



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2011년09월21일
(11) 등록번호 10-1066472
(24) 등록일자 2011년09월15일

(51) Int. Cl.
G10L 15/06 (2006.01) G10L 15/08 (2006.01)
(21) 출원번호 10-2009-0087154
(22) 출원일자 2009년09월15일
심사청구일자 2009년09월15일
(65) 공개번호 10-2011-0029471
(43) 공개일자 2011년03월23일
(56) 선행기술조사문헌
KR1020080043035 A*
KR1020030061797 A*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
국민대학교산학협력단
서울특별시 성북구 정릉동 861-1 국민대학교내
(72) 발명자
김남운
경기도 의정부시 호원동 한승아파트 101-304
김기두
서울특별시 강남구 대치동 선경아파트 5동 603호
(74) 대리인
특허법인남춘

전체 청구항 수 : 총 11 항

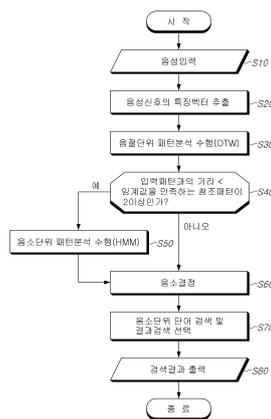
심사관 : 정성윤

(54) 초성 기반 음성인식장치 및 음성인식방법

(57) 요약

본 발명은 초성 기반 음성인식장치 및 음성인식방법에 관한 것으로서, 음성을 입력받는 음성입력부, 수신된 음성 신호의 특징벡터를 산출하여 입력패턴을 생성하는 특징분석부, 참조패턴들이 미리 저장되는 참조패턴저장부, 입력패턴과 참조패턴을 비교하여 입력패턴과 유사한 참조패턴에 대응하는 초성정보를 출력하는 패턴인식부, 그리고 초성정보에 대응하는 단어를 검색하여 출력할 단어를 결정하는 결정부를 포함하는 초성 기반 음성인식장치에 있어서, 상기 참조패턴은, 각각의 자음에 동일한 모음이 조합된 음절을 기초로 하여 생성된 패턴이고; 상기 패턴인식부는, 상기 입력패턴에 대하여 음절단위 비교 알고리즘을 수행하여 유사한 참조패턴을 검출하되, 유사한 참조패턴이 둘 이상 검출되는 경우에 한하여 음소단위 비교 알고리즘을 수행함으로써 최종적으로 유사한 참조패턴 하나를 검출한다. 이와 같은 본 발명에 의하면 임베디드 시스템에 적합한 음성인식 기술을 채용하여 간단하면서도 정확성 있는 초성 기반의 음성인식을 가능하게 하고, 초성 기반 음성인식에서 검색된 단어의 중요도에 따라 결과를 출력함으로써 보다 향상된 검색 기능을 제공할 수 있다는 이점이 있다.

대표도 - 도2



특허청구의 범위

청구항 1

음성을 입력받아 음성신호로 변환하는 음성입력부와;
 상기 음성입력부로부터 수신된 상기 음성신호의 특징 벡터를 산출하여 입력패턴을 생성하는 특징분석부와;
 상기 입력패턴과 비교하여 음성인식을 수행하기 위한 참조패턴들이 미리 저장되는 참조패턴저장부와;
 상기 입력패턴과 상기 참조패턴을 비교하여 상기 입력패턴과 유사한 참조패턴에 대응하는 초성정보를 출력하는 패턴인식부; 그리고
 상기 패턴인식부에서 검출된 초성정보에 대응하는 단어를 검색하여 출력할 단어를 결정하는 결정부를 포함하는 초성 기반 음성인식장치에 있어서,
 상기 참조패턴저장부에 저장되는 참조패턴은,
 각각의 자음에 동일한 모음이 조합된 음절을 기초로 하여 생성된 패턴이고;
 상기 패턴인식부는,
 상기 입력패턴에 대하여 우선적으로 음절단위 비교 알고리즘을 수행하여 유사한 참조패턴을 검출하되, 유사한 참조패턴이 둘 이상 검출되는 경우에 한하여 선택적으로 음소단위 비교 알고리즘을 수행함으로써 최종적으로 유사한 참조패턴 하나를 검출하고;
 상기 결정부는,
 복수의 단어가 저장된 데이터베이스를 포함하여 구성되고, 상기 데이터베이스에 저장된 단어 중 상기 패턴인식부에서 검출된 초성정보에 대응하는 단어를 검색하여 출력하되, 상기 패턴인식부에서 검출된 초성정보에 대응하는 단어가 둘 이상 검색된 경우 상기 검색된 모든 단어의 중요도를 산출하는 것을 특징으로 하는 초성 기반 음성인식장치.

청구항 2

제1항에 있어서,
 상기 음절단위 비교 알고리즘은,
 동적 시간 워핑(DTW; Dynamic Time Warping) 방법이고;
 상기 음소단위 비교 알고리즘은,
 은닉 마르코프 모델(Hidden Markov model) 방법이며;
 상기 참조패턴저장부에는,
 상기 음절단위 비교 알고리즘을 위한 참조패턴과, 상기 음소단위 비교 알고리즘을 위한 참조패턴이 별도로 저장되는 것을 특징으로 하는 초성 기반 음성인식장치.

청구항 3

제2항에 있어서,
 상기 패턴인식부는,
 상기 음절단위 비교 알고리즘의 수행에 의하여 상기 입력패턴과 각각의 참조패턴 사이의 유클리드 제곱 거리 (Squared Euclidean Distance)를 산출하고, 산출된 유클리드 제곱 거리가 미리 저장된 임계거리 미만인 참조패턴이 둘 이상인 경우, 상기 음소단위 비교 알고리즘을 수행하는 것을 특징으로 하는 초성 기반 음성인식장치.

