



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2011년04월08일
(11) 등록번호 10-1027548
(24) 등록일자 2011년03월30일

(51) Int. Cl.
G06F 15/00 (2006.01) G06F 17/00 (2006.01)
G10L 21/00 (2006.01)
(21) 출원번호 10-2005-7012829
(22) 출원일자(국제출원일자) 2004년01월08일
심사청구일자 2008년12월31일
(85) 번역문제출일자 2005년07월08일
(65) 공개번호 10-2005-0100608
(43) 공개일자 2005년10월19일
(86) 국제출원번호 PCT/US2004/000497
(87) 국제공개번호 WO 2004/064299
국제공개일자 2004년07월29일
(30) 우선권주장
10/339,067 2003년01월09일 미국(US)
(56) 선행기술조사문헌
US5915001 A
US20030200080 A1
전체 청구항 수 : 총 10 항

(73) 특허권자
모토로라 모빌리티, 인크.
미국 60048 일리노이주 리버티빌 노쓰 유에스 하
이웨이 45 600
(72) 발명자
페란스, 제임스
미국 60187 일리노이주 휘튼 크리스티나 씨클 140
앵겔스마, 조나단
미국 49426 미시건주 허드슨빌 우드클리프 드라이
브 7434
(뒷면에 계속)
(74) 대리인
양영준, 정은진, 백만기

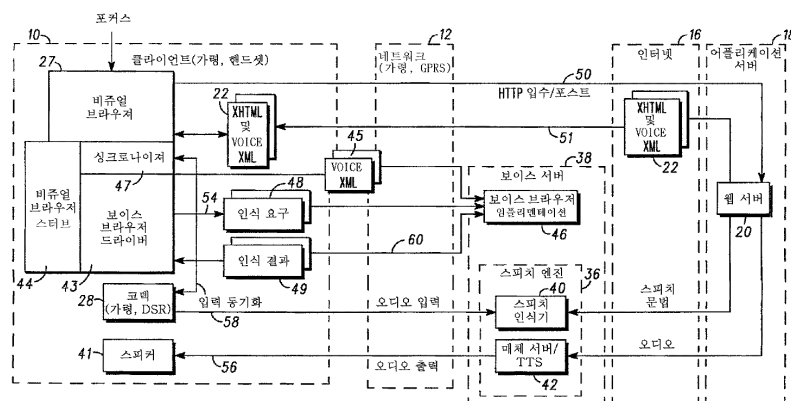
심사관 : 권오성

(54) 통신 시스템용 보이스 브라우저 다이얼로그 인에이블러

(57) 요약

멀티모드 다이얼로그를 위한 보이스 브라우저 다이얼로그 인에이블러는 각 필드와 관련된 마크업 기반 폼 필드를 구비하고 프래그먼트(45)를 정의하는 멀티모드 마크업 문서(22)를 사용한다. 보이스 브라우저 드라이버(43)는 통신 장치(10) 상에 상주하고 프래그먼트(45) 및 프래그먼트(45)를 식별하는 식별자(48)를 제공한다. 보이스 브라우저 임플리멘테이션(46)은 원격 보이스 서버(38)에 상주하고 상기 드라이버(43)으로부터 프래그먼트를 수신하며, 복수의 스피치 문법을 다운로드한다. 입력 스피치는 인식 요구시에 보이스 브라우저 드라이버(43)로부터 수신된 대응 식별자와 관련된 이러한 스피치 문법에 매칭된다.

대표도



(72) 발명자

피어스, 마이클

미국 60010 일리노이주 베링톤 버크셔 레인 1223

란돌프, 마크

미국 60047 일리노이주 킬디어 바클리 코트 21922

보게데스, 제롬

미국 53185 위스콘신주 윈드 레이크 프란시스 웨이
8525

특허청구의 범위

청구항 1

통신 시스템용 보이스 브라우저 다이얼로그 인에이블러(voice browser dialog enabler)로서,

복수의 어플리케이션 상호작용 유닛(units of application interaction)을 포함하는 스피치 인식 어플리케이션 (speech recognition application) - 상기 복수의 어플리케이션 상호작용 유닛 각각은 프래그먼트들 (fragments)을 정의하는 연관 보이스 다이얼로그 폼(associated voice dialog form)들을 가짐 - 과,

보이스 브라우저 드라이버 - 상기 보이스 브라우저 드라이버는 모바일 통신 장치에 상주하며, 상기 보이스 브라우저 드라이버는 상기 스피치 인식 어플리케이션으로부터 상기 프래그먼트들을 제공하고 상기 프래그먼트들을 식별하는 식별자들을 생성함 - 와,

원격 보이스 서버에 상주하는 보이스 브라우저 임플리멘테이션(voice browser implementation) - 상기 보이스 브라우저 임플리멘테이션은 상기 보이스 브라우저 드라이버로부터 상기 프래그먼트들을 수신하며 복수의 스피치 문법을 다운로드함 -

을 포함하고,

후속하는 입력 스피치는 상기 보이스 브라우저 드라이버로부터 스피치 인식 요구시 수신되는 대응 식별자들과 연관된 그 스피치 문법들에 대해 매칭되는, 통신 시스템용 보이스 브라우저 다이얼로그 인에이블러.

청구항 2

제1항에 있어서,

상기 스피치 인식 요구 및 후속하는 스피치 인식 결과는 마크업 기반(markup-based)인, 통신 시스템용 보이스 브라우저 다이얼로그 인에이블러.

청구항 3

제1항에 있어서,

상기 프래그먼트들은 식별된 폼들의 VoiceXML 페이지로 구성된, 통신 시스템용 보이스 브라우저 다이얼로그 인에이블러.

청구항 4

제1항에 있어서,

상기 스피치 인식 어플리케이션은 멀티모드 마크업 문서들을 처리하는 멀티모드 브라우저이고, 상기 보이스 브라우저 드라이버는, 멀티모드 마크업 문서에 대해, 상기 멀티모드 마크업 문서를 디스플레이가능한 마크업 부분 및 보이스 마크업 부분으로 분할하도록 동작하는 보이스 브라우저 스터브(voice browser stub) 내에 상주하며, 상기 보이스 브라우저 드라이버 및 상기 보이스 브라우저 임플리멘테이션은 상기 보이스 마크업 부분에 대해 동작가능한, 통신 시스템용 보이스 브라우저 다이얼로그 인에이블러.

청구항 5

제4항에 있어서,

상기 멀티모드 마크업 문서 및 상기 스피치 문법들을 포함하는 웹 서버를 구비한 인터넷 어플리케이션 서버를 더 포함하며, 상기 멀티모드 마크업 문서의 상기 디스플레이가능한 마크업 부분 및 보이스 마크업 부분 양쪽 모두에 대해 동작가능한 비주얼 브라우저를 상기 통신 장치 내에 더 포함하는, 통신 시스템용 보이스 브라우저 다이얼로그 인에이블러.

