(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2011년06월17일
(11) 등록번호 10-1042119
(24) 등록일자 2011년06월09일

(51) Int. Cl.
G10L 15/08 (2006.01)
(21) 출원번호 10-2004-0038497
(22) 출원일자 2004년05월28일
심사청구일자 2009년04월27일
(65) 공개번호 10-2004-0103445
(43) 공개일자 2004년12월08일
(30) 우선권주장
10/448.018 2003년05월29일 미국(US)
(56) 선행기술조사문헌
Kuansan Wang, 'SALT: A spoken language interface for web-based multimodal dialog systems', ICSLP 2002, pp.2241-2244, September 2002*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
마이크로소프트 코포레이션
미국 워싱턴주 (우편번호 : 98052) 레드몬드 원 마이크로소프트 웨이
(72) 발명자
왕쿠안산
미국98006워싱턴주벨레뷰사우스이스트48 번코트16470
(74) 대리인
주성민, 이중희, 백만기

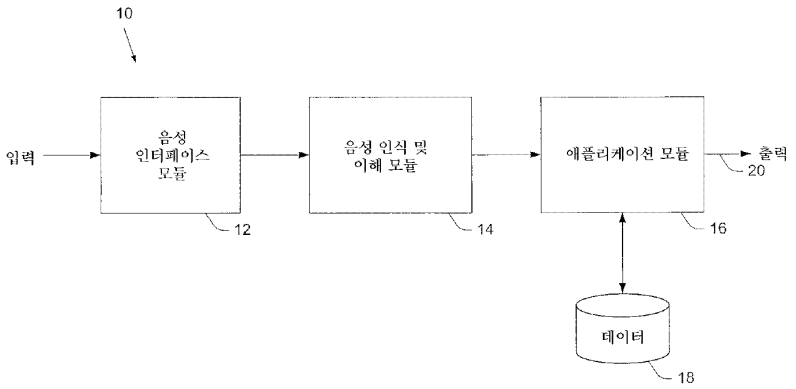
전체 청구항 수 : 총 8 항 심사관 : 정성윤

(54) 음성 이해 시스템, 및 컴퓨터 판독가능 기록 매체

(57) 요 약

음성 이해 시스템(speech understanding system)은 N-그램(N-gram) 언어 모델과 문맥 무관 문법(context-free grammar) 언어 모델의 조합을 포함한다. 상기 언어 모델은 인식될 단어와 의미 정보에 관련된 정보를 저장한다. 모듈은 사용자로부터 입력을 수신하고 처리를 위해 상기 입력을 캡쳐하도록 구성되어 있다. 상기 모듈은 또한 상기 입력의 인식에 관련된 음성 응용 언어 태그(SALT; Speech Application Language Tags) 애플리케이션 프로그램을 수신하도록 구성되어 있다. 상기 모듈은 상기 모듈은 상기 음성 응용 언어 태그(SALT) 애플리케이션 프로그램 인터페이스와 상기 입력을 처리하여 상기 입력의 제1 부분(first portion)에 관련된 의미 정보를 확인하고 상기 언어 모델을 액세스함으로써 상기 제1 부분에 대한 텍스트와 의미 정보를 포함하는 의미 객체를 출력하고, 상기 인식의 실행과 상기 의미 객체의 출력은 다음 입력 부에 대하여 계속 캡쳐하면서 수행되도록 구성된다.

대 표 도



특허청구의 범위

청구항 1

음성 이해 시스템(speech understanding system)에 있어서,

N-그램(N-gram) 언어 모델과 문맥 무관 문법(context-free grammar) 언어 모델의 조합을 포함하는 언어 모델 - 상기 언어 모델은 인식될 단어들과 의미 정보(semantic information)에 관련된 정보를 저장함 -; 및

사용자로부터 입력을 수신하고, 처리를 위해 상기 입력을 캡쳐하도록 구성된 모듈

을 포함하고,

상기 모듈은 또한 상기 입력의 인식에 관한 음성 응용 언어 태그(SALT; Speech Application Language Tags) 애플리케이션 프로그램 인터페이스들을 수신하도록 구성되고, 처리는 상기 수신된 입력으로부터 텍스트를 인식하는 것을 포함하며,

상기 모듈은 상기 입력의 제1 부분(first portion)에 대한 의미 정보를 확인하기 위해 상기 입력의 제1 부분 및 상기 음성 응용 언어 태그(SALT) 애플리케이션 프로그램 인터페이스들을 처리하도록 구성되고,

상기 모듈은 상기 언어 모델에 액세스함으로써 상기 입력의 제1 부분에 대한 의미 정보 및 텍스트를 포함하는 부분적 의미 객체(partial semantic object)를 출력하도록 구성되고, 텍스트의 인식을 수행하는 것과 부분적 의미 객체들을 출력하는 것은 상기 입력의 다음 부분들(subsequent portions of the input)에 대한 캡쳐가 계속되고 있는 동안 수행되는,

음성 이해 시스템.

청구항 2

제1항에 있어서,

상기 언어 모델은 통합 언어 모델(unified language model)을 포함하는, 음성 이해 시스템.

청구항 3

제1항에 있어서,

상기 언어 모델은 의미 언어 모델(semantic language model)을 포함하는, 음성 이해 시스템.

청구항 4

제1항 내지 제3항 중 어느 한 항에 있어서,

상기 모듈은 상기 입력의 다음 부분들에 대한 캡쳐가 계속되고 있는 동안 텍스트의 인식을 수행하고 부분적 의미 객체들을 출력하기 위해 다중 모드(multiple mode)로 음성 응용 언어 태그(SALT)의 듣기 객체(listen object)를 식별하도록 구성된, 음성 이해 시스템.

청구항 5

제1항 내지 제3항 중 어느 한 항에 있어서,

상기 모듈은 문법 객체들을 식별하여 상기 언어 모델의 검색 공간(search space)을 정의하도록 구성된, 음성 이해 시스템.

청구항 6

삭제

청구항 7

구현시에 컴퓨팅 장치가 단계들을 수행하여 정보를 처리하도록 하는, 상기 컴퓨팅 장치에 의해 판독가능한 명령

어들을 저장하는 컴퓨터 판독가능 기록 매체로서,

상기 단계들은,

사용자로부터 입력을 수신하고, 처리를 위해 상기 입력을 캡쳐하는 단계 - 처리는 상기 수신된 입력으로부터 텍스트를 인식하는 것을 포함함 -;

언어 모델을 식별하여 텍스트의 인식 및 이해를 수행하기 위해 음성 응용 언어 태그(SALT) 애플리케이션 프로그램 인터페이스들을 수신하는 단계 - 상기 언어 모델은 인식된 입력의 텍스트와 상기 수신된 입력에 대한 의미 정보를 제공하도록 구성됨 -; 및

상기 입력에 대한 인식을 수행하여 상기 입력의 제1 부분에 대한 의미 정보를 확인하기 위해, 상기 언어 모델에 액세스하여 상기 입력을 처리하고, 상기 입력의 제1 부분에 대한 의미 정보와 상기 인식된 입력의 텍스트를 포함하는 부분적 의미 객체를 출력하는 단계 - 텍스트의 인식을 수행하는 것과 부분적 의미 객체들을 출력하는 것은 상기 입력의 다음 부분들에 대한 캡쳐가 계속되고 있는 동안 수행됨 -

를 포함하는, 컴퓨터 판독가능 기록 매체.

청구항 8

제7항에 있어서,

처리는, 상기 입력의 다음 부분들에 대한 캡쳐가 계속되고 있는 동안, 텍스트의 인식을 수행하고 부분적 의미 객체들을 출력하기 위해서 다중 모드로 음성 응용 언어 태그(SALT)의 듣기 객체를 식별하는 것을 포함하는, 컴퓨터 판독가능 기록 매체.

청구항 9

제7항 또는 제8항에 있어서,

상기 음성 응용 언어 태그(SALT) 애플리케이션 프로그램 인터페이스들을 수신하는 것은, 문법 객체들을 식별하여 상기 언어 모델의 검색 공간을 정의하는 것을 포함하는, 컴퓨터 판독가능 기록 매체.

청구항 10

삭제

명 세 서

발명의 상세한 설명

발명의 목적

발명이 속하는 기술 및 그 분야의 종래기술

[0014] 본 발명은 컴퓨터 시스템에서의 정보의 액세스 및 렌더링에 관한 것이다. 보다 상세하게는, 인식(recognition) 및 이해(understanding)를 사용한 정보의 액세스에 관한 것이다.

[0015] 최근, 사용자가 음성 명령을 제공하여 컴퓨터 시스템에 대한 정보를 액세스를 할 수 있게 하는 기술이 발전되어 왔다. 사용자 명령의 수신시에, 컴퓨터 시스템이 원하는 동작을 컴퓨터 시스템은 사용자 입력에 대하여 음성 인식을 수행하고 수행하기 위해서는 사용자의 의도를 확인하기 위해 입력을 더 처리한다.

[0016] 몇몇 상황에서, 사용자에 의해 제공된 입력이 불충분하거나 불명확한 경우에는, 컴퓨터는 시각(visual) 또는 가청(audible) 프롬프트의 형태로 사용자에게 추가 정보를 요청할 수 있다. 따라서, 사용자와 컴퓨터 시스템 사이에 대화가 설정될 수 있으며, 이 경우에는 사용자의 의도가 확인되어 동작이 수행될 수 있을 때까지 각자 번갈아가며 질문, 응답, 및/또는 인식 등을 제공한다. 다른 상황에서, 이러한 대화를 하는 것은 컴퓨터 시스템과 상호작용하는데 있어 바람직한 모드이다.

[0017] 음성 응용 언어 태그(Speech Application Language Tags; SALT)는 최신 사용자 인터페이스 설계에 있어서 경쟁력있는 입출력 모드로서 음성을 이용하도록 도입되어 왔다. 음성 응용 언어 태그(SALT)에서 설계 목표는 일반 음성 작업이 프로그램하기에는 용이하지만, 간단한 구현을 통해 개선된 성능을 구비할 수 있도록 하는 것이다.

음성 응용 언어 태그(SALT)는 많은 애플리케이션에서 설계되었다. 예를 들어, 오로지 음성 대화만으로 사용자와 상호작용하는 전화 기반의 음성 전용 애플리케이션이 그것이다.

[0018] 음성 응용 언어 태그(SALT)는 음성 입출력 객체("듣기(listen)" 및 "프롬프트(prompt)")를 포함하며, 이들은 사용자 순번(turn)의 시작과 끝을 탐지하는 기술을 통합하는 모드 설계를 구비한다. 따라서, 많은 음성 애플리케이션은 사용자에게 사용자 순번의 개시를 알리도록 요구하는 사용자 인터페이스를 채용한다. 몇몇 컴퓨터 시스템은 착용가능 컴퓨터, 음성가능 모드(speech enabled mode) 또는 다중 모드(마우스와 같은 입력 장치에 의해 선택된 필드에 대하여 제공되는 음성입력) 장치 및 기타 시각에 의존하지 않는(eyes-free) 애플리케이션을 포함한다. 그럼에도 불구하고, 이들 각각의 환경에서, 대화에서 사용자와 컴퓨터 시스템의 순번에 대한 명확한 정의가 여전히 존재한다.

