

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
26 February 2004 (26.02.2004)

PCT

(10) International Publication Number
WO 2004/017603 A1

(51) International Patent Classification⁷: **H04L 29/06**,
H04M 3/00, H04L 29/08, G10L 15/28, 15/26

(74) Agent: **LOWE, Scott, L.**; Birch, Stewart, Kolasch &
Birch, LLP, P.O. Box 747, Falls Church, VA 22040-0747
(US).

(21) International Application Number:
PCT/US2003/024443

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,
MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC,
SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA,
UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(22) International Filing Date: 5 August 2003 (05.08.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10/218,608 15 August 2002 (15.08.2002) US

(84) Designated States (*regional*): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE,
ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO,
SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM,
GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(71) Applicant: **AVAYA TECHNOLOGY CORPORATION**
[US/US]; 211 Mount Airy Road, Basking Ridge, NJ 07920
(US).

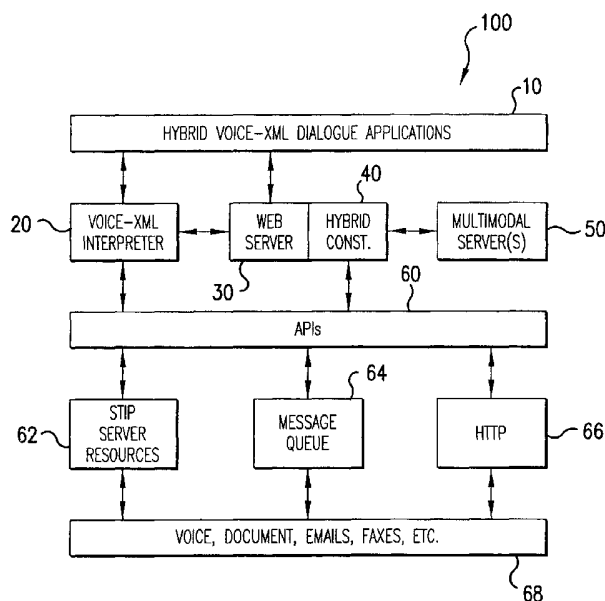
(72) Inventors: **CHOU, Wu**; 57 Sentinel Drive, Basking
Ridge, NJ 07920 (US). **LI, Li**; 331 Tammy Lane, Bride-
water, NJ 08807 (US). **Liu, Feng**; 405B Sherman Avenue,
Raritan, NJ 08869 (US). **SAAD, Antoine**; 3806 Hana
Road, Edison, NJ 08817 (US).

Published:

— with international search report

For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.

(54) Title: DISTRIBUTED MULTIMODAL DIALOGUE SYSTEM AND METHOD



(57) Abstract: A system and method for providing distributed multimodal interaction are provided. The system is a hybrid multi-modal dialogue system that includes one or multiple hybrid constructs to form sequential and joint events in multimodal interaction. It includes as application interface receiving a multimodal interaction request for conducting a multimodal interaction over at least two different modality channels; and at least one hybrid construct communicating with multimodal servers corresponding to the multiple modality channels to execute the multimodal interaction request.

WO 2004/017603 A1

DISTRIBUTED MULTIMODAL DIALOGUE SYSTEM AND METHOD**BACKGROUND OF THE INVENTION****Field of the Invention**

[01] The invention relates to techniques of providing a distributed multimodal dialogue system in which multimodal communications and/or dialogue types can be integrated into one dialogue process or into multiple parallel dialogue processes as desired.

Discussion of the Related Art

[02] Voice Extensible Markup Language or VoiceXML is a standard set by World Wide Web Committee (W3C) and allows users to interact with the Web through voice-recognizing applications. Using VoiceXML, a user can access the Web or application by speaking certain commands through a voice browser or a telephone line. The user interacts with the Web or application by entering commands or data using the user's natural voice. The interaction or dialogue between the user and the system is over a single channel – voice channel. One of the assumptions underlying such VoiceXML-based systems is that a communication between a user and the system through a telephone line follows a single modality communication model where events or communications occur sequentially in time as in a stream line synchronized process.

[03] However, conventional VoiceXML systems using the single modality communication model are not suitable for multimodal interactions where multiple communication processes need to occur in parallel over different modes of communication (modality channels) such as voice, e-mail, fax, web form, etc. More specifically, the single modality communication model of the conventional VoiceXML systems is no longer adequate for use in a multimodal interaction because it follows a stream line synchronous communication model.

[04] In a multimodal interaction system, the following four level hierarchies of various types of multimodal interactions, which cannot be provided by a single streamline modality communication of the related art, would be desired:

(Level 1) Sequential Multimodal Interaction: Although the system would allow multiple modalities or modes of communication, only one modality is active at any given time instant, and two or more modalities are never active simultaneously.

(Level 2) Uncoordinated, Simultaneous Multimodal Interaction: The system would allow a concurrent activation of more than one modality. However, if an input needs to be provided by more than one modality, such inputs are not integrated, but are processed in isolation, in random or specified order.

(Level 3) Coordinated, Simultaneous Multimodal Interaction: The system would allow a concurrent activation of more than one modality for integration and forms joint events based on time stamping or other process synchronization information to combine multiple inputs from multiple modalities.