청구항 6

통신 시스템용 보이스 브라우저와의 다이얼로그를 인에이블시키는 방법으로서,

통신 장치상에 상주하는 보이스 브라우저 드라이버 및 원격 보이스 서버상에 상주하는 복수의 스피치 문법을 포함하는 보이스 브라우저 임플리멘테이션을 제공하는 단계와,

복수의 어플리케이션 상호작용 유닛을 포함하는 스피치 인식 어플리케이션을 구동하는 단계 - 상기 복수의 어플리케이션 상호작용 유닛 각각은 프래그먼트들을 정의하는 연관된 보이스 다이얼로그 품들을 구비함 -와,

각 프래그먼트와 연관된 식별자들을 정의하는 단계와,

상기 프래그먼트들을 상기 보이스 브라우저 임플리멘테이션에 공급하는 단계와,

상기 어플리케이션 상호작용 유닛들 중 하나의 필드에 포커싱하는 단계와,

상기 포커싱된 필드와 연관된 상기 품의 식별자를 포함하는 스피치 인식 요구를 상기 보이스 브라우저 드라이버로부터 상기 보이스 브라우저 임플리멘테이션으로 전송하는 단계와,

스피치를 입력 및 인식하는 단계와,

상기 스피치를 상기 식별자와 연관된 허용가능한 스피치 문법에 매칭하는 단계와,

스피치 인식 결과들을 취득하는 단계

를 포함하는, 통신 시스템용 보이스 브라우저와의 다이얼로그를 인에이블시키는 방법.

청구항 7

제6항에 있어서,

상기 전송 단계에서의 상기 스피치 인식 요구 및 상기 취득 단계에서의 상기 스피치 인식 결과들은 마크업 기반 이고, 상기 공급 단계는 상기 보이스 브라우저 임플리멘테이션에 식별된 품들의 VoiceXML 페이지를 공급하는 단계를 포함하는, 통신 시스템용 보이스 브라우저와의 다이얼로그를 인에이블시키는 방법.

청구항 8

제6항에 있어서,

상기 제공 단계는, 싱크로나이저(synchronizer)를 이용하여 상기 통신 장치상의 상기 보이스 브라우저 임플리멘테이션 및 비주얼 브라우저와 인터페이싱하는 보이스 브라우저 스템브에 상기 보이스 브라우저 드라이버가 통합되는 것을 포함하는, 통신 시스템용 보이스 브라우저와의 다이얼로그를 인에이블시키는 방법.

청구항 9

제6항에 있어서,

상기 구동 단계는, 멀티모드 마크업 문서를 상기 스피치 인식 어플리케이션 문서로서 다운로드하는 단계를 포함하는, 통신 시스템용 보이스 브라우저와의 다이얼로그를 인에이블시키는 방법.

청구항 10

제9항에 있어서,

상기 구동 단계 후, 상기 멀티모드 마크업 문서를, 상기 복수의 어플리케이션 상호작용 유닛을 포함하는 디스플레이가능한 마크업 부분 및 보이스 마크업 부분으로 분할하는 단계를 더 포함하고, 후속하는 단계들은 상기 문서의 상기 보이스 마크업 부분에 대해서만 동작할 수 있으며, 상기 제공 단계는 상기 멀티모드 마크업 문서의 상기 디스플레이가능한 마크업 부분 및 보이스 마크업 부분 양쪽 모두에 대해 동작할 수 있는 비주얼 브라우저를 상기 통신 장치 내에 제공하는 단계를 포함하는, 통신 시스템용 보이스 브라우저와의 다이얼로그를 인에이블시키는 방법.

명세서

기술 분야

본 발명은 무선 통신 장치의 동작 모드의 제어에 관한 것이다. 특히, 본 발명은 상이한 시스템상에서 다중 무

선 통신 장치(a multiple radio communication device)를 동작하는 방법에 관한 것이다.

배경 기술

- [0002] 셀룰러 폰과 같은 무선 통신 장치는 계속 커지는 처리 능력과 후속하여 무선 통신 장치상에서 구동하는 소프트웨어 어플리케이션을 구비한다. 그러나, 장치의 크기로 인해, 가령 통상적으로 이용가능한 사용자 인터페이스 하드웨어를 컴퓨터에 부착하는 것이 어렵다. 셀룰러 폰은 작은 키보드 및 디스플레이를 구비한다. 그러나, 셀룰러 폰에 내재된 기본적 음성 통신 능력을 이용하도록 하는 기술이 개발되었다. 스피치 인식 기술은 이제 무선 통신 장치에 공통적으로 사용된다. 이제는 보이스 활성 다이얼링(voice activated dialing)도 쉽게 이용가능하다. 인터넷 사용을 포함하는 데이터 서비스의 출현으로, 스피치 가능 서비스(speech-enabled service)가 통신 장치의 기능성을 매우 향상시킬 수 있다는 것이 명백해졌다. 이를 위해, 무선 통신 장치에 대해 스피치 가능 서비스를 용이하게 하는 VoiceXML(Voice Extensible Markup Language)이 개발되었다. 그러나, 소비자에게 이용가능한 스피치 가능 서비스의 출현으로, 휴대형 통신 장치에는 몇몇 심각한 문제가 발생하였다.
- [0003] 스피치 가능 서비스는 멀티모드 서비스(multimodal services)와 공동으로 사용될 경우, 어려운 문제를 제공한다. 멀티모드 다이얼로그에서, 입력은 음성, 키보드, 마우스 및 다른 입력 형식으로부터 입력될 수 있는 반면, 출력은 스피커, 디스플레이 및 다른 출력 형식일 수 있다. 표준 웹 브라우저(standard web browser)는 키보드 및 마우스 입력과 디스플레이 출력을 구현할 수 있다. 표준 보이스 브라우저는 음성 입력 및 오디오 출력을 구현할 수 있다. 멀티모드 시스템은 두 개의 브라우저(및 가능하다면 다른 것들)가 어떤 형식으로 결합될 것을 요구한다. 전형적으로, 이는 상이한 모드를 갖는 어플리케이션을 적절하게 동기화하는 다양한 기술을 요구한다. 이러한 어플리케이션 중 일부는 3GPP TR22.977, 3rd General Partnership Project; Technical Specification Group Services and Systems Aspects; Feasibility study for speech enabled services; (Release 6), v2.0.0(2002-09)에 기술되어 있다.
- [0004] 제1 해결 방법으로서, "로컬 스피치 리소스(local speech resources)를 구비한 팻 클라이언트(a fat client)" 해결 방법은 웹(비주얼) 브라우저, 보이스 브라우저, 내재적 스피치 인식(underlying speech recognition) 및 스피치 합성(텍스트 투 스피치) 엔진(speech synthesis (text-to-speech) engines)을 동일한 장치(컴퓨터, 모바일 폰, 셋톱 박스 등) 상에 두는 것이다. 이러한 해결 방법은, 많은 양의 소프트웨어 및 필요한 프로세싱 전력(processing power)으로 인해, 무선 통신 장치에 구현하는 것이 불가능할 것이다. 제2 해결 방법은 "서버 기반 스피치 리소스를 구비한 팻 클라이언트"로서, 여기서 스피치 엔진은 네트워크상에 상주하지만, 비주얼 브라우저 및 보이스 브라우저는 여전히 장치상에 상주할 것이다. 이는, 소형 장치에 있어서 제1 해결 방법보다는 어느 정도 보다 실질적이지만, 모바일 폰과 같은 소형 장치상에 구현하기에는 여전히 매우 어렵다. 제3 해결 방법은 "신 클라이언트(thin client)"인데, 여기서 장치는 단지 비주얼 브라우저만을 구비하며, 이는 네트워크상에 위치한 보이스 브라우저 및 스피치 엔진과 통합되어야 한다. 이러한 해결 방법은 모바일 폰과 같은 장치에 적합하지만, 두 브라우저가 통합되도록 유지하는데 필요한 동기화(synchronization)는 전체 시스템을 매우 복잡하게 만든다. 이러한 모든 해결 방법에서도 문제는 여전히 존재하는데, 즉, 해결 방법은 보다 작은 장치에 제공하기에는 실질적이지 못하거나, 보다 복잡한 동기화를 필요로 한다는 점이다.
- [0005] 그러므로, 보이스 브라우저 기술 및 멀티모드 기술을 무선 통신 장치 내에 통합하 것에 대한 문제를 제거할 필요성이 존재한다. 또한, 확장된 프로세싱 능력을 확장할 필요 없이, 통신 장치 내에 문제에 대한 해결 방법을 제공하는 것이 유리할 것이다. 또한, 통신 장치에 어떤 상당한 부가적 하드웨어 또는 비용이 없도록 하여 복잡성을 피하는 것이 유리할 것이다.