발명이 이루고자 하는 기술적 과제

[0019] 그러나, 사람들 간의 대화는 통상 참가자들 사이에 명확하고 번갈아가며 하는 대화는 아니다. 그보다는, 대화는 한 참가자가 응답, 확인, 질문 등을 하는 반면, 다른 참가자는 화자(speaker)가 정보를 제공하는 태도에 과도하게 영향을 주거나 영향을 거의 주지 않거나 영향을 전혀 주지 않는 정보를 제공하는 식으로 이루어질 수 있다. 말하는 사람은 이러한 자연스러운 형태의 대화를 즐긴다. 유사하게, 전화 시스템도 이러한 대화가 이루어지도록 양방향(full duplex) 기술을 채용하고 있다.

[0020] 이와 달리, 대화 기반 인터페이스는 사용자와 컴퓨터 시스템 간의 동작의 경직된 순번을 사용하여, 다음 동작을 취하여 처리하기 전에는, 컴퓨터 시스템이 사용자의 대화가 끝날 때까지 기다리게 한다. 비록 컴퓨터 화면에 걸쳐 일련의 도트 처리와 같은 시각적 표시 따위의 단순 피드백이 사용자에게 컴퓨터 시스템이 적어도 무언가를 처리하고 있다는 일종의 확신을 줄 수는 있지만, 사용자가 그/그녀의 순번을 끝내고 컴퓨터 시스템이 응답할 때까지는, 컴퓨터 시스템이 어느 정도 이해하고 있는지는 모르게 된다.

[0021] 따라서, 인식 및 이해에 기초한 컴퓨터 시스템의 개선이 요청된다. 이러한 개선은 사용자에게 보다 자연스럽게 됨으로써 사용하기에 보다 용이할 수 있는 정보를 액세스하는 시스템 및 방법을 제공한다.

발명의 구성 및 작용

[0022] 본 발명의 방법 및 시스템은 오디오 캡셔닝(audio captioning)은 여전히 진행하면서 부분적 의미 구문분석을 동적으로 보고하는 음성 입력 모드를 제공한다. 상기 의미 구문분석(semantic parse)은 상기 사용자에게 즉시 보고되는 출력을 사용하여 평가될 수 있다.

[0023] 일 양태에서, 음성 이해 시스템(speech understanding system)은 N-그램(N-gram) 언어 모델과 문맥 무관 문법(context-free grammar) 언어 모델의 조합을 포함한다. 상기 언어 모델은 인식될 단어와 의미 정보에 관련된 정보를 저장한다. 모듈은 사용자로부터 입력을 수신하고 처리를 위해 상기 입력을 캡쳐하도록 구성되어 있다. 상기 모듈은 또한 상기 입력의 인식에 관련된 음성 응용 언어 태그(SALT) 애플리케이션 프로그램을 수신하도록 구성되어 있다. 상기 모듈은 상기 모듈은 상기 음성 응용 언어 태그(SALT) 애플리케이션 프로그램 인터페이스와 상기 입력을 처리하여 상기 입력의 제1 부분(first portion)에 관련된 의미 정보를 확인하고 상기 언어 모델을 액세스함으로써 상기 제1 부분에 대한 텍스트와 의미 정보를 포함하는 의미 객체를 출력하고, 상기 인식의 실행과 상기 의미 객체의 출력은 다음 입력 부분에 대하여 계속 캡쳐하면서 수행되도록 구성된다.

[0024] <실시예>

[0025] 도 1은 음성 입력에 기초한 데이터를 렌더링하는 데이터 프리젠테이션 시스템(10)의 블록도이다. 시스템(10)은 음성 인터페이스 모듈(12), 음성 인식 및 이해 모듈(14) 및 데이터 렌더링 모듈(16)을 포함한다. 사용자는 음성 인터페이스 모듈(12)에 음성 질의의 형태로 입력을 제공한다. 음성 인터페이스 모듈(12)은 사용자로부터 음성 정보를 수집하여 이를 나타내는 신호를 제공한다. 입력 음성이 음성 인터페이스 모듈(12)에 의해 수집된 후, 음성 인식 및 이해 모듈(14)이 음성 인식기를 사용하여 음성을 인식하고, 또한, 음성 이해를 수행하는데, 여기서, 본 발명의 일 양태는 오디오 음성 입력이 여전히 캡쳐되는 동안, 그때까지 수신된 입력의 부분적 의미 구문분석을 제공한다.

[0026] 통상, 확인된 의미 정보 뿐만 아니라 수신된 입력에 대한 텍스트(또는 입력의 텍스트를 나타내는 다른 데이터)도 포함하는 부분적 의미 구문분석으로서, 다수의 서로 다른 형태를 가질 수 있는 애플리케이션 모듈(16)에 제공된다. 예를 들어, 일 실시예에서 애플리케이션(16)은 이메일을 전송, 수신 및 응답하고 일정을 조정하는 등

에 사용되는 개인 정보 관리자일 수 있다. 이러한 방식으로, 사용자는 이들 작업을 수행하기 위한 가청 명령(audible command)을 제공할 수 있다. 그러나, 보다 중요한 것은, 애플리케이션 모듈(16)이 대화형 피드백을 제공하고 및/또는 부분적 의미 구문분석 정보의 수신시에 이에 대한 동작을 수행함으로써, 애플리케이션 모듈(16)에 대한 매우 고도의 대화형 인터페이스를 사용자에게 제공할 수 있다. 예를 들어, 음성 전용 동작 모드에서, 물론 애플리케이션에 관련된 다른 작업을 수행할 수 있으면서, 출력(20)은 사용자에게 들릴 수 있는 말(statement)을 포함할 수 있다. 부분적 의미 구문분석 또는 의미 객체(partial semantic parses or semantic objects)는 애플리케이션 내의 대화 논리(dialog logic)를 실행하는데 사용될 수 있다. 예를 들어, 대화 논리는 하나 이상의 의미 객체에 기반하여 사용자에게 하나의 옵션 또는 복수개 또는 일련의 옵션을 제공할 수 있다.

[0027] 이는 시스템(10)이 부분적인 발언(partial utterance)에 기초하여, 즉, 사용자 순번이 끝나기 전에, 출력을 보고할 수 있게 한다. 다시 말하면, 이면(back channel) 통신을 사용하여 시스템 순번과 관련된 작업을 정상적으로 보고 및 수행함으로써, 사용자와 시스템의 순번에 대한 정의가 불명확해진다. 대부분은 종래의 대화 연구에서, 특히, 사람들 간의 대화에 기초한 것에는, 이면 통신을 긍정, 부정 또는 중립 응답과 같은 단순 신호만을 전달하는 비침입적 피드백(non-intrusive feedback)으로 종종 판단하였다. 그러나, 출력(20)에 의해 제공된 피드백은 진행중인 사용자 발언에 대하여 다소 침입적일 수 있도록 보다 많은 정보를 잠재적으로 전송할 수 있게 되어, 사용자가 사용자의 의도 또는 지시를 명확하게 하거나 하지 않게 할 수 있다. 그럼에도 불구하고, 이러한 접근 방식은 사용자와 시스템(10) 간의 보다 현실적인 사람들 간의 대화를 제공하여, 많은 경우 성가시다기보다는 사용자에게 보다 편안하고 사용자의 요구가 충족될 수 있다는 확신을 심어주게 된다.

[0028] 이 점에서, 본 발명은 음성 전용 운영 환경에 국한되는 것이 아니라, 부분적 의미 구문분석 또는 객체(partial semantic parses or objects)의 처리에 기초하여 사용자에게 다른 형태의 피드백을 제공할 수 있다. 예를 들어, 애플리케이션 모듈(16)이 이메일 작업을 수행하는 상술한 애플리케이션에서, 출력(20)은 "이메일 전송 Bob에게"를 포함하는 사용자로부터의 인터럽트되지 않은 명령에서 "이메일 전송"과 같은 어구(phrase)의 수신에만 기초하여 이메일 모듈을 활성화하는 등의 시각적 피드백(visual feedback)을 포함하며, 여기서, "Bob 에게"라는 어구의 처리는 애플리케이션 모듈이 데이터 저장매체(18)에서 추가 정보를 액세스하여 "Bob"이라는 이름을 갖는 사람의 리스트를 렌더링하게 한다. 이 리스트를 보자마자, 사용자는 의도된 수신자가 "Bob Green"임을 용이하게 식별는데, 시스템이 부분적 발언 "Bob Green"에 대한 다른 의미 객체를 제공하여 선택될 수 있으며, 애플리케이션에 의한 수신시에, "Bob Green"이 선택될 수 있도록 처리된다.

[0029] 상술한 바와 같이, 애플리케이션 모듈(16)은 후술하는 본 발명의 양태들이 이로울 수 있는 많은 형태를 취할 수 있다. 제한없이, 애플리케이션 모듈(16)은 또한 사용자의 구두 입력(spoken input)의 텍스트 출력을 제공하는 구술 모듈(dictation module)일 수 있다. 그러나, 부분적 입력 뿐만 아니라 입력의 어구에 대한 의미 정보를 처리함으로써, 보다 정확한 복사본(transcription)이 획득될 수 있다.

[0030] 이상, 음성 명령을 포함하는 사용자로부터의 입력에 대하여 설명하였지만, 본 발명의 양태들은 수기, DTMF, 몸동작 또는 시각 표시 등의 다른 형태의 입력에 또한 적용될 수 있다.

[0031] 상기 넓은 응용 가능성의 부분적 의미 구문분석 또는 객체를 처리하는 경우, 상술한 시스템(10)에서 동작할 수 있는 일반적인 컴퓨팅 장치를 설명하는 것이 유용할 수 있다. 당업자가 이해할 수 있는 바와 같이, 시스템(10)의 콤포넌트는 네트워크 접속 및 프로토콜을 사용하여 분산 컴퓨팅 환경에 걸쳐 분산되거나 단일 컴퓨터 내에 위치할 수 있다.

[0032] 도 2를 이제 참조하면, 데이터 관리 장치(PIM, PDA 등)와 같은 이동 장치의 예시적인 형태가 30에 예시되어 있다. 그러나, 본 발명은 후술하는 다른 컴퓨팅 장치를 사용하여 또한 실시될 수 있다. 예를 들어, 전화기 및/또는 데이터 관리 장치는 또한 본 발명으로부터 이익을 얻을 수 있다. 이들 장치는 기존 휴대용 개인 정보 관리 장치 및 기타 휴대 전자 장치에 비해 개선된 유용성을 가질 수 있다.

[0033] 도 2에 예시되어 있는 데이터 관리 이동 장치(30)의 예시적인 형태에서, 이동 장치(30)는 하우징(32)을 포함하고, 스타일러스(33)와 함께 접촉형 디스플레이 화면을 사용하는 디스플레이(34)를 포함하는 사용자 인터페이스를 가질 수 있다. 스타일러스(33)는 지정된 좌표에서 디스플레이(34)를 누르거나 접촉하는데 사용되어, 필드를 선택하고, 커서의 시작 위치를 선택적으로 이동하거나 몸동작 또는 수기 등을 통해 명령 정보를 제공한다. 대안으로 또는 추가적으로, 하나 이상의 버튼(35)이 네비게이션을 위해 장치(30) 상에 포함될 수 있다. 또한, 회전가능 휠, 롤러 등과 같은 다른 입력 메터니즘이 제공될 수 있다. 그러나, 본 발명은 이러한 형태들의 입력 메커니즘에 국한시키려는 것이 아님이 이해되어져야 한다. 예를 들어, 다른 형태의 입력이 컴퓨터 비전

(computer vision) 등에 의하는 것과 같은 시각적 입력을 포함할 수 있다.

