

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>7</sup> :  H04M 3/493		A1	(11) International Publication Number: <b>WO 00/33548</b>
			(43) International Publication Date: 8 June 2000 (08.06.00)
(21) International Application Number: PCT/US99/28480	(22) International Filing Date: 1 December 1999 (01.12.99)	(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, 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, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
(30) Priority Data: 09/203,155 1 December 1998 (01.12.98) US			
(71) Applicant (for all designated States except US): NUANCE COMMUNICATIONS [US/US]; 1380 Willow Road, Menlo Park, CA 94025 (US).			
(72) Inventors; and			
(75) Inventors/Applicants (for US only): COHEN, Michael, H. [US/US]; 2928 Hillside Drive, Burlingame, CA 94010 (US). WAX, Tracy, D. [US/US]; Apartment 174, 39199 Guardino Drive, Fremont, CA 94538 (US).			
(74) Agents: HAVERSTOCK, Thomas, B. et al.; Haverstock & Owens LLP, Suite 420, 260 Sheridan Avenue, Palo Alto, CA 94306 (US).			
(54) Title: SYSTEM FOR AND METHOD OF CREATING AND BROWSING A VOICE WEB			
(57) Abstract			
<p>The present invention allows a user to audibly and interactively browse through a network of audio information, forming a seamless integration of the world wide web and the entire telephone network browsable from any telephone set. Preferably a browser controller (102) allows the user (100) to receive audio information and to transmit verbal instructions. The browser controller (102) links the user (100) to voice pages (108, 112, 114, 116, 118, 120), which can be any telephone station (108, 112, 114, 116) or world wide web page (120), in response to voice commands. Upon linking, certain information is played with an audio indicia which identifies a linking capability. If the user (100) repeats the information set off by the audio indicia, the telephone number or URL of the selected link is transmitted to the browser controller (112). The browser controller (112) establishes a new link with the identified telephone number or URL, and if successful, disconnects the previous link. The originator (100) no longer needs to know of the existence of the receiver nor the telephone number or URL of the receiver because this invention provides a method to browse the entire telephone network and world wide web and to connect to a receiver by saying the name of the hyperlink. This brings the power of the world wide web to the telephone network. In effect, this invention takes the PSTN from its current state as a set of more than 800 million nodes including means to make pairwise connections and converts it to a highly interconnected browsable web, as well as integrating it with the entire world wide web.</p>			

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

**SYSTEM FOR AND METHOD OF  
CREATING AND BROWSING A VOICE WEB**

Field of the Invention

The present invention relates to the field of information systems. More specifically, the present invention relates to the field of interactive voice response systems.

5

Background to the Invention

A variety of services are available over the telephone network. Initially these services required a human operator. With the introduction of touch tone telephones, the caller could make selections and provide information using the telephone buttons. Recent 10 developments have allowed users to make selections and provide information using natural speech. Such an interface in general makes it far easier for the user to gain access to such services. Examples of technology to implement such a voice system are found in U.S. patent application entitled A SYSTEM ARCHITECTURE FOR AND METHOD OF VOICE PROCESSING, serial number 09/039,203, filed on March 31, 1998, and in U.S. 15 patent application entitled METHOD OF ANALYZING DIALOGS IN A NATURAL LANGUAGE SPEECH RECOGNITION SYSTEM, serial number 09/105,837, filed on June 26, 1998, and also in provisional patent application entitled A METHOD AND APPARATUS FOR PROCESSING AND INTERPRETING NATURAL LANGUAGE IN A VOICE ACTIVATED APPLICATION serial number 60/091,047, filed on June 29, 20 1998. These three patent documents are incorporated in their entireties herein by reference.

With the advent of natural language recognition systems, users could respond to interactive telephone systems using more natural spoken responses. Such systems are used

for a variety of applications and are known as interactive voice response (IVR) systems.

One known example is for providing information and services regarding flight availability,

flight times, flight reservations and the like for a predetermined airline. Another well

known use for such systems includes gaining information regarding stocks, bonds and other

5 securities, purchasing and selling such securities, and gaining information regarding a

user's stock account. Also, systems exist for controlling transactions in accounts at a bank.

Other applications are also available.

While using such systems provides dramatic improvement over other voice

information and voice services systems, there are still drawbacks. Each such system

10 accessed by a user requires that the user make a separate telephone call. Often,

information exists on related topics. For example, in the event a user contacts a voice

service to obtain airline information and travel tickets, they may also desire a hotel room

and dinner reservations in the destination city. Even if hotels are located in the destination

city that provide a voice system of room rate and availability information and allow callers

15 to reserve rooms automatically or manually, the user must hang up the telephone call

during which they made airline reservations, somehow discover the telephone number for a

hotel in the destination city and only then place the desired call. This procedure is

cumbersome at best. The procedure can be dangerous when undertaken from an

automobile in commute hour traffic.

20 Other automatic information and service systems are also available. The World

Wide Web (also known as and hereinafter referred to as the "Internet") is a rapidly

expanding network of computers which provide users with numerous services and a wealth

of information. Unlike the voice systems discussed above, the Internet is primarily a

visually based system which allows a user to graphically interact with an image or series of

images on a display screen.

The Internet was originally created as a non-commercial venue to provide communication links between government institutions as well as institutions of higher learning. Today, the Internet has evolved to become a universal network of computers which now includes private industry as well as government institutions. The Internet has become accessible to many people from computers located in their homes, offices, or public libraries. People are able to locate updated information regarding the weather, stock prices, news and many other topics. Further, people are able to locate a wide variety of information regarding products and services.

The Internet offers many advantages over other media. The Internet seamlessly links information stored on geographically distant servers together. Thus, users are capable of seamlessly accessing information stored on geographically distant servers. Similarly, the information on a server can be remotely updated from any geographic point that has access to the Internet.

When the user accesses information on a server, the user interfaces with the server through a website. Many websites offer hyperlinks to other websites, which makes the Internet user-friendly. When a current website has a hyperlink to another website, the user is enabled to jump directly from a current website to this other website without entering an address of this other website. In use, a hyperlink is a visually discernable notation. The user activates the hyperlink by "clicking" on the hyperlink notation or icon also called point-and-click. The user's computer is programmed to automatically access the website identified by the hyperlink as a result of the user's point-and-click operation.

Unfortunately, Internet techniques are not readily applicable to a voice system. In a visual Internet system the graphical image remains on the display screen until changed by

the user. This allows the user ample opportunity to carefully read all the images on the display screen as many times as desired before making an appropriate point-and-click choice. With a voice system, once the message is spoken it cannot be readily reviewed by the user. Thus, there is no previously known analogous operation to point-and-click in a voice system. Further, hyperlinking is not available for voice systems. Telephone calls are made through the central office on a call-by-call basis. In contrast, in the Internet, once connected computers are functionally connected to all Internet addresses concurrently. 5 Different sites are accessed by requesting information which is located at different addresses. At least these differences make ordinary Internet techniques inapplicable to a 10 voice system. What is needed is a system for browsing an audio network.

The PSTN (Public Switched Telephone Network) provides means for more than 800 million individual 'stations' to make any pairwise connection by one party (the originator) dialing the telephone number of another party (the receiver). A station can be any person with a telephone, an IVR system or an information service among others. The current 15 approach has two disadvantages. First, the originator must know of the existence of the receiver. There is no easy way to browse or discover information or receivers that may be of interest to the originator. Second, the originator must know the telephone number of the receiver. Furthermore, from the telephone there is no convenient way to browse web pages that may or may not be audio enabled. Additionally, there is no integration between the 20 PSTN and the world wide web that would allow seamless browsing of both as an integrated web.

Summary of the Invention

The present invention is a system and method for allowing a user to audibly and interactively browse through a network of audio information. This system and method preferably includes a browser controller which allows the user to receive audio information and to transmit verbal instructions. The browser controller preferably links the user to telephone stations, voice-enabled world wide web pages, and regular world wide web pages, in response to voice commands. Upon linking to a voice page, telephone station or world wide web, certain information is played with an audio indication of linking capability of that information. For example, the information could be set apart by using earcons, which consist of special sounds played before and after the link text. Other audio indications besides an earcon may be used, such as speaking the link in a different voice than the main text, setting the link text apart from the surrounding text with pauses, or playing a background sound mixed in with the link. If the user repeats the information set off by the earcon or other means, the voice page transmits the telephone number or the URL of the selected link to the browser controller. The browser controller establishes a new link with the identified new telephone number or world wide web page. If the new link is made successfully, the first call is dropped.

The present invention overcomes the limitations of requiring the originator to know of the existence of the receiver as well as the telephone number or URL of the receiver by providing a method to browse the entire telephone network, as well as to audio browse the world wide web, and to connect to a receiver by saying the name of the hyperlink. This brings the power of the world wide web to the telephone network. In effect, this invention takes the PSTN from its current state as a set of more than 800 million nodes including means to make pairwise connections and converts it to a highly interconnected browsable

web. In addition, this invention integrates the entire PSTN with the entire world wide web into one large audio-browsable network.