발명의 상세한 설명

- [0011] 본 발명은 보이스 브라우저 어플리케이션을 단일체로 다루기보다는 보이스 브라우저 어플리케이션을 두 개의 구성 요소로 나눈다. 이런 방법에서는, 장치상의 소프트웨어의 양이 매우 감소되며, 멀티모드 다이얼로그가 적은 비용으로 보다 더 작은 장치에서 구동되게 한다. 상기 장치상에서 브라우저 동기화를 행함으로써, 종래 기술의 해결 방법의 많은 복잡성이 회피된다. 또한, 공통 보이스 브라우저 드라이버를 구비함으로써, 멀티모드 어플리케이션이 브라우저 어플리케이션 대신에 자립형 프로그램(stand-alone program)으로서 기록될 수 있다. 이러한 개선은 통신 장치에 있어서 매우 작은 비용으로 달성된다. 본 발명은, 비용을 부가하고 장치 크기를 증가시키는 프로세싱 전력을 부가하는 대신에, 소프트웨어 멀티모드 다이얼로그에 필요한 보이스 브라우저용 소프트웨어 솔루션과 함께 통신 장치의 기존의 프로세싱 전력을 유리하게 이용한다.

[0012] 도 1을 참조하면, 멀티모드 통신에 대한 대부분의 또는 모든 프로세싱이 (신(thin)) 통신 장치에서 행해지는 종래 기술의 아키텍처가 제공된다. 간단히 하기 위해, 도시되지 않은 멀티모드 다이얼로그(multimodal dialog)의 적절한 동작에 요구되는 많은 상호접속들(interconnections)이 존재한다는 것이 이해되어야 한다. 도시된 예에서, 클라이언트 통신 장치(10)는 어플리케이션 서버(18)에 상주하는 멀티모드 어플리케이션에 액세스하기를 원한다. 전형적으로, 어플리케이션 서버(18)는 인터넷(16) 상에서 통신하기 위해 현행 상주 웹 서버(an existing, resident web sever)(20)를 사용한다. 예를 들어, 서비스 제공자의, 통신 시스템 내의 멀티모드/보이스 서버(14)는 인터넷에 연결되며, 후속하여 클라이언트 통신 장치(10)에 연결되는 셀룰러 네트워크(12)에 서비스를 제공한다. 웹 서버는, 사용자에게 인터페이스를 제공하기 위해 비주얼(XHTML) 마크업 및 보이스(VoiceXML) 마크업을 포함하는 멀티모드 마크업 문서(multimodal markup document)(22)를 제공한다. 공지된 바와 같이, XHTML 마크업 파일은 정보 상호작용(information interaction)을 위한 몇몇 필드(fields)를 사용자에게 제공할 수 있는 비주얼 폼(visual form)이다. 예를 들어, 사용자는 선택을 나타내기 위해 "라디오 버튼(radio button)" 필드 위를 지정 및 클릭할 수 있거나, 정보를 기입하기 위해 빈 필드 안에 텍스트를 타이핑할 수 있다. VoiceXML은 마크업 문서의 필드 내에 정보를 입력하기 위한 보이스 인터페이스를 제공하기 위해 XHTML과 연계하여 작동한다. 예를 들어, VoiceXML 마크업은 사용자에게 정보를 필드 내에 기입할 것을 요구하는 오디오 프롬프트(audio prompt)를 명시할 수 있다. 그러면, 사용자는 무언가를 말할 수 있고(또는 원하는 경우 텍스트를 기입할 수 있고), 보이스 브라우저 VoiceXML은 이러한 말을 듣거나 변환하여 프롬프트에 대해 허용가능한 응답(acceptable responses)을 정의한 VoiceXML 마크업에 의해 특정된 또는 참조된 문법(grammers)에 대해 이를 비교할 수 있다. VoiceXML 마크업은 문서의 어떤 필드, 즉 포커스의 필드와 연관될 수 있다. XHTML 및 VoiceXML을 포함하는 마크업 문서의 동작은 이미 현재의 표준 내에 제공되어 있다.

[0013] 셀룰러 네트워크(12)는, 본 기술 분야에 공지되어 있는 바와 같이, RTP 또는 유사한 전송 프로토콜 내에서 표준화되어 있는 오디오 패킷을 이용하며 DSR(Distributed Speech Recognition)을 포함하는 코덱(codec)(28)을 통해 클라이언트 장치(10)에 표준 오디오 입력 및 출력을 제공한다. 또한, 네트워크(12)는 멀티모드 정보를 클라이언트 장치의 비주얼 브라우저(26)에 제공하기 위해 사용되는 채널을 제공한다. 멀티모드 정보는 XHTML 파일(24)로서 전송된다. 본 예에서, 멀티모드/보이스 서버(14)는 클라이언트(10) 및 웹 서버(20) 사이의 통신의 보이스(VoiceXML) 및 비주얼(XHTML) 부분들을 분할 및 결합한다. 분할 및 결합은, 클라이언트 장치가 정보의 보이스 부분과 함께 통합된 방식으로 멀티모드 정보를 수신 및 제공하도록, 멀티모드 문서(22)의 보이스 및 비주얼 부분의 멀티모드 동기화에 의해 제공되는 통합(coordination)을 요한다. 클라이언트 장치는 상주하는 비주얼 브라우저(26)를 통해 멀티모드 정보(24)를 처리한다. 보이스 및 비주얼 정보의 개별적 처리는, 사용자에게 적절한 동작을 제공하기를 원하는 경우, 로컬 인터록(local interlock)의 사용을 필요로 하는 몇몇 통합 문제를 야기할 수 있다. 예를 들어, 사용자는 필드 포커스가 확립되기 전에 버튼을 누르고 있을 수 있다. 로컬 인터록은 필드 포커스가 확립될 때까지 스크린을 멈추게 할 수 있다. 다른 예로서, XHTML 폼이 클라이언트 장치상에 나타난 경우, 그리고 보이스 정보가 입력된 경우, 로컬 장치는 보이스 정보가 승인될 때까지 장치의 스크린을 잠금(lock out) 수 있다. 스크린을 잠그는 것은 사용자가 폼의 동일 필드 내에 텍스트 정보를 제공하지 못하도록 하는데, 이는 멀티모드/보이스 서버(14)를 통해 충돌하는 보이스 및 텍스트 정보의 경쟁을 초래한다.

