



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2007년11월19일
(11) 등록번호 10-0776729
(24) 등록일자 2007년11월08일

(51) Int. Cl.

G10L 15/00 (2006.01) G10L 15/20 (2006.01)

G10L 15/14 (2006.01)

(21) 출원번호 10-2006-0082521

(22) 출원일자 2006년08월29일

심사청구일자 2006년08월29일

(56) 선행기술조사문헌

JP2002091484 A

(뒷면에 계속)

(73) 특허권자

울산대학교 산학협력단

울산 남구 무거동 산29

(72) 발명자

이정철

대전 중구 태평2동 버드내아파트 215동 801호

김민제

경남 마산시 월영동 주공아파트 113동 502호

(74) 대리인

장한특허법인

전체 청구항 수 : 총 6 항

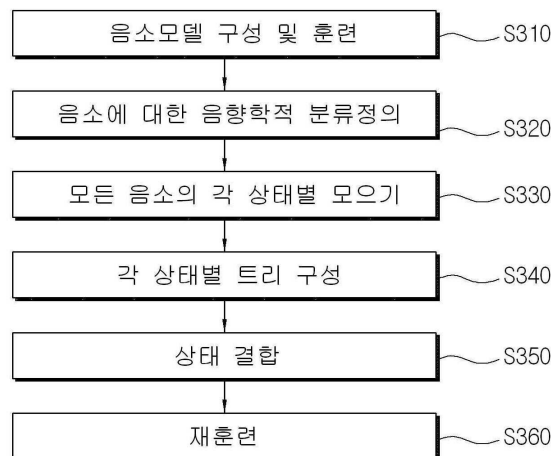
심사관 : 경연정

(54) 결정 트리 기반의 상태 군집화 방법을 사용하는 비핵심어모델링부를 포함하는 화자독립 가변어휘 핵심어 검출시스템 및 그 방법

(57) 요약

본 발명은, 입력되는 음성 구간을 검출하고, 검출된 음성의 특징을 추출하는, 전처리부와, 상기 전처리부로부터의 음성 특징을 이용하고, 핵심어 모델부와 비핵심어 모델부와 목음 모델부를 통해서, 핵심어를 검출하는 가변어휘 핵심어 인식 시스템을 포함하고, 상기 비핵심어 모델링은, 각 음소 간의 음향학적 정보와 통계적인 정보를 모두 사용한 결정 트리 기반 상태 군집화 방법(Decision Tree-based state clustering)을 사용하는 비핵심어 모델링부인 것을 특징으로 하는, 화자독립 가변어휘 핵심어 검출 시스템을 제공한다.

대표도 - 도3



(56) 선행기술조사문헌

KR1019990040215 A

KR1020030080156 A

KR1020060072504 A

US20050228666 A1

한국음향학회지 Vol.25 No.7 pp.319-324

특허청구의 범위

청구항 1

입력되는 음성 구간을 검출하고, 검출된 음성의 특징을 추출하는, 전처리부와,

상기 전처리부로부터의 음성 특징을 이용하고, 핵심어 모델부와 비핵심어 모델부와 묵음 모델부를 통해서, 핵심어를 검출하는 가변어휘 핵심어 인식 시스템을 포함하고,

상기 비핵심어 모델링은, 각 음소 간의 음향학적 정보와 통계적인 정보를 모두 사용한 결정 트리 기반 상태 군집화 방법(Decision Tree-based state clustering)을 사용하는 비핵심어 모델링부인 것을 특징으로 하는, 화자독립 가변어휘 핵심어 검출 시스템.

청구항 2

제 1 항에 있어서,

상기 결정 트리 기반 상태 군집화 방법을 사용하는 비핵심어 모델링부는,

각 음소에 대하여 HMM 모델을 구성하고,

상기 음향학적 정보를 이용하여 복수 개의 문맥 질의(context question)를 구성하고,

상기 수학식 1의 통계학적 평가 함수를 이용하여, 상기 통계학적 평가 함수가 최대값이 되는 상기 문맥 질의를 선택하여 2 개의 부분 집합으로 나누어 가는 과정을 반복하는 것을 통해서 각 상태(state) 별로 트리(tree)를 구성하고,

상기 부분 집합으로 나눈 경우, 상기 평가 함수를 통한 관측 확률 값의 증가가 소정의 임계값보다 작아지게 되면, 분할을 중지하고,

상기 구성된 트리의 리프 노드(leaf node)에 해당하는 음소들의 상태를 결합시키고,

모든 상태가 결합되면, 가우시안 믹스처를 순차적으로 증가시키면서 상기 비핵심어 모델링의 파라미터를 재추정함으로써, 비핵심어 모델이 이루어지는 것을 특징으로 하는, 화자독립 가변어휘 핵심어 검출 시스템.

수학식 1

$$L = \sum_{s \in S} -\frac{1}{2} (n + \ln[(2\pi)^n | \sum s |]) \sum_{t=1}^T \gamma_s(t)$$

여기서, $\sum s$ 는 상태(state) s의 공분산을 의미하며, $\sum \gamma_s(t)$ 는 상태(state) s가 발생하는 프레임(frame) 수를 의미하고, n은 특징 벡터(characteristic vector)의 차수를 의미하고, S는 문맥 질의에 따른 집합($s \in S$)을 나타낸다.

