



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2017-0033398
(43) 공개일자 2017년03월24일

(51) 국제특허분류(Int. Cl.)
G10L 15/16 (2006.01) G10L 15/04 (2006.01)
(52) CPC특허분류
G10L 15/16 (2013.01)
G10L 15/02 (2013.01)
(21) 출원번호 10-2017-7004672
(22) 출원일자(국제) 2015년08월18일
심사청구일자 2017년02월20일
(85) 번역문제출일자 2017년02월20일
(86) 국제출원번호 PCT/US2015/045750
(87) 국제공개번호 WO 2016/048486
국제공개일자 2016년03월31일
(30) 우선권주장
14/493,434 2014년09월23일 미국(US)

(71) 출원인
인텔 코퍼레이션
미합중국 캘리포니아 95054 산타클라라 미션 칼리
지 블러바드 2200
(72) 발명자
바우어 요셉
독일 비와이 80797 뮌헨 빈저러스트라세 96
로젠 피오토르
폴란드 피엠 피엘-80-034 그단스크 애니 자지엘론
키 36/17
스테머 조지
독일 비와이 80797 뮌헨 슐레이스하이머 스트라세
59아
(74) 대리인
제일특허법인

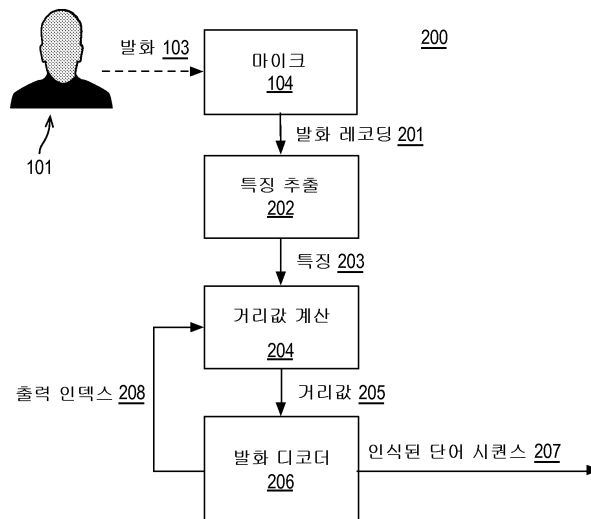
전체 청구항 수 : 총 25 항

(54) 발명의 명칭 자동 발화 인식을 위한 외삽 이용 프레임 스키핑 및 수요 출력 신경망

(57) 요약

발화 인식 시스템을 위해 신경망을 구현하는 것에 관련된 기법이 논의된다. 그러한 기법은 발화 디코더에 의해 필요로 되는 출력만이 신경망 또는 근사계산 기법을 통하여 제공되도록, 근사계산된 스킵 프레임을 이용한 프레임 스키핑 및/또는 수요 거리를 구현하는 것을 포함할 수 있다.

대표도 - 도2



(52) CPC특허분류
G10L 15/04 (2013.01)

명세서

청구범위

청구항 1

자동 발화 인식(automatic speech recognition)을 제공하는 컴퓨터로 구현된(computer-implemented) 방법으로서,

제1 시간 인스턴스(time instance)와 연관된 적어도 하나의 제1 거리값(distance value)을 판정하기 위해 상기 제1 시간 인스턴스에 대해 신경망(neural network)을 평가하는(evaluating) 단계 - 상기 적어도 하나의 제1 거리값은 상기 신경망으로부터의 출력을 포함함 - 와,

제2 시간 인스턴스에 대해, 상기 제1 거리값에 적어도 부분적으로 기반하여 적어도 하나의 제2 거리값을 근사계산하는(approximating) 단계 - 상기 신경망은 상기 제2 시간 인스턴스에 대해 평가되지 않음 - 와,

상기 제1 거리값 및 상기 제2 거리값에 적어도 부분적으로 기반하여 텍스트 요소(textual element)의 시퀀스(sequence)를 판정하는 단계를 포함하는

방법.

청구항 2

제1항에 있어서,

상기 제1 시간 인스턴스에 대해 하나 이상의 출력 인덱스를 생성하는 단계를 더 포함하되, 상기 제1 거리값은 상기 출력 인덱스 중의 제1 출력 인덱스와 연관되고, 상기 신경망은 복수의 출력층 노드를 가지는 출력층을 포함하며, 상기 제1 시간 인스턴스에 대해 상기 신경망을 평가하는 단계는 상기 출력 인덱스와 연관된 상기 복수의 출력층 노드의 서브세트(subset)만을 평가하는 단계를 포함하는

방법.

청구항 3

제2항에 있어서,

상기 신경망은 최종 은닉층 노드를 가지는 최종 은닉층을 더 포함하고, 상기 제1 시간 인스턴스에 대해 상기 신경망을 평가하는 단계는 상기 최종 은닉층 노드 전부를 평가하는 단계를 포함하는

방법.

청구항 4

제1항에 있어서,

상기 제2 거리값을 근사계산하는 단계는 선형 함수(linear function), 비선형 함수(non-linear function), 또는 분산 함수(variance function) 중 적어도 하나를 기반으로 상기 제1 거리값에 적어도 부분적으로 기반하여 상기 제2 거리값을 외삽하는(extrapolating) 단계를 포함하는

방법.

청구항 5

제1항에 있어서,

상기 제2 거리값을 근사계산하는 단계는 상기 제1 거리값 및 제3 시간 인스턴스와 연관된 제3 거리값에 기반하여 상기 제2 거리값을 외삽하는 단계를 포함하되, 상기 제1 시간 인스턴스는 상기 제2 시간 인스턴스 이전이고 상기 제3 시간 인스턴스는 상기 제1 시간 인스턴스 이전인

방법.

청구항 6

제5항에 있어서,

상기 제2 거리값을 외삽하는 단계는 상기 제1 거리값 및 상기 제3 거리값에 기반하여 선형 함수를 통하여 상기 제2 거리값을 외삽하는 단계를 포함하되, 상기 제3 거리값은 상기 신경망의 평가에 기반하여 판정되는

방법.

청구항 7

제6항에 있어서,

상기 선형 함수는 상기 제1 거리값과 상기 제3 거리값 사이의 차이의 절반에 상기 제1 거리값이 더해진 것을 포함하는

방법.

청구항 8

제1항에 있어서,

상기 제2 시간 인스턴스에 대해 하나 이상의 출력 인덱스를 생성하는 단계 - 상기 제2 시간 인스턴스에 대한 제3 거리값이 상기 출력 인덱스 중의 제1 출력 인덱스와 연관되고, 상기 신경망은 복수의 출력층 노드를 가지는 출력층 및 복수의 최종 은닉층 노드를 가지는 최종 은닉층을 포함하며, 상기 제1 시간 인스턴스에 대해 상기 신경망을 평가하는 단계는 복수의 최종 은닉층 값을 생성하기 위해 상기 최종 은닉층의 모든 최종 은닉층 노드를 평가하는 단계를 포함함 - 와,

상기 제2 시간 인스턴스에서, 상기 제1 시간 인스턴스에 대한 제4 거리값을, 상기 최종 은닉층 값에 기반하여 상기 제4 거리값과 연관된 상기 복수의 출력층 노드 중의 출력층 노드를 평가함으로써 판정하는 단계와,

상기 제2 시간 인스턴스에 대해, 상기 제3 거리값을 상기 제4 거리값에 적어도 부분적으로 기반하여 근사계산하는 단계를 더 포함하는

방법.

청구항 9

제1항에 있어서,

상기 신경망은 복수의 출력층 노드를 가지는 출력층을 포함하고, 상기 제1 시간 인스턴스에 대해 상기 신경망을 평가하는 단계는 상기 복수의 출력층 노드 전부를 평가하는 단계를 포함하는

방법.

청구항 10

제1항에 있어서,

상기 제1 시간 인스턴스는 신경망 평가 프레임(evaluation frame)과 연관되고, 상기 제2 시간 인스턴스는 스킵 프레임(skip frame)과 연관되며, 1개, 2개 또는 3개의 추가적인 스킵 프레임이 상기 평가 프레임과 상기 스킵 프레임 사이에 있는

방법.

청구항 11

제1항에 있어서,

상기 제1 시간 인스턴스는 신경망 평가 프레임과 연관되고, 상기 제2 시간 인스턴스는 스킵 프레임과 연관되며, 상기 방법은 가용 컴퓨팅 리소스(available computing resource) 또는 현재의 실시간 지수(current real time factor) 중 적어도 하나에 기반하여 프레임 스킵핑 레이트(frame skipping rate)를 판정하는 단계와,

상기 프레임 스킵핑 레이트에 기반하여 상기 평가 프레임과 상기 스킵 프레임 사이에 추가적인 스킵 프레임을 제공하는 단계를 더 포함하는

방법.

청구항 12

제1항에 있어서,

텍스트 요소의 상기 시퀀스를 판정하는 단계는 비터비 빔 탐색 디코더(Viterbi beam searching decoder)를 통하여 텍스트 요소의 상기 시퀀스를 판정하는 단계를 포함하는

방법.

청구항 13

제1항에 있어서,

수신된 발화(received speech)를 발화 레코딩(speech recording)으로 변환하는 단계와,

상기 발화 레코딩의 시간 윈도우와 연관된 특징 벡터를 추출하는 단계와,

상기 신경망으로의 입력으로서 상기 특징 벡터를 제공하는 단계를 더 포함하는

방법.

청구항 14

자동 발화 인식을 제공하는 시스템으로서,

발화 인식 데이터를 저장하도록 구성된 메모리와,

상기 메모리에 커플링된 중앙 처리 유닛을 포함하되, 상기 중앙 처리 유닛은

제1 시간 인스턴스와 연관된 적어도 하나의 제1 거리값을 판정하기 위해 상기 제1 시간 인스턴스에 대해 신경망을 구현하도록 구성된 신경망 회로와,

제2 시간 인스턴스에 대해, 상기 제1 거리값에 적어도 부분적으로 기반하여 적어도 하나의 제2 거리값을 근사계산하도록 구성된 거리값 근사계산 회로와,

상기 제1 거리값 및 상기 제2 거리값에 적어도 부분적으로 기반하여 텍스트 요소의 시퀀스를 판정하도록 구성된

발화 디코더 회로를 포함하는
시스템.

청구항 15

제14항에 있어서,

상기 발화 디코더 회로는 상기 제1 시간 인스턴스에 대해 하나 이상의 출력 인덱스를 생성하도록 더 구성되되,
상기 제1 거리값은 상기 출력 인덱스 중의 제1 출력 인덱스와 연관되고, 상기 신경망은 복수의 출력층 노드를
가지는 출력층을 포함하며, 상기 신경망 회로는 상기 제1 시간 인스턴스에 대해 상기 출력 인덱스와 연관된 상
기 복수의 출력층 노드의 서브세트만을 평가하도록 구성된
시스템.

청구항 16

제15항에 있어서,

상기 신경망은 최종 은닉층 노드를 가지는 최종 은닉층을 더 포함하고, 상기 신경망 회로는 상기 제1 시간 인스
턴스에 대해 상기 최종 은닉층 노드 전부를 평가하도록 구성된
시스템.

청구항 17

제14항에 있어서,

상기 거리값 근사계산 회로가 상기 제2 거리값을 근사계산하도록 구성되는 것은 상기 거리값 근사계산 회로가
상기 제1 거리값 및 제3 시간 인스턴스와 연관된 제3 거리값에 기반하여 선형 함수를 통하여 상기 제2 거리값을
외삽하도록 구성되는 것을 포함하되, 상기 제1 시간 인스턴스는 상기 제2 시간 인스턴스 이전이고 상기 제3 시
간 인스턴스는 상기 제1 시간 인스턴스 이전이며, 상기 제3 거리값은 상기 신경망의 평가에 기반하여 판정되는
시스템.

청구항 18

제14항에 있어서,

상기 발화 디코더 회로는 상기 제2 시간 인스턴스에 대해 하나 이상의 출력 인덱스를 생성하도록 더 구성되되,
상기 제2 시간 인스턴스에 대한 제3 거리값이 상기 출력 인덱스 중의 제1 출력 인덱스와 연관되고, 상기 신경망
은 복수의 출력층 노드를 가지는 출력층 및 복수의 최종 은닉층 노드를 가지는 최종 은닉층을 포함하며, 상기
신경망 회로는 복수의 최종 은닉층 값을 생성하기 위해 상기 제1 시간 인스턴스에 대해 모든 최종 은닉층 노드
를 평가하도록 구성되되, 상기 신경망 회로는 상기 제2 시간 인스턴스에서, 상기 제1 시간 인스턴스에 대한 제4
거리값을, 상기 최종 은닉층 값에 기반하여 상기 제4 거리값과 연관된 상기 복수의 출력층 노드 중의 출력층 노
드를 평가함으로써 판정하도록 더 구성되고, 상기 거리값 근사계산 회로는 상기 제2 시간 인스턴스에 대해, 상
기 제3 거리값을 상기 제4 거리값에 적어도 부분적으로 기반하여 근사계산하도록 구성된
시스템.

청구항 19

자동 발화 인식을 제공하는 시스템으로서,

제1 시간 인스턴스와 연관된 적어도 하나의 제1 거리값을 판정하기 위해 상기 제1 시간 인스턴스에 대해 신경망을 평가하는 수단 - 상기 적어도 하나의 제1 거리값은 상기 신경망으로부터의 출력을 포함함 - 과,

제2 시간 인스턴스에 대해, 상기 제1 거리값에 적어도 부분적으로 기반하여 적어도 하나의 제2 거리값을 근사계산하는 수단 - 상기 신경망은 상기 제2 시간 인스턴스에 대해 평가되지 않음 - 과,

상기 제1 거리값 및 상기 제2 거리값에 적어도 부분적으로 기반하여 텍스트 요소의 시퀀스를 판정하는 수단을 포함하는

시스템.

청구항 20

제19항에 있어서,

상기 제1 시간 인스턴스에 대해 하나 이상의 출력 인덱스를 생성하는 수단을 더 포함하되, 상기 제1 거리값은 상기 출력 인덱스 중의 제1 출력 인덱스와 연관되고, 상기 신경망은 복수의 출력층 노드를 가지는 출력층 및 최종 은닉층 노드를 가지는 최종 은닉층을 포함하며, 상기 제1 시간 인스턴스에 대해 상기 신경망을 평가하는 것은 상기 출력 인덱스와 연관된 상기 복수의 출력층 노드의 서브세트만을 평가하는 것 및 상기 최종 은닉층 노드 전부를 평가하는 것을 포함하는

시스템.

청구항 21

제19항에 있어서,

상기 제2 거리값을 근사계산하는 것은 상기 제1 거리값 및 제3 시간 인스턴스와 연관된 제3 거리값에 기반하여 선형 함수를 통하여 상기 제2 거리값을 외삽하는 것을 포함하되, 상기 제1 시간 인스턴스는 상기 제2 시간 인스턴스 이전이고 상기 제3 시간 인스턴스는 상기 제1 시간 인스턴스 이전이며, 상기 제3 거리값은 상기 신경망의 평가에 기반하여 판정되는

시스템.

청구항 22

컴퓨팅 디바이스 상에서 실행되는 것에 응답하여 상기 컴퓨팅 디바이스로 하여금,

제1 시간 인스턴스와 연관된 적어도 하나의 제1 거리값을 판정하기 위해 상기 제1 시간 인스턴스에 대해 신경망을 평가하는 것 - 상기 적어도 하나의 제1 거리값은 상기 신경망으로부터의 출력을 포함함 - 과,

제2 시간 인스턴스에 대해, 상기 제1 거리값에 적어도 부분적으로 기반하여 적어도 하나의 제2 거리값을 근사계산하는 것 - 상기 신경망은 상기 제2 시간 인스턴스에 대해 평가되지 않음 - 과,

상기 제1 거리값 및 상기 제2 거리값에 적어도 부분적으로 기반하여 텍스트 요소의 시퀀스를 판정하는 것에 의해 자동 발화 인식을 제공하게 하는 복수의 명령어를 포함하는

적어도 하나의 머신 판독가능 매체.

청구항 23

제22항에 있어서,

상기 컴퓨팅 디바이스 상에서 실행되는 것에 응답하여 상기 컴퓨팅 디바이스로 하여금 상기 제1 시간 인스턴스에 대해 하나 이상의 출력 인덱스를 생성하는 것에 의해 발화 인식을 수행하게 하는 명령어를 더 포함하되, 상

기 제1 거리값은 상기 출력 인덱스 중의 제1 출력 인덱스와 연관되고, 상기 신경망은 복수의 출력층 노드를 가지는 출력층을 포함하며, 상기 제1 시간 인스턴스에 대해 상기 신경망을 평가하는 것은 상기 출력 인덱스와 연관된 상기 복수의 출력층 노드의 서브세트만을 평가하는 것을 포함하는
머신 판독가능 매체.

청구항 24

제22항에 있어서,

상기 제2 거리값을 근사계산하는 것은 상기 제1 거리값 및 제3 시간 인스턴스와 연관된 제3 거리값에 기반하여 선형 함수를 통하여 상기 제2 거리값을 외삽하는 것을 포함하되, 상기 제1 시간 인스턴스는 상기 제2 시간 인스턴스 이전이고 상기 제3 시간 인스턴스는 상기 제1 시간 인스턴스 이전이되, 상기 제3 거리값은 상기 신경망의 평가에 기반하여 판정되는

머신 판독가능 매체.

청구항 25

제22항에 있어서,

상기 컴퓨팅 디바이스 상에서 실행되는 것에 응답하여 상기 컴퓨팅 디바이스로 하여금,

상기 제2 시간 인스턴스에 대해 하나 이상의 출력 인덱스를 생성하는 것 - 상기 제2 시간 인스턴스에 대한 제3 거리값이 상기 출력 인덱스 중의 제1 출력 인덱스와 연관되고, 상기 신경망은 복수의 출력층 노드를 가지는 출력층 및 복수의 최종 은닉층 노드를 가지는 최종 은닉층을 포함하며, 상기 제1 시간 인스턴스에 대해 상기 신경망을 평가하는 것은 복수의 최종 은닉층 값을 생성하기 위해 상기 최종 은닉층의 모든 최종 은닉층 노드를 평가하는 것을 포함함 - 과,

상기 제2 시간 인스턴스에서, 상기 제1 시간 인스턴스에 대한 제4 거리값을, 상기 최종 은닉층 값에 기반하여 상기 제4 거리값과 연관된 상기 복수의 출력층 노드 중의 출력층 노드를 평가함으로써 판정하는 것과,

상기 제2 시간 인스턴스에 대해, 상기 제4 거리값에 적어도 부분적으로 기반하여 상기 제3 거리값을 근사계산하는 것에 의해 발화 인식을 수행하게 하는 명령어를 더 포함하는

머신 판독가능 매체.

발명의 설명

기술 분야

우선권 주장

이 출원은, 이로써 참조에 의해 전체로서 포함되는, "FRAME SKIPPING WITH EXTRAPOLATION AND OUTPUTS ON DEMAND NEURAL NETWORK FOR AUTOMATIC SPEECH RECOGNITION"라는 표제로 2014년 11월 23일 출원된 미국 특허 출원 제14/493,434호에 대한 우선권을 주장한다.

배경 기술

머신 러닝(machine learning)과 지각 및 인지 시스템을 위해 딥 신경망(deep neural network)을 비롯한 신경망이 사용될 수 있다. 그러한 신경망은 발화 인식(speech recognition) 시스템과 같은 다양한 구현에서 사용될 수 있다. 예컨대, 신경망은 뉴런(neuron) 또는 노드(node)의 상호연결된 층(interconnected layer)을 포함할 수 있다. 각각의 층을 위한 입력값(input value)은 (가령, 입력층(input layer)에서의) 시스템으로의 입력 또는 신경망 내의 앞의 층으로부터의 출력을 포함한다. 신경망의 출력층(output layer)으로부터의 출력값(output value)은 출력값, 거리값(distance value) 또는 분류값(classification value) 또는 유사한 것을 포함할 수 있으니 신경망으로의 입력은 신경망 및/또는 추가적인 처리를 통하여 분류될 수 있다. 그러한 신경망 처리 및 분류는 더한 표준적인 규칙 기반 처리 시스템으로써 수행하기가 곤란하거나 불가능할 수 있는 분류 또는 다른 작

업을 수행하는 데에 사용될 수 있다.

[0004] 논의된 바와 같이, 그러한 신경망은 자동 발화 인식(Automatic Speech Recognition: ASR) 시스템 내에 구현될 수 있고, 몇몇 구현에서, 그것은 그러한 시스템의 가장 중요한 컴포넌트일 수 있다. 실시간(real-time) 애플리케이션 내의 현재의 신경망에 있어서의 문제는 신경망을 평가하는(evaluate) 데에 필요한 대규모 계산작업(computational effort)이다. 이 문제를 다루기 위하여, 몇몇 현재의 구현은 신경망 계산을 디바이스의 중앙 처리 유닛(Central Processing Unit: CPU)으로부터 디바이스의 그래픽 처리 유닛(Graphics Processing Unit: GPU)으로 분담시킨다(offload). 그러나, 그러한 분담은 디바이스 상에서 구동되는 게임과 같은 다른 GPU 집약적(GPU-intensive) 애플리케이션 또는 디바이스의 카메라 또는 유사한 것과의 충돌을 유발할 수 있다. 나아가, 디바이스의 GPU의 집약적 사용은 전력 사용을 증가시키고 이로써 모바일 디바이스를 위한 배터리 수명을 제한할 수 있다. 다른 구현에서, 단일 명령어 다중 데이터(Single Instruction, Multiple Data: SIMD) 플랫폼 및/또는 최적화, 예를 들어 일괄 해태 평가 모델(batched lazy evaluation model)(이는 필요할 때까지 계산을 지연할 수 있음)가 사용될 수 있다. 그러나, 그러한 구현은 감소된 분류 정확도를 가질 수 있다.

[0005] 이와 같이, 기존의 기법은 실시간의 효율적이고 정확한 신경망 구현을 제공하지 않는다. 그러한 문제는, 예를 들어 발화 인식에서, 신경망을 통하여 분류를 활용하려는 요망이 더 널리 퍼지게 됨에 따라 중대하게 될 수 있다.

도면의 간단한 설명

[0006] 본 문서에 기술된 제재(material)는 첨부된 도면에서 한정으로서가 아니고 예로서 보여진다. 예시의 단순성 및 명확성을 위하여, 도면 내에 예시된 요소는 반드시 축척에 맞게 그려지지는 않는다. 예컨대, 몇몇 요소의 치수는 명확성을 위해 다른 요소에 비해 과장될 수 있다. 또한, 적절하다고 간주되는 경우, 대응하거나 비슷한 요소를 나타내기 위해 도면 중에서 참조 라벨이 반복되었다. 도면에서,

도 1은 발화 인식을 제공하기 위한 예시적 세팅(setting)의 설명도(illustrative diagram)이고,

도 2는 발화 인식을 제공하기 위한 예시적 시스템의 설명도이며,

도 3은 예시적 거리값 계산 모듈(distance values computation module)의 설명도이고,

도 4는 예시적 신경망의 설명도이며,

도 5는 평가 및 스킵(skip) 프레임의 예시적 시간선(timeline)을 보여주고,

도 6은 프레임 스킵핑(frame skipping) 및 거리값 근사계산(distance value approximation)을 사용하여 거리값을 판정하기 위한 예시적 프로세스를 보여주는 흐름도(flow diagram)이며,

도 7은 예시적 거리값 계산 모듈의 설명도이고,

도 8은 예시적 신경망의 설명도이며,

도 9는 프레임 스킵핑, 수요 거리(distance on demand) 및 거리값 근사계산을 사용하여 거리값을 판정하기 위한 예시적 프로세스를 보여주는 흐름도이고,

도 10은 프레임 스킵핑 및/또는 수요 거리를 구현하기 위한 예시적 시스템의 설명도이며,

도 11은 자동 발화 인식을 제공하기 위한 예시적 프로세스를 보여주는 흐름도이고,

도 12는 발화 인식을 제공하기 위한 예시적 시스템의 설명도이고,

도 13은 예시적 시스템의 설명도이며,

도 14는 온전히 본 개시의 적어도 몇몇 구현에 따라 마련된 예시적 디바이스를 보여준다

발명을 실시하기 위한 구체적인 내용

[0007] 동봉된 도면을 참조하여 하나 이상의 실시예 또는 구현이 이제 기술된다. 특정 구성 및 조성이 논의되나, 이것은 단지 예시적인 목적으로 행해짐이 이해되어야 한다. 관련 업계에서 숙련된 자는 설명의 사상 및 범주로부터 벗어나지 않고서 다른 구성 및 조성이 이용될 수 있음을 인식할 것이다. 본 문서에 기술된 기법 및/또는 조성은 본 문서에 기술된 것이 아닌 다양한 다른 시스템 및 응용에서 또한 이용될 수 있음은 관련 업계에서 숙련된 이

에게 명백할 것이다.