[0014] 멀티모드/보이스 서버(14)는 멀티모드 정보를 클라이언트 서버(10)와 교환하는 대부분의 또는 모든 처리를 포함한다. 그러한 처리는 동기화 매니저(30)에 의해 제어된다. 동기화 매니저(30)는 문서(22)를 보이스 다이얼로그 정보(32)(가령, VoiceXML) 및 멀티모드 정보(XHTML)로 분할 및 분리하고, 이 정보를 전송한 바와 같이 동기화한다. 보이스 다이얼로그 정보는 클라이언트 장치(10)에 적절하게 포맷팅된 오디오 정보를 제공하기 위해 서버(14) 상에 스피치 엔진(36)과 인터페이스된 보이스 브라우저(34)에 전송된다. 불행히도, 두 개의 브라우저(26, 34)가 통합되도록 하는데 필요한 동기화는 전체 시스템을 매우 복잡하게 만들고, 클라이언트 장치(10)에 대한 로컬 록아웃 여전히 요구한다. 더욱이, 특정 목적의 멀티모드 서버(14)는 브라우저를 동기화하기 위한 특정 프로토콜도 필요로 한다.

[0015] 스피치 엔진(36)은, 본 기술 분야에 공지되어 있는 바와 같이, 오디오를 재생하고 스피치 인식을 제공한다. 스피치 엔진은 계산적으로 막대하며 많은 양의 RAM(random access memory)을 요한다. 그러한 리소스는 전형적으로 무선 전화와 같은 클라이언트 장치에 이용불가능한데, 이것이 개별 멀티모드/보이스 서버(14)가 본 예에 사용되는 이유이다. 보이스 브라우저(34)는 다이얼로그를 다루며, 마크업 문서의 관련 이벤트를 취하며 스피치 엔진에 오디오 프롬프트를 재생하고 보이스 응답을 듣도록 지시하는 상위 레벨의 프로세서이다. 그런 다음, 스피치 엔진은 소정의 보이스 응답(voiced response)을 보이스 브라우저 매치 폼 필드(voice browser match form

field)에 전송한다. 보이스 브라우저는 스피치 엔진으로부터의 응답에 어울리는 허용가능한 문법에 대한 사전 저장된 리스트(pre-stored list)를 유지하는 메모리를 포함한다. 예를 들어, XHTML 문서상의 필드는 허용가능한 응답으로서, "예" 또는 "아니오"라는 응답만을 요구할 수 있다. 스피치 엔진은 들어오는 보이스 입력(incoming voice input)을, 현재 문법에 의해 허용될 수 있는 인식된 발성(recognized utterance) 또는 에러 코드(error code) 중 어느 하나를 지정할 수 있는 인식 결과(recognition result)에 맵핑할 수 있다. 그런 다음, 이후 각각의 필드에 발성(utterance)을 할당함으로써, 상기 결과를 반영하기 위해 그 내부 상태를 업데이트할 보이스 브라우저에 상기 인식 결과를 전송할 것이다. 보이스 브라우저는 후속하여 상기 인식 결과를 동기화 매니저에 통보할 것이다. 이 경우, 스피치 엔진은 보이스 응답을 허용가능한 문법 리스트 내의 "예" 또는 "아니오"라는 응답 중 어느 하나에 매칭하고자 할 것이며, 상기 결과를 예/아니오 결과를 적절한 필드에 할당하고 동기화 매니저에게 통보하는 보이스 브라우저에 포워딩할 것이다.

[0016] 동기화 매니저(30)는, 응답을 조정하기 위해 현재 문서 내의 무슨 필드가 현재 행해지는 지를 웹 및 보이스 브라우저에 알린다. 달리 말하면, 동기화 매니저는 브라우저의 포커스의 필드를 결정한다. 비록 이것이 문자 그대로 동기화는 아니더라도, 효과는 동일하다. 정의에 의하면, 멀티모드 다이얼로그는 코텍(28)을 통하는 오디오 또는 비주얼 브라우저(26)를 통하는 키 스트로크 텍스트 엔트리(key stroke text entry) 중 어느 하나일 수 있는 유효 응답(valid response)을 포함할 수 있다. 동기화 매니저는, 멀티모드 정보의 통합 전송을 제공하기 위해, 이러한 이벤트들 모두의 가능성을 처리한다.

[0017] 도 2는 멀티모드 통신에 대한 대부분의 또는 모든 처리가 (팻) 통신 장치((fat) communication device) 상에서 행해지는 종래 기술의 아키텍처를 나타낸다. 이전과 같이, 클라이언트 통신 장치(10)는 어플리케이션 서버(18) 상에 존재하는 멀티모드 어플리케이션에 액세스하기를 원하는데, 여기서 어플리케이션 서버(18)는 통신하기 위해 상주 웹 서버(resident web server)(20)를 사용한다. 웹 서버(20)는 (전형적으로 셀룰러 네트워크(12)를 통해 서버 제공자, 가령, GPRS(General Packet Radio Service)를 통해 인터넷 접속을 제공함으로써) 클라이언트 장치(10)에 직접적으로 멀티모드 마크업 문서(22) 교환을 제공한다. 이전 예의 모든 멀티미디어/비주얼 서버 프로세스는 이제 클라이언트 장치(10) 상에 상주하며, 전술한 바와 동일하게 동작한다. 불행히도, (팻) 장치(10)는 이제 매우 확장된 프로세싱 전력 및 메모리를 요구하는데, 이는 비용을 많이 들게 한다.

[0018] 도 3은 멀티모드 통신의 프로세싱 중 몇몇이 통신 장치(10) 상에서 제한된 프로세싱 및 메모리 제한을 수용하기 위해 원격으로 행해지는 종래 기술의 아키텍처를 나타낸다. 이전과 같이, 클라이언트 통신 장치(10)는 어플리케이션 서버(18) 상에 존재하는 멀티모드 어플리케이션에 액세스하기를 원하는데, 여기서, 어플리케이션 서버(18)는 통신하기 위해 상주 웹 서버(20)를 사용한다. 웹 서버(20)는 (전형적으로 셀룰러 네트워크(12)를 통해 서버 제공자를 통한 인터넷 접속을 제공함으로써) 클라이언트 서버(10)에 직접적으로 멀티모드 파일(20) 교환을 제공한다. 이전 예의 대부분의 멀티모드/보이스 서버 프로세스는 여전히 클라이언트 서버(10) 상에 상주하며, 전술한 바와 동일하게 동작한다. 그러나, 이제 원격 보이스 서버(38)에는 그 내부에 상주하는 스피치 엔진(36)이 제공된다. 원격 보이스 서버(38)는, 현재에도 존재하는 것처럼, 서버 제공자 또는 기업에 의해 공급될 수 있다. 보이스 브라우저(34)는 정의된 MRCP(Media Resource Control Protocol)를 통해 스피치 엔진(36)과 통신한다. 불행히도, 원격 리소스를 구비한 (팻) 장치(10)는 여전히 실질적으로 확장된 프로세싱 전력 및 메모리를 요구하며, 이는 여전히 비용이 많이 들게 한다. 더욱이, 보이스 브라우저 및 스피치 엔진 사이에 전송될 많은 양의 코드가 존재하는데, 이는 네트워크에 부담을 주며 통신을 느리게 할 것이다.