청구항 3

제 2 항에 있어서,

상기 파라미터는, 가우시안 분포의 평균과 분산값 및 각 상태 간의 천이확률인 것을 특징으로 하는, 화자독립 가변어휘 핵심어 검출 시스템.

청구항 4

입력되는 음성 구간을 검출하고, 검출된 음성의 특징을 추출하는, 전처리 단계와,

상기 전처리 단계로부터의 음성 특징을 이용하고, 핵심어 모델과 비핵심어 모델과 묵음 모델을 통해서, 핵심어를 검출하는 가변어휘 핵심어 인식 단계를 포함하고,

상기 비핵심어 모델은, 각 음소 간의 음향학적 정보와 통계적인 정보를 모두 사용한 결정 트리 기반 상태 군집

화 방법(Decision Tree-based state clustering)을 사용하는 비핵심어 모델인 것을 특징으로 하는, 화자독립 가변어휘 핵심어 검출 방법.

청구항 5

제 4 항에 있어서,

상기 결정 트리 기반 상태 군집화 방법을 사용하는 비핵심어 모델은,

각 음소에 대하여 HMM 모델을 구성하는 단계,

상기 음향학적 정보를 이용하여 복수 개의 문맥 질의(context question)를 구성하는 단계,

상기 수학적 1의 통계학적 평가 함수를 이용하여, 상기 평가 함수가 최대값이 되는 상기 문맥 질의를 선택하여 2 개의 부분 집합으로 나누어 가는 과정을 반복하는 것을 통해 각 상태(state) 별로 트리(tree)를 구성하는 단계,

상기 부분 집합으로 나눈 경우, 상기 평가 함수를 통한 관측 확률 값의 증가가 소정의 임계값보다 작아지게 되면, 분할을 중지하는 단계,

상기 구성된 트리의 리프 노드(leaf node)에 해당하는 음소들의 상태를 결합시키는 단계, 및

모든 상태가 결합되면, 가우시안 믹스처를 순차적으로 증가시키면서 상기 비핵심어 모델링의 파라미터를 재추정하는 단계로 이루어지는 것을 특징으로 하는, 화자독립 가변어휘 핵심어 검출 방법.

수학적 1

$$L = \sum_{s \in S} -\frac{1}{2} (n + \ln[(2\pi)^n | \sum s |]) \sum_{t=1}^T \gamma_s(t)$$

여기서, $\sum_s$ 는 상태(state) s의 공분산을 의미하며, $\sum \gamma_s(t)$ 는 상태(state) s가 발생하는 프레임(frame) 수를 의미하고, n은 특징 벡터(characteristic vector)의 차수를 의미하고, S는 문맥 질의에 따른 집합($s \in S$)을 나타낸다.

청구항 6

제 5 항에 있어서,

상기 파라미터는, 가우시안 분포의 평균과 분산값 및 각 상태 간의 전이확률인 것을 특징으로 하는, 화자독립 가변어휘 핵심어 검출 방법.

명세서

발명의 상세한 설명

발명의 목적

발명이 속하는 기술 및 그 분야의 종래기술

<13> 본 발명은 결정 트리 기반의 상태 군집화 방법을 사용하는 비핵심어 모델링부를 포함하는 화자독립 가변어휘 핵심어 검출 시스템 및 그 방법에 관한 것으로서, 더 자세하게는, 기존의 음성인식 시스템에 변경을 가하지 않으면서도 음성인식에서 핵심어 검출 성능을 향상시킬 수 있도록, 비핵심어 모델링부를 본 발명이 제안하는 결정 트리 기반의 상태 군집화 방법을 사용하는 비핵심어 모델링부로 교체한 화자독립 가변어휘 핵심어 검출 시스템 및 그 방법에 관한 것이다.

<14> 최근의 음성인식 기술은 과거와 비교하여 많은 주목을 받고 있는 상황이다. 이것은, 음성인식 기술이 구현될 경우, 인터넷 정보 이용이나 전자 상거래, 각종 유비쿼터스 환경을 보다 편리하게 음성을 통하여 조작할 수 있을

뿐만 아니라, 손을 사용하기 어려운 차량 운행 등과 같은 경우에도 인터넷을 액세스한다든가 사무나 기타 서비스를 이용할 수 있게 되어, 사용자의 다양한 니즈(needs)를 충족시켜줄 수 있기 때문이다. 또한, 이와 같은 음성인식은 각 사용자의 독특한 음성 주파수를 인식하기 때문에, 보안성(security) 문제도 해결할 수 있는 장점을 갖고 있다.