It is an object of the present invention to provide a system which allows a user to audibly browse an audio network.

5

#### Brief Description of the Drawings

The following drawings illustrate an embodiment of the present invention. Other embodiments are possible and are described herein.

Figure 1 illustrates an overall block diagram of the present invention.

10 Figure 2 shows a flow chart of the operation of the browser controller of the present invention.

Figure 3 shows a flow chart of the operation of the execute link block of Figure 2.

Figure 4 shows a flow chart of the operation of the return from link block of Figure 2.

15 Figure 5 shows a flow chart of the operation of the voice page of the present invention.

Figure 6 shows simplified block diagram of a speech recognition and interpretation system of the present invention.

20 Detailed Description of the Preferred Embodiment

The present invention is directed toward a voice activated system that is designed to allow a user to request, navigate, retrieve, and store information from a network of telephone stations, "interactive voice response" (IVR) stations voice-enabled world wide web pages and regular world wide web pages. This complete set of telephone numbers,

and URLs is referred to as voice pages. Voice pages that have been designed to operate cooperatively with the present invention as well as all regular world wide web pages can take advantage of the features of the present invention, including hyperlinking.

Conventional telephone stations or currently existing IVR stations cannot take advantage of hyperlinking, but other browsing capabilities will still be available. These voice pages will link to other voice pages to form a network. All conventional telephone stations are also part of this network. The various voice web pages are capable of being linked together to form the pseudo-network of voice web pages by having a browser which can connect to any voice web page, be it either a telephone number on the PSTN or a URL on the world wide web.

The present invention contemplates several principal uses of a system incorporating these teachings. First, is contemplated that users will generate specific applications which include voice pages or world wide web pages which are specially configured to take advantage of the voice browsing capacity of the present invention. In addition, the present invention also includes the capability to allow access to the more than 800 million already existing conventional telephone nodes plus voice pages in addition to world wide web pages which can be read by a text-to-speech converter. This is true regardless of whether or not the intended recipient of the call is aware of the existence of the present invention.

Thus, a user can access any of a plurality of known telephone numbers, IVRs, voice information services or world wide web pages which were not designed to take advantage of the present invention and a user can also access IVR voice pages or world wide web pages designed with the present invention in mind. When accessing a voice page designed according to the present invention, the user will be presented with hyperlink options to other telephone numbers or URLs. When accessing other types of telephone

numbers including a conventional telephone set, an IVR or a voice information service no hyperlink capability will be present. Nevertheless, certain features of the present invention will still be available to the caller. For example, the caller can return to their start page, return to a previous page or visit a bookmarked voice page.

5       With the present invention, the user could access a personal start page by dialing a telephone number assigned to that browser controller. The browser controller can connect sequentially to each desired voice page. The browser controller preferably maintains its connection to the user, makes connections to the desired voice page and joins those calls together. This allows the browser controller to monitor the calls between the user and each  
10      voice page. If a current voice page contains an audiolink to another voice page which is selected by the user, the browser controller makes a connection to the selected voice page, and then if this connection is successful, severs the call to the current voice page, all the while maintaining its connection to the user. The voice activated browser controller system thus allows the user to scout information or services available on a collection of available  
15      voice pages. The preferred embodiment of the present invention includes a browser controller and a variety of voice pages. The user is capable of directing the browser controller via the user's voice which frees the user's hands for other tasks such as driving. Further, the present invention relays information to the user via an audio source. The browser controller is also configured to contact any of the voice pages.

20       Figure 1 shows an exemplary network which incorporates the preferred embodiment of the present invention. This representation of the present invention is not intended to be limited to a specific number or type of system. A user can access the system using any conventional telephone 100 system including a stand-alone analog telephone, a digital telephone, a node on a PBX and alike. The system includes a browser controller 102

which is accessed by the user using their conventional telephone 100. It is anticipated that the browser controller 102 will be provided as a service to many users requiring similar access thereto. Corporations could also use this invention to provide voice access to their websites for their customers, also potentially providing a linking ability to those sites to 5 which they choose to allow such linking. Thus, typical users will access the browser controller 102 via the PSTN 104. However, in certain corporate or institutional environments a browser controller 102 may be made available to the users at the institution, such as in a PBX, thereby eliminating the need to connect the user's conventional telephone 100 to the browser controller 102 via the PSTN 104 and instead 10 allowing a direct connection. Additionally, the browser controller 102 could be implemented in hardware and/or software in a personal computer, which also eliminates the need to connect the user's conventional telephone 100 to the browser controller 102 via the PSTN 104 and instead allowing a direct connection via the internet.

The browser controller 102 includes a pointer to a start page 106 for each user of 15 the system. The start page 106 can operate in conjunction with the browser controller 102, being a personal home page for the user. The start page could also be any voice page on the voice web. The browser controller 102 possesses a static grammar to assist the user with navigation and other browser functions. In addition the browser controller 102 generates a dynamic grammar of the browsable links on each voice page that is visited. 20 The browser controller 102 also includes a number of dynamic grammars which are modified according to the needs of each particular user. These grammars will be described in more detail below.

An originating user can place a telephone call using their conventional telephone 100 directly to a second conventional telephone 108 of another user 110 via the PSTN 104

in the usual way. There is nothing unusual in this and is conventionally done. In the alternative, using the present invention the originating user places a telephone call using their conventional telephone 100 to the browser controller 102. Once the link is established to the browser controller 102, the originating user is recognized and then 5 instructs the browser controller 102 to place the call to the conventional telephone 108 via the PSTN 104. The originating user can be recognized using any known method. The browser controller 102 dials the telephone number of the second conventional telephone 108 to establish the link. The originating user is linked to the receiving user through the browser controller 102. In this way, the browser controller 102 has two links via the 10 PSTN 104, one the originating user and one to the receiving user.

This linking through the browser controller 102 allows the originating user advantages over a conventional telephone call. The browser controller 102 includes a natural language speech recognition engine to 'listen' to the originating user. Each originating user speaks a known assigned 'browser wake-up' word to provide commands to 15 the browser controller 102. The browser wake-up word is preferably not a commonly used word. Under certain circumstances a user may select their own browser wake-up word, but this is not preferred. When the browser controller 102 recognizes the browser wake-up word spoken by the originating user, the browser reverts to a command mode, to be discussed in more detail below. The browser controller 102 can be configured to simply 20 wait for a command subsequent to the browser wake-up word, or the browser controller 102 can respond by saying, 'How can I help you?', for example. Depending upon the nature of the command, the link to the receiving user can be maintained or severed. Other calls can then be placed, such as those described below.

The originating user can establish other types of communications links using the browser controller 102. For example, the originating user might desire to receive audio information such as the time or weather. Other types of audio information are also available. As is well known, the originating user can dial the information service directly using the telephone number for that service. In the alternative, the originating user can dial the browser controller 102 and instruct it to call a predetermined information service 112. Once the call is established to the information service 112 by the browser controller 102, the originating user hears the desired information in the usual way. At any time, the originating user recites the browser wake-up word and can sever the call the information service 112 and have the browser controller 102 place another call.

The originating user can use the browser controller 102 to call an IVR system 114 which only recognizes dtmf tones. Once the browser controller 102 connects the originating user to the IVR dtmf system 114, the user can extract or provide information as necessary using the key pad on the conventional telephone 100. Upon concluding the desired transaction or communication, or at any time the user speaks the browser wake-up word, the originating user states the browser wake-up word, control is returned to the browser controller 102. Thereafter, the connection the IVR dtmf system 114 can be severed or reasserted.

Similarly, the originating user can use the browser controller 102 to call an IVR system 116 which includes a natural language speech recognitions system. Once the browser controller 102 connects the originating user to the IVR speech system 116, the user can extract or provide information as necessary using natural language. Upon concluding the desired transaction or communication or at any time the user speaks the browser wake-up word, the originating user states the browser wake-up word and control is

returned to the browser controller 102. Thereafter, the connection the IVR dtmf system 116 can be severed or reasserted. For example, the user may speak the browser wake-up word to return control to the browser controller 102, but still desire to return to the present 5 telephone link. Once the control is returned to the browser controller 102 any appropriate action can be taken. As one example, the user could request that a bookmark be made for the present page. Then, upon uttering the appropriate command the browser controller 102 returns the user to the pending link. All of the links described above are accessed via the PSTN using a conventional telephone number to initiate contact.

As another example, the originating user can use the browser controller 102 to call 10 a page on the world wide web which has been audio-enabled and voice-enabled. The configuration of such a voice-enabled world wide web page 120 will vary according to the needs or desires of the developer. For example, the developer could include an ability to determine whether contact originated via the computer world wide web or from the PSTN and audio web. The developer can configure the page to include voice content in much the 15 same way as a voice page 118. Any hyperlink that exists on the world wide web page could be identified by an audio indication in the way described herein. This hyperlinking could also take place in a world wide web page that has not been voice-enabled. The browser controller 102 could include a text to speech converter and read the contents of the page to the originating user which has made contact with the browser controller 102 via the 20 PSTN. The same audio indications can be used to indicate the hyperlinks on the world wide web page.

