(19)대한민국특허청(KR)
(12) 등록특허공보(B1)

(51) 。Int. Cl.
G10L 15/22 (2006.01)
G10L 15/00 (2006.01)

(45) 공고일자 2007년07월25일
(11) 등록번호 10-0742888
(24) 등록일자 2007년07월19일

(21) 출원번호 10-2006-0021863
(22) 출원일자 2006년03월08일
심사청구일자 2006년03월08일

(65) 공개번호 10-2006-0097647
(43) 공개일자 2006년09월14일

(30) 우선권주장 JP-P-2005-00065355 2005년03월09일 일본(JP)

(73) 특허권자 캐논 가부시끼가이샤
일본 도꾜도 오오따꾸 시모마루꼬 3쪼메 30방 2고

(72) 발명자 후까다 도시아끼
일본 도꾜도 오오따꾸 시모마루꼬 3-30-2 캐논 가부시끼가이샤 내

(74) 대리인 구영창
이중희
장수길

(56) 선행기술조사문헌
KR1020020033791A
KR100482477 B1
KR1020040051317 A
US 5191635
KR1020050015586A
KR100569612 B1
EP 1083545

심사관 : 경연정

전체 청구항 수 : 총 13 항

(54) 음성 인식 방법

(57) 요약

본 발명의 음성 인식 방법은 이용자가 만들어낸 음성의 취입을 포함하고 있다. 이용자의 조작 또는 동작에 따라서 이러한 취입이 개시된다. 이어서, 취입된 음성의 시작부분이 존재하는지 누락하고 있는지의 여부가 판정된다. 음성 판정 유닛의 결과에 기초하여 인식 대상 단어의 발음 정보가 설정되며, 그 설정된 발음 정보를 이용하여 취입된 음성이 인식된다.

대표도

도 5

특허청구의 범위

청구항 1.

음성 인식 방법으로서,

이용자 입력에 따라서 이용자가 만들어낸 음성의 취입을 개시하는 단계와;

취입된 음성의 시작부분이 누락되고 있는지의 여부를 판정하는 단계와;

상기 판정 단계의 결과에 기초하여 인식 대상 단어(target word to be recognized)의 발음 정보를 설정하는 단계와;

상기 설정된 발음 정보를 이용하여 취입된 음성을 인식하는 단계

를 포함하는 음성 인식 방법.

청구항 2.

음성 인식 방법으로서,

이용자 입력에 따라서 이용자가 만들어낸 음성의 취입을 개시하는 단계와;

상기 음성의 취입이 상기 이용자가 만들어낸 음성 도중에 개시되는지의 여부를 판정하는 단계와;

상기 판정 단계의 결과에 기초하여 인식 대상 단어의 발음 정보를 설정하는 단계와;

상기 설정된 발음 정보를 이용하여 상기 취입된 음성을 인식하는 단계

를 포함하는 음성 인식 방법.

청구항 3.

제1항에 있어서,

상기 판정 단계에서의 음성의 시작부분이 누락되고 있는지의 여부의 판정은, 음성 파형의 파워, 영교차율(zero-crossing rate), 스펙트럼 정보 및 고조파 구조를 나타내는 특징 파라미터 중 적어도 하나에 대한 정보를 이용하여 행해지는 음성 인식 방법.

청구항 4.

제1항에 있어서,

상기 발음 정보는 인식 대상 단어의 판독에 관련된 발음 계열(pronunciation sequence), 발음 계열에 관련된 상세 발음 계열, 및 인식 대상 단어에 관련된 참조 패턴 계열 중 적어도 하나인 음성 인식 방법.

청구항 5.

제4항에 있어서,

상기 상세 발음 계열은 상태 천이 모델로 모델화된 발음 계열의 상태 계열인 음성 인식 방법.

청구항 6.

제4항에 있어서,

상기 참조 패턴 계열은 등록 요구형의 음성 인식 장치에서의 등록 음성의 특징 파라미터 계열인 음성 인식 방법.

청구항 7.

제1항에 있어서,

상기 음성을 취입하는 단계는, 이용자의 조작에 따라서 음성의 취입을 개시하는 음성 인식 방법.

청구항 8.

제1항에 있어서,

상기 음성을 취입하는 단계는, 센서에 의해 검지된 이용자의 동작에 따라서 음성의 취입을 개시하는 음성 인식 방법.

청구항 9.

컴퓨터로 하여금 제1항 또는 제2항의 음성 인식 방법을 실행하도록 하는 컴퓨터 실행가능 명령을 저장한 컴퓨터 판독가능 매체.

청구항 10.

음성 인식 장치로서,

이용자 입력에 따라서 이용자가 만들어낸 음성의 취입을 개시하도록 구성된 음성 취입 유닛과;

상기 취입된 음성의 시작부분이 누락되고 있는지의 여부를 판정하도록 구성된 판정 유닛과;

상기 판정 유닛의 결과에 기초하여 인식 대상 단어에 대한 발음 정보를 설정하도록 구성된 설정 유닛과;

상기 설정된 발음 정보를 이용하여 상기 취입된 음성을 인식하도록 구성된 음성 인식 유닛

을 포함하는 음성 인식 장치.

청구항 11.

음성 인식 장치로서,

이용자 입력에 따라서 이용자가 만들어낸 음성의 취입을 개시하도록 구성된 음성 취입 유닛과;

상기 음성의 취입이 상기 이용자의 음성 도중에 개시되는지의 여부를 판정하는 판정 유닛과;

상기 판정 유닛의 결과에 기초하여 인식 대상 단어의 발음 정보를 설정하도록 구성된 설정 유닛과;

상기 설정된 발음 정보를 이용하여 상기 취입된 음성을 인식하도록 구성된 음성 인식 유닛

을 포함하는 음성 인식 장치.

청구항 12.

제10항에 있어서,

상기 판정 유닛에서 상기 음성의 시작부분이 누락되고 있는지의 여부의 판정은, 음성 파형의 파워, 영교차율, 스펙트럼 정보 또는 고조파 구조를 나타내는 특징 파라미터 중 적어도 하나에 대한 정보를 이용하여 행해지는 음성 인식 장치.

청구항 13.

제10항에 있어서,

상기 설정 유닛 내의 발음 정보는, 인식 대상 단어의 판독에 관련된 발음 계열, 발음 계열에 관련된 상세 발음 계열, 및 인식 대상 단어에 관련된 참조 패턴 계열 중 적어도 하나인 음성 인식 장치.

명세서

발명의 상세한 설명

발명의 목적

발명이 속하는 기술 및 그 분야의 종래기술

본 발명은, 버튼 누름과 같은 음성 개시 커맨드의 입력을 수반한 음성 인식을 수행하여, 버튼을 누르기 전에 음성이 만들어질 수 있는 고정밀도의 음성 인식을 실현하기 위한 방법에 관한 것이다.