(Level 4) Collaborative, Information-overlay-based Multimodal Interaction: In addition to Level 3 above, the interaction provided by the system would utilize a common shared multimodal environment (e.g., white board, shared web page, and game console) for multimodal collaboration, thereby allowing collaborative interaction be shared and overlaid on top of each other with the common collaborating environment.

[05] Each level up in the hierarchy above represents a new challenge for dialogue system design and departs farther away from the single modality communication by an existing voice model. Thus, if a multimodal communication is desired, i.e., if interaction through multiple modes of communication is desired, new approaches are needed.

SUMMARY OF THE INVENTION

[06] The present invention provides a method and system for providing distributed multimodal interaction, which overcome the above-identified problems and limitations of the related art. The system of the present invention is a hybrid VoiceXML dialogue system, and includes an application interface receiving a multimodal interaction request for conducting a multimodal interaction over at least two different modality channels; and at least one hybrid construct communicating with multimodal servers corresponding to the multiple modality channels to execute the multimodal interaction request.

[07] Advantages of the present invention will become more apparent from the detailed description given hereinafter. However, it should be understood that the detailed description and specific examples, while indicating preferred embodiments of the invention, are given by way of illustration only, since various changes and modifications within the spirit and scope of the invention will become apparent to those skilled in the art from this detailed description.

BRIEF DESCRIPTION OF THE DRAWINGS

[08] The present invention will become more fully understood from the detailed description given hereinbelow and the accompanying drawings which are given by way of illustration only, and thus do not limit the present invention.

[09] Fig. 1 is a functional block diagram of a system for providing distributed multimodal communications according to an embodiment of the present invention;

[10] Fig. 2 is a more detailed block diagram of a part of the system of Fig. 1 according to an embodiment of the present invention; and

[11] Fig. 3 is a function block diagram of a system for providing distributed multimodal communications according to an embodiment of the present invention, wherein it is adapted for integrating finite-state dialogue and natural language dialogue.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

[12] The use of the term “dialogue” herein is not limited to voice dialogue, but is intended to cover a dialoging or interaction between multiple entities using any modality channel including voice, e-mail, fax, web form, documents, web chat, etc. Same reference numerals are used in the drawings to represent the same or like parts.

[13] Generally, a distributed multimodal dialogue system according to the present invention follows a known three-tier client-server architecture. The first layer of the system is the physical resource tier such as a telephone server, internet protocol (IP) terminal, etc. The second layer of the system is the application program interface (API) tier, which wraps all the physical resources of the first tier as APIs. These APIs are exposed to the third, top-level application tier for dialogue applications. The present invention focuses on the top application layer by modifying it to support multimodal interaction. This configuration provides an extensible and flexible environment for application development so that any new issues, current and potentially future ones, can be addressed without requiring extensive modifications to the existing infrastructure. It also provides sharable cross multiple platforms with reusable and distributed components that are not tied to specific platforms. In this process, although not necessary, VoiceXML may be used as its voice modality if voice dialogue is involved as one of the multiple modalities involved.

[14] Fig. 1 is a functional block diagram of a dialogue system 100 for providing distributed multimodal communications according to an embodiment of the present invention. As shown in Fig. 1, the dialogue system 100 employs components for multimodal interaction including hybrid VoiceXML based dialogue applications 10 for controlling multimodal interactions, a VoiceXML interpreter 20, application program interfaces (APIs) 60, speech technology integration platform (STIP) server resources 62, and a message queue 64, and a server such as HyperText Transfer Protocol (HTTP) server 66. The STIP server resources 62, the message queue 64 and the HTTP 66 receive inputs 68 of various modalities such as voice, documents, e-mails, faxes, web-forms, etc.

The hybrid VoiceXML based dialogue applications 10 are multimodal, multimedia dialogue applications such as multimodal interaction for direction assistance, customer relation management, etc., and the VoiceXML interpreter 20 is a voice browser known in the art. VoiceXML products such as VoiceXML 2.0 System (Interactive Voice Response 9.0) from Avaya Inc. would provide these known components.

[15] The operation of each of the components 20, 60, 62, 64 and 66 is known in the art. For instance, the resources needed to support voice dialogue interactions are provided in the STIP server resources 62. Such resources include, but are not limited to, multiple ports of automatic speech recognition (ASR), text-to-speech engine (TTS), etc. Thus, when a voice dialogue is involved, a voice command from a user would be processed by the STIP server resources 62, e.g., converted into text information. The processed information is then processed (under the dialogue application control and management provided by the dialogue applications 10) through the APIs 60 and VoiceXML interpreter 20. The message queue 64, HTTP 66 and socket or other connections are used to form an interface communication tier to communicate with external devices. These multimodal resources are exposed through the APIs 60 to the application tier of the system (platform) to communicate with the VoiceXML interpreter 20 and the multimodal hybrid-VoiceXML dialogue applications 10.

[16] More importantly, the dialogue system 100 further includes a web server 30, a hybrid construct 40, and multimodal server(s) 50. The hybrid construct 40 is an important part of the dialogue system 100 and allows the platform to integrate distributed multimodal resources which may not physically reside on the platform. In another embodiment, multiple hybrid constructs 40 may be provided to perform sets of multiple multimodal interactions either in parallel or in some sequence, as needed. These components of the system 100, including the hybrid construct(s) 40, are implemented as computer software using known computer programming languages.