[0008] 이하의 설명은 예컨대 시스템 온 칩(System-on-a-Chip: SoC) 아키텍처와 같은 아키텍처 내에 현출될 수 있는 다양한 구현을 개진하나, 본 문서에 기술된 기법 및/또는 조성의 구현은 특정한 아키텍처 및/또는 컴퓨팅 시스템에 제한되지 않으며 유사한 목적으로 임의의 아키텍처 및/또는 컴퓨팅 시스템에 의해 구현될 수 있다. 예를 들면, 셋톱 박스(set top box), 스마트폰(smart phone) 등등과 같은 다양한 컴퓨팅 디바이스 및/또는 가전(Consumer Electronic: CE) 디바이스, 그리고/또는 예컨대 집적 회로(Integrated Circuit: IC) 칩 및/또는 패키지를 이용하는 다양한 아키텍처가 본 문서에 기술된 기법 및/또는 조성을 구현할 수 있다. 또한, 이하의 설명은 로직(logic) 구현, 시스템 컴포넌트의 유형 및 상호관계, 로직 분리(partitioning)/통합(integration) 선택 등등과 같은 다수의 특정 세부사항을 개진할 수 있으나, 청구된 대상물(claimed subject matter)은 그러한 특정 세부사항 없이 실시될 수 있다. 다른 사례에서, 예컨대 제어 구조(control structure) 및 온전한 소프트웨어 명령어 시퀀스(full software instruction sequence)와 같은 어떤 제재는 본 문서에 개시된 제재를 모호하게 하지 않기 위해서 상세히 보여지지 않을 수 있다.

[0009] 본 문서에 개시된 제재는 하드웨어, 펌웨어, 소프트웨어, 또는 이의 임의의 조합으로 구현될 수 있다. 본 문서에 개시된 제재는 머신 판독가능 매체(machine-readable medium) 상에 저장된 명령어로서 또한 구현될 수 있는데, 이는 하나 이상의 프로세서에 의해 판독되고 실행될 수 있다. 머신 판독가능 매체는 머신(가령, 컴퓨팅 디바이스(computing device))에 의해 판독가능한 형태로 정보를 저장하거나 송신하기 위한 임의의 매체 및/또는 메커니즘을 포함할 수 있다. 예컨대, 머신 판독가능 매체는 판독 전용 메모리(Read Only Memory: ROM), 랜덤 액세스 메모리(Random Access Memory: RAM), 자기적 디스크 저장 매체, 광학적 저장 매체, 플래시 메모리(flash memory) 디바이스, 전기적이거나 광학적이거나 음향적이거나 다른 형태의 전파되는 신호(가령, 반송파, 적외선 신호, 디지털 신호 등등) 및 다른 것을 포함할 수 있다.

[0010] 명세서 내에서의 "하나의 구현", "일 구현", "일 예시적 구현" 등등에 대한 참조는 기술된 구현이 특정한 특징, 구조 또는 특성을 포함할 수 있으나, 모든 실시예가 반드시 그 특정한 특징, 구조 또는 특성을 포함하지는 않을 수 있음을 나타낸다. 더욱이, 그러한 문구는 반드시 동일한 구현을 참조하고 있는 것은 아니다. 또한, 특정한 특징, 구조 또는 특성이 일 실시예와 관련하여 기술되는 경우, 본 문서에 명시적으로 기술되든 또는 그렇지 않은 다른 구현과 관련하여 그러한 특징, 구조 또는 특성을 야기하는 것은 당업자의 지식 내에 있다고 체론된다.

[0011] 발화 인식을 위해 구현된 신경망에 관하여, 그리고 몇몇 예에서, 신경망을 통하여 구현된 프레임 스키핑 기법 및 수요 출력값 기법에 관하여 방법, 디바이스, 장치, 컴퓨팅 플랫폼 및 물품이 본 문서에 기술된다.

[0012] 전술된 바와 같이, 실시간으로 신경망을 구현하는 것은 사용자에게 유리할 수 있지만 제한된 계산상의 리소스 및 배터리 리소스의 집약적 사용으로 인해 어려울 수 있다. 나아가, 그러한 리소스 사용을 줄이려는 시도는 부정확한 분류 결과를 제공할 수 있다. 신경망을 최적화하는 것은 데이터 센터를 통하여 호스팅되는(hosted) 서버에서의 소유의 총 비용에 대해 그리고 모바일 디바이스를 통하여 구현된 애플리케이션에서의 배터리 수명에 대해 직접적인 상관을 가질 수 있다.

[0013] 본 문서에서 논의된 몇몇 실시예에서, 신경망을 통하여 프레임 스키핑 기법이 구현될 수 있다. 예컨대, 프레임 스키핑을 사용하는 경우, 신경망 출력(가령, 거리값)이 매 N번째 시간 인스턴스(time instance) 또는 프레임(frame)에 대해 산정되거나 판정될 수 있다. 신경망 거리값이 판정되지 않는 시간 인스턴스에 대해, 그러한 거리값은 하나 이상의 이전의 시간 인스턴스 또는 프레임으로부터의 신경망 판정된 거리값에 기반하여 근사계산될(approximated) 수 있다. 예컨대, 신경망을 평가하는 것은, 몇몇 예에서, 전체 신경망이 그 시간 인스턴스에서 또는 그 프레임에 대해 평가되기에 계산상으로 복잡할 수 있다. 근사계산 방법에 의해 판정되는 거리값은 훨씬 더 낮은 계산상의 복잡도로써 (그리고 신경망의 평가 없이) 계산될 수 있다. 예컨대, 근사계산된 거리값은 거리값의 하나 이상의 이전의 프레임을 사용하여 외삽(extrapolation) 기법에 의해 판정될 수 있다. 몇몇 예에서, 외삽은 두 개의 앞의 프레임으로부터의 거리값에 기반하는 선형 외삽을 포함할 수 있다. 본 문서에서 사용되는 바와 같이, 평가 프레임 또는 비-스킵(non-skip) 프레임은 신경망의 평가에 의해 판정된 거리값을 나타내고 스킵 프레임은 근사계산 기법에 의해 판정된 거리값을 나타낸다. 스킵 프레임을 위한 거리값의 근사계산과 조합된 그러한 프레임 스키핑은 본 문서에서 추가로 논의되는 바와 같이 발화 인식 정확도의 최소의 희생으로써 또는 이의 어떤 희생도 없이 실질적인 계산상의 절감을 제공할 수 있다.

[0014] 나아가, 본 문서에서 논의된 몇몇 실시예에서, 수요에 따른(on demand) 신경망 (가령, 거리값인) 출력(가령, 수요 거리(Distance On Demand) 또는 DOD) 기법이 신경망을 통하여 구현될 수 있다. 예컨대, 수요 거리 기법을 사용하는 경우, 발화 디코더(speech decoder)와 같은 다운스트림 디코더(downstream decoder)는, 시간 인스턴스

또는 프레임에 대해, 모든 이용가능한 거리값의 서브세트(subset)인 요청된 거리값을 제공할 수 있다. 예컨대, 비터비 빔 탐색 디코더(Viterbi beam searching decoder)와 같은 발화 디코더는, 특정한 시간 인스턴스에 대해, 신경망으로부터 이용가능한 거리값의 서브세트를 요구할 뿐일 수 있다. 그러한 예에서, 발화 디코더는 출력 인덱스(output index)(가령, 특정한 시간 인스턴스에 대해 어느 출력 또는 거리값이 필요한지의 지시자(indicator))를 신경망에 제공할 수 있다. 신경망은, 본 문서에서 추가로 논의되는 바와 같이, 입력층, 하나 이상의 은닉층(hidden layer), 그리고 출력층을 포함할 수 있다. 예컨대, 출력층으로부터의 출력 또는 거리값이 발화 디코더에 제공될 수 있다. 몇몇 예에서 출력층의 각각의 노드는 최종 은닉층(가령, 출력층에 연결된 은닉층)의 노드 하나하나에 연결되므로, 출력층의 심지어 하나의 노드를 평가하기 위해 최종 은닉층은 전적으로 평가되어야만 한다. 따라서, 몇몇 예에서, 신경망은 최종 은닉층을 통해 전적으로 평가될 수 있지만, 요청된 출력 인덱스와 연관된 노드의 서브세트만이 평가될 수 있다. 요청되지 않은 출력 노드를 평가하지 않음으로써, 특히나 많은 구현에서 그러하듯이 출력층이 전체 신경망의 실질적인 일부본인 경우, 실질적인 계산상의 절감이 이루어질 수 있다.

[0015] 몇몇 실시예에서, 그러한 프레임 스킵핑 기법 및 그러한 수요 거리 기법은 조합될 수 있다. 예컨대, 발화 디코더는 매 시간 인스턴스에 대해 논의된 바와 같은 출력 인덱스를 제공할 수 있다. 평가 또는 비-스킵 프레임에 대해, 출력 인덱스와 연관된 거리값의 서브세트는 논의된 바와 같이 판정될 수 있다(가령, 신경망은 최종 은닉층을 통해 전적으로 평가될 수 있고 출력 인덱스와 연관된 출력 노드만이 평가될 수 있되 결과적인 거리값은 발화 디코더에 제공됨). 나아가, 최종 은닉층으로부터의 거리값 및 결과적인 값의 서브세트는 아래에서 논의되는 바와 같이 차후의 사용을 위해 메모리 내에 세이브될(saved) 수 있다. 스킵 프레임에 대해, 출력 인덱스와 연관된 거리값의 서브세트는 예컨대 외삽 기법을 사용하여 근사계산될 수 있다. 그러한 기법은, 특정한 거리값이 근사계산되도록, 이전의 프레임(들)으로부터의 연관된(가령, 동일한 출력층 노드로부터의) 거리값을 요할 수 있다. 몇몇 예에서, 이전의 프레임(들)으로부터의 그러한 거리값은 논의된 바와 같이 이전에 발화 디코더에 의해 요청되고 신경망을 통하여 판정되며 메모리에 세이브되었을 수 있다.

[0016] 다른 예에서, 이전의 프레임(들)으로부터의 그러한 거리값은 신경망을 통하여 이전에 판정되지 않았을 수 있다. 그러한 예에서, 세이브된 최종 은닉층 값은 필요한 이전의 프레임(들) 거리값을 위해 신경망의 출력층의 노드를 재평가하는 데에 신경망을 통하여 사용될 수 있다. 예컨대, 필요한 거리값은 앞의 프레임 또는 시간 인스턴스로부터의 것일 수 있지만 현재의 프레임 또는 시간 인스턴스에 대한 현재의 거리값을 근사계산하기 위해 필요할 수 있다. 그러면 현재의 거리값 또는 값들은, 예컨대 선형 외삽 기법 또는 유사한 것을 사용하여 논의된 바와 같이 근사계산될 수 있다. 프레임 스킵핑 및 수요 거리 기법의 그러한 조합은 스킵 프레임 또는 스킵 시간 인스턴스에 대한 신경망의 평가를 제거할 수 있고 평가 프레임에 대한 신경망의 출력층의 평가(그리고 논의된 바와 같은 차후의 재평가)를 실질적으로 줄일 수 있다.

[0017] 논의된 거리값은 단어 또는 구문 또는 n-그램(n-gram) 또는 유사한 것과 같은 텍스트(textual) 요소의 시퀀스(sequence)를 판정하기 위해 발화 디코더에 의해 사용될 수 있다. 본 문서에서 논의된 기법은 계산상의 리소스, 모바일 디바이스 구현을 위한 배터리 수명, 클라우드(cloud) 또는 원격 서버 구현을 위한 소유의 비용, 또는 유사한 것을 절감할 수 있다. 나아가, 그러한 기법은 본 문서에서 추가로 논의되는 바와 같이 발화 인식을 위한 실시간 구현을 제공할 수 있다.

[0018] 도 1은 본 개시의 적어도 몇몇 구현에 따라 마련된, 발화 인식을 제공하기 위한 예시적 세팅(100)의 설명도이다. 도 1에 도시된 바와 같이, 세팅(100)은 디바이스(102)에 의한 평가를 위한 발화(103)를 제공하는 사용자(101)를 포함한다. 예컨대, 디바이스(102)는 발화(103)가 단어, 문장, n-그램, 또는 유사한 것과 같은 텍스트 또는 텍스트 요소로 변환될 수 있도록 발화 인식을 제공할 수 있다. 도시된 바와 같이, 몇몇 예에서, 발화 인식 시스템은 디바이스(102)와 같은 디바이스를 통하여 구현될 수 있다. 예시된 바와 같이, 디바이스(102)는 스마트폰일 수 있다. 그러나, 디바이스(102)는 컴퓨터, 랩톱(laptop), 울트라북(ultrabook), 태블릿(tablet), 또는 유사한 것과 같은 임의의 적합한 디바이스일 수 있다. 몇몇 예에서, 디바이스(102)는 스마트워치(smart watch) 또는 스마트안경(smart glasses) 또는 유사한 것과 같은 착용가능 디바이스(wearable device)일 수 있다. 다른 예에서, 클라우드 발화 인식 시스템 내의 서버 또는 서버들과 같이 디바이스(102)에 원격인 시스템을 통하여 발화 인식이 제공될 수 있다. 몇몇 예에서, 발화(103)는 (디바이스(102)의 맨 아래로 예시된) 디바이스(102)의 마이크(microphone)(104)를 통하여 수신될 수 있다. 다른 예에서, 발화(103)는 발화의 사전 레코딩(pre-recording) 또는 발화 신호(speech signal) 또는 유사한 것으로서 수신될 수 있다. 나아가, 몇몇 예에서, 텍스트 요소는 디바이스(102)의 디스플레이(105)를 통하여 사용자(101)에게 제공될 수 있다. 다른 예에서, 텍스트 요소는 디바이스(102)의 메모리에 또는 원격 클라우드 메모리에 또는 유사한 것에 세이브될 수 있다. 몇몇

예에서, 디바이스(102)는 본 문서에서 사용되는 바와 같이 컴퓨팅 디바이스로서 기술될 수 있다.

[0019] 도 2는 본 개시의 적어도 몇몇 구현에 따라 마련된, 발화 인식을 제공하기 위한 예시적 시스템(200)의 설명도이다. 도 2에 도시된 바와 같이, 시스템(200)은 마이크(104), 특징 추출 모듈(feature extraction module)(202), 거리값 계산 모듈(204) 및 발화 디코더 모듈(206)을 포함할 수 있다. 몇몇 예에서, 발화 디코더 모듈(206)은 메모리를 통하여 구현된 통계적 모델(도시되지 않음)에 커플링될(coupled) 수 있는데, 예컨대, 인식된 단어 시퀀스(recognized word sequence)((207)를 판정하기 위해 거리값(205)과 비교될 수 있다. 도시된 바와 같이, 마이크(104)는 사용자(101)로부터 발화(103)를 수신할 수 있다. 발화(103)는 사용자(101)에 의해 발생될 수 있고 마이크(104)는 (가령, 공기 중의 음파로서) 발화(103)를 수신하고 발화(103)를 디지털 신호와 같은 전기적 신호로 변환하여 발화 레코딩(speech recording)(201)을 생성할 수 있다. 예컨대, 발화 레코딩(201)은 (도 2에 도시되지 않은) 메모리 내에 저장될 수 있다. 다른 예에서, 발화 레코딩(201)은 사전 레코딩될(pre-recorded) 수 있고 발화 레코딩(201)은 다른 디바이스를 통하여 시스템(200)에 의해 수신될 수 있다.

[0020] 특징 추출 모듈(202)은 마이크(104)로부터 또는 시스템(200)의 메모리로부터 발화 레코딩(201)을 수신할 수 있고 특징 추출 모듈(202)은 발화(103)와 연관된 특징(203)을 생성할 수 있다. 특징(203)은 발화(103)를 나타내는 임의의 적합한 특징을 포함할 수 있고 특징(203)은 특징 벡터(feature vector) 포맷 또는 유사한 것과 같은 임의의 적합한 포맷으로 표현될 수 있다. 예컨대, 특징(203)은 수신된 발화의 전력 스펙트럼(power spectrum)을 나타내는 계수(coefficient) 또는 다른 스펙트럼 분석 계수나 파라미터일 수 있다. 몇몇 예에서, 특징(203)은 멜 주파수 cepstrum 계수(Mel Frequency Cepstrum Coefficient: MFCC)일 수 있다. 몇몇 예에서, 특징 추출 모듈(202)은 특징 벡터를 생성하기 위해 발화 레코딩(201)의 발화 파 신호(speech wave signal)를 처리할 수 있다. 특징(203)이 특징 벡터를 통하여 표현되는 예에서, 특징(203)의 각각의 특징 벡터는 발화(103)(그리고/또는 발화 레코딩(201))의 시간 윈도우(time window)에 기반할 수 있다. 예컨대, 시간 윈도우는 발화 레코딩(201)을 가로질러 활주하는 발화 레코딩(201)의 어떤 시간 인스턴스 또는 레코딩 지속기간(recording duration)(가령, 10 밀리초(milliseconds) 또는 유사한 것)일 수 있다. 예컨대, 특징(203)의 각각의 특징 벡터는 이로써 연관된 시간 윈도우의 평가(가령, 전력 스펙트럼 분석 또는 유사한 것)에 기반하여 판정될 수 있다. 나아가, 몇몇 예에서, 특징(203)은 한 더미(stack)의 특징 벡터(가령, 여러 시간 인스턴스로부터의 특징 벡터)를 포함할 수 있다. 특징(203)은 임의의 수의 특징을 포함할 수 있다. 예컨대, 특징(203)은 200개 내지 260개의 특징, 250개 내지 300개의 특징, 또는 300개 내지 400개의 특징 또는 유사한 것을 포함할 수 있다. 일 실시예에서, 특징(203)은 253개의 특징을 포함한다. 다른 실시예에서, 특징(203)은 256개의 특징을 포함한다. 본 문서에서 추가로 논의되는 바와 같이, 특징(203)은 신경망의 입력층에 제공될 수 있다. 특징 추출 모듈(202)은 특징(203)을 거리값 계산 모듈(204) 및/또는 시스템(200)의 메모리에 전송할 수 있다.

[0021] 거리값 계산 모듈(204)은 추출 특징 모듈(202)로부터 또는 메모리로부터 특징(203)을 수신할 수 있다. 거리값 계산 모듈(204)은 특징(203)을 (가령, 도시되지 않은 프리-프로세서(pre-processor)를 통한 것이든, 또는 신경망 자체를 통한 것이든) 신경망으로의 입력으로서 취할 수 있다. 나아가, 몇몇 예에서, 거리값 계산 모듈(204)은 발화 디코더 모듈(206)을 통하여 출력 인덱스를 수신할 수 있다. 예컨대, 특징(203)은 현재의 특징 벡터 및 현재의 특징 벡터에 선행하고/하거나 후속하는 사전결정된 수의 특징 벡터를 포함할 수 있는 특징 벡터의 더미를 포함할 수 있다. 일 실시예에서, 특징(203)은 한 더미의 11개의 특징 벡터를 위해 현재의 특징 벡터, 5개의 선행 특징 벡터 및 5개의 후속 특징 벡터를 포함한다. 몇몇 예에서, 각각의 특징 벡터는 23개의 특징을 포함한다. 더미가 11개의 특징 벡터(각각은 23개의 특징을 가짐)를 포함하는 예에서, 신경망(301)으로의 입력의 수는 253개의 입력일 수 있다(가령, 23×11 개의 입력). 그러한 예에서, 거리값 계산 모듈(204)을 통하여 구현된 신경망은 본 문서에서 추가로 논의되는 바와 같이 253개의 노드 또는 뉴런(가령, 신경망으로의 입력의 수와 동일한 수의 입력층 노드)을 포함하는 입력층을 가질 수 있다. 몇몇 예에서, 그러한 특징은 각각의 시간 인스턴스에서 (가령, 위에서 논의된 바와 같은 각각의 시간 윈도우에 대해) 거리값 계산 모듈(204)에 제공될 수 있다.

[0022] 아래에서 추가로 논의되는 바와 같이, 거리값 계산 모듈(204)은 거리값(205)을 생성하기 위해 신경망 및/또는 거리값 근사계산 모듈(distance value approximation module)을 구현할 수 있다. 논의된 바와 같이, 수요 거리 기법을 구현하는 예에서, 거리값 계산 모듈(204)은 발화 디코더 모듈(206)로부터 출력 인덱스(208)를 수신할 수 있고 거리값 계산 모듈(204)은 출력 인덱스(208)와 연관된 것에 대해서만 거리값을 제공할 수 있다. 몇몇 예에서, 어떤 수요 거리도 활용되지 않고, 그러한 예에서, 출력 인덱스(208)는 구현되지 않을 수 있다. 나아가, 거리값 계산 모듈(204)은 몇몇 시간 인스턴스에서 (가령, 평가 프레임에 대해) 거리값(205)이 신경망을 통하여 제공되고 다른 시간 인스턴스에서 (가령, 스킵 프레임에 대해)거리(205)가 신경망을 통하여 평가된 하나 이상의 이전의 프레임(들)으로부터의 거리값에 기반하여 근사계산 기법을 통하여 제공되도록 프레임 스킵핑 기법을 구

현할 수 있다. 도시된 바와 같이, 거리값 계산 모듈(204)은 발화 디코더 모듈(206)에 그리고/또는 시스템(200)의 메모리에 거리값(205)을 전송할 수 있다.

[0023] 발화 디코더 모듈(206)은 거리값 계산 모듈(204)로부터 또는 메모리로부터 거리값(205)을 수신할 수 있다. 발화 디코더 모듈(206)은 거리값(205)을 디코딩하고 가장 가능성 있는 텍스트 요소 및/또는 인식된 단어 시퀀스 매치(match)를 찾아 탐색할 수 있다. 예컨대, 발화 디코더 모듈(206)은 매 시간 인스턴스(가령, 10 밀리초 또는 유사한 것)에 대한 거리값(205)을 수신하고 인식된 단어 시퀀스(207)를 발화의 끝이 검출된 후에 전달할 수 있다. 발화 디코더 모듈(206)은 임의의 적합한 발화 디코더를 포함할 수 있다. 일례에서, 발화 디코더 모듈(206)은 비터비 빔 탐색 디코더이다. 도시된 바와 같이, 몇몇 예에서, 발화 디코더 모듈(206)은 인식된 단어 시퀀스(207)를 출력으로서 제공할 수 있다. 인식된 단어 시퀀스(207)는 시스템(200)의 메모리에 저장되고/되거나 디스플레이(105) 또는 유사한 것을 통하여 사용자(101)에게 디스플레이될 수 있다. 몇몇 예에서, 인식된 단어 시퀀스(207)는 다른 모듈 또는 소프트웨어 애플리케이션 또는 유사한 것에 그 모듈 또는 소프트웨어 애플리케이션에 의한 사용을 위해 제공될 수 있다. 본 문서에서 사용되는 바와 같은 인식된 단어 시퀀스(207) 또는 텍스트 요소는 단어, 조단어(sub-word) 유닛, n-그램, 음절, 문자 또는 유사한 것의 임의의 적합한 시퀀스를 포함할 수 있다. 도시된 바와 같이, 발화 디코더 모듈(206)은 인식된 단어 시퀀스(207)를 거리값(205)에 기반하여 생성할 수 있다. 나아가, 수요 거리 구현에서, 발화 디코더 모듈(206)은 출력 인덱스(208)를 생성하고 그러한 출력 인덱스(208)를 거리값 계산 모듈(204)에 제공할 수 있다. 예컨대, 출력 인덱스(208)는 특정한 시간 인스턴스에서 필요한 이용가능한 거리값(가령, 거리값 계산 모듈(204)을 통하여 이용가능함)의 서브세트(가령, 그 중의 하나 이상)를 나타낼 수 있다. 예컨대, 발화 디코더 모듈(206)은 다양한 실시예에서 각각의 시간 인스턴스에서 모든 이용가능한 거리값을 필요로 하지는 않을 수 있다.

[0024] 도 2에 도시된 바와 같이, 몇몇 예에서, 거리값 계산 모듈(204)은 발화 인식 시스템의 일부로서 구현될 수 있다. 그러나, 거리값 계산 모듈(204)은 시각적 컴퓨팅 시스템(perceptual computing system), 머신 러닝 시스템, 인지적 컴퓨팅 시스템(cognitive computing system), 이미지 처리 시스템(image processing system), 또는 광학적 문자 인식 시스템(optical character recognition system) 또는 유사한 것과 같은 임의의 적합한 시스템 내에 구현될 수 있다. 나아가, 거리값 계산 모듈(204)의 신경망은 신경망의 가중치(weight) 및/또는 바이어스(bias)를 판정하기 위해 시스템(200)을 통한 구현 전에 훈련 세트(training set) 또는 유사한 것에 기반하여 사전 훈련될(pre-trained) 수 있다. 몇몇 예에서, 사전 훈련이 시스템(200) 자체를 통하여 구현될 수 있다. 다른 예에서, 그러한 사전 훈련 또는 다른 사전 구현 단계가 별개의 시스템에 의해 수행될 수 있다.