청구항 4

삭제

청구항 5

삭제

청구항 6

제1항에 있어서,

상기 결정부는,

검색된 단어의 범위에서, 검색된 각각의 단어에 포함된 단순명사의 빈도수를 검출하고, 검출된 단순명사의 빈도수에 기초하여 각 단어의 중요도를 산출하는 것을 특징으로 하는 초성 기반 음성인식장치.

청구항 7

제6항에 있어서,

상기 결정부는,

검색된 각각의 단어에 미리 설정된 가중치가 부여된 단순명사가 포함되어 있는지 여부를 검출하여, 상기 빈도수와 상기 가중치에 기초하여 각 단어의 중요도를 산출하는 것을 특징으로 하는 초성 기반 음성인식장치.

청구항 8

제1항에 있어서,

상기 결정부는,

산출된 중요도가 높은 단어부터 출력하거나, 중요도가 가장 높은 단어만을 출력하는 것을 특징으로 하는 초성 기반 음성인식장치.

청구항 9

(A)음성을 입력받아 음성신호로 변환하는 단계와;

(B)상기 음성신호의 특징 벡터를 산출하여 입력 패턴을 생성하는 단계와;

(C)음절단위 비교 알고리즘에 따라 입력패턴을 미리 저장된 참조패턴들과 비교하여 유사한 참조패턴을 검출하는 단계와;

(D)상기 (C)단계에서 유사한 참조패턴이 둘 이상 검출된 경우에 한해 선택적으로 음소단위 비교 알고리즘에 따라 하나의 유사한 참조패턴을 검출하는 단계와;

(E)상기 (C)단계 또는 상기 (D)단계에서 검출된 하나의 유사한 참조패턴에 대응하는 초성을 갖는 단어를 검색하여 출력하는 단계를 포함하며,

상기 (E)단계는,

(E1)검출된 하나의 유사한 참조패턴에 대응하는 초성을 갖는 단어를 모두 검색하는 단계와;

(E2)상기 (E1)단계에서 검색된 단어의 범위에서, 검색된 각각의 단어에 포함된 단순명사의 빈도수를 검출하는 단계와;

(E3)상기 (E2)단계에서 검출된 단순명사의 빈도수에 기초하여 각 단어의 중요도를 산출하는 단계; 그리고

(E4)상기 (E2)단계에서 산출된 중요도에 따라 검색된 단어를 출력하는 단계를 포함하여 수행됨을 특징으로 하는 초성 기반 음성인식방법.

청구항 10

제9항에 있어서,

상기 참조패턴은,

각각의 자음에 동일한 모음을 조합한 음절을 기초로 하여, 상기 음절단위 비교 알고리즘을 위한 참조패턴과, 상

기 음소단위 비교 알고리즘을 위한 참조패턴에 대하여 별도로 미리 생성 및 저장됨을 특징으로 하는 초성 기반 음성인식방법.

청구항 11

제10항에 있어서,
 상기 음절단위 비교 알고리즘은,
 동적 시간 워핑(DTW; Dynamic Time Warping) 방법이고;
 상기 음소단위 비교 알고리즘은,
 은닉 마르코프 모델(Hidden Markov model) 방법인 것을 특징으로 하는 초성 기반 음성인식방법.

청구항 12

삭제

청구항 13

제9항에 있어서,
 상기 (E3)단계는,
 검색된 각각의 단어에 미리 설정된 가중치가 부여된 단순명사가 포함되어 있는지 여부를 검출하여, 상기 빈도수와 상기 가중치에 기초하여 각 단어의 중요도를 산출함으로써 수행됨을 특징으로 하는 초성 기반 음성인식방법.

청구항 14

제13항에 있어서,
 상기 가중치가 부여된 단순명사는,
 공공기관을 나타내는 단순명사임을 특징으로 하는 초성 기반 음성인식방법.

명세서

발명의 상세한 설명

기술분야

[0001] 본 발명은 초성 기반 음성인식장치 및 음성인식방법에 관한 것으로서, 보다 구체적으로는, 음성인식을 통한 초성 검색에서 자음에 동일한 모음을 조합한 음절을 이용하여 음성인식이 수행되도록 함으로써 임베디드 시스템의 자원을 절약함과 동시에, 초성 검색시 중요도 높은 단어가 우선적으로 출력되도록 함으로써 검색 기능을 향상시킬 수 있는 음성인식장치 및 음성인식방법을 제공하는 것이다.

배경기술

[0002] 운전자의 편의와 주행의 안정성을 위해 음성인식기술에 의한 차량 내 기기의 필요성이 대두되고 있다. 특히 차량용 내비게이션 장치(Navigator)의 음성인식은 실용화 단계에 있다.

[0003] 그러나 상용화된 내비게이션 시스템의 음성인식 성능은 만족할 만한 수준에 이르고 있지 못하다. 이는 차량 내의 잡음과 임베디드 시스템이라는 열악한 조건에 의한 것이다. 잡음이 많은 환경에서의 음성인식 시스템은 잡음 처리, 음성신호 특징 모델링, 음성인식 알고리즘 등을 모두 구현해야 하는 반면, 내비게이션 장치 내의 음성인식 시스템과 같은 임베디드 시스템은 작은 용량의 데이터베이스 내에서 구현될 것이 요구된다.