[17] Fig. 2 is a more detailed block diagram showing the hybrid construct 40. As shown in Fig. 2, the hybrid construct 40 includes a server page 42 interacting with the web server 30, a plurality of synchronizing modules 44, and a plurality of dialogue agents (DAs) 46 communicating with a plurality of multimodal servers 50. The sever page 42 can be a known server page such as active server page (ASP) or java server page (JSP). The synchronizing modules 44 can be known message queues (e.g., sync threads, etc.) used for asynchronous-type synchronization such as for e-mail processing, or can be function calls known for non-asynchronous type synchronization such as for voice processing.

[18] The multimodal servers 50 include servers capable of communication over different modes of communication (modality channels). The multimodal servers 50 may include, but are not limited to, one or multiple e-mail servers, one or multiple fax servers, one or multiple web-form servers, one or multiple voice servers, etc. The synchronizing modules 44 and the DAs 46 are designated to communicate with the multimodal servers 50 such that the server page 42 has information on which synchronizing module and/or DA should be used to get to a particular type of the multimodal server 50. The server page 42 prestores and/or preassigns this information.

[19] An operation of the dialogue system 100 is as follows.

[20] The system 100 can receive and process different multiple modal communication requests either simultaneously or sequentially in some random or sequenced manner, as needed. For example, the system 100 can conduct multimodal interaction simultaneously using three modalities (three modality channels) – voice channel, email channel and web channel. In this case, a user may use voice (voice channel) to activate other modality communications such as e-mail and web channel, such that the user can begin dialogue actions over the three (voice, e-mail and web) modality channels in a parallel, sequenced or collaborated processing manner.

[21] The system 100 can also allow cross-channel, multimedia multimodal interaction. For instance, a voice interaction response that uses the

voice channel can be converted into text using known automatic speech recognition techniques (e.g., via the ASR of the STIP server resources 62), and can be submitted to a web or email channel through the web server 30 for a web/email channel interaction. The web/email channel interaction can also be converted easily into voice using the TTS of the STIP server resources 62 for the voice channel interaction. These multimodal interactions, including the cross-channel and non cross-channel interactions, can occur simultaneously or in some other manner as requested by a user or according to some preset criteria.

[22] Although a voice channel is one of main modality channels often used by end-users, multimodal interaction that does not include the use of the voice channel is also possible. In such a case, the system 100 would not need to use voice channel and the voice channel related STIP server resources 62, and the hybrid construct 40 would communicate directly with the APIs 60.

[23] In the operation of the system 100 according to one example of application, when the system 100 receives a plurality of different modality communication requests either simultaneously or in some other manner, they would be processed by one or more of the STIP server resources 62, message queue 64, HTTP 66, APIs 60, and VoiceXML interpreter 20, and the multimodal dialogue applications 10 will be launched to control the multimodal interactions. If one of the modalities of this interaction involves voice (voice channel), then STIP server resources 62 and the VoiceXML interpreter 20, under control of the dialogue applications 10, would be used in addition to other components as needed. On the other hand, if none of the modalities of this interaction involves voice, then the components 20 and 62 may not be needed.

[24] The multimodal dialogue applications 10 can communicate interaction requests to the hybrid construct 40 either through the VoiceXML interpreter 20 or through the web server 30 (e.g., if the voice channel is not used). Then the server page 42 of the hybrid construct 40 is activated so that it formats or packs these requests into 'messages' to be processed by the requested multimodal servers 50. A 'message' here is a specially formatted information bearing data packet, and the formatting/packing of the request

involves embedding the appropriate request into a special data packet. The server page 42 then sends these messages simultaneously to the corresponding synchronizing modules 44 depending on the information indicating which synchronizing module 44 is designated to serve a particular modality channel. Then the synchronizing modules 44 may temporarily store the messages and send the messages to the corresponding DAs 46 when they are ready.

[25] When each of the corresponding DAs 46 receives the corresponding message, it unpacks the message to access the request, translates the request into a predetermined proper format recognizable by the corresponding multimodal server 50, and sends the request in the proper format to the corresponding server 50 for interaction. Then each of the corresponding servers 50 receives the request and generates a response to that request. As one example only, if a user orally requested the system to obtain a list of received e-mails pertaining to a particular topic, then the multimodal server 50 which would be an e-mail server, would generate a list of received emails about the requested topic as its response.

[26] Each of the corresponding DAs 46 receives the response from the corresponding multimodal server 50 and converts the response into an XML page using known XML page generation techniques. Then each of the corresponding DAs 46 transmits the XML page with channel ID information to the server page 42 through the corresponding message queues 44. The channel ID information identifies the channel type or modality type that is processed in the corresponding DA 46. Channel ID information identifies a channel ID of each modality which is assigned to each DA as the server page resources. It also identifies the modality type to which the DA is assigned. The modality type may be preassigned and the channel ID numbering can be either preassigned or dynamic as long as the server page 42 keeps an updated record of the channel ID information.

[27] The server page 42 receives all returned information as the response of the multimodal interaction from all related DAs 46. These pieces of the interaction response information, which can be represented in the format of

XML pages, are received with the channel ID information and type of modality it pertains to. The server page 42 then integrates or compiles all the received interaction responses into a joint response or joint event which can also be in the form of a joint XML page. This can be achieved by using the server side scripting or programming to combine and filter the received information from the multiple DAs 46, or by integrating these responses to form a joint multimodal interaction event based on multiple inputs from the different multimodal servers 50. According to another embodiment, the joint event can be formed at the VoiceXML interpreter 20.

