(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2012-0129421
(43) 공개일자 2012년11월28일

(51) 국제특허분류(Int. Cl.)
G10L 15/14 (2006.01) *G10L 15/02* (2006.01)
(21) 출원번호 10-2011-0047663
(22) 출원일자 2011년05월20일
심사청구일자 2011년05월20일

(71) 출원인
고려대학교 산학협력단
서울 성북구 안암동5가 1
(72) 발명자
육동석
서울특별시 용산구 이촌로 347, 1동 102호 (서빙고동, 신동아아파트)
조선호
서울특별시 광진구 자양로32길 22-7 (구의동)
백향순
서울특별시 영등포구 선유로 207, 현대 6차 아파트 606동 1704호 (양평동3가)
(74) 대리인
특허법인충현

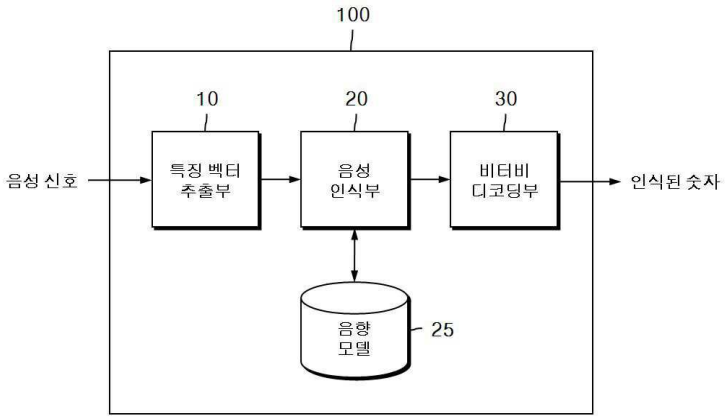
전체 청구항 수 : 총 5 항

(54) 발명의 명칭 **숫자 음성 인식 장치 및 방법**

(57) 요 약

본 발명은 숫자 음성 인식 장치, 방법 및 그 방법을 기록한 기록매체에 관한 것으로, 본 발명에 따른 숫자 음성 인식 방법은, 음성 신호를 입력받고, 입력된 음성 신호로부터 음성의 특징을 나타내는 특징 벡터를 추출하고, 추출된 특징 벡터를 이용하여 좌, 우 문맥을 고려한 음절에 기반한 음향 모델로부터 적어도 하나 이상의 숫자 음절 후보를 생성하며, 생성된 숫자 음절 후보의 조합으로부터 산출된 확률에 따라 음성 인식된 숫자를 출력한다.

대 표 도 - 도1



특허청구의 범위

청구항 1

음성 신호를 입력받는 단계;

상기 입력된 음성 신호로부터 음성의 특징을 나타내는 특징 벡터를 추출하는 단계;

상기 추출된 특징 벡터를 이용하여 좌, 우 문맥을 고려한 음절(syllable)에 기반한 소정 음향 모델로부터 적어도 하나 이상의 숫자 음절 후보를 생성하는 단계; 및

상기 생성된 숫자 음절 후보의 조합으로부터 산출된 확률에 따라 음성 인식된 숫자를 출력하는 단계를 포함하는 숫자 음성 인식 방법.

청구항 2

제 1 항에 있어서,

상기 소정 음향 모델은 음절의 좌, 우 문맥을 고려한 트라이실러블(trisyllable)을 이용하여 학습된 HMM(Hidden Markov Model)인 것을 특징으로 하는 숫자 음성 인식 방법.

청구항 3

제 1 항에 있어서,

상기 소정 음향 모델은 하향식(top-down) 결정 트리 기반 상태 클러스터링(decision tree-based state clustering)을 이용하여 상기 좌, 우 문맥을 고려한 음절들 중 특성이 유사한 음절의 상태를 클러스터링함으로써 학습된 것을 특징으로 하는 숫자 음성 인식 방법.

청구항 4

제 3 항에 있어서,

상기 결정 트리 기반 상태 클러스터링은 연속하는 한국어 숫자의 음절 특성을 고려하여 발음의 유사도를 기준으로 생성된 음절 단위의 질의어 집합(question set)을 이용하여 수행되는 것을 특징으로 하는 숫자 음성 인식 방법.

청구항 5

제 1 항 내지 제 4 항 중에 어느 한 항의 방법을 컴퓨터에서 실행시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체.

명 세 서

기 술 분 야

[0001] 본 발명은 숫자 음성 인식 장치 및 방법에 관한 것으로, 특히 연속하는 한국어 숫자의 음절을 이용하여 해당 한국어 숫자에 대한 음성을 인식하는 장치, 방법 및 그 방법을 기록한 기록매체에 관한 것이다.