음성 인식을 수행할 때, 주위 잡음에 기인한 에러를 방지하기 위해, 이용자의 입과 마이크로폰 간의 거리 및 입력 레벨을 적절하게 설정하고, 음성 개시 커맨드를 (통상 버튼을 누름으로써) 적절하게 입력할 필요가 있다. 이들이 적절하게 행해지지 않으면, 인식 성능의 상당한 저하를 초래할 것이다. 그러나, 이용자가 이러한 설정 또는 입력을 항상 적절하게 유지할 수는 없기 때문에, 이러한 경우의 성능 저하를 방지하기 위한 대책이 필요하게 된다. 특히, 예를 들어, 버튼을 누르기 전에 음성이 만들어지는 등과 같이, 간혹 음성 개시 커맨드가 정확하게 입력되지 않는 경우가 있다. 그러한 경우에는, 음성 개시 커맨드가 입력된 후에 마이크로폰을 통해 음성을 취득하기 때문에, 음성의 시작부분이 생략될 것이다. 그 생략된 음성에 기초하여 종래의 음성 인식을 수행할 경우, 음성 개시 커맨드가 정확하게 입력되는 경우와 비교하여, 인식율이 크게 떨어질 것이다.

발명이 이루고자 하는 기술적 과제

이러한 문제를 고려하여, 일본특허공보 제2829014호에서는, 인식 처리 개시 커맨드가 입력된 후에 취입된 음성 데이터를 저장하는 데이터 버퍼외에도, 항상 일정 길이의 음성을 취입하는 링 버퍼를 제공하는 방법을 개시하고 있다. 그리고, 커맨드가 입력된 후, 데이터 버퍼에 의해 취입된 음성을 이용하여 음성의 헤드를 검출한다. 음성의 헤드가 검출되지 않을 경우, 링 버퍼에 저장되어 있는 커맨드 입력전의 음성을 추가로 사용함으로써 음성 헤드의 검출을 수행한다. 이 방법에서, 링 버퍼가 음성의 취입 처리를 지속적으로 수행해야 하기 때문에, 데이터 버퍼만을 이용하는 경우와 비교하면, 추가적인 CPU 부하를 요구한다. 즉, 이동 장치와 같은 배터리 구동 장치에 사용하기에 반드시 적절한 방법은 아니다.

또한, 일본특허공보 제3588929호에서는 단어의 시작부분에 반음절 또는 단음절 생략되어 있는 단어를 인식 대상으로 하는 방법을 개시하고 있다. 이 방법으로, 잡음 환경에서의 음성 인식률의 저하를 방지할 수 있다. 또한, 일본특허공보 제3588929호에서는 헤드 부분이 생략된 단어를 잡음 레벨에 따라서 인식 대상의 단어로 해야 하는지를 판정하는 제어를 수

행하는 방법을 개시하고 있다. 이 방법으로, 단어 시작부분에 반음절 또는 단음절의 종류 또는 잡음 레벨에 따라, 단어 시작부분에 반음절 또는 단음절 생략되어 있는지의 여부에 관한 판정이 수행된다. 생략된 것으로 판정된 경우, 생략되지 않은 단어는 인식 대상의 단어로 지정되지 않는다. 부가적으로, 단어의 시작부분이 생략되고 있는지의 여부가 판정되면, 이용자의 조작 또는 동작에 의해 입력된 음성 시작 커맨드를 정확하게 수행하고 있는지의 여부는 고려하지 않는다. 따라서, 일본특허공보 제3588929호에서는, 단어의 시작부분에서의 생략은 1음절까지이며, 조용한 환경에서는 단어의 시작부분을 생략하지 않는다. 결과적으로, 버튼을 누르기 전에 음성이 만들어진 경우, 예를 들어, 조용한 분위기에서 2음절 정도의 음성이 생략된 경우에는, 인식 성능의 저하가 방지될 수 없다.

상기 문제에 관하여, 본 발명의 목적은 음성의 시작부분이 누락 또는 생략된 경우에, 간단하고 쉬운 처리로 인식 성능의 저하를 방지하는 방법을 제공하는 것이다. 그러한 생략은 음성 개시 커맨드가 이용자에 의해 부적절하게 입력된 경우에 발생한다.

발명의 구성

본 발명의 일양상은, 이용자 입력에 따라서 이용자가 만들어낸 음성의 취입을 개시하는 단계와, 취입된 음성의 시작부분이 누락되고 있는지의 여부를 판정하는 단계와, 상기 판정 단계의 결과에 기초하여 인식 대상 단어의 발음 정보를 설정하는 단계와, 상기 설정된 발음 정보를 이용하여 취입된 음성을 인식하는 단계를 포함하는 음성 인식 방법이다.

본 발명의 또 다른 양상은, 이용자 입력에 따라서 이용자가 만들어낸 음성의 취입을 개시하는 단계와, 상기 음성의 취입이 상기 이용자가 만들어낸 음성 도중에 개시되는지의 여부를 판정하는 단계와, 상기 판정 단계의 결과에 기초하여 인식 대상 단어의 발음 정보를 설정하는 단계와, 상기 설정된 발음 정보를 이용하여 상기 취입된 음성을 인식하는 단계를 포함하는 음성 인식 방법이다.

본 발명의 또 다른 양상은, 이용자 입력에 따라서 이용자가 만들어낸 음성의 취입을 개시하도록 구성된 음성 취입 유닛과, 상기 취입된 음성의 시작부분이 누락되고 있는지의 여부를 판정하도록 구성된 판정 유닛과, 상기 판정 유닛의 결과에 기초하여 인식 대상 단어에 대한 발음 정보를 설정하도록 구성된 설정 유닛과, 상기 설정된 발음 정보를 이용하여 상기 취입된 음성을 인식하도록 구성된 음성 인식 유닛을 포함하는 음성 인식 장치이다.

본 발명의 또 다른 양상은, 이용자 입력에 따라서 이용자가 만들어낸 음성의 취입을 개시하도록 구성된 음성 취입 유닛과, 상기 음성의 취입이 상기 이용자의 음성 도중에 개시되는지의 여부를 판정하는 판정 유닛과, 상기 판정 유닛의 결과에 기초하여 인식 대상 단어의 발음 정보를 설정하도록 구성된 설정 유닛과, 상기 설정된 발음 정보를 이용하여 상기 취입된 음성을 인식하도록 구성된 음성 인식 유닛을 포함하는 음성 인식 장치이다.

본 발명의 다른 특징들은 첨부 도면을 참조하여 예시적인 실시예에 대한 다음의 상세한 설명으로부터 명백해질 것이다.

명세서에 통합되어 일부를 구성하고 있는 첨부 도면은 본 발명의 예시적인 실시예를 설명하며, 상세한 설명과 함께 본 발명의 원리들을 설명해 줄 것이다.

<실시예>

이하 도면을 참조하여 본 발명의 예시적인 실시예를 상세하게 설명한다.