Unlike the other links described above, a conventional page on the world wide web is not accessed using a telephone number over the PSTN. Rather, a page on the world wide web is accessed using an internet address. Further, with the internet world wide web

information is generally transferred directly using a conventional protocol such as http.

Communications via the internet are generally not carried out using data signals exchanged between pairs of modems over the PSTN. This is true in spite of the fact that many users access the internet via an ISP (internet service provider). The communications between the user and the ISP are carried out using data signals exchanged between a pair of modems; however, the communications for the same information transaction from the ISP to a site on the internet is carried out using an internet protocol such as TCP/IP or HTTP.

For at least this reason, without a direct internet connection, the browser controller 102 described above cannot interact directly with a conventional page on the world wide web. To overcome this obstacle, the browser controller 102 preferably includes a secondary dedicated internet connection to interface the browser controller 102 to the internet. The browser controller 102 is configured to bi-directionally communicate data between the browser controller 102 and the internet. Additionally, the browser controller 102 is also configured as a gateway to bi-directionally couple audio voice information between the user through the PSTN on the one hand, and the world wide web page via the internet on the other hand. As an alternative, each server that serves world wide web pages that include voice information configured to interact with a browser controller 102 according to the present invention could be configured to have access by telephone and a modem and further include its own gateway. However, it is clear that such a construction would require considerable duplication of equipment and software across all of the appropriate world wide web servers and pages.

A further alternative would entail providing PSTN access to world wide web pages. This approach overcomes the well known latency problems with the internet. As internet latency issues are resolved, this approach will become even less desirable.

As another approach to providing access from the browser controller 102 to a world wide web page or IVR system, one can include an interface using the so-called I.P. telephony protocols. As is well known, I.P. telephony allows simultaneous transmission of both voice and digital data. Alternatively, a parallel telephone line and internet connection 5 can be provided to emulate I.P. telephony. Yet another alternative allows the use of XML or another similar voice/data protocol (such as the Motorola® VoxML or Microsoft's HTML extensions) to provide internet access to a PSTN application such as the browser controller 102.

As clear from the discussion above, all of the features of the present invention, 10 except for hyperlinking, can be utilized even when accessing conventional telephony services. This provides the originating user access to the existing more than 800 million telephone numbers by using the improved features of the present invention. The full power of the present invention can be achieved by connecting to a voice page 118 specifically designed to accommodate all the advantages of this invention, including hyperlinking as 15 defined herein. The voice page 118 can be formed on an IVR system or as a world wide web page. As information is presented to the originating user certain voice items are specially identified to the user. For example, a particular audio information segment is configured to inform the user of the latest stock prices by stating, "The current trading price of <Apple> is \$xx.xx dollars per share. The current trading price of <IBM> is 20 \$yy.yy dollars per share." The "less than" character ("<") represents an audible beginning marker, such as an earcon (defined below), to inform the user that a custom audiolink is beginning. Similarly, "greater than" character (">") represents an audible ending marker to inform the user that a custom audiolink is ending. Following this example, if the user wanted to learn more about the company Apple, then the user is able to say, "Apple". If

limited to the above audio information segment and the user wanted to know more about "trading price", by verbally saying, "trading price," the user would not receive details on "trading price" because there is no audiolink to "trading price." The user would know that "Apple" is a valid audiolink, because the user can hear the beginning marker ("<") before 5 the word "Apple" is read and can hear the ending marker (">") after the word "Apple". By way of example, the beginning marker ("<") can be represented as a sequence of three rising audible notes arranged such that the pitch for each note rises in the sequence. Additionally, the ending marker (">") can be represented as a sequence of three falling audible notes arranged such that the pitch for each note falls in the sequence. The term 10 "earcon" is used for this process of audibly marking a custom grammar audiolink. The previous example is merely for demonstrative purposes only and should not be construed to limit the scope of the present invention. It will be apparent to those skilled in the art that there is no meaningful way to limit the number of ways to audibly mark a custom grammar audiolink. For example, the text of an audio link could be spoken in a different 15 voice, a background sound could be mixed with the audio link, or the text of the audio link could be surrounded by pauses.

When the browser controller 102 hears the originating user repeat an audiolink, a new telephone number is dialed or a world wide web page is accessed in accordance with the repeated audiolink. If successful, the connection with the currently accessed voice page 20 118 is severed. The browser controller 102 knows the telephone number or the world wide web URL corresponding to the repeated audiolink, because that information was transmitted by the voice page 118 to the browser controller 102. There is sufficient bandwidth even on the PSTN 104 to allow such information to be transmitted between such equipment, transparently to the originating user and without any loss of ordinary

speech functionality or quality.

Another type of connection can be established to a world wide web page which includes audio capability, or plain text that can be read via a text-to-speech converter. Such a page is configured to provide either graphical data, voice data or both, depending upon the type of equipment that accesses it. In this way, links can be shown as hypertext links in the usual way or as voice links with an earcon, or other audio indication, or both. Certain links will only be available to a computer user logged onto the internet and only provide a graphical information. Such links will not be presented with an earcon to the originating user of the present invention. Other links will only be to voice services and will only provide audio information. Such links will not be presented with a hypertext link on the graphical world wide web page. Still other links will be to a data provider that offers both graphical and audio data and both an audio indication and a hypertext link will be available.

The link to the world wide web page can be made by the browser controller 102 through a modem via the PSTN or via a gateway as is well known, though clearly a gateway is preferable. In either case such connections are utilized to provide the advantages of the present invention.

The originating user can perform a number of functions by using the browser controller 102. All originating users will have a predetermined suite of functions and commands available to them upon connection to their respective browser controller 102. Such functions are listed in Table 1, below. This list is exemplary and an implementation of the present invention can include more or less commands.

## GRAMMARS

Residing on Browser/Controller

	<u>Static</u>	<u>Dynamic</u>
5		
	next page	bookmark 1
	previous page	bookmark 2
	go back	
	go home	
10	go to my start page	bookmark n
	what are my choices	telephone number 1
	help	telephone number 2
	where am I	
	add this to my bookmarks	
15	delete this from my bookmarks	telephone number n
	go to my bookmarks	preference 1
	go to bookmark _____	preference 2
	search	
	personal information	
20		preference n

telephone number n

TABLE 1

5

Further, each originating user could develop a series of individual tasks for their browser controller 102 to perform which would be stored on their personal start page. Such dynamic information allows the originating user to make calls or connect to known services without having to remember telephone numbers. For example, while driving to work, an originating user could access their browser controller 102 and state the command 'weather'. The browser controller 102 will then dial the number it knows for the local weather report and allow the user to listen to the report. The browser controller 102 will maintain the connection until it 'hears' the browser wake-up word. Upon hearing the browser wake-up word, the browser controller 102 waits for a command. Our sample originating user then asks for her stock list. The connection to the weather report is severed and a new connection is established to the service that gives stock information. That connection is maintained until the browser controller 102 again hears the browser wake-up word. Our sample originating user then commands 'call mom'. Whereupon the browser controller 102 severs the connection to the stock list and dials the desired person. Our sample originating user concludes her call and then accesses a voice page 118 news report. During an advertisement, an audio indication announces a link to a local <restaurant>. Our sample originating user then says the name of the <restaurant>. The browser controller 102 automatically connects our sample originating user to the restaurant,

then disconnects the present call, and then the originating user makes a lunch reservation.

All these communication transactions occurred without our sample originating user having to dial any number except the first call to the browser controller 102. Further, she accessed both conventional telephones, IVRs, audio information services and voice pages in a single call to her browser controller 102.

5 There will be a set of static and dynamic grammars that will be active on each voice page 118. Depending on the implementation, voice recognition for the items in these grammars could reside as part of either the browser controller 102 or the voice page 118.

Table 2 sets forth what these grammars might be. It is clear to anyone involved in the art

10 that more or less items can be included in these grammars.

## GRAMMARS

Active on Voice Pages

	<u>Static</u>	<u>Dynamic</u>
5		
	help	dynamic links
	what are my choices	
	static links	

10

TABLE 2

15

There are dynamic grammars in the voice page because certain items may change periodically. For example, on a news voice page it is recognized that the news changes continually. The news reports will contain audio links to other voice pages, telephone numbers or audio information services and the like which correspond to the news report. Thus, these links will necessarily be dynamic. Either the voice page 118 or the browser controller 102 will generate the dynamic grammar links. For example, if the voice page 118 is a world wide web page, then the dynamic grammar will be generated by the text of the links that are denoted by the audio cues such as earcons.

