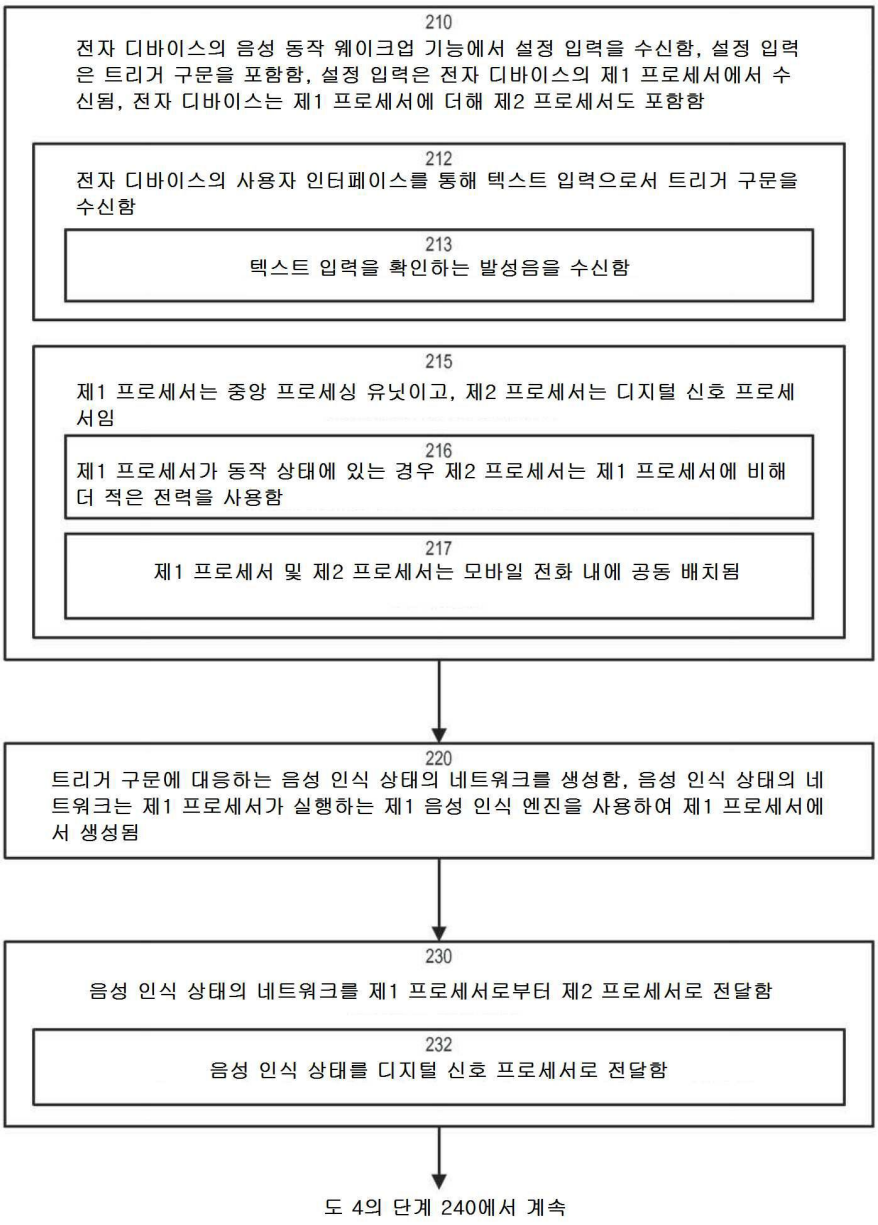
도면1



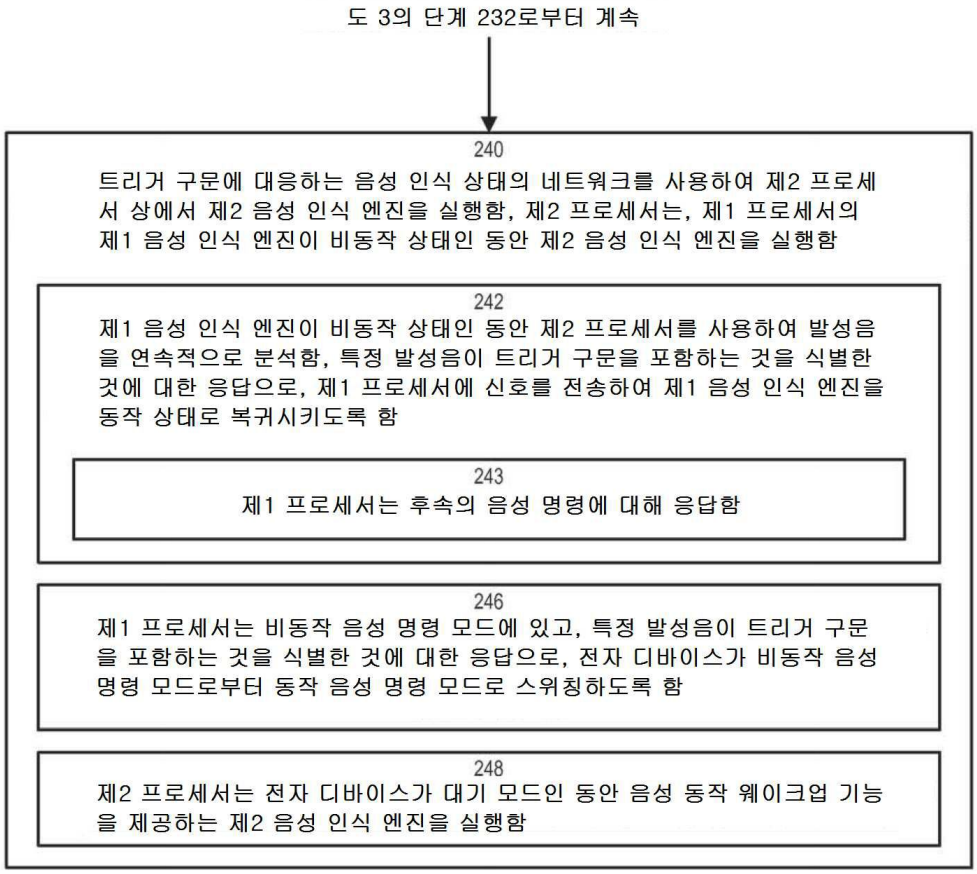
도면2


210
전자 디바이스의 음성 동작 웨이크업 기능에서 설정 입력을 수신함, 설정 입력은 트리거 구문을 포함함, 설정 입력은 전자 디바이스의 제1 프로세서에서 수신됨, 전자 디바이스는 제1 프로세서에 더해 제2 프로세서도 포함함
220
트리거 구문에 대응하는 음성 인식 상태의 네트워크를 생성함, 음성 인식 상태의 네트워크는 제1 프로세서가 실행하는 제1 음성 인식 엔진을 사용하여 제1 프로세서에서 생성됨
230
음성 인식 상태의 네트워크를 제1 프로세서로부터 제2 프로세서로 전달함
240
트리거 구문에 대응하는 음성 인식 상태의 네트워크를 사용하여 제2 프로세서 상에서 제2 음성 인식 엔진을 실행함, 제2 프로세서는, 제1 프로세서의 제1 음성 인식 엔진이 비동작 상태인 동안 제2 음성 인식 엔진을 실행함