배 경 기 술

[0002] 음성 인식은 음성 처리 기술의 발달과 함께 오래 전부터 연구되어 온 주제이다. 음성 인식과 관련하여 인식률을 향상시키기 위한 다양한 연구들이 진행되어 왔고, 현재까지도 계속 연구되고 있다.

[0003] 음성은 사람마다 다르기 때문에 음성 인식을 위해, 음성을 적당한 전처리를 거친 뒤 디지털 신호로 바꾸어 특징을 추출한다. 추출된 음성 패턴을 기준으로 하여 음성 입력 신호를 인식하는 기술은 넓게는 인공지능(Artificial Intelligence)을 구현하기 위해 필요한 기술이고, 좁게는 음성인식 전화기 등 각종 자동화 기기에 응용될 수 있다. 이러한 음성 인식 기술은 다양한 제품에 채용될 수 있고, 이로 인해 사용자는 수동 조작으로부터 자유로와지므로 행동의 동시성을 이룰 수 있으며, 특히 손사용이 불편한 장애인들이 편리하게 사용할 수 있

어 사람과 기계의 친숙성의 효과를 기대할 수 있다.

[0004] 한편, 이러한 음성 인식 기술은 다양한 음향 모델들과 인식 방법을 제안하고 있으나, 이러한 종래의 음향 모델들은 여전히 음성 인식률이 저하되거나, 개별적인 음소나 단어에만 치중하고 있어 활용도가 떨어진다.

발명의 내용

해결하려는 과제

[0005] 본 발명이 해결하고자 하는 기술적 과제는 음향 모델의 구축을 간소화할 경우 음성 인식률이 저하되는 문제점을 해결하고, 단순한 음절(monosyllable) 단위의 음향 모델을 사용할 경우 좌, 우 문맥과 동떨어진 음성 인식 결과가 나타나는 한계를 극복하며, 음향 모델의 학습 과정에서 부족한 학습 데이터의 보충이 음소에 한정되는 제한을 해소하고자 한다.

과제의 해결 수단

[0006] 상기 기술적 과제를 해결하기 위하여, 본 발명에 따른 숫자 음성 인식 방법은, 음성 신호를 입력받는 단계; 상기 입력된 음성 신호로부터 음성의 특징을 나타내는 특징 벡터를 추출하는 단계; 상기 추출된 특징 벡터를 이용하여 좌, 우 문맥을 고려한 음절(syllable)에 기반한 소정 음향 모델로부터 적어도 하나 이상의 숫자 음절 후보를 생성하는 단계; 및 상기 생성된 숫자 음절 후보의 조합으로부터 산출된 확률에 따라 음성 인식된 숫자를 출력하는 단계를 포함한다.

[0007] 상기된 숫자 음성 인식 방법에서, 상기 소정 음향 모델은 음절의 좌, 우 문맥을 고려한 트라이실러블(trisyllable)을 이용하여 학습된 HMM(Hidden Markov Model)이다.

[0008] 상기된 숫자 음성 인식 방법에서, 상기 소정 음향 모델은 하향식(top-down) 결정 트리 기반 상태 클러스터링(decision tree-based state clustering)을 이용하여 상기 좌, 우 문맥을 고려한 음절들 중 특성이 유사한 음절의 상태를 클러스터링함으로써 학습된다.

[0009] 또한, 상기된 숫자 음성 인식 방법에서, 상기 결정 트리 기반 상태 클러스터링은 연속하는 한국어 숫자의 음절 특성을 고려하여 발음의 유사도를 기준으로 생성된 음절 단위의 질의어 집합(question set)을 이용하여 수행된다.

[0010] 나아가, 이하에서는 상기 기재된 숫자 음성 인식 방법을 컴퓨터에서 실행시키기 위한 프로그램을 기록한 컴퓨터로 읽을 수 있는 기록매체를 제공한다.

발명의 효과

[0011] 본 발명은 음향 모델의 구축의 복잡도 및 난이도를 크게 증가시키지 않으면서도 음성 인식의 성능을 향상시키고, 좌, 우 문맥을 고려한 음절에 기반한 음성 인식 모델을 통하여 문맥에 적합한 숫자 후보를 제공할 수 있으며, 음향 모델의 학습 과정에서 부족한 학습 데이터의 보충을 음절 단위로 확장할 수 있다.

도면의 간단한 설명

[0012] 도 1은 본 발명의 일 실시예에 따른 트라이실러블(trisyllable)을 이용한 음향 모델에 기반한 숫자 음성 인식 장치를 도시한 블록도이다.

도 2는 본 발명의 일 실시예에 따른 트라이실러블을 이용한 음향 모델에 기반한 숫자 음성 인식 방법을 도시한 흐름도이다.