- <15> 이하에서는, 이와 같이 최근에 부각되고 있는 음성인식 기술이 종래에 갖고 있는 문제점을 살펴보도록 한다.
- <16> 일반적으로 음성인식 시스템에서 핵심어 검출이란 연속적인 음성 중에서 상대적으로 중요성이 높은 특정 단어를 인식함으로써 일을 수행하는 것을 말한다. 핵심어 검출은, 입력 음성 중 핵심어를 처리하는 핵심어 모델들과, 핵심어 이외의 나머지 부분을 처리하는 비핵심어 모델(garbage model)로 나뉘어져서, 인식대상 어휘 내의 핵심어를 검출하게 된다.
- <17> 따라서, 핵심어 검출 시스템의 성능 향상을 위하여 적절한 비핵심어 모델의 선택이 필요하고, 비핵심어 모델의 성능은 이것을 어떻게 정의하고 구현할 것이며, 또한 핵심어 부분을 잠식하지 않으면서 비핵심어 부분 및 배경 잡음 부분을 얼마만큼 효과적으로 표현해 줄 수 있는가에 따라 크게 좌우된다.
- <18> 종래의 음성인식 기술에는 비핵심어 모델을 구성하기 위해서 두 가지 방법을 사용하고 있었다. 즉, 통계적인 방법을 사용하여 monophone 모델을 군집화(clustering) 하기 위하여 가중치 유클리드 거리함수(weighted Euclidean distance)를 사용한 변환된 K 평균 알고리즘(modified k-means algorithm)을 적용하여 군집화하는 방법을 이용하고 있었다.
- <19> 그리고, 비핵심어 음성 구간 전체를 하나의 상태(state)로 두어 다중 믹스처(mixtures)로 표현하는 방식이 있었다. 이를 위하여 먼저 훈련 데이터베이스(DB)를 이용하여 비핵심어 구간 전체를 단일 상태 단일 믹스처(single state single mixture)를 가지는 HMM(Hidden Markov Model)으로 구성한 후, 상태(state)의 믹스처 개수를 하나씩 증가시키면서 재훈련하는 과정을 반복하는 절차를 거치게 되는 방법을 채용하고 있었다.
- <20> 상기와 같은 종래의 K 평균 군집화(K-means clustering) 방법은, K개의 그룹으로 군집화하여 K개의 비핵심어 모델을 사용한다. 하지만, 이 방법은 몇 가지 문제점이 발생한다. 먼저, 초기 K개의 중심값과 거리측정 함수에 민감한 특징을 가지고 있으며, 단순히 각 모델에서 모든 상태에서 발생하는 평균과 분산을 이용하여 가중치 유클리드 거리함수를 구하게 되어, 각 상태에 대한 특징을 반영하지 못하게 되는 문제점을 갖고 있었다. 두 번째로, HMM을 구성할 경우, 중요한 파라미터인 상태(state) 사이의 천이 확률은 거리 계산시 참여하지 못하는 단점이 있었다. 세 번째로, 데이터의 발생 빈도와 상관없이 단순히 음소 44개의 모델을 K개로 군집화하기 때문에 군집을 위한 데이터가 너무 적어 신뢰도 높은 모델을 만들기 어렵다는 문제점을 또한 갖고 있었다. 네 번째로, 결과에 대한 해석이 어렵기 때문에 이후 새로운 데이터를 이용하여 재훈련이 어렵다는 문제점을 갖고 있었다.
- <21> 한편, 상기의 두 번째 방법인 GMM(Gaussian Mixture Model) 방법은 음성구간을 단일상태(single state)로 구성한 다음, 가우시안 믹스처의 개수를 증가시키며 비핵심어모델을 구성하게 된다. 이 방법은 훈련과정은 간단하지만 단어 전체를 하나의 모델로 구성하였기 때문에 배경잡음이나 간투사처럼 음절의 특성을 가지는 부분을 잘 표현하지 못한다는 문제점을 안고 있었다. 상술한 K 평균 군집화 방법 및 GMM 방법은, 각각의 조건과 경우에 따라 그 성능이 우위를 보이는 것이 달라 지게 되기 때문에, 성능상 어느 쪽이 우위에 있다고 단정적으로 얘기할 수 없지만, 종래의 핵심어 검출에서 현재 주류를 이루는 것은 상기 2가지 방법이다.
- <22> 한편, 핵심어 검출 성능을 높이기 위해서, 비핵심어 모델의 성능을 향상시키는 것과 대조적으로, 대한민국 공개특허공보 특2002-0045960(2002년 6월 20일 공개)에 기재되어 있는 종래의 기술에 따르면, 반응소 모델을 생성하는 방법을 제안하여 이를 안티-모델로 사용하고, 핵심어 검증 방법으로는 입력된 단어의 각 음소마다 정상적인 음소 모델과 안티-모델과의 유사도를 이용한 신뢰도 측정 방법을 사용함으로써 반응소 모델과 발화 검증 방법을 사용하여 핵심어 검증 성능을 향상시키고자 하고 있다. 다시 말해, 상기 공보에 공개된 종래 기술은, 핵심어 검출 시스템에서 핵심어 인식의 성능을 향상시키기 위하여, 후처리 방법을 제안하여 핵심어 인식의 성능 향상을 꾀하였다.

발명이 이루고자 하는 기술적 과제

- <23> 본 발명은 상기와 같은 종래 기술들의 문제점 및 단점들을 해결, 극복하기 위하여 안출된 것으로서, 본 발명은, 개선된 K 평균 군집화 방법, 즉 결정 트리 기반의 상태 군집화 방법을 사용하는 비핵심어 모델부를 포함하는 독자 독립 가변어휘 핵심어 검출 시스템 및 그 방법을 제공하는 것을 그 목적으로 한다.