[0034] 도 3을 이제 참조하면, 블록도는 이동 장치(30)를 포함하는 기능 콤포넌트를 예시한다. 중앙 처리부(CPU; 50)는 소프트웨어 제어 평션을 구현한다. CPU(50)가 디스플레이(34)에 접속됨으로써 제어 소프트웨어에 따라 생성된 텍스트와 그래픽 아이콘이 디스플레이(34) 상에 나타나게 된다. 스피커(43)는 가청 출력을 제공하기 위해 디지털 아날로그 컨버터(59)를 통상 사용하여 CPU(50)에 바인딩될 수 있다. 사용자에 의해 이동 장치(30)에 입력되거나 다운로드되는 데이터는 CPU(50)에 양방향으로 바인딩된 비휘발성 판독/기입 랜덤 액세스 메모리 저장(54) 내에 저장된다. 랜덤 액세스 메모리(RAM; 54)는 CPU(50)에 의해 실행되는 명령에 대한 휘발성 스토리지, 및 레지스터 값과 같은 임시 데이터를 저장하는 스토리지를 제공한다. 구성 옵션 및 기타 변수에 대한 기준값은 판독 전용 메모리(ROM; 58)에 저장된다. ROM(58)은 또한 이동 장치의 기본 기능(30) 및 다른 운영 체체 커널 평션(예를 들어, 소프트웨어 콤포넌트의 RAM(54)으로의 로딩)을 제어하는 장치에 대한 운영 시스템 소프트웨어를 저장하는데 사용될 수 있다.

[0035] RAM(54)은 또한 애플리케이션 프로그램을 저장하는데 사용되는 PC 상의 하드 드라이브의 기능과 유사한 방식으로 코드에 대한 스토리지로서 역할할 수 있다. 코드를 저장하는데 비휘발성 메모리가 사용되지만, 그와 달리 코드의 실행에 사용되지 않는 휘발성 메모리에도 저장될 수도 있음이 인식되어져야 한다.

[0036] 무선 신호는 CPU(50)에 바인딩되어 있는 무선 트랜시버(52)를 통해 이동 장치에 의해 전송/수신될 수 있다. 선택적 통신 인터페이스(60)는 컴퓨터(예를 들어, 데스크탑 컴퓨터)로부터 또는 원하는 경우 유선 네트워크로부터 직접 데이터를 다운로드받기 위해 제공될 수 있다. 따라서, 인터페이스(60)는 예를 들어, 적외선 링크, 모뎀, 네트워크 카드 등의 여러 형태의 통신 장치를 포함할 수 있다.

[0037] 이동 장치(30)는 마이크로폰(29), 및 아날로그 디지털(A/D) 컨버터(37) 및 저장매체(54)에 저장되어 있는 선택적 인식 프로그램(음성, DTMF, 수기, 몸동작 또는 컴퓨터 비전)을 포함한다. 예를 들면, 장치(30)의 사용자로부터의 가청 정보, 지시 또는 명령에 응답하여, 마이크로폰(29)은 A/D 컨버터(37)에 의해 디지털화된 음성 신호를 제공한다. 음성 인식 프로그램은 디지털화된 음성 신호에 대한 정규화 및/또는 특징 추출을 수행하여 중간의 음성 인식 결과를 획득한다. 무선 트랜시버(52) 또는 통신 인터페이스(60)를 사용하여, 음성 데이터는 후술되고 도 4의 아키텍처에 예시되어 있는 원격 인식 서버(204)에 전송될 수 있다. 인식 결과는 그 후 이동 장치(30)에 그에 대한 렌덩링을 위해(예를 들어, 시각 및/또는 가청) 리턴되어 종국적으로 웹서버(202; 도 6)에 전송될 수 있으며, 여기서, 웹서버(202)와 이동 장치(30)는 클라이언트/서버의 관계로서 동작한다.

[0038] 유사한 처리가 다른 형태의 입력에 대하여 사용될 수 있다. 예를 들어, 수기 입력(handwriting input)은 장치(30) 상의 전처리를 하거나 또는 전처리 없이 디지털화될 수 있다. 음성 데이터와 유사하게, 이러한 형태의 입력은 인식을 위해 인식 서버(204)에 전송될 수 있으며, 여기서, 인식 결과는 장치(30) 및/또는 웹 서버(202) 중의 적어도 하나에 리턴된다. 유사하게, DTMF 데이터, 몸동작 데이터 및 시각 데이터는 유사하게 처리될 수 있다. 입력 형태에 따라, 장치(30) (및 후술하는 다른 형태의 클라이언트)는 시각적 입력을 위해 카메라와 같은 필요한 하드웨어를 포함할 수 있다.

[0039] 도 4는 휴대 전화(80)의 일 실시예의 평면도이다. 전화기(80)는 디스플레이(82)와 키패드(84)를 포함한다. 통상, 도 3의 블록도는 도 4의 전화기에 적용되지만, 다른 평션을 수행하는데 필요한 추가적인 회로가 요구될 수 있다. 예를 들어, 전화기로서 동작하는데 요구되는 트랜시버가 도 3의 실시예에 대하여 요구될 수 있지만, 이러한 회로는 본 발명과는 관련되어 있지 않다.

[0040] 상술한 휴대용 또는 이동 컴퓨팅 장치 뿐만 아니라, 본 발명은 범용 데스크탑 컴퓨터와 같은 수많은 다른 컴퓨팅 장치를 가지고 사용될 수 있음이 이해되어져야 한다. 예를 들어, 본 발명은 장애를 갖고 있는 사용자가 영숫자(alpha-numeric) 키보드와 같은 다른 종래의 입력 장치를 작동하기에 매우 어려울 경우에 컴퓨터 또는 다른 컴퓨팅 장치에 텍스트를 입력 또는 엔터할 수 있게 할 것이다.

[0041] 또한, 본 발명은 다수의 다른 범용 또는 특정 목적의 컴퓨팅 시스템, 환경 또는 구성을 사용하여 동작할 수 있다. 공지의 컴퓨팅 시스템, 환경 및/또는 본 발명에 사용하기에 적합(suitable)할 수 있는 구성의 예들은 일반적인 전화기(화면이 없는 것), 개인용 컴퓨터, 서버 컴퓨터, 핸드헬드 또는 랩탑 장치, 태블릿 컴퓨터, 멀티프로세서 시스템, 마이크로프로세서 기반 시스템, 셋탑 박스, 프로그래머블 소비자 전자제품, 및 상기 시스템 또는 장치 중의 임의의 것을 포함하는 분산 컴퓨팅 환경을 포함하지만 이에 국한된 것은 아니다.

[0042] 이하, 도 5에 예시된 범용 컴퓨터(120)을 간략하게 설명한다. 그러나, 컴퓨터(120)는 적합한 컴퓨팅 환경의 단지 일 예이며, 본 발명의 기능의 범위에 대한 어떠한 제한도 암시하려는 것은 아니다. 또한, 컴퓨터(120)는 여

기서 설명된 콤포넌트 중의 임의의 하나 또는 그 조합에 관한 임의의 의존성 또는 요건을 갖는 것으로 파악되어서는 안된다.

[0043] 본 발명은 컴퓨터에 의해 실행되는 프로그램 모듈과 같은 컴퓨터 실행가능 명령의 경우에 대하여 설명할 것이다. 일반적으로, 프로그램 모듈은 특정 작업을 수행하거나 특정 추상 데이터형을 구현하는 루틴, 프로그램, 오브젝트, 콤포넌트, 데이터 구조를 포함한다. 또한, 본 발명은 통신 네트워크를 통해 연결된 원격 처리 장치에 의해 작업이 수행되는 분산 컴퓨팅 환경에서 실시될 수 있다. 분산 컴퓨팅 환경에서, 프로그램 모듈은 메모리 스토리지 장치를 포함하는 로컬 및 원격 컴퓨터 스토리지 매체에 위치할 수 있다. 프로그램 및 모듈에 의해 수행되는 작업은 도면을 참조하여 이하 설명한다. 당업자는 임의 형태의 컴퓨터 판독가능 매체 상에 기재될 수 있는 프로세서 실행가능 명령으로 설명 및 도면을 구현할 수 있다.

[0044] 도 5를 참조하면, 컴퓨터(120)의 콤포넌트는 처리부(140), 시스템 메모리(150), 및 시스템 메모리를 포함하는 여러 시스템 콤포넌트를 처리부(140)에 바인딩시키는 시스템 버스(141)를 포함할 수 있지만 이에 국한되지는 않는다. 시스템 버스(141)는 메모리 버스 또는 메모리 컨트롤러, 병렬 버스 및 다양한 버스 구조 중의 임의의 것을 사용하는 로컬 버스를 포함하는 여러 유형의 버스 구조 중의 임의의 것일 수 있다. 예를 들면 - 한정이 아님 -, 이러한 아키텍쳐는 산업 표준 아키텍쳐 (ISA) 버스, 범용 직렬 버스(USB), 마이크로 채널 아키텍쳐(MCA) 버스, 개선된 ISA(EISA) 버스, 비디오 전자 표준 협회(VESA) 로컬 버스, 및 메자닌(Mezzanine) 버스로도 불리는 주변 콤포넌트 상호접속(PCI) 버스를 포함한다. 컴퓨터(120)는 통상 다양한 컴퓨터 판독가능 매체를 포함한다. 컴퓨터 판독가능 매체는 컴퓨터(120)에 의해 액세스될 수 있는 임의의 입수 가능 매체일 수 있으며, 휘발성 및 비휘발성, 분리형 및 비분리형 매체를 모두 포함하는 임의의 입수가능 매체일 수 있다. 예를 들면 - 한정이 아님 -, 컴퓨터 판독가능 매체는 컴퓨터 스토리지 매체 및 통신 매체를 포함할 수 있다. 컴퓨터 스토리지 매체는 컴퓨터 판독가능 명령, 데이터 구조, 프로그램 모듈 또는 기타 데이터와 같은 정보의 저장을 위한 임의의 방법 또는 기술로 구현되는 휘발성 및 비휘발성, 분리형 및 비분리형 매체를 포함한다. 컴퓨터 저장 매체는 RAM, ROM, EEPROM, 플래시 메모리 또는 다른 메모리 기술, CD-ROM, 디지털 다기능 디스크(DVD) 또는 다른 광디스크 스토리지, 자기 카세트, 자기 테이프, 자기 디스크 스토리지 또는 다른 자기 스토리지 장치, 도는 원하는 정보를 저장하는데 사용될 수 있고 컴퓨터(120)에 의해 액세스될 수 있는 임의의 다른 매체를 포함하지만 이에 국한되지 않는다.

[0045] 통신 매체는 통상 컴퓨터 판독가능 명령, 데이터 구조, 프로그램 모듈 또는 반송파 혹은 다른 전송 메커니즘 등의 변조된 데이터 신호에서의 다른 신호를 구체화하며 임의의 정보 전달 매체를 포함한다. "변조된 데이터 신호"라는 용어는 신호 내의 정보를 암호화하기 위한 방식으로 설정 또는 변경되는 하나 이상의 특성을 갖는 신호를 의미한다. 예를 들어 - 한정이 아님 -, 통신 매체는 유선 네트워크 또는 다이렉트 유선 접속과 같은 유선 매체, 그리고 어쿠스틱, FR, 적외선 및 기타 무선 매체와 같은 무선 매체를 포함한다. 상술한 것의 임의의 조합은 컴퓨터 판독가능 매체의 범주에 또한 포함될 수 있다.