20

Figure 2 shows a flow chart of the operation of the browser controller 102 (Figure 1). The originating user calls the browser controller 102 which identifies the caller using any known method in block 200. Once the originating user is identified, the browser controller 102 may load the start page 106 (Figure 1) for that originating user in block 202. The browser controller 102 executes a dialog with the originating user in order to receive a

command. For example, the prompt "Hello, Steve. How may I help you?" could be played. Depending upon this interaction including commands replied by the originating user, an activity will be performed. For example, if the originating user says, preferences 206, the browser controller 102 initiates a program related to their preferences. The 5 originating user can then command add 208, delete 210 or list 212 to execute those functions. Upon concluding this activity, the browser controller 102 returns to the execute dialog block 204 and a new dialog exchange occurs between the originating user and the browser controller 102.

As another example, the user could command 'bookmarks' 214. The originating 10 user can then command add 216, delete 218 or go to 220 to execute those functions. Upon concluding this activity, the browser controller 102 returns to the execute dialog block 204 and a new dialog exchange occurs between the originating user and the browser controller 102. In the alternative, the originating user could provide a 'go to' command or request an 15 audio link which requires making a new telephone call. The browser controller 102, then enters the execute link block 222. This operation couples the originating user to another telephone number or world wide web page via the browser controller 102. Upon completion of the link the browser controller 102 will return to the execute dialog block 204 via the return from link block 224.

From the execute dialog block 204 the originating user can instruct the browser 20 controller 102 to replay the originating user's start page. If no audio link is recited, the control is returned to the execute dialog block 204. If an audio link is recited, the execute link block 222 makes the appropriate connection. As mentioned before, the audio link could be set apart from the rest of the voice page by earcons; however, there are also other means for distinguishing an audio link.

The originating user can instruct the browser controller 102 in the execute dialog block to perform a search of voice pages 118. The search request can be performed by an appropriate search engine. Finally, the interaction can be concluded and the call placed on hook in the exit block 226.

5 Figure 3 shows a flow chart of the operation of the execute link block 222 (Figure 2).

A list is maintained of the calls placed during the session. This allows the originating user to return to a previous call. Once the execute link block 222 is entered the forward/backward list is updated in the block 300 with the link information communicated with the command to execute link. The call is made to the link telephone number in the block 302. The call is connected to the desired telephone number in the block 304.

10 Thereafter, while the call is in progress, the browser controller 102 (Figure 1) listens for either dtmf or the browser wake-up word in the block 306. If a dtmf command is executed in block 308, the link is disconnected in the block 310, the forward/backward list is updated in the block 300 and a new call is made as before in the block 302. As an alternative to dtmf, as mentioned before, the browser controller's 102 telephone station and the voice page's telephone station could communicate via I.P. telephony or could include a parallel internet connection to emulate I.P. telephony. In this case, rather than using dtmf, the destination telephone number or world wide web URL could be communicated over

15 this data channel. Furthermore, additional information such as the state of the user's interaction may be communicated. If the browser wake-up word is heard in the block 306, the recognize command block 312 identifies the command which is executed in the execute command block 314. If the command is not for a new link, control returns to the block 306 to continuing listening for dtmf or the browser wake-up word. If the command is for

a new link, the current link is disconnected in the block 316, the link is disconnected in the block 310, the forward/backward list is updated in the block 300 and a new call is made as before in the block 302. Or, instead of making a call, a world wide web page could be downloaded off of the internet.

5       Figure 4 shows a flow chart of the operation of the return from link block 224 (Figure 2). First, the telephone call is disconnected in the disconnect from link block 400. Then, the forward/backward list is updated in the update forward/backward list block 402.

Figure 5 shows a flow chart of the operation of the voice page 118 (Figure 1). Upon being accessed, the voice page 118 plays audio text or prompts in the play text block 500. The prompts can include a link names or a list of link names. The speech of the originating user is recognized in the recognition block 502. Upon recognizing a command, an action is undertaken in the action block 504. If the action was stating the name of a hyperlink, the telephone number for that link is dtmf transferred to the browser controller 102 (Figure 1) in the block 506. Alternatively, the link could be communicated to the browser controller 102 via I.P. telephony or an internet connection, as shown in the block 506. Thereafter, the voice page 118 is exited in the block 508 to return control to the browser controller 102. If the action was not a link and not the browser wake-up word, then the voice page 118 returns to the play text block 500. If the action was the browser wake-up word, control is returned to the browser controller 102 in the block 510. The line 10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60  
65  
70  
75  
80  
85  
90  
95  
is maintained in an on-hold condition in the block 512. If the browser controller 102 returns control to the voice page 118, the operation returns to the play text block 500. If the browser controller 102 cuts the link, then the voice page 118 exits in the block 514.

Figure 6 shows a flow chart of the operation of the recognize and interpret steps in the browser controller 102 (Figure 1) and the voice page 118 (Figure 1). A memory 600

includes a dictionary memory 602, an acoustic model memory 604 and a grammar, and their pronunciations memory 606. The dictionary memory 602 contains all the words in the grammar. The acoustic model memory 604 contains all the statistical models of all phonetic units that make up the words. An input signal 608 of digitized speech is input to 5 a front end analysis module 610. The front end analysis module 610 separates feature vectors from the digitized speech, each covering a predetermined length of speech. In the preferred embodiment, a feature vector is output for each 10 mS in length of the speech signal. The feature vectors are provided to the search engine 612 which compares the feature vectors to the language model. The search engine 612 uses the grammar memory 10 which defines all of the word strings that the originating user might say, the dictionary memory defines how those words might be said and the acoustic memory stores the phonetic segments for the dictionary of words. A best guess is made for the words. This string of words is provided to the natural language interpreter 616 which assigns a meaning to those words.

15 It is possible for the present invention to be implemented and utilized by users that do not have their own browser controller or access an account on a service provider's browser controller. By way of example, consider an airline, car rental agency and hotel chain that agree market cooperatively. A user could call the airline to make travel arrangements to a city. The flight arrangements can be made and tickets can be purchased 20 using an automated system. The automated system can include a browser controller. In such a case, the user could be prompted by appropriate earcons or other audio cues to then reserve a rental automobile with the cooperating car rental agency. The browser controller in the airline's automated system will then automatically connect the user to the car rental agency in just the way described above. Once the automobile is rented, the car rental

agency's browser controller can then connect the user to the hotel chain to reserve a room.

As will be readily understood, the user in this example is daisy-chained to the hotel chain through both the airline browser controller and the car rental agency browser controller. When a user is daisy-chained, each call in the chain remains active and thus 5 billed by the telephone service provider. Thus, it is preferable that the browser controller operate as described above wherein it establishes a new call upon hearing the user repeat an audio link and then disconnects the previous call rather than daisy-chaining the calls through one another.

By way of another example of how the present invention can be utilized by users 10 that do not have their own browser controller 102, consider that the airline described above does not wish to link to the hotel and the rental car voice pages. Even so, it is still to the airline's advantage to use the present invention. The browser controller 102 could read the airline's information as a voice-enabled world wide web page, thereby eliminating the need on the part of the airline for a separate IVR system with separate database integration. If a 15 user has their own browser controller, then the airline does not need to provide telephone access to its world wide web page. However, if the user does not have their own browser controller 102, the airline can provide it for them. The airline could also lease time on a browser controller 102 that exists at an external call center, eliminating the need for the airline to have its own call center for telephone access to its world wide web page. This 20 provides for considerable economies of scale. With intelligent caching of the airline's voice data, prompts and grammars, latency can still be kept to a minimum.

The present invention has been described in terms of specific embodiments incorporating details to facilitate the understanding of the principles of construction and operation of the invention. Such reference herein to specific embodiments and details

thereof is not intended to limit the scope of the claims appended hereto. It will be apparent to those skilled in the art that modifications can be made in the embodiments chosen for illustration without departing from the spirit and scope of the invention. For example, the Browser Controller 102 could be configured to first disconnect a present 5 before establishing a new link.

Specifically, it will be apparent to one of ordinary skill in the art that the device of the present invention could be implemented in several different ways and the apparatus disclosed above is only illustrative of the preferred embodiment of the invention and is in no way a limitation. It will be apparent that the various aspects of the above-described 10 invention can be utilized singly or in combination with one or more of the other aspects of the invention described herein. In addition, the various elements of the present invention could be substituted with other elements.

**C L A I M S**

What is claimed is:

1       1. An apparatus configured to allow a user to interactively browse an audio telephony  
2       network, the apparatus comprising:

3           a. means for coupling an originating user to a first telephony service at a first  
4           telephone number;

5           b. means for providing a first audio indication with a first associated text having  
6           an associated second telephone number, wherein the audio indication is configured  
7           to be sensed by the originating user; and

8           c. means for sensing the originating user repeating the first associated text and in  
9           response thereto coupling the originating user to a second telephony service at the  
10          second telephone number.

1       2. The apparatus according to claim 1 further comprising means for disconnecting the  
2       first telephony service upon sensing a successful coupling to the second telephony service.

1       3. The apparatus according to claim 1 wherein the first telephony service includes a  
2       plurality of audio indications with associated texts, each active upon access to the first  
3       telephony service.