<24> 다시 말해, 종래 기술에서 상술한 기술 중, 종래의 음소를 K 평균 알고리즘에 의해 K개의 클러스터로 그룹핑(grouping)하는 방법의 단점을 보완한 것으로서, 본 발명은 결정 트리 기반의 상태 군집화 방법(Decision Tree-based state clustering)을 이용하여 상태(state)를 묶는(tying) 방법을 사용하는 방법을 이용하여 비핵심어 모델링부의 성능을 향상시키는 것을 목적으로 하고 있다.

<25> 본 발명의 다른 목적 및 장점들은 하기에 설명될 것이며, 본 발명의 실시예에 의해 알게 될 것이다. 또한, 본 발명의 목적 및 장점들은 특허청구범위에 나타난 수단 및 조합에 의해 실현될 수 있다.

발명의 구성 및 작용

<26> 상술한 목적을 달성하기 위해서, 본 발명에 따른 일 실시예는, 입력되는 음성 구간을 검출하고, 검출된 음성의 특징을 추출하는, 전처리부와; 상기 전처리부로부터의 음성 특징을 이용하고, 핵심어 모델부와 비핵심어 모델부와 묵음 모델을 통해서, 핵심어를 검출하는 가변어휘 핵심어 인식 시스템;을 포함하고, 상기 비핵심어 모델링은, 각 음소 간의 음향학적 정보와 통계적인 정보를 모두 사용한 결정 트리 기반 상태 군집화 방법(Decision Tree-based state clustering)을 사용하는 비핵심어 모델링부인 것을 특징으로 하는, 화자독립 가변어휘 핵심어 검출 시스템을 제공한다.

<27> 또한, 상기 결정 트리 기반 상태 군집화 방법을 사용하는 비핵심어 모델링부는, 각 음소에 대하여 HMM모형을 구성하고, 상기 음향학적 정보를 이용하여 복수 개의 문맥 질의(context question)를 구성하고, 상기 수학적 1의 통계학적 평가 함수를 이용하여, 상기 통계학적 평가 함수가 최대값이 되는 상기 문맥 질의를 선택하여 2 개의 부분 집합으로 나누어 가는 과정을 반복하는 것을 통해서 각 상태(state) 별로 트리(tree)를 구성하고, 상기 부분 집합으로 나눈 경우, 상기 평가 함수를 통한 관측 확률 값의 증가가 소정의 임계값보다 작아지게 되면, 분할을 중지하고, 상기 구성된 트리의 리프 노드(leaf node)에 해당하는 음소들의 상태를 결합시키고, 모든 상태가 결합되면, 가우시안 믹스처를 순차적으로 증가시키면서 상기 비핵심어 모델링의 파라미터를 재추정함으로써, 비핵심어 모델이 이루어진다.

<28> 수학적 1

$$L = \sum_{s \in S} -\frac{1}{2} (n + \ln[(2\pi)^n | \sum s |]) \sum_{t=1}^T \gamma_s(t)$$

<29> 여기서, $\sum s$ 는 상태(state) s의 공분산을 의미하며, $\sum \gamma_s(t)$ 는 상태(state) s가 발생하는 프레임(frame) 수를 의미하고, n은 특징 벡터(characteristic vector)의 차수를 의미하고, S는 문맥 질의에 따른 집합(s ∈ S)을 나타낸다.

<31> 또한, 상기 파라미터는, 가우시안 분포의 평균과 분산값 및 각 상태 간의 전이확률이다.

<32> 한편, 본 발명의 일 실시예에 따르면, 입력되는 음성 구간을 검출하고, 검출된 음성의 특징을 추출하는, 전처리 단계와, 상기 전처리 단계로부터의 음성 특징을 이용하고, 핵심어 모델과 비핵심어 모델과 묵음 모델을 통해서, 핵심어를 검출하는 가변어휘 핵심어 인식 단계를 포함하고, 상기 비핵심어 모델은, 각 음소 간의 음향학적 정보와 통계적인 정보를 모두 사용한 결정 트리 기반 상태 군집화 방법(Decision Tree-based state clustering)을 사용하는 비핵심어 모델인 것을 특징으로 하는, 화자독립 가변어휘 핵심어 검출 방법을 제공한다.

<33> 또한, 상기 결정 트리 기반 상태 군집화 방법을 사용하는 비핵심어 모델은, 각 음소에 대하여 HMM 모델을 구성하는 단계, 상기 음향학적 정보를 이용하여 복수 개의 문맥 질의(context question)를 구성하는 단계, 상기 수학적 1의 통계학적 평가 함수를 이용하여, 상기 평가 함수가 최대값이 되는 상기 문맥 질의를 선택하여 2 개의 부분 집합으로 나누어 가는 과정을 반복하는 것을 통해 각 상태(state) 별로 트리(tree)를 구성하는 단계, 상기 부분 집합으로 나눈 경우, 상기 평가 함수를 통한 관측 확률 값의 증가가 소정의 임계값보다 작아지게 되면, 분할을 중지하는 단계, 상기 구성된 트리의 리프 노드(leaf node)에 해당하는 음소들의 상태를 결합시키는 단계, 및 모든 상태가 결합되면, 가우시안 믹스처를 순차적으로 증가시키면서 상기 비핵심어 모델링의 파라미터를 재추정하는 단계로 이루어진다.