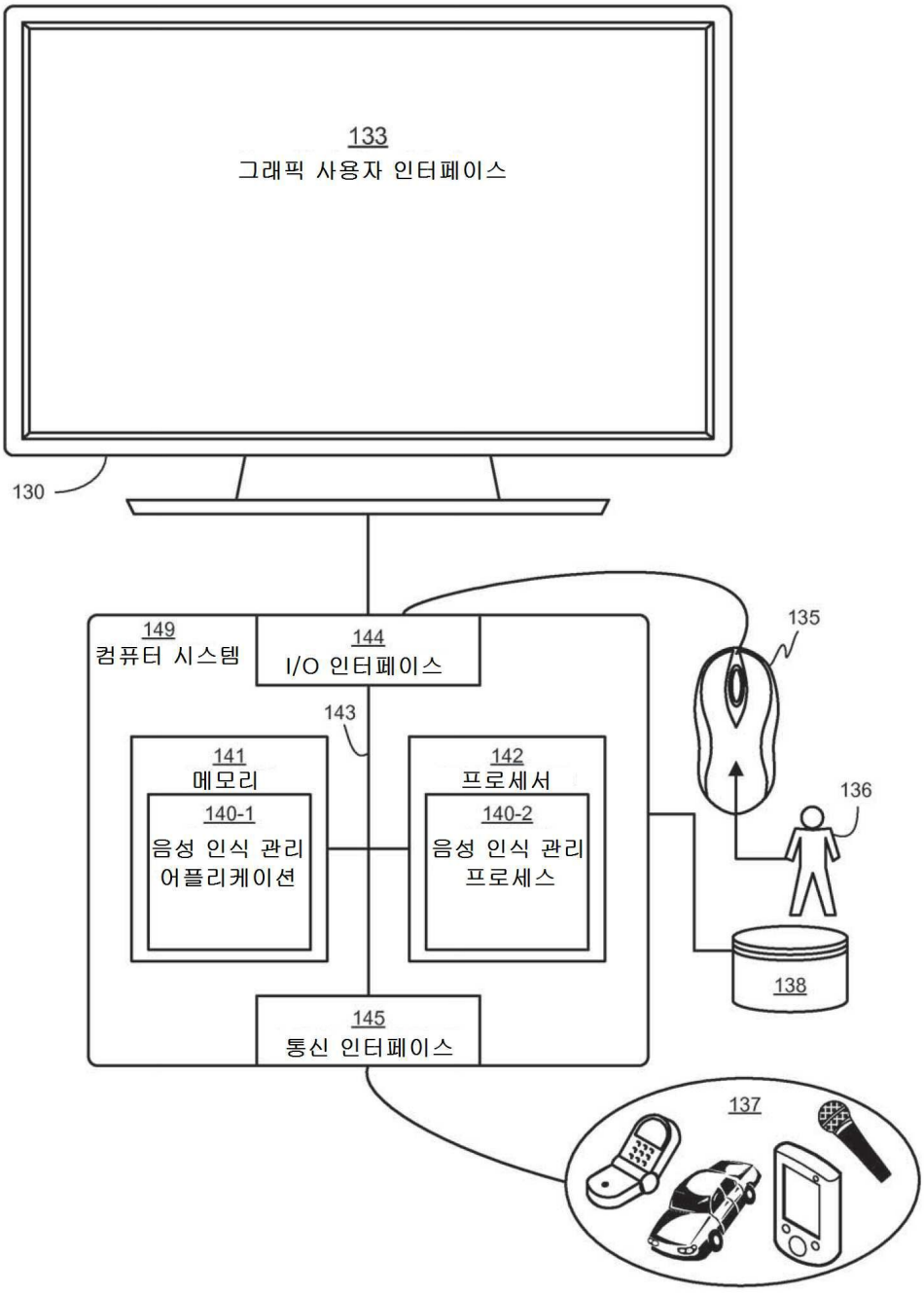
도면3


210
전자 디바이스의 음성 동작 웨이크업 기능에서 설정 입력을 수신함, 설정 입력은 트리거 구문을 포함함, 설정 입력은 전자 디바이스의 제1 프로세서에서 수신됨, 전자 디바이스는 제1 프로세서에 더해 제2 프로세서도 포함함
212
전자 디바이스의 사용자 인터페이스를 통해 텍스트 입력으로서 트리거 구문을 수신함
213
텍스트 입력을 확인하는 발성음을 수신함
215
제1 프로세서는 중앙 프로세싱 유닛이고, 제2 프로세서는 디지털 신호 프로세서임
216
제1 프로세서가 동작 상태에 있는 경우 제2 프로세서는 제1 프로세서에 비해 더 적은 전력을 사용함
217
제1 프로세서 및 제2 프로세서는 모바일 전화 내에 공동 배치됨
220
트리거 구문에 대응하는 음성 인식 상태의 네트워크를 생성함, 음성 인식 상태의 네트워크는 제1 프로세서가 실행하는 제1 음성 인식 엔진을 사용하여 제1 프로세서에서 생성됨
230
음성 인식 상태의 네트워크를 제1 프로세서로부터 제2 프로세서로 전달함
232
음성 인식 상태를 디지털 신호 프로세서로 전달함
도 4의 단계 240에서 계속

도면4


도 3의 단계 232로부터 계속
240
트리거 구문에 대응하는 음성 인식 상태의 네트워크를 사용하여 제2 프로세서 상에서 제2 음성 인식 엔진을 실행함, 제2 프로세서는, 제1 프로세서의 제1 음성 인식 엔진이 비동작 상태인 동안 제2 음성 인식 엔진을 실행함
242
제1 음성 인식 엔진이 비동작 상태인 동안 제2 프로세서를 사용하여 발성음을 연속적으로 분석함, 특정 발성음이 트리거 구문을 포함하는 것을 식별한 것에 대한 응답으로, 제1 프로세서에 신호를 전송하여 제1 음성 인식 엔진을 동작 상태로 복귀시키도록 함
243
제1 프로세서는 후속의 음성 명령에 대해 응답함
246
제1 프로세서는 비동작 음성 명령 모드에 있고, 특정 발성음이 트리거 구문을 포함하는 것을 식별한 것에 대한 응답으로, 전자 디바이스가 비동작 음성 명령 모드로부터 동작 음성 명령 모드로 스위칭하도록 함
248
제2 프로세서는 전자 디바이스가 대기 모드인 동안 음성 동작 웨이크업 기능을 제공하는 제2 음성 인식 엔진을 실행함

도면5


133
그래픽 사용자 인터페이스
130
149
컴퓨터 시스템
144
I/O 인터페이스
143
141
메모리
140-1
음성 인식 관리
어플리케이션
142
프로세서
140-2
음성 인식 관리
프로세스
145
통신 인터페이스
135
136
138
137