(제1 실시예)

도 1은 본 발명의 제1 실시예에 따른 음성 인식 장치의 블록도이다. CPU(101)는 ROM(102)에 저장되거나 외부 저장 장치(104)로부터 RAM(103)에 로딩된 제어 프로그램에 따른 음성 인식 장치의 블록도이다. ROM(102)은 각종 파라미터 및 CPU(101)에 의해 실행되는 제어 프로그램을 저장하고 있다. RAM(103)은 각종 제어 기능을 수행할 때의 작업 영역을 제공하며, CPU(101)에 의해 실행되는 제어 프로그램을 저장하고 있다. 도 5의 흐름도에 도시된 방법은, 바람직하게는 CPU(101)에 의해 실행되는 프로그램이며, 이는 ROM(102), RAM(103) 또는 저장 장치(104)에 저장된다.

참조번호 104는 하드디스크, 플로피(등록상표) 디스크, CD-ROM, DVD-ROM, 메모리 카드와 같은 외부 저장 장치를 나타낸다. 외부 저장 장치(104)가 하드디스크인 경우에는, CD-ROM 또는 플로피(등록상표) 디스크 등으로부터 설치된 각종 프로그램을 저장한다. 마이크로폰과 같은 음성 입력 장치(105)는 음성 인식을 수행할 음성을 취입한다. CRT 또는 LCD와 같은 디스플레이 장치(106)는 처리 내용의 설정을 수행하고, 입력 정보를 디스플레이하고, 처리 결과를 출력한다. 버튼, 텐

키, 키보드, 마우스 또는 펜과 같은 보조 입력 장치(107)는 이용자가 만들어낸 음성을 취입하기 시작하라는 명령을 제공하는데 사용된다. 스피커와 같은 보조 출력 장치(108)는 음성 인식 결과를 소리(voice)로 확인하는데 이용된다. 버스(109)는 상기 모든 장치들을 접속한다. 인식 대상 음성은 음성 입력 장치(105)를 통해 입력될 수도 있으며, 다른 장치 또는 유닛에 의해 획득될 수도 있다. 다른 장치 또는 유닛에 의해 획득된 대상 음성은 ROM(102), RAM(103), 외부 저장 장치(104) 또는 네트워크를 통해 접속된 외부 장치에 보유된다.

도 2는 음성 인식 방법의 모듈 구성의 블록도이다. 음성 취입 유닛(201)은 음성 입력 장치(105)인 마이크로폰을 통해 입력된 음성을 취입한다. 음성 취입을 개시하라는 명령은 보조 입력 장치(107)의 버튼을 누르는 것과 같은 이용자 조작에 의해 제공된다. 취입된 음성 판정 유닛(202)은 음성 취입 유닛에 의해 취입된 음성의 시작 또는 시작부분이 누락 또는 생략되고 있는지를 판정한다. 발음 정보 설정 유닛(203)은 취입된 음성 판정 유닛(202)의 결과에 기초하여, 대상 단어의 발음 정보를 설정한다. 음성 인식 유닛(204)은 발음 정보 설정 유닛(203)에 의해 설정된 발음 정보를 이용하여 음성 취입 유닛(201)에 의해 취입된 음성을 인식한다.

도 3은 비등록 음성 또는 화자 독립형 음성(speaker-independent speech)을 인식할 때에 이용되는 일반적인 음성 인식 방법의 모듈의 블록도이다. 음성 입력 유닛(301)은 음성 입력 장치(105)를 통해 입력된 음성을 인식한다. 음성 특징 파라미터 추출 유닛(302)은 음성 입력 유닛(301)에 의해 입력된 음성에 스펙트럼 분석을 수행하여, 특징 파라미터를 추출한다. 발음 사전(305)은 인식 대상 단어의 발음 정보를 보유하고 있다. 음향 모델(306)은 음소 모델(또는 음절 모델 또는 단어 모델)을 보유하고 있으며, 인식 대상 단어의 참조 패턴은 발음 사전(305)의 발음 정보에 따라서 음향 모델을 이용하여 구성된다. 언어 모델(307)은 단어 리스트 및 단어 접속 확률(또는 문법 제약)을 보유한다. 탐색 유닛(303)은 발음 사전(305)으로부터 언어 모델(307)을 이용하여 구성되는 참조 패턴과, 음성 특징 파라미터에 의해 추출 유닛(302)에 의해 얻어지는 음성의 특징 파라미터 간의 거리를 계산한다. 탐색 유닛(303)은 우도(likelihood)를 계산하거나, 탐색 처리를 수행한다. 결과 출력 유닛(304)은 탐색 유닛(303)에 의해 얻어진 결과를 디스플레이 장치(106)에 디스플레이하거나, 그 결과를 보조 출력 장치(108)에 음성으로 출력하거나, 소정의 조작을 수행하기 위해 인식 결과를 출력한다. 발음 정보 설정 유닛(203)에 의한 발음 정보의 설정은 발음 사전(305)의 설정에 대응한다.

도 5는 음성 인식 방법의 전체 처리의 흐름도이다. 흐름도를 이용하여 전체 처리를 상세하게 설명한다. 단계 S501에서, 음성 개시 커맨드 입력을 대기한다. 이 커맨드는 이용자의 조작 또는 동작에 따라서 입력된다. 커맨드 입력은 사용자로 하여금 예를 들어, 텐키, 키보드 또는 스위치 등의 버튼을 누르거나, 마우스를 클릭하거나, 터치 패널을 누름으로써 음성 개시 명령을 제공하도록 하는 임의의 수단을 이용할 수 있다. 부가적으로, 적외선 센서를 포함하는 광 센서, 촉각 센서, 초음파 센서 등의 센서를 이용하면, 음성 인식 장치에 근접하고 있는 이용자의 동작을 검지할 수 있다. 이러한 이용자 동작이 음성 개시 커맨드로서 간주되면, 센서에 의한 검출을 음성 개시 커맨드로서 이용할 수 있다. 단계 S501에서의 커맨드는 단계 S502의 마이크로폰을 통해 음성 취입을 트리거한다. 단계 S504에서, 취입된 음성의 시작부분이 생략되고 있는지의 여부를 판정하여, 이 판정에 필요한 음성 분석을 단계 S503에서 수행한다.

도 6a 및 도 6b는 음성 개시 커맨드를 입력하는 타이밍 차에 기인한 음성 생략의 개략도이다. 횡축은 시간 눈금이며, 시각 S에서 음성이 시작된다. 도 6a는 음성을 개시하라는 커맨드가 시각 P(P<S)에서 입력되는 경우이다. 시각 P(또는 P 직후)에서 음성 취입이 개시될 수 있기 때문에, 음성은 생략되지 않으며 적절하게 취입된다. 한편, 도 6b는 음성을 개시하라는 커맨드가 시각 Q(S<Q)에서 입력되는 경우이다. 시각 Q(또는 Q 직후)에서 음성 취입이 개시되기 때문에, 음성의 시작부분이 생략된다. 음성 분석 및 음성의 시작부분이 생략되는지의 여부 판정은 다음의 방법으로 수행된다.