[0025] 도 3은 본 개시의 적어도 몇몇 구현에 따라 마련된 예시적 거리값 계산 모듈(204)의 설명도이다. 도시된 바와 같이, 거리값 계산 모듈(204)은 신경망(301), 제어기(302) 및 거리값 근사계산 모듈(303)을 포함할 수 있다. 도 3의 실시예에서, 거리값 계산 모듈(204)은 수요 거리 기법이 아니라 거리값 근사계산으로써 프레임 스키핑을 구현할 수 있고 따라서, 어떤 출력 인덱스도 거리값 계산 모듈(204)을 통하여 수신될 수 없다. 도시된 바와 같이, 거리값 계산 모듈(204)은 신경망(301)을 통하여 특징(203)을 수신할 수 있다. 신경망(301)은 딥 신경망 또는 유사한 것과 같은 임의의 적합한 신경망을 포함할 수 있다.

[0026] 도 4는 본 개시의 적어도 몇몇 구현에 따라 마련된 예시적 신경망(301)의 설명도이다. 도시된 바와 같이, 신경망(301)은 입력층(401), 은닉층(402, 403, 404, 405) 및 출력층(406)을 포함할 수 있다. 나아가, 은닉층(405)은 그것이 출력층(406)에 인접하므로 최종 은닉층으로서 특징지어질 수 있다. 또한 도시된 바와 같이, 입력층(401)은 입력층 노드(407)를 포함할 수 있다. 논의된 바와 같이, 입력층(401)은 임의의 수의 입력층 노드(407)를 포함할 수 있다. 예컨대, 입력층(401)은 요소 특징(203)의 개수와 동일한 수의 노드를 포함할 수 있다. 예컨대, 입력층(401)은 253개 또는 256개 또는 기타의 입력층 노드(407)를 가질 수 있다.

[0027] 나아가, 보여진 예에서와 같이, 신경망(301)은 4개의 은닉층(402 내지 405)을 포함한다. 그러나, 다른 예에서, 신경망은 3개, 5개, 6개 또는 더 많은 은닉층을 포함할 수 있다. 은닉층(402 내지 405)은 임의의 수의 은닉층 노드(408, 409, 410, 411)를 포함할 수 있다. 예컨대, 은닉층(402 내지 405)은 각각 100개 내지 200개의 노드, 200개 내지 300개의 노드, 또는 유사한 것을 포함할 수 있다. 일 실시예에서, 신경망(301)은 각각 192개의 노드를 가지는 4개의 은닉층(402 내지 405)을 포함한다. 몇몇 예에서, 은닉층(402 내지 405)은 각각 동일한 수의 노드를 가지며, 다른 예에서, 은닉층(402 내지 405) 중 하나 이상은 상이한 수의 노드를 가질 수 있다.

[0028] 출력층(406)은 텍스트 요소 또는 인식된 단어 시퀀스 또는 유사한 것을 판정하기 위해 비교 및/또는 탐색을 위한 값을 거리값(Distance Value: DV)(205)이 포함하도록 임의의 적합한 수의 출력층 노드(412)를 포함할 수 있다. 예컨대, 출력층(406)은 400개 내지 800개의 노드, 800개 내지 1,500개의 노드, 또는 1,500개 내지 2,500개

의 노드 또는 더 많은 노드를 포함할 수 있다. 일 실시예에서, 출력층(406)은 512개의 출력층 노드(412)를 포함한다. 일 실시예에서, 출력층(406)은 1,015개의 출력층 노드(412)를 포함한다. 보여진 예에서, 데이터는 좌측에서 우측으로 입력층(401)으로부터, 은닉층(402 내지 405)을 통해, 그리고 출력층(406)을 통해 도시된 바와 같이 흐르며 입력층(401)의 출력은 은닉층(402)으로의 입력이고, 은닉층(402)의 출력은 은닉층(403)으로의 입력이며 기타 등등이고, 출력층(405)의 출력은 신경망(301)의 출력(가령, 거리값(205))이다. 몇몇 예에서, 층 내의 노드 하나하나는 인접한 층 내의 노드 하나하나에 연결될 수 있다(가령, 그 층들은 전적으로 연결될(fully connected) 수 있음). 일례에서, h 개의 노드를 갖는 층은 hh 개의 노드를 갖는 그것의 이웃 층에 $h \times hh$ 개의 가중치를 통해 연결될 수 있다. 일례에서, 입력층(401)은 253개의 입력층 노드(407)를 가지고, 은닉층(402 내지 405)은 각각 192개의 은닉층 노드(408 내지 411)를 가지며, 출력층(406)은 1,105개의 출력층 노드(412)를 가지고, 신경망(301)은 약 354,000개의 가중치를 가진다. 예컨대, 입력층(401)의 매 입력층 노드(407)는 은닉층(402)의 매 은닉층 노드(408)에 연결될 수 있고, 은닉층(402)의 매 은닉층 노드(408)는 은닉층(403)의 매 은닉층 노드(409)에 연결될 수 있으며, 기타 등등이다. 다른 예에서, 노드 간의 몇몇 연결은 만들어지지 않을 수 있다.

[0029] 신경망(301)의 평가(가령, 계산)은 임의의 적합한 기법 또는 기법들을 포함할 수 있다. 예컨대, 입력층(401)의 입력층 노드(407)는 특징(203), 특징(203) 중의 각각의 특징과 연관된 가중치, 그리고/또는 입력층 노드(407) 각각을 위한 활성화(activation) 함수에 기반하여 산정될 수 있다. 일례에서, 입력층 노드(407) 각각은 특징(203) 및 그것의 연관된 가중치(가령, 상이한 특징을 위한 가중치는 상이할 수 있음)의 곱의 가중화된 합을 생성하는 것 및 가중화된 합에 활성화 함수를 적용하는 것에 의해 판정될 수 있다. 은닉층 노드(408)는 입력층 노드(407), 입력층 노드(407) 각각과 연관된 가중치(가령, 입력층 노드(407) 및 은닉층 노드(408)의 상이한 연결 간의 가중치가 상이할 수 있음), 은닉층 노드(408) 각각을 위한 바이어스, 그리고/또는 은닉층 노드(408) 각각을 위한 활성화 함수에 기반하여 판정될 수 있다. 일례에서, 은닉층 노드(408) 각각은 입력층 노드(407) 및 연관된 가중치의 곱의 가중화된 합을 생성하는 것, 가중화된 합에 바이어스를 적용하는 것 및 바이어스된(biased) 가중화된 합에 활성화 함수를 적용하는 것에 의해 판정된다. 은닉층 노드(409, 410, 411)는 은닉층 노드와 유사하게 그러나 선행 층을 각 은닉층으로의 입력으로서 사용하여 판정될 수 있다. 나아가, 출력층 노드는 최종 은닉층 노드(411), 최종 은닉층 노드(411) 각각과 연관된 가중치(가령, 가중치는 상이할 수 있음), 그리고/또는 출력층 노드(412) 각각을 위한 바이어스에 기반하여 판정될 수 있다. 일례에서, 출력층 노드(412) 각각은 최종 은닉층 노드(411) 및 연관된 가중치의 곱의 가중화된 합을 생성하는 것 및 가중화된 합에 바이어스를 적용하는 것에 의해 판정된다. 논의된 바와 같이, 신경망(301)의 노드를 평가하기 위해 다른 기법이 사용될 수 있고 본문서에서 논의된 기법은 임의의 신경망 평가 기법 또는 기법들에 한정되지 않는다.

[0030] 몇몇 예에서, 신경망(301)이 훈련 국면(training phase)에서 훈련된 후 신경망(301)은 테스트 또는 구현 국면에서 발화 인식을 위해 구현될 수 있다. 그러한 훈련 국면은 신경망(301)의 노드를 위한 가중치, 신경망(301)의 노드를 위한 바이어스, 그리고 유사한 것을 판정할 수 있다. 몇몇 예에서, 신경망(301)의 교차 엔트로피(cross-entropy) 훈련 동안(가령, 훈련 국면 동안), 출력층(406)은 구현 또는 테스트 국면 동안에 생략될 수 있는 소프트맥스(softmax) 활성화 함수를 가질 수 있다. 몇몇 예에서, 구현 동안, 출력층(406)으로부터의 출력은 거리값(304)으로서 제공되기 전에 클래스 확률(class probability)에 기반하여 스케일링될(scaled) 수 있다.

[0031] 또한 도 4에 도시된 바와 같이, 신경망(301)은 신경망(Neural Network: NN) 제어 신호(305)를 통하여 제어기(302)의 제어 하에서 동작될 수 있다. 예컨대, 논의된 바와 같은 프레임 스키핑 구현에서, 신경망 제어 신호(305)는 신경망(301)이 특정한 시간 인스턴스에서 평가되는지 여부를 제어할 수 있다. 예컨대, 평가 프레임에 대해, 신경망 제어 신호(305)는 신경망(301)의 평가를 위해 시그널링할(signal) 수 있고, 스킵 프레임에 대해, 신경망 제어 신호(305)는 신경망(301)의 비 평가(no evaluation)를 위해 시그널링할 수 있다.

[0032] 도 3으로 되돌아가면, 논의된 바와 같이, 평가 프레임 또는 시간 인스턴스에 대해, 제어기(302)는 신경망(301)의 평가를 위한 신경망 제어 신호(305)를 신경망(301)에 제공할 수 있다. 그러한 시간 인스턴스에서 또는 그러한 프레임에 대해, 신경망(301)은 거리값(304)을 생성하기 위해 신경망(301)의 특징(203) 및 다른 특성(가령, 가중치, 바이어스, 활성화 함수 및 유사한 것)에 기반하여 거리값(304)을 생성할 수 있는데, 이는 도시된 바와 같이 그러한 평가 프레임 또는 시간 인스턴스에 대해 거리값(205)으로서 제공될 수 있다. 그러한 평가 프레임 또는 시간 인스턴스 동안, 제어기(302)는 어떤 근사계산도 행해지지 않도록 거리값 근사계산 모듈(303)에 또한 시그널링할 수 있다. 도시된 바와 같이, 거리값(304)은 차후의 시간 인스턴스에서의 사용을 위해 거리값 근사계산(303)에 또한 제공될 수 있다.

[0033] 스킵 프레임에 대해, 제어기(302)는 신경망(301)의 비 평가를 나타내는 신경망 제어 신호(305)를 제공할 수 있

다. 나아가, 제어기(302)는 그러한 스킵 프레임 또는 스킵 프레임과 연관된 시간 인스턴스에 대해 거리값 근사 계산 제어 신호(306)를 통하여 거리값(307)이 생성되어야 한다는 표시(indication)를 거리값 근사 계산 모듈(303)에 제공할 수 있다. 예컨대, 그러한 시간 인스턴스에서, 신경망(301)은 평가되지 않을 수 있어(가령 그것은 오프(off)일 수 있음) 실질적인 컴퓨팅 리소스를 절감한다. 나아가, 그러한 시간 인스턴스에서, 거리값(205)은 근사 계산된 거리값(307)으로서 제공할 수 있다.

[0034] 예컨대, 특징(203)은 시간에 걸쳐 대단히 상관될(correlated) 수 있고, 결과적으로, 특징(203)의 연이은(consecutive) 특징 벡터는 실질적으로 유사할 수 있다. 나아가, 만일 그러한 특징에 대해 전적으로 평가되면, 신경망(301)은 시간에 걸쳐 실질적으로 유사하거나 상관된 거리값을 제공할 것이다. 시간에 걸친 그러한 유사성은 본 문서에서 논의된 바와 같이 (가령, 프레임 스키핑이 구현될 수 있도록) 각각의 시간 인스턴스에서 신경망(301)을 평가하는 것을 방지하는 데에 활용될 수 있다. 예컨대, 신경망(301)은 매 프레임 대신 매 N번째 프레임에 대해 평가될 수 있고(여기서 N=2, 3, 4 또는 더 큰 수임) 앞서 평가된 시간 인스턴스(가령, 평가 프레임)으로부터의 거리값이 그러한 스킵 프레임에 대한 거리값을 근사 계산하는 데에 사용될 수 있다.

[0035] 도 5는 본 개시의 적어도 몇몇 구현에 따라 마련된, 평가 및 스킵 프레임의 예시적 시간선(500)을 보여준다. 도 5에서, 시간선(500)의 x축은 예시적 시간 인스턴스 0, 1, 2, 3, 4 및 5에 걸쳐 시간을 증가시키는 것을 보여준다. 도시된 바와 같이, 시간 인스턴스 0, 2, 4에서, 평가 프레임(501, 502, 503)은 신경망(301)과 같은 신경망을 통하여 생성될 수 있다. 나아가, 도 5는 예시적인 신경망 판정된 거리값(504, 505)을 보여준다. 예컨대, 신경망 판정된 거리값(504, 505)은 양자 모두 출력층 노드(412) 중의 출력 노드와 연관될 수 있다. 예컨대, 거리값(504)은 시간 t=0에서 노드의 출력일 수 있고 거리값(505)은 시간 t=2에서 노드의 출력일 수 있다. 도 5는 근사 계산된(가령, 스킵) 프레임(506, 507, 508)을 또한 보여준다. 예컨대, 시간 인스턴스 1, 3, 5에서, 근사 계산된 프레임(506, 507, 508)은 앞의 평가 프레임(501, 502, 503)에 대한 신경망 판정된 거리값에 기반하여 거리값 근사 계산 모듈(303)을 통하여 생성될 수 있다. 예컨대, 도 5는 신경망 판정된 거리값(504, 505)에 기반하여 판정된 근사 계산된 거리값(509)을 보여준다.

[0036] 근사 계산된 거리값(509)은 임의의 적합한 근사 계산 기법 또는 기법들을 사용하여 판정될 수 있다. 근사 계산된 거리값(509)은 (가령, 근사 계산된 거리값(509)은 근사 계산된 것이고 신경망(301)의 출력력이 아니지만) 출력층 노드(412) 중의 동일한 출력 노드와 연관될 수 있다. 보여진 예에서, 근사 계산된 거리값(509)은 두 개의 앞선 신경망 판정된 거리값(504, 505)을 사용하는 외삽에 기반하여 판정된다. 예컨대, 근사 계산된 거리값(509)은 앞의 신경망 판정된 거리값(504, 505)에 기반하여 외삽될 수 있고, 도시된 바와 같이, 앞의 신경망 판정된 거리값(505)을 위한 시간 인스턴스는 근사 계산된 거리값(509)을 위한 시간 인스턴스 앞일 수 있고 앞의 신경망 판정된 거리값(504)을 위한 시간 인스턴스는 앞의 신경망 판정된 거리값(505)을 위한 시간 인스턴스 앞일 수 있다. 다른 예에서, 근사 계산된 거리값(509)은 오직 하나의 앞선 신경망 판정된 거리값(가령, 앞의 신경망 판정된 거리값(505))을 사용하여 판정될 수 있다. 또 다른 예에서, 근사 계산된 거리값(509)은 오직 셋 이상의 앞선 신경망 판정된 거리값을 사용하여 판정될 수 있다. 외삽 기법에 관해서 예시되고 논의되나, 근사 계산된 거리값(509)은 (가령, 앞의 신경망 판정된 거리값(505) 및 차후의 평가 프레임(503)으로부터의 비슷한 차후의 신경망 판정된 거리값에 기반하여) 보간(interpolation) 기법을 사용하여 판정될 수 있다.

[0037] 몇몇 예에서, 근사 계산된 거리값(509)은 선형 외삽 기법을 사용하여 판정될 수 있다. 예컨대, 근사 계산된 거리값(509)은 앞선 신경망 판정된 거리값(505)을 앞의 신경망 판정된 거리값(505) 및 앞의 신경망 판정된 거리값(504) 간의 차이의 절반에 더하는 것에 기반하여 판정될 수 있다. 몇몇 예에서, 근사 계산된 거리값(509)은 식 1에 도시된 바와 같이 판정될 수 있다:

수학식 1

[0038]
$$o(3) = o(2) + 0.5 \times (o(2) - o(0))$$

[0039] 여기서 o(3)은 근사 계산된 거리값(509)일 수 있고, o(2)는 앞의 신경망 판정된 거리값(505)일 수 있으며, o(0)은 앞의 신경망 판정된 거리값(504)일 수 있다.

[0040] 도 5에 도시된 바와 같이, 몇몇 예에서, 프레임은 하나 걸러 하나씩 근사 계산된 또는 스킵 프레임(가령, N=2인 매 N번째 프레임)일 수 있다. 예컨대, 그러한 프레임 스키핑은 N=2인 경우에 모듈로(modulo) 동작을 통하여 구현될 수 있고 만일 2로 모듈로가 된 시간 인스턴스가 0인 경우, 프레임은 평가 프레임이다. 다른 예에서, 프레

임은 둘 걸러 하나씩(가령, $N=3$), 셋 걸러 하나씩(가령, $N=4$), 또는 넷 걸러 하나씩(가령, $N=5$), 또는 유사하게, 근사계산된 또는 스킵 프레임일 수 있다. 또 다른 예에서, 스킵핑 패턴(skipping pattern)은 스킵되는(skipped) 프레임의 개수가 발화 디코더 시스템(206)의 정확도 필요성 또는 유사한 것에 기반하여 증가되거나 감소되도록 어림법으로(heuristically) 판정될 수 있다. 예컨대, 제어기(302)는 발화 디코더 시스템(206)의 정확도 필요성, 발화 디코더 시스템(206)의 가용 컴퓨팅 리소스(available computing resource), 또는 현재의 실시간 지수(current real time factor) 중 하나 이상에 기반하여 스킵핑 패턴 또는 프레임 스킵핑 레이트(frame skipping rate)를 정할 수 있다. 예컨대, 실시간 지수는 발화 디코더 시스템(206)의 속도를 측정할 수 있다(가령, 만일 지수기간 I 의 입력을 처리하는 데에 시간 P 가 걸리는 경우, 실시간 지수(real time factor) RTF는 P/I 로서 정의될 수 있음). 몇몇 예에서, 프레임 스킵핑 레이트는 발화 디코더 시스템(206)의 정확도 필요성, 발화 디코더 시스템(206)의 가용 컴퓨팅 리소스, 또는 현재의 실시간 지수, 또는 유사한 것에 기반하여 동적으로 판정되거나 조절될 수 있다. 예컨대, 만일 더 높은 정확도가 필요한 경우 프레임 스킵핑 레이트는 감소될 수 있거나, 만일 컴퓨팅 리소스가 이용가능하지 않은 경우 프레임 스킵핑 레이트는 증가될 수 있거나, 만일 현재의 실시간 지수가 너무 높거나 증가하는 경우 프레임 스킵핑 레이트는 증가될 수 있다. 몇몇 예에서, 제어기(302) 또는 시스템(200)의 다른 모듈을 통하여 프레임 스킵핑 레이트를 정하는 데에 모든 세 요인 및/또는 추가적인 요인이 사용될 수 있다. 프레임 스킵핑 레이트는 본 문서에서 논의된 바와 같이 신경망(301) 및 거리값 근사계산 모듈(303)을 제어하기 위해 제어기(302)를 통하여 구현될 수 있다. 예컨대, 프레임 스킵핑 레이트를 증가시키는 것은 평가 프레임 및 스킵 프레임 간에 하나 이상의 추가적인 스킵 프레임을 제공하는 것을 포함할 수 있고 프레임 스킵핑 레이트를 줄이는 것은 평가 프레임 및 스킵 프레임 사이로부터 하나 이상의 추가적인 스킵 프레임을 제거하는 것을 포함할 수 있다.

[0041] 나아가, 논의된 바와 같이, 근사계산된 거리값(509)을 판정하는 데에 선형 외삽이 사용될 수 있다. 다른 예에서, 외삽은 비선형 함수 또는 분산 함수(variance function) 또는 유사한 것에 기반하여 수행될 수 있다. 몇몇 예에서, 근사계산 프레임의 모든 거리값은 동일한 근사계산 기법(가령, 선형 외삽)을 사용하여 판정될 수 있고, 다른 예에서, 근사계산 프레임의 거리값은 상이한 근사계산 기법(가령, 일부는 선형, 일부는 비선형)을 사용하여 판정될 수 있다. 나아가, 도 5는 $t=1$ 프레임(506)을 근사계산된 프레임으로서 보여준다. 예컨대, 프레임(506)은 참조를 위해 앞의 평가 프레임(501)만 가질 수 있고 이에 따라, 프레임(506)은 앞의 평가 프레임(501)과 동일한 것으로 근사계산될 수 있다. 다른 예에서, $t=1$ 프레임(506)은 신경망(301)의 평가를 통하여 판정될 수 있다. 본 문서에서 다른 데에서 논의되는 바와 같이, 평가 프레임(501, 502, 503)은 근사계산된 프레임(506, 507, 508)의 인출(retrieval) 및 생성을 위해 메모리 내에 세이브될 수 있다. 나아가, 선형 외삽 기법을 포함하는 그러한 외삽 기법은 실시간 발화 인식 결과를 위해 작동시간(runtime)에 처리될 수 있다.

[0042] 도 6은 본 개시의 적어도 몇몇 구현에 따라 마련된, 프레임 스킵핑 및 거리값 근사계산을 사용하여 거리값을 판정하기 위한 예시적 프로세스(600)를 보여주는 흐름도이다. 프로세스(600)는 도 6에 보여진 바와 같이 하나 이상의 동작(601 내지 608)을 포함할 수 있다. 프로세스(600)는 디바이스, 예를 들어 디바이스(102) 또는 원격 서버 또는 유사한 것에 의해 수행될 수 있다. 프로세스(600)는 본 문서에서 논의된 바와 같이 발화 디코더에 의한 사용을 위해 거리값을 판정하고 제공하는 데에 사용될 수 있다. 예컨대, 프로세스(600)는 제어기(302), 신경망(301) 및 거리값 근사계산 모듈(303)에 의해 구현될 수 있다.

[0043] 도시된 바와 같이, 프로세스(600)는 시작 동작(601)에서 시작할 수 있고 프로세스(600)는 판단 동작(602), "평가 또는 스킵 프레임"(Evaluation or Skip Frame)에서 계속될 수 있는데, 여기서 현재의 프레임이 평가 프레임 또는 스킵 프레임인지에 대해 판정이 행해질 수 있다. 예컨대, 제어기(302)는 현재의 프레임이 평가 프레임인지 또는 스킵 프레임인지 판정할 수 있다. 도시된 바와 같이, 만일 현재의 프레임이 평가 프레임인 경우, 프로세스(600)는 동작(603), "거리값을 판정하기 위해 신경망을 평가함"(Evaluate Neural Network to Determine Distance Values)에서 계속될 수 있는데, 여기서 거리값은 신경망의 평가에 기반하여 현재의 프레임에 대해 판정될 수 있다. 예컨대, 동작(603)에서 거리값 계산 모듈(204)을 통하여 구현된 바와 같은 신경망(301)에 의해 거리값(304)이 판정될 수 있다.

[0044] 만일 현재의 프레임이 스킵 프레임인 경우, 프로세스(600)는 동작(604), "이전의 프레임(들)의 거리값에 기반하여 거리값을 근사계산함"(Approximate Distance Values based on Distance Values of Prior Frame(s))에서 계속될 수 있는데, 여기서 거리값은 이전의 신경망 산정된 프레임의 거리값에 기반하여 근사계산될 수 있다. 예컨대, 거리값은 본 문서에서 논의된 바와 같이 선형 외삽을 사용하여 근사계산될 수 있다. 일례에서, 동작(604)에서 거리값 계산 모듈(204)을 통하여 구현된 바와 같은 거리값 근사계산 모듈(303)에 의해 거리값(307)이 판정될 수 있다.

[0045] 도시된 바와 같이, 평가 프레임의 경우에서든 또는 스킵 프레임의 경우에서든, 프로세스(600)는 동작(605), "거리값을 제공함"(Provide Distance Values)에서 계속될 수 있는데, 여기서 거리값은 본 문서에서 논의된 바와 같이 텍스트 요소의 시퀀스의 판정을 위해 발화 디코더에 제공될 수 있다. 예컨대, 거리값 계산 모듈(204)은 인식된 단어 시퀀스(207)의 생성을 위해 발화 디코더 모듈(206)에 (가령, 프레임 유형에 따라 거리값(304) 또는 거리값(307)을 포함하는) 거리값(205)을 제공할 수 있다.

[0046] 프로세스(600)는 판단 동작(606), "마지막 시간 인스턴스/프레임?"(Last Time Instance / Frame?)에서 계속될 수 있는데, 여기서 현재의 프레임이 마지막 프레임인지에 대해 판정이 행해질 수 있다. 만일 현재의 프레임이 마지막 프레임이 아닌 경우, 프로세스(600)는 동작(607), "다음 시간 인스턴스/프레임으로 감"(Go To Next Time Instance / Frame)에서 계속될 수 있는데, 여기서 프로세스(600)는 도시된 바와 같이 판단 동작(602)에서 다음 프레임에 대해 다음 시간 인스턴스에서 계속될 수 있다. 만일 현재의 프레임이 마지막 프레임인 경우, 프로세스(600)는 종료 동작(608)에서 종료할 수 있다.