[0004] 잡음 처리 기술에는 음성향상과 음성강화 기법이 있으며, 그 종류로는 통계모델을 기반으로 한 Spectral Subtraction(SS) 기술과, Minimum Mean Square Error(MMSE) 추정을 이용한 Wiener 필터 기술이 있다. 또한 인간 청각 특성을 고려한 음성향상 등의 방법으로는 Signal subspace 음성 향상기법과 마스킹효과를 이용한 기법이 있다. 통계 모델 기법 중 SS는 간단하지만 Musical noise에 약한 단점이 있으며, Wiener 필터를 이용한 기술에서 역시 음성이 왜곡(Speech distortion)되며 잔여잡음(Residual noise)이 발생한다. 인간 청각 특성을 이용

한 기법 중 Signal subspace 음성향상 기법과 마스킹효과를 이용한 기법은 성능 자체는 좋지만 계산량이 많은 단점이 있다. 이와 같은 음성강화 기법은 내 주위의 잡음을 제어하는 기법으로서 기본적으로 전력소모가 많기 때문에 임베디드 시스템에의 적용이 어렵다.

발명의 내용

해결 하고자하는 과제

- [0005] 따라서 본 발명은 상기와 같은 종래의 문제점을 해결하기 위하여 안출된 것으로, 본 발명의 목적은 임베디드 시스템에 적합한 음성인식장치 및 음성인식방법을 제공하는 것이다.
- [0006] 본 발명의 다른 목적은 간단하면서도 정확성 있는 초성 기반의 음성인식장치 및 음성인식방법을 제공하는 것이다.
- [0007] 본 발명의 또 다른 목적은 초성 기반 음성인식에서 검색된 단어의 중요도에 따라 결과를 출력하여 검색 성능이 향상된 음성인식장치 및 음성인식방법을 제공하는 것이다.

과제 해결수단

- [0008] 상기한 바와 같은 목적을 달성하기 위한 본 발명의 특징에 따르면, 본 발명은 음성을 입력받아 음성신호로 변환하는 음성입력부와; 상기 음성입력부로부터 수신된 상기 음성신호의 특징 벡터를 산출하여 입력패턴을 생성하는 특징분석부와; 상기 입력패턴과 비교하여 음성인식을 수행하기 위하여 참조패턴들이 미리 저장되는 참조패턴저장부와; 상기 입력패턴과 상기 참조패턴을 비교하여 상기 입력패턴과 유사한 참조패턴에 대응하는 초성정보를 출력하는 패턴인식부; 그리고 상기 패턴인식부에서 검출된 초성정보에 대응하는 단어를 검색하여 출력할 단어를 결정하는 결정부를 포함하여 구성되는 초성 기반 음성인식장치에 있어서, 상기 참조패턴저장부에 저장되는 참조패턴은, 각각의 자음에 동일한 모음이 조합된 음절을 기초로 하여 생성된 패턴이고; 상기 패턴인식부는, 상기 입력패턴에 대하여 우선적으로 음절단위 비교 알고리즘을 수행하여 유사한 참조패턴을 검출하되, 유사한 참조패턴이 둘 이상 검출되는 경우에 한하여 선택적으로 음소단위 비교 알고리즘을 수행함으로써 최종적으로 유사한 참조패턴 하나를 검출하고, 상기 결정부는, 복수의 단어가 저장된 데이터베이스를 포함하여 구성되고, 상기 데이터베이스에 저장된 단어 중 상기 패턴인식부에서 검출된 초성정보에 대응하는 단어를 검색하여 출력하되, 상기 패턴인식부에서 검출된 초성정보에 대응하는 단어가 둘 이상 검색된 경우 상기 검색된 모든 단어의 중요도를 산출하는 것을 특징으로 한다.
- [0009] 여기서 상기 음절단위 비교 알고리즘은, 동적 시간 워핑(DTW; Dynamic Time Warping) 방법이고; 상기 음소단위 비교 알고리즘은, 은닉 마르코프 모델(Hidden Markov model) 방법이며; 상기 참조패턴저장부에는, 상기 음절단위 비교 알고리즘을 위한 참조패턴과, 상기 음소단위 비교 알고리즘을 위한 참조패턴이 별도로 저장될 수 있다.
- [0010] 또한 상기 패턴인식부는, 상기 음절단위 비교 알고리즘의 수행에 의하여 상기 입력패턴과 각각의 참조패턴 사이의 유클리드 제곱 거리(Squared Euclidean Distance)를 산출하고, 산출된 유클리드 제곱 거리가 미리 저장된 임계거리 미만인 참조패턴이 둘 이상인 경우, 상기 음소단위 비교 알고리즘을 수행할 수 있다.
- [0011] 삭제
- [0012] 삭제
- [0013] 이때 상기 결정부는, 검색된 단어의 범위에서, 검색된 각각의 단어에 포함된 단순명사의 빈도수를 검출하고, 검출된 단순명사의 빈도수에 기초하여 각 단어의 중요도를 산출할 수 있다.
- [0014] 나아가, 상기 결정부는, 검색된 각각의 단어에 미리 설정된 가중치가 부여된 단순명사가 포함되어 있는지 여부를 검출하여, 상기 빈도수와 상기 가중치에 기초하여 각 단어의 중요도를 산출할 수도 있다.
- [0015] 여기서 상기 결정부는, 산출된 중요도가 높은 단어부터 출력하거나, 중요도가 가장 높은 단어만을 출력할 수도 있다.
- [0016] 한편 본 발명은, (A)음성을 입력받아 음성신호로 변환하는 단계와; (B)상기 음성신호의 특징 벡터를 산출하여 입력 패턴을 생성하는 단계와; (C)음절단위 비교 알고리즘에 따라 입력패턴을 미리 저장된 참조패턴들과 비교하