음성 분석 및 판정을 수행하기 위한 여러가지 방법이 있다. 쉽고 간편한 한 방법은 취입된 음성 파형(예컨대, 300 샘플)의 헤드 부분을 이용하여 파형 파워를 계산하여, 그 결과를 소정의 임계값과 비교하는 것이다. 그 결과가 임계값을 초과하는 경우에는, 음성의 시작부분이 생략되고 있는 것이라고 판정될 수 있다. 영교차율, 분석, 스펙트럼 분석 또는 기본 주파수 분석과 같은 다른 분석을 수행함으로써 판정을 행할 수도 있다.

영교차율은 취입된 음성 데이터를 코드(예를 들어, 16비트의 signed short 연산의 경우, -32768과 32767 사이의 값을 취함)로 표현하고, 그 코드가 변화한 횟수를 카운트함으로써 얻어질 수 있다. 이 영교차율은 음성 파형의 헤드 부분에 대해 얻어지며, 그 결과는 상술된 파형 파워인 임계값과 비교된다. 따라서, 음성의 시작부분은 결과가 임계값보다 큰 경우에는 생략되는 것으로, 결과가 임계값보다 작거나 같은 경우에는 생략되지 않는 것으로 판정될 수 있다.

스펙트럼 분석은, 예를 들어 음성 인식 특징 파라미터 추출 유닛(302)의 음성 인식의 특징 파라미터 추출과 동일한 방식으로 수행될 수 있다. 다음으로, 추출된 특징 파라미터를 이용하여 음성 모델과 비음성 모델의 우도(또는 확률)를 획득하여, 음성 모델의 우도가 비음성 모델의 우도보다 크면, 음성이 생략되고 있다고 판정한다. 음성 모델의 우도가 비음성 모델의 우도보다 작은 경우에는, 생략되고 있지 않다고 판정한다. 음성 모델과 비음성 모델은 음성 부분의 특징 파라미터 및 비음

성 부분의 특징 파라미터로부터 미리 통계 모델로서 준비되어 있다. 이 모델들은 어떤 기존의 방법, 예를 들면 GMM (Gaussian Mixture Model)에 의해 생성될 수 있다. 또한, 음성 특징 파라미터 추출 유닛(302)의 음성 인식의 특징 파라미터 추출과는 다른 분석에 의해 획득된 다른 스펙트럼을 나타내는 특징 파라미터를 이용한 방법을 이용할 수도 있다.

기본 주파수 분석에 대해서는, 자기 상관 기술(autocorrelation technique) 또는 캡스트럼 기술(Cepstrum technique)과 같은 기존의 분석 기술을 이용할 수 있다. 생략은, 기본 주파수 값을 직접적으로 이용하는 대신에, 주기성에 관련된 값을 이용하여 판정한다. 보다 정확하게는, 예를 들면, 캡스트럼 기술에 기초한 기본 주파수 분석의 경우, 큐프렌시(quefrency) (대수 진폭 스펙트럼의 역이산 푸리에 변환) 시의 계열의 소정의 범위 내(사람의 소리의 피치의 범위 내)의 최대값을 이용할 수 있다. 이 값은 음성 파형의 헤드 부분에 대해 획득되어, 파형 파워의 경우에서와 같이 임계값과 비교된다. 그 값이 임계값보다도 큰 경우에는 음성이 생략된 것으로 판정하고, 그 값이 임계값보다 작은 경우에는 음성이 생략되지 않는 것으로 판정한다. 그 외에도, 기본 주파수 대신에 고조파 구조를 획득하도록 분석을 수행하는 방법을 이용할 수 있으며, 그 결과는 특징 파라미터로서 이용된다.

음성이 단계 S504에서 생략된 것으로 판정되면, 단계 S505에서 생략된 음성의 발음 정보를 설정한다. 이어서, 단계 S506에서 이 발음 정보를 이용하여 음성 인식을 수행한다. 음성이 단계 S504에서 생략되지 않은 것으로 판정되면, 단계 S506에서 통상의 음성 인식을 수행한다. S505에서 수행된 처리를 도 7 내지 도 11을 참조하여 설명한다. S505의 처리시, 인식 대상 단어는 "Tokyo", "Hiroshima", "Tokushima", "Tu"이다. 도 7은 인식 대상 단어의 예이며, 단어 ID, 표기, 발음(음소)의 정보를 유지하고 있다. 발음(음소) 계열("Tokyo"의 경우에는, /t o o k y o o/의 7음소)에 따라서 음향 모델(306)(예를 들어, 음소 HMM)에 접속함으로써 음성 인식 처리의 참조 패턴을 생성한다. 도 8은 제1 음소가 도 7의 발음 정보로부터 삭제된 경우의 인식 대상 단어를 도시한다. 예를 들면, "Tokyo"의 경우에는, 제1 음소 /t/가 삭제되어, 인식 대상 단어가 /o o k y o o/가 된다. 도 9 및 도 10은 제2 및 제4 음소가 삭제된 경우의 인식 대상 단어를 도시한다. "Tu"의 경우, 발음 계열은 /ts u/의 2음소이다. 그러므로, 2 이상의 음소가 삭제되면 발음 계열이 없어져 버릴 것이다. 이러한 경우에는, 무음 모델(SIL)을 발음 계열로서 할당한다. 부가적으로, 도 10의 "Hiroshima" 및 "Tokushima"의 경우에는, 처음 4개의 음소가 삭제되면, 동일한 발음 계열(/shima/)이 될 것이다. 단계 S504에서 음성이 생략되지 않는 것으로 판정되면, 단계 S506에서 단지 도 7의 대상 단어에 대해서만 음성 인식을 수행한다. 한편, 단계 S504에서 음성이 생략된 것으로 판정되면, 단계 S505에서, 도 7의 인식 대상 단어에 추가하여 도 8 내지 도 10의 대상 단어에 대해서도 음성 인식을 수행한다. 도 8 및 도 10의 대상 단어에서, 발음 계열의 헤드 부분이 삭제되었다. 단계 S503의 음성 분석과 단계 S504의 음성 생략 판정을 수행함으로써, 음성이 생략되고 있는지의 여부가 판정될 수 있다. 그러나, 생략된 음성의 길이 또는 음소의 수를 추정할 수는 없다. 따라서, 추가되어야 할 대상 단어의 삭제된 음소의 적절한 수에 대해 미리 결정할 필요가 있다. 그 수는 경험적으로 설정될 수도 있고, 또는 이용자의 조작 또는 동작에 따라 생략된 음성의 경향을 고려하여 설정될 수도 있으며, 또는 인식 성능을 고려하여 설정될 수도 있다. 제1 내지 제4 음소의 발음 계열이 삭제된 단어의 모든 조합이 인식 대상이 될 수 있다. 이러한 경우, 도 11에 도시된 바와 같은 대산 단어는 음성 생략에 대한 발음 정보와 같이 설정된다.