1 4. The apparatus according to claim 1 further comprising:

2 a. means for providing an audio indication with associated text, and having an  
3 associated URL for a voice-enabled world wide web page, wherein the audio  
4 indication is configured to be sensed by the originating user; and  
5 b. means for sensing the originating user repeating the second associated text and  
6 in response thereto coupling the originating user to the voice-enabled world wide  
7 web page at the associated URL.

1 5. The apparatus according to claim 4 further comprising means for disconnecting the  
2 first telephony service upon sensing a successful coupling to the voice-enabled world wide  
3 web page.

1 6. The apparatus according to claim 1 further comprising:

2 a. means for providing an audio indication with associated text, and having an  
3 associated URL for a world wide web page operating in conjunction with a text-to-  
4 speech converter, wherein the audio indication is configured to be sensed by the  
5 originating user; and  
6 b. means for sensing the originating user repeating the second associated text and  
7 in response thereto coupling the originating user to the world wide web page at the  
8 associated URL.

1 7. The apparatus according to claim 6 further comprising means for disconnecting the  
2 first telephony service upon sensing a successful coupling to the world wide web page.

1       8. An apparatus configured to allow a user to interactively browse an audio telephony  
2       network, the apparatus comprising:  
3           a.       means for coupling an originating user to a voice-enabled world wide web page  
4               at a first URL;  
5           b.       means for providing a first audio indication with a first associated text, and  
6               having an associated telephone number, wherein the audio indication is configured  
7               to be sensed by the originating user; and  
8           c.       means for sensing the originating user repeating the first associated text and in  
9               response thereto coupling the originating user to a first telephony service at the first  
10          telephone number.

1       9. The apparatus according to claim 8 wherein the voice-enabled world wide web page  
2       includes a plurality of audio indications, each active upon access to the voice-enabled  
3       world wide web page.

1       10. The apparatus according to claim 8 further comprising:  
2           a.       means for providing an audio indication with a second associated text, and  
3               having an associated URL for a voice-enabled world wide web page, wherein the  
4               audio indication is configured to be sensed by the originating user;  
5           b.       means for sensing the originating user repeating the second associated text and  
6               in response thereto coupling the originating user to the voice-enabled world wide  
7               web page at the associated URL.

1 11. The apparatus according to claim 10 further comprising means for disconnecting the  
2 first telephony service upon sensing a successful coupling to the voice-enabled world wide  
3 web page.

1 12. The apparatus according to claim 8 further comprising:  
2 a. means for providing an audio indication with a second associated text, and  
3 having an associated URL for a world wide web page configured to operate in  
4 conjunction with a text-to-speech converter, wherein the audio indication is  
5 configured to be sensed by the originating user;  
6 b. means for sensing the originating user repeating the second associated text and  
7 in response thereto coupling the originating user to the world wide web page at the  
8 associated URL.

1 13. The apparatus according to claim 12 further comprising means for disconnecting the  
2 first telephony service upon sensing a successful coupling to the world wide web page.

1 14. An apparatus configured to allow a user to interactively browse an audio telephony  
2 network, the apparatus comprising:  
3 a. means for coupling an originating user to a world wide web page at a first  
4 URL configured to operate in conjunction with a text-to-speech converter;  
5 b. means for providing a first audio indication with a first associated text, and  
6 having an associated telephone number, wherein the audio indication is configured  
7 to be sensed by the originating user; and  
8 c. means for sensing the originating user repeating the first associated text and in

9 response thereto coupling the originating user to a first telephony service at the first  
10 telephone number.

15. The apparatus according to claim 14 wherein the world wide web page includes a plurality of audio indications, each active upon access to the world wide web page.

16. The apparatus according to claim 14 further comprising:

- a. means for providing an audio indication with a second associated text, and having an associated URL for a voice-enabled world wide web page, wherein the audio indication is configured to be sensed by the originating user;
- b. means for sensing the originating user repeating the second associated text and in response thereto coupling the originating user to the voice-enabled world wide web page at the associated URL.

17. The apparatus according to claim 16 further comprising means for disconnecting the first telephony service upon sensing a successful coupling to the voice-enabled world wide web page.

18. The apparatus according to claim 14 further comprising:

- a. means for providing an audio indication with a second associated text, and having an associated URL for a world wide web page configured to operate in conjunction with a text-to-speech converter, wherein the audio indication is configured to be sensed by the originating user;
- b. means for sensing the originating user repeating the second associated text and

7           in response thereto coupling the originating user to the world wide web page at the  
8           associated URL.

1       19.    The apparatus according to claim 18 further comprising means for disconnecting the  
2           first telephony service upon sensing a successful coupling to the world wide web page.

1       20.    A method of interactively browsing an audio telephony network comprising the  
2           steps of:

3           a.       coupling an originating user to a first telephony service at a first telephone  
4           number;  
5           b.       providing an audio indication having an associated text, and within the first  
6           telephony service having an associated second telephone number, wherein the audio  
7           indication is configured to be sensed by the originating user; and  
8           c.       sensing the originating user repeating the associated text and in response thereto  
9           disconnecting the first telephony service and for coupling the originating user to a  
10           second telephony service at the second telephone number.

1       21.    The method according to claim 20 wherein the first telephony service includes a  
2           plurality of audio indications with associated texts, each active upon access to the first  
3           telephony service.

1       22.    A system for a user to interactively browse a network of audio information  
2           comprising:

3           a.       a browser controller, comprising:

- (1) means for allowing the user to receive audio information and to transmit verbal instructions; and
- (2) means for linking the user to a first telephone station in response to a voice command; and

b. a voice page, comprising:

- (1) means for playing information, wherein certain information is played with an audio indication of a linking capability for that information;
- (2) means for sensing if the user repeats the information set off by the audio indication; and
- (3) means for transmitting a telephone number associated with the certain information and control signal the browser controller in response to sensing that the user repeated the information set off by the audio indication,

23. The system according to claim 22 wherein the browser controller further comprises a plurality of predetermined voice commands available to the user.

24. The system according to claim 22 wherein the browser controller further comprises a start page of information regarding preferred links desired by the user.

25. The system according to claim 22 wherein the browser controller further comprises means for monitoring a telephone call and for allowing the browser controller to recapture control of the call in the event the user speaks a predetermined control word.

1       26.    The system according to claim 22 wherein the first telephone station comprises a  
2       predetermined information service.

1       27.    The system according to claim 22 wherein the first telephone station comprises an  
2       IVR speech system.

1       28.    The system according to claim 22 wherein the first telephone station comprises an  
2       IVR dtmf system.

1       29.    The system according to claim 22 wherein the first telephone station comprises a  
2       telephony service configured for interacting with the means for linking.

1       30.    The system according to claim 22 wherein the first telephone station comprises a  
2       conventional telephone set.

1       31.    The system according to claim 22 wherein the first telephone station comprises a  
2       voice-enabled world wide web page.

1       32.    The system according to claim 22 wherein the first telephone station comprises a  
2       world wide web page configured to operate in conjunction with a text-to-speech converter.

1       33.    A system for a user to interactively browse a network of audio information  
2       comprising:

3            a.       a browser controller, comprising:

4 (1) means for allowing the user to receive audio information and to transmit verbal  
5 instructions; and  
6 (2) means for linking the user to a voice-enabled world wide web page in response  
7 to a voice command; and  
8 b. a voice page, comprising:  
9 (1) means for playing information, wherein certain information is played with an  
10 audio indication of a linking capability for that information;  
11 (2) means for sensing if the user repeats the information set off by the audio  
12 indication; and  
13 (3) means for transmitting a telephone number associated with the certain  
14 information and control signal the browser controller in response to sensing that  
15 the user repeated the information set off by the audio indication,  
16 such that the browser controller disconnects the user from the first telephone station and  
17 establishes a new link to a second telephone station with the telephone number.

1 34. The system according to claim 33 wherein the browser controller further comprises  
2 a plurality of predetermined voice commands available to the user.

1 35. The system according to claim 33 wherein the browser controller further comprises  
2 a start page of information regarding preferred links desired by the user.

1 36. The system according to claim 33 wherein the browser controller further comprises  
2 means for monitoring a telephone call and for allowing the browser controller to recapture  
3 control of the call in the event the user speaks a predetermined control word.

1       37. The system according to claim 33 wherein the first telephone station comprises a  
2       predetermined information service.

1       38. A method of allowing a user to interactively browse a network of audio information  
2       comprising:

- 3           a.       allowing the user to receive audio information and to transmit verbal  
4           instructions;
- 5           b.       linking the user to a first telephone station in response to a voice command;  
6           and
- 7           c.       at a remote location:
  - 8           (1) playing information, wherein certain information is played with an audio  
9           indication of a linking capability for that information; and
  - 10           (2) sensing if the user repeats the information set off by the audio indication;
  - 11           (3) transmitting a telephone number associated with the certain information and  
12           control signal the browser controller in response to sensing that the user  
13           repeated the information set off by the audio indication,  
14           such that the user is disconnected from the first telephone station and a new link is  
15           established to a second telephone station with the telephone number.