[0047] 논의된 바와 같이, 신경망(301), 제어기(302) 및 거리값 근사계산 모듈(303)을 포함하는 거리값 계산 모듈(204)은 자동 발화 인식 구현에서 계산상의 부하(computational load)를 실질적으로 줄이기 위해 프레임 스킵핑을 구현할 수 있다. 그러한 구현은 다양한 컴퓨팅 환경에서 빠르고 정확한 발화 인식 결과를 줄 수 있다. 예컨대, 그러한 프레임 스킵핑 기법은 정확도의 어떤 손실도 없이 계산 비용에서 50% 이상의 감소를 가능케 할 수 있다. 나아가, 작동시간 전에 또는 작동시간 동안에 스킵 프레임의 수를 설정함으로써든 또는 처리 동안에 스킵 프레임의 수를 조절함으로써든 속도와 정확도 간의 트레이드오프(tradeoff)가 이루어질 수 있다. 나아가, 이전의 (prior) 프레임 스킵핑 또는 다중 프레임(multi-frame) 신경망 구현과 대조적으로, 기술된 프레임 스킵핑 기법은 (가령, 신경망을 훈련하기 위한) 추가적인 특별히 훈련된 지식 소스(specifically trained knowledge source)나 구현 전에 신경망의 조절 또는 최적화를 요구하지 않는다. 표 1은 기술된 프레임 스킵핑 기법의 예시적 결과를 보여준다.

표 1

	N=1	N=2	N=3	N=4
방법	단어 에러 레이트에서의 증가 [%]			
이전의 프레임 스킵핑 기법	0	2	4	8
거리 값 근사계산을 이용한 프레임 스킵핑	0	0	0	4

[0048] 단어 에러 레이트에서의 증가의 비교

[0049] 표 1은 모든 시험을 위해 동일한 발화 인식 엔진을 사용하여 결과를 제공한다. 보고된 결과는 통틀어 대략 총 25000개의 발화된 발언(spoken utterance)에 기반한 6개의 인식 실험의 평균을 나타낸다. 표 1의 결과에서, 거리값 근사계산을 이용한 프레임 스킵핑은 수학적 1에 관해서 논의된 바와 같이 2개의 이전의 신경망 거리를 기반으로 선형 외삽에 기반한다. 도시된 바와 같이, 이전의 프레임 스킵핑 기법에 있어서, 단어 에러 레이트(word error rate)는 (가령, N=1에서 어떤 스킵핑도 없는) 기저선(baseline)으로부터 프레임 스킵핑이 N=4에서일 때 8%로 꾸준히 증가한다. 논의된 기법으로써, N=2 및 N=3에서의 프레임 스킵핑에 있어서 에러 레이트의 어떤 증가도 없다. N=4에 대해, 선형 외삽을 사용하는 거리값 근사계산을 사용하여 단어 에러 레이트의 증가가 4%인데, 이는 이전의 프레임 스킵핑 기법에 기반한 증가의 절반이다. 예컨대, N=3에서, 거리값 근사계산 기법을 이용한 논의된 프레임 스킵핑을 사용하는 자동 발화 인식 시스템은 정확도에서의 손실 없이 50% 넘게 계산상의 비용을 줄일 수 있다. 예컨대, 신경망은 자동 발화 인식 시스템의 전체적인 풋프린트(footprint)에 70%만큼 많이 기여할 수 있고 신경망의 사용의 그러한 감소는 상당히 자동 발화 인식 시스템의 계산 풋프린트를 줄일 수 있다.

[0050] 나아가, 의사코드(Pseudo-code) (1)은 선형 근사계산으로써 프레임 스킵핑을 제공하는 예시적 프로세스를 제공한다.

Pseudo-code (1):

```
Integer N; // modulo
Integer t; // time frame
Integer m; // modulo rest
Integer s; // frame index of last computed frame
Integer p; // frame index of computed frame before frame s
Float fac;
FloatVector feature_vectors[];
FloatVector o[]; // matrix with outputs
FloatVector delta;

for ( t=0; t<T; t++) { // loop over time frames t from 0 to T-1
    m = t % N; // modulo residual
    if ( m == 0 ) { // a non-skipped frame
        o(t) = compute_DNN( feature_vectors(t) );
    } else {
        s = ( t / N ) * N; // index of last computed frame
        p = s - N; // index of computed frame before frame s
        fac = (float)m / (float)N;
        if ( p < 0 ) {
            fac = 0.0; // skip these
            p = 0; // value does not matter
        }
        delta = o(s) - o(p); // delta vector
        o(t) = o(s) + fac * delta;
    }
}
```

[0051]

[0052]

의사코드 (1)에 도시된 바와 같이, 스킵되지 않은(non-skipped) 프레임(가령, 평가 프레임)에 대해, 시간 프레임(가령, 시간 인스턴스 또는 프레임) 및 스킵 레이트(skip rate)(가령, N 값)의 모듈로는 0일 수 있고 그러한 시간 프레임에 대해, 거리값(가령, 출력)은 신경망을 평가함(가령, compute_DNN)으로써 판정될 수 있다. 스킵 프레임에 대해, 시간 프레임 및 스킵 레이트의 모듈로는 0이 아닐 수 있고 앞선 평가 프레임의 인덱스는 (가령, 의사코드 (1) 내의 s 및 p를 통하여) 판정될 수 있다. 그러면 스킵 프레임에 대한 거리값은 사전에 판정된 신경망 출력 간의 델타(delta)에 인자(factor)(가령, fac)를 적용하는 것 및 델타를 가장 최근의 사전에 판정된 신경망 출력에 더하는 것에 의해 사전에 판정된 신경망 출력에 기반하여 판정될 수 있다. 예컨대, 그 인자는 이전의 신경망 계산된 프레임에 대한 스킵 프레임 위치에 기반할 수 있다. N=1인 수학적 1의 예에서, 모듈로(가령, m)은 1이고 스킵 레이트(가령 N)는 2이므로 그 인자는 예컨대 0.5일 수 있다. 평가 프레임 간에 더 많은 스킵 프레임이 있는 예에서, 그 인자는 평가 프레임이 스킵 프레임에 대해 얼마나 최근인지에 따라서 달라질 수 있다. 예컨대, 만일 N=3이고 스킵 프레임이 평가 프레임에 바로 후속하고 있는 경우 그 인자는 1/3일 수 있고 만일 스킵 프레임이 평가 프레임 후의 제2의 스킵 프레임인 경우, 그 인자는 예컨대 2/3일 수 있다.

[0053]

논의된 바와 같이, 신경망(301), 제어기(302) 및 거리값 근사계산 모듈(303)을 포함하는 거리값 계산 모듈(204)은 자동 발화 인식 구현에서 계산상의 부하를 실질적으로 줄이기 위해 프레임 스킵핑을 구현할 수 있다. 다른 실시예에서, 그러한 프레임 스킵핑 기법은 수요 거리 기법과 조합될 수 있다.

[0054]

도 7은 본 개시의 적어도 몇몇 구현에 따라 마련된 예시적 거리값 계산 모듈(204)의 설명도이다. 논의된 바와 같이, 거리값 계산 모듈(204)은 신경망(301), 제어기(302) 및 거리값 근사계산 모듈(303)을 포함할 수 있다. 도 7의 실시예에서, 거리값 계산 모듈(204)은 거리값 근사계산을 이용한 프레임 스킵핑 및 수요 거리 기법을 구현

할 수 있다. 예컨대, 거리값 계산 모듈(204)은 발화 디코더 모듈(206)로부터 출력 인덱스(208)를 수신할 수 있다(도 2를 참조하시오). 그러한 예에서, 거리값 계산 모듈(204) 및 발화 디코더 모듈(206)은 양방향으로(bi-directionally) 커플링될 수 있다. 그러한 출력 인덱스(208)는 발화 디코더 모듈(206)이 특정한 시간 인스턴스에서 (또는 특정한 프레임 또는 유사한 것에 대해) 어느 거리값(가령, 출력)을 요청하고 있는지의 지시자를 포함할 수 있다. 출력 인덱스(208)는 출력층 노드(412)와 연관된 지시자 및/또는 요청을 위한 시간 인스턴스를 나타내는 타임스탬프(time stamp)와 같은 임의의 적합한 지시자를 포함할 수 있다. 도시된 바와 같이, 거리값 계산 모듈(204)은 신경망(301)을 통하여 특징(203)을 그리고 제어기(302)를 통하여 출력 인덱스(208)를 수신할 수 있다. 신경망(301)은 딥 신경망 또는 유사한 것과 같은 임의의 적합한 신경망을 포함할 수 있다. 예컨대, 신경망(301)은 본 문서에서 논의된 바와 같은 임의의 신경망을 포함할 수 있다.

[0055] 도 8은 본 개시의 적어도 몇몇 구현에 따라 마련된 예시적 신경망(301)의 설명도이다. 도시된 바와 같이, 신경망(301)은 입력층(401), 은닉층(402, 403, 404, 405) 및 출력층(406)을 포함할 수 있다. 나아가, 논의된 바와 같이, 은닉층(405)은 그것이 출력층(406)에 인접하기에 최종 은닉층으로서 특징지어질 수 있다. 또한 도시된 바와 같이, 입력층(401)은 입력층 노드(407)를 포함할 수 있고, 은닉층(402, 403, 404)은 각각 은닉층 노드(408, 409, 410)를 포함할 수 있으며, 최종 은닉층(405)은 최종 은닉층 노드(411)를 포함할 수 있다. 또한, 출력층(406)은 출력층 노드(412)를 포함할 수 있다. 신경망(301)의 특성은 도 4에 관해서 논의되었고 간결성을 위해 반복되지 않을 것이다.

[0056] 도 7로 되돌아가면, 제어기(302)는 출력 인덱스(208)를 수신할 수 있다. 제어기(302)는 현재의 시간 인스턴스가 평가 프레임 또는 스킵 프레임과 연관되는지를 또한 판정할 수 있다. 만일 현재의 시간 인스턴스가 평가 프레임과 연관된 경우(가령, 거리값은 신경망(301)의 평가에 기반하여 판정될 것임), 제어기(302)는 신경망(Neural Network: NN) 제어 신호(305)를 신경망(301)에 제공할 수 있다. 신경망 제어 신호(305)는 거리값이 요청된 출력 노드 및 현재의 시간 인스턴스에서 신경망이 평가되어야 함을 나타낼 수 있다. 예컨대, 출력 인덱스(208)는 모든 이용가능한 거리값의 서브세트를 나타낼 수 있고 신경망은 그러한 거리값을 제공하도록 요청받을 뿐일 수 있다.

[0057] 그러한 평가 프레임 예에서, 신경망(301)은 최종 은닉층(405)을 통해 네트워크의 모든 층을 평가할 수 있다. 예컨대, 출력층 노드(412) 중의 심지어 하나의 출력 노드를 평가하기 위하여, 최종 은닉층(405)을 통한 모든 층이 평가될 필요가 있을 수 있다. 도 8로 되돌아가면, 도시된 바와 같이, 그러한 예에서, 신경망(301)은 입력층 노드(407), 은닉층 노드(408, 409, 410) 및 최종 은닉층 노드(411) 전부를 평가할 수 있다. 최종 은닉층 노드(411)를 통하여 판정된 최종 은닉층 값은 스킵 프레임 예에 관해서 본 문서에서 추가로 논의되는 바와 같이 향후의 사용을 위해 메모리 내에 세이브될 수 있다. 나아가, 신경망(301)은 (가령, 출력 인덱스(208)를 통하여) 요청되는 출력층 노드(412)만을 평가할 수 있다. 출력층(406)은 평가되지 않는 출력층 노드(801 내지 805)를 또한 포함할 수 있어서 평가되지 않는 출력층 노드(801 내지 805)(가령, 도 8에서 까맣게 된 것)는 계산되지 않으며 요청된 출력층 노드(412)(가령, 하얗게 되어 있는 것)만 계산된다.

[0058] 도 7로 되돌아가면, 신경망(301)은 제어기(302)를 통하여 요청된 거리값을 요청된 거리값(Requested Distance Value: RDV)(701)으로서 제공할 수 있는데, 이는 거리값 계산 모듈(204)을 통하여 거리값(205)의 일부분으로서 발화 디코더 모듈(206)에 제공될 수 있다(도 2를 참조하시오).

[0059] 논의된 바와 같이, 만일 현재의 프레임이 평가 프레임인 경우, 신경망(301)은 요청된 거리값(701)을 판정하도록 구현될 수 있다. 만일 대신에 현재의 프레임(또는 차후의 프레임)이 스킵 프레임인 경우, 제어기(302)는 거리값 근사계산 제어 신호(306)를 거리값 근사계산 모듈(303)에 제공할 수 있으니 출력 인덱스(208)를 통하여 요청된 거리값의 근사계산을 요청하는 것이다. 그러나, 본 문서에서 논의된 바와 같이, 거리값 근사계산 모듈(303)은 근사계산된 거리값을, 신경망(301)을 통하여 산정된 앞의 거리값에 기반하여 생성할 수 있다. 나아가, 평가 프레임 예에 관해서 논의된 바와 같이, 신경망 출력층 노드의 서브세트만이 평가될 수 있고 신경망 판정된 거리값의 대응하는 서브세트만이 이용가능할 수 있다. 만일 스킵 프레임 요청된 출력 인덱스(208)가 앞의 평가 프레임에서 판정된 거리값에 대응하는 경우, 거리값 근사계산 모듈(303)은 본 문서에서 논의된 바와 같은 근사계산 기법을 통하여 요청된 거리값(702)을 생성하기 위해 그러한 사전에 판정된 거리값을 사용할 수 있다.

[0060] 그러나, 만일 (가령, 신경망(301)을 통한) 그러한 사전에 판정된 거리값이 이용가능하지 않은 경우(가령, 그것이 사전에 산정되어 메모리를 통하여 세이브되지 않았음), 제어기(302)는 앞의 프레임에 대한 필요한 거리값을 판정하기 위해 신경망(301)에 요청을, 신경망 제어 신호(305)를 통하여, 제공할 수 있다. 예컨대, 신경망(301)은 앞선 평가 프레임에 대한 세이브된 최종 은닉층 값을 적재하고(load) 출력층(406)의 새로 요청된 노드를 평

가할 수 있다. 예컨대, 도 8을 참조하면, 사전에 요청되지 않은 출력층 노드(804)와 연관된 거리값이 이제 (현재의) 스킵 프레임에 대한 거리값을 근사계산하는 데에 필요할 수 있다. 신경망(301)은 요청된 노드를 평가하고 도 7에 도시된 바와 같이 요청된 거리값(701)을 거리값 근사계산 모듈(303)에 제공할 수 있다. 그러한 프로세스는 임의의 수의 필요한 신경망 판정된 거리값에 대해 그리고 임의의 수의 앞선 평가 프레임(가령, 통상적으로 2개의 앞선 평가 프레임)에 대해 반복될 수 있다.

[0061] 도 7로 되돌아가면, 거리값 근사계산 모듈(303)은, 이제 필요한 신경망 판정된 거리값으로써, 현재의 스킵 프레임에 대한 요청된 거리값(702)을 생성할 수 있다. 예컨대, 거리값 근사계산 모듈(303)은 본 문서에 기술된 바와 같이 선형, 비선형 또는 분산 함수를 기반으로 외삽 또는 보간 기법을 사용하여 그러한 요청된 거리값(702)을 판정할 수 있다. 거리값 근사계산 모듈(303)은 본 문서에서 논의된 임의의 기법 또는 특성을 사용하여 그러한 요청된 거리값(702)을 판정할 수 있고 그러한 기법 또는 특성은 간결성을 위해 반복되지 않을 것이다.

[0062] 도 2를 참조하면, 발화 디코더 모듈(206)은 그러한 (요청된) 거리값(205)을 수신할 수 있고 인식된 단어 시퀀스를 디코딩하고/하거나 이를 찾아 탐색하기를 계속할 수 있다. 나아가, 발화 디코더 모듈(206)은, 다음의 시간 프레임에 대해, 출력 인덱스(208)를 생성할 수 있는데, 이는 발화 디코더가 기술된 디코딩/탐색을 계속하기 위해 다음의 프레임에 대해 필요로 하는 거리값(205)을 나타낼 수 있다. 예컨대, 발화 디코더 모듈(206)은 인식된 단어 시퀀스(207) 또는 이의 일부분을 효과적으로 찾아 탐색하는 데에 이용가능한 거리값의 서브세트가 사용될 수 있도록 평가되는 가설의 개수 또는 일람표(inventory)를 한정할 수 있는 비터비 빔 탐색(beam searching) 또는 프루닝(pruning) 발화 디코더일 수 있다.

[0063] 도 9는 본 개시의 적어도 몇몇 구현에 따라 마련된, 프레임 스키핑, 수요 거리 및 거리값 근사계산을 사용하여 거리값을 판정하기 위한 예시적 프로세스(900)를 보여주는 흐름도이다. 프로세스(900)는 도 9에 예시된 바와 같이 하나 이상의 동작(901 내지 913)을 포함할 수 있다. 프로세스(900)는 디바이스, 예를 들어 디바이스(102) 또는 원격 서버 또는 유사한 것에 의해 수행될 수 있다. 프로세스(900)는 본 문서에서 논의된 바와 같이 발화 디코더에 의한 사용을 위한 거리값을 판정하고 제공하는 데에 사용될 수 있다. 예컨대, 프로세스(900)는 제어기(302), 신경망(301) 및 거리값 근사계산 모듈(303)에 의해 구현될 수 있다.

[0064] 도시된 바와 같이, 프로세스(900)는 시작 동작(901)에서 시작할 수 있고 프로세스(900)는 동작(902), "시간 인스턴스/프레임에 대한 출력 인덱스를 수신함"(Receive Output Indices for Time Instance / Frame)에서 계속될 수 있는데, 여기서 현재의 시간 인스턴스 또는 프레임에 대해 출력 인덱스가 수신될 수 있다. 예컨대, 발화 디코더 모듈(206)은 출력 인덱스(208)를 생성할 수 있는데, 이는 거리값 계산 모듈(204)에 제공되고 이에 의해 수신될 수 있다. 프로세스(900)는 판단 동작(903), "평가 또는 스킵 프레임"(Evaluation or Skip Frame)에서 계속될 수 있는데, 여기서 현재의 프레임이 평가 프레임 또는 스킵 프레임인지에 대해 판정이 행해질 수 있다. 예컨대, 제어기(302)는 프레임 스킵 레이트를 사용하는 모듈로 연산에 기반하여 현재의 프레임이 평가 프레임 또는 스킵 프레임인지를 판정할 수 있다.

[0065] 도시된 바와 같이, 만일 현재의 프레임이 평가 프레임인 경우, 프로세스(900)는 동작(904), "최종 은닉층을 통해 신경망을 평가함"(Evaluate Neural Network through Final Hidden Layer)에서 계속될 수 있는데, 여기서 최종 은닉층을 통해 신경망이 평가될 수 있다. 예컨대, 신경망(301)은 입력층(401)으로부터 최종 은닉층(405)을 거쳐 전적으로 평가될 수 있다. 프로세스(900)는 동작(905), "출력 인덱스와 연관된 출력층 노드를 통하여 거리값을 판정함"(Determine Distance Values via Output Layer Nodes Associated with Output Indices)에서 계속될 수 있는데, 여기서 신경망의 출력층 노드는 출력 인덱스와 연관된 거리값(가령, 요청된 거리값)을 판정하기 위해 평가될 수 있다. 예컨대, 신경망(301)을 사용하여, 출력층(406)의 출력층 노드(412)의 서브세트가 요청된 거리값을 판정하기 위해 평가될 수 있다. 예컨대, 요청된 거리값(701)이 동작(905)에서 거리값 계산 모듈(204)을 통하여 구현된 것으로 신경망(301)에 의해 판정될 수 있다. 프로세스(900)는 동작(906), "최종 은닉층 값 및 거리값을 세이브함"(Save Final Hidden Layer Values and Distance Values)에서 계속될 수 있는데, 여기서 신경망의 최종 은닉층과 연관된 최종 은닉층 값 및 신경망 판정된 거리값은 향후의 사용을 위해 세이브될 수 있다. 예컨대, 신경망(301)의 최종 은닉층 노드(411)를 통하여 판정된 최종 은닉층 값은 본 문서에서 추가로 논의되는 바와 같이 차후의 거리값의 근사계산에서의 사용을 위해 메모리를 통하여 세이브될 수 있다. 유사하게, 출력층(406)의 활성화된 서브세트를 통하여 판정된 거리값은 차후의 거리값의 근사계산에서의 사용을 위해 세이브될 수 있다. 그러한 세이브된 최종 은닉층 값 및 신경망 판정된 거리값은 더 이상 필요하지 않은 경우(가령, 현재의 프레임이 거리값의 근사계산을 위해 그러한 평가 프레임을 더 이상 재호출하지 않을 수 있는 경우) 폐기될 수 있다.

- [0066] 만일 현재의 프레임이 스킵 프레임인 경우, 프로세스(900)는 동작(907), "이전의 프레임(들)에 대한 최종 은닉층 값 및/또는 거리값을 인출함"(Retrieve Final Hidden Layer Values and/or Distance Values for Prior Frame(s))에서 계속될 수 있는데, 여기서 최종 은닉층 값 및/또는 이전의 신경망 판정된 거리값(동작(906)에서 세이브된 바와 같은 것)이 인출될 수 있다. 예컨대, 그러한 값은 제어기(302) 또는 거리값 근사계산 모듈(303)을 통하여 메모리로부터 인출될 수 있다. 예컨대, 만일 현재의 프레임에 대한 출력 인덱스 중의 출력 인덱스가 사전에 신경망 판정된 거리값과 연관된 경우, 그러한 사전에 신경망 판정된 거리값은 인출될 수 있다. 만일 현재의 프레임에 대한 출력 인덱스의 출력 인덱스 값이 사전에 판정되지 않은 신경망 노드와 연관된 경우, 최종 은닉층 값이 인출될 수 있다.
- [0067] 프로세스(900)는 동작(908), "연관된 출력층 노드를 통하여, 필요에 따라, 이전의 거리값을 판정함"(Determine Prior Distance Values, as Needed, via Associated Output Layer Nodes)에서 계속될 수 있는데, 여기서 이전의 거리값은 연관된 출력층 노드를 통하여 출력층 인덱스에 대해 판정될 수 있다. 예컨대, 근사계산을 위해 현재 필요하지만 신경망을 통하여 사전에 판정되지 않은 거리값에 대해, 그러한 거리값은 신경망의 출력층 노드를 통하여 판정될 수 있다. 예컨대, 동작(907)에서 인출된 최종 은닉층 값이 출력층(406)의 출력층 노드(412)의 서브세트(가령, 그 중의 하나 이상)를 평가하는 데에 사용될 수 있다. 몇몇 예에서, 그러한 거리값은 동작(906)에서 이미 세이브되었을 수 있고 동작(908)은 스킵될 수 있다.
- [0068] 프로세스(900)는 동작(909), "이전의 거리값에 기반하여 출력 인덱스와 연관된 거리값을 근사계산함"(Approximate Distance Values Associated with Output Indices based on Prior Distance Values)에서 계속될 수 있는데, 여기서 거리값은 이전의 신경망 산정된 거리값에 기반하여 근사계산될 수 있다. 근사계산된 거리값은 수신된 출력 인덱스와 연관된 것일 수 있고 이전의 거리값은 논의된 바와 같이 동작(906)을 통하여 세이브되거나 동작(908)을 통하여 판정될 수 있다. 예컨대, 거리값은 본 문서에서 논의된 바와 같이 선형 외삽을 사용하여 근사계산될 수 있다. 일례에서, 요청된 거리값(702)은 동작(909)에서 거리값 계산 모듈(204)을 통하여 구현된 것으로 거리값 계산 모듈(303)에 의해 판정될 수 있다.
- [0069] 도식된 바와 같이, 평가 프레임의 경우에서든 또는 스킵 프레임의 경우에서든, 프로세스(900)는 동작(910), "거리값을 제공함"(Provide Distance Values)에서 계속될 수 있는데, 여기서 본 문서에서 논의된 바와 같이 텍스트 요소의 시퀀스의 판정을 위해 발화 디코더에 거리값이 제공될 수 있다. 예컨대, 거리값 계산 모듈(204)은 인식된 단어 시퀀스(207)의 생성을 위해 발화 디코더 모듈(206)에 거리값(205)(가령, 프레임 유형에 따라서 요청된 거리값(701) 또는 요청된 거리값(702)을 포함함)을 제공할 수 있다.
- [0070] 프로세스(900)는 판단 동작(911), "마지막 시간 인스턴스/프레임?"(Last Time Instance / Frame?)에서 계속될 수 있는데, 여기서 현재의 프레임이 마지막 프레임인지에 대해 판정이 행해질 수 있다. 만일 현재의 프레임이 마지막 프레임이 아닌 경우, 프로세스(900)는 동작(912), "다음의 시간 인스턴스/프레임으로 감"(Go to Next Time Instance / Frame)에서 계속될 수 있는데, 여기서 프로세스(900)는 도식된 바와 같이 동작(902)에서 다음의 프레임에 대해 다음의 시간 인스턴스에서 계속될 수 있다. 만일 현재의 프레임이 마지막 프레임인 경우, 프로세스(900)는 종료 동작(913)에서 종료할 수 있다.
- [0071] 논의된 바와 같이, 신경망(301), 제어기(302) 및 거리값 근사계산 모듈(303)을 포함하는 거리값 계산 모듈(204)은 자동 발화 인식 구현에서 계산상의 부하를 실질적으로 줄이기 위해 프레임 스킵핑 및 수요 거리를 구현할 수 있다. 몇몇 예에서, 그러한 수요 거리 기법은 그러한 프레임 스킵핑 기법의 구현 없이 구현될 수 있다. 예컨대, 도 7을 참조하면, 거리값 계산 모듈(204)은 거리값 근사계산 모듈(303) 없이 구현될 수 있고, 각각의 시간 인스턴스 또는 프레임에 대해, 제어기(302)는 출력 인덱스(208)와 연관된 거리값(가령, 요청된 거리값(701))만을 제공하도록 신경망을 제어할 수 있다. 도 9를 참조하면, 그러한 수요 거리 프로세스는 동작(902, 904, 905, 910, 911, 912 및 913)을 포함할 수 있어서 출력 인덱스가 수신될 수 있고, 최종 은닉층을 통해 신경망이 전적으로 평가될 수 있으며, 출력 인덱스에 대응하는 출력층 노드의 서브세트만이 평가될 수 있고, 판정된 거리값(가령, 그 서브세트)은 발화 디코더에 제공될 수 있다. 그러한 프로세스는 최종 은닉층 값 또는 거리값을 예컨대 향후의 사용을 위해 세이브하는 것도, 스킵 프레임 기법의 통합을 가능케 하는 동작의 분기(branch)(가령, 동작(903, 907, 908))의 구현도 필요케 하지 않을 수 있다.
- [0072] 그러한 프레임 스킵핑 기법은 논의된 바와 같이 거리값이 외삽 또는 보간 기법을 통하여 근사계산되기에 근사계산 기법이라고 간주될 수 있다. 그러한 수요 거리 기법은 발화 디코더가 현재의 프레임에 대해 그것이 필요로 하는 거리값만을 수신하고 있고 (신경망 판정된 거리값의 경우에) 그러한 거리값은 근사계산되지 않으므로 비-근사계산(non-approximation) 기법이라고 간주될 수 있다. 따라서, 그러한 수요 거리 기법은 정확도를 감소시키

지 않고서 계산상의 비용을 줄일 수 있다. 예컨대, 프레임 스키핑만과 비교하여, 수요 거리 기법의 추가는 22%만큼 계산을 감소시킬 수 있다. 몇몇 예에서, 출력층은 신경망의 약 50%일 수 있고 출력층의 요청된 서브세트만을 평가하는 것은 신경망의 계산상의 비용의 약 45%를 절감할 수 있다. 다양한 예에서, 출력층의 요청된 서브세트를 평가하는 것은 신경망의 계산상의 비용의 0 내지 50%를 절감할 수 있다. 표 2는 기술된 프레임 스키핑 및 수요 거리 기법의 예시적 결과를 보여준다.