[28] The joint response is then communicated to the user or other designated device in accordance with the user's request through known techniques, e.g., via the APIs 60, message queues, HTTP 66, client's server, etc.

[29] The server page 42 also communicates with the dialogue applications 10 (e.g., through the web server 30) to generate new instructions for any follow-up interaction which may accompany the response. If the follow-up interaction involves the voice channel, the server page 42 will generate a new VoiceXML page and make it available to the VoiceXML interpreter 20 through the web server 30, in which the desired interaction through the voice channel is properly described using the corresponding VoiceXML language. The VoiceXML interpreter 20 interprets the new VoiceXML page and instructs the platform to execute the desired voice channel interaction. If the follow-up interaction does not involve the voice channel, then it would be processed by other components such as the message queues 64 and the HTTP 66.

[30] Due to the specific layout of the system 100 or 100a, one of the important features of the hybrid construct 40 is that it can be exposed as a distributed multimodal interaction resource and is not tied to any specific platform. Once it is constructed, it can be hosted and shared by different processes or different platforms.

[31] As an example only, it is discussed below one application of the system 100 to perform email management when two modality channels are

used. In this example, the two modality channels are voice and email. If a user speaks a voice command such as "please open and read my e-mail" into a known client device, then this request from the voice channel is processed at the Application API 60, which in turn communicates this request to the VoiceXML interpreter 20. The VoiceXML interpreter 20 under control of the dialogue applications 10 then recognizes that the current request involves opening a second modality channel (e-mail channel), and submits the email channel request to the web server 30.

[32] The server page 42 is then activated and packages the request with related information (e.g., email account name, etc.) in a message and sends the message through the synchronizing module 44 to one of its email channel DAs 46 to execute it. The e-mail channel DA 46 interacts with the corresponding e-mail server 50 and accesses the requested e-mail content from the e-mail server 50. Once the email content is extracted by the email channel DA 46 as the result of the email channel interaction, the extracted e-mail content is transmitted to the server page 42 through the synchronizing module 44. The server page 42 in turn generates a VoiceXML page which contains the email content as well as the instructions to the VoiceXML interpreter 20 on how to read the e-mail content through the voice channel as a follow-up voice channel interaction. Obviously, this example can be modified or expanded to provide cross-channel multimodal interaction. In such a case, instead of providing instructions to the VoiceXML interpreter 20 on how to read the e-mail content through the voice channel, the server page 42 would provide instructions to send an e-mail to the designated e-mail address which carries the extracted e-mail content. Accordingly, using a single modality (voice channel in this example), multiple modality channels can be activated and used to conduct multimodal interaction of various types.

[33] Fig. 3 shows a diagram of a dialogue system 100a which corresponds to the dialogue system 100 of Fig. 1 that has been applied to integrate natural language dialogue and finite-state dialogue as two modalities according to one embodiment of the present invention. Natural language

dialogue and finite-state dialogue are two different types of dialogues. Existing VoiceXML programs are configured to support only the finite-state dialogue. Finite-state dialogue is a limited computer-recognizable dialogue which must follow certain grammatical sequences or rules for the computer to recognize. On the other hand, natural language dialogue is an everyday dialogue spoken naturally by a user. A more complex computer system and program is needed for machines to recognize the natural language dialogue.

[34] Referring to Fig. 3, the system 100a contains components of the system 100 as indicated by the same reference numerals and thus, these components will not be discussed in detail.

[35] The system 100a is capable of integrating not only multiple different physical modalities but also capable of integrating different interactions or processes as special modalities in a joint multimodal dialogue interaction. In this embodiment, two types of voice dialogues (i.e., finite-state dialogue as defined in VoiceXML and natural language dialogue which is not defined in VoiceXML) are treated as two different modalities. The interaction is through the voice channel but it is a mix of two different types (or modes) of dialogue. When the natural language dialogue is called (e.g., by the oral communication of the user), the system 100a recognizes that a second modality (natural language dialogue) channel needs to be activated. This request is submitted to the web server 30 for the natural language dialogue interaction through the VoiceXML interpreter 20 over the same voice channel used for the finite-state dialogue.

[36] The server page 42 of a hybrid construct 40a packages the request and send it as a message to a natural language call routing DA (NLCR DA) 46a. A NLCR dialogue server 50a receives a response from the designated NLCR DA 46a with follow-up interaction instructions. A new VoiceXML page is then generated that instructs the VoiceXML interpreter 20 to interact according to the NLCR DA 46a. As this process continues, the dialogue control is shifted from VoiceXML to the NLCR DA 46a. The same voice channel and the same VoiceXML interpreter 20 are used to provide both natural language dialogue and finite-state dialogue interactions. But the role has been changed and the

interpreter 20 acts as a slave process controlled and handled by the NLCR DA 46a. In the similar setting, the same approach applies to other generic cases involves multiple modalities and multiple processes.

[37] As one example of implementation, <object> tag extensions can be used to allow the VoiceXML interpreter 20 to recognize the natural language speech. The <object> tag extensions are known VoiceXML programming tools that can be used to add new platform functionalities to the existing VoiceXML system.