여 유사한 참조패턴을 검출하는 단계와; (D)상기 (C)단계에서 유사한 참조패턴이 2개 이상 검출된 경우에 한하여 선택적으로 음소단위 비교 알고리즘에 따라 하나의 유사한 참조패턴을 검출하는 단계와; (E)상기 (C)단계 또는 상기 (D)단계에서 검출된 하나의 유사한 참조패턴에 대응하는 초성을 갖는 단어를 검색하여 출력하는 단계를 포함하며,

이때, 상기 (E)단계는, (E1)검출된 하나의 유사한 참조패턴에 대응하는 초성을 갖는 단어를 모두 검색하는 단계와; (E2)상기 (E1)단계에서 검색된 단어의 범위에서, 검색된 각각의 단어에 포함된 단순명사의 빈도수를 검출하는 단계와; (E3)상기 (E2)단계에서 검출된 단순명사의 빈도수에 기초하여 각 단어의 중요도를 산출하는 단계; 그리고 (E4)상기 (E2)단계에서 산출된 중요도에 따라 검색된 단어를 출력하는 단계를 포함하여 수행될 수 있다.

[0017] 이때 상기 참조패턴은, 각각의 자음에 동일한 모음을 조합한 음절을 기초로 하여, 상기 음절단위 비교 알고리즘을 위한 참조패턴과, 상기 음소단위 비교 알고리즘을 위한 참조패턴에 대하여 별도로 미리 생성 및 저장될 수 있다.

[0018] 또한 상기 음절단위 비교 알고리즘은, 동적 시간 워핑(DTW; Dynamic Time Warping) 방법이고; 상기 음소단위 비교 알고리즘은, 은닉 마르코프 모델(Hidden Markov model) 방법일 수 있다.

[0019] 삭제

[0020] 상기 (E3)단계는, 검색된 각각의 단어에 미리 설정된 가중치가 부여된 단순명사가 포함되어 있는지 여부를 검출하여, 상기 빈도수와 상기 가중치에 기초하여 각 단어의 중요도를 산출함으로써 수행될 수도 있다.

[0021] 여기서 상기 가중치가 부여된 단순명사는, 공공기관을 나타내는 단순명사가 될 수 있다.

효과

[0022] 이상에서 상세히 설명한 바와 같이 본 발명에 의한 음성인식장치 및 음성인식방법에 의하면 다음과 같은 효과를 기대할 수 있다.

[0023] 즉, 임베디드 시스템에 적합한 음성인식 기술을 채용하여 간단하면서도 정확성 있는 초성 기반의 음성인식이 가능하게 한다는 장점이 있다.

[0024] 또한, 본 발명에 의하면 초성 기반 음성인식에서 검색된 단어의 중요도에 따라 결과를 출력함으로써 보다 향상된 검색 기능을 제공할 수 있다는 이점이 있다.

발명의 실시를 위한 구체적인 내용

[0025] 이하에서는 상기한 바와 같은 본 발명에 의한 음성인식장치 및 음성인식방법의 구체적인 실시예를 첨부된 도면을 참고하여 상세하게 설명한다.

[0026] 도 1은 일반적인 음성인식 시스템의 개략적인 구성을 도시한 블록도이다. 일반적인 음성인식 시스템에는, 우선 사용자의 음성을 마이크로폰으로 입력받아 신호 형태로 변환하는 음성입력부(10)가 구비된다.

[0027] 그리고 상기 음성입력부(10)에 의해 얻어진 음성신호를 분석하여 인식에 필요한 특징 벡터를 추출하고, 이를 이용하여 추후 참조패턴과 비교될 입력패턴을 생성하는 특징분석부(11)가 구비된다.

[0028] 상기 특징분석부(11)에 의한 입력패턴은 다시 패턴인식부(12)로 입력된다. 상기 패턴인식부(12)는 후술할 참조패턴저장부(13)에 미리 저장된 참조패턴과 상기 입력패턴에 대한 비교 알고리즘을 수행한다. 그리고 상기 패턴인식부(12)는 상기 입력패턴과 상기 참조패턴의 비교 결과 가장 유사도가 높은 참조패턴에 대한 정보를 음성인식 결과로 출력한다.

[0029] 여기서 상기 참조패턴저장부(13)는 음성인식을 위하여 미리 정해진 참조패턴들이 저장되는 기억수단이다.

[0030] 한편 상기 패턴인식부(12)에서 가장 유사도가 높은 것으로 최종 결정된 참조패턴에 대한 정보를 출력하면, 이를 결정부(14)에서 수신하여 최종 결정된 유사한 참조패턴에 대응하는 문자를 검색어로 하여 단어를 검색한다. 그리고 검색된 단어를 최종 출력함으로써 음성인식 동작의 수행이 종료된다.

[0031] 이와 같은 음성인식 시스템에서 상기 참조패턴저장부(13)에 저장되어 입력패턴과 비교되는 참조패턴의 수가 많을수록 정확한 음성인식이 가능하나, 이 경우 상기 참조패턴저장부(13)의 용량이 방대해야할 뿐 아니라, 입력패

턴과 참조패턴의 비교 알고리즘 수행시 계산량이 크게 증가하기 때문에 상기 패턴인식부(12)의 성능 또한 높아야만 한다. 그러나 이미 설명한 바와 같이 임베디드 시스템에서는 시스템 자원이 제한되므로 최소한의 자원으로 정확한 음성인식 결과를 도출하는 것이 요구되며 이를 위하여 본 발명에서는 초성 기반의 음성인식을 지원한다.