단계 S503에서의 스펙트럼 분석 또는 기본 주파수 분석은 음성 인식 처리에서의 음성 특징 파라미터 추출과 동일하거나 유사한 처리이다. 그러므로, 이 처리들은 음성 인식 유닛(204)에 포함될 수도 있고, 음성 인식 유닛(204) 내에 구성되어 있는 것으로서 실행될 수도 있다. 도 17은 음성 인식 처리시 취입된 음성 판정 및 발음 정보 설정을 포함하는 음성 인식 방법의 모듈 구성의 블록도이다. 취입된 음성 판정 유닛(202) 및 발음 정보 설정 유닛(203)은 취입된 음성 판정 유닛(603) 및 발음 정보 설정 유닛(604)으로서 도 3의 처리에 각각 포함된다. 음성 입력 유닛(601) 내지 언어 모델(609)은 도 2 및 도 3의 것과 동일한 것이므로, 그 설명은 생략한다.

또한, 음성 분석은 첫 음성 프레임만을 이용하여 단계 S503에서 반드시 수행될 필요는 없으나, 복수 프레임(예를 들면, 처음 5개의 프레임)에 대한 정보가 이용될 수도 있다. 부가적으로, 음성이 생략되고 있는지를 판정하기 위해, 본 발명은 단계 S504에 도시된 바와 같이, 임계값이 비교될 때 소정의 값을 이용하는 것으로 한정되지는 않는다. 예를 들어, 첫 프레임과 10번째 프레임의 파형 파워를 비교하는 등의 다른 처리를 수행할 수도 있다. 이 경우, 첫 프레임의 파형 파워가 10번째 프레임보다 훨씬 작을 경우(예를 들어, 10% 미만일 경우), 어떤 음성 생략도 없는 것으로 판정한다.

단계 S504에, 음성이 생략되고 있는지의 여부를 판정하는 예를 제공했다. 그러나, 본 발명은 이 예에 한정되는 것은 아니고, 음성 취입이 이용자의 음성 도중에 개시되는지의 여부를 판정하도록 구성될 수 있다.

상기 실시예에 따르면, 이용자가 음성 개시 커맨드를 정확한 타이밍에 입력하지 않아도, 인식 성능의 저하를 방지할 수 있다. 결과적으로, 음성 인식 장치를 조작하는데 익숙하지 않은 이용자라도 조작 수행시 쉽게 여길 수 있게 된다.

(제2 실시예)

제1 실시예에서는, 인식 대상 단어의 발음을 음소화하고, 판독을 위한 발음 계열이 삭제되어, 단계 S505에서 그 생략된 음성에 대한 발음 정보를 설정했다. 그러나, 본 발명은 본 실시예에 제한되는 것이 아니다. 인식 대상 단어의 발음을 음소에 비해 보다 상세한 발음 계열을 이용하여 표현할 수 있으며, 그 상세화된 발음 계열을 삭제할 수도 있다. 보다 자세히 말하면, 히든 마르코프 모델(HMM:Hidden Markov Model)에 기초하여 음성 인식을 수행할 때, 음소는 통상적으로 복수의 상태로 모델화된다. 이 상태 계열은 상세화된 발음 계열로 간주되어, 상태 레벨에서 삭제된다. 이러한 방식으로, 음소 레벨에서의 삭제에 비해 보다 정밀하게 발음 정보를 설정할 수 있다. 도 12는 음소/t/가 HMM의 3가지 상태(t1, t2, t 3)로 모델화된 예이다. 도 7의 발음을 그러한 상태 계열로 설명할 경우, 도 13에 도시된 바와 같은 표현이 가능해진다. 이 경우, 도 13의 상태 계열에서 제1 상태 계열을 삭제하면, 도 14를 얻을 수 있다.

도 15a, 도 15b 및 도 15c는 발음(음소) 계열의 삭제와 상태 계열의 삭제의 차이를 설명하는 개략도이다. 모든 음소가 HMM의 3가지 상태로 모델화되는 경우, "Tokyo"의 발음 계열인 /t o o k y o o/는 도 15a에 도시된 바와 같이 HMM의 링크로 표현된다. 제1 음소(/t/)가 삭제되면, 도 15b에 도시된 바와 같이 /t/의 3가지 HMM 상태가 모두 삭제된다. 그러나, "Tokyo"의 상세화된 발음 계열이 HMM의 상태 계열로 표현되면, 도 15c에 도시된 바와 같이 /t/의 HMM의 제1 상태 t1만을 삭제할 수 있다. 즉, 음소 레벨 대신에 상태 레벨이 삭제됨으로써, 보다 상세화된 발음 정보가 설정될 수 있다. 대안으로, 상술된 HMM 대신에 일반적인 상태 천이 모델을 이용하여도 동일한 처리를 수행할 수 있다.

(제3 실시예)

상기 실시예에 따른 발음 정보는 인식 대상 단어가 발음 계열 또는 상세화된 발음 계열로서 표현될 수 있는 경우에 설정될 수 있다. 그러나, 상기 설정은 널리 이용되고 있는 음소 HMM에 기초한 불특정 화자 음성 인식(비등록 요구형의 음성 인식 방법)에도 이용될 수 있다. 보다 구체적으로는, 특정 화자 음성 인식(등록 요구형 음성 인식 방법)에서는 참조 패턴으로부터 음소 또는 상태 계열을 식별할 수 없다. 특정 화자 음성 인식에서는, 음성 인식을 이용하기 전에 참조 패턴이 음성에 의해 등록된다. 따라서, 상기 실시예에서 설명된 방법은 이용할 수 없다. 그러나, 참조 패턴의 특징 파라미터 계열을 직접 이용하면, 생략된 음성에 대한 발음 정보를 설정하는 것이 가능해진다.

도 4는 등록 요구형의 음성 인식 방법의 모듈 구성을 도시한 블록도이다. 음성 입력 유닛(401)에서 결과 출력 유닛(404)까지의 블록들은 음성 입력 유닛(301)에서 결과 출력 유닛(304)까지의 블록과 동일하기 때문에, 이 유닛들의 설명은 생략한다. 인식 대상 단어는 음성에 의해 미리 등록된다. 참조 패턴(405)은 그 등록된 음성의 특징 파라미터 계열로서 보유된다. 12차 켑스트럼과 그 12차 캡스트럼의 1차 회귀 계수인 델타켑스트럼(c1~c12, Δc1~Δc12)으로 특정 파라미터 계열이 유지되어 있다고 가정한다. 이 경우, "Tokyo"라는 단어에 대한 등록된 음성의 특징 파라미터 계열은, 도 16a(T1은 등록된 음성을 분석할 때의 프레임 수임)에 도시된 바와 같이 참조 패턴 계열(24차원의 벡터 계열)로서 보유된다. 음성이 단계 S504에서 생략되고 있다고 판정되면, 도 16b(첫 프레임이 삭제됨) 또는 도 16c(첫 프레임과 둘째 프레임이 삭제됨)에 도시된 바와 같이, 참조 패턴으로부터 처음 몇개의 프레임이 삭제된다. 그 삭제된 프레임을 포함한 특징 파라미터 계열의 음성 인식에 의해, 음성의 시작부분이 생략된 음성 입력에 대해서도 거의 저하 없이 음성 인식을 수행할 수 있다.