[0019] 가장 간단한 실시예로서, 본 발명은 통신 시스템용 보이스 브라우저 다이얼로그 인에이블러(voice browser dialog enabler)이다. 보이스 브라우저 인에이블러는 복수의 어플리케이션 상호작용 유닛(units of application interaction)을 포함하는 스피치 인식 어플리케이션을 포함하는데, 이는 복수의 관련 사용자 인터페이스 입력 엘리먼트(related user input elements)이다. 예를 들어, 어드레스 북(address book)에서, 사용자가 새로운 어드레스 엔트리를 생성하기를 원하는 경우에는 이름 및 전화 번호를 입력하는 것이 필요할 것이다. 이 경우, 어플리케이션 상호작용 유닛은 밀접하게 관련된 두 개의 입력 필드(즉, 이름 필드 및 어드레스 필드)일 것이다. 각각의 어플리케이션 상호작용 유닛은 프래그먼트(fragments)를 정의하는 보이스 다이얼로그 포맷(voice dialog formats)을 연관시킬 것이다. 예를 들어, 스피치 인식 어플리케이션은 XHTML+VoiceXML 문서를 처리하는 멀티모드 브라우징 어플리케이션일 수 있다. 각각의 XHTML+VoiceXML 문서는 단일 어플리케이션 상호작용 유닛을 구성할 수 있고 하나 이상의 필드와 연관된 하나 이상의 VoiceXML 폼을 포함할 수 있다. 각각의 VoiceXML 폼은 프래그먼트를 정의한다. 통신 장치상에 상주하는 보이스 브라우저 드라이버(voice browser driver)는 어플리케이션으로부터 프래그먼트를 제공하며 프래그먼트를 식별하는 식별자를 생성한다. 원격 보이스 서버 상에 상주하는 보이스 브라우저 임플리멘테이션은, 보이스 브라우저 드라이버로부터

터 프래그먼트를 수신하고 복수의 스피치 문법을 다운로드하는데, 여기서 후속하는 입력 스피치는, 보이스 브라우저 드라이버로부터 스피치 인식 요구로 수신된, 대응하는 식별자와 관련된 이들 스피치 문법에 대해 매칭된다.

[0020] 도 4는 본 발명에 따른, 멀티모드 다이얼로그를 용이하게 하기 위해 보이스 브라우저 인에이블러를 사용하는 보이스 브라우저의 실질적인 구성을 나타낸다. 이 예에서, 어플리케이션 서버(18), 웹 서버(20), 인터넷 접속(16) 및 마크업 문서(22)는 전술한 바와 동일하지만, 본 발명을 설명하기 위해 보다 상세히 도시되어 있는데, 여기서 보이스 브라우저의 기능은 분할되어 있다. 예를 들어, 마크업 문서(22)는 스피치 문법 및 오디오 파일에 대한 방향을 나타내는 URLs도 포함한다. 게다가, 스피치 엔진(36)은 전술한 바와 동일하지만, 보다 상세하다. 예를 들어, 스피치 엔진(26)은 어플리케이션 서버에 의해 제공되는 스피치 문법에 사용하기 위한 스피치 인식 유닛(40), 및 기록된 오디오 URL로부터 오디오 프롬프트를 제공할 수 있거나 TTS(Text-To-Speech)를 사용할 수 있는 미디어 서버(42)를 포함하는데, 이는 본 기술 분야에 공지되어 있는 것이다.

[0021] 본 발명의 새로운 특징은 본 발명이 통신 장치상의 보이스 브라우저를 "스터브(stub)(44)" 및 원격 보이스 서버(38) 상의 보이스 브라우저 "임플리멘테이션(implementation)(46)"으로 분할된다는 것이다. 바람직한 실시예에서, 보이스 브라우저 스터브(44)는 보이스 브라우저 임플리멘테이션(46)과 인터페이스되는 보이스 브라우저 드라이버(43)와, 보이스 브라우저 스터브(44) 및 비주얼 브라우저(26)를 통합하는 싱크로나이저(synchronizer)(47)로 재분할된다. 또한 싱크로나이저(47)는, 사용자가 코텍(28)에 말하고 있는지 여부에 기초하여 비주얼 브라우저(27)에 대한 입력을 선택적으로 인에이블 및 디스에이블시킨다(입력 동기화). 이러한 보이스 브라우저 스터브의 재분할은 클라이언트 장치(가령, J2ME 머플리케이션)를 구동하는 독립형 어플리케이션(stand-alone applications)이 비주얼 브라우저(27) 및/또는 싱크로나이저(47) 대신에 사용되도록 하며, 보이스 브라우저 스터브(44)의 나머지의 능력을 재사용하도록 한다.

[0022] 본 발명의 또 다른 새로운 특징은 비주얼 브라우저(27)가 이제는 전체 마크업 문서(full markup document), 보이스 및 오디오 상에서 동작하는데, 이는 원격 동기화에 대한 필요성을 제거한다. 결과적으로, 싱크로나이저(47)는 종래 기술의 동기화 매니저(종래 기술의 도면에 30으로 나타나 있음) 보다 작고 간단한 임플리멘테이션을 갖는다. 또한, 보이스 브라우저(43, 46)는 종래 기술과 같이 입력 필드 및 값을 사용하지 않는다. 대신 음성 브라우저는 포커스 필드(focused field)와 함께 동작한다. 이는 보이스 브라우저 임플리멘테이션(46)을 단순화하는 것을 돕는데, 이에 대해서는 이하에서 설명하기로 한다.

[0023] 동작에 있어서, 멀티모드 마크업 문서(22)가 웹 서버(20)로부터 페치된 후, 비주얼 브라우저는 그 카피(copy)를 보이스 브라우저 스터브에 전송한다. 비주얼 브라우저는 그 카피를 보이스 브라우저 스터브에 전송한다. 보이스 브라우저 스터브(44)는 보이스 브라우저 마크업(가령, VoiceXML)을 문서에서 분리 또는 분할하여, 디스플레이 가능한 마크업(가령, XHTML) 및 보이스 브라우저 마크업(가령, VoiceXML)으로 만든다. 그런 다음, 보이스 브라우저 스터브(44)는, 전술한 바와 같은 클라이언트 장치상에서와 같이 프로세싱 및 디스플레이하기 위해 비주얼 마크업을 비주얼 브라우저에 전송한다. 그러나, 보이스 브라우저 스터브(44)의 보이스 브라우저 드라이버(43)는 보이스 브라우저 마크업에 대해 종래 기술에서 행해지는 것과는 다르게 동작한다. 본 발명에서, 보이스 브라우저 드라이버는 프래그먼트 마크업 문서의 프래그먼트에 대해 동작한다. 프래그먼트는 단일 VoiceXML 폼이며(XHTML 폼과 혼동되지 않음; XHTML과 유사하나 그들 사이에 1대1 관계가 없음), 큰 XHTML+VoiceXML 문서의 개별 조각(individual piece)으로 간주될 수 있다. 폼은 단지 VoiceXML 내의 다이얼로그 유닛인데, 그 목적은 사용자에게 프롬프트하고 전형적으로는 그 폼 내의 하나 이상의 필드를 기입하는 것이다. XHTML 내의 하나의 입력 필드 폼은 하나의 VoiceXML 폼 또는 그와 관련된 프래그먼트를 가질 수 있다. 또한, 밀접하게 연관된 XHTML 폼 입력의 세트가 XHTML 폼 입력 모두를 기입할 수 있는 단일 VoiceXML을 가지도록 하는 것이 가능하다. 보이스 브라우저 드라이버는, 전체 문서 폼 및 값에 대해 동작하는 종래 기술의 보이스 브라우저 대신, 마크업 문서의 하나의 포커스 필드 또는 프래그먼트에 대해 한번에 동작한다.