도 3은 본 발명의 다른 실시예에 따른 숫자 음성 인식을 위한 음향 모델을 트라이실러블을 이용하여 학습시키는 장치를 도시한 블록도이다.

도 4는 본 발명의 다른 실시예에 따른 도 3에서 결정 트리 기반 상태 클러스터링에 사용될 수 있는 질의어 집합의 일부를 예시한 도면이다.

도 5a 및 도 5b는 종래의 음성 인식 방법을 통한 실험 결과와 본 발명의 일 실시예에 따른 숫자 음성 인식 방법을 통한 실험 결과를 비교하여 나타낸 도표이다.

발명을 실시하기 위한 구체적인 내용

[0013] 본 발명의 실시예들을 설명하기에 앞서 실시예들이 활용되는 음성 인식 기술 분야에 대해 개괄적으로 소개하고, 실시예들이 공통적으로 채용하고 있는 기본 아이디어를 제시하고자 한다.

[0014] 음성 인식 분야에서 사용되는 음향 모델의 기본 단위에는 모노폰(monophone), 트라이폰(triphone) 및 음절(syllable) 등이 알려져 있다. 이 중, 모노폰은 문맥을 고려하지 않은 단순한 음소를 의미하고, 트라이폰은 음소의 좌, 우 문맥을 고려한 음소를 의미한다. 또한, 음절은 하나의 종합된 음의 느낌을 주는 말소리의 단위를 의미하는 것으로, 하나 이상의 음소를 포함한다. 하나의 음절은 독립적인 인식 단위로 활용될 수 있으며, 음절 그 자체 이외에 여타의 고려 요소가 없으므로 모노실러블(monosyllable)이라고 명명될 수 있다.

[0015] 모노폰의 경우, 음향 모델은 간단하게 구축될 수 있는 반면에 같은 음소도 문맥에 따라 다르게 발음되는 특성을 고려하지 않았기 때문에 음성 인식률은 저조하게 나타난다. 트라이폰의 경우, 음소의 좌, 우 문맥을 고려하기 때문에 음소의 개수는 모노폰의 경우에 비해 음소 개수가 세 제곱이 되며, 이를 학습시키기 위해서는 많은 데이터가 필요하다는 단점이 존재한다. 한편, 음절 자체만을 음성 인식의 단위로 활용하는 모노실러블의 경우에도 모노폰의 경우와 마찬가지로 여전히 좌, 우 문맥을 고려하지 않으므로 상황에 맞는 음성 인식이 이루어질 수 없다는 문제점이 존재한다.

[0016] 이상과 같은 문제 인식에 기초하여 이하에서 소개될 본 발명의 실시예들은 공통적으로 음절을 음성 인식의 기본 단위로 활용할 뿐만 아니라, 음절의 좌, 우 문맥을 고려한 음향 모델을 활용하여 음성 인식을 수행한다. 따라서, 이하에서는 이상에서 언급한 모노실러블의 개념에 대응하여 음절의 좌, 우 문맥을 고려한 트라이실러블(trisyllable)의 개념을 제안하고자 한다. 예를 들어, '아버지'라는 단어에 대해 트라이실러블을 이용한 음성 인식의 경우, '-아+버', '아-버+지', '버-지+' 등과 같은 음절의 좌, 우 문맥을 고려한 후보들이 활용될 수 있으며, 이로 인해 음성 인식의 정확도를 향상시킬 수 있다.

[0017] 도 1은 본 발명의 일 실시예에 따른 트라이실러블(trisyllable)을 이용한 음향 모델에 기반한 숫자 음성 인식 장치(100)를 도시한 블록도로서, 특징 벡터 추출부(10), 음성 인식부(20), 음향 모델(25) 및 비터비 디코딩부(30)를 포함한다.

[0018] 특징 벡터 추출부(10)는 사용자로부터 입력된 음성 신호로부터 음성의 특징을 나타내는 특징 벡터를 추출한다. 이 때, 특징 벡터로는 Mel-Frequency Cepstral Coefficient(MFCC)를 포함하여 음향 인식 분야에서 활용될 수 있는 다양한 기법들이 활용될 수 있다.

[0019] 음성 인식부(20)는 특징 벡터 추출부(10)를 통해 추출된 특징 벡터를 이용하여 좌, 우 문맥을 고려한 음절(즉, 트라이실러블을 의미한다.)에 기반한 소정 음향 모델(25)로부터 적어도 하나 이상의 숫자 음절 후보를 생성한다. 여기서 음향 모델이란, 음성 인식에서 사람의 음성을 입력받아 설정된 단위를 기준으로 그 음성이 어떤 소리인지를 인식하기 위해 사용되는 모델을 의미한다. 이러한 음성 인식부(20)는 특징 벡터를 이용하여 활성 노드를 산출하고, 산출된 활성 노드의 수와 미리 정해진 기준치와의 관계를 이용하여 관측 확률 계산 방식을 결정하며, 결정된 방식을 이용하여 관측 확률을 산출함으로써 음성 인식을 수행한다.