또한, 본 발명의 목적은 상기 실시예의 기능을 실현하는 소프트웨어의 프로그램 코드를 저장하고 있는 저장 매체를 시스템 또는 장치에 공급하고, 그 시스템 또는 장치의 컴퓨터(또는 CPU 또는 MPU)에 의해 저장 매체에 저장된 프로그램 코드를 검색 및 실행함으로써 달성될 수 있다.

이 경우, 저장 매체로부터 판독된 프로그램 코드 자체는 상기 실시예의 기능을 실현하여, 그 프로그램 코드를 저장하는 저장 매체는 본 발명을 구성할 수 있게 된다.

프로그램 코드를 공급하기 위한 저장 매체의 예들로서는, 플렉시블 디스크, 하드디스크, 광 디스크, 광 자기 디스크, CD-ROM, CD-R, 자기 테이프, 불휘발성 메모리 카드 및 ROM을 들 수 있다.

또한, 컴퓨터에 의해 검색된 프로그램 코드를 실행함으로써 상기 실시예의 기능을 실현할 뿐만 아니라, 본 발명은, 컴퓨터에서 가동중인 운영체제(OS)가 그 프로그램 코드의 명령에 따라 실제 처리의 일부 또는 전부를 수행하여, 그 처리가 상기 실시예의 기능들을 실현하는 경우까지도 포함하고 있다.

더구나, 프로그램 코드가 저장 매체로부터 검색되어, 컴퓨터에 삽입된 기능 확장 보드 또는 컴퓨터에 접속된 기능 확장 유닛 내의 메모리에 로딩된 후, 그 기능 확장 보드 또는 기능 확장 유닛 유닛 내의 CPU가 그 프로그램 코드의 명령에 따라 실제 처리의 일부 또는 전부를 수행하여, 그 처리가 상기 실시예의 기능들까지 실현하는 경우까지 포함하고 있다.

본 발명은 하드웨어는 물론 하드웨어와 소프트웨어의 조합에 의해서도 구현될 수 있다.

본 발명을 실시예들을 참조하여 설명하였지만, 본 발명이 기술된 실시예들에만 제한되는 것이 아님을 이해해야 한다. 이하의 청구범위는 실시예들에 대한 모든 변경들, 등가 구조들 및 기능들을 포함하도록 가장 넓게 해석되어야 한다.

발명의 효과

본 발명은 음성의 시작부분이 누락 또는 생략되더라도, 간단하고 쉬운 공정으로 음성 인식 성능의 저하를 방지할 수 있다.

도면의 간단한 설명

도 1은 본 발명의 제1 실시예에 따른 음성 인식 방법이 탑재된 정보 장치의 하드웨어 구성의 블록도.

도 2는 본 발명의 제1 실시예에 따른 음성 인식 방법의 모듈 구성의 블록도.

도 3은 통상의 등록 비요구형 음성 인식 방법의 모듈 구성의 블록도.

도 4는 통상의 등록 요구형 음성 인식 방법의 모듈 구성의 블록도.

도 5는 본 발명의 제1 예시적인 실시예에 따른 음성 인식 방법의 전체 처리의 흐름도.

도 6a 및 도 6b는 발성을 개시하라는 커맨드를 입력하는 타이밍 차에 기인한 음성 생략의 개략도.

도 7은 인식 대상 단어의 일례를 도시한 도면.

도 8은 도 7의 인식 대상 단어를, 제1 발음 계열을 삭제한 일례를 도시한 도면.

도 9는 도 7의 인식 대상 단어를, 제1 및 제2 발음 계열을 삭제한 일례를 도시한 도면.

도 10은 도 7의 인식 대상 단어를, 제1 내지 제4 발음 계열을 삭제한 일례를 도시한 도면.

도 11은 도 7의 인식 대상 단어를, 제1 내지 제4 발음 계열을 삭제한 모든 조합의 일례를 도시한 도면.

도 12는 음소 /t/를 3 상태의 히든 마르코프 모델(HMM)로 모델화한 예를 도시한 도면.

도 13은 인식 대상 단어의 일례로서, 도 7의 인식 대상 단어의 발음 정보를 HMM의 상태 계열로 표현한 도면.

도 14는 도 13의 인식 대상 단어를, 제1 상태 계열을 삭제한 일례를 도시한 도면.

도 15a, 도 15b, 도 15c는 발음 계열의 삭제 및 상태 계열의 삭제 간의 차이를 설명하는 개략도.

도 16a, 도 16b, 도 16c는 발음 정보가 참조 패턴 계열의 삭제에 의해 어떻게 설정되는지를 설명하는 개략도.

도 17은 음성 인식 처리에 취입된 음성의 판정 및 발음 정보의 설정을 포함하는 음성 인식 방법의 모듈 구성의 블록도.