- [0032] 특히 초성 'ㄱ', 'ㄴ', 'ㄷ' 등을 음성으로 입력함에 있어서, '기역', '니은', '디귤'과 같이 초성의 명칭을 사용하지 않고, '가', '나', '다' 같이 초성에 하나의 통일된 모음을 조합하여 발음하여 입력하도록 하고, 상기 참조패턴 또한 이와 같이 초성과 통일된 하나의 모음이 조합된 형태의 음성신호에 대응하여 생성하고 저장되도록 한다.
- [0033] 이에 의하면, 상기 참조패턴저장부(13)에 저장되어야할 참조패턴의 수가 크게 줄어들 뿐 아니라, 그에 따라 상기 패턴인식부(12)에서의 계산량도 현저하게 줄어들게 된다.
- [0034] 이러한 본 발명에 의한 음성인식방법을 도 2를 참조하여 보다 상세하게 설명한다. 도 2는 본 발명의 실시예에 따른 초성 기반 음성인식방법을 순차적으로 도시한 흐름도이다.
- [0035] 도 2에 도시된 바와 같이 본 발명의 실시예에 의한 음성인식방법에서는, 우선 사용자로부터 음성이 입력되는 단계(S10)로부터 시작된다. 상기 S10단계에서 사용자로부터 입력되는 음성은 어떤 것이라도 될 수 있으나, 앞서 설명한 바와 같이, 본 발명의 실시예에서는 초성 기반 음성인식이 수행되고, 아울러 초성에 공통된 하나의 모음을 조합한 발음을 기초로 생성된 참조패턴을 이용하여 음성인식이 수행되므로 '가', '나', '다' 와 같이 공통된 모음을 갖고, 종성을 포함하지 않는 음절들의 조합으로 입력될 것이다. 예를 들어, 사용자가 '홍길동'을 검색하고자 하는 경우, '하가다'와 같은 음성이 입력될 것이다.
- [0036] 이와 같이 사용자로부터 음성이 입력되면, 입력된 음성을 음성신호로 변환하고, 그로부터 특징 벡터를 추출하는 단계(S20)가 수행된다. 여기서 특징 벡터의 추출은 선형 예측 부호화(LPC; Linear Predictive Coding)를 수행함으로써 이루어질 수 있다. 그리고 상기 S20단계에서 특징 벡터가 추출되면 특징 벡터를 이용하여 입력된 음성신호에 대한 입력 패턴을 생성하게 된다. 여기서 상기 입력 패턴은 상기 음성 신호를 기저장된 참조 패턴과의 비교 알고리즘 수행이 가능하도록 처리한 데이터이다.
- [0037] 그리고 이어서 S30단계에서는 상기 입력 패턴과 상기 참조 패턴 사이의 비교 및 분석이 이루어진다. 여기서 비교 및 분석되는 입력 패턴과 참조 패턴은 모두 음절단위로 생성되어 서로 비교된다. 특히 상기 S30단계는 동적 시간 워핑(DTW; Dynamic Time Warping) 방법을 사용할 수 있다. 상기 동적 시간 워핑은 입력된 음성과 참조 음성 사이의 발음 속도와 길이의 차이를 보상하기 위하여 입력 패턴과 참조 패턴을 비선형적으로 정합하여 가장 유사도가 높은 참조 패턴의 음성으로 입력된 음성을 인식하는 방법이다. 유사도를 판단하는 기준은 입력 패턴과 각각의 참조 패턴 사이의 유클리드 제곱 거리(Squared Euclidean Distance)를 산출하여 그 거리가 가장 적은 참조 패턴을 입력 패턴과 가장 유사한 패턴으로 인식하는 것이다.
- [0038] 이와 같이 상기 S30단계에서 동적 시간 워핑을 수행하여 입력 패턴과 각각의 참조 패턴 사이의 유사도를 산출함에 있어서, 입력 패턴과의 유사도가 미리 정한 일정 범위 내인 참조 패턴이 2 이상인지 여부를 판단하는 단계(S40)가 함께 수행된다. 즉, 상기 동적 시간 워핑 수행 중, 입력 패턴과 각각의 참조 패턴 사이의 유클리드 제곱 거리를 산출한 결과, 미리 설정된 임계값보다 작은 유클리드 제곱 거리를 갖는 참조 패턴이 둘 이상인지 여부를 판단한다.
- [0039] 현재 입력된 음성이 둘 이상의 유사한 음성으로 인식될 가능성이 있는 경우에 해당하므로 보다 정확한 패턴 분석이 요구된다. 예를 들어, '가'와 '카', '다'와 '타' 등은 발음의 유사성으로 인해 신호 패턴도 어느 정도 유사하므로 이를 동적 시간 워핑 방식만으로 비교하는 경우, 사용자가 의도한 바와 다르게 인식될 가능성이 있다.
- [0040] 따라서 이와 같이 유사한 참조 패턴이 2 이상인지 여부를 판별하여, 2 이상인 경우에는 동적 시간 워핑 방식보다 더 인식률이 좋은 방식의 패턴 분석을 다시 수행한다(S50). 즉, 음성 신호를 음소단위로 분리한 후 은닉 마르코프 모델(Hidden Markov model)과 같은 방식에 의하여 음소단위의 패턴 비교 알고리즘을 수행한다.
- [0041] 여기서 상기 은닉 마르코프 모델은, 모델링하는 시스템이 미지의 파라미터를 가진 Markov process일 것이라고 가정하여, 그 가정에 기초해서 관측된 파라미터로부터 숨겨진 파라미터를 결정하는 하나의 통계모델로서, 음성 인식분야에서 널리 사용되는 방식 중 하나이다.
- [0042] S60단계에서는 상기 S30단계와 상기 S50단계에서 수행된 패턴 분석 결과에 따라 음소를 결정한다. 즉, S50단계를 거쳐 음소단위의 패턴 비교 알고리즘을 수행한 결과 가장 유사한 것으로 판단된 음소를 입력된 음소로 결정하거나, 상기 S40단계에서, 유사한 참조 패턴이 하나만 검출된 경우에는 해당 참조 패턴에 해당하는 음성에 대