<34> 수학적 식 1

$$L = \sum_{s \in S} -\frac{1}{2} (n + \ln[(2\pi)^n | \Sigma_s |]) \sum_{t=1}^T \gamma_s(t)$$

<35>

<36> 여기서, Σ_s 는 상태(state) s의 공분산을 의미하며, $\Sigma \gamma_s(t)$ 는 상태(state) s가 발생하는 프레임(frame) 수를 의미하고, n은 특징 벡터(characteristic vector)의 차수를 의미하고, S는 문맥 질의에 따른 집합($s \in S$)을 나타낸다.

<37> 또한, 상기 파라미터는, 가우시안 분포의 평균과 분산값 및 각 상태 간의 전이확률이다.

<38> 이하 첨부된 도면을 참조로 본 발명의 바람직한 실시예를 상세히 설명하기로 한다. 이에 앞서, 본 명세서 및 청구범위에 사용된 용어나 단어는 통상적이거나 사전적인 의미로 한정해서 해석되어서는 아니되며, 발명자는 그 자신의 발명을 가장 최선의 방법으로 설명하기 위해 용어의 개념을 적절하게 정의할 수 있다는 원칙에 입각하여 본 발명의 기술적 사상에 부합하는 의미와 개념으로 해석되어야만 한다.

<39> 따라서, 본 명세서에 기재된 실시예와 도면에 도시된 구성은 본 발명의 가장 바람직한 일 실시예에 불과할 뿐이고 본 발명의 기술적 사상을 모두 대변하는 것은 아니므로, 본 출원 시점에 있어서 이들을 대체할 수 있는 다양한 균등물과 변형예들이 있을 수 있음을 이해하여야 한다.

<40> 도 1은 본 발명에 따른 가변어휘 핵심어 인식 시스템의 네트워크를 나타내는 도면이며, 도 2는 본 발명의 일 실시예에 따른 도 2는 본 발명의 일 실시예에 따른 결정 트리 기반의 상태 군집화 방법을 사용하는 비핵심어 모델링부를 포함하는 화자독립 가변어휘 핵심어 검출 시스템을 나타내는 블록 다이어그램이다. 이하에서, 도 1 및 도 2를 참조하여 설명하기로 한다.

<41> 본 발명에 따른 가변어휘 핵심어 검출 시스템에서의 인식 네트워크는 핵심어 모델과 비핵심어 모델, 그리고 묵음모델의 결합으로 구성하였으며, 입력음성에는 핵심어가 반드시 하나만이 들어 있다고 가정하여 구성되었다. 그래서 입력음성이 들어오면 인식과정을 통하여 하나의 핵심어를 검출하게 된다.

<42> 여기서 언급되고 있는 음성인식 기술에 대하여 간략하게 살펴보면, 음성인식 기술은 크게 세 가지로 구분할 수 있는데, 인식대상에 따른 분류와 어휘에 따른 분류, 그리고 발음에 따른 분류로 나눌 수 있다.

<43> 인식 대상에 따라 '화자종속'과 '화자독립'으로 다시 나누어지는데, '화자종속'이란 학습된 화자의 명령만을 수행하는 기술이다. 즉, 화자가 미리 녹음을 통해 기계를 학습시킨 후에 실행하게 하는 것으로 화자 이외의 다른 사람의 명령은 인식을 보장할 수 없다. '화자독립'은 불특정 다수의 화자음성을 인식하게 하는 기술로 여러 명이 함께 사용할 수 있다. 당연히 화자독립 기술이 화자종속 기술보다 난이도가 높은 기술이다.

<44> 어휘는 다시 어휘종속과 어휘독립 기술로 나뉘는데, 어휘종속은 인식시키고자 하는 특정 단어, 즉 '우리집'과 같이 연결되지 않는 단독 어휘를 학습시킨 후에 인식하는 것이다. 어휘종속은 어휘의 크기에 따라 소용량 어휘 인식, 대용량 어휘인식으로 재분류되는데, 대용량 어휘인식은 Dictation(낭독) 기술에서 볼 수 있듯이 수만 단어 이상을 인식하는 기술이다.

<45> 어휘독립은 가변 어휘인식으로도 불리는데, 가변어휘 인식은 단어가 한정된 것이 아니라 기존에 학습된 어휘의 음소 특징을 추출, 조합하여 인식된 어휘 외의 어휘도 인식할 수 있는 기술이다.

<46> 발음에 따른 분류는 다시 분절음 인식과 연속음 인식으로 나뉘는데, 분절음 인식이란, 고립단어, 즉, '우리집'과 같은 고립된 단어를 인식하는 것과 고립단어가 연결된 형태, 예를 들어 "Check E-mail" 등을 인식하는 연결단어 인식, 그리고 핵심어만 인식하는 핵심어 인식으로 문장에서 핵심어를 추출하는 기술로 구성된다.

<47> 핵심어 인식을 예를 들면 "동경에서 뉴욕까지..."의 문장에서 '동경'과 '뉴욕'이라는 핵심어를 추출하여 인식하는 것이다.

<48> 연속음 인식은 자연스런 문장을 인식하는 기술로 문장 전체를 모두 인식하는 것으로 음성인식 기술의 최종 관문으로 인식되고 있다.