1       39. A method of allowing a user to interactively browse a network of audio information  
2       comprising:

- 3           a.       allowing the user to receive audio information and to transmit verbal  
4           instructions;
- 5           b.       linking the user to a first telephone station in response to a voice command;

6 and

7 c. at a remote location:

- 8 (1) playing information, wherein certain information is played with an audio  
9 indication of a linking capability for that information; and
- 10 (2) sensing if the user repeats the information set off by the audio indication;
- 11 (3) transmitting a URL associated with the certain information and control signal the  
12 browser controller in response to sensing that the user repeated the information  
13 set off by the audio indication,

14 such that the user is disconnected from the first telephone station and a new link is  
15 established to a voice-enabled world wide web page with the URL.

1 40. A method of allowing a user to interactively browse a network of audio information  
2 comprising:

- 3 a. allowing the user to receive audio information and to transmit verbal  
4 instructions;
- 5 b. linking the user to a first telephone station in response to a voice command;
- 6 and
- 7 c. at a remote location:
  - 8 (1) playing information, wherein certain information is played with an audio  
9 indication of a linking capability for that information; and
  - 10 (2) sensing if the user repeats the information set off by the audio indication;
  - 11 (3) transmitting a URL associated with the certain information and control signal  
12 the browser controller in response to sensing that the user repeated the  
13 information set off by the audio indication,

14 such that the user is disconnected from the first telephone station and a new link is  
15 established to a world wide web page configured to operate in conjunction with a text-to-  
16 speech converter with the URL.

1 41. A method of allowing a user to interactively browse a network of audio information  
2 comprising:

3 a. allowing the user to receive audio information and to transmit verbal  
4 instructions;

5 b. linking the user to a voice-enabled world wide web page in response to a voice  
6 command; and

7 c. at a remote location:

8 (1) playing information, wherein certain information is played with an audio  
9 indication of a linking capability for that information; and

10 (2) sensing if the user repeats the information set off by the audio indication;

11 (3) transmitting a telephone number associated with the certain information and  
12 control signal the browser controller in response to sensing that the user  
13 repeated the information set off by the audio indication,

14 such that the user is disconnected from the first telephone station and a new link is  
15 established to a second telephone station with the telephone number.

1 42. A method of allowing a user to interactively browse a network of audio information  
2 comprising:

3 a. allowing the user to receive audio information and to transmit verbal  
4 instructions;

5       b.       linking the user to a voice-enabled world wide web page in response to a voice  
6        command; and  
7       c.       at a remote location:  
8           (1) playing information, wherein certain information is played with an audio  
9            indication of a linking capability for that information; and  
10          (2) sensing if the user repeats the information set off by the audio indication;  
11          (3) transmitting a URL associated with the certain information and control signal the  
12            browser controller in response to sensing that the user repeated the information  
13            set off by the audio indication,  
14        such that the user is disconnected from the first telephone station and a new link is  
15        established to a voice-enabled world wide web page with the URL.

1       43.      A method of allowing a user to interactively browse a network of audio information  
2        comprising:  
3           a.       allowing the user to receive audio information and to transmit verbal  
4            instructions;  
5           b.       linking the user to a voice-enabled world wide web page in response to a voice  
6            command; and  
7           c.       at a remote location:  
8            (1) playing information, wherein certain information is played with an audio  
9            indication of a linking capability for that information; and  
10          (2) sensing if the user repeats the information set off by the audio indication;  
11          (3) transmitting a URL associated with the certain information and control signal  
12            the browser controller in response to sensing that the user repeated the

13                   information set off by the audio indication,  
14                   such that the user is disconnected from the first telephone station and a new link is  
15                   established to a world wide web page configured to operate in conjunction with a text-to-  
16                   speech converter with the URL.

1       44.    A method of allowing a user to interactively browse a network of audio information  
2                   comprising:

- 3               a.    allowing the user to receive audio information and to transmit verbal  
4                   instructions;
- 5               b.    linking the user to a world wide web page configured to operate in conjunction  
6                   with a text-to-speech converter in response to a voice command; and
- 7               c.    at a remote location:

- 8               (1) playing information, wherein certain information is played with an audio  
9                   indication of a linking capability for that information; and
- 10              (2) sensing if the user repeats the information set off by the audio indication;
- 11              (3) transmitting a telephone number associated with the certain information and  
12                   control signal the browser controller in response to sensing that the user  
13                   repeated the information set off by the audio indication,

14                   such that the user is disconnected from the first telephone station and a new link is  
15                   established to a second telephone station with the telephone number.

1 45. A method of allowing a user to interactively browse a network of audio information  
2 comprising:

3 a. allowing the user to receive audio information and to transmit verbal  
4 instructions;

5 b. linking the user to a world wide web page configured to operate in conjunction  
6 with a text-to-speech converter in response to a voice command; and

7 c. at a remote location:  
8 (1) playing information, wherein certain information is played with an audio  
9 indication of a linking capability for that information; and  
10 (2) sensing if the user repeats the information set off by the audio indication;  
11 (3) transmitting a URL associated with the certain information and control signal the  
12 browser controller in response to sensing that the user repeated the information  
13 set off by the audio indication,  
14 such that the user is disconnected from the first telephone station and a new link is  
15 established to a voice-enabled world wide web page with the URL.

1 46. A method of allowing a user to interactively browse a network of audio information  
2 comprising:

3 a. allowing the user to receive audio information and to transmit verbal  
4 instructions;

5 b. linking the user to a world wide web page configured to operate in conjunction  
6 with a text-to-speech converter in response to a voice command; and

7 c. at a remote location:  
8 (1) playing information, wherein certain information is played with an audio

9 indication of a linking capability for that information; and

10 (2) sensing if the user repeats the information set off by the audio indication;

11 (3) transmitting a URL associated with the certain information and control signal

12 the browser controller in response to sensing that the user repeated the

13 information set off by the audio indication,

14 such that the user is disconnected from the first telephone station and a new link is

15 established to a world wide web page configured to operate in conjunction with a text-to-

16 speech converter with the URL.