표 2

수요 거리 구현됨?	프레임 스키핑 구현됨?	전체적인 실시간 지수	신경망 계산 비용
아니오	아니오	2.49	~800 MCPS
예	아니오	1.97	~615 MCPS
아니오	예	1.18	~290 MCPS
예	예	1.00	~225 MCPS

실시간 지수 및 계산 비용의 비교

[0073]

[0074]

표 2는 253개의 입력층 노드, 각각 192개의 은닉층 노드를 갖는 4개의 은닉층, 그리고 1,015개의 출력층 노드를 가지는 예에서 논의된 바와 같이 약 354,000개의 가중치를 가지는 신경망에 대한 예시적 결과를 제공한다. 나아가, 그 결과는 모든 시험에 대해 동일한 발화 인식 엔진으로써 획득되었다. 적용된 통계적 언어 모델 내의 단어의 수는 1,000개였다. 프레임 스키핑 레이트는 $N=3$ (가령, 모듈로 3)으로 설정되었다. 발화 디코더는 최선의 가능한 발화 인식 성능을 위한 보수적인 설정과 함께 제공된 빔 폭 발화 디코더(beam width speech decoder)였다. 도식된 바와 같이, 전체적인(가령, 특징 추출, 신경망 및/또는 값 근사계산 모듈, 그리고 발화 디코더) 시스템 성능은 약 800 MPCPS(초당 메가 사이클(mega cycles per second))의 계산 비용(가령, 시스템이 실시간으로 작동하기 위한 초당 중앙 처리 유닛(Central Processing Unit: CPU) 사이클의 요구되는 수)으로써 1초의 발화를 처리하는 데 2.49초라는 전수(brute force) 신경망 계산(가령, 프레임 스키핑도 수요 거리도 없음)에 대한 실시간 지수(Real Time Factor: RTF)를 제공하였다. 수요 거리만은 약 22%만큼 실시간 지수를 감소시켰고 프레임 스키핑만은 약 64%만큼 실시간 지수를 감소시켰다. 프레임 스키핑에다가 수요 거리를 적용하는 것은 약 22%의 추가적인 감소를 제공하였고 전체적인 시스템을 실시간 지수나 1.00이게 하였다. 주어진 예에서, 선형 외삽으로써의 프레임 스키핑 및 수요 거리의 조합은 자동화된 발화 인식 시스템으로 하여금 정확도의 손실 없이 실시간으로 작동할 수 있게 한다.

[0075]

나아가, 의사코드 (2)는 수요 거리와 조합된 선형 근사계산으로써 프레임 스키핑을 제공하기 위한 예시적 프로세스를 제공한다.

```

// l(t) nodes values of the uppermost hidden layer, filled for (t%N)==0
// d output index
// t time frame
// DNN structure containing the DNN
getOutput(d, t, l, DNN)
{
    m = t % N;
    if ( m==0 ) {
        if ( undefined o(t, d) ) o(t, d) = compute_output(DNN, l(t), d);
        return( o(t, d) );
    } // implicit else : a skipped frame
    s = ( t / N ) * N; // index of last computed frame
    p = s - N; // index of computed frame before frame s
    if (p<0) p = 0;
    if ( undefined o(s, d) ) o(s, d) = compute_output(DNN, l(s), d);
    if ( undefined o(p, d) ) o(p, d) = compute_output(DNN, l(p), d);
    return( linear_extrapolate(N, m, o(s, d), o(p, d) ) );
}

```

[0076]

[0077]

의사코드(1)에 도시된 바와 같이, 스킵되지 않는 프레임(가령, 평가 프레임)에 대해, 모듈로는 0일 수 있고 요청된 거리값은 신경망(가령, DNN)을 통하여 계산될 수 있다. 스킵되는 프레임에 대하여, 이전의 필요한 거리값은 신경망을 통하여 계산되고 현재의 거리값의 외삽(가령, linear_extrapolate)을 통한 근사계산을 위해 사용될 수 있다.

[0078]

논의된 바와 같이, 신경망(301), 제어기(302) 및 거리값 근사계산 모듈(303)을 포함하는 거리값 계산 모듈(204)은 발화 인식 정확도에서의 감소가 전혀 또는 별로 없이 자동 발화 인식 구현에서 계산상의 부하를 실질적으로 줄이기 위해 프레임 스킵핑 및/또는 수요 거리를 구현할 수 있다. 나아가, 신경망(301), 제어기(302) 및 거리값 근사계산 모듈(303)을 포함하는 그러한 거리값 계산 모듈(204)은 동일한 처리 디바이스(가령, 중앙 처리 유닛, 그래픽 처리 유닛, 신호 프로세서(signal processor), 또는 유사한 것)를 통하여 구현될 수 있거나 신경망의 다양한 부분은 상이한 처리 디바이스를 통하여 구현될 수 있다.

[0079]

도 10은 본 개시의 적어도 몇몇 구현에 따라 마련된, 특징(203) 및 출력 인덱스(208)로부터 거리값(205)을 생성하기 위해 프레임 스킵핑 및/또는 수요 거리를 구현하기 위한 예시적 시스템(1000)의 설명도이다. 도 10에 도시된 바와 같이, 시스템(1000)은 신경망(301), 제어기(302) 및 거리값 근사계산 모듈(303)을 가지는 거리값 계산 모듈(204) 및 신경망 데이터, 거리값(가령, 현재의 프레임에 관련된 이전의 거리값) 및 은닉층 값(가령, 현재의 프레임에 관련된 이전의 은닉층 값)을 저장하도록 구성된 메모리 스토어(memory store)(1001), 또는 유사한 것을 포함할 수 있다. 몇몇 예에서, 거리값 계산 모듈(204)은 본 문서에서 추가로 논의되는 바와 같이 중앙 처리 유닛 또는 다른 프로세서를 통하여 구현될 수 있다. 논의된 바와 같이, 몇몇 예에서, 거리값 계산 모듈(204) 모듈의 일부 또는 전부는 상이한 프로세서를 통하여 구현될 수 있다.

[0080]

도시된 바와 같이 그리고 본 문서에서 다른 데에서 논의된 바와 같이, 거리값 계산 모듈(204)은 (가령, 특징 추출 모듈(202)을 통하여) 특징(203)을 그리고 (가령, 발화 디코더 모듈(206)을 통하여) 출력 인덱스(208)를 수신할 수 있다. 거리값 계산 모듈(204)은 메모리 스토어(1001)를 통하여 신경망 가중치, 바이어스 및 정정(가령, 신경망 데이터)을 또한 수신할 수 있다. 나아가, 거리값 계산 모듈(204)은 이전의 거리값 및/또는 이전의 은닉층 값을 메모리 스토어(1001)를 통하여 수신할 수 있다. 예컨대, 특징(203)은 신경망(301)의 입력층에 입력을 제공할 수 있다. 신경망(301)은 신경망의 층을 위한 노드 스코어(score)를 판정할 수 있는 노드 스코어링 모듈(node scoring module), 그러한 노드 스코어를 생성된 바이어싱된 스코어로 바이어싱할 수 있는 스코어 바이어스 모듈(score bias module), 그리고 바이어싱된 스코어에 기반하여 노드를 위한 출력을 생성할 수 있는 출력/활성화 함수 모듈(output/activation function module)을 통하여 구현될 수 있다. 예컨대, 은닉층 노드 및 입력층 노드에 대해, 출력/활성화 함수 모듈은 출력을 생성하기 위해 활성화 함수를 구현할 수 있고 출력층 노드에 대해, 출력/활성화 함수 모듈은 정정된 바이어싱된 스코어를 노드 출력으로서 제공할 수 있다. 논의된 바와

같이, 몇몇 예에서, 출력층은 단지 출력 인덱스(208)와 연관된 거리값을 제공하기 위해 수요 거리 구현에서 제어될 수 있다.

[0081] 나아가, 거리값 근사계산 모듈(303)은 메모리 스토어(1001)로부터든 또는 신경망(301)으로부터든 신경망 판정된 거리값을 수신할 수 있고 거리값 근사계산 모듈(303)은, 예컨대, 수신된 신경망 판정된 거리값에 기반하여 거리값을 외삽할 수 있다. 신경망(301) 및/또는 거리값 근사계산 모듈(303)을 통하여 판정된 거리값은 본 문서에서 논의된 바와 같이 거리값(205)으로서 제공될 수 있다.

[0082] 도 11은 본 개시의 적어도 몇몇 구현에 따라 마련된, 자동 발화 인식을 제공하기 위한 예시적 프로세스(1100)를 보여주는 흐름도이다. 프로세스(1100)는 도 11에 예시된 바와 같이 하나 이상의 동작(1101 내지 1103)을 포함할 수 있다. 프로세스(1100)는 자동 발화 인식을 제공하기 위한 컴퓨터 구현된(computer-implemented) 방법의 적어도 일부를 형성할 수 있다. 비한정적인 예로서, 프로세스(1100)는 본 문서에서 논의된 바와 같이 시스템(200 또는 1000)에 의해 시행되는 발화 레코딩(201)과 같은 획득된 발화 레코딩을 위한 자동 발화 인식 프로세스의 적어도 일부를 형성할 수 있다. 또한, 프로세스(1100)는 도 12의 시스템(1200)을 참조하여 본 문서에 기술될 것이다.

[0083] 도 12는 본 개시의 적어도 몇몇 구현에 따라 마련된, 발화 인식을 제공하기 위한 예시적 시스템(1200)의 설명도이다. 도 12에 도시된 바와 같이, 시스템(1200)은 하나 이상의 중앙 처리 유닛(Central Processing Unit: CPU)(1201), 그래픽 처리 유닛(Graphics Processing Unit: GPU)(1202), 시스템 메모리(1203) 및 마이크(104)를 포함할 수 있다. 또한 도시된 바와 같이, CPU(1201)는 특징 추출 모듈(202), 거리값 계산 모듈(204) 및 발화 디코더 모듈(206)을 포함할 수 있다. 나아가, 거리값 계산 모듈(204)은 신경망(301), 제어기(302) 및 거리값 근사계산 모듈(303)을 포함할 수 있다. 도시된 바와 같이, 시스템(1200)의 예에서, 시스템 메모리(1203)는 메모리 스토어(1001)를 포함할 수 있는데, 이는 신경망 데이터, 거리값 및/또는 은닉층 값을 저장할 수 있다. 나아가, 시스템 메모리(1203)는 발화 레코딩, 특징, 특징 벡터, 거리값, 인식된 단어 시퀀스, 또는 유사한 것과 같이 본 문서에서 논의된 바와 같은 임의의 다른 데이터를 저장할 수 있다. 마이크(104)는, 발화(103)(가령, 공기 중의 음파와 같은 것, 도 1을 참조하시오)를 수신하고 발화(103)를 디지털 신호와 같은 전기 신호로 변환할 수 있는 임의의 적합한 디바이스 또는 디바이스들을 포함할 수 있다. 일 실시예에서, 마이크는 발화(103)를 발화 레코딩(201)으로 변환한다. 일 실시예에서, 발화 레코딩(201)은 CPU(1201)에 의한 액세스를 위해 시스템 메모리(1203) 내에 저장될 수 있다.

[0084] CPU(1201) 및 그래픽 처리 유닛(1202)은 본 문서에서 논의된 바와 같은 동작을 제공할 수 있는 임의의 수 및 유형의 처리 유닛을 포함할 수 있다. 그러한 동작은 소프트웨어 또는 하드웨어 또는 이의 조합을 통하여 구현될 수 있다. 예컨대, 그래픽 처리 유닛(1202)은 시스템 메모리(1203) 또는 전용 그래픽 메모리(도시되지 않음)로부터 획득된 데이터를 조작하는 데 전용인 회로를 포함할 수 있다. 나아가, 중앙 처리 유닛(1201)은 본 문서에서 논의된 바와 같은 동작뿐만 아니라 시스템(1200)을 위한 제어 및 다른 고수준 기능을 제공할 수 있는 임의의 수 및 유형의 처리 유닛 또는 모듈을 포함할 수 있다. 시스템 메모리(1203)는 휘발성 메모리(가령, 정적 랜덤 액세스 메모리(Static Random Access Memory: SRAM), 동적 랜덤 액세스 메모리(Dynamic Random Access Memory: DRAM) 등등) 또는 비휘발성 메모리(가령, 플래시 메모리(flash memory) 등등) 및 기타 등등과 같은 임의의 유형의 메모리일 수 있다. 비한정적인 예에서, 시스템 메모리(1203)는 캐시 메모리(cache memory)에 의해 구현될 수 있다. 도시된 바와 같이, 일 실시예에서, 특징 추출 모듈(202), 거리값 계산 모듈(204) 및 발화 디코더 모듈(206)은 CPU(1201)를 통하여 구현될 수 있다. 몇몇 예에서, 특징 추출 모듈(202), 거리값 계산 모듈(204) 및 발화 디코더 모듈(206)은 CPU(1201)를 통하여 구현된 바와 같은 소프트웨어에 의해 제공될 수 있다. 다른 예에서, 특징 추출 모듈(202), 거리값 계산 모듈(204) 및 발화 디코더 모듈(206) 중 하나 이상은 디지털 신호 프로세서 또는 유사한 것을 통하여 구현될 수 있다. 다른 실시예에서, 특징 추출 모듈(202), 거리값 계산 모듈(204) 및 발화 디코더 모듈(206) 중 하나 이상은 그래픽 처리 유닛(1202)의 실행 유닛(Execution Unit: EU)을 통하여 구현될 수 있다. EU는, 예컨대, 광범위한 배열의 프로그램가능(programmable) 로직 기능을 제공할 수 있는 로직 코어 또는 코어들과 같은 프로그램가능 로직 또는 회로를 포함할 수 있다.

[0085] 도 11의 논의로 되돌아가면, 프로세스(1100)는 동작(1101), "신경망의 출력으로서 제1 거리값을 판정하기 위해 신경망을 평가함"(Evaluate a Neural Network to Determine a First Distance Value as an Output of the Neural Network)에서 시작할 수 있는데, 여기서 신경망의 출력으로서 제1 거리값을 판정하기 위해 신경망이 평가될 수 있다. 예컨대, 제1 거리값은 제1 시간 인스턴스와 연관될 수 있다. 예컨대, CPU(1201)를 통하여 구현된 바와 같은 거리값 계산 모듈(204)의 신경망(301)은 신경망(301)의 출력으로서 거리값을 판정할 수 있다.

- [0086] 몇몇 예에서, 프로세스(1100)는 본 문서에서 논의된 바와 같이 프레임 스키핑 기법을 구현하나 수요 거리 기법은 그렇지 않을 수 있다. 그러한 예에서, 신경망은 여러 출력층 노드를 가지는 출력층을 포함할 수 있고 동작(1101)에서 신경망을 평가하는 것은 신경망의 출력 노드 전부를 평가하는 것을 포함할 수 있다. 다른 예에서, 프로세스(1100)는 프레임 스키핑과 함께 또는 프레임 스키핑 없이 수요 거리 기법을 구현할 수 있다. 그러한 예에서, 동작(1101) 전에, (가령, 발화 디코더를 통하여) 제1 시간 인스턴스에 대해 출력 인덱스가 생성될 수 있다. 예컨대, 제1 거리값은 출력 인덱스 중의 출력 인덱스와 연관될 수 있다. 그러한 예에서, 신경망은 여러 출력층 노드를 가지는 출력층을 포함할 수 있어서 제1 시간 인스턴스에 대해 신경망을 평가하는 것은 여러 출력층 노드의 서브세트를 평가하는 것을 포함하니 그 서브세트는 출력 인덱스와 연관된다. 논의된 바와 같이, 그러한 예에서, 최종 은닉층 노드를 가지는 신경망의 최종 은닉층은 제1 시간 인스턴스에 대해 전적으로 평가될 수 있고 최종 은닉층 노드 값은 세이브될 수 있다.
- [0087] 프로세스(1100)는 동작(1102), "신경망 판정된 거리값에 기반하여 차후의 시간 인스턴스에 대해 제2 거리값을 근사계산함"(Approximate a Second Distance Value for a Subsequent Time Instance based on the Neural Network Determined Distance Value)에서 계속될 수 있는데, 여기서, 제1 시간 인스턴스에 후속하는 제2 시간 인스턴스에 대해, 제1 거리값에 적어도 부분적으로 기반하여 제2 거리값이 근사계산될 수 있고 그리하여 신경망은 제2 시간 인스턴스에 대해 평가되지 않는다. 예컨대, 제2 시간 인스턴스에서, 어떤 거리값도 신경망의 직접적인 출력이 아닐 수 있다. 예컨대, CPU(1201)를 통하여 구현된 바와 같은 거리값 계산 모듈(204)의 거리값 근사계산 모듈(303)이 제2 거리값을 근사계산할 수 있다. 제2 거리값은 선형 함수, 비선형 함수, 분산 함수, 또는 유사한 것을 기반으로 외삽을 통하여 근사계산될 수 있다. 몇몇 예에서, 제2 거리값의 근사계산은 제1 거리값 및 제1 거리값과 연관된 시간 인스턴스에 선행하는 시간 인스턴스로부터의 제3 거리값을 사용하는 외삽에 기반할 수 있다. 몇몇 예에서, 제1 및 제3 거리값에 기반한 그러한 외삽은 선형 외삽에 기반할 수 있고 제3 거리값은 신경망을 통하여 사전에 판정되었을 수 있다. 몇몇 예에서, 그러한 외삽은 제1 거리값 및 제3 거리값 간의 차이의 절반에 제1 거리값이 더해진 것을 선형 외삽이 포함하도록 수학적 식 1에 관해서 도시된 바와 같이 제공될 수 있다.
- [0088] 논의된 바와 같이, 몇몇 예에서, 프레임 스키핑 및 수요 거리는 함께 구현될 수 있다. 그러한 예에서, 제2 거리값을 근사계산하기 전에, 제2 시간 인스턴스에 대해 (가령, 발화 디코더를 통하여) 출력 인덱스가 생성될 수 있다. 예컨대, 제2 거리값은 출력 인덱스 중의 출력 인덱스와 연관될 수 있다. 그러한 예에서, 제1 시간 인스턴스에서 신경망을 평가하는 것은 세이브될 수 있는 최종 은닉층 값을 생성하기 위해 최종 은닉층의 모든 최종 은닉층 노드를 평가하는 것을 포함할 수 있다. 제2 시간 인스턴스에서 제1 시간 인스턴스에 대한 신경망은, 제2 거리값을 근사계산하는 데에 사용될 수 있는 신경망 판정된 거리값을 판정하기 위해 제1 시간 인스턴스에 대한 세이브된 은닉층 값에 기반하여 제2 시간 인스턴스에 대한 출력 인덱스와 연관된 출력층 노드를 평가함으로써 재평가될 수 있다.
- [0089] 프로세스(1100)는 동작(1103), "제1 및 제2 거리값에 기반하여 텍스트 요소의 시퀀스를 판정함"(Determine a Sequence of Textual Elements based on the First and Second Distance Values)에서 계속될 수 있는데, 여기서 제1 및 제2 거리값에 기반하여 텍스트 요소의 시퀀스가 판정될 수 있다. 예컨대, CPU(1201)를 통하여 구현된 바와 같은 발화 디코더 모듈(206)은 본 문서에서 논의된 바와 같은 인식된 단어 시퀀스(207)를 판정할 수 있다. 몇몇 예에서, 발화 디코더는 비터비 빔 탐색 디코더를 포함한다.
- [0090] 논의된 바와 같이, 프로세스(1100)는 프레임 스키핑 기법을 구현할 수 있다. 그러한 기법은 임의의 수의 프레임을 스킵할 수 있다. 예컨대, 신경망을 사용하여 판정된 프레임은 신경망 평가 프레임으로서 기술될 수 있고 근사계산 기법을 사용하여 판정된 프레임은 스킵 프레임으로서 기술될 수 있다. 몇몇 예에서, 제1 시간 인스턴스는 신경망 평가 프레임과 연관될 수 있고, 제2 시간 인스턴스는 스킵 프레임과 연관될 수 있으며, 평가 프레임 및 스킵 프레임은 인접한 프레임일 수 있다. 다른 예에서, 평가 프레임 및 스킵 프레임 사이에 1개, 2개, 3개, 또는 더 많은 스킵 프레임이 있을 수 있다. 나아가, 몇몇 예에서, (가령, 정확도 필요성, 이용가능한 컴퓨팅 리소스, 또는 현재의 실시간 지수에 기반하는) 프레임 스키핑 레이트가 평가 프레임 사이로부터 스킵 프레임을 제거하거나 추가적인 스킵 프레임을 제공하기 위해 CPU(1201)를 통하여 구현된 바와 같은 거리값 계산 모듈(204)의 제어기(302)를 통하여 판정되고 구현될 수 있다.
- [0091] 프로세스(1100)는 임의의 수의 시간 인스턴스 및/또는 발화 레코딩에 대해 직렬로 또는 병렬로 임의의 횟수로 반복될 수 있다. 프로세스(1100)는 본 문서에서 논의된 바와 같은 디바이스(102)와 같은 디바이스를 통하여 또는 클라우드 컴퓨팅 플랫폼 또는 유사한 것의 일부로서의 서버를 통하여 거리값을 판정하는 것 및 텍스트 요소의 시퀀스를 생성하는 것을 가능케 할 수 있다. 또한 본 문서에서 논의된 바와 같이, 실시간으로 그러한 처리를

하기 전에, 신경망의 다양한 컴포넌트는 사전훈련될 수 있거나, 바이어스 및/또는 가중치가 판정될 수 있거나, 기타 등등인데, 몇몇 예에서, 별개의 시스템을 통하여 그러하다. 논의된 바와 같이, 몇몇 예에서, 프로세서(1100)는 CPU(1201)를 통하여 구현될 수 있다. 다른 예에서, 프로세서(1100)(그리고 연관된 모듈)는 코프로세서(co-processor) 또는 유사한 것과 같은 전용 프로세서를 통하여 구현될 수 있다.