<49> 여기서, 본 발명이 다루는 음성인식기술은 화자독립 가변어휘 핵심어 검출 시스템을 다루고 있으며, 이것은 본

명세서 전반에 해당되는 것이다.

- <50> 다시 도 2로 돌아가서 설명을 하면, 본 발명에 따른 스위칭 가능한 비핵심어 모델링부를 포함하는 화자독립 가변어휘 핵심어 검출 시스템은, 전처리부(100)와, 가변어휘 핵심어 인식 시스템(200)을 포함한다.
- <51> 상기 전처리부(100)는, 입력되는 음성 구간을 검출하고 검출된 음성의 특징을 추출하는 구성요소(element)이다. 또한, 상기 가변어휘 핵심어 인식 시스템(200)은, 상기 전처리부(100)로부터의 음성 특징을 이용하여 핵심어를 검출하는 구성요소이다.
- <52> 그럼, 상기 가변어휘 핵심어 인식 시스템(200)에 대해서 좀 더 자세히 살펴보면, 상기 인식 시스템(2)은, 변경어휘 입력부(250), 인식 대상 목록부(255), 발음 변환기(230), 스피치 데이터베이스(290), 은닉 마르코브 추측부(280), 서브 폰 모델부(270), 핵심어 모델부(240), 비핵심어 모델부(260), 묵음 모델부(265), 네트워크 구성부(220), 및 핵심어 검출부(210)를 포함한다.
- <53> 상기 변경 어휘 입력부(250)는, 핵심어 모델부를 구성하는 기존 핵심어에 추가하거나 다른 핵심어로 변경하고자 할 경우 사용되는 부분으로, 변경 어휘가 인식 대상 단어 목록부(255)에 저장되도록 입력되는 유닛이다. 또한, 인식 대상 단어 목록부(255)는 변경 어휘 입력부(250)를 통해서 입력되는 변경 어휘를 반영하여 저장하는 유닛이다.
- <54> 또한, 발음 변환기(230)는, 상기 인식단어 목록부(255)에 저장되어 있는 단어를 한국어 표준 발음법을 바탕으로 하여 인식 어휘를 발음 표기로 변환하는 유닛이다. 이와 같은 발음 표기로 변환된 인식 어휘를 음소열로 변환시킨 후, 스피치 데이터베이스(290)에 저장된 음성 데이터로부터 훈련되어 지는 서브 폰(sub-phone) 모델부(270)를 이용하여 핵심어 모델부(240)에서 핵심어 모델을 구성하게 된다.
- <55> 그 이후 핵심어 모델부(240)의 구성된 핵심어 모델과 비핵심어 모델부(260)에서 모델링된 비핵심어 모델을 이용하여, 도 1과 같은 구조를 가지는 네트워크를 구성하여 핵심어를 검출하게 된다.
- <56> 이하에서, 본 발명의 가변어휘 핵심어 인식 시스템(200)에서 사용되는 비핵심어 모델을 효율적으로 모델링하는 방법을 도 3을 참조하여 상세히 설명하기로 한다.
- <57> 도 3은 본 발명에 따른 결정 트리 기반의 상태 군집화 방법(Decision Tree-based state clustering)의 순서도이다. 이 방법은, 음소 모델의 결정 트리 기반 상태 군집화(Decision Tree-based State Clustering)를 이용한 비핵심어 모델링 방법이며, 이것은 각 음소 사이의 음성학적 정보와 통계적인 정보를 모두 사용한 결정 트리 기반 비핵심어 모델의 상태 군집화(Decision Tree-based State of Garbage model Clustering) 방법이다. 이 방법은 각 음소의 음향학적 특징과 평가함수를 고려하여 각 모델의 상태를 유사도에 따라 결합시켜 각 상태의 특징을 반영하였고, 상태 사이의 천이확률은 음소 모델의 것을 그대로 사용할 수 있다. 음향학적을 고려하여 45개의 문맥 질의(context question) 구성하였다. 그리고 평가함수는 하기 수학식 1과 같이 정의하였다.
- <58> 수학식 1

$$L = \sum_{s \in S} -\frac{1}{2} (n + \ln[(2\pi)^n | \sum s |]) \sum_{t=1}^T \gamma_s(t)$$

- <59>
- <60> 여기서 $\sum s$ 는 상태(state) s의 공분산을 의미하며, $\sum \gamma_s(t)$ 는 상태(state) s가 발생하는 프레임(frame) 수를 의미하고, n은 특징 벡터(characteristic vector)의 차수를 의미한다. 그리고 S는 문맥 질의(context question)에 따른 집합($s \in S$)을 나타낸다. 여기서, 공분산과 발생 수는 계산의 편리함을 위해 각 음소의 훈련 과정에서 구해진다.
- <61> 상기과 같이 본 발명에서 제안하고 있는 비핵심어 모델 구성 방법은 다음과 같은 단계를 거친다.
- <62> 먼저, 충분한 훈련 데이터를 이용하여 각 음소에 대하여 3 상태(state)를 가지는 HMM모델을 구성한다. 이것은 기존의 문맥종속 모델로 구성하는 과정에서 얻을 수 있기 때문에, 특별한 훈련 절차 없이 구성할 수 있다(S310). 또한, 음소에 대한 음향학적 분류를 정의한다(S320). 즉, 음향학적 특성을 고려하여 45개의 문맥 질의(context question)를 구성한다.