[0046] 시스템 메모리(150)는 판독 전용 메모리(ROM; 151) 및 랜덤 액세스 메모리(RAM; 152)와 같은 휘발성 및/또는 비휘발성 메모리의 형태의 컴퓨터 스토리지 매체를 포함한다. 시동 중과 같이 컴퓨터(120) 내 요소들 간의 정보를 전송할 수 있게 하는 기본 루틴을 포함하는 기본 입출력 시스템(153; BIOS)은 통상 ROM(151) 내에 저장되어 있다. RAM(152)은 통상 즉시 액세스가능하고 및/또는 처리부(140)에 의해 현재 동작되는 데이터 및/또는 프로그램 모듈을 포함한다. 예를 들어 - 한정이 아님 -, 도 5는 운영 체제(154), 애플리케이션 프로그램(155), 다른 프로그램 모듈(156), 및 프로그램 데이터(157)를 예시하고 있다.

[0047] 컴퓨터(120)는 다른 분리형/비분리형 휘발성/비휘발성 컴퓨터 스토리지 매체를 또한 포함할 수 있다. 단지 예를 들면, 도 5는 비분리형 비휘발성 자기 매체로부터 판독하거나 이에 기입하는 하드 디스크 드라이브(161), 분리형 비휘발성 자기 디스크(172)로부터 판독하거나 이에 기입하는 자기 디스크 드라이브(171), 및 CD ROM 또는 다른 광매체와 같이 분리형 비휘발성 광디스크(176)로부터 판독하거나 이에 기입하는 광디스크 드라이브(175)를 예시하고 있다. 예시적인 운영 환경에서 사용될 수 있는 다른 분리형/비분리형, 휘발성/광휘발성 컴퓨터 스토리지 매체는 자기 테이프 카세트, 플래시 메모리 카드, 디지털 다기능 디스크, 디지털 비디오 테이프, 고체상태 RAM, 고체상태 ROM 등을 포함하며 이에 국한되지 않는다. 하드 디스크 드라이브(161)는 인터페이스(160)와 같은 비분리형 메모리 인터페이스를 통해 시스템 버스(141)에 통상 접속되며, 자기 디스크 드라이브(171)와 광디스크 드라이브(175)는 인터페이스(170)와 같은 분리형 메모리 인터페이스에 의해 시스템 버스(141)에 통상 접속된다.

[0048] 위에서 설명되고 도 5에서 예시되어 있는 드라이브 및 관련 컴퓨터 스토리지 매체는 컴퓨터(120)에 컴퓨터 판독

가능 명령, 데이터 구조, 프로그램 모듈, 및 기타 데이터를 제공한다. 도 5에서, 예를 들면, 하드 디스크 드라이브(161)는 운영 체제(164), 애플리케이션 프로그램(165), 다른 프로그램 모듈(166), 및 프로그램 데이터(167)를 저장하는 것으로 예시되어 있다. 이들 콤포넌트는 운영 체제(154), 애플리케이션 프로그램(155), 다른 프로그램 모듈(156), 및 프로그램 데이터(157)와 동일 또는 상이할 수 있다. 운영 체제(164), 애플리케이션 프로그램(165), 다른 프로그램 모듈(166), 및 프로그램 데이터(167)는 적어도 그들이 서로 다른 복사본임을 나타내기 위해 여기서 서로 다른 번호들이 부여된다.

[0049] 사용자는 키보드(182), 마이크로폰(183), 및 마우스, 트랙볼 또는 터치패드와 같은 포인팅 장치(181) 등의 입력 장치를 통해 컴퓨터(120)에 명령과 정보를 입력한다. 다른 입력 장치(미도시)는 조이스틱, 게임 패드, 위성 접시, 스캐너 등을 포함할 수 있다. 이들 및 다른 입력 장치는 시스템 버스에 바인딩되어 있는 사용자 입력 인터페이스(180)를 통해 처리부(140)에 종종 접속되지만, 병렬 포트, 게임 포트 또는 범용 직렬 버스(USB)와 같은 다른 인터페이스와 버스 구조에 의해 접속될 수도 있다. 모니터(184) 또는 다른 유형의 디스플레이 장치는 비디오 인터페이스(185)와 같은 인터페이스를 통해 시스템 버스(185)에 또한 접속된다. 모니터 뿐만 아니라, 컴퓨터도 또한 출력 병렬 인터페이스(188)에 접속될 수 있는 스피커(187) 및 프린터(186)와 같은 다른 주변 출력 장치를 포함할 수 있다.

[0050] 컴퓨터(120)는 원격 컴퓨터(194)와 같은 하나 이상의 원격 컴퓨터에 대한 논리적 접속을 사용하여 네트워크화된 환경에서 동작할 수 있다. 이 원격 컴퓨터(194)는 개인용 컴퓨터, 핸드헬드 장치, 서버, 라우터, 네트워크 PC, 피어 장치 또는 다른 공통 네트워크 노드일 수 있으며, 통상 컴퓨터(120)에 대하여 상술한 요소의 다수 또는 모두를 포함한다. 도 5에 도시되어 있는 논리적 접속은 근거리 네트워크(LAN; 191) 및 원거리 네트워크(WAN; 193)를 포함하지만 다른 네트워크를 포함할 수도 있다. 이러한 네트워킹 환경은 사무실, 범사내망, 인트라넷 및 인터넷에서는 흔한 것이다.

[0051] LAN 네트워킹 환경에서 사용되는 경우, 컴퓨터(120)는 네트워크 인터페이스 또는 어댑터(190)를 통해 LAN(191)에 접속된다. WAN 네트워킹 환경에서 사용되는 경우, 컴퓨터(120)는 통상 모뎀(192) 또는 인터넷과 같은 WAN(193)을 통해 통신을 설정하는 다른 수단을 포함한다. 모뎀(192)은 내장형 또는 외장형일 수 있으며, 사용자 입력 인터페이스(180) 또는 다른 적절한 메커니즘을 통해 시스템 버스(141)에 접속될 수 있다. 네트워크화된 환경에서, 컴퓨터(120) 또는 그 일부에 대하여 도시된 프로그램 모듈은 원격 메모리 스토리지 장치에 저장될 수 있다. 예를 들면 - 한정이 아님 -, 도 5는 원격 컴퓨터(194)에 상주하는 것으로 원격 애플리케이션 프로그램(195)을 예시하고 있다. 도시된 네트워크 접속은 예시적이며 컴퓨터들 간의 통신 링크를 설정하는 다른 수단이 사용될 수 있음이 이해될 것이다.

[0052] 도 6은 본 발명의 예시적인 환경인 웹기반 인식 및 데이터 렌더링을 위한 아키텍처(200)을 예시한다. 일반적으로, 웹서버(202)에 저장된 정보는 이동 장치(30) 또는 컴퓨터(120)(여기서, 입력 형태에 따라 요구되는 디스플레이 스크린, 마이크로폰, 카메라, 터치형 패널 등을 구비한 다른 형태의 컴퓨팅 장치를 나타냄)와 같은 클라이언트(100)를 통해, 또는 정보가 들릴 수 있게 요구되는 전화기(80)를 통해, 또는 눌려진 키에 응답하여 폰(80)에 의해 생성된 톤을 통해 액세스될 수 있으며, 여기서 웹서버(202)로부터의 정보는 이 사용자에게만 들릴 수게 제공된다.

[0053] 이 실시예에서, 정보가 음성 인식을 사용하여 클라이언트(100)를 통해서든 또는 전화기(80)를 통해서든 획득되어, 단일 인식 서버(204)는 어떤 동작모드이든 지원할 수 있다는 점에서 아키텍처(200)는 통합된다. 또한, 아키텍처(200)는 공지의 마크업(markup) 언어의 확장(extension)을 사용하여 동작한다(예를 들어, HTML, XHTML, cHTML, XML, WML 등). 따라서, 웹서버(202) 상에 저장된 정보는 이들 마크업 언어에서 발견되는 공지의 GUI 방법을 사용하여 또한 액세스될 수 있다. 공지의 마크업 언어의 확장을 사용하여, 웹서버(202) 상에서의 저작(authoring)은 보다 용이하며, 현재 존재하는 레거시 애플리케이션은 음성 인식을 포함하도록 용이하게 변형될 수 있다.

[0054] 통상, 클라이언트(100)는 브라우저를 사용하여 웹서버(202)에 의해 제공되는 206에서 일반적으로 나타낸 HTML 페이지, 스크립트 등을 실행한다. 음성 인식이 요구되는 경우, 예를 들어, 디지털화된 오디오 신호일 수 있는 음성 데이터 또는 상술한 바와 같이 클라이언트(100)에 의해 오디오 신호가 전처리되는 음성 특징은 음성 인식 동안 사용하기 위해 클라이언트(100)에 의해 제공될 수 있는 문법 또는 언어 모델(220)의 지시로 인식 서버(204)에 제공된다. 다르게는, 음성 서버(204)는 언어 모델(220)을 포함할 수 있다. 인식 서버(204)의 구현은 많은 형태를 취할 수 있지만 - 이 중 하나가 설명됨 -, 통상 인식기(211)를 포함한다. 인식의 결과는 원하거나 적절한 경우 로컬 렌더링을 위해 클라이언트(100)에 다시 제공된다. 원하는 경우, 텍스트 대 음성 모듈(text-

to-speech module; 222)은 구두 텍스트를 클라이언트(100)에 제공하는데 사용될 수 있다. 사용된다면, 인식 및 임의의 그래픽 사용자 인터페이스에 의한 정보를 컴파일할 때, 필요하다면, 클라이언트(100)는 추가 처리 및 추가 HTML 페이지/스크립트의 수신을 위해 웹 서버(202)에 정보를 전송한다.

[0055] 도 6에 나타낸 바와 같이, 클라이언트(100), 웹 서버(202) 및 인식 서버(204)는 네트워크(205) - 여기서는, 인터넷과 같은 광역 네트워크 - 를 통해 공통 접속되고, 개별 어드레싱가능하다. 따라서, 이들 장치중 임의의 것이 서로 물리적으로 인접 배치될 필요는 없다. 특히, 웹 서버(202)가 인식 서버(204)를 포함할 필요는 없다. 이러한 방식으로, 웹 서버(202)에서의 저작은 저작자가 인식 서버(204)의 복잡함을 인식할 필요없이 의도되는 애플리케이션에 집중할 수 있다. 그 대신, 인식 서버(204)는 네트워크에 독립적으로 설계되고 접속되어, 그에 따라 웹 서버(202)에서 요구되는 추가 변경없이 갱신되고 개선될 수 있다. 웹 서버(202)는 또한 클라이언트측 마크업 및 스크립트를 동적으로 생성할 수 있는 저작을 포함할 수 있다. 다른 실시예에서, 웹 서버(202), 인식 서버(204) 및 클라이언트(100)는 구현 머신의 성능에 의존하여 바인딩될 수 있다. 예를 들어, 클라이언트(100)가 범용 컴퓨터, 예를 들면, 개인용 컴퓨터를 포함하면, 클라이언트는 인식 서버(204)를 포함할 수 있다. 유사하게, 원하는 경우, 웹 서버(202) 및 인식 서버(204)는 단일 머신에 통합될 수 있다.