[0092] 나아가, 동작(1101) 전에, 몇몇 예에서, 수신된 발화가 발화 레코딩으로 변환될 수 있다. 예컨대, 발화(103)는 시스템(1100)의 마이크(104) 및/또는 관련된 회로를 통하여 발화 레코딩(201)으로 변환될 수 있다. 나아가, 특징(203)(가령, 특징 벡터)은 CPU(1201)를 통하여 구현된 바와 같은 특징 추출 모듈(202)에 의해 발화 레코딩(201)에 기반하여 판정되거나 추출될 수 있고 그러한 특징은 CPU(1201)를 통하여 구현된 바와 같은 거리값 계산 모듈(204)의 신경망(301)에 제공될 수 있다. 몇몇 예에서, 특징 추출 모듈(202)은 시스템(1200)의 디지털 신호 프로세서(도시되지 않음)를 통하여 구현될 수 있다. 몇몇 예에서, CPU(1201)를 통하여 구현된 바와 같은 발화 디코더 모듈(206)은 거리값(205)을 시스템 메모리(1203)를 통하여 획득된 바와 같은 통계적 모델(도시되지 않음)과 비교함으로써 인식된 단어 시퀀스(207)를 판정할 수 있다.

[0093] 본 문서 내에 논의된 예시적 프로세스의 구현은 예시된 순서로 도시된 모든 동작의 시행을 포함할 수 있으나, 본 개시는 이 점에 한정되지 않고, 다양한 예에서, 본 문서 내의 예시적 프로세스의 구현은 도시된 동작의 단지 서브세트, 예시된 것과는 상이한 순서로 수행되는 동작, 또는 추가적인 동작을 포함할 수 있다.

[0094] 추가로, 하나 이상의 컴퓨터 프로그램 제품(computer program product)에 의해 제공되는 명령어에 응답하여 본 문서 내에 논의된 동작 중 임의의 하나 이상이 시행될 수 있다. 그러한 프로그램 제품은, 예컨대 프로세서에 의해 실행되는 경우, 본 문서 내에 기술된 기능을 제공할 수 있는 명령어를 제공하는 신호 함유 매체(signal bearing medium)를 포함할 수 있다. 컴퓨터 프로그램 제품은 하나 이상의 머신 판독가능 매체(machine-readable medium)의 임의의 형태로 제공될 수 있다. 그러므로, 예컨대, 하나 이상의 그래픽 처리 유닛(들) 또는 프로세서 코어(들)를 포함하는 프로세서는 하나 이상의 머신 판독가능 매체에 의해 프로세서에 전달되는 프로그램 코드 및/또는 명령어 또는 명령어 세트에 응답하여 본 문서 내의 예시적 프로세스의 블록 중 하나 이상을 시행할 수 있다. 일반적으로, 머신 판독가능 매체는 본 문서 내에 기술된 디바이스 및/또는 시스템 중 임의의 것으로 하여금 시스템(200, 1000 또는 1200)의 적어도 일부분, 또는 본 문서 내에 논의된 바와 같은 임의의 다른 모듈 또는 컴포넌트를 구현하게 할 수 있는 프로그램 코드 및/또는 명령어 또는 명령어 세트의 형태로 소프트웨어를 전달할 수 있다.

[0095] 본 문서 내에 기술된 임의의 구현에서 사용되는 바와 같이, 용어 "모듈"은 본 문서 내에 기술된 기능을 제공하도록 구성된 소프트웨어 로직, 펌웨어 로직, 하드웨어 로직 및/또는 회로의 임의의 조합을 나타낸다. 소프트웨어는 소프트웨어 패키지, 코드 및/또는 명령어 세트 또는 명령어로서 실체화될(embodied) 수 있고, "하드웨어"는, 본 문서 내에 기술된 임의의 구현에서 사용되는 바와 같이, 예컨대, 고정배선된 회로(hardwired circuitry), 프로그램가능 회로(programmable circuitry), 상태 머신 회로(state machine circuitry), 고정 기능 회로(fixed function circuitry), 실행 유닛 회로(execution unit circuitry), 그리고/또는 프로그램가능 회로에 의해 실행되는 명령어를 저장하는 펌웨어를, 단독으로 또는 임의의 조합으로, 포함할 수 있다. 모듈들은, 집합적으로 또는 개별적으로, 더 큰 시스템, 예컨대, 집적 회로(Integrated Circuit: IC), 시스템 온 칩(System on-Chip: SoC) 및 기타 등등의 일부를 형성하는 회로로서 실체화될 수 있다.

[0096] 도 13은 본 개시의 적어도 몇몇 구현에 따라 마련된 예시적 시스템(1300)의 설명도이다. 다양한 구현에서, 시스템(1300)은 모바일 시스템(mobile system)일 수 있는데 다만 시스템(1300)은 이 맥락에 한정되지 않는다. 예컨대, 시스템(1300)은 개인용 컴퓨터(Personal Computer: PC), 랩톱 컴퓨터(laptop computer), 울트라 랩톱 컴퓨터(ultra-laptop computer), 태블릿(tablet), 터치 패드(touch pad), 휴대가능 컴퓨터(portable computer), 핸드헬드 컴퓨터(handheld computer), 팜톱 컴퓨터(palmtop computer), 개인용 디지털 보조기기(Personal Digital Assistant: PDA), 셀룰러 전화(cellular telephone), 셀룰러 전화/PDA 조합, 텔레비전, 스마트 디바이스(smart device)(가령, 스마트폰, 스마트 태블릿, 스마트 watch, 스마트 안경 또는 스마트 텔레비전), 모바일 인터넷 디바이스(Mobile Internet Device: MID), 메시징 디바이스(messaging device), 데이터 통신 디바이스(data communication device), 카메라(가령, 보고 찍는 카메라(point-and-shoot camera), 슈퍼 줌 카메라(super-zoom camera), 디지털 단일 렌즈 반사(Digital Single-Lens Reflex: DSLR) 카메라) 및 기타 등등 내에 포함될 수 있다.

[0097] 다양한 구현에서, 시스템(1300)은 디스플레이(1320)에 커플링된(coupled) 플랫폼(1302)을 포함한다. 플랫폼(1302)은 콘텐츠 서비스 디바이스(들)(1330) 또는 콘텐츠 전달 디바이스(들)(1340) 또는 다른 유사한 콘텐츠 소

스와 같은 콘텐츠 디바이스로부터 콘텐츠를 수신할 수 있다. 도시된 바와 같이, 몇몇 예에서, 시스템(1300)은 플랫폼(1302)을 통하여 구현된 마이크(104)를 포함할 수 있다. 플랫폼(1302)은 본 문서에 논의된 바와 같이 마이크(104)를 통하여 발화(103)와 같은 발화를 수신할 수 있다. 예컨대 플랫폼(1302) 및/또는 디스플레이(1320)와 상호작용하는 데에 하나 이상의 내비게이션 특징부(navigation feature)를 포함하는 내비게이션 제어기(1350)가 사용될 수 있다. 이 컴포넌트들 각각은 아래에서 더욱 상세히 기술된다.

- [0098] 다양한 구현에서, 플랫폼(1302)은 칩셋(chipset)(1305), 프로세서(1310), 메모리(1312), 안테나(1313), 스토리지(storage)(1314), 그래픽 서브시스템(graphics subsystem)(1315), 애플리케이션(1316) 및/또는 무선기기(radio)(1318)의 임의의 조합을 포함할 수 있다. 칩셋(1305)은 프로세서(1310), 메모리(1312), 스토리지(1314), 그래픽 서브시스템(1315), 애플리케이션(1316) 및/또는 무선기기(1318) 사이의 상호통신(intercommunication)을 제공할 수 있다. 예컨대, 칩셋(1305)은 스토리지(1314)와의 상호통신을 제공할 수 있는 스토리지 어댑터(storage adapter)(묘사되지 않음)를 포함할 수 있다.
- [0099] 프로세서(1310)는 복합 명령어 세트 컴퓨터(Complex Instruction Set Computer: CISC) 또는 축소 명령어 세트 컴퓨터(Reduced Instruction Set Computer: RISC) 프로세서, x86 명령어 세트 호환가능 프로세서, 다중 코어(multi-core), 또는 임의의 다른 마이크로프로세서 또는 중앙 처리 유닛(Central Processing Unit: CPU)으로서 구현될 수 있다. 다양한 구현에서, 프로세서(1310)는 이중 코어(dual-core) 프로세서(들), 이중 코어 모바일 프로세서(들) 및 기타 등등일 수 있다.
- [0100] 메모리(1312)는, 랜덤 액세스 메모리(Random Access Memory: RAM), 동적 랜덤 액세스 메모리(Dynamic Random Access Memory: DRAM), 또는 정적 RAM(Static RAM: SRAM)과 같은 것이나 이에 한정되지 않는 휘발성 메모리 디바이스로서 구현될 수 있다.
- [0101] 스토리지(1314)는, 자기 디스크 드라이브(magnetic disk drive), 광학 디스크 드라이브(optical disk drive), 테이프 드라이브(tape drive), 내부 저장 디바이스(internal storage device), 부착 저장 디바이스(attached storage device), 플래시 메모리(flash memory), 배터리 예비 동기식 DRAM(battery backed-up SDRAM(Synchronous DRAM)) 및/또는 네트워크 액세스가능 저장 디바이스(network accessible storage device)와 같은 것이나 이에 한정되지 않는 비휘발성 저장 디바이스로서 구현될 수 있다. 다양한 구현에서, 예컨대, 스토리지(1314)는 여러 하드 드라이브가 포함된 경우 가치 있는 디지털 매체를 위한 저장 성능 향상된 보호를 증가시키는 기술을 포함할 수 있다.
- [0102] 그래픽 서브시스템(1315)은 디스플레이를 위한 정지(still) 또는 비디오와 같은 이미지의 처리를 수행할 수 있다. 그래픽 서브시스템(1315)은 예컨대 그래픽 처리 유닛(Graphics Processing Unit: GPU) 또는 시각적 처리 유닛(Visual Processing Unit: VPU)일 수 있다. 그래픽 서브시스템(1315) 및 디스플레이(1320)를 통신가능하게 커플링하는 데에 아날로그 또는 디지털 인터페이스가 사용될 수 있다. 예컨대, 인터페이스는 고선명 멀티미디어 인터페이스(High-Definition Multimedia Interface), 디스플레이포트(DisplayPort), 무선 HDMI 및/또는 무선 HD 준수 기법 중 임의의 것일 수 있다. 그래픽 서브시스템(1315)은 프로세서(1310) 또는 칩셋(1305) 내로 통합될 수 있다. 몇몇 구현에서, 그래픽 서브시스템(1315)은 칩셋(1305)에 통신가능하게 커플링된 독립형(stand-alone) 디바이스일 수 있다.
- [0103] 본 문서 내에 기술된 그래픽 및/또는 비디오 처리 기법은 다양한 하드웨어 아키텍처 내에 구현될 수 있다. 예컨대, 그래픽 및/또는 비디오 기능은 칩셋 내에 통합될 수 있다. 대안적으로, 별도의 그래픽 및/또는 비디오 프로세서가 사용될 수 있다. 또 다른 구현으로서, 그래픽 및/또는 비디오 기능은 다중코어 프로세서를 포함하는 일반 목적 프로세서에 의해 제공될 수 있다. 추가의 실시예에서, 그 기능은 가전 디바이스(consumer electronics device) 내에 구현될 수 있다.
- [0104] 무선기기(1318)는 다양한 적합한 무선 통신 기법을 사용하여 신호를 송신하고 수신할 수 있는 하나 이상의 무선 기기를 포함할 수 있다. 그러한 기법은 하나 이상의 무선 네트워크에 걸친 통신을 수반할 수 있다. 예시적 무선 네트워크는 무선 로컬 영역 네트워크(Wireless Local Area Network: WLAN), 무선 개인 영역 네트워크(Wireless Personal Area Network: WPAN), 무선 대도시 영역 네트워크(Wireless Metropolitan Area Network: WMAN), 셀룰러 네트워크 및 위성 네트워크를 포함한다(그러나 이에 한정되지 않음). 그러한 네트워크에 걸쳐 통신하는 중에, 무선기기(1318)는 임의의 버전의 하나 이상의 적용가능한 표준에 따라 동작할 수 있다.
- [0105] 다양한 구현에서, 디스플레이(1320)는 임의의 텔레비전 유형 모니터 또는 디스플레이를 포함할 수 있다. 디스플레이(1320)는, 예컨대, 컴퓨터 디스플레이 스크린, 터치 스크린 디스플레이, 비디오 모니터, 텔레비전과 비슷한

디바이스 및/또는 텔레비전을 포함할 수 있다. 디스플레이(1320)는 디지털 및/또는 아날로그일 수 있다. 다양한 구현에서, 디스플레이(1320)는 홀로그램 디스플레이(holographic display)일 수 있다. 또한, 디스플레이(1320)는 시각적 투영(visual projection)을 수신할 수 있는 투명한 표면(transparent surface)일 수 있다. 그러한 투영은 다양한 형태의 정보, 이미지 및/또는 객체를 전달할 수 있다. 예컨대, 그러한 투영은 모바일 증강 현실(Mobile Augmented Reality: MAR) 애플리케이션을 위한 시각적 오버레이(visual overlay)일 수 있다. 하나 이상의 소프트웨어 애플리케이션(1316)의 제어 하에서, 플랫폼(1302)은 디스플레이(1320) 상에 사용자 인터페이스(user interface)(1322)를 디스플레이할 수 있다.

[0106] 다양한 구현에서, 예컨대, 콘텐츠 서비스 디바이스(들)(1330)는 임의의 전국적, 국제적 및/또는 독립적 서비스에 의해 호스팅되고(hosted) 따라서 플랫폼(1302)이 인터넷을 통하여 액세스할 수 있는 것일 수 있다. 콘텐츠 서비스 디바이스(들)(1330)는 플랫폼(1302)에 그리고/또는 디스플레이(1320)에 커플링될 수 있다. 플랫폼(1302) 및/또는 콘텐츠 서비스 디바이스(들)(1330)는 미디어 정보(media information)를 네트워크(1360)에 그리고 네트워크(1360)로부터 통신(가령, 발신 및/또는 수신)하기 위해 네트워크(1360)에 커플링될 수 있다. 콘텐츠 전달 디바이스(들)(1340)는 플랫폼(1302)에 그리고/또는 디스플레이(1320)에 또한 커플링될 수 있다.

[0107] 다양한 구현에서, 콘텐츠 서비스 디바이스(들)(1330)는 케이블 텔레비전 박스(cable television box)와, 개인용 컴퓨터와, 네트워크와, 전화와, 디지털 정보 및/또는 콘텐츠를 전달할 수 있는 인터넷 가능화된(Internet enabled) 디바이스 또는 기기, 네트워크(1360)를 통하여 또는 직접적으로 콘텐츠 제공자 및 플랫폼(1302) 및/또는 디스플레이(1320) 간에 콘텐츠를 단방향으로(uni-directionally) 또는 양방향으로(bi-directionally) 통신할 수 있는 임의의 다른 유사한 디바이스를 포함할 수 있다. 콘텐츠는 네트워크(1360)를 통하여 콘텐츠 제공자 및 시스템(1300) 내의 컴포넌트 중 임의의 하나로 그리고 이로부터 단방향으로 그리고/또는 양방향으로 통신될 수 있음이 인식될 것이다. 콘텐츠의 예는, 예컨대, 비디오, 음악, 의료 및 게이밍(gaming) 정보, 그리고 기타 등등을 포함하는 임의의 미디어 정보를 포함할 수 있다.

[0108] 콘텐츠 서비스 디바이스(들)(1330)는 미디어 정보, 디지털 정보 및/또는 다른 콘텐츠를 포함하는 케이블 텔레비전 프로그래밍(cable television programming)과 같은 콘텐츠를 수신할 수 있다. 콘텐츠 제공자의 예는 임의의 케이블 또는 위성 텔레비전 또는 라디오 또는 인터넷 콘텐츠 제공자를 포함할 수 있다. 제공된 예는 어떤 식으로든 본 개시에 따른 구현을 한정하도록 의도되지 않는다.

[0109] 다양한 구현에서, 플랫폼(1302)은 하나 이상의 내비게이션 특징부를 갖는 내비게이션 제어기(1350)로부터 제어 신호를 수신할 수 있다. 제어기(1350)의 내비게이션 특징부는 예컨대 사용자 인터페이스(1322)와 상호작용하는 데에 사용될 수 있다. 다양한 실시예에서, 내비게이션 제어기(1350)는 사용자로 하여금 공간적(가령, 연속적이고 다차원적인) 데이터를 컴퓨터 내에 입력할 수 있게 하는 컴퓨터 하드웨어 컴포넌트(특히, 인간 인터페이스 디바이스(human interface device))일 수 있는 포인팅 디바이스(pointing device)일 수 있다. 많은 시스템, 예를 들어 그래픽 사용자 인터페이스(Graphical User Interface: GUI)와, 텔레비전과 모니터는 사용자로 하여금 물리적 제스처를 사용하여 컴퓨터 또는 텔레비전을 제어하고 이에 데이터를 제공할 수 있게 한다.

[0110] 제어기(1350)의 내비게이션 특징부의 움직임은 디스플레이(가령, 디스플레이(1320)) 상에 포인터(pointer), 커서(cursor), 포커스 링(focus ring), 또는 디스플레이 상에 디스플레이되는 다른 시각적 지시자(visual indicator)의 움직임에 의해 모사될(replicated) 수 있다. 예컨대, 소프트웨어 애플리케이션(1316)의 제어 하에서, 내비게이션 제어기(1350) 상에 위치된 내비게이션 특징부는 예컨대 사용자 인터페이스(1322) 상에 디스플레이된 가상 내비게이션 특징부로 맵핑될(mapped) 수 있다. 다양한 실시예에서, 제어기(1350)는 별개의 컴포넌트가 아닐 수 있으나 플랫폼(1302) 및/또는 디스플레이(1320) 내로 통합될 수 있다. 그러나, 본 개시는 본 문서 내에 도시되거나 기술된 요소로 또는 맥락 내에 한정되지 않는다.

[0111] 다양한 구현에서, 드라이버(도시되지 않음)는 사용자로 하여금, 예컨대, 가능화된 경우, 초기 시동(boot-up) 후에 버튼의 터치로써 텔레비전과 같이 플랫폼(1302)을 즉시 켜고 끌 수 있게 하기 위한 기술을 포함할 수 있다. 프로그램 로직은 플랫폼(1302)으로 하여금 플랫폼이 "꺼진" 경우에도 콘텐츠를 미디어 어댑터 또는 다른 콘텐츠 서비스 디바이스(들)(1330) 또는 콘텐츠 전달 디바이스(들)(1340)에 스트리밍(stream)할 수 있도록 할 수 있다. 추가로, 칩셋(1305)은 예컨대 5.1 서라운드 사운드 오디오(5.1 surround sound audio) 및/또는 고선명 7.1 서라운드 사운드 오디오(high definition 7.1 surround sound audio)를 위한 하드웨어 및/또는 소프트웨어 지원을 포함할 수 있다. 드라이버는 통합된 그래픽 플랫폼을 위한 그래픽 드라이버(graphics driver)를 포함할 수 있다. 다양한 실시예에서, 그래픽 드라이버는 주변 컴포넌트 상호연결 익스프레스 그래픽 카드(peripheral component interconnect (PCI) Express graphics card)를 포함할 수 있다.

- [0112] 다양한 구현에서, 시스템(1300) 내에 도시된 컴포넌트 중 임의의 하나 이상이 통합될 수 있다. 예컨대, 가령, 플랫폼(1302) 및 콘텐츠 서비스 디바이스(들)(1330)가 통합될 수 있거나, 플랫폼(1302) 및 콘텐츠 전달 디바이스(들)(1340)가 통합될 수 있거나, 플랫폼(1302), 콘텐츠 서비스 디바이스(들)(1330) 및 콘텐츠 전달 디바이스(들)(1340)가 통합될 수 있다. 다양한 실시예에서, 플랫폼(1302) 및 디스플레이(1320)는 통합된 유닛(integrated unit)일 수 있다. 예컨대, 디스플레이(1320) 및 콘텐츠 서비스 디바이스(들)(1330)는 통합될 수 있거나, 디스플레이(1320) 및 콘텐츠 전달 디바이스(들)(1340)는 통합될 수 있다. 이 예들은 본 개시를 한정하도록 의도되지 않는다.
- [0113] 다양한 실시예에서, 시스템(1300)은 무선 시스템, 유선 시스템, 또는 양자 모두의 조합으로서 구현될 수 있다. 무선 시스템으로서 구현되는 경우, 시스템(1300)은, 하나 이상의 안테나, 송신기, 수신기, 송수신기, 증폭기, 필터, 제어 로직 및 기타 등등과 같은, 무선 공유 매체(wireless shared media) 상에서 통신하는 데에 적합한 컴포넌트 및 인터페이스를 포함할 수 있다. 무선 공유 매체의 일례는 무선 스펙트럼(wireless spectrum)의 부분, 예를 들어 RF 스펙트럼 및 기타 등등을 포함할 수 있다. 유선 시스템으로서 구현되는 경우, 시스템(1300)은, 입력/출력(Input/Output: I/O) 어댑터, I/O 어댑터를 대응하는 유선 통신 매체와 연결하기 위한 물리적 커넥터(physical connector), 네트워크 인터페이스 카드(Network Interface Card: NIC), 디스크 제어기, 비디오 제어기, 오디오 제어기 및 유사한 것과 같은, 유선 통신 매체 상에서 통신하는 데에 적합한 컴포넌트 및 인터페이스를 포함할 수 있다. 유선 통신 매체의 예는 배선(wire), 케이블(cable), 금속 리드(metal lead), 인쇄 회로 보드(Printed Circuit Board: PCB), 백플레인(backplane), 스위치 패브릭(switch fabric), 반도체 재료(semiconductor material), 꼬임쌍선(twisted-pair wire), 동축 케이블(co-axial cable), 광섬유(fiber optics) 및 기타 등등을 포함할 수 있다.
- [0114] 플랫폼(1302)은 정보를 통신하기 위해 하나 이상의 논리적 또는 물리적 채널을 수립할 수 있다. 정보는 미디어 정보 및 제어 정보를 포함할 수 있다. 미디어 정보는 사용자를 위해 의도된 콘텐츠를 나타내는 임의의 데이터를 나타낼 수 있다. 콘텐츠의 예는, 예컨대, 음성 대화(voice conversation), 비디오컨퍼런스(videoconference), 스트리밍 비디오(streaming video), 전자 메일("이메일"(email)) 메시지, 음성 메일 메시지, 문자숫자 심볼(alphanumeric symbol), 그래픽, 이미지, 비디오, 텍스트 및 기타 등등으로부터의 데이터를 포함할 수 있다. 음성 대화로부터의 데이터는, 예컨대, 발화 정보(speech information), 묵음 기간(silence period), 배경 잡음(background noise), 안정 잡음(comfort noise), 톤 및 기타 등등일 수 있다. 제어 정보는 자동화된 시스템을 위해 의도된 커맨드(command), 명령어 또는 제어 워드(control word)를 표현하는 임의의 데이터를 나타낼 수 있다. 예컨대, 제어 정보는 사전결정된 방식으로 미디어 정보를 처리하도록 노드(node)에 지시하거나, 시스템을 통해 미디어 정보를 라우팅하는(route) 데에 사용될 수 있다. 그러나, 실시예는 도 13에 도시되거나 기술된 맥락 내에 또는 요소로 한정되지 않는다.
- [0115] 위에서 기술된 바와 같이, 시스템(1300)은 갖가지 물리적 스타일 또는 폼 팩터(form factor)로 실체화될 수 있다. 도 13은 시스템(1300)이 실체화될 수 있는 소형 폼 팩터 디바이스(1300)의 구현을 보여준다. 다양한 실시예에서, 예컨대, 디바이스(1300)는 무선 능력을 갖는 모바일 컴퓨팅 디바이스로서 구현될 수 있다. 모바일 컴퓨팅 디바이스는 예컨대 처리 시스템 및 모바일 전력 소스 또는 공급부, 예를 들어 하나 이상의 배터리를 가지는 임의의 디바이스를 나타낼 수 있다. 몇몇 실시예에서, 디바이스(1300)는 본 문서에 기술된 바와 같은 신경망의 구현을 통한 실시간 발화 인식을 위해 마이크(가령, 마이크(104))를 포함하고/하거나 발화(가령, 발화(103))를 수신할 수 있다.
- [0116] 위에서 기술된 바와 같이, 모바일 컴퓨팅 디바이스의 예는 개인용 컴퓨터(Personal Computer: PC), 랩톱 컴퓨터, 울트라 랩톱 컴퓨터, 태블릿, 터치 패드, 휴대가능 컴퓨터, 핸드헬드 컴퓨터, 팜톱 컴퓨터, 개인용 디지털 보조기기(Personal Digital Assistant: PDA), 셀룰러 전화, 셀룰러 전화/PDA 조합, 텔레비전, 스마트 디바이스(가령, 스마트폰, 스마트 태블릿 또는 스마트 텔레비전), 모바일 인터넷 디바이스(Mobile Internet Device: MID), 메시징 디바이스, 데이터 통신 디바이스, 카메라(가령, 보고 찍는 카메라, 수퍼 줌 카메라, 디지털 단일 렌즈 반사(Digital Single-Lens Reflex: DSLR) 카메라) 및 기타 등등을 포함할 수 있다.
- [0117] 모바일 컴퓨팅 디바이스의 예는 또한 사람이 착용하도록 마련된 컴퓨터, 예를 들어 손목 컴퓨터(wrist computer), 손가락 컴퓨터(finger computer), 반지 컴퓨터(ring computer), 안경 컴퓨터(eyeglass computer), 벨트 클립 컴퓨터(belt-clip computer), 암 밴드 컴퓨터(arm-band computer), 신발 컴퓨터(shoe computer), 의복 컴퓨터(clothing computer) 및 다른 착용가능 컴퓨터(wearable computer)를 포함할 수 있다. 다양한 실시예에서, 예컨대, 모바일 컴퓨팅 디바이스는 음성 통신 및/또는 데이터 통신뿐만 아니라, 컴퓨터 애플리케이션을 실행하는 것이 가능한 스마트폰으로서 구현될 수 있다. 몇몇 실시예는 예로서 스마트폰으로서 구현된 모바일 컴

퓨팅 디바이스로써 기술될 수 있으나, 다른 무선 모바일 컴퓨팅 디바이스를 사용하여 다른 실시예가 구현될 수도 있음이 인식될 수 있다. 실시예는 이 맥락 내에 한정되지 않는다.