- <63> 다음으로, 각 음소에 대한 모델이 구성되면, 모든 음소모델을 상태(state) 별로 모은다(S330).
- <64> 그 다음, 첫 번째 상태에 대하여 평가함수가 최대가 되게 하는 문맥 질의(context question)을 선택하여 두 개의 부분집합으로 나누어 가는 일련의 과정을 반복하는 것을 통해서 트리(tree)를 구성하여 간다(S340).
- <65> 이 경우, 부분 집합으로 분리했을 때, 평가함수를 통한 관측 확률 값의 증가가 미리 정의한 임계값(threshold)보다 작아지는 시점에서 분할을 멈추게 된다. 이러한 과정은 모든 상태(state)에 대해 수행하게 된다. 임계값을 적절하게 두어 상태 수를 줄이게 된다.
- <66> 여기서 말하는 관측 확률 값이라는 것은 문맥질의(context question)에 의해 분리되었을 때 각 분류의 확률값을 말한다. 또한, 미리 정의한 임계값은 보통 실험을 통하여 상태의 수가 원하는 만큼 줄어들 수 있도록 임계값을 설정하는 것이 바람직하다. 임계값은 훈련 데이터에 따라 달라지며, 상태수를 얼마만큼 줄이느냐에 따라 달라지는 값이므로, 특정한 값을 설정할 수 없고 실험을 통하여 변화하는 조건에 따라서 임계값을 조정할 수 있도록 하는 것이 바람직하다.
- <67> 다음으로, 상태(state)가 묶여지면(tying), 훈련 데이터를 이용하여 믹스처(mixture)의 개수를 순차적으로 증가시키면서 재훈련(즉, 파라미터를 재추정)하여 비핵심어 모델을 구성하였다(S360). 여기서, 언급하고 있는 믹스처 개수는 통상 1부터 순차적으로 증가시켜 나가는 것이 바람직하다. 또한, 파라미터는 통상 가우시안 분포의 평균과 분산값 및 각 상태 간의 천이확률을 의미한다.
- <68> 이와 같은 구성의 결과, 각 상태(state)에 따라 다른 음소들로 결합되는 것을 확인하였다.
- <69> 이렇게 구성된 상기 비핵심어 모델부를 본 발명의 핵심어 인식 시스템(200)의 인식 네트워크(6)에 적용하였을 경우, 기존의 종래 핵심어 인식 시스템 또는 방법에 비하여 가변어휘 핵심어 검출의 성능을 높여줄 수 있는 실험 결과를 도출함을 알 수 있었다.
- <70> 이상과 같이, 본 발명은 비록 한정된 실시예와 도면에 의해 설명되었으나, 본 발명은 이것에 의해 한정되지 않으며 본 발명이 속하는 기술 분야에서 통상의 지식을 가진 자에 의해 본 발명의 기술 사상과 아래에 기재될 특허청구범위의 균등 범위 내에서 다양한 수정, 변경 및 변형이 가능함은 물론이다.

발명의 효과

- <71> 이상에서 설명한 바와 같이, 본 발명에 따른 화자독립 가변어휘 핵심어 검출 시스템은, 종래의 핵심어 검출 시스템에서 비핵심어 모델링부를 비핵심어 부분과 배경잡음 부분의 제거 성능이 양호한, 본 발명이 제안하는 결정 트리 기반의 상태 군집화 방법을 사용하는 비핵심어 모델부로 교체함으로써, 손쉽게 핵심어 검출 시스템의 성능을 향상시킬 수 있다.
- <72> 다시 말해, 이는 기존의 핵심어 인식 시스템의 구조를 그대로 유지하면서 인식 네트워크 구성시 비핵심어 모델을 본 발명이 제안하는 결정 트리 기반의 상태 군집화 방법을 사용하는 비핵심어 모델부로 교체함으로써 성능을 향상시킬 수 있어서, 음성인식 서비스의 질을 손쉽게 향상시킬 수 있는 효과가 있다.
- <73> 또한, 본 발명에 따르면, 새로운 비핵심어 모델링 방법을 제안하여, 연속적인 음성 중에서 상대적으로 중요성이 높은 특정 단어를 인식함으로써 일을 수행하는 핵심어 검출 시스템에서 비핵심어를 효율적으로 모델링하여 오인식 결과를 감소시킬 수 있는 결과를 얻을 수 있다.
- <74> 또한, 본 발명에 따른 핵심어 검출 시스템은 핵심어 검출 기반 음성인식 시스템을 사용하는 모든 분야의 HCI의 입력 장치로 활용될 수 있다.