[0056] 전화기(80)에 의한 웹 서버(202)로의 액세스는 전화기(800)의 유선 또는 무선 전화망(208)으로의 접속을 포함하며, 그에 따라, 전화기(800)를 제3자 게이트웨이(210)에 접속하는 것도 포함한다. 게이트웨이(210)는 전화기(80)를 전화 음성 브라우저(212)에 접속시킨다. 전화 음성 브라우저(212)는 전화 인터페이스를 제공하는 매체 서버(214)와 음성 브라우저(216)를 포함한다. 클라이언트(100)와 같이, 또는 웹 서버(202)에서와 같이, 전화 음성 브라우저(212)는 HTML 페이지/스크립트를 수신한다. 일 실시예에서, HTML 페이지/스크립트는 클라이언트(100)에 제공되는 HTML 페이지/스크립트와 유사한 형태이다. 이러한 방식으로, 웹 서버(202)는 클라이언트(100)와 전화기(80)를 별도로 지원할 필요는 없거나 또는 심지어 표준 GUI 클라이언트를 별도로 지원할 필요가 없다. 그 대신, 일반 마크업 언어가 사용될 수 있다. 또한, 클라이언트(100)와 같이, 전화기(800)에 의해 전송되는 가청 신호로부터의 음성 인식이 네트워크(205)를 통해 또는 전용선(207), 예를 들어, TCP/IP를 통해 음성 브라우저(216)로부터 인식 서버(204)에 제공된다. 웹 서버(202), 인식 서버(204), 및 전화 음성 브라우저(212)는 도 5에 나타낸 범용 데스크탑 컴퓨터와 같은 임의의 적절한 컴퓨팅 환경에서 구현될 수 있다.

[0057] 시스템(10)에서 동작하는 여러 환경 및 계층 구조를 기술함으로써, 시스템(10)의 다양한 콤포넌트와 펑션의 상세한 설명이 제공된다. 도 7은 음성 인식기 및 이해 모듈(14)의 블록도를 나타낸다. 음성 인터페이스 모듈(12)로부터 수신되는 입력 음성은 음성 인식 및 이해 모듈(14)에 전송된다. 음성 인식 및 이해 모듈(14)은 관련 언어 모델(310)을 구비하는 인식 엔진(306)을 포함한다. 인식 엔진(306)은 언어 모델(310)을 사용하여 입력을 구성하는 각각의 어구를 나타내도록 가능한 표면 의미 구조를 식별하여, 입력이 수신됨에 따라 부분적 의미 구문분석 또는 객체를 제공한다. 사용자가 발언을 끝낸 것을 후에야 수신된 입력의 완성을 처리하는 시스템과는 달리, 모듈(14)은 그때까지 수신된 것에만 기초하여 의미 객체를 연속적으로 제공한다.

[0058] 인식 엔진(306)은 부분적 발언에 기초하여 적어도 하나의 의미 출력 객체를 제공한다. 몇몇 실시예에서, 인식 엔진(306)은 각 다른 구조에 대한 하나 이상의 다른 표면 의미 객체를 제공할 수 있다.

[0059] 도 7에는 음성 입력이 제공되는 것으로 도시되어 있지만, 본 발명은 수기 인식, 몸동작 인식 또는 그래픽 사용자 인터페이스(사용자가 키보드 또는 다른 입력 장치와 상호작용하는 경우)에 사용될 수 있다. 이들 다른 실시예에서, 음성 인식기(306)는 당업계에 공지된 바와 같은 적절한 인식 엔진으로 대체된다. 그래픽 사용자 인터페이스에 대해서는, 문법(언어 모델을 구비)은 입력 상자와 같은 사용자 입력에 관련된다. 따라서, 사용자의 입력은 입력의 상태에 따라 중요한 변경없이 일정한 방식으로 처리된다.

[0060] 부분적 의미 구문분석 또는 객체에 기초하는 시스템(10)에 의한 다른 형태의 정보 피드백을 또한 포함하는 상술한 상호 대화는 SALT(음성 애플리케이션 언어 태그) 또는 기타 음성, 수기 및 해당 애플리케이션 및 의미 객체 동기 디코딩에 대한 선택된 스키마(schema)에 기초하여 의미 정보를 제공할 수 있는 언어 모델 구성을 지원하는 패턴 인식 API(애플리케이션 프로그램 인터페이스)를 사용하여 구현될 수 있다. 음성 응용 언어 태그(SALT)는 예를 들어, 개인용 컴퓨터, 전화기, 태블릿 PC 및 무선 이동 장치로부터 정보, 애플리케이션, 웹 서비스로의 액세스를 가능하게 하는 개발 중인 표준이지만, 또한, 네트워크에 걸쳐 상호접속 없이 또한 애플리케이션 인터페이스에 인가될 수 있다. 음성 응용 언어 태그(SALT)는 HTML, XHTML 및 XML과 같은 기존의 마크업 언어를 확장한다. SALT 1.0 규격은 http://www.SALTforum.org에서 온라인으로 구입할 수 있다. 음성 응용 언어 태그(SALT)는, 예를 들어, 음성 서버(204)로부터의 사용자 입력에 기초하는 의미 정보를 제공할 수 있으며, 이러한 정보는 발언 종료 후에 데이터 렌더링 모듈(16)에 제공되는 객체를 형성한다; 그러나, 아래에서 상세히 설명하

는 바와 같이 음성 응용 언어 태그(SALT)는 종래에는 고찰되지 않은 방식으로 부분적 의미 구문분석 또는 객체를 제공하는데 사용될 수 있다. 음성 응용 언어 태그(SALT) 확장의 사용, 다른 API에서 유사한 확장의 사용은 고도의 대화형 이벤트 구동(event-driven) 사용자 상호작용에 대한 지원을 제공한다.

[0061] 예를 들어, 음성 응용 언어 태그(SALT)를 사용하면, 음성 응용 언어 태그(SALT) 듣기 객체는 음성 인식 및 이해 작업 모두를 수행하는데 사용될 수 있다. 이는 디자인이 음성 이해를 음성 인식과 같이 패턴 인식의 문제로서 처리하는 관점 및 공식을 따르기 때문이다. 이들 양자는 모두 해당 음성 신호와 가장 잘 일치하는 것의 가능한 출력의 수집으로부터의 패턴을 발견하려 한다. 음성 인식에 있어서 발견될 패턴은 단어열(word string)이지만, 이해에 있어서는 의미 객체의 트리이다. 종래의 음성 인식 작업은 가능성 있는 단어열로 이루어진 언어모델을 사용하여 검색 처리를 지시한다. 유사한 방식으로, 음성 인식 작업은 의미 모델을 사용하여 적절한 의미 객체 트리를 만들 수 있도록 동일한 검색 엔진을 유도할 수 있다. 어휘집(lexicon)과 이 어휘집 입력으로부터 어구 분절(phase segment)의 작성 규칙을 포함하는 언어 모델과 같이, 의미 모델은 모든 의미 객체의 사전과 이들의 작성 규칙을 포함한다. 인식 출력은 텍스트열(text string)이지만, 이해 결과는 의미 객체의 트리이다.

[0062] 구조화된 검색 결과를 리턴하기 위해 N-그램을 확장할 수 있지만, 대부분의 통상의 음성 이해 애플리케이션은 확률 문맥 무관 문법(PCFG)에 기초하며, 여기서 설계자는 큰 트리 뱅크 주석의 트레이닝 데이터(massive tree-bank annotated training data)없이도 의미 객체를 생성하는 규칙을 규정할 수 있다. 이러한 규칙들을 규정하는 방법들 중 하나는, 부분적 PCFG 파싱 트리를 의미 객체 트리로 변환하는 방식에 대한 검색 엔진에 대한 생성 지침과 각 PCFG 규칙을 관련시키는 것이다. 마이크로소프트 음성 인식 인터페이스(SAPI)(본 발명에서 또한 사용될 수 있는 음성 API의 일례) 포맷에서 기재된 일례는 다음과 같다.

```
<rule name="nyc">
 <list>
   <phrase>new york ?city</phrase>
   <phrase>?the big apple</phrase>
 </list>
 <output>
  <city_location>
   <city>New York</city>
   <state>New York</state>
   <country>USA</country>
  </city_location>
 </output>
</rule>
...
<rule name="NewMeeting">
 <ruleref min="0"   name="CarrierPhrase"/>
 <ruleref max="inf" name="ApptProperty"/>
 <output>
  <NewMeeting>
   <DateTime>
    <xsl:apply-templates select="//Date"/>
    <xsl:apply-templates select="//Time"/>
    <xsl:apply-templates select="//Duration"/>
   </DateTime>
   <Invitees>
     <xsl:apply-templates select="//Person"/>
   </Invitees>
     ....
  </NewMeeting>
```

[0063]

```
  </output>
 </rule>

 <rule name="ApptProperty"/>
  <list>
    <ruleref name="Date"/>
    <ruleref name="Duration"/>
    <ruleref name="Time"/>
    <ruleref name="Person" max="inf"/>
    <ruleref name="ApptSubject"/>
   .. ..
  </list>
 </rule>
 .. ..
```

[0064]

[0065] 문법 세그먼트는 3개의 규칙을 포함한다. 첫번째는, "nyc"로 지정된 프리 터미널(pre-terminal)은 뉴욕시에 대한 표현을 열거한다. 이 예에서, <output> 태그는 의미 객체를 구성하는 규칙을 포함한다. 이들은 검색 경로가 이를 즉시 진행하는 토큰에 의해 지정되는 문법 노드에 나가는 경우 호출된다. 이러한 경우에, <city_location> 요소로 XML에서 표현된 의미 객체는, 검색 경로가 "nyc" 규칙을 나갈 때 생성된다. 의미 객체는 순번대로 3개의 의미 객체: 도시명, 주 명칭, 국각 명칭 약어를 각각 구성한다.

[0066] 또한, 의미 객체의 구성은 예를 들어, 새로운 모임을 스케쥴하는 동적 프로세스일 수 있다. 예를 들어, Newmeeting 의미 객체는 사용자가 사용자가 날짜, 시간, 기간 및 참석자와 같은 모임 속성의 규정을 종료할 때 생성될 수 있다. 다른 의미 객체를 구성요소로서 Newmeeting 의미 객체에 붙이기하는데 템플릿이 사용될 수 있다. 동일한 원리가 또한 여기에 나타내지 않은 다른 규칙에 적용될 수 있다. 예를 들면, "schedule a meeting with Li Deng and Alex Acero on January first for one hour"라는 발언은 다음의 의미 객체가 될 것이다.

```
<NewMeeting>
 <DateTime>
   <Date>01/01/2003</Date>
   <Duration>3600</Duration>
 </DateTime>
 <Invitees>
   <Person>Li Deng</Person>
   <Person>Alex Acero</Person>
 </Invitees>
</NewMeeting>
```

[0067]

[0068] 실제 애플리케이션에서, PCFG 커버리지(coverage)의 개선은 힘든 작업이다. 따라서, 무엇보다도 주요 의미 정보를 포함하지는 않지만 구문 구조(예를 들어, "May I...", "Could you show me...", "Please show me...")에서의 상당한 변화를 갖는 기능성 어구(functioal phrase)를 모델링하는데 N-그램(N-gram)을 사용할 수 있는 것이 바람직하다. 일 실시예에서, 언어 모델(310)은 PCFG를 N-그램과 바인딩시키는 의미 언어 모델(semantic language model)을 포함한다. 이 기술은 역시 사용될 수 있는 통합 언어 모델(unified language model)과는 약간 상이하다. 통합 언어 모델은 N-그램 내에 개별 토큰(token)으로서 모델링되는, 단지 단어 리스트가 아닌, CFG 단편(fragment)을 허용한다는 점을 제외하면, 종래의 클래스 N-그램에 대한 자연 확장(natural extension)이다. 이러한 모델을 사용하는 인식기(306)는 다음에 파싱되어야 할 텍스트열을 또한 생성한다. 통합 언어 모델은 따라서 텍스트 복사본을 지원하는 특정 언어학적 구조를 통합하도록 설계된다.