[0118] 도 14에 도시된 바와 같이, 디바이스(1400)는 하우징(housing)(1402), 디스플레이(1404), 입력/출력(Input/Output: I/O) 디바이스(1406) 및 안테나(1408)를 포함할 수 있다. 디바이스(1400)는 또한 내비게이션 특징부(1412)를 포함할 수 있다. 디스플레이(1404)는 모바일 컴퓨팅 디바이스에 적절한 정보를 디스플레이하기 위한 임의의 적합한 디스플레이 유닛을 포함할 수 있다. 디스플레이(1404)는 I/O 커맨드를 수신하기 위한 터치 스크린 영역(1410)을 포함할 수 있다. 몇몇 예에서, 터치스크린 영역(1410)은 디스플레이(1404)와 실질적으로 동일한 크기일 수 있다. I/O 디바이스(1406)는 모바일 컴퓨팅 디바이스 내로 정보를 입력하기 위한 임의의 적합한 I/O 디바이스를 포함할 수 있다. I/O 디바이스(1406)에 대한 예는 문자숫자 키보드(alphanumeric keyboard), 숫자 키보드(numeric keyboard), 터치 패드, 입력 키, 버튼, 스위치, 로커 스위치(rocker switch), 마이크, 스피커, 음성 인식 디바이스 및 소프트웨어, 그리고 기타 등등을 포함할 수 있다. 정보가 또한 마이크(도시되지 않음)에 의해서 디바이스(1400) 내로 입력될 수 있다. 그러한 정보는 음성 인식 디바이스(도시되지 않음)에 의해 디지털화될 수 있다. 실시예는 이 맥락 내에 한정되지 않는다.

[0119] 다양한 실시예는 하드웨어 요소, 소프트웨어 요소 또는 양자 모두의 조합을 사용하여 구현될 수 있다. 하드웨어 요소의 예는 프로세서, 마이크로프로세서, 회로, 회로 요소(가령, 트랜지스터(transistor), 저항기(resistor), 커패시터(capacitor), 인덕터(inductor) 및 기타 등등), 집적 회로, 애플리케이션 특정 집적 회로(Application Specific Integrated Circuit: ASIC), 프로그램가능 로직 디바이스(Programmable Logic Device: PLD), 디지털 신호 프로세서(Digital Signal Processor: DSP), 필드 프로그램가능 게이트 어레이(Field Programmable Gate Array: FPGA), 로직 게이트(logic gate), 레지스터(register), 반도체 디바이스, 칩, 마이크로칩, 칩셋 및 기타 등등을 포함할 수 있다. 소프트웨어의 예는 소프트웨어 컴포넌트, 프로그램, 애플리케이션, 컴퓨터 프로그램, 애플리케이션 프로그램, 시스템 프로그램, 머신 프로그램, 운영 체제 소프트웨어, 미들웨어(middleware), 펌웨어, 소프트웨어 모듈, 루틴(routine), 서브루틴(subroutine), 함수(function), 메소드(method), 프로시저(procedure), 소프트웨어 인터페이스, 애플리케이션 프로그램 인터페이스(Application Program Interface: API), 명령어 세트, 컴퓨팅 코드, 컴퓨터 코드, 코드 세그먼트(code segment), 컴퓨터 코드 세그먼트, 워드(word), 값(value), 심볼(symbol), 또는 이들의 임의의 조합을 포함할 수 있다. 하드웨어 요소 및/또는 소프트웨어 요소를 사용하여 일 실시예가 구현되는지를 정하는 것은 임의의 개수의 인자, 예를 들어 원하는 계산상의 속도(computational rate), 전력 레벨(power level), 내열성(heat tolerance), 처리 사이클 예산(processing cycle budget), 입력 데이터 레이트(input data rate), 출력 데이터 레이트(output data rate), 메모리 리소스(memory resource), 데이터 버스 속도(data bus speed) 및 다른 설계 또는 성능 제약에 따라 달라질 수 있다.

[0120] 적어도 하나의 실시예의 하나 이상의 양상은, 머신에 의해 관독되는 경우 머신으로 하여금 본 문서 내에 기술된 기법을 수행하기 위한 로직을 지어내게 하는, 프로세서 내의 다양한 로직을 나타내는 머신 판독가능 매체 상에 저장된 표상적인 명령어에 의해 구현될 수 있다. "IP 코어"로 알려진 그러한 표상은 유형적인(tangible) 머신 판독가능 매체 상에 저장되고 다양한 고객 또는 제조 설비에 제공되어 실제로 로직 또는 프로세서를 만드는 제조 머신 내에 적재될(load) 수 있다.

[0121] 본 문서 내에 개진된 어떤 특징은 다양한 구현을 참조하여 기술되었으나, 이 설명은 한정적인 의미로 해석되도록 의도된 것이 아니다. 따라서, 다른 구현뿐만 아니라 본 문서 내에 기술된 구현의 다양한 수정은, 본 개시가 관련된 업계에서 숙련된 자들에게 명백한 것으로서 본 개시의 사상 및 범주 내에 있는 것으로 간주된다.

[0122] 하나 이상의 제1 실시예에서, 자동 발화 인식(automatic speech recognition)을 제공하는 컴퓨터 구현된(computer-implemented) 방법은, 제1 시간 인스턴스(time instance)와 연관된 적어도 하나의 제1 거리값(distance value)을 판정하기 위해 위 제1 시간 인스턴스에 대해 신경망(neural network)을 평가하는(evaluating) 단계(위 적어도 하나의 제1 거리값은 위 신경망으로부터의 출력을 포함함)와, 제2 시간 인스턴스에 대해, 위 제1 거리값에 적어도 부분적으로 기반하여 적어도 하나의 제2 거리값을 근사계산하는(approximating) 단계(위 신경망은 위 제2 시간 인스턴스에 대해 평가되지 않음)와, 위 제1 거리값 및 위 제2 거리값에 적어도 부분적으로 기반하여 텍스트(textual) 요소의 시퀀스(sequence)를 판정하는 단계를 포함한다.

[0123] 제1 실시예에 덧붙여, 위 방법은 위 제1 시간 인스턴스에 대해 하나 이상의 출력 인덱스를 생성하는 단계를 더 포함하되, 위 제1 거리값은 위 출력 인덱스 중의 제1 출력 인덱스와 연관되고, 위 신경망은 복수의 출력층 노드를 가지는 출력층을 포함하며, 위 제1 시간 인스턴스에 대해 위 신경망을 평가하는 단계는 위 출력 인덱스와 연

관된 위 복수의 출력층 노드의 서브세트(subset)만을 평가하는 단계를 포함한다.

- [0124] 제1 실시예에 덧붙여, 위 방법은 위 제1 시간 인스턴스에 대해 하나 이상의 출력 인덱스를 생성하는 단계를 더 포함하되, 위 제1 거리값은 위 출력 인덱스 중의 제1 출력 인덱스와 연관되고, 위 신경망은 복수의 출력층 노드를 가지는 출력층을 포함하며, 위 제1 시간 인스턴스에 대해 위 신경망을 평가하는 단계는 위 출력 인덱스와 연관된 위 복수의 출력층 노드의 서브세트만을 평가하는 단계를 포함하되, 위 신경망은 최종 은닉층 노드를 가지는 최종 은닉층을 더 포함하고, 위 제1 시간 인스턴스에 대해 위 신경망을 평가하는 단계는 위 최종 은닉층 노드 전부를 평가하는 단계를 포함한다.
- [0125] 제1 실시예에 덧붙여, 위 방법은 위 제1 시간 인스턴스에 대해 하나 이상의 출력 인덱스를 생성하는 단계를 더 포함하되, 위 제1 거리값은 위 출력 인덱스 중의 제1 출력 인덱스와 연관되고, 위 신경망은 복수의 출력층 노드를 가지는 출력층을 포함하며, 위 제1 시간 인스턴스에 대해 위 신경망을 평가하는 단계는 위 출력 인덱스와 연관된 위 복수의 출력층 노드의 서브세트만을 평가하는 단계를 포함하고/하거나, 위 신경망은 최종 은닉층 노드를 가지는 최종 은닉층을 더 포함하고, 위 제1 시간 인스턴스에 대해 위 신경망을 평가하는 단계는 위 최종 은닉층 노드 전부를 평가하는 단계를 포함한다.
- [0126] 제1 실시예에 덧붙여, 위 제2 거리값을 근사계산하는 단계는 선형 함수(linear function), 비선형 함수(non-linear function), 또는 분산 함수(variance function) 중 적어도 하나를 기반으로 위 제1 거리값에 적어도 부분적으로 기반하여 위 제2 거리값을 외삽하는(extrapolating) 단계를 포함한다.
- [0127] 제1 실시예에 덧붙여, 위 제2 거리값을 근사계산하는 단계는 위 제1 거리값 및 제3 시간 인스턴스와 연관된 제3 거리값에 기반하여 위 제2 거리값을 외삽하는 단계를 포함하되, 위 제1 시간 인스턴스는 위 제2 시간 인스턴스 전이고 위 제3 시간 인스턴스는 위 제1 시간 인스턴스 전이다.
- [0128] 제1 실시예에 덧붙여, 위 제2 거리값을 근사계산하는 단계는 위 제1 거리값 및 제3 시간 인스턴스와 연관된 제3 거리값에 기반하여 위 제2 거리값을 외삽하는 단계를 포함하되, 위 제1 시간 인스턴스는 위 제2 시간 인스턴스 전이고 위 제3 시간 인스턴스는 위 제1 시간 인스턴스 전이되, 위 제2 거리값을 외삽하는 단계는 위 제1 거리값 및 위 제3 거리값에 기반하여 선형 함수를 통하여 위 제2 거리값을 외삽하는 단계를 포함하되, 위 제3 거리값은 위 신경망의 평가에 기반하여 판정된다.
- [0129] 제1 실시예에 덧붙여, 위 제2 거리값을 근사계산하는 단계는 위 제1 거리값 및 제3 시간 인스턴스와 연관된 제3 거리값에 기반하여 위 제2 거리값을 외삽하는 단계를 포함하되, 위 제1 시간 인스턴스는 위 제2 시간 인스턴스 전이고 위 제3 시간 인스턴스는 위 제1 시간 인스턴스 전이되, 위 제2 거리값을 외삽하는 단계는 위 제1 거리값 및 위 제3 거리값에 기반하여 선형 함수를 통하여 위 제2 거리값을 외삽하는 단계를 포함하되, 위 제3 거리값은 위 신경망의 평가에 기반하여 판정되되, 위 선형 함수는 위 제1 거리값 및 위 제3 거리값 간의 차이의 절반에 위 제1 거리값이 더해진 것을 포함한다.
- [0130] 제1 실시예에 덧붙여, 위 제2 거리값을 근사계산하는 단계는 위 제1 거리값 및 제3 시간 인스턴스와 연관된 제3 거리값에 기반하여 위 제2 거리값을 외삽하는 단계를 포함하되, 위 제1 시간 인스턴스는 위 제2 시간 인스턴스 전이고 위 제3 시간 인스턴스는 위 제1 시간 인스턴스 전이고/이거나, 위 제2 거리값을 외삽하는 단계는 위 제1 거리값 및 위 제3 거리값에 기반하여 선형 함수를 통하여 위 제2 거리값을 외삽하는 단계를 포함하고/하거나, 위 제3 거리값은 위 신경망의 평가에 기반하여 판정된다.
- [0131] 제1 실시예에 덧붙여, 위 방법은, 위 제2 시간 인스턴스에 대해 하나 이상의 출력 인덱스를 생성하는 단계(위 제2 시간 인스턴스에 대한 제3 거리값이 위 출력 인덱스 중의 제1 출력 인덱스와 연관되고, 위 신경망은 복수의 출력층 노드를 가지는 출력층 및 복수의 최종 은닉층 노드를 가지는 최종 은닉층을 포함하며, 위 제1 시간 인스턴스에 대해 위 신경망을 평가하는 위 단계는 복수의 최종 은닉층 값을 생성하기 위해 위 최종 은닉층의 모든 최종 은닉층 노드를 평가하는 단계를 포함함)와, 위 제2 시간 인스턴스에서, 위 제1 시간 인스턴스에 대한 제4 거리값을, 위 최종 은닉층 값에 기반하여 위 제4 거리값과 연관된 위 복수의 출력층 노드 중의 출력층 노드를 평가함으로써 판정하는 단계와, 위 제2 시간 인스턴스에 대해, 위 제3 거리값을 위 제4 거리값에 적어도 부분적으로 기반하여 근사계산하는 단계를 더 포함한다.
- [0132] 제1 실시예에 덧붙여, 위 신경망은 복수의 출력층 노드를 가지는 출력층을 포함하고, 위 제1 시간 인스턴스에 대해 위 신경망을 평가하는 단계는 위 복수의 출력층 노드 전부를 평가하는 단계를 포함한다.
- [0133] 제1 실시예에 덧붙여, 위 제1 시간 인스턴스는 신경망 평가 프레임(evaluation frame)과 연관되고, 위 제2 시간 인스턴스는 스킵 프레임(skip frame)과 연관되며, 1개, 2개 또는 3개의 추가적인 스킵 프레임이 위 평가 프레임

및 위 스킵 프레임 사이에 있다.

- [0134] 제1 실시예에 덧붙여, 위 신경망은 복수의 출력층 노드를 가지는 출력층을 포함하고, 위 제1 시간 인스턴스에 대해 위 신경망을 평가하는 단계는 위 복수의 출력층 노드 전부를 평가하는 단계를 포함하고/하거나, 위 제1 시간 인스턴스는 신경망 평가 프레임과 연관되고, 위 제2 시간 인스턴스는 스킵 프레임과 연관되며, 1개, 2개 또는 3개의 추가적인 스킵 프레임이 위 평가 프레임 및 위 스킵 프레임 사이에 있고/있거나, 텍스트 요소의 위 시퀀스를 판정하는 단계는 비터비 빔 탐색 디코더(Viterbi beam searching decoder)를 통하여 텍스트 요소의 위 시퀀스를 판정하는 단계를 포함한다.
- [0135] 제1 실시예에 덧붙여, 위 제1 시간 인스턴스는 신경망 평가 프레임과 연관되고, 위 제2 시간 인스턴스는 스킵 프레임과 연관되며, 위 방법은 가용 컴퓨팅 리소스(available computing resource) 또는 현재의 실시간 지수(current real time factor) 중 적어도 하나에 기반하여 프레임 스킵핑 레이트(frame skipping rate)를 판정하는 단계와, 위 프레임 스킵핑 레이트에 기반하여 위 평가 프레임 및 위 스킵 프레임 간에 추가적인 스킵 프레임을 제공하는 단계를 더 포함한다.
- [0136] 제1 실시예에 덧붙여, 텍스트 요소의 위 시퀀스를 판정하는 단계는 비터비 빔 탐색 디코더를 통하여 텍스트 요소의 위 시퀀스를 판정하는 단계를 포함한다.
- [0137] 제1 실시예에 덧붙여, 위 방법은 수신된 발화(received speech)를 발화 레코딩(speech recording)으로 변환하는 단계와, 위 발화 레코딩의 시간 윈도우와 연관된 특징 벡터를 추출하는 단계와, 위 신경망으로의 입력으로서 위 특징 벡터를 제공하는 단계를 더 포함한다.
- [0138] 하나 이상의 제2 실시예에서, 자동 발화 인식을 제공하는 시스템은 발화 인식 데이터를 저장하도록 구성된 메모리와, 위 메모리에 커플링된 중앙 처리 유닛을 포함하되, 위 중앙 처리 유닛은 제1 시간 인스턴스와 연관된 적어도 하나의 제1 거리값을 판정하기 위해 위 제1 시간 인스턴스에 대해 신경망을 구현하도록 구성된 신경망 회로와, 제2 시간 인스턴스에 대해, 위 제1 거리값에 적어도 부분적으로 기반하여 적어도 하나의 제2 거리값을 근사계산하도록 구성된 거리값 근사계산 회로와, 위 제1 거리값 및 위 제2 거리값에 적어도 부분적으로 기반하여 텍스트 요소의 시퀀스를 판정하도록 구성된 발화 디코더 회로를 포함한다.
- [0139] 제2 실시예에 덧붙여, 위 발화 디코더 회로는 위 제1 시간 인스턴스에 대해 하나 이상의 출력 인덱스를 생성하도록 또한 구성되되, 위 제1 거리값은 위 출력 인덱스 중의 제1 출력 인덱스와 연관되고, 위 신경망은 복수의 출력층 노드를 가지는 출력층을 포함하며, 위 신경망 회로는 위 제1 시간 인스턴스에 대해 위 출력 인덱스와 연관된 위 복수의 출력층 노드의 서브세트만을 평가하도록 구성된다.
- [0140] 제2 실시예에 덧붙여, 위 발화 디코더 회로는 위 제1 시간 인스턴스에 대해 하나 이상의 출력 인덱스를 생성하도록 또한 구성되되, 위 제1 거리값은 위 출력 인덱스 중의 제1 출력 인덱스와 연관되고, 위 신경망은 복수의 출력층 노드를 가지는 출력층을 포함하며, 위 신경망 회로는 위 제1 시간 인스턴스에 대해 위 출력 인덱스와 연관된 위 복수의 출력층 노드의 서브세트만을 평가하도록 구성되되, 위 신경망은 최종 은닉층 노드를 가지는 최종 은닉층을 더 포함하고, 위 신경망 회로는 위 제1 시간 인스턴스에 대해 위 최종 은닉층 노드 전부를 평가하도록 구성된다.
- [0141] 제2 실시예에 덧붙여, 위 발화 디코더 회로는 위 제1 시간 인스턴스에 대해 하나 이상의 출력 인덱스를 생성하도록 또한 구성되되, 위 제1 거리값은 위 출력 인덱스 중의 제1 출력 인덱스와 연관되고, 위 신경망은 복수의 출력층 노드를 가지는 출력층을 포함하며, 위 신경망 회로는 위 제1 시간 인스턴스에 대해 위 출력 인덱스와 연관된 위 복수의 출력층 노드의 서브세트만을 평가하도록 구성되고/되거나, 위 신경망은 최종 은닉층 노드를 가지는 최종 은닉층을 더 포함하되, 위 신경망 회로는 위 제1 시간 인스턴스에 대해 위 최종 은닉층 노드 전부를 평가하도록 구성된다.
- [0142] 제2 실시예에 덧붙여, 위 거리값 근사계산 회로가 위 제2 거리값을 근사계산하도록 구성되는 것은 위 거리값 근사계산 회로가 위 제1 거리값 및 제3 시간 인스턴스와 연관된 제3 거리값에 기반하여 위 제2 거리값을 외삽하도록 구성되는 것을 포함하되, 위 제1 시간 인스턴스는 위 제2 시간 인스턴스 전이고 위 제3 시간 인스턴스는 위 제1 시간 인스턴스 전이다.
- [0143] 제2 실시예에 덧붙여, 위 거리값 근사계산 회로는 위 제1 거리값 및 위 제3 거리값에 기반하여 선형함수를 통하여 위 제2 거리값을 외삽하도록 구성되되, 위 신경망 회로는 위 제3 거리값을 판정하기 위해 위 신경망을 구현하도록 구성된다.

- [0144] 제2 실시예에 덧붙여, 위 거리값 근사계산 회로가 위 제2 거리값을 근사계산하도록 구성되는 것은 위 거리값 근사계산 회로가 위 제1 거리값 및 제3 시간 인스턴스와 연관된 제3 거리값에 기반하여 선형 함수를 통하여 위 제2 거리값을 외삽하도록 구성되는 것을 포함하되, 위 제1 시간 인스턴스는 위 제2 시간 인스턴스 전이고 위 제3 시간 인스턴스는 위 제1 시간 인스턴스 전이고/이거나, 위 제3 거리값은 위 신경망의 평가에 기반하여 판정된다.
- [0145] 제2 실시예에 덧붙여, 위 거리값 근사계산 회로가 위 제2 거리값을 근사계산하도록 구성되는 것은 위 거리값 근사계산 회로가 위 제1 거리값 및 제3 시간 인스턴스와 연관된 제3 거리값에 기반하여 선형 함수를 통하여 위 제2 거리값을 외삽하도록 구성되는 것을 포함하되, 위 제1 시간 인스턴스는 위 제2 시간 인스턴스 전이고 위 제3 시간 인스턴스는 위 제1 시간 인스턴스 전이되, 위 제3 거리값은 위 신경망의 평가에 기반하여 판정된다.
- [0146] 제2 실시예에 덧붙여, 위 발화 디코더 회로는 위 제2 시간 인스턴스에 대해 하나 이상의 출력 인덱스를 생성하도록 또한 구성되되, 위 제2 시간 인스턴스에 대한 제3 거리값이 위 출력 인덱스 중의 제1 출력 인덱스와 연관되고, 위 신경망은 복수의 출력층 노드를 가지는 출력층 및 복수의 최종 은닉층 노드를 가지는 최종 은닉층을 포함하며, 위 신경망 회로는 복수의 최종 은닉층 값을 생성하기 위해 위 제1 시간 인스턴스에 대해 모든 최종 은닉층 노드를 평가하도록 구성되되, 위 신경망 회로는 위 제2 시간 인스턴스에서, 위 제1 시간 인스턴스에 대한 제4 거리값을, 위 최종 은닉층 값에 기반하여 위 제4 거리값과 연관된 위 복수의 출력층 노드 중의 출력층 노드를 평가함으로써 판정하도록 또한 구성되고, 위 거리값 근사계산 회로는 위 제2 시간 인스턴스에 대해, 위 제3 거리값을 위 제4 거리값에 적어도 부분적으로 기반하여 근사계산하도록 구성된다.
- [0147] 제2 실시예에 덧붙여, 위 신경망은 복수의 출력층 노드를 가지는 출력층을 포함하고, 위 신경망 회로가 위 제1 시간 인스턴스에 대해 위 신경망을 평가하도록 구성되는 것은 위 신경망 회로가 위 복수의 출력층 노드 전부를 평가하도록 구성되는 것을 포함한다.
- [0148] 제2 실시예에 덧붙여, 위 제1 시간 인스턴스는 신경망 평가 프레임과 연관되고, 위 제2 시간 인스턴스는 스킵 프레임과 연관되며, 1개, 2개 또는 3개의 추가적인 스킵 프레임이 위 평가 프레임 및 위 스킵 프레임 사이에 있다.
- [0149] 제2 실시예에 덧붙여, 위 발화 디코더 회로는 비터비 빔 탐색 디코더를 포함한다.
- [0150] 제2 실시예에 덧붙여, 위 시스템은 발화 레코딩의 시간 윈도우와 연관된 특징 벡터를 추출하도록 구성된 특징 추출 회로와, 위 시스템의 가용 컴퓨팅 리소스 또는 현재의 실시간 지수 중 적어도 하나에 기반하여 프레임 스킵핑 레이트를 판정하도록 구성된 제어기 회로를 더 포함한다.
- [0151] 하나 이상의 제3 실시예에서, 자동 발화 인식을 제공하는 시스템은, 제1 시간 인스턴스와 연관된 적어도 하나의 제1 거리값을 판정하기 위해 위 제1 시간 인스턴스에 대해 신경망을 평가하는 수단(위 적어도 하나의 제1 거리값은 위 신경망으로부터의 출력을 포함함)과, 제2 시간 인스턴스에 대해, 위 제1 거리값에 적어도 부분적으로 기반하여 적어도 하나의 제2 거리값을 근사계산하는 수단(위 신경망은 위 제2 시간 인스턴스에 대해 평가되지 않음)과, 위 제1 거리값 및 위 제2 거리값에 적어도 부분적으로 기반하여 텍스트 요소의 시퀀스를 판정하는 수단을 포함한다.
- [0152] 제3 실시예에 덧붙여, 위 시스템은 위 제1 시간 인스턴스에 대해 하나 이상의 출력 인덱스를 생성하는 수단을 더 포함하되, 위 제1 거리값은 위 출력 인덱스 중의 제1 출력 인덱스와 연관되고, 위 신경망은 복수의 출력층 노드를 가지는 출력층 및 최종 은닉층 노드를 가지는 최종 은닉층을 포함하며, 위 제1 시간 인스턴스에 대해 위 신경망을 평가하는 것은 위 출력 인덱스와 연관된 위 복수의 출력층 노드의 서브세트만을 평가하는 것 및 위 최종 은닉층 노드 전부를 평가하는 것을 포함한다.
- [0153] 제3 실시예에 덧붙여, 위 제2 거리값을 근사계산하는 것은 위 제1 거리값 및 제3 시간 인스턴스와 연관된 제3 거리값에 기반하여 선형 함수를 통하여 위 제2 거리값을 외삽하는 것을 포함하되, 위 제1 시간 인스턴스는 위 제2 시간 인스턴스 전이고 위 제3 시간 인스턴스는 위 제1 시간 인스턴스 전이며, 위 제3 거리값은 위 신경망의 평가에 기반하여 판정된다.
- [0154] 하나 이상의 제4 실시예에서, 적어도 하나의 머신 판독가능 매체(machine readable medium)는, 컴퓨팅 디바이스(computing device) 상에서 실행되는 것에 응답하여 위 컴퓨팅 디바이스로 하여금, 제1 시간 인스턴스와 연관된 적어도 하나의 제1 거리값을 판정하기 위해 위 제1 시간 인스턴스에 대해 신경망을 평가하는 것(위 적어도 하나의 제1 거리값은 위 신경망으로부터의 출력을 포함함)과, 제2 시간 인스턴스에 대해, 위 제1 거리값에 적어도

부분적으로 기반하여 적어도 하나의 제2 거리값을 근사계산하는 것(위 신경망은 위 제2 시간 인스턴스에 대해 평가되지 않음)과, 위 제1 거리값 및 위 제2 거리값에 적어도 부분적으로 기반하여 텍스트 요소의 시퀀스를 판정하는 것에 의해 자동 발화 인식을 제공하게 하는 복수의 명령어를 포함한다.