[0020] 음성 인식부(20)가 음성 인식을 수행함에 있어서 다양한 음성 인식 방법들이 활용될 수 있으나, 본 발명의 실시예를 통해서는 HMM(Hidden Markov Model)을 사용하는 방법을 예시하고자 한다. 그러나, 이러한 HMM을 이용한 숫자 음성 인식은 일례에 불과한 것으로서, 본 발명의 다양한 실시예들은 트라이실러블을 이용하여 좌, 우 문맥을 고려한 음향 모델에 기반한 음성 인식을 활용하는 한도 내에서, 다른 어떠한 음성 인식 방법에도 적용될 수 있음을 밝혀둔다.

[0021] HMM에서는 음성 인식 과정으로 비터비(Viterbi) 탐색을 실시하는데, 이는 인식 대상 후보 단어들에 대한 미리 훈련하여 구축한 음향 모델(HMM)과 현재 입력된 음성의 특징들과의 차이를 비교하여 가장 유력한 후보 단어를 결정하는 과정이다. 통상적으로 HMM은 음성 인식을 위한 기본 단위로 음소를 모델링하나, 본 발명의 실시예들은 좌, 우 문맥을 고려한 음절에 기초하여 모델링을 수행한다. 이를 위해, 본 실시예의 음향 모델(25)은 음성 인식 장치(100)에 입력되는 음절과 음성 인식 장치(100) 내의 음향 모델에서 보유하고 있는 데이터베이스 내의 음절을 결합하여 단어와 문장을 생성한다. 즉, 음성 인식부(20)에서 활용하는 HMM은 관측 불가능한 프로세스를 관측 가능한 다른 프로세스를 통해 추정하는 이중 확률 처리 방법을 채택하고 있으므로, 본 실시예가 채택하고 있는 음성 인식의 최소 단위인 음절을 좌, 우 문맥을 고려하여 모델링하고, 이를 이용해 음성 인식부(20)를 통해 음성 인식을 수행한다.

[0022] 비터비 디코딩부(30)는 음성 인식부(20)를 통해 생성된 숫자 음절 후보의 조합으로부터 산출된 확률에 따라 비터비 디코딩(Viterbi Decoding)을 수행함으로써 음성 인식된 숫자를 출력한다.

[0023] 보다 구체적으로, 본 실시예에 따른 숫자 음성 인식 장치(100)는 인식 대상 어휘들을 HMM을 통해 확률적으로 모델링 한 후, 인식 대상 어휘들의 HMM으로 구성된 트리(tree) 구조를 따라서 입력된 음성 신호로부터 추출된 특징 벡터의 관측 확률을 계산하는 비터비 탐색 알고리즘을 사용할 수 있다. 이때, 특징 벡터의 관측 확률은 HMM을 구성하는 다수의 가우시안 분포에서의 관측 확률의 누적 합으로 표현될 수 있다. 즉, 특징 벡터를 이용하여 구해진 활성 노드는 고립 단어 인식기의 경우 처음 수십 프레임 동안은 수십 개의 활성 노드를 갖고, 점차적으로 증가하여 인식 대상 어휘 수에 따라서 수천 내지 수만 개까지 증가하다가 점점 감소하기 시작하여 음성 종료 시점의 수십 프레임 동안은 다시 수십 개의 활성 노드를 통상적으로 갖는다.

[0024] 도 2는 본 발명의 일 실시예에 따른 트라이실러블을 이용한 음향 모델에 기반한 숫자 음성 인식 방법을 도시한 흐름도로서, 도 1의 구성에 대응되는 각각의 단계들을 포함하므로 여기서는 그 개요를 중심으로 간략하게 설명한다.

[0025] 210 단계에서 음성 신호를 입력받는다. 이러한 음성 신호는 통상적인 아날로그 형태의 데이터가 아니라 디지털 신호로 가공된 데이터이며, 특히 음성 신호를 프레임화하고, 노이즈가 제거된 신호로부터 잔여 에너지를 대상으로 음향 프레임별로 절대 에너지의 크기와 영교차율을 분석하여 음성 구간을 검출하는 전처리 과정을 통해 가공된 신호인 것이 바람직하다.

[0026] 220 단계에서는 210 단계를 통해 입력된 음성 신호로부터 음성의 특징을 나타내는 특징 벡터를 추출한다. 이 단계는 도 1에서 설명한 특징 벡터 추출부(10)에 대응하는 것으로 여기서는 구체적인 설명을 생략한다.