[0069] 반면에, 의미 언어 모델은 PCFG에 의해 캡쳐되는 것보다 통상 우수한 의미 구조에 대한 검색을 위해 디코더 또는 인식기를 사용하는 것을 목적으로 한다. 따라서, N-그램으로 CFG 단편을 임베딩(embedding)하는 대신, PCFG는 관심있는 N-그램에 대응하는 특정 PCFG 프리터미널을 생성함으로써 N-그램을 포함하는데 사용된다. 마이크로소프트 SAPI 문법 포맷에서, 이는 다음과 같이 XML <dictation> 태그를 갖는 전처리를 사용하여 나타낼 수 있다.

[0070] LCFG <dictaion max="inf"/> RCFG

[0071] 여기서, LCFG와 RCFG는 임베디드된 N-그램의 좌측 및 우측 내용을 각각 나타낸다. 검색 프로세스는 <dictation> 태그를 토큰으로 처리하여 정상적인 비종단을 입력하는 것과 같이 N-그램에 확장한다. 태그에 대한 max 속성은 N-그램이 처리할 수 있는 소모될 수 있는 최대 단어수를 규정한다. 이러한 N-그램 내에서, 단어열 확률은 PCFG를 갖는 백오프(back-off) N-그램을 삽입함으로써 계산되며, 보다 상세하게는,

[0072] <식 1>

$$P(w_n \mid w_{n-1}, w_{n-2}, \ldots) = \lambda P(w_n \mid Ngram, w_{n-1}, w_{n-2}, \ldots) + (1-\lambda) P(w_n \mid RCFG) P(RCFG \mid w_{n-1}, w_{n-2}, \ldots)$$

[0073]

[0074] 여기서, λ는 N-그램 삽입 가중치(insertion weight)이고, P(RCFG|w_{n-1}, ...)는 N-그램의 백오프 확률을 사용하며, w_n은 단어집에서 있는 것과 같이 취급된다. 일 실시예에서, P(w_n|RCFG)항은 최대 N-그램 단어열 크기에 도달했는지 그리고 단어가 CFG 단편의 커버리지 내에 있는지에 따라 단지 이진값을 가정한다. PCFG에서 도출되는 단어는 보다 높은 확률을 가지므로, 실제 CFG에 의해 커버되는 것에 속하는 경로는 최대 N-그램 단어수가 무한정으로 설정되는 경우에도 그들의 N-그램 대응부분(N-gram counterparts)보다 우선인 경향이 있다. 기능적 어구에 더하여, 임베디드되는 N-그램은 구술와 유사한 속성을 갖는 의미 객체를 모델링하는데 사용될 수 있다. 예를 들어, 모임 주제(subject)는 다음과 같은 작업에서의 모델이다.

[0075] <rule name="ApptSubject">

[0076] <dictation max="inf"/>

[0077] 의미 언어 모델에 대한 세부사항은 2001년, 이탈리아 토렌토, Proc. ASRU-2001에서 K. Wang의 "Semantic modeling for dialog systems in a pattern recognition framework"에 설명되어 있으며, 이는 그 전체가 여기에 참조로서 통합된다.

[0078] 본 발명의 다른 양태는 음성 응용 언어 태그(SALT) 듣기(listen) 객체의 새로운 사용을 포함한다. 음성 응용 언어 태그(SALT)는 일련의 XML 요소에 관련 속성과 DOM 객체 속성, 이벤트 및 방법들을 제공하며, 이들은 소스 마크업 문서에 대하여 사용되고 소스 페이지를 음성 인터페이스에 인가할 수 있다. 통상, 주요 요소는 다음 사항을 포함한다:

[0079] <prompt ...> 음성 합성 구성 및 즉시 재생용

[0080] <listen ...> 음성 인식기 구성, 인식 실행과 후처리, 및 기록용

[0081] <dtmf ...> DTMF 구성 및 제어용

[0082] <smex ...> 플랫폼 콤포넌트와의 범용 통신용

[0083] 또한, 듣기 및 dtmf 객체는 다음의 문법 및 바인딩 컨트롤(binding control)을 포함한다:

[0084] <grammar ...> 입력 문법 리소스의 규정용

[0085] <bind ...> 인식 결과의 처리용

[0086] 듣기 요소는 3개의 인식 모드를 구별하는 "모드" 속성을 포함할 수 있으며, 이는 어떻게 그리고 언제 결과를 리턴할지에 대하여 인식 서버(예를 들어, 204)를 지시한다. 결과의 리턴값은 "onReco" 이벤트를 제공하거나 적절하다면 "bind" 요소를 활성화하는 것을 의미한다.

[0087] 제1 모드인 "자동(automatic)"에서, 애플리케이션이 아닌 음성 플랫폼은 인식 프로세스를 정지할 때를 제어한다. 이 모드는 전화 또는 핸즈프리 시나리오에 대하여 개선되어 왔다. 인식 결과가 이용가능하게 되자마자 및/또는 침묵을 의미하는 기간이 경과하는 경우에, 음성 플랫폼은 인식기를 중단시켜 그 결과를 리턴하므로, 바인딩 요소를 통해 적절한 필드에 관련될 수 있다.

[0088] 제2 모드인 "단일(single)" 동작에서, 인식 결과의 리턴은 명시적인 "중단(stop)" 호의 제어 하에 있다. 이 중단 호출은 사용자에 의한 "펜 업(pen-up)"과 같은 이벤트에 대응하며, 이 모드는 장치가 음성 입력을 허용하는 다중 모드에서 사용되도록 개발되었지만, 예를 들어, 스타일러스(33; 도 1)의 사용을 통해 필드를 언제 그리고

어느 것을 선택하는지가 제어된다.

[0089] 음성 인식기의 제3 동작 모드는, "다중 모드(multiple mode)"이다. 이 동작 모드는 "오픈 마이크로폰(open microphone)"에 대하여 또는 구술 시나리오에서 사용된다. 통상, 이 동작모드에서, 인식 결과는 명시적 중단 호가 수신되거나 또는 미인식 입력에 관련된 기간 또는 듣기에 대한 최대 시간이 초과될 때까지의 구간에 리턴된다. 통상, 이 동작 모드에서, 인식되는 각 어구에 대하여, "onReco" 이벤트가 발행되어 중단() 호가 수신될 때까지 그 결과는 리턴된다.

[0090] 그러나, 이 동작모드는 본 발명의 다른 양태와 같이, 검색 엔진이 사용자에게 보다 상호작용적인 성능을 제공할 수 있도록 그들을 현저한 언어적 경계표(salient linguistic landmark)가 도래할 때마다 즉시 보고할 수 있게 한다. 시간 동기 디코딩(time synchronous decoding)에 기반한 검색 알고리즘은 공지되어 있으며, 직접적 방식으로 이러한 모드에 대하여 채용될 수 있다. 이러한 알고리즘 중의 하나는 1999년 IEEE Signal Processing Magazine, pp. 64-83에 개시된 H. Ney와 S. Ortmanns에 의한 "Dynamic programming search for continuous speech recognition"에 기재되어 있다. 음성 인식에 있어서, 언어학 경계표는 단어 또는 어구 경계(boundary)에 항상 대응한다. 따라서, 음성 응용 언어 태그(SALT) 다중 모드 인식은 다수의 상용 구술 소프트웨어에서 흔히 볼 수 있는 UI 효과인 단어열 가정(word string hypotheses)을 가능한한 빨리 동적으로 표시하는데 사용될 수 있다. 그러나, 본 발명에서는, 다중 모드는 언어적 경계표 및 리포트 등의 의미 객체의 인스터스화(instantiation)를 처리할 수 있으며, 즉, 일부 응답을 동적 방식으로 그들에 대한 애플리케이션에 파악선 평션으로 다시 제공한다. 이는 애플리케이션 디자이너에게는 음성 응용 언어 태그(SALT)가 의미 객체 동기 이해를 수행하는 것처럼 보인다.

[0091] 이 동작 모드는 다중 모드 시나리오를 이와 비교함으로써 보다 잘 이해될 수 있다. 다중 모드 시나리오에서는, 사용자가 예를 들어, 말하면서 입력 필드 내에 스타일러스를 포인팅하여 이를 유지하여 필드를 가리킨다. 사용자는 일반 필드 상으로 태핑(tapping)을 하고 단일 문장 내에 많은 필드를 채우도록 정교한 문장을 발언하지만, 그럼에도 불구하고, 탭-앤-톡(tap-and-talk) 인터페이스는 사용자의 눈과 손에 관련되어 많은 상황에 있어서 적절하지 않은 방식이다. 또한, 탭-앤-톡은 구두 언어 처리 기반의 프로세스 바와 볼륨을 표시하는 리치 백 채널 통신(rich back channel communication)을 특징으로 하더라도, 이들 피드백은 속도와 정확성에 있어서 구두 언어 처리의 품질에 대한 단지 매우 원시적인 단서만을 제공한다. 이는, 결국 인식 및 이해 출력만을 검증하여 보정하는 데 보다 많은 노력을 결국 요구하는 보다 넓은 영역으로 에러가 전파할 수 있는 보다 장문에 있어서는 잠재적으로 큰 문제가 될 수 있다. 유용성 연구는 키보드 기능향상 또는 대안 대신 음성을 사용하는 것이 장문에서는 주요 차별화 요소라는 것을 나타내므로 충분한 UI 경험은 경쟁력 있는 모드로서 연속적인 음성을 사용하는 것이 절대적으로 필요하다.

[0092] 공통 목적을 달성하는데 보다 밀접하게 협력하는 동업자로서 사람-컴퓨터의 인지를 증진하기 위해서, 의미 객체 동기 이해는, 그들이 입수가능하자마자 부분적 의미 구문분석 또는 객체를 보고함으로써, 효과적이게 된다. 일 실시예에서, 이는 음성 응용 언어 태그(SALT)에서 듣기 요소의 다중 모드를 사용함으로써 달성된다. 특히, 듣기 요소에 있어서, 다중 모드가 지정되어, 후에 인식될 입력 음성에 대하여 모든 인식 문법이 특정된다. 그 결과 또한 듣기 객체 내에 규정되어 있다. 예를 들어, 새로운 모임을 생성하기 위해서 필요한 정보를 획득하기 위한 날짜, 시간, 위치, 주제 및 모임 참가자 등과 같은 HTML 입력 코드는 다음과 같은 형태를 취할 수 있다:

```
<listen mode="multiple" …>
  <grammar src="subject.grm"/>
  <grammar src="date.grm"/>
  <grammar src="time_duration.grm"/>
  <grammar src="attendees.grm"/>
  <bind targetElement="subject"
   value="//ApptSubject"/>
  <bind targetElement="date"
   value="//DateTime"/>
  <bind targetElement="start_time"
   value="//start_time"
   targetElement="end_time"
   value="//end_time"
   targetElement="duration"
   value="//DateTime/duration"/>
  …
</listen>
```

[0093]

[0094] 다중 문법(multiple grammar)은 입력 지점으로 돌아오는 널 전환(null tranition)을 사용하여 인식을 위한 병렬 검색 공간(parallel search space)을 구성한다. 이 모드에서, 음성 응용 언어 태그(SALT)는 듣기 객체가 문법이 나오자마자 이벤트를 올릴 수 있게 한다. 그 이벤트는 병렬 프로세스를 채용하여 하부의 오디오 수집과 인식을 진행하면서 순서대로 바인딩 지침을 호출함으로써, 사용자에게 포맷 상의 관련 필드는 채워지는 반면 구두 명령은 필드의 시각적 렌더링을 갖는 애플리케이션에 대하여 여전히 발언되는 효과를 낸다.