[0024] 또한, 이러한 폼/프래그먼트는 이미 문서의 헤드 섹션 내에 함께 모여있으므로, XHTML+VoiceXML 문서로부터 VoiceXML 폼/프래그먼트를 생성하는 것이 어렵지 않기 때문에 보이스 브라우저 드라이버는 종래 기술의 보이스 브라우저보다 적은 프로세싱을 사용한다. 보이스 브라우저 드라이버가 할 필요가 있는 모든 것은 프래그먼트/폼을 찾고, 유일한 식별자를 그들과 연관시키며, 보이스 브라우저 임플리멘테이션으로의 전송을 위해 보이스 브라우저 스터브가 그들을 랩업(wrap up)하는 것이다. 식별자는 단지 단일 VoiceXML 폼(여기서 유일함은, 단일 멀티모드 마크업 문서로부터 생성된 것처럼, 보이스 브라우저에 주어진 프래그먼트의 세트의 범주 내에서만 요구됨)을 유일하게 식별하는 스트링(string)이다. 프래그먼트 및 식별자의 사용은 네트워크(12)를 통한 클라이

언트 장치(10) 및 원격 서버(38) 사이의 데이터 전송량을 줄인다.

[0025] 특히, 포커스 필드에 있어서, 프래그먼트가 연관되어 있다. 보이스 브라우저 드라이버는 필드가 XHTML 또는 VoiceXML인지 여부에 독립적으로 동작할 수 있다는 점이 유의되어야 한다. 예를 들어, XHTML 폼은 사용자에게 스트리트 어드레스(street address)에 대하여 물을 수 있다. 이 경우, 스트리트 어드레스(숫자 및 거리)에 대한 텍스트 필드, (선택적인) 아파트 번호에 대한 또 다른 텍스트 필드, 도시에 대한 또 다른 텍스트 필드, 주(state)에 대한 팝업 메뉴 및 zip코드(zipcode)에 대한 마지막 텍스트 필드가 존재할 것이다. 이제, 이러한 XHTML 폼이 주어진 경우, 이 필드를 기입하기 위해 함께 동작하는 VoiceXML 폼의 세트가 존재할 수 있다. 예를 들어, 하나의 VoiceXML 폼은 스트리트 어드레스 및 아파트 번호 필드 모두를 기입할 수 있을 것이며, 또 다른 VoiceXML 폼은 zip 코드만 기입할 수 있을 것이다. 이러한 폼들은 페이지의 프래그먼트로 정의된다.

[0026] 이러한 세 개의 VoiceXML 폼 각각은 자신의 유일한 식별자(즉, 명명된 VoiceXML 폼)를 갖는다. 예를 들어, 이러한 식별자는 "스트리트+아파트", "시(city)+주(state)" 및 "zip코드"로 각각 불릴 수 있다. "스트리트+아파트" VoiceXML 폼은, 활성화되었을 경우, 사용자가 "스트리트 어드레스 및 아파트 번호를 말하십시오."를 듣도록 하는 오디오 프롬프트를 포함할 것이다. 또한, 스트리트 어드레스 및 선택적인 아파트 번호를 이해하는 것이 가능하게 되어 있는 문법이 존재할 것이다. "시+주" VoiceXML 폼은 "시 이름과 주를 말하십시오."와 같은 오디오 프롬프트 및 그에 대한 적절한 문법을 포함할 것이다. zip코드에 대해서도 유사하게 적용될 수 있다.

[0027] 보이스 브라우저 스택은 관련 VoiceXML 프래그먼트(45)의 페이지를 보이스 브라우저 임플리멘테이션(46)에 전송한다. 그런 다음, 보이스 브라우저 스택(44)가 사용자 입력에 대해 들을 필요가 있을 경우에는 인식 요구(48)를, 인식을 위한 사용을 위해 폼에 대한 이름 및 식별자를 알리는 보이스 브라우저 임플리멘테이션(46)에 전송한다. 이전과 같이, 보이스 서버(38)는 스피치 문법을 포함하지만, 본 실시예에서, 식별자는 이미 전송된 보이스 프래그먼트와의 매칭을 찾기 위해 단지 "스트리트+아파트", "시+주", "zip코드" 문법을 보기 위해서 보이스 브라우저 임플리멘테이션을 코드화하는 식별자가 전송될 수 있다. VoiceXML 폼은 보이스 서버(38)에 한번 전송, 처리 및 그 후 캐쉬될(cached) 수 있다. 후속하는 요구는 그들의 식별자에 의해 캐쉬된 VoiceXML 폼을 식별할 수 있다. 이는 VoiceXML을 매 요구마다 전송 및 처리할 필요성을 제거한다. 그 결과, 문법 조사(grammar search)는 단순화되어, 프로세싱 전원 및 시간이 절약된다. 문서의 스트리트+아파트 필드에 대한 폼/프래그먼트를 위한 식별자가 보이스 브라우저 임플리멘테이션에 인식 요구(48)로서 전송되는 경우, 보이스 브라우저 임플리멘테이션은 스피치를 입력할 것이며, 보이스 브라우저(46)는, 가령, "메인 스트리트"의 입력 스피치에 대한 매치와 같은, 매치에 조사에 대한 적절한 문법으로 스피치 인식기(40)를 활성화할 것이다. 매치가 발견되고 나면, 보이스 브라우저 임플리멘테이션은 사용자가 텍스트로 말한 것("M-a-i-n-S-t-r-e-e-t")을, 보이스 브라우저 드라이버(43)에 인식 결과(49)로서 다시 전달하는데, 이는 종래 기술과 유사하다. 다음으로, 보이스 브라우저 스택(44)는 결과를 취하고 결과를 디스플레이하기 위해 비주얼 브라우저(27)를 업데이트한다. 보이스 브라우저 임플리멘테이션(46)은 보이스 브라우저 스택(44)를 위한 인터페이스를 구비한 종래 기술의 보이스 브라우저와 동일할 수 있으나, 본 발명은, 보이스 브라우저가 이제 단순한 VoiceXML 마크업의 작은 프래그먼트만을 처리하기 때문에, VoiceXML 언어 내에서 많은 태그 및 특징을 이용하지 않는, 보다 단순한 임플리멘테이션을 제공한다.

[0028] 실제로, 보이스 브라우저 스택(44)는 페이지의 모든 필드에 대해 관련 프래그먼트(45)를 한번에 보이스 브라우저 임플리멘테이션(46)에 전송할 수 있다. 이후, 보이스 브라우저 스택(44)는 임의의 포커스 필드에 대한 멀티모드 상호작용의 보이스 부분을 통합하고, 보이스 브라우저 임플리멘테이션(46)에 필요한 임의의 스피치 인식 요구 식별자(48)를 전송하며, 그 프래그먼트에 대응하는 인식 결과(49)를 획득한다. 바람직하게, 인식 요구(48) 및 인식 결과(49)를, 종래 기술에서 사용되는 것과 같이 MRCP처럼 하위 레벨의 API를 사용하는 것보다는 오히려, 마크업 기반(가령, XML)으로 만드는 것이 요구된다.