<도면의 주요 부분에 대한 부호의 설명>

104: 외부 저장 장치

105: 음성 입력 장치

106: 디스플레이 장치

107: 보조 입력 장치

108: 보조 출력 장치

201: 음성 취입 유닛

202: 취입된 음성 판정 유닛

203: 발음 정보 설정 유닛

204: 음성 인식 유닛

301: 음성 입력 유닛

302: 음성 특징 파라미터 추출 유닛

303: 탐색 유닛

304: 결과 출력 유닛

305: 발음 사전

306: 음향 모델

307: 언어 모델

401: 음성 입력 유닛

402: 음성 특징 파라미터 추출 유닛

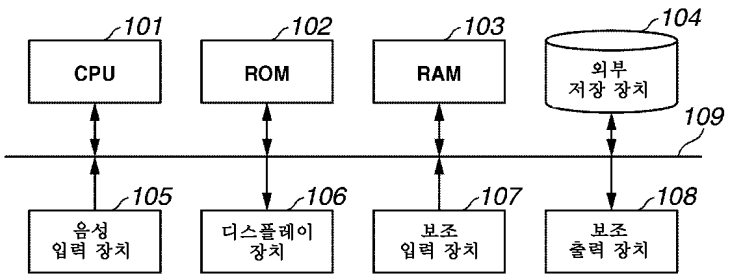
403: 탐색 유닛

404: 결과 출력 유닛

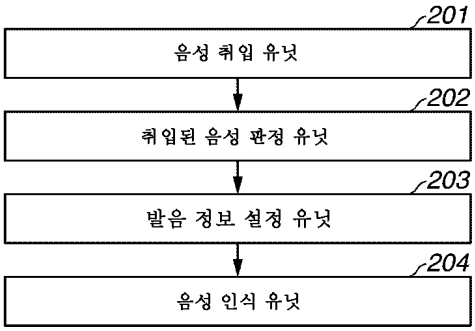
405: 참조 패턴

도면

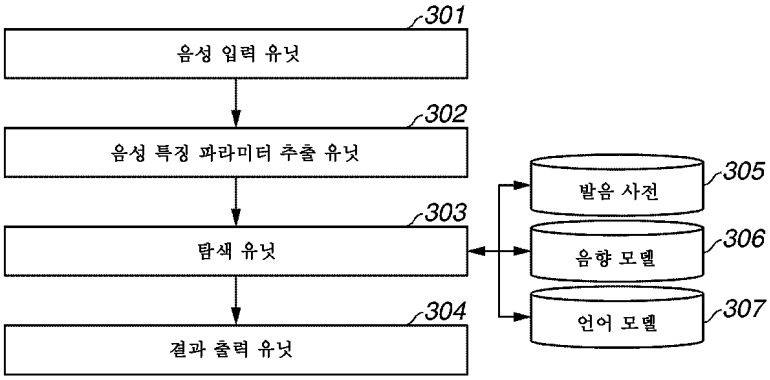
도면1



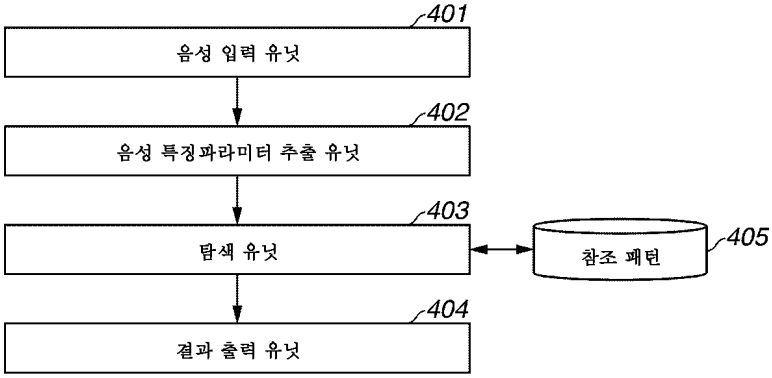
도면2


201
음성 취입 유닛
202
취입된 음성 판정 유닛
203
발음 정보 설정 유닛
204
음성 인식 유닛

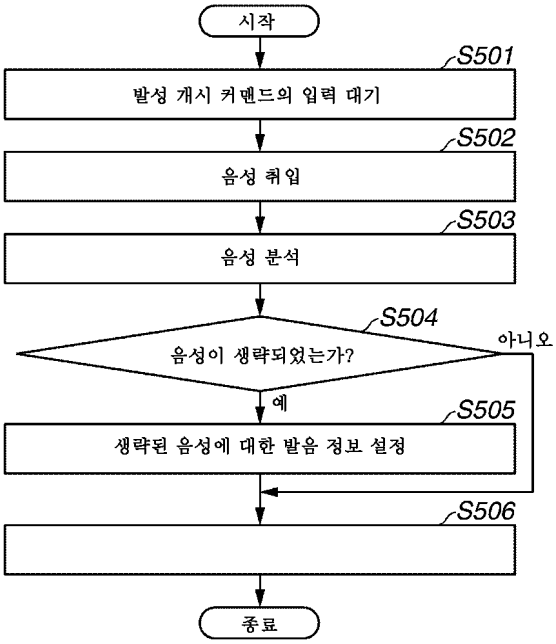
도면3


301
음성 입력 유닛
302
음성 특징 파라미터 추출 유닛
303
탐색 유닛
304
결과 출력 유닛
305
발음 사전
306
음향 모델
307
언어 모델

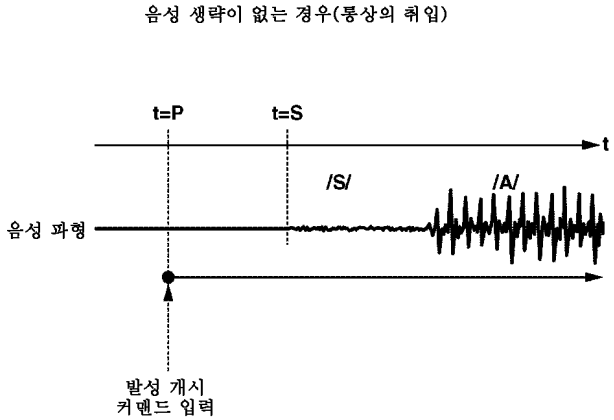
도면4


401
음성 입력 유닛
402
음성 특징파라미터 추출 유닛
403
탐색 유닛
404
결과 출력 유닛
405
참조 패턴

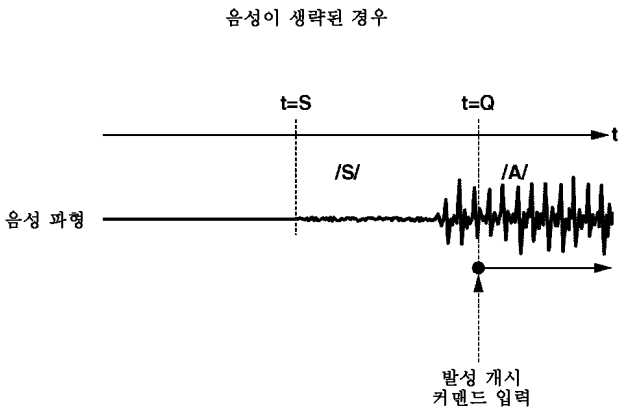
도면5



도면6a



도면6b



도면7

단어 ID	전사	발음
001	TOKYO	tookyoo
002	HIROSHIMA	hiroshima
003	TOKUSHIMA	tokushima
004	TU	tsu

도면8

단어 ID	전사	발음
101	TOKYO	ookyoo
102	HIROSHIMA	iroshima
103	TOKUSHIMA	okushima
104	TU	u

도면9

단어 ID	전사	발음
201	TOKYO	okyoo
202	HIROSHIMA	roshima
203	TOKUSHIMA	kushima
204	TU	SIL