[0095] 아이 프리(eye free) 애플리케이션에 대한 사용자 인터페이스에 있어서, 음성 출력이 수반되는 것이 바람직할 수 있다. 이 경우, 음성 응용 언어 태그(SALT) 프롬프트 객체는 즉시 피드백을 부여하는데 사용될 수 있다. 예를 들면, 다음 음성 응용 언어 태그(SALT) 프롬프트 객체는 데이터 필드 내의 동적 컨텐츠에 기초하여 응답을 합성하는데 사용될 수 있으며, 음성 합성은 추가 음성 응용 언어 태그(SALT) 바인딩 지침을 사용하여 다음과 같이 트리거될 수 있다.

```
<prompt id="say_date">
  on <value targetElement="date"/>
</prompt>
…
<listen …>
  …
  <bind targetElement="date"
   value="//date"
   targetElement="say_date"
   targetMethod="Start"/>
  …
</listen>
```

[0096]

[0097] 그 결과, 사용자가 그/그녀가 "Schedule a meeting (new meeting) at two (starting at two o'clock PM) next Tuesday (on 10/29/02) for two hours (duration: two hours)"와 같이 들리는 것을 간단히 메모할 뿐만 아니라 반복하는 다른 상대방과 대화하고 있다고 느끼며, 여기서, 괄호 안에 제공된 어구는 사용자에게 들릴 수 있거나 및/또는 볼 수 있는 프롬프트(동기화될 수도 있음)를 의미한다.

[0098] 음성 응용 언어 태그(SALT)는 설계자가 음성 응용 언어 태그(SALT) 바인딩 지침을 사용하는 바와 같이 단순 과제를 넘는 정교한 계산을 수행하는 개별화된 인식 이벤트 핸들러(customized recognition event handler)를 부착하게 할 수 있음이 인식되어져야 한다. 상기 예에서, 데이터 정규화(data normalization)는 의미 문법에서 달성될 수 있지만, 이는 기준 결정의 개선을 용이하게 하지는 않는다 (예를 들어, "Schedule a meeting with Li Deng and his manager"). 이러한 경우에 있어서, 알고리즘은 불명확한 기준을 확인하기 위해 저장된 데이터를

액세스하기 위한 적절한 이벤트에 대하여 액세스가능한 스크립트 객체로서 구현될 수 있다. 이러한 알고리즘은 전체로 여기에 참조로 통합되는, 2000년 중국 베이징 Proc. ICSLP-2000에서 K. Wang에 의한 "A plan based dialog system with probabilistic inferences" 및 2002년 4월 24일에 공개된 유럽특허출원번호 제EP 1199630A2에 기재되어 있다.

[0099] 본 발명은 특정 실시예를 참조하여 설명하였지만, 당업자는 본 발명의 취지 및 범위를 벗어남이 없이 형태상 그리고 세부적으로 변경이 행해질 수 있음을 인식할 것이다.

발명의 효과

[0100] 기존 구현예에서는 듣기 객체에 대하여 다중 모드 연산이 존재하지만, 이 모드는 구술 시나리오와 같은 수신 입력에 대하여 텍스트만을 제공한다. 그러나, 본 발명은 이 점에서, 입력이 수신됨에 따라 부분적 결과는 단지 텍스트 뿐만 아니라 텍스트에 관련되는 대응 의미 정보를 포함하며, 따라서, 컴퓨터가 수신된 것을 적절하게 이해하는 보다 우수한 피드백을 사용자에게 제공하도록, 출력은 상술한 바와 같이 사용될 수 있는 부분적 의미 구문분석 또는 객체를 포함한다. 부분적 의미 구문분석 또는 객체를 수신하는 애플리케이션의 정교함에 따라, 시스템은 수신된 부분적 의미 구문분석에 기초하여 확인, 대안, 정정 및 설명을 사용자에게 다시 제공할 수 있다.

[0101] 다중 모드 애플리케이션에서 다중 문법을 포함하는 것은 사용자에게 미리 말할 수 있게 하여, 표시하지 않았던 정보를 제공할 수 있는 능력을 부여하는 것으로 알려져 있지만, 다중 모드 연산에서 듣기 요소를 사용하는 것이 사용자에게 보다 높은 이해정도를 제공하기 때문에 보다 우수할 수 있다. 음성 전용 애플리케이션에서, 자연스러운 대화가 이루어지는 반면, 시각적 렌더링의 사용이 채용되면, 애플리케이션은 프로세서(예를 들어, 팝업 윈도우를 통해 동작을 취하고, 중간 결과 또는 옵션을 표시)를 사용자가 그때까지 제공한 것의 부분적 의미 구문분석에만 기초하여 계속 발언을 하면서 개시할 수 있다.

도면의 간단한 설명

[0001] 도 1은 데이터 프리젠테이션 시스템의 블록도.

[0002] 도 2는 컴퓨팅 장치 운영 환경의 평면도.

[0003] 도 3은 도 2의 컴퓨팅 장치의 블록도.

[0004] 도 4는 전화기의 평면도.

[0005] 도 5는 범용 컴퓨터의 블록도.

[0006] 도 6은 클라이언트/서버 시스템에서의 아키텍처의 블록도.

[0007] 도 7은 음성 인식 및 이해 모듈의 블록도.