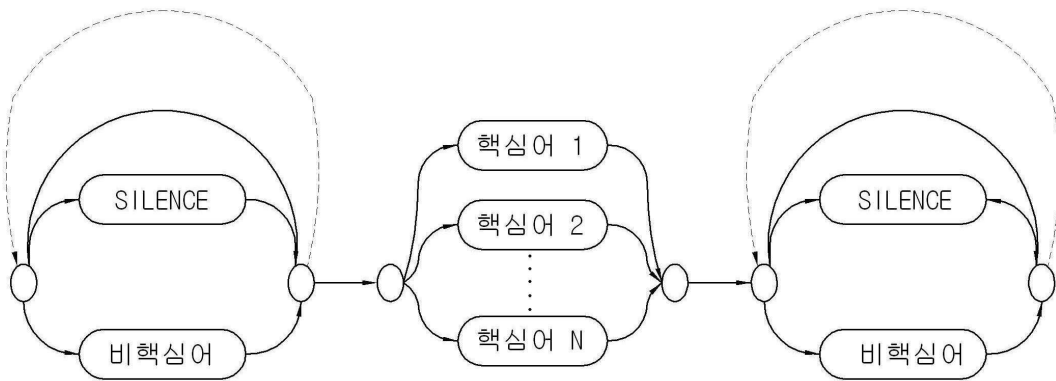
도면의 간단한 설명

- <1> 도 1은 본 발명에 따른 화자독립 가변어휘 핵심어 인식 시스템의 네트워크를 나타내는 도면이며,
- <2> 도 2는 본 발명의 일 실시예에 따른 결정 트리 기반의 상태 군집화 방법을 사용하는 비핵심어 모델링부를 포함하는 화자독립 가변어휘 핵심어 검출 시스템을 나타내는 블록 다이어그램이며,
- <3> 도 3은 본 발명에 따른 비핵심어 모델링의 일례로서, 결정 트리 기반의 상태 군집화 방법(Decision Tree-based state clustering)의 순서도이고,
- <4> <도면의 주요 부분에 대한 부호의 설명>

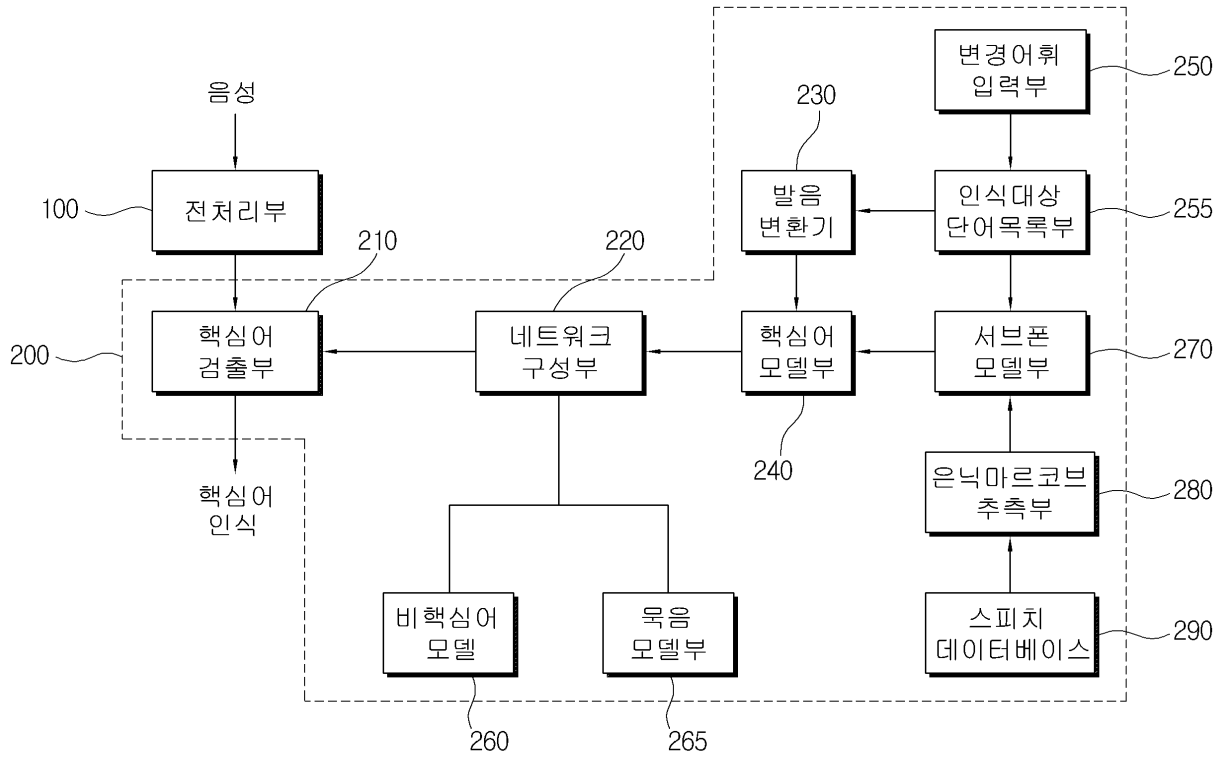
<5>	100 전처리부	200 가변어휘 핵심어 인식 시스템
<6>	210 핵심어(keyword) 검출부	220 네트워크 구성부(network construction)
<7>	230 발음 변환기	240 핵심어 모델부
<8>	250 변경 어휘 입력부	255 인식 대상 단어 목록부
<9>	260 비핵심어 모델(garbage model)	265 묵음 모델부
<10>	270 서브 폰(sub-phone) 모델부	
<11>	280 은닉 마르코브 추측부(Hidden Markov Model estimation)	
<12>	290 스피치 데이터베이스(speech database)	

도면

도면1



도면2



도면3