[0029] 도 5는 도 4와 함께, 본 발명에 따른, 멀티모드 다이얼로그의 상호작용을 설명하기 위해 사용될 수 있다. 도 5는 마크업 문서 내에 두 개의 텍스트 필드, 즉, 보이스에 의해 기입되는 필드(A)와 텍스트로서 기입되는 필드(B)를 구비한 단순화된 상호작용을 나타낸다. 다수의 보이스 필드 또는 텍스트 필드가 멀티모드 다이얼로그에 사용될 수 있다는 것이 인식되어야 한다. 사용자는, 가령, 인터넷 어드레스를 클릭함으로써 다이얼로그를 개시한다. 이는 비주얼 브라우저에 HTTP 입수/포스트(GET/POST) 요구(50)를 어플리케이션 웹 서버(20)에 전송하도록 지시하여, 원하는 마크업 문서(22)를 얻게 한다(51). 또한, 문서는 문서에 대해 허용가능한 문법에 대한 URL을 포함하는데, 이는 보이스 서버(38)에 다운로드될 수 있다. 일단 수신되고 나면, 비주얼 브라우저(27)가 구동되어 클라이언트 장치(10)의 스크린 상에 문서를 만든다(52). 다음으로, 오디오 및 비주얼 문서는 보이스

브라우저 스택(44)에 전송되는데, 이는 보이스(VoiceXML) 마크업을 문서에서 분할된다. 또한, 보이스 브라우저 스택(44)은 마크업의 VoiceXML 폼(프래그먼트)을 식별하며 이 프래그먼트를 보이스 서버(38)에 전송한다. 이 시점에서, 보이스 서버(36)의 보이스 브라우저 임플리멘테이션(46) 및 스피치 엔진(36)은 문서가 잘 형성되었는지에 대한 선택적 백그라운드 검사(optional background check)를 할 수 있고, 문서를 전처리(즉, 컴파일)도 할 수 있으며, 어떤 외부 스피치 문법 또는 문서가 참조할 수 있는 오디오 프롬프트를 폐지/전처리(즉, 컴파일, 디코딩/인코딩)하고, 텍스트를 스피치에 통합할 수 있다.

[0030] 다음으로, 사용자는 포커스(53)를 정의하는 디스플레이된 마크업 문서의 필드를 선택할 수 있다. 비주얼 브라우저(27)는 포커스 변화(focus change)를 수신하고, 포커스 필드 우측으로 점프하며, 필드 포커스를 보이스 브라우저 스택(44)로 옮긴다. 다음으로, 보이스 브라우저 스택(44)의 보이스 브라우저 드라이버(43)는 인식 요구(48)로서 폼의 그 필드 포커스에 대한 식별자를 보이스 서버(38)에 전송하는데(54), 이는 요구를 승인한다(55). 이 시점에서, 보이스 서버(38)는 하나 이상의 오디오 프롬프트를 RTP(real time streaming protocol) 오디오 패킷(57)으로서 사용자에게 전송함으로써(56), 스피치 입력에 대해 사용자에게 선택적으로 프롬프트할 수 있다. 오디오는 클라이언트 장치의 스피커(41) 오디오 리소스로 전달된다. 다음으로, 사용자는 PTT(push-to-talk) 버튼을 누르고 클라이언트 장치의 코덱(28) 오디오 리소스를 통해 보이스 서버(38)에 전송함으로써 보이스에 의해 응답할 수 있다. 코덱은 RTP DSR 패킷(59)으로서의 스피치를 보이스 서버의 스피치 엔진에 전송 - 이는 스피치를 허용가능한 그 폼 및 필드와 관련된 관련 식별자에 따른 문법에 매칭시킴 - 하고, 인식 결과로서의 텍스트 응답을 보이스 브라우저 스택(44)의 보이스 브라우저 드라이버(43)에 전송한다. 보이스 브라우저 스택(44)은 장치상의 디스플레이 스크린과 필드 및 값의 맵(map)을 업데이트하기 위해 비주얼 브라우저(27)와 인터페이스된다.

[0031] 또한, 사용자는 텍스트를 기입하기 위해 포커스(61)를 정의하는 디스플레이된 마크업 문서의 필드를 선택한다. 이전과 같이, 비주얼 브라우저(27)는 포커스 변화를 수신하고, 포커스된 필드의 우측으로 점프하며, 필드 포커스를 보이스 브라우저 스택(44)로 전송한다(62). 다음으로, 보이스 브라우저 스택(44)의 보이스 브라우저 드라이버(43)는 인식 요구(48)로서 포커스의 그 필드에 대해 폼(63)을 보이스 서버(44)에 구동하는데, 이는 요구(64)를 승인한다. 여기서 어떠한 오디오 프롬프트가 사용되지 않는다고 가정하면(비록 전술한 바와 같이 사용된다 하더라도), 사용자는, 필드 및 값의 맵을 집적적으로 업데이트하는 텍스트를 기입하는 것(65)에 응답할 수 있다. 또한, 폼 데이터(이는 필드 및 값의 맵 내에 유지됨)를 웹 서버(20)에 제출(66) 시, 비주얼 브라우저는, 보이스 서버에 문법에 대한 듣기를 멈추도록 지시하기 위해, 보이스 브라우저 스택(44)을 통해 보이스 서버에 취소 커맨드(67)를 전송한다. 다음으로, 다이얼로그 프로세스는 반복될 수 있다.

[0032] 본 발명은 멀티보드 다이얼로그를 제한된 리소스에 제공하는 솔루션을 제공한다. 본 발명은 동기화된 멀티모드 통신을 유지함에 있어서 특별한 어플리케이션을 발견한다. 본 발명의 방법은 통신 장치상에 최소한의 프로세서 및 메모리 요구를 이용하여, 보이스 브라우저의 프로세싱 요구를 분할하는 프로세스를 제공한다. 이는 단지 최소한의 소프트웨어 변형으로 달성되며, 여기서 외부 동기화 또는 특정된 멀티모드 서버에 대한 필요성은 존재하지 않는다.

[0033] 본 발명이 상기 설명 및 도면으로 기술 및 설명되었지만, 이러한 설명은 단지 예에 의한 것이며, 본 발명의 범주 및 사상 내에서 수많은 변형 및 변경이 당업자에 의해 이루어질 수 있다. 본 발명이 휴대형 셀룰러 무선 전화에서의 특정 사용을 찾았지만, 본 발명은, 페이지, 전자 오거나이저 및 컴퓨터를 포함하는 임의의 통신 장치에서의 멀티모드 다이얼로그에 적용될 수 있다. 출원인의 발명은 청구범위로 한정된다.

[0034]

도면의 간단한 설명

[0006] 도 1은 제1 종래 기술의 멀티모드 통신 시스템의 블록 다이어그램이다.

[0007] 도 2는 제2 종래 기술의 멀티모드 통신 시스템의 블록 다이어그램이다.

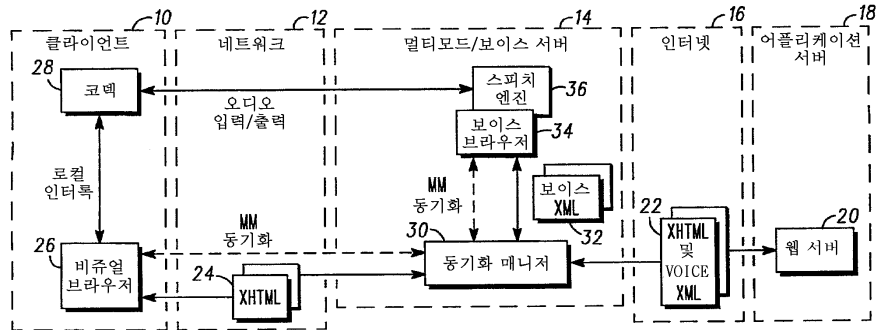
[0008] 도 3은 제3 종래 기술의 멀티모드 통신 시스템의 블록 다이어그램이다.