[0008] <도면의 주요 부분에 대한 부호의 설명>

[0009] 12 : 음성 인터페이스 모듈

[0010] 14 : 음성 인식 및 이해 모듈

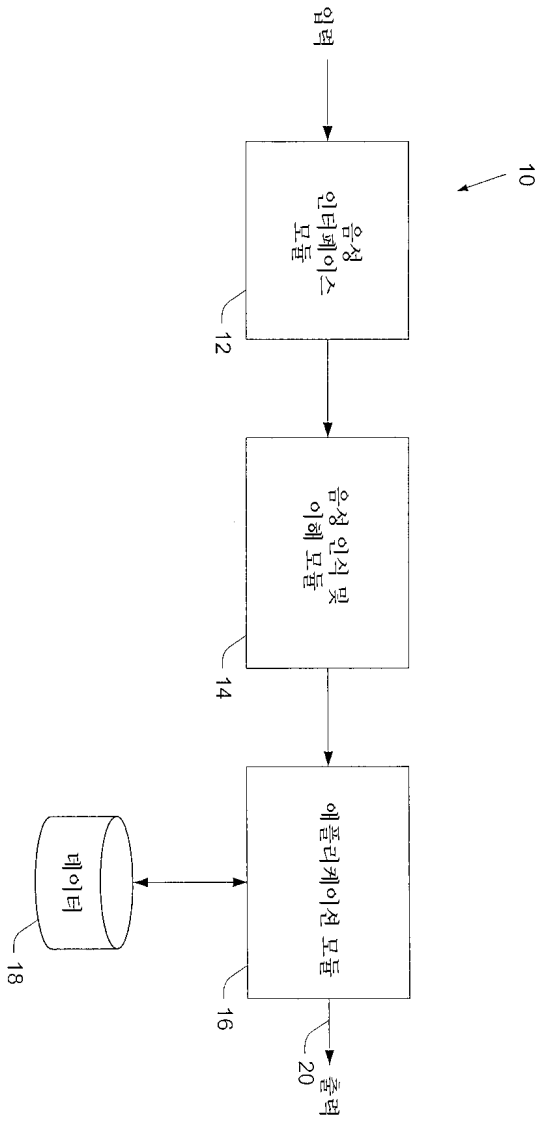
[0011] 16 : 애플리케이션 모듈

[0012] 306 : 음성 인식기

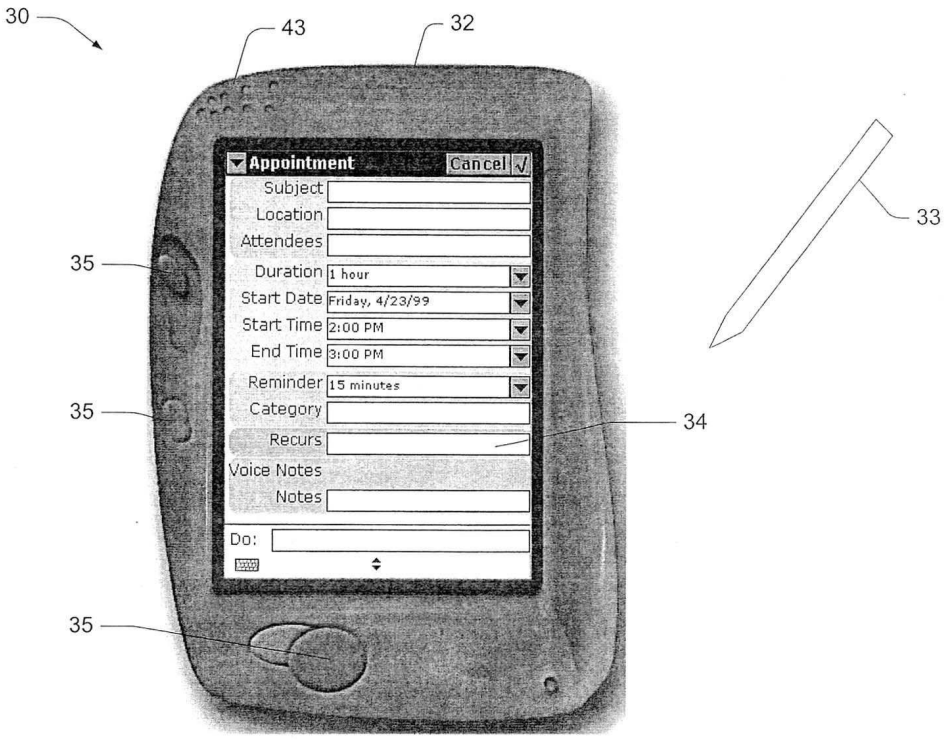
[0013] 310 : 언어 모델

도면

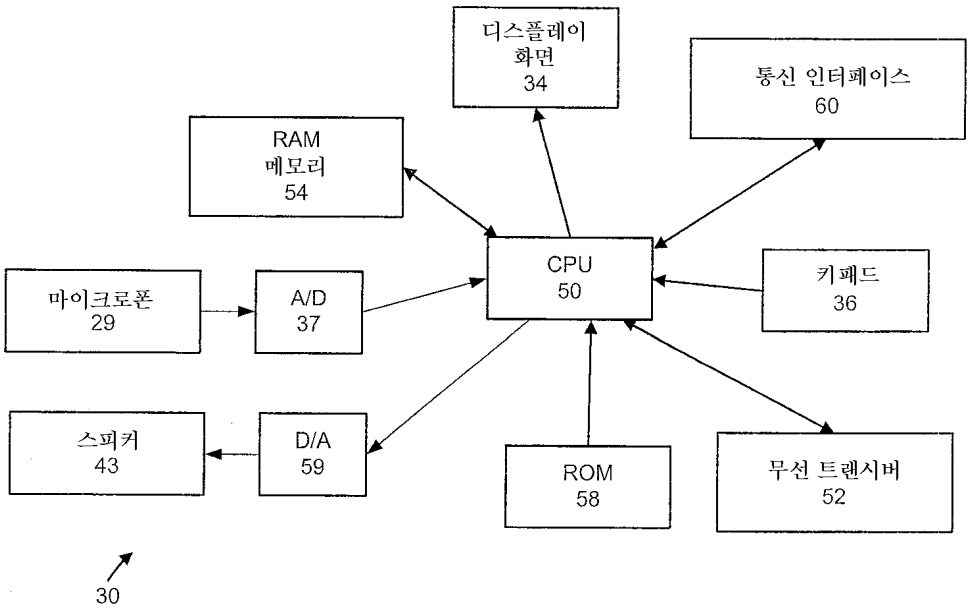
도면1



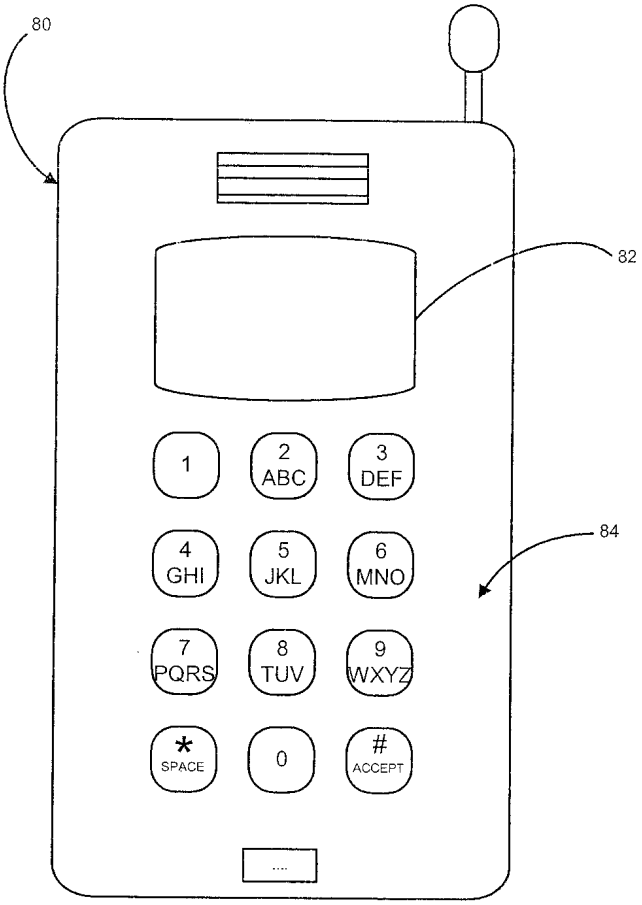
도면2


30
43
32
33
35
35
35
34
Appointment
Cancel
Subject
Location
Attendees
Duration
1 hour
Start Date
Friday, 4/23/99
Start Time
2:00 PM
End Time
3:00 PM
Reminder
15 minutes
Category
Recurs
Voice Notes
Notes
Do:

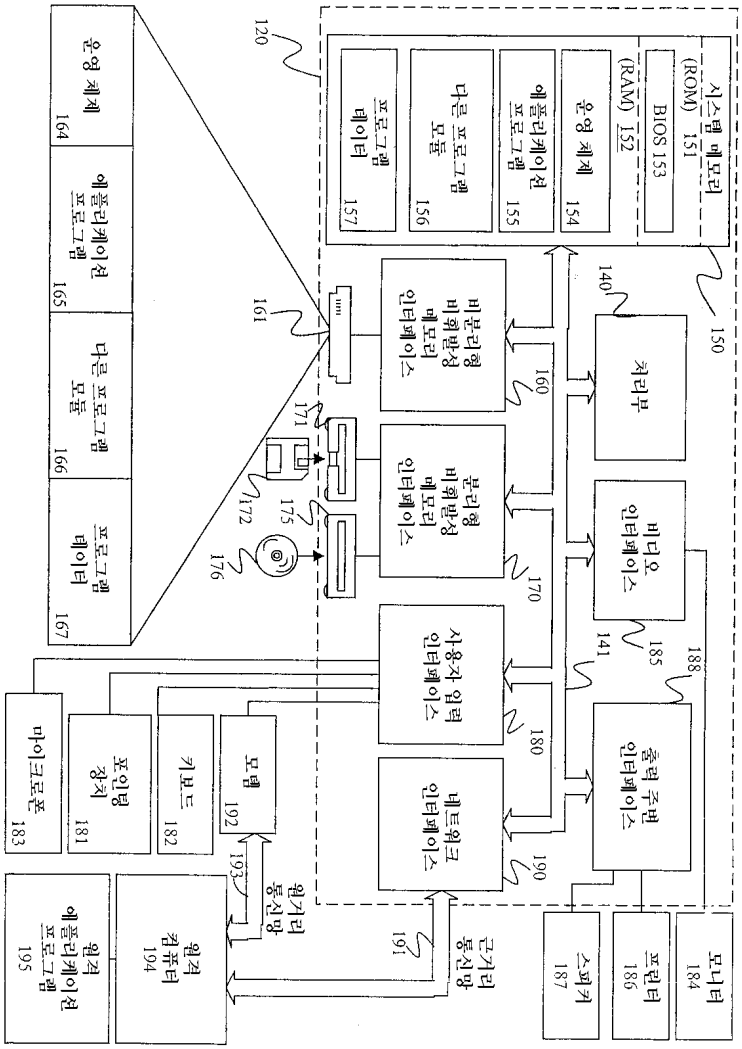
도면3


디스플레이
화면
34
통신 인터페이스
60
RAM
메모리
54
CPU
50
키패드
36
마이크로폰
29
A/D
37
스피커
43
D/A
59
ROM
58
무선 트랜시버
52
30

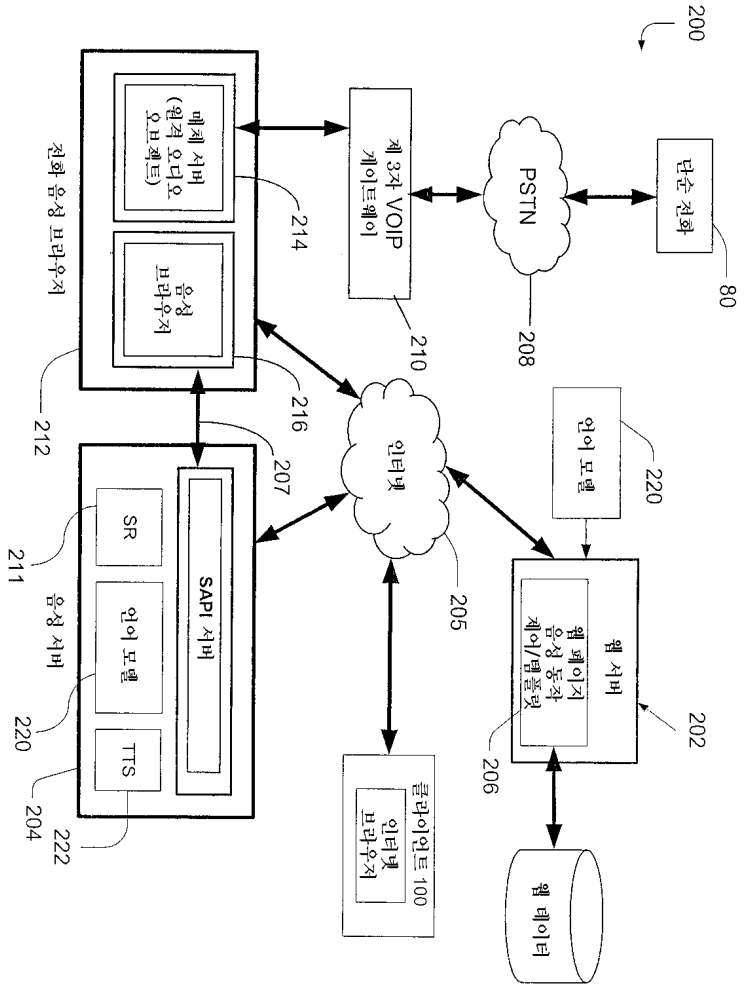
도면4


80
82
84
1
2 ABC
3 DEF
4 GHI
5 JKL
6 MNO
7 PQRS
8 TUV
9 WXYZ
* SPACE
0
ACCEPT

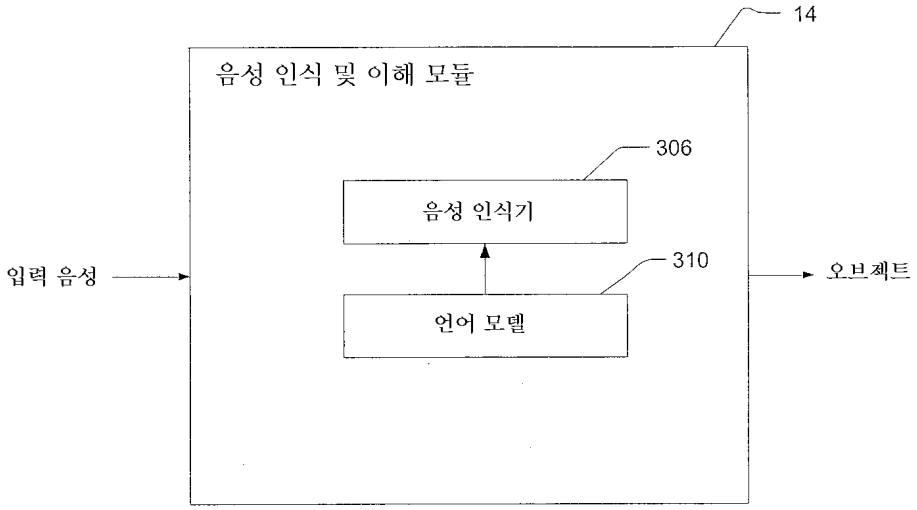
도면5


시스템 메모리
(ROM) 151
BIOS 153
(RAM) 152
운영 체제 154
애플리케이션 프로그램 155
다른 프로그램 모듈 156
프로그램 데이터 157
120
150
140
처리부
비분리형 비휘발성 메모리 인터페이스
160
분리형 비휘발성 메모리 인터페이스
170
비디오 인터페이스
141
185
188
출력 주변 인터페이스
사용자 입력 인터페이스
180
네트워크 인터페이스
190
191
161
171
172
175
176
운영 체제 164
애플리케이션 프로그램 165
다른 프로그램 모듈 166
프로그램 데이터 167
마이크로폰 183
포인팅 장치 181
키보드 182
모뎀 192
193
원거리 통신망
원격 컴퓨터 194
원격 애플리케이션 프로그램 195
근거리 통신망
스피커 187
프린터 186
모니터 184

도면6


200
전화 음성 브라우저
매체 서버
(원격 오디오 오브젝트)
음성 브라우저
214
212
216
207
제 3자 VOIP 게이트웨이
210
PSTN
208
단순 전화
80
인터넷
205
언어 모델
220
웹 서버
웹 페이지
음성 동작 제어/템플릿
202
206
음성 서버
SR
211
언어 모델
220
TTS
204
222
SAPI 서버
클라이언트 100
인터넷 브라우저
웹 데이터

도면7


14
음성 인식 및 이해 모듈
306
음성 인식기
310
언어 모델
입력 음성
오브젝트