응하는 음소를 입력된 음소로 결정한다.

[0043] 예를 들어, 사용자가 음성 '가'를 입력하여 상기 S30단계에서 각각 '가'와 '카'에 해당하는 참조 패턴이 유사한 패턴으로 인식된 경우에는 다시 S50단계에서 다시 저장된 음성 신호의 음소 부분만을 따로 처리하여 은닉 마르코프 모델을 수행함으로써 사용자가 실질적으로 입력하고자 하였던 초성 'ㄱ'을 인식된 음소로 결정하게 된다. 한편 사용자가 음성 '나'를 입력하여 상기 S30단계에서 유사한 참조패턴이 '나'로 인식되었다면 S50단계를 거치지 않고 바로 'ㄴ'이 입력된 것으로 결정된다.

[0044] 순차적으로 입력된 음성 신호에 대하여 이와 같이 각각 음소가 결정되면, 해당 음소들을 이용하여 단어를 검색하고 검색된 단어 중 최종 결과를 선택하는 단계가 수행된다(S70). 여기서 S70단계의 보다 구체적인 수행에 대해서는 도 4를 참조하여 후술한다.

[0045] 그리고 상기 S70단계에서 검색된 결과를 디스플레이 장치에 표시하는 등 최종적으로 인식된 결과를 출력하는 단계가 수행된다(S80).

[0046] 이와 같은 방식에 의하면, 우선 초성 기반의 음성인식을 통해 비교되는 참조패턴의 개수를 줄일 수 있어, 메모리를 절약함과 동시에 계산량을 감소시킬 수 있다. 또한 비교적 계산량이 많지 않은 음절 단위 패턴에 대한 동적 시간 워핑을 기본적으로 사용하되, 정확성이 요구되는 경우에만 음소 단위 패턴에 대한 은닉 마르코프 모델 방식을 보조적으로 사용함으로써 시스템에 과도한 로드를 주지않으면서 음성인식의 정확성을 담보할 수 있다는 장점이 있다.

[0047] 또한 본 발명의 실시예에 의한 음성인식방법에서의 음성신호 그래프 및 음성신호 스펙트로그램을 도시하고 있는 도 3을 참조하면, (a)는 '가, 다, 타'의 남성 음성 신호를 나타낸 것이고, (b)는 '가, 다, 타'의 여성 음성 신호를 나타낸 것이며, (c)와 (d)는 '가' 음절에 대한 남성과 여성의 음성 신호를 20ms의 윈도우(Window)로 윈도우링(windowing)한 결과와, 그 스펙트로그램, 그리고 선형 예측 부호화(10차)의 스펙트로그램을 함께 나타낸 것이다.

[0048] 도 3의 (a), (b)에 도시된 '가, 다, 타' 음성 신호에 대한 음절단위 동적 시간 워핑을 수행한 결과를 보면, 각 음절에 대한 선형 예측 부호화를 통한 특징 벡터 추출에서 선형 예측 부호화 계수는 10차까지 산출하였으며, 각 음절간, 남성과 여성의 동적 시간 워핑 거리는 다음의 표와 같이 나타났다.

표 1

		남			여		
		가	다	타	가	다	타
남	가	0	-	-	-	-	-
	다	0.56	0	-	-	-	-
	타	0.58	0.55	0	-	-	-
여	가	0.9	1.65	2.23	0	-	-
	다	1.91	0.35	0.52	1.89	0	-
	타	1.79	0.39	0.83	1.84	0.33	0

[0049] [0050] 표에서 보는 것처럼 '가' 와 '다' 같은 경우에는 인식률이 높았으나, '다'와 '타'와 같은 유사 음절에 대한 동적 시간 워핑 방식에 의한 비교 알고리즘 수행은 인식률이 높지 않았다. 따라서 은닉 마르코프 모델 등의 방식

에 의한 음소단위의 패턴 비교 알고리즘을 보조적으로 추가 수행하여 인식 정확성을 높일 수 있다.