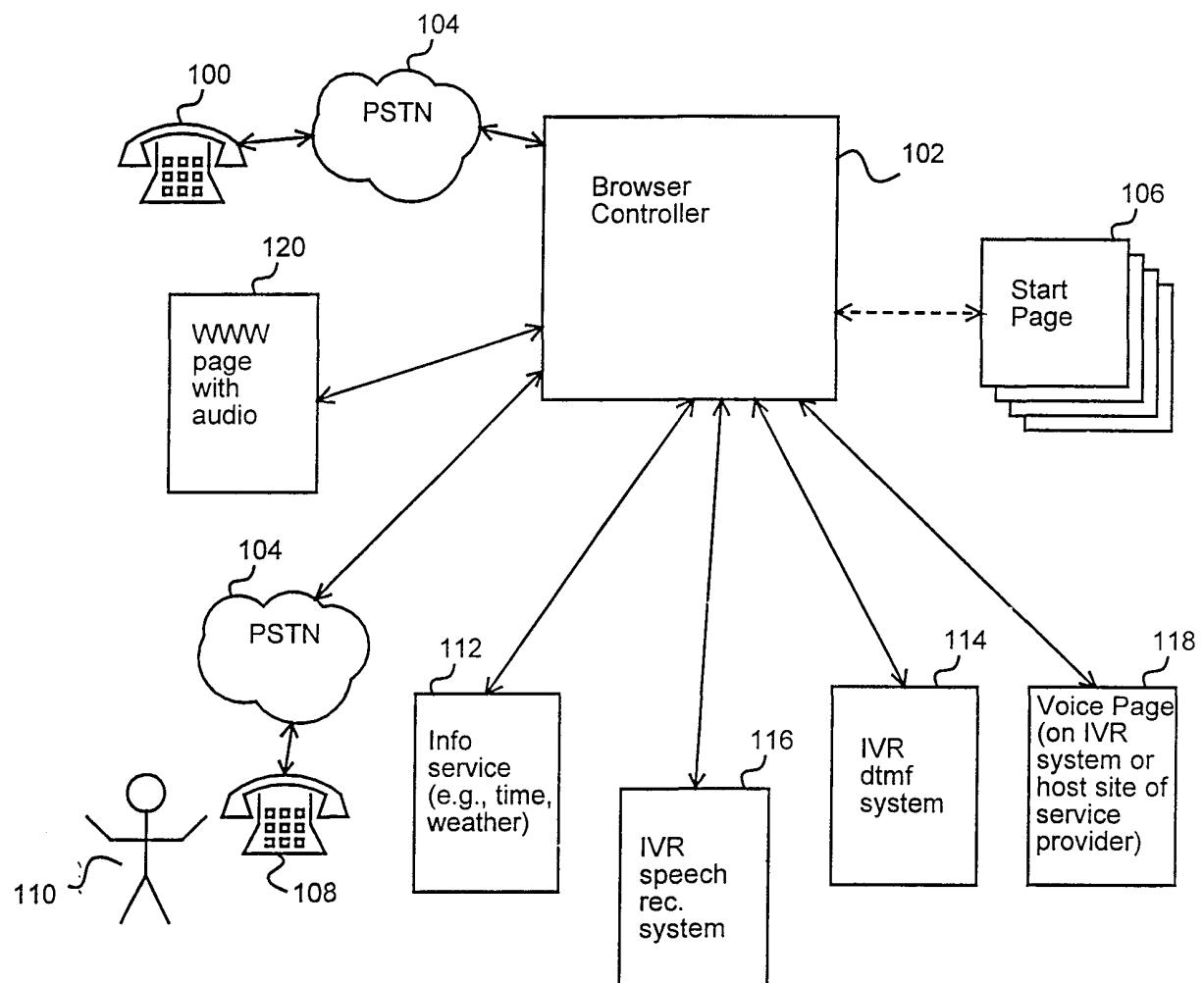


Fig. 1

# BROWSER CONTROLLER

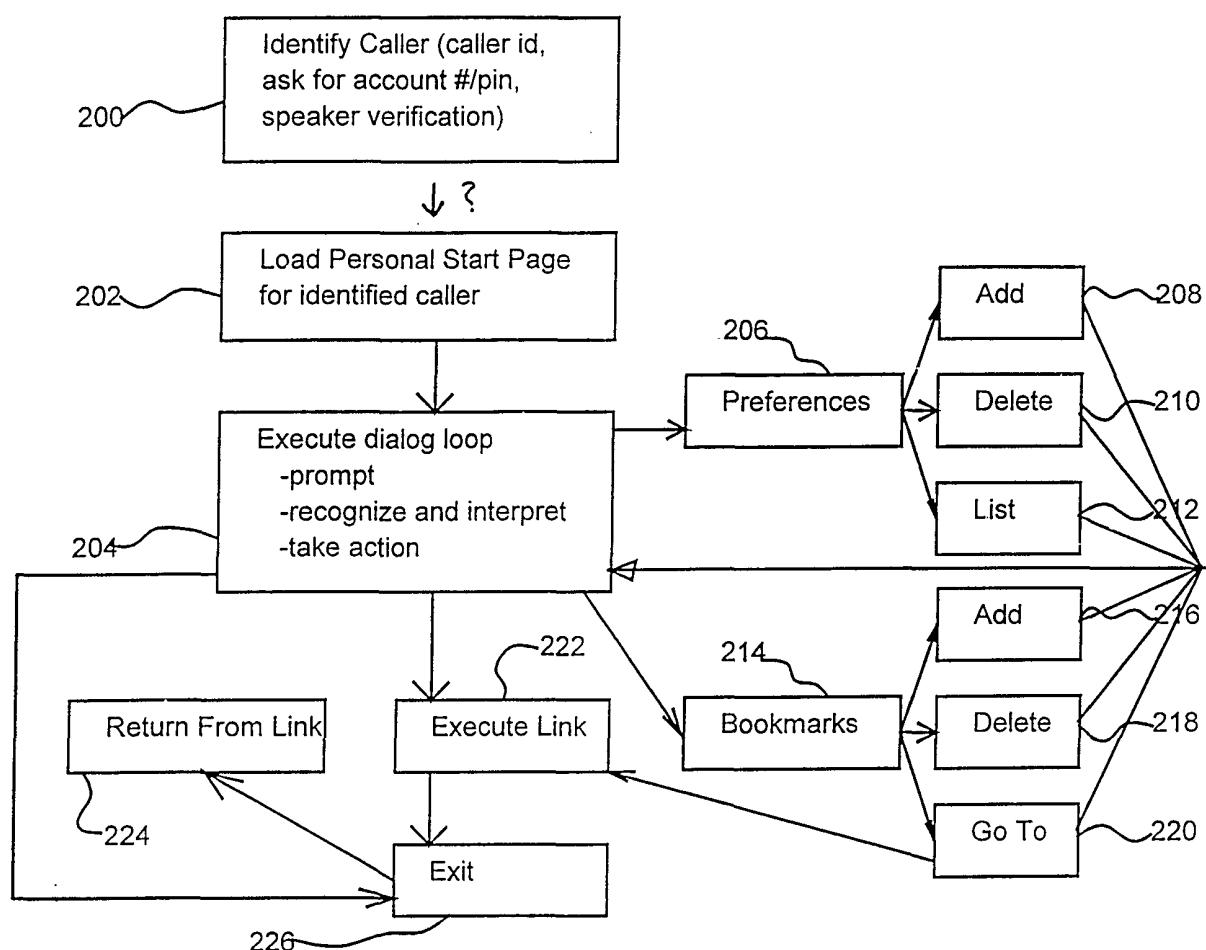
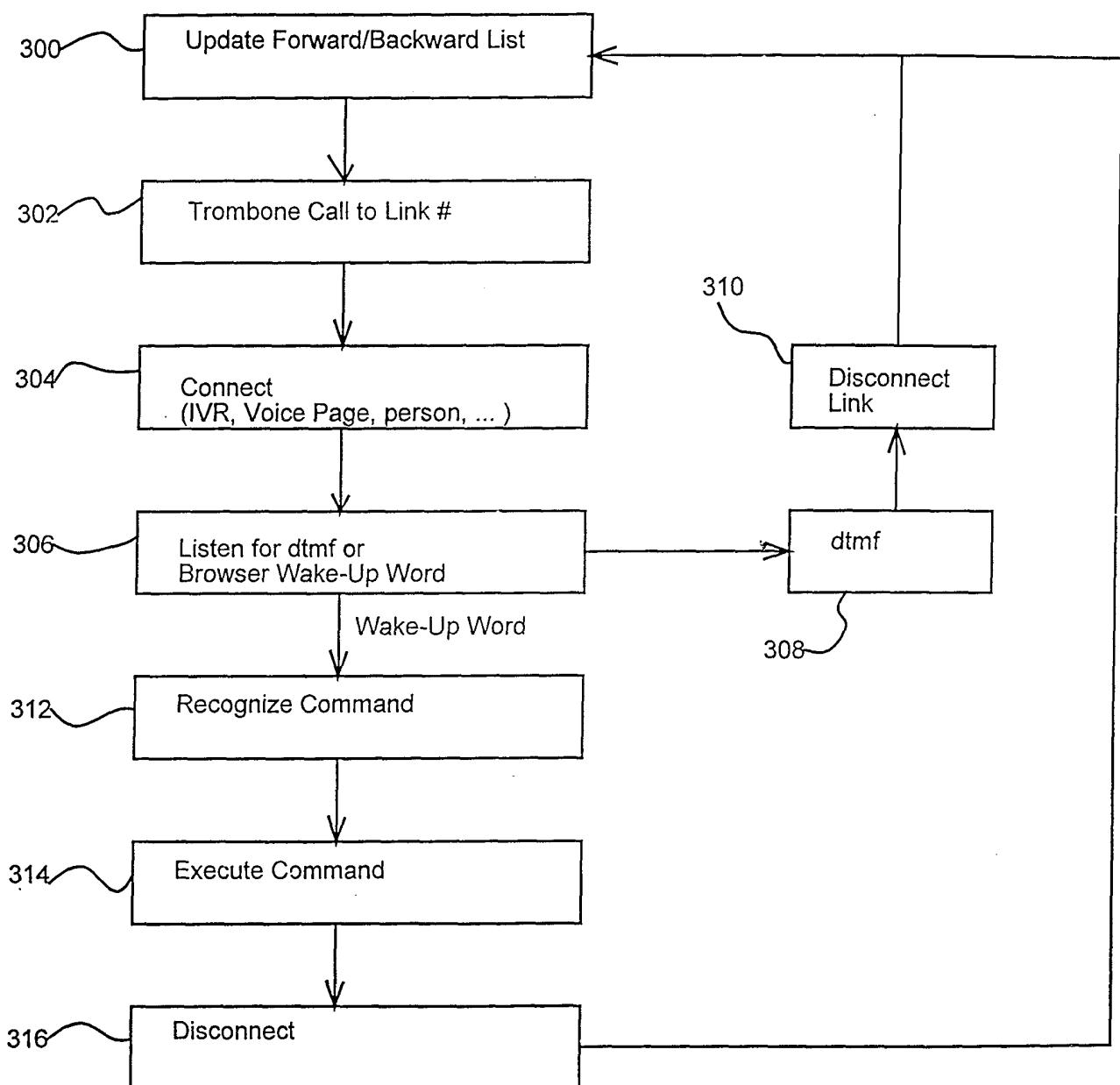