[38] The system 100a can be configured such that the finite-state dialogue interaction is the default to the alternative, natural language dialogue interaction. In this case, the system would first engage automatically in the finite-state dialogue interaction mode, until it determines that the received dialogue corresponds to the natural language dialogue and requires the activation of the natural language dialogue interaction mode.

[39] It should be noted that the system 100a can also be integrated into the dialogue system 100 of Fig. 1 such that the natural language dialogue interaction can be one of many multimodal interactions possible by the system 100. For instance, the NLCR DA 46a can be one of the DAs 46 in the system 100, and the NLCR dialogue server 50a can be one of the multimodal servers 50 in the system 100. Other modifications can be made to provide this configuration.

[40] The components of the dialogue systems shown in Figs. 1 and 3 can reside all at a client side, or all at a server side, or across the server and client sides. Further, these components may communicate with each other and/or other devices over known networks such as internet, intranet, extranet, wired network, wireless network, etc. or over any combination of the known networks.

[41] The present invention can be implemented using any known hardware and/or software. Such software may be embodied on any computer-readable medium. Any known computer programming language can be used to implement the present invention.

[42] The invention being thus described, it will be obvious that the same may be varied in many ways. Such variations are not to be regarded as a departure from the spirit and scope of the invention, and all such modifications as would be obvious to one skilled in the art are intended to be included within the scope of the following claims.

WHAT IS CLAIMED:

1. A distributed multimodal interaction system comprising:
an application interface receiving a multimodal interaction request for conducting a multimodal interaction over at least two different modality channels;
and
at least one hybrid construct communicating with multimodal servers corresponding to the modality of channels to execute the multimodal interaction request.
2. The system of claim 1, wherein the system is a hybrid voice extensible markup language (VoiceXML) system including one or multiple hybrid constructs.
3. The system of claim 1, wherein the hybrid construct receives responses to the multimodal interaction request from the multiple modality channels, and compiles a joint event response based on the responses from each individual modality, and transmits the joint event response to the application interface to conduct the multimodal interaction.
4. The system of claim 3, wherein the joint event response is compiled in the form of an extensible markup language (XML) page.
5. The system of claim 1, wherein the at least two modality channels include a voice channel, and the system further comprises an interpreter and a web server for processing voice dialogue over the voice channel.
6. The system of claim 1, wherein the hybrid construct includes:
a server page communicating with the application interface or a voice browser;

at least one synchronizing modules distributing the multimodal interaction request to the appropriate multimodal servers over the different modality channels; and

at least one dialogue agent communicating the multimodal interaction request with the appropriate multimodal servers, receiving the responses from the multimodal servers, and delivering the responses to the server page.

7. The system of claim 1, wherein the at least two modality channels include different types of voice dialogue channels.

8. The system of claim 7, wherein the types of voice dialogue channels include a natural language dialogue channel and a finite-state dialogue channel.

9. The system of claim 1, wherein the at least two modality channels include at least two of the following: voice, e-mail, fax, web-form, and web chat.

10. The system of claim 1, wherein the system conducts the multimodal interaction over at least the two modality channels, simultaneously and in parallel.

11. A method of providing distributed multimodal interaction in a dialogue system, the dialogue system including an application interface and at least one hybrid construct, the method comprising:

receiving, by the application interface, a multimodal interaction request for conducting a multimodal interaction over at least two different modality channels; and

communicating, by the hybrid construct, with multimodal servers corresponding to the modality channels to execute the multimodal interaction request.

12. The method of claim 11, wherein the dialogue system is a hybrid voice extensible markup language (VoiceXML) system with one or multiple hybrid constructs.

13. The method of claim 11, wherein the communicating step includes:
receiving, by the hybrid construct, responses to the multimodal interaction request from the modality channels;
compiling a joint event response based on the responses; and
transmitting the joint event response to the application interface to conduct
the multimodal interaction.

14. The method of claim 13, wherein the joint event response is compiled in the form of an extensible markup language (XML) page.

15. The method of claim 11, wherein the at least two modality channels include a voice channel, and the method further comprises processing voice dialogue over the voice channel.

16. The method of claim 11, wherein the communicating step includes:
communicating by a server page with the application interface or a voice browser;
distributing the multimodal interaction request to the appropriate multimodal servers over the modality channels using at least one synchronizing module; and
communicating the multimodal interaction request with the appropriate multimodal servers using at least one dialogue agent, receiving the responses from the multimodal servers, and delivering the responses to the server page.

17. The method of claim 11, wherein the at least two modality channels include different types of voice dialogue channels.

18. The method of claim 17, wherein the types of voice dialogue channels include a natural language dialogue channel and a finite-state dialogue channel.

19. The method of claim 11, wherein the at least two modality channels include at least two of the following: voice, e-mail, fax, web-form, and web chat.

20. The method of claim 11, wherein the multimodal interaction is conducted over at least the two modality channels, simultaneously and in parallel.

21. A computer program product embodied on computer-readable media, for providing distributed multimodal interaction in a dialogue system, the dialogue system including an application interface and at least one hybrid construct, the computer program product comprising computer-executable instructions for;
receiving, by the application interface, a multimodal interaction request for conducting a multimodal interaction over at least two different modality channels;
and
communicating, by the hybrid construct, with multimodal servers corresponding to the modality channels to execute the multimodal interaction request.

