PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : G01L 9/00	A2	(11) International Publication Number: WO 99/08084
		(43) International Publication Date: 18 February 1999 (18.02.99)

(21) International Application Number: PCT/US98/15528

(22) International Filing Date: 22 July 1998 (22.07.98)

(30) Priority Data:
08/907,628 8 August 1997 (08.08.97) US

(71) Applicant: BOARD OF TRUSTEES, LELAND STANFORD, JR., UNIVERSITY [US/US]; Suite 350, 900 Welch Road, Palo Alto, CA 94304–1850 (US).

(72) Inventors: SCOTT, Brian, L.; 3401 E. University, #104, Denton, TX 76208 (US). MILLER, Clint, L.; 1914 W. Oak Street, Denton, TX 76201 (US).

(74) Agents: MINSK, Alan, D. et al.; Four Embarcadero Center, Suite 3300, San Francisco, CA 94111 (US).

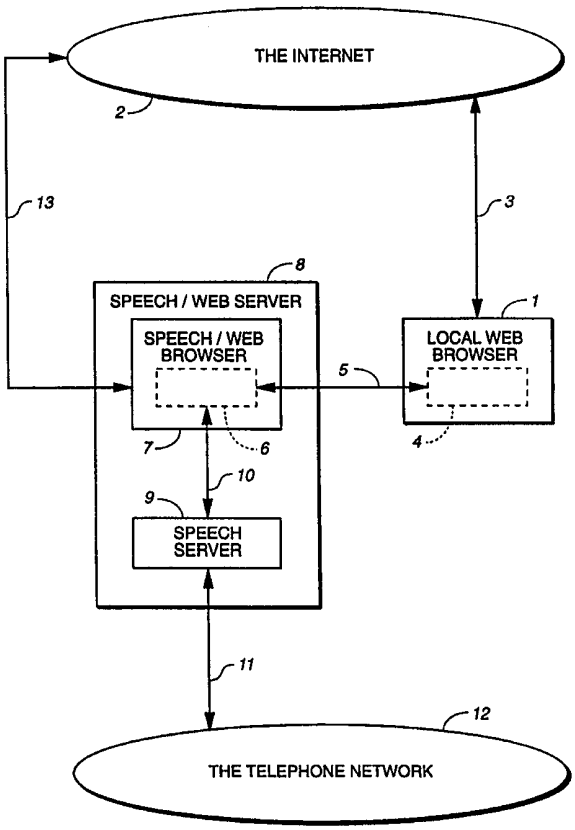
(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: METHOD AND SYSTEM FOR USING SPEECH RECOGNITION TO ACCESS THE INTERNET, INCLUDING ACCESS VIA A TELEPHONE

(57) Abstract

Voice activation of functions on a network such as the Internet are accomplished using a speech recognition system running synchronously with standard desktop–based Internet functions. This synchronous operation allows voice–based control to be exercised for all operations on the Internet. System functions are based on a unique combination of a local web browser, a remotely–located speech/web server, and control links between a web browser and a speech/web server. The control links provide a mechanism for controlling a speech server from a web page and a mechanism for driving both the local, as well as a remote, web browser.



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			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

METHOD AND SYSTEM FOR USING SPEECH RECOGNITION TO ACCESS THE INTERNET, INCLUDING ACCESS VIA A TELEPHONE

FIELD OF THE INVENTION

The present invention relates to the field of computerized communication on the Internet. The general purpose of this invention is to enable speech access to the Internet over standard telephone lines and Internet control of telephony functions through standard web pages. This is accomplished through a unique combination of speech server, web browser, and control links. The control links provide a mechanism for controlling the speech server from a web page and a mechanism for driving both the local, as well as a remote, web browser.

BACKGROUND OF THE INVENTION

The Internet is essentially a network of servers containing information that users can obtain using personal computers. Users generally connect to a server, a computer equipped with information and capabilities that assist the user with contacting other servers and obtaining additional information. Users typically execute these functions, also referred to as "navigating" on the Internet, using a mouse and Windows-based software. The user's navigation of the Internet is thus essentially graphically-based (looking at a screen) with functions activated using a mouse.

Speech recognition software and hardware for use in conjunction with personal computers and other environments, like the Internet, is a rapidly developing technology. With speech recognition, a user's voice commands are recognized by a computer and then converted, based on the speech pattern, into an electronic signal. For example, speech recognition has been highly successful in the field of long-distance telephone calling for the purpose of allowing collect calls. Typically, with this application, a caller will provide a name and a phone number to a computer when making a collect call. The computer will then place the caller on hold and call the number to be reached. The person receiving the collect call will answer "yes" or "no"

in response to the computer message and the collect caller's name. The voice recognition hardware and software, which is also known as a speech recognition engine, either signals a switch to complete the call upon recognizing the "yes" response, or to disconnect upon recognizing the "no" response.