[0009] 도 4는 본 발명에 따른 개선된 보이스 브라우저를 구비한 멀티모드 통신 시스템의 블록 다이어그램이다.

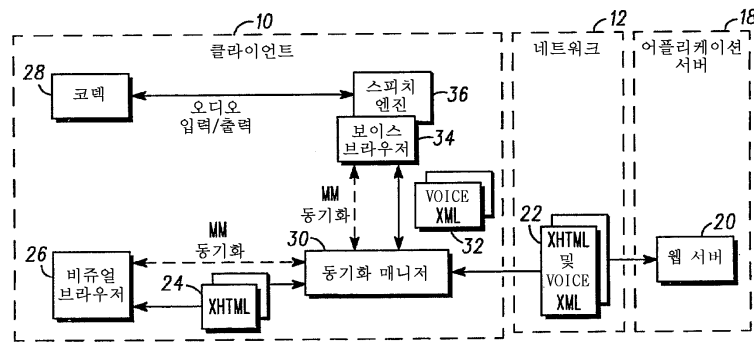
[0010] 도 5는 본 발명의 바람직한 실시예에 따른 멀티모드 다이얼로그의 단계를 도시한 플로우 차트이다.

도면

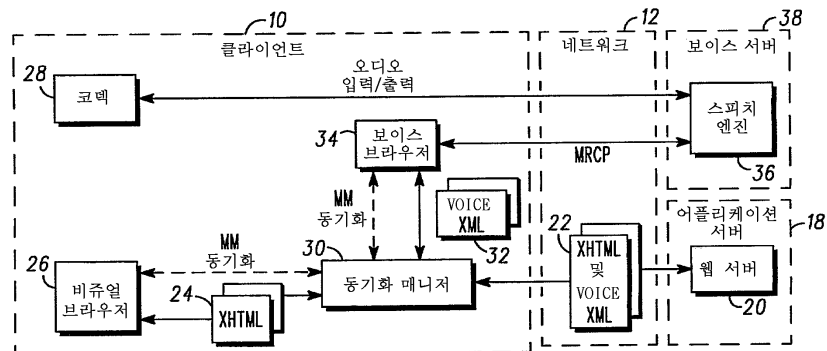
도면1



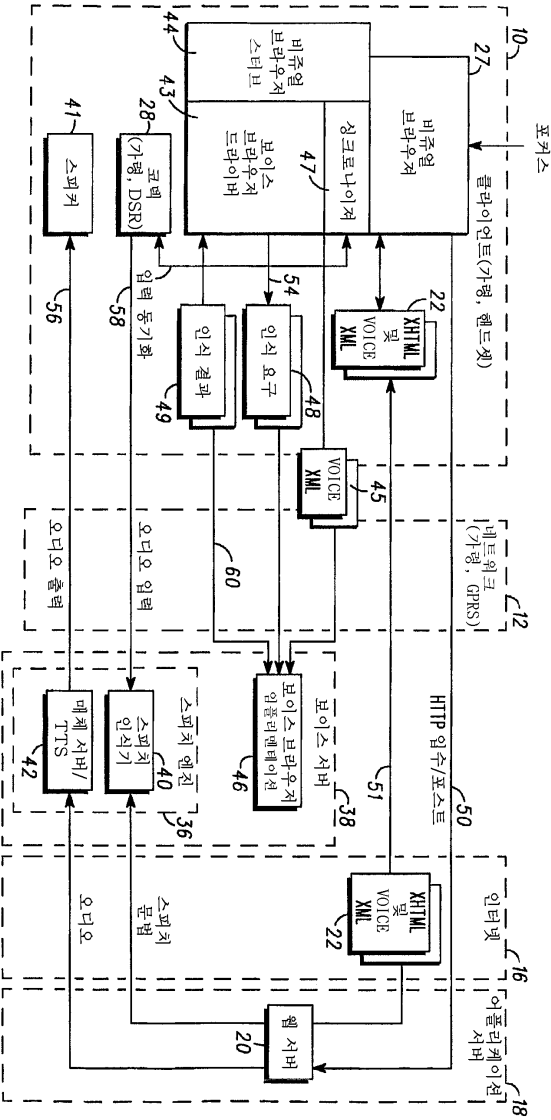
도면2



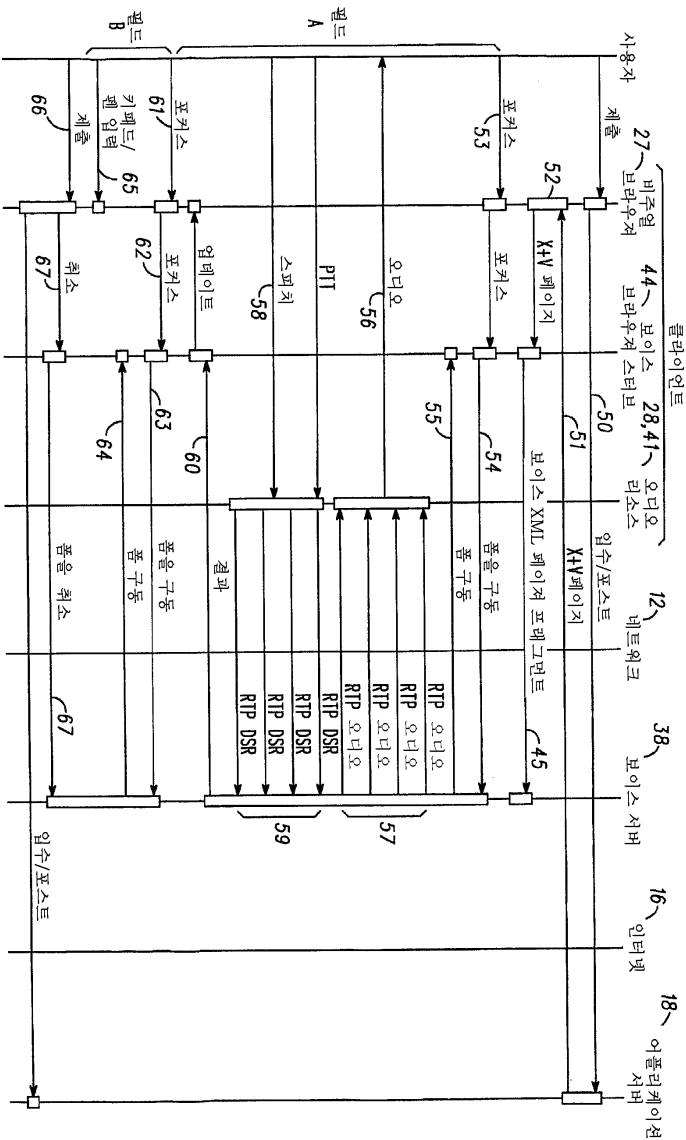
도면3



도면4



도면5



【심사관 직권보정사항】

【직권보정 1】

【보정항목】 청구범위

【보정세부항목】 제1항

【변경전】

복수의 스피치 문법을 다운로드함

【변경후】

복수의 스피치 문법을 다운로드함