[0027] 230 단계에서는 220 단계를 통해 추출된 특징 벡터를 이용하여 좌, 우 문맥을 고려한 음절에 기반한 음향 모델로부터 적어도 하나 이상의 숫자 음절 후보를 생성한다. 이 단계는 도 1에서 설명한 음성 인식부(20)에 대응하는 것으로 여기서는 구체적인 설명을 생략한다. 한편, 음향 모델은 음절의 좌, 우 문맥을 고려한 트라이실러블(trisyllable)을 이용하여 학습된 HMM(Hidden Markov Model)인 것이 바람직하다. 음향 모델을 학습시키는 과정에 대한 실천적인 구성은 이후 도 3을 통해 보다 구체적으로 설명한다.

[0028] 240 단계에서는 230 단계를 통해 생성된 숫자 음절 후보의 조합으로부터 산출된 확률에 따라 음성 인식된 숫자를 출력한다. 이 단계는 도 1에서 설명한 비터비 디코딩부(30)에 대응하는 것으로 여기서는 구체적인 설명을 생략한다.

[0029] 도 3은 본 발명의 다른 실시예에 따른 숫자 음성 인식을 위한 음향 모델을 트라이실러블을 이용하여 학습시키는 장치(300)를 도시한 블록도로서, 특징 벡터 추출부(21), 음향 모델 학습부(22) 및 결정 트리 기반 상태 클러스터링부(23)를 포함한다.

[0030] 특징 벡터 추출부(21)는 표본으로서 입력된 숫자에 대한 음성 신호들로부터 특징 벡터를 추출한다.

[0031] 음향 모델 학습부(22)는 특징 벡터와 트라이실러블을 이용하여 음향 모델을 학습시킨다. 본 실시예에서는 HMM 음향 모델의 기본 인식 단위로 단순히 하나의 음절이 아닌 좌, 우 문맥을 고려한 트라이실러블을 사용하며, 한국어 연속 숫자음 인식을 기준으로 예시하였다. 한국어 숫자음은 '일', '이', '삼', '사', '오', '육', '칠', '팔', '구', '공', '영' 이라는 숫자음에 대해 인식을 수행하였다.

[0032] 한편, 숫자 음절을 인식하기 위한 모노실러블의 초기 HMM 모델로 모노폰의 HMM 모델을 조합하여 사용할 수 있다. 예를 들어, 모노실러블 모델 '삼'의 경우는 모노폰의 'ㅅ', 'ㅏ', 'ㅁ'의 조합이 되고, '구'의 경우에는 모노폰의 'ㄱ'와 'ㅜ'의 조합이 된다. 조합을 이루는 모노폰의 개수에 따라 모델별 상태 개수도 달라지게 된다.

[0033] 트라이실러블의 경우에는 모노실러블의 HMM 모델을 복제하여 초기 모델로 사용할 수 있는데, 트라이실러블의 경우, 트라이폰과 유사하게 음절의 좌, 우 문맥을 고려하기 때문에 모노실러블의 음절 개수의 세제곱만큼의 모델이 필요하다. 그 결과 트라이실러블의 연산을 위해서는 많은 양의 학습 데이터를 필요로 하게 되지만, 현실적으로 많은 양의 충분한 학습 데이터를 확보하기란 쉽지 않다는 점이 지적될 수 있다. 즉, 모노실러블에서 트라이실러블로 음절 인식 단위를 확장할 경우 학습 데이터가 부족하여 인식율이 높지 않은 모델이 구출될 가능성이 있다. 이를 극복하기 위해 본 실시예에서는 결정 트리 기반 상태 클러스터링부(23)를 통해 상태 클러스터링 방법을 활용함으로써 이러한 문제를 해결하였다.

[0034] 결정 트리 기반 상태 클러스터링부(23)는 음절 기반 질의어 집합을 이용하여 상태 클러스터링을 수행한다. 상태 클러스터링(state clustering)이란, HMM에서 사용되는 각각의 상태(state)를 유사 정도에 따라 군집화하는 것을

의미하며, 상태 클러스터링된 상태들을 각각 서로의 학습 데이터를 공유하게 된다.

[0035] 보다 구체적으로, 본 실시예의 음향 모델은 하향식(top-down) 결정 트리 기반 상태 클러스터링(decision tree-based state clustering)을 이용하여 좌, 우 문맥을 고려한 음절들 중 특성이 유사한 음절의 상태를 클러스터링함으로써 학습된다. 즉, 결정 트리 기반 상태 클러스터링은 연속하는 한국어 숫자의 음절 특성을 고려하여 발음의 유사도를 기준으로 생성된 음절 단위의 질의어 집합(question set)을 이용하여 수행된다. 여기서, 하향식 클러스터링이란, 이미 가지고 있는 질의어 집합을 바탕으로 결정 트리 알고리즘을 이용하여 상태를 클러스터링하는 것을 의미하며, 이러한 질의어 집합은 언어학적 특징을 이용하여 구축될 수 있다.