One issue with using speech recognition is selecting the appropriate speech recognition engine to use for a particular application. These speech recognition engines include speaker dependent and independent dictation machines, continuous speech systems, large vocabulary systems, and small vocabulary systems. Further, these systems can be Windows based, Macintosh based, UNIX based, Windows NT based, or based on another platform, depending on the preferred operating system.

Speech recognition operating in conjunction with computer connection with the Internet, also known as speech enabling of the Internet, appears to have promising application possibilities. One possible application of this technology is for navigational purposes on the Internet. For example, speech recognition has been successfully utilized at the desktop level generally. Voice macros have been created for a number of Windows functions for use on the Internet. A macro is a series of functions on the computer activated by a single command. For a voice macro, the speech server's recognition of an inputted voice command activates a series of commands.

Two prior art methods for speech-enabling the Internet have been explored by various companies and research entities. In general terms, researchers have approached the problem from either the perspective of speech-enabling the Internet, or from the perspective of Internet-enabling the telephone system.

The first method is the most common approach and the one being pursued by Texas Instruments, Apple Computer, and Microsoft. In this approach, the speech recognition engine is located on the local host, along with the web browser. This approach allows such activities as those described above -- voice macros for Windows functions that can be used when browsing the Internet.

Texas Instruments further refined this approach by using the text associated with hotlinks to supply the vocabularies for the recognizer. Apple has taken the approach of making both the web browser and the speech recognition engine scriptable

(controllable with the AppleScript language). Microsoft has taken the approach of providing tools for web page developers to allow them to speech-enable their web pages. These tools provide a mechanism for supplying the recognizer with grammars and their speech synthesizers with spoken prompts.

The advantages of the present invention over this method include: (1) telephone access serves a far greater potential audience than speech access limited to desktop operations; (2) no additional requirements of the user's computer, such as a speech recognition engine, are required; (3) the system uses a migration path starting with an immediate utility with no long-term limitations; and (4) direct benefits are available from telephony integration.

Internet-enabling the telephone system is primarily being investigated as a research effort. Demonstrations from MIT and the Sun SpeechActs group have shown potential for using a speech-only interface for retrieving personal information (voice e-mail) over the phone and for using the Internet as an up-to-date repository of information available over the phone. For example, ALTech, a commercial spin-off of MIT, has demonstrated the use of a speech server for obtaining information about local movies.

Advantages of the present invention over this method include: (1) an optional Graphical User Interface (GUI) makes using the system with today's World Wide Web much more practical and simple than attempting to do it with speech alone; (2) the potential user base is just as large over the long term; and (3) providing tools to other developers is expected to lead to much more rapid progress than attempting to build speech-only interfaces from the ground up.

SUMMARY OF THE INVENTION

This invention links networks such as the Internet and the World Wide Web to a speech recognition server, which resides on the telephone system, to provide for speech access to these networks over standard telephone lines and control of telephony functions through standard web pages. These capabilities are accomplished through a combination of speech server (typical of those found in Interactive Voice Response (IVR) applications), web browser, and control links. The control links consist of

software that provides a mechanism for controlling the speech server from a web page, and a mechanism for driving both the local, as well as a remote, web browser.

An example of the capabilities of the system is as follows. A user seeking a service to provide stock quotes can access these quotes by graphically browsing the Internet to a web page that continually carries the quotes. Once at the web page, the user can activate the present invention, telling the speech server to, for example, "mark this" or "show me the stock quote." The server can then be set to either tell the user the stock price or go to that web page upon recognizing of the selected speech pattern.

The general purpose of this invention is thus to provide a method for linking a remote speech recognition device operating over the telephone network to any web browser operating over the Internet. This link enables the user's web browser to be controlled by the remote speech recognition device, and, in turn, enables telephony functions to be controlled by any web browser. In addition to providing an immediate solution to accessing the web by voice, the invention provides tools and motivation for web page authors to generate web pages that are tailored to speech-only interfaces. This is expected to transform the nature of the web, and, over time, to support a truly multi-modal interface with the Internet.

The significance of the invention is that it provides both a means for immediately speech-enabling the Internet and a means for gradually Internet-enabling the telephone system. Other systems have approached the problem of linking speech technology and the Internet from either one perspective or the other (that is, speech-enabling the net or net-enabling the telephone). The approach of the present invention, however, can be viewed from either perspective, and, in so doing, leads to an immediate speech-enabling of the Internet, and to a process of Internet-enabling the telephone. In addition, the present invention leads to functionality completely unobtainable from either of the other approaches taken alone.

The control of both the server's web browser and the user's remote web browser also enables an optional GUI for the user of the Speech/web server. The GUI link is not required for the system to operate; however, because the web is currently graphically-oriented, the ability to use the local web browser as a GUI for the speech-

driven browser is expected to be beneficial when surfing the web by voice. The concept of a telephony-based web browser with an optional GUI constitutes a significant attribute of the system because it provides a common platform that can be used for simple applications by anyone with a telephone. In addition, it can be used for more difficult tasks when a PC or workstation is available to the user.