- [0051] 한편 이하에서는 도 4를 참조하여 도 2에 도시된 단어검색 및 최종단어 선택 단계를 보다 자세하게 살펴본다. 도 4는 본 발명의 실시예에 따른 초성 기반 음성인식방법에서의 단어검색 및 최종단어선택단계를 보다 구체적으로 도시한 흐름도이다.
- [0052] 도 2에 순차적으로 도시된 S10단계 내지 S60단계의 수행에 의하여, 사용자가 입력한 음성에 대한 음소가 인식되면, 도 1의 결정부(14)에 해당하는 음성인식 시스템의 제어수단에서는 시스템 내의 데이터베이스에서, 결정된 음소에 대응하는 모든 단어를 검색한다(S71). 여기서 상기 데이터베이스는 예를 들어, 내비게이션 장치에서는 목적지 명칭에 대한 데이터베이스 등이 될 수 있다.
- [0053] 그리고나서, 상기 S71단계에서 검색된 단어가 둘 이상인지 여부를 결정한다(S72). 검색된 단어가 하나뿐이면, 검색된 단어를 최종 단어로 선택하여 출력하면 되지만, 검색된 단어가 둘 이상인 경우에는, 다시 검색된 단어 중 어떤 단어를 최종 단어로 선택하여 출력할지 여부를 결정하기 위하여 이를 구분하는 것이다.
- [0054] S72단계에서 검색된 단어가 둘 이상인 것으로 판별되면, 다시 검색된 단어들을 단순명사 단위로 구분한다(S73). 내비게이션 장치의 목적지 명칭 등은 특히 '서울관광호텔'과 같이 복합명사로 구성되는 경우가 매우 많다. 예로 든 경우, '서울', '관광', '호텔'과 같은 단순명사가 한 단어 내에 여러 개 구성되는데 상기 S73단계에서는 검색된 모든 단어들을 단순명사 단위로 구분한다.
- [0055] 사용자가 '사, 아, 아, 사, 하, 가'를 입력하여 음성인식장치에서 '사, 오, 오, 사, 흥, 가'으로 인식되고, 이에 대해 '서울예술학교', '상암예술회관', '서울여성학교' 총 3개의 단어가 검색된 경우를 예로 들면, 이들 각각을 단순명사 단위로 '서울/예술/학교', '상암/예술/회관', '서울/여성/학교'로 구분 가능하다.
- [0056] 이와 같이 검색된 단어 내에 포함된 단순명사를 구분한 후에는, 각 단순명사들의 검색 범위 내에서의 빈도수를 검출한다(S74). 즉, 동일한 예에서, 검색된 총 단어의 수는 3개이고, 이중 '서울'이 포함된 단어는 2개, '예술'이 포함된 단어는 2개, '학교'가 포함된 단어는 2개이며, '상암', '회관', '여성'은 각각 1개씩 포함되었다.
- [0057] 그 후, 선택적으로 각 단순명사에 대한 가중치를 부여하는 단계가 수행된다(S75). 각 단어 중 중요도나 검색 빈도는 높으나, 실제로 데이터베이스 내에서의 빈도수는 적은 단어, 예를 들어 병원, 소방서, 학교, 구청, 시청 등과 같은 관공서 내지 공공기관 등은 지역 내에 한정된 수로 존재하므로 그 검색 빈도에 비하여 데이터베이스에 포함된 목적지 명칭 내에 포함된 수는 적으므로 별도로 가중치를 부여할 수 있다.
- [0058] 예를 들어, 관공서를 나타내는 단순명사에 대한 가중치는 아래의 식에 의하여 산출되도록 할 수 있다.

수학식 1

$$G_{gov} = \frac{N_{max}}{N_{gov}}$$

- [0059]
- [0060] 여기서 상기 N_{gov} 는 공공기관을 나타내는 단순명사를 포함하는 단어의 전체 수이고, N_{max} 는 공공기관을 나타내는 단순명사 중 해당 단순명사를 포함하는 단어수가 가장 많은 단순명사를 포함하는 단어의 수이다.
- [0061] 그리고 상기 S74단계에서 검출된 빈도수와 S75단계에서 부여된 가중치에 기초하여 검색된 단어들의 중요도 지수를 산출하는 단계(S76)가 수행된다. 중요도 지수를 산출하기 위한 수학식은 아래와 같이 정해질 수 있다.

수학식 2a

[0062]

$$C_{(n)} = \sum \left(\frac{N_{w(n)}}{N_t} \right) \quad \text{또는} \quad C_{(n)} = \sum \left(\frac{N_{w(n)}}{N_t} + G_{gov} \right)$$

[0063]

여기서 $C_{(n)}$ 는 검색된 각 단어의 중요도 지수이고, $N_{w(n)}$ 는 검색된 모든 단어 중 특정 단순명사가 포함된 단어의 수이며, N_t 는 검색된 모든 단어의 수이다. 또한 G_{gov} 는 위에서 설명한 가중치이다.

[0064]

왼쪽에 도시된 수학적식은 가중치를 고려하지 않은 경우, 즉 S75단계를 수행하지 않은 경우의 중요도 지수 식이고, 오른쪽에 도시된 수학적식은 가중치를 고려한 경우이다.

[0065]

보다 쉬운 이해를 위하여 예를 들어 설명하면, 이미 설명한 바와 같은 예에서 '사, 아, 아, 사, 하, 가'를 입력한 경우 검색된 모든 단어 '서울예술학교', '서울여성학교', '상암예술회관' 각각에 대한 중요도 지수를 계산하면, 우선 '서울예술학교'는 세 개의 단순명사를 포함하고 있다. 따라서 N_t 는 3이다. 그리고 '서울'이라는 단어의 빈도수는 2이므로, '서울'이라는 단어에 대한 $N_{w(n)}$ 는 2이다. 이와 같은 식으로, '서울예술학교'에 포

함된 각 단어, '서울', '예술', '학교'에 대한 $\frac{N_{w(n)}}{N_t}$ 값을 각각 산출하면, 2/3, 2/3, 2/3이고, 따라서 총합

$$C_{(n)} = \sum \left(\frac{N_{w(n)}}{N_t} \right) \quad \text{은 2가 된다.}$$

[0066]

한편 '서울여성학교'에 대해서도 동일한 방식으로 중요도 지수를 산출하면 2/3+1/3+2/3으로서 5/3이 된다. 그리고 '상암예술회관'은 1/3+2/3+1/3으로 4/3이 된다.