[0036] 도 4는 본 발명의 다른 실시예에 따른 도 3에서 결정 트리 기반 상태 클러스터링에 사용될 수 있는 질의어 집합의 일부를 예시한 도면이다. 종래의 음향 모델에서 상태 클러스터링을 위해 사용하는 질의어 집합은 트라이폰에서 사용될 수 있을 뿐, 음절 단위의 음성 인식에서는 활용될 수 없는 한계가 존재하였다. 따라서, 본 실시예에서는 음절의 좌, 우 문맥을 고려한 트라이실러블을 위해 새로운 질의어 집합을 제작하였다.

[0037] 도 4를 참조하면, 오른쪽 음절이 'ㄹ'로 끝나는 경우의 예시로서, '*+칠', '*+팔', '*+일'을 예시하고 있고, 또한 오른쪽 음절이 'ㅁ'으로 끝나는 경우의 예시로서, '*+삼'을 예시하고 있다. 이러한 질의어 집합은 연속적으로 발음되는 한국어 음성 인식을 위해 연속하는 음절의 특성을 반영할 수 있도록 안출된 것으로서, 와일드카드(wildcard)로 표현된 바와 같이 대상 음절의 좌측에 어떠한 다른 음절이 위치할 경우에도 음성 인식이 가능하도록 제시된 것이다.

[0038] 마찬가지로, 왼쪽 음절이 'ㄹ'로 끝나는 경우의 예시로서, '칠+*', '팔+*', '일+*'을 예시하고 있고, 또한 왼쪽 음절이 'ㅁ'으로 끝나는 경우의 예시로서, '삼+*'을 예시하고 있다.

[0039] 상기된 본 발명의 다양한 실시예들에 의하면 음향 모델의 구축의 복잡도 및 난이도를 크게 증가시키지 않으면서도 음성 인식의 성능을 향상시키고, 좌, 우 문맥을 고려한 음절에 기반한 음성 인식 모델을 통하여 문맥에 적합한 숫자 후보를 제공할 수 있으며, 음향 모델의 학습 과정에서 부족한 학습 데이터의 보충을 음절 단위로 확장할 수 있다. 이하에서는 제안된 실시예들을 통해 얻어진 성능 실험 결과를 제시하도록 한다.

[0040] 도 5a 및 도 5b는 종래의 음성 인식 방법을 통한 실험 결과와 본 발명의 일 실시예에 따른 숫자 음성 인식 방법을 통한 실험 결과를 비교하여 나타낸 도표이다.

[0041] 실험을 위해 사용된 학습 데이터를 34,560 문장(약 14 시간)이며, 테스트 데이터를 8,640 문장(약 3 시간)의 사연 숫자이다. 음절 기반 인식 단위를 사용한 한국어 연속 숫자음 인식 시스템의 성능을 평가하기 위하여 음소 기반의 음향 모델을 사용하는 베이스라인(baseline) 시스템을 구축하여, 음절 기반의 시스템과 성능을 비교하였다. 물론 음절 기반이라 함은 모노실러블과 트라이실러블을 모두 포함하며, 본 발명의 실시예가 제안하고 있는 방법은 트라이실러블에 기반한 음성 인식 방법이다.

[0042] 우선, 도 5a는 베이스라인 시스템의 성능 측정 결과를 나타낸 도표로서, 각각 모노폰, 트라이폰과 단어 경계를 넘어서 문맥을 고려하는 크로스-워드(cross-word) 트라이폰의 경우도 측정 대상에 포함시켰다. 다음으로, 도 5b는 음절 기반의 시스템의 성능 측정 결과를 나타낸 도표로서, 모노실러블과 트라이실러블의 경우를 측정 대상에 포함시켰다.

[0043] 성능 측정의 기준 요소로 HMM 상태의 개수, 상태(state)당 가우시안(Gaussian) 확률 분포의 개수(G/S), 전체 가우시안 확률 분포의 개수, 오인식률, 인식 속도 등을 사용하였다. 오인식률은 크게 문장 단위의 오류와 단어 단위의 오류로 구분할 수 있다. 실험에 사용된 컴퓨터 시스템의 사양은 모두 동일하다.

[0044] 도 5a 및 도 5b를 참조하면, 음절 기반 시스템의 경우 베이스라인 시스템보다 상대적으로 적은 가우시안 확률 분포를 사용함을 알 수 있다. 모노실러블은 모노폰에 비해 문장 인식률이 저조하였으며, 단어 인식률에서는 좋은 것으로 나타났다. 트라이실러블의 경우에는 크로스-워드 트라이폰을 사용하는 경우보다 약 8.2% 내지8.3% 정도 오인식률이 감소하는 것을 볼 수 있다.