FIGURE 2

# EXECUTE LINK TO A TELEPHONE STATION



## FIGURE 3

## RETURN FROM LINK

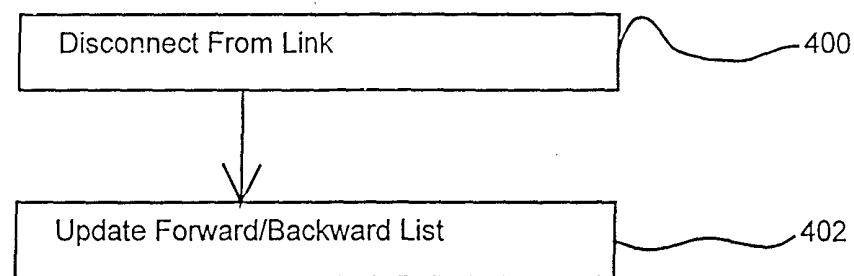


FIGURE 4

## VOICE PAGE

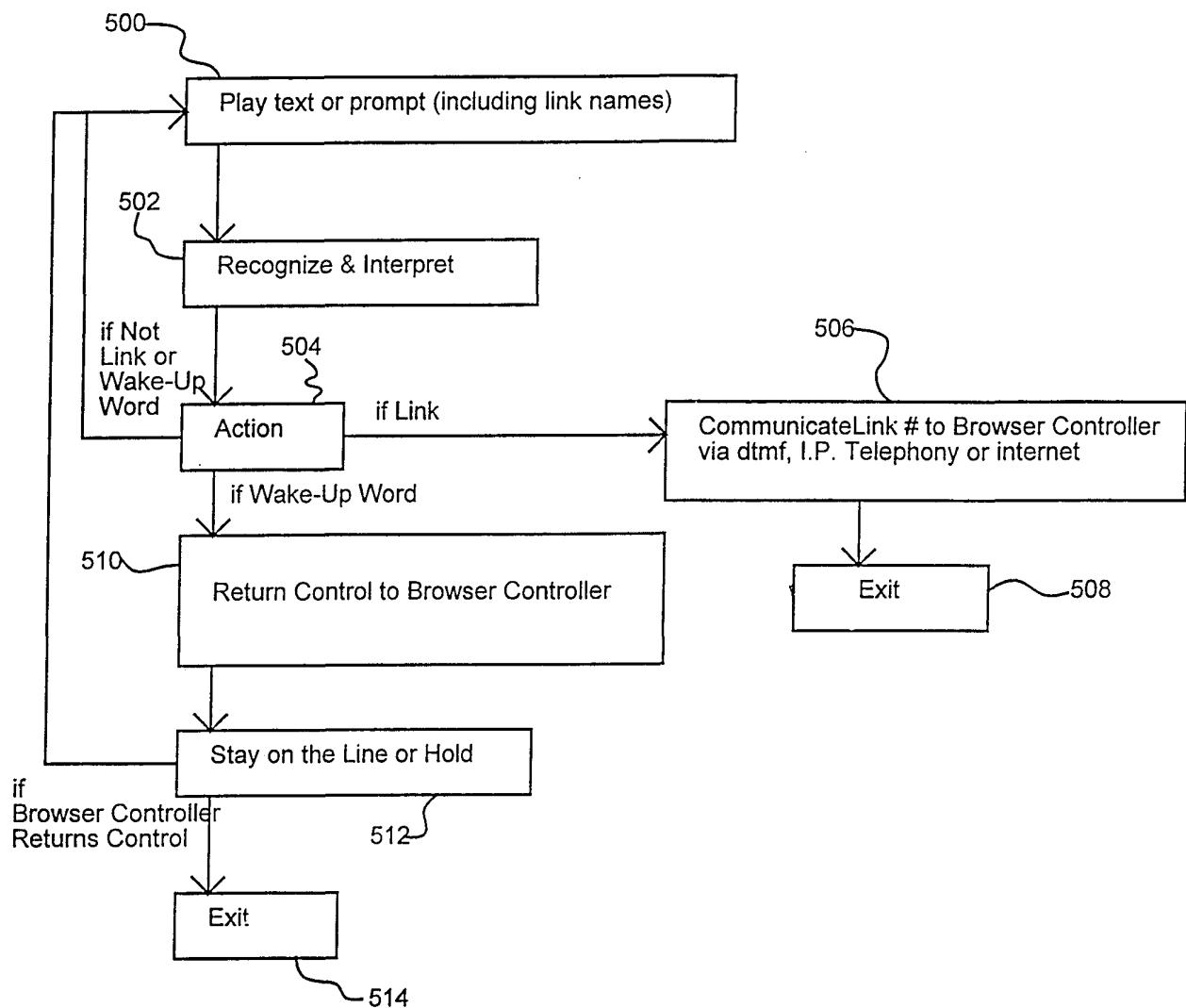


FIGURE 5

## RECOGNIZE & INTERPRET

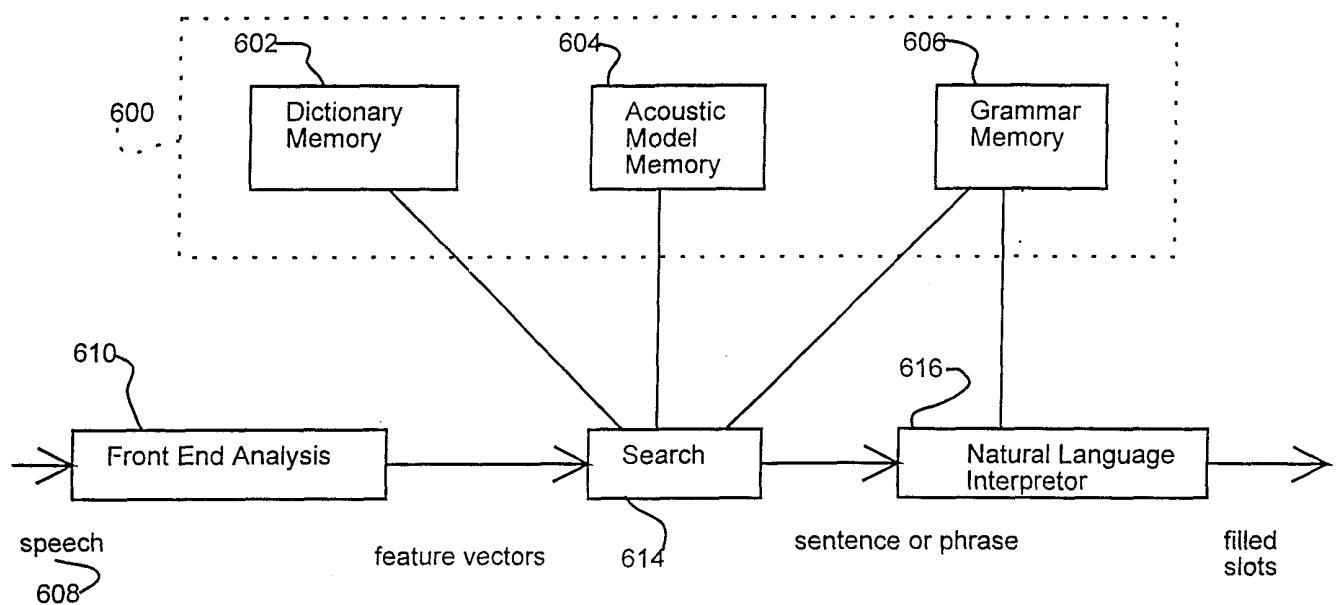


FIGURE 6

## INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 99/28480

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 H04M3/493

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 H04M

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)

C. DOCUMENTS CONSIDERED TO BE RELEVANT
--

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 0 847 179 A (AT & T CORP) 10 June 1998 (1998-06-10) column 1, line 1 -column 2, line 44 column 4, line 15 -column 6, line 25 ---	42,43,46
X	EP 0 848 373 A (SIEMENS CORP RES INC) 17 June 1998 (1998-06-17) column 1, line 21-45 column 5, line 40 -column 6, line 4 column 10, line 18-27 ---	42,43, 45,46
X	GB 2 317 070 A (IBM) 11 March 1998 (1998-03-11) page 3, line 13 -page 6, line 20 ---	42,43, 45,46 -/-

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	Date of mailing of the international search report
---	--

16 March 2000

05/04/2000

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

Cremer, J

## INTERNATIONAL SEARCH REPORT

International Application No  
PCT/US 99/28480

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5 479 491 A (HERRERO GARCIA JOSE E ET AL) 26 December 1995 (1995-12-26) column 6, line 24 -column 7, line 14 column 18, line 30 -column 19, line 40 -----	1,20,22, 38
X	EP 0 822 698 A (ROLM SYSTEMS) 4 February 1998 (1998-02-04) column 3, line 1 -column 5, line 22 -----	1,20,38

## INTERNATIONAL SEARCH REPORT

Information on patent family members

Int'l Application No

PCT/US 99/28480

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
EP 0847179	A 10-06-1998	CA 2213591 A	04-06-1998	JP 10207685 A 07-08-1998
EP 0848373	A 17-06-1998	US 6018710 A	25-01-2000	
GB 2317070	A 11-03-1998	NONE		
US 5479491	A 26-12-1995	US 5187735 A	16-02-1993	
EP 0822698	A 04-02-1998	US 5943410 A	24-08-1999	US 5950165 A 07-09-1999