22. The computer program product of claim 21, wherein the dialogue system is a hybrid voice extensible markup language (VoiceXML) system with one or multiple hybrid constructs.

23. The computer program product of claim 21, wherein the computer-executable instructions for communicating include computer-executable instructions for:
receiving, by the hybrid construct, responses to the multimodal interaction request from the modality channels;
compiling a joint event response based on the responses; and

transmitting the joint event response to the application interface to conduct the multimodal interaction.

24. The computer program product of claim 23, wherein the joint event response is compiled in the form of an extensible markup language (XML) page.

25. The computer program product of claim 21, wherein the at least two modality channels include a voice channel, and the computer program product further comprises computer-executable instructions for processing voice dialogue over the voice channel.

26. The computer program product of claim 21, wherein the computer-executable instructions for communicating include computer-executable instructions for:

communicating by a server page with the application interface or a voice browser;

distributing the multimodal interaction request to the appropriate multimodal servers over the modality channels using at least one synchronizing module; and

communicating the multimodal interaction request with the appropriate multimodal servers using at least one dialogue agent, receiving the responses from the multimodal servers, and delivering the responses to the server page.

27. The computer program product of claim 21, wherein the at least two modality channels include different types of voice dialogue channels.

28. The computer program product of claim 27, wherein the types of voice dialogue channels include a natural language dialogue channel and a finite-state dialogue channel.

29. The computer program product of claim 21, wherein the at least two modality channels include at least two of the following: voice, e-mail, fax, web-form, and web chat.

30. The computer program product of claim 21, wherein the multimodal interaction is conducted over at least the two modality channels, simultaneously and in parallel.