[0045] 한편, 본 발명은 컴퓨터로 읽을 수 있는 기록 매체에 컴퓨터가 읽을 수 있는 코드로 구현하는 것이 가능하다. 컴퓨터가 읽을 수 있는 기록 매체는 컴퓨터 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록 장치를 포함한다.

[0046] 컴퓨터가 읽을 수 있는 기록 매체의 예로는 ROM, RAM, CD-ROM, 자기 테이프, 플로피디스크, 광 데이터 저장장치 등이 있으며, 또한 캐리어 웨이브(예를 들어 인터넷을 통한 전송)의 형태로 구현하는 것을 포함한다. 또한, 컴퓨터가 읽을 수 있는 기록 매체는 네트워크로 연결된 컴퓨터 시스템에 분산되어, 분산 방식으로 컴퓨터가 읽을

수 있는 코드가 저장되고 실행될 수 있다. 그리고 본 발명을 구현하기 위한 기능적인(functional) 프로그램, 코드 및 코드 세그먼트들은 본 발명이 속하는 기술 분야의 프로그래머들에 의하여 용이하게 추론될 수 있다.

[0047] 이상에서 본 발명에 대하여 그 다양한 실시예들을 중심으로 살펴보았다. 본 발명에 속하는 기술 분야에서 통상의 지식을 가진 자는 본 발명이 본 발명의 본질적인 특성에서 벗어나지 않는 범위에서 변형된 형태로 구현될 수 있음을 이해할 수 있을 것이다. 그러므로 개시된 실시예들은 한정적인 관점이 아니라 설명적인 관점에서 고려되어야 한다. 본 발명의 범위는 전술한 설명이 아니라 특허청구범위에 나타나 있으며, 그와 동등한 범위 내에 있는 모든 차이점은 본 발명에 포함된 것으로 해석되어야 할 것이다.