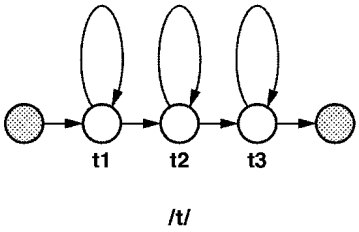
도면10

단어 ID	전사	발음
401	TOKYO	yoo
402	HIROSHIMA	shima
403	TOKUSHIMA	shima
404	TU	SIL

도면11

단어 ID	전사	발음
001	TOKYO	tookyoo
101	TOKYO	ookyoo
201	TOKYO	okyoo
301	TOKYO	kyoo
401	TOKYO	yoo
002	HIROSHIMA	hiroshima
102	HIROSHIMA	iroshima
202	HIROSHIMA	roshima
302	HIROSHIMA	oshima
402	HIROSHIMA	shima
003	TOKUSHIMA	tokushima
103	TOKUSHIMA	okushima
203	TOKUSHIMA	kushima
303	TOKUSHIMA	ushima
403	TOKUSHIMA	shima
004	TU	tsu
104	TU	u
204	TU	SIL

도면12



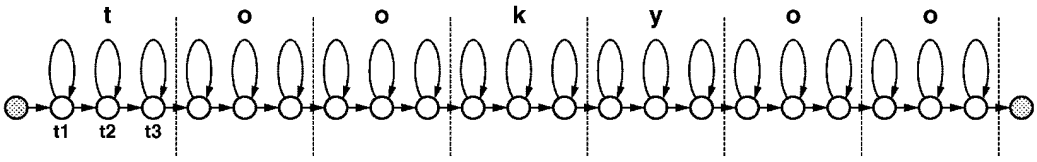
도면13

단어 ID	전사	상태 계열
001	TOKYO	t1 t2 t3 o1 o2 o3 o1 ...
002	HIROSHIMA	h1 h2 h3 i1 i2 i3 r1 ...
003	TOKUSHIMA	t1 t2 t3 o1 o2 o3 k1 ...
004	TU	ts1 ts2 ts3 u1 u2 u3
:	:	:

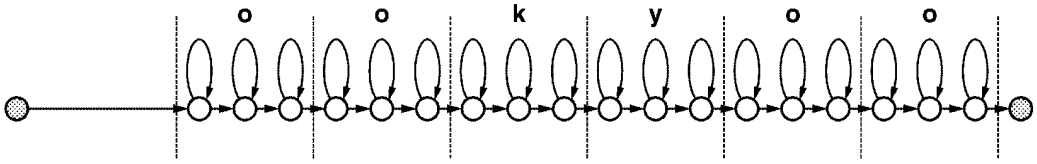
도면14

단어 ID	전사	상태 계열
101	TOKYO	t2 t3 o1 o2 o3 o1 ...
102	HIROSHIMA	h2 h3 i1 i2 i3 r1 ...
103	TOKUSHIMA	t2 t3 o1 o2 o3 k1 ...
104	TU	ts2 ts3 u1 u2 u3
:	:	:

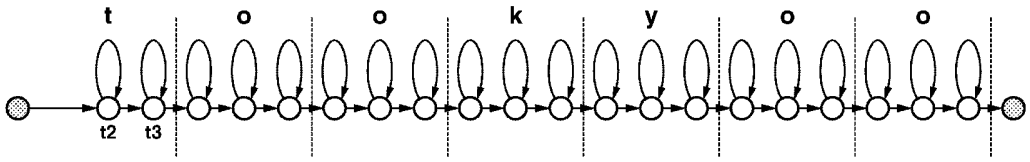
도면15a



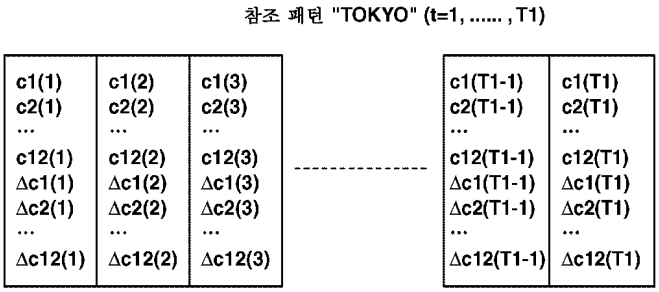
도면15b



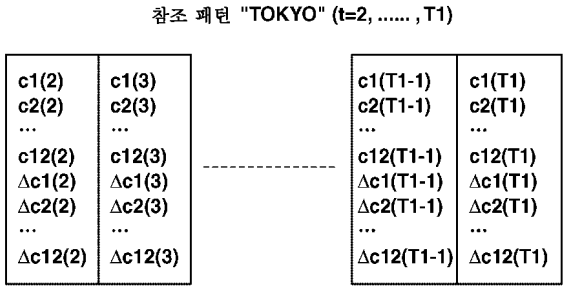
도면15c



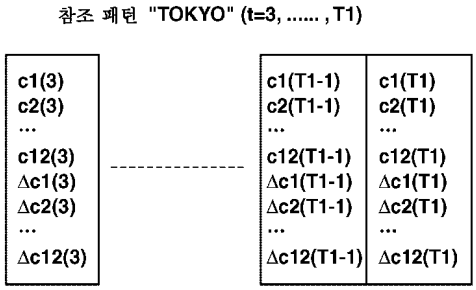
도면16a



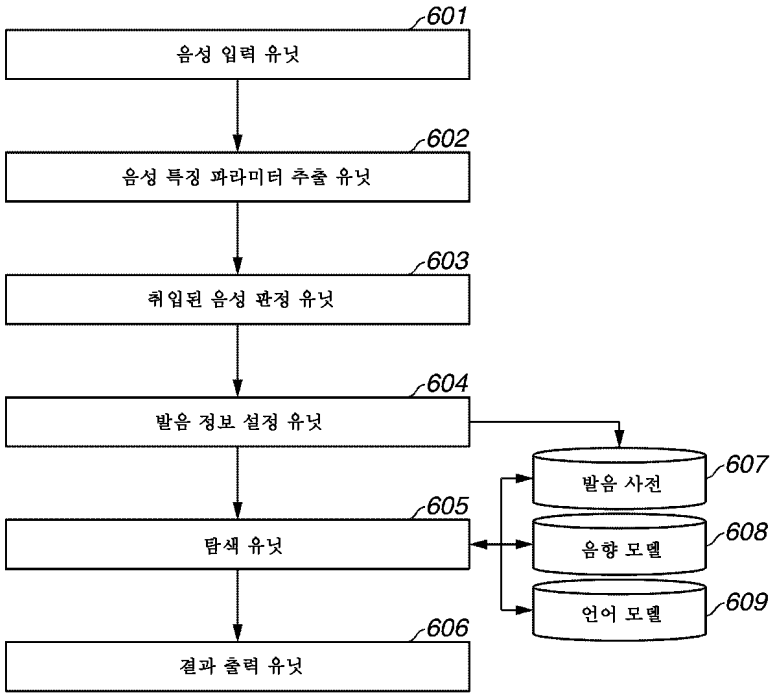
도면16b



도면16c



도면17


601
음성 입력 유닛
602
음성 특징 파라미터 추출 유닛
603
취입된 음성 판정 유닛
604
발음 정보 설정 유닛
605
탐색 유닛
606
결과 출력 유닛
607
발음 사전
608
음향 모델
609
언어 모델