[0067]

따라서 위 세 개의 단어 중 중요도 지수가 가장 높은 단어는 '서울예술학교'가 된다.

[0068]

가중치를 고려하는 경우에도, 각 단어에 대한 가중치를 함께 더하여 전체값을 구하면 된다.

[0069]

그러나, 상기 S76단계에서 중요도 지수를 산출하기 위한 식은 반드시 위에 기재된 수학적식 2a와 같이 정의되는 것은 아니다. 다른 예로 아래의 수학적식을 예시한다.

수학적식 2b

[0070]

$$C_{(n)} = \frac{\sum \left(\frac{N_{w(n)}}{N_t} \right)}{N_n} \quad C_{(n)} = \frac{\sum \left(\frac{N_{w(n)}}{N_t} + G_{gov} \right)}{N_n}$$

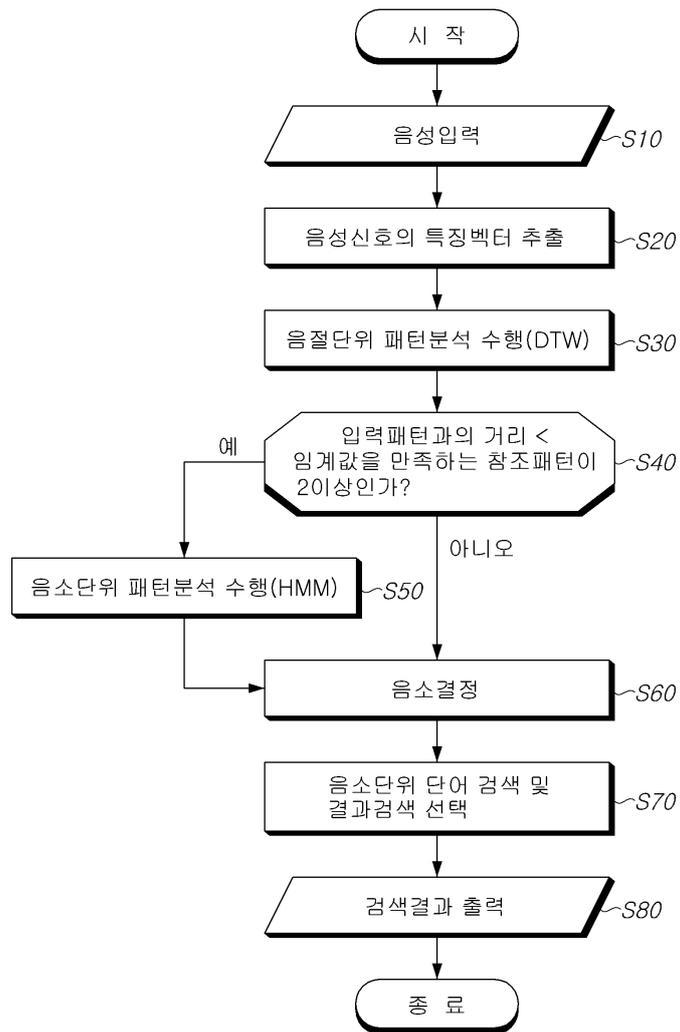
[0071]

여기서 N_n 은 검색된 단어에 포함된 단순명사의 수이다.

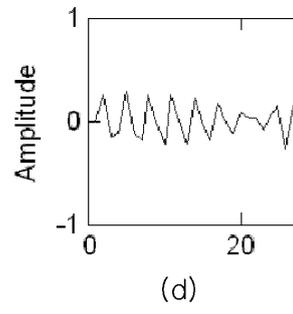
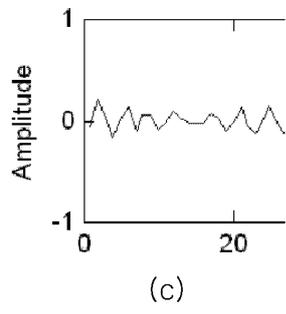
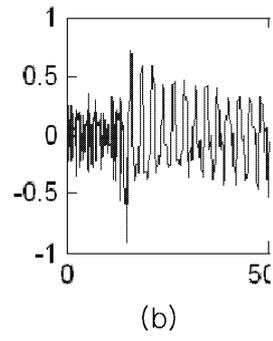
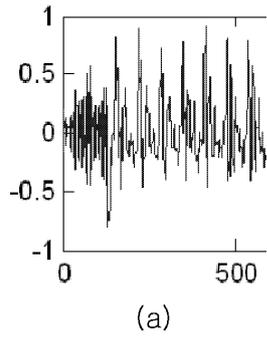
[0072]

예를 들어, 초성 'ㄱ, ㅋ, ㆁ, ㆁ'가 입력되고, 이에 대하여 검색된 단어가 '문경새재'와 '면곡시장' 두 개인 경우, 수학적식 2a에 의한다면, '문경새재'는 포함된 단순명사가 하나로서, 중요도 지수를 산출하면 1/2인 반면, '면곡시장'은 두 개의 검색된 전체 단어 중 '면곡'이 들어간 단어가 1개, '시장'이 들어간 단어 1개로 중요도 지수가 1이 되므로, 빈도수와 무관하게 단어에 포함된 단순명사의 수가 많을수록 중요도 지수가 높게 나타날 수 있다. 따라서 산출된 빈도수의 합을 해당 단어에 포함된 단순명사의 수로 나눠서 중요도 지수를 산출할 수도 있다.

도면2



도면3



도면4