부호의 설명

[0048] 100 : 숫자 음성 인식 장치

10 : 특징 벡터 추출부

20 : 음성 인식부 25 : 음향 모델

30 : 비터비 디코딩부

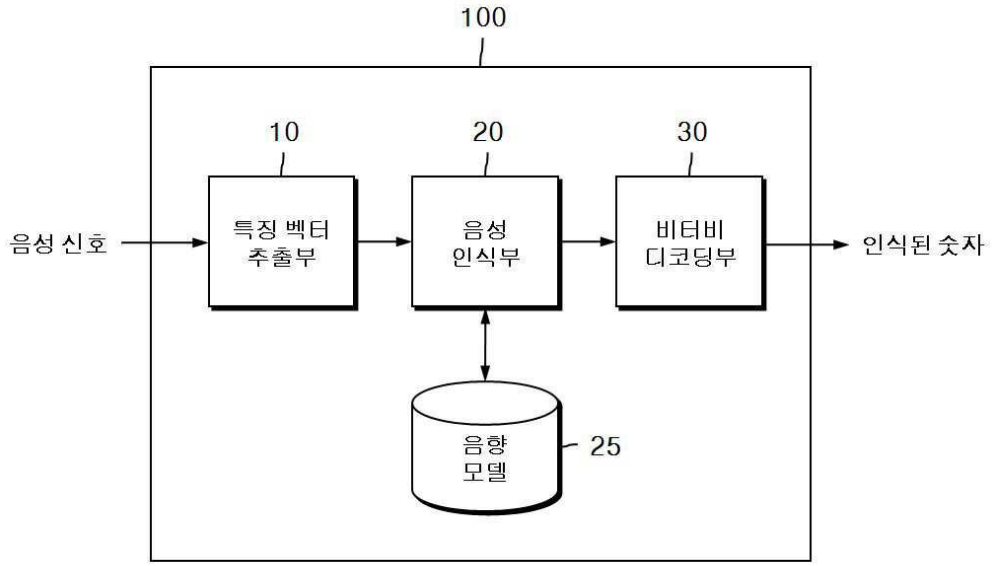
300 : 음향 모델 학습 장치

21 : 특징 벡터 추출부 22 : 음향 모델 학습부

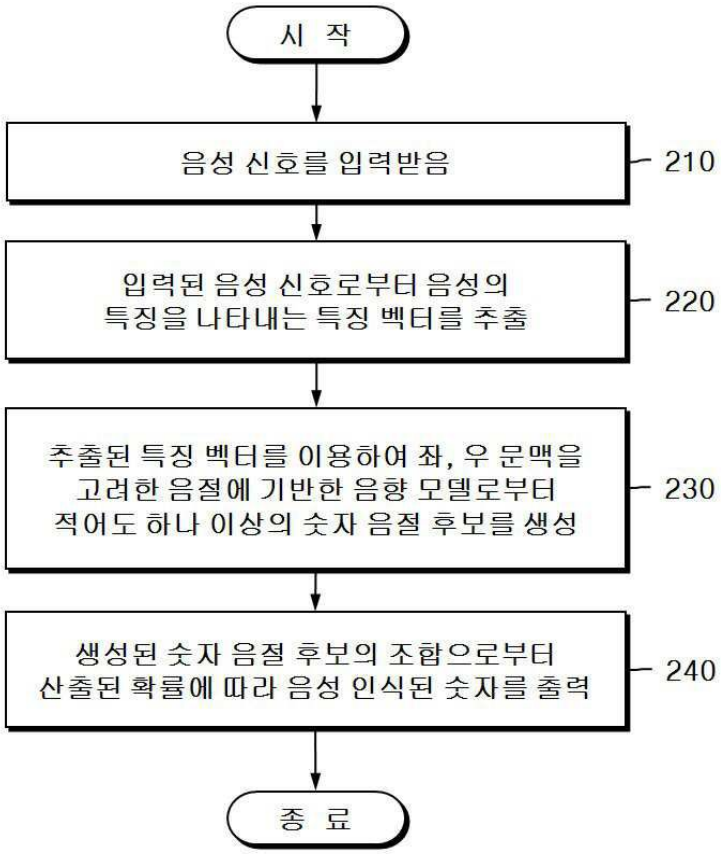
23 : 결정 트리 기반 상태 클러스터링부

도면

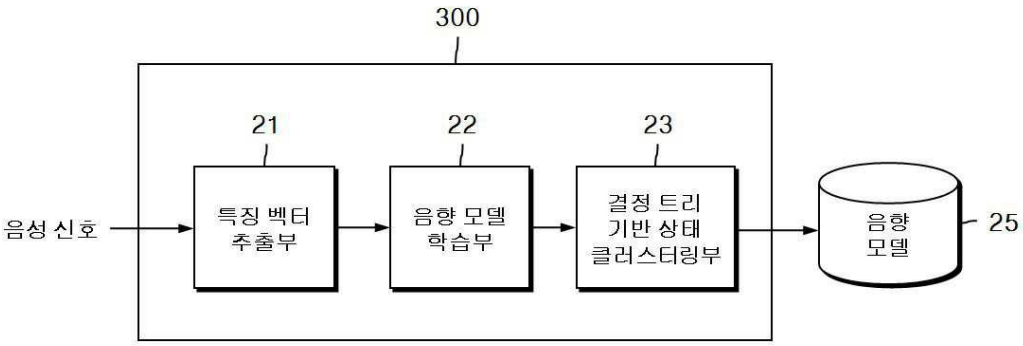
도면1



도면2


시 작
음성 신호를 입력받음
210
입력된 음성 신호로부터 음성의
특징을 나타내는 특징 벡터를 추출
220
추출된 특징 벡터를 이용하여 좌, 우 문맥을
고려한 음절에 기반한 음향 모델로부터
적어도 하나 이상의 숫자 음절 후보를 생성
230
생성된 숫자 음절 후보의 조합으로부터
산출된 확률에 따라 음성 인식된 숫자를 출력
240
종 료

도면3


300
21
22
23
음성 신호
특징 벡터
추출부
음향 모델
학습부
결정 트리
기반 상태
클러스터링부
음향
모델
25

도면4

```
...
QS "R_C-L-Back" { *+chil, *+pal, *+il } // 오른쪽 음절이 'ㄹ' 로 끝나는 경우
QS "R_C-M-Back" { *+sam } // 오른쪽 음절이 'ㅁ' 으로 끝나는 경우
...
QS "L_C-L-Back" { chil-*, pal-*, il-* } // 왼쪽 음절이 'ㄹ' 로 끝나는 경우
QS "L_C-M-Back" { sam-* } // 왼쪽 음절이 'ㅁ' 으로 끝나는 경우
...
```

도면5a

	State	G/S	Gaussian	오인식률 (%)		속도 (RT)
				문장	단어	
Monophone	5	40	1,800	9.49	3.01	0.027
Triphone	5	27	1,214	7.99	2.40	0.055
Cross-word Triphone	5	19	855	4.17	1.45	0.333

도면5b

	State	G/S	Gaussian	오인식률 (%)		속도 (RT)
				문장	단어	
Monosyllable	5,8,11	13	1,014	9.72	2.89	0.166
Trisyllable	5,8,11	12	936	3.82	1.33	0.500