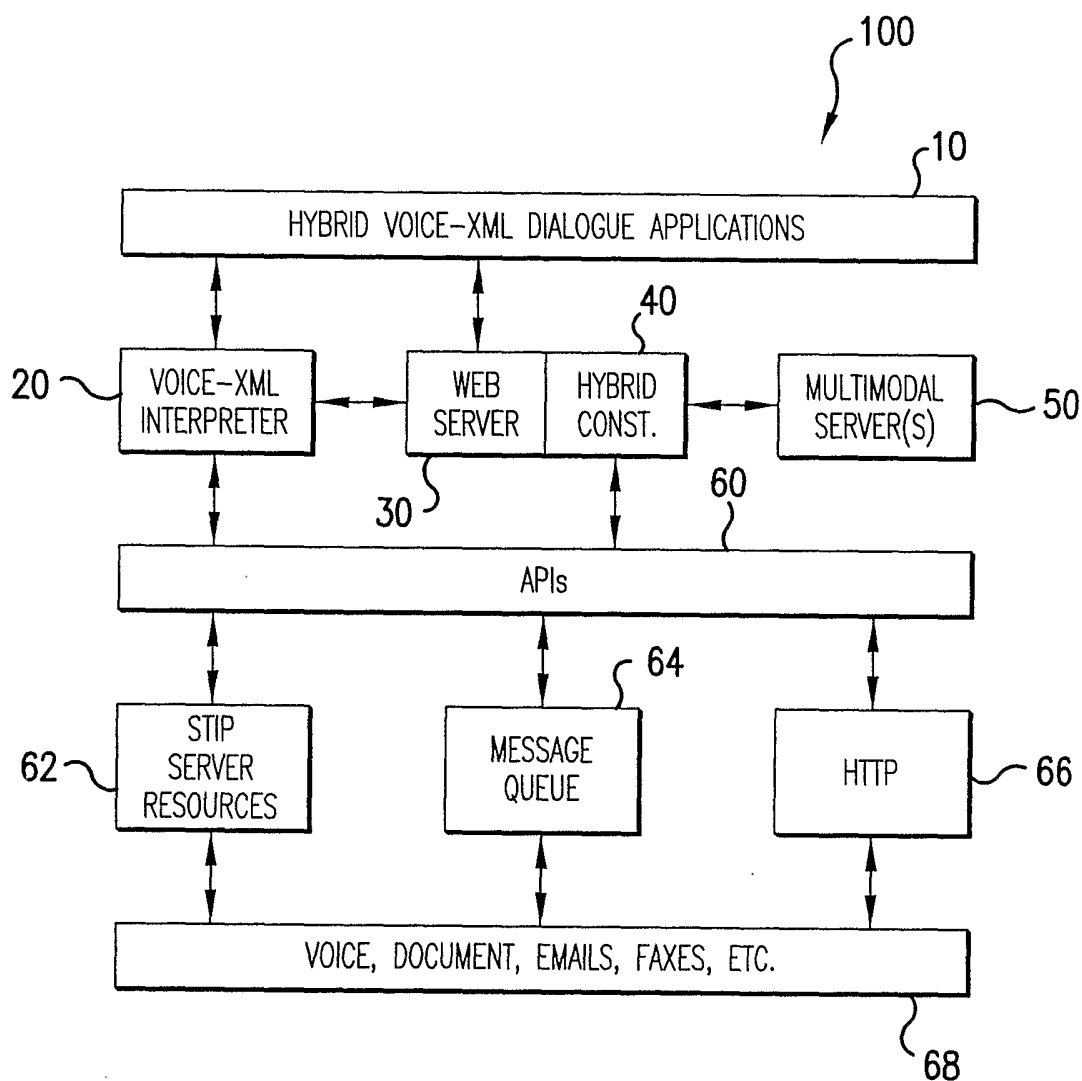


FIG.1

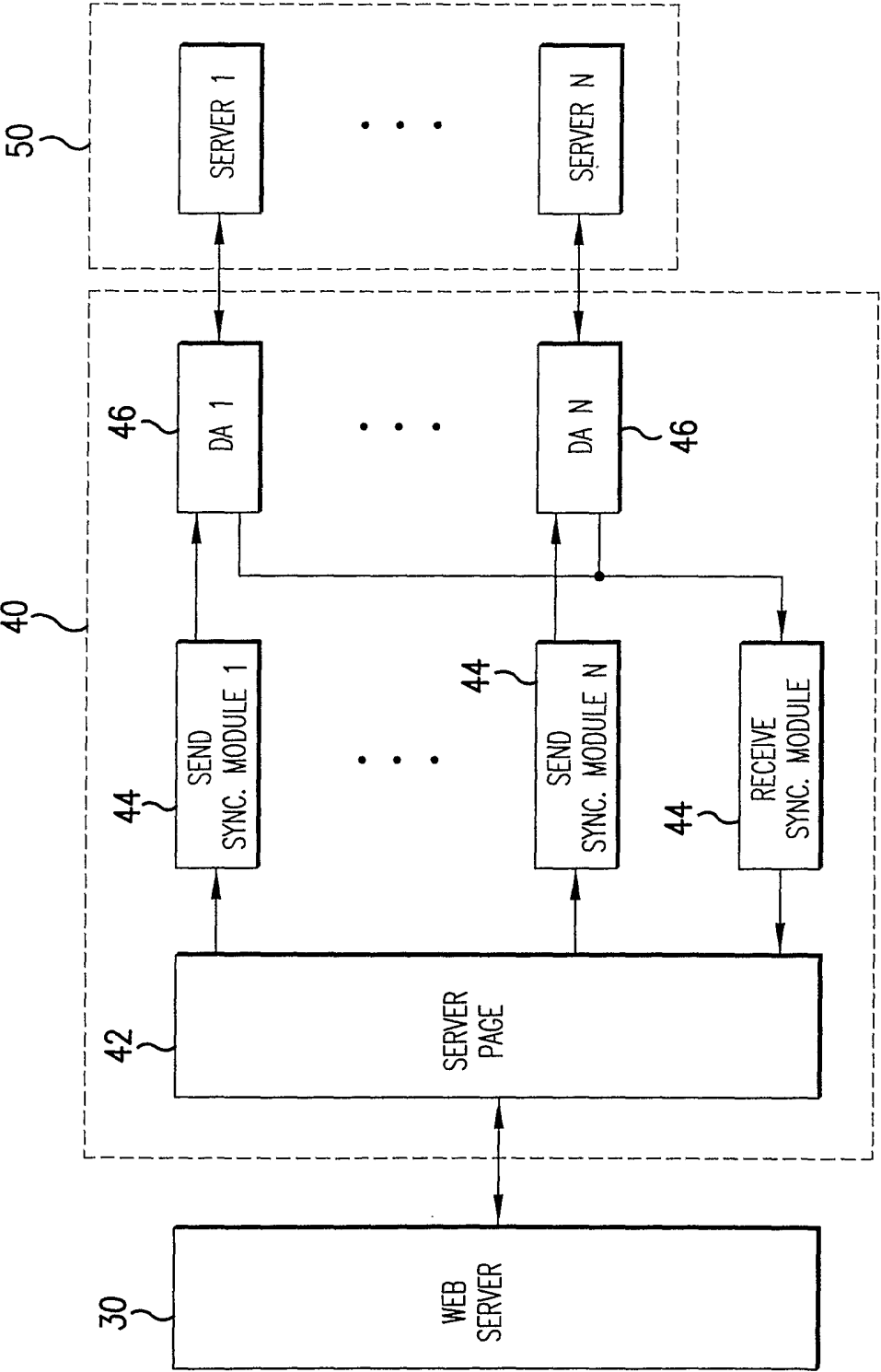


FIG.2

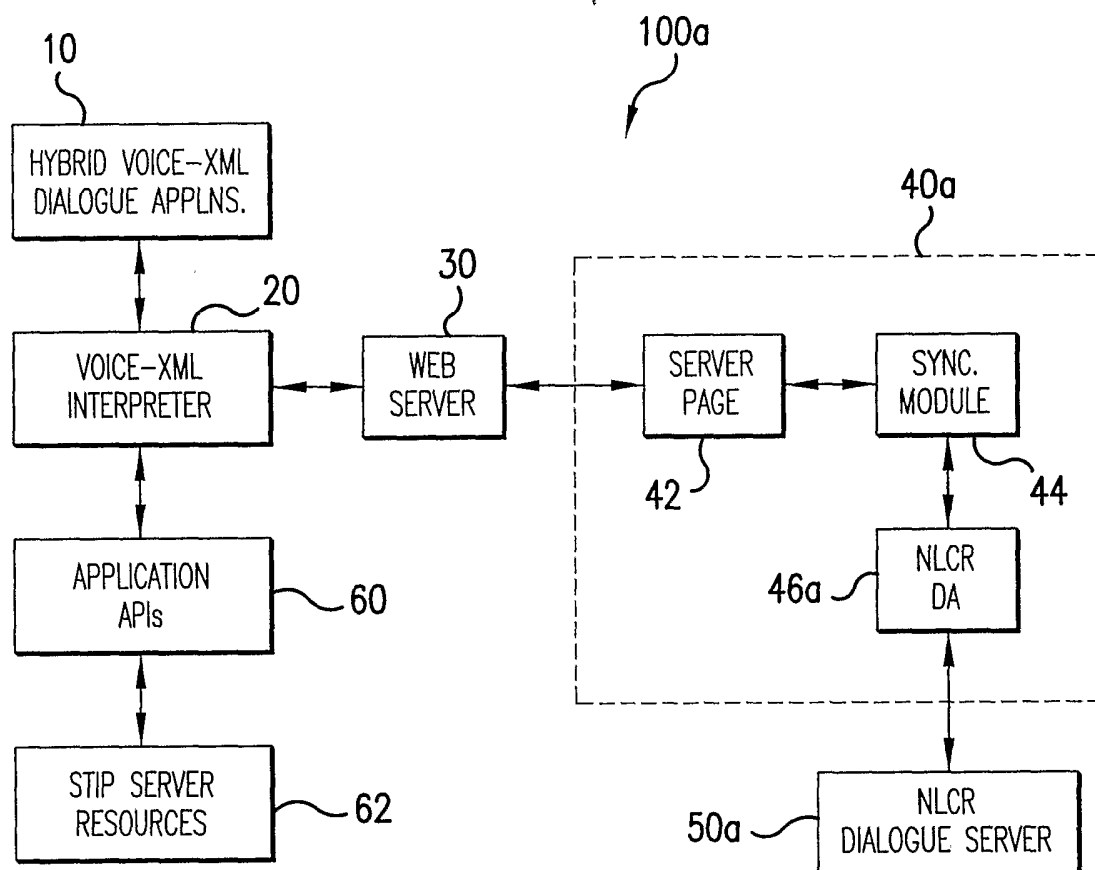


FIG.3

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 03/24443

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 H04L29/06 H04M3/00 H04L29/08 G10L15/28 G10L15/26

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 H04L H04M G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

WPI Data, PAJ, EPO-Internal, INSPEC, COMPENDEX

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	FENG LIU, ANTOINE SAAD, LI LI, WU CHOU: "A Distributed Multimodal Dialogue System based on Dialogue System and Web Convergence" -, 'Online! April 2002 (2002-04), XP002262844 Retrieved from the Internet: <URL:http://www.research.avaya.com/tec hreport/ALR-2002-017-paper.pdf> 'retrieved on 2003-11-25! the whole document --- -/--	1-30

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

26 November 2003

Date of mailing of the international search report

09/12/2003

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Mannekens, J

INTERNATIONAL SEARCH REPORT

Intel I Application No
PCT/US 03/24443

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	INTERNATIONAL BUSINESS MACHINES CORPORATION: "Multi-modal data access" RESEARCH DISCLOSURE, KENNETH MASON PUBLICATIONS, HAMPSHIRE, GB, vol. 426, no. 114, October 1999 (1999-10), XP007124971 ISSN: 0374-4353 the whole document	1-30