[0155] 제4 실시예에 덧붙여, 위 머신 판독가능 매체는, 위 컴퓨팅 디바이스 상에서 실행되는 것에 응답하여 위 컴퓨팅 디바이스로 하여금 위 제1 시간 인스턴스에 대해 하나 이상의 출력 인덱스를 생성하는 것에 의해 발화 인식을 수행하게 하는 명령어를 더 포함하되, 위 제1 거리값은 위 출력 인덱스 중의 제1 출력 인덱스와 연관되고, 위 신경망은 복수의 출력층 노드를 가지는 출력층을 포함하며, 위 제1 시간 인스턴스에 대해 위 신경망을 평가하는 것은 위 출력 인덱스와 연관된 위 복수의 출력층 노드의 서브세트만을 평가하는 것을 포함한다.

[0156] 제4 실시예에 덧붙여, 위 제2 거리값을 근사계산하는 것은 위 제1 거리값 및 제3 시간 인스턴스와 연관된 제3 거리값에 기반하여 선형 함수를 통하여 위 제2 거리값을 외삽하는 것을 포함하되, 위 제1 시간 인스턴스는 위 제2 시간 인스턴스 전이고 위 제3 시간 인스턴스는 위 제1 시간 인스턴스 전이되, 위 제3 거리값은 위 신경망의 평가에 기반하여 판정된다.

[0157] 제4 실시예에 덧붙여, 위 머신 판독가능 매체는, 위 컴퓨팅 디바이스 상에서 실행되는 것에 응답하여 위 컴퓨팅 디바이스로 하여금, 위 제2 시간 인스턴스에 대해 하나 이상의 출력 인덱스를 생성하는 것(위 제2 시간 인스턴스에 대한 제3 거리값이 위 출력 인덱스 중의 제1 출력 인덱스와 연관되고, 위 신경망은 복수의 출력층 노드를 가지는 출력층 및 복수의 최종 은닉층 노드를 가지는 최종 은닉층을 포함하며, 위 제1 시간 인스턴스에 대해 위 신경망을 평가하는 것은 복수의 최종 은닉층 값을 생성하기 위해 위 최종 은닉층의 모든 최종 은닉층 노드를 평가하는 것을 포함함)과, 위 제2 시간 인스턴스에서, 위 제1 시간 인스턴스에 대한 제4 거리값을, 위 최종 은닉층 값에 기반하여 위 제4 거리값과 연관된 위 복수의 출력층 노드 중의 출력층 노드를 평가함으로써 판정하는 것과, 위 제2 시간 인스턴스에 대해, 위 제4 거리값에 적어도 부분적으로 기반하여 위 제3 거리값을 근사계산하는 것에 의해 발화 인식을 수행하게 하는 명령어를 더 포함한다.

[0158] 제4 실시예에 덧붙여, 위 머신 판독가능 매체는, 위 컴퓨팅 디바이스 상에서 실행되는 것에 응답하여 위 컴퓨팅 디바이스로 하여금, 수신된 발화를 발화 레코딩으로 변환하는 것과, 위 발화 레코딩의 시간 윈도우와 연관된 특징 벡터를 추출하는 것과, 위 신경망으로의 입력으로서 위 특징 벡터를 제공하는 것에 의해 발화 인식을 수행하게 하는 명령어를 더 포함한다.

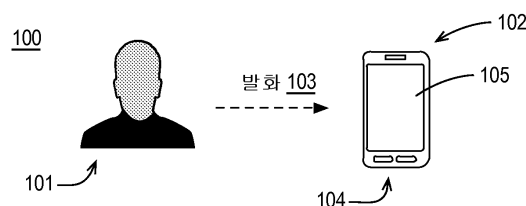
[0159] 하나 이상의 제5 실시예에서, 적어도 하나의 머신 판독가능 매체는, 컴퓨팅 디바이스 상에서 실행되는 것에 응답하여 위 컴퓨팅 디바이스로 하여금 위 실시예 중 임의의 것에 따른 방법을 수행하게 하는 복수의 명령어를 포함할 수 있다.

[0160] 하나 이상의 제6 실시예에서, 장치는 위 실시예 중 임의의 것에 따른 방법을 수행하는 수단을 포함할 수 있다.

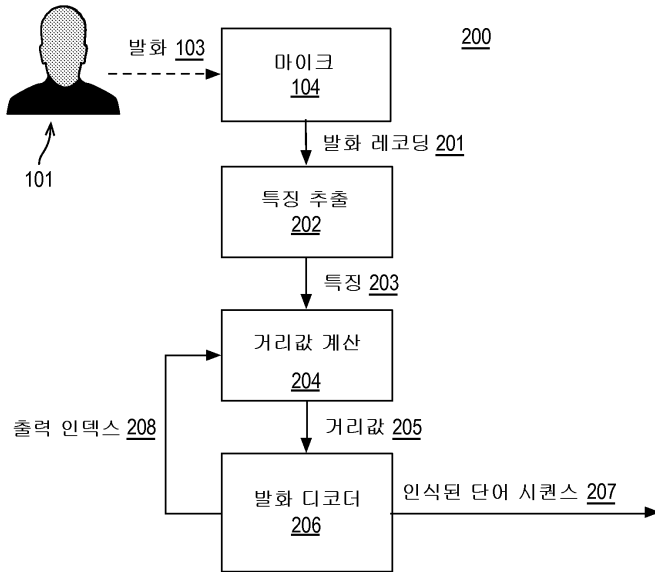
[0161] 실시예는 이와 같이 기술된 실시예에 한정되지 않으나, 부기된 청구항의 범주로부터 벗어나지 않고서 수정 및 변경과 함께 실시될 수 있음이 인식될 것이다. 예컨대, 위의 실시예는 특징의 특정 조합을 포함할 수 있다. 그러나, 위의 실시예는 이 점에 한정되지 않고, 다양한 구현에서, 위의 실시예는 그러한 특징의 서브세트만을 시행하는 것, 그러한 특징의 다른 순서를 시행하는 것, 그러한 특징의 상이한 조합을 시행하는 것, 그리고/또는 명시적으로 나열된 특징 외의 추가적인 특징을 시행하는 것을 포함할 수 있다. 따라서, 실시예의 범주는, 부기된 청구항을 그러한 청구항에 주어질 만한 균등물의 전 범주와 더불어 참조하여 판정되어야 한다.

도면

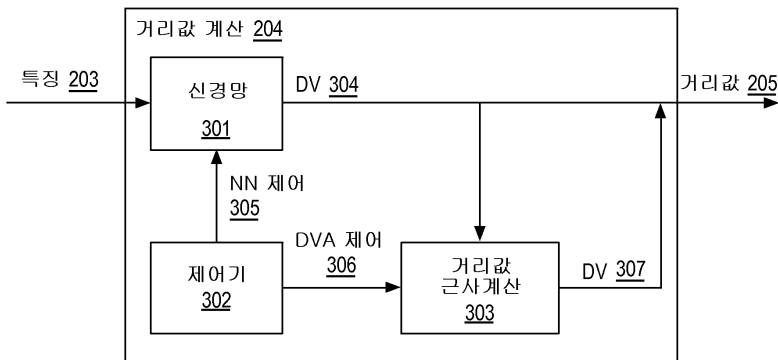
도면1



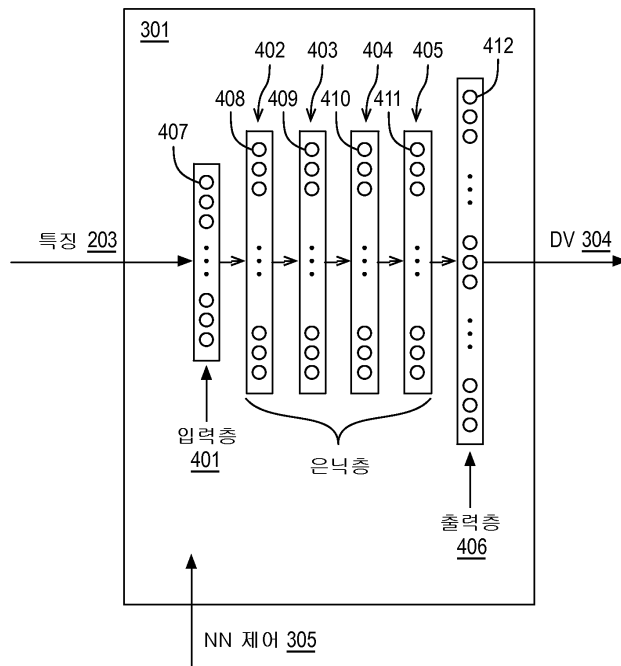
도면2



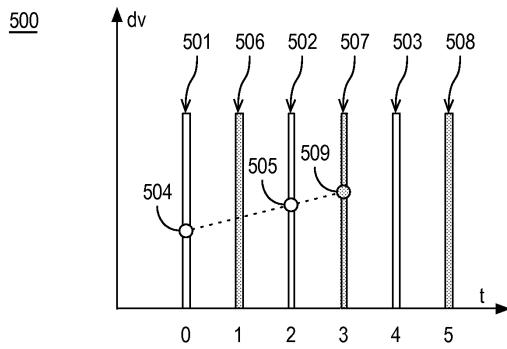
도면3



도면4

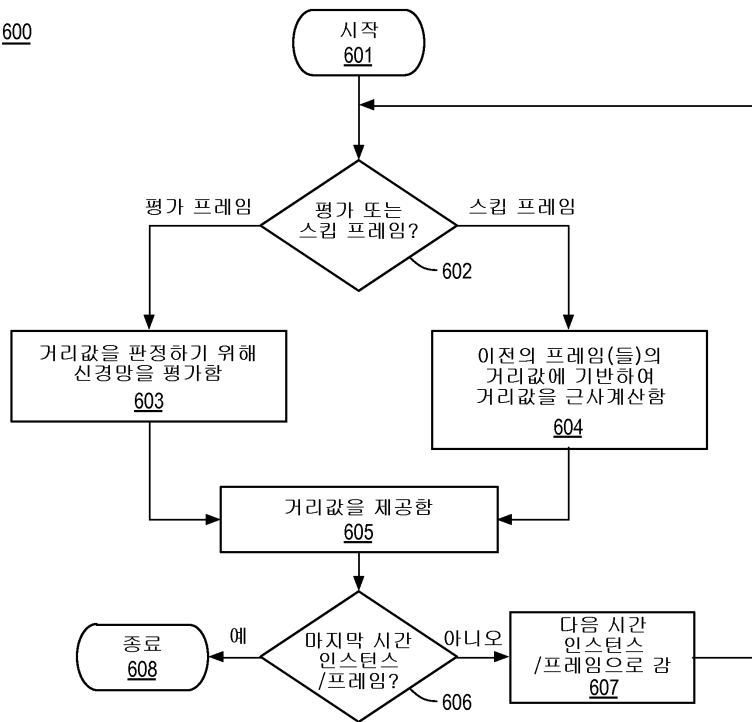


도면5

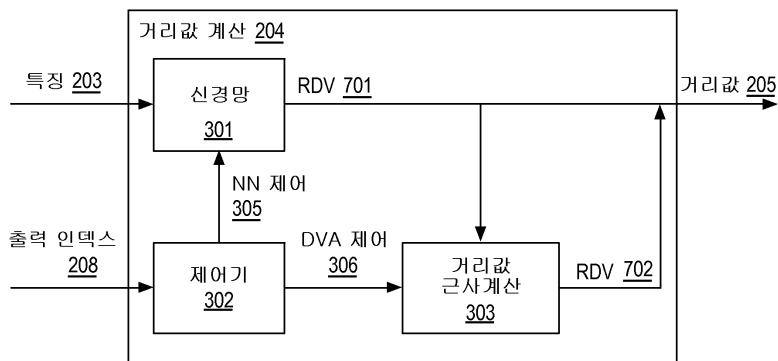


도면6

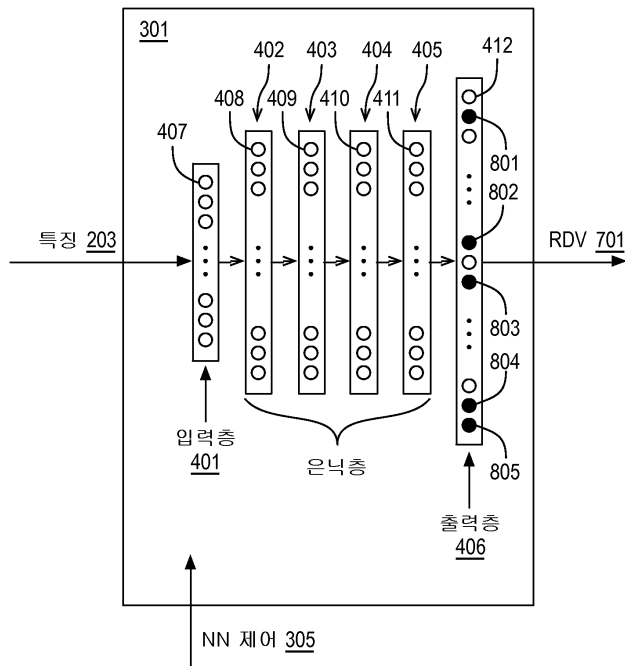
600



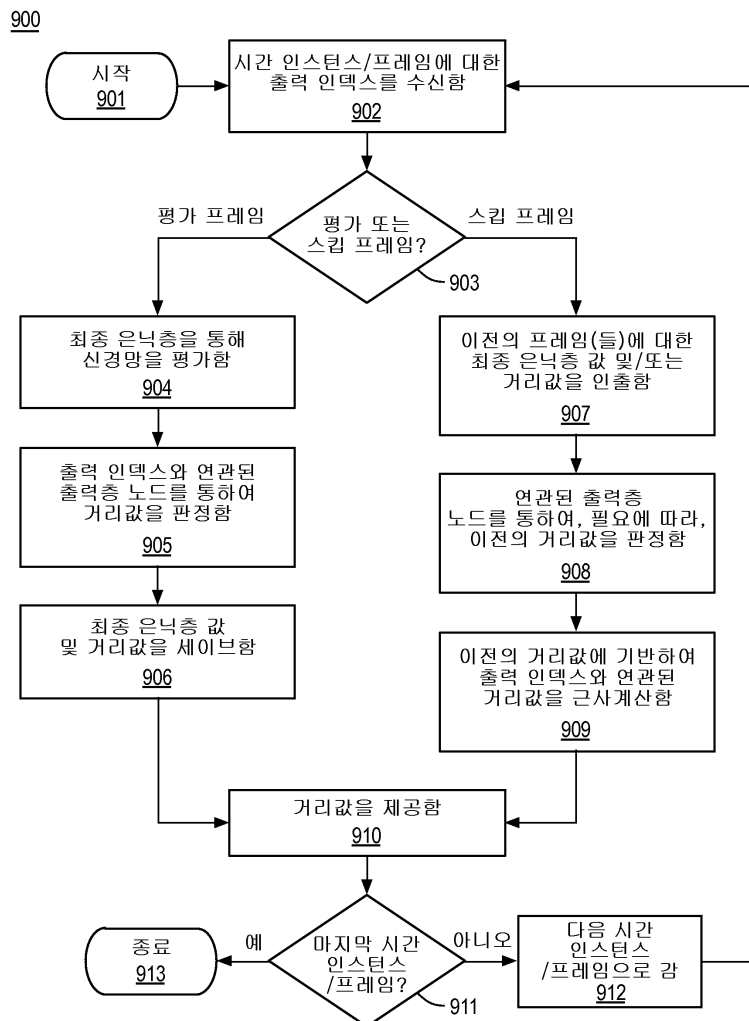
도면7



도면8

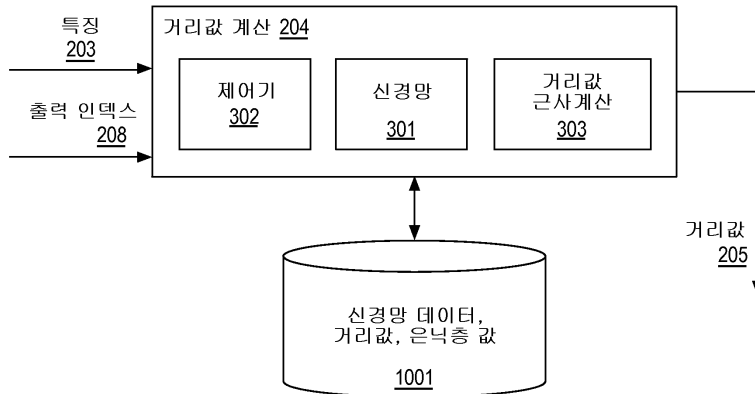


도면9



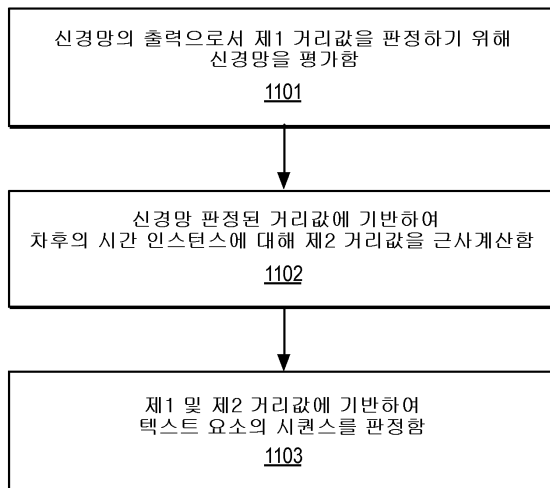
도면10

1000

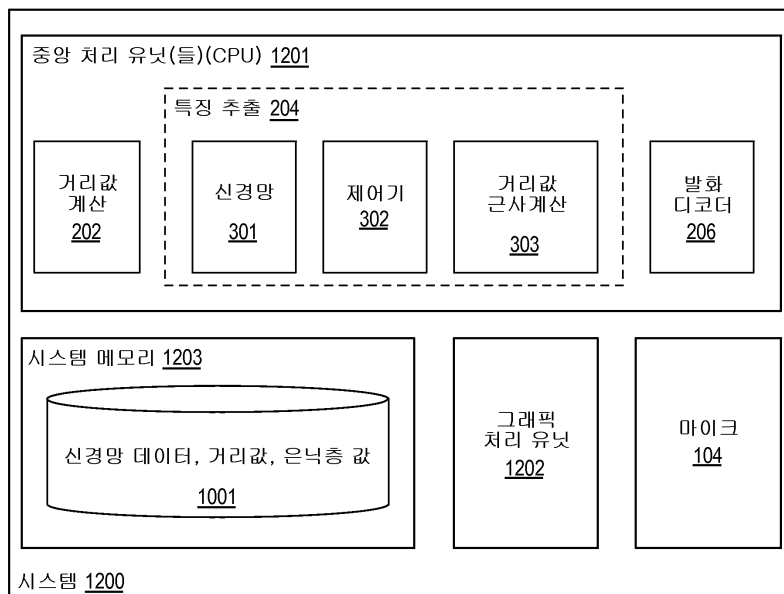


도면11

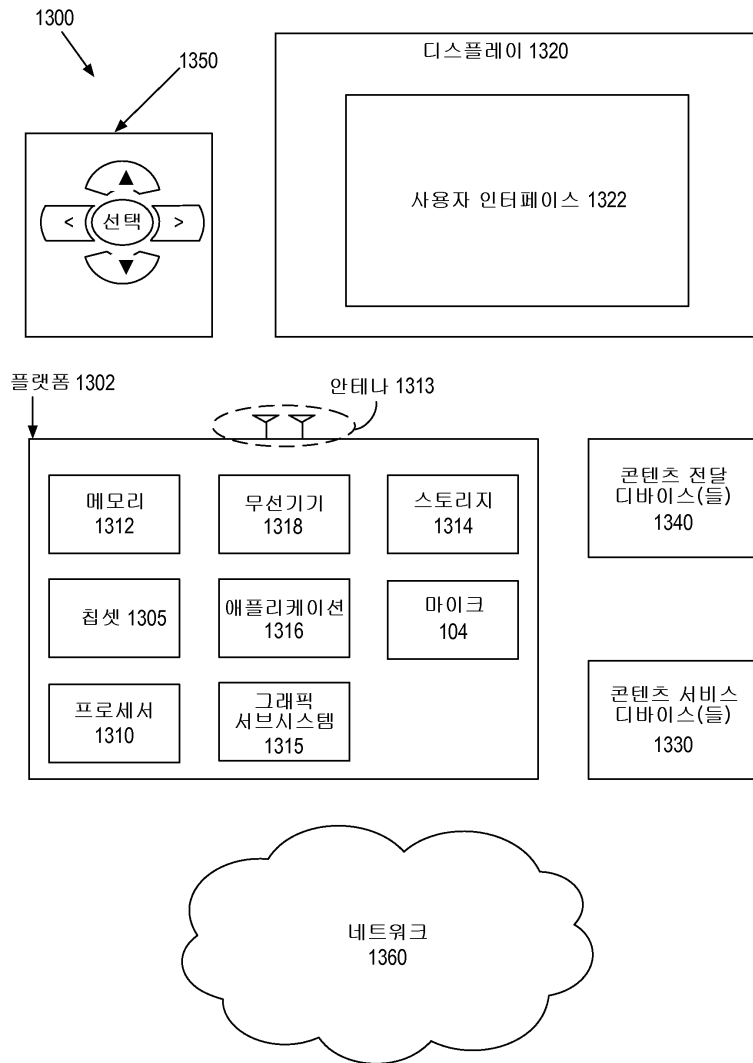
1100



도면12



도면13



도면14