A	COHEN P R ET AL: "QUICKSET: MULTIMODAL INTERACTION FOR DISTRIBUTED APPLICATIONS" PROCEEDINGS ACM MULTIMEDIA 97. SEATTLE, NOV. 9 - 13, 1997, READING, ADDISON WESLEY, US, vol. CONF. 5, 9 November 1997 (1997-11-09), pages 31-40, XP000765765 ISBN: 0-201-32232-3 paragraph '0006! - paragraph '06.3!	1-30
A	ABU-HAKIMA S ET AL: "A COMMON MULTI-AGENT TESTBED FOR DIVERSE SEAMLESS PERSONAL INFORMATION NETWORKING APPLICATIONS" IEEE COMMUNICATIONS MAGAZINE, IEEE SERVICE CENTER. PISCATAWAY, N.J, US, vol. 36, no. 7, 1 July 1998 (1998-07-01), pages 68-74, XP000778129 ISSN: 0163-6804 the whole document	1-30
A	ARONS B: "TOOLS FOR BUILDING ASYNCHRONOUS SERVERS TO SUPPORT SPEECH AND AUDIOAPPLICATIONS" UIST '92. 5TH ANNUAL SYMPOSIUM ON USER INTERFACE SOFTWARE AND TECHNOLOGY. PROCEEDINGS OF THE ACM SYMPOSIUM ON USER INTERFACE SOFTWARE AND TECHNOLOGY. MONTEREY, NOV. 15 - 18, 1992, ACM SYMPOSIUM ON USER INTERFACE SOFTWARE AND TECHNOLOGY, NEW YORK, NY:, 15 November 1992 (1992-11-15), pages 71-78, XP000337552 ISBN: 0-89791-549-6 the whole document	1-30
A	EP 1 100 013 A (IBM) 16 May 2001 (2001-05-16) the whole document	1-30

International Application No
PCT/US 03/24443

Form PCT/ISA/210 (patent family annex) (July 1992)