Another example of the use of the present invention pertains to speech input and output over telephone lines as the additional modality that can be linked to the conventional web browser interface. Thus, rather than placing a call, hanging up, and placing another call, a user will be able to browse using the telephone. This browsing includes such activities as seamlessly speaking to one person, and then connecting to another, and then checking messages and ordering a pizza, all without hanging up and without ever dialing a number. The same method links any alternative user interface to the user's standard web browser. This pertains to browsers with teletypewriter (TTY) interfaces, browsers that understand and speak other languages, or even browsers capable of providing a sense of smell, sight, taste, and touch.

Additional objects, advantages and novel features of the invention will be set forth in part in the description which follows, and in part will become more apparent to those skilled in the art upon examination of the following or may be learned by practice of the invention. The objects and advantages of the invention may be realized and attained by means of the instrumentalities and combinations particularly pointed out in the appended claims.

To achieve the stated and other objects of the present invention, as embodied and described below, the invention may comprise the steps of:

accessing a voice recognition server through a voice transmission device;

such device translating voice transmissions into electronic signals; and

using said translated voice transmissions to perform functions on the Internet via voice translation being performed by said server.

BRIEF DESCRIPTION OF THE DRAWINGS

A block diagram of the invention is shown in Figure 1.

Figure 2 shows how a user that happens across the web page containing connection information on the present invention initiates the process of speech enabling his or her web browser using the preferred embodiment.

Figure 3 illustrates the exchange of information necessary to speech enable a web browser.

Figure 4 shows the connections in place for operation of the preferred embodiment.

Figure 5 illustrates all of the components of the system in operation.

Figure 6 contains an alternative embodiment, in which the local web browser is a slave to the speech/web server.

Figure 7 contains a second alternative embodiment, in which the speech/web server is a slave to the local web browser.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

Using the drawings, the preferred embodiment of the present invention will now be explained.

A block diagram of the invention is shown in Figure 1. A local web browser 1, such as Netscape on a PC, is used to browse the Internet 2 using a conventional Transport Control Protocol (TCP) link 3. The local web browser 1 contains an Applied Speech Technology Protocol (ASTP) plug-in 4, which communicates by ASTP link 5 with an ASTP controller 6 located within a speech/web browser 7 of a speech/web server 8, such as a Pentium processor-based PC running Windows NT. This PC also hosts, or a separate PC coupled to the speech/web server 8 hosts, the speech server 9, which is coupled 10 to the ASTP controller 6. These couples can consist of such connections as an electronic circuit, a fiber optic line, an electromagnetic signal, or any other means of coupling known in the art. A Dialogic line card located in the backplane of the speech server 9 PC couples 11 the speech server 4 to a telephone network 12. The speech/web browser 7 is also TCP linked 13 to the Internet 2.

The three major components of the speech/web server thus are the speech/web browser 7, the speech server 9 with telephony functions, and the ASTP controller software 4 and 6.

The speech/web browser 7 is a standard, off-the-shelf web browser with an ASTP plug-in 6. The ASTP plug-in 6, as described below, is a software program written in a language, such as JAVA, that allows the program to run within a web browser, such as Netscape. However, since the speech/web browser 7 is driven by speech-only, it is always run in text-only mode. This gives it a considerable response time advantage over a browser that must download and display graphics. The time normally devoted to graphics can thus be used by the recognizer (speech server 4) to compile the grammar for the new web page.

The speech server 9 is typical of those used for IVR and operator assist applications. These systems vary considerably in the number of simultaneous channels of speech recognition they can support, but are most often built from off-the-shelf components that plug into a PC (AT bus). A typical configuration for a speech server would be a Pentium class PC running UNIX or Windows NT, loaded with a speech recognizer such as ALTech, PureSpeech, or Nuance, with a Dialogic line card capable of handling multiple simultaneous telephone lines, and two speech recognition boards, each with four channels of recognition. Speech output is either from pre-recorded prompts or a speech synthesizer. The telephone line card enables the system to dial out, receive calls, and to conference calls.

The ASTP software 4 and 6 is the heart of the system. As noted, this software is written and distributed as a plug-in module to Netscape or other browsers and is written in a typical software that can operate in Netscape, such as JAVA. The protocol is a superset of the Common Client Interface (CCI), which provides the mechanism for establishing a persistent link between the speech/web browser 7 and the user's browser (local web browser 1). The persistent link enables the speech/web browser 7 to remotely control the user's web browser 1, the user's web browser 1 to control the speech/web browser 7, and also allows the two browsers 1 and 7 to traverse the web in tandem.

In addition to the CCI-like capability, the ASTP protocols provide the interface to the speech server 9, telling the recognizer what grammar to compile for the next web page. This function is typically fulfilled by simply stripping the text associated with each hotlink and sending it to the recognizer's grammar compiler. Alternatively, versions of the protocol support calls to high-level routines, called "speech behaviors," that handle all of the dialog between the user and the machine. These high-level routines allow users to supply, by voice, specific kinds of information when using the Internet, such as credit card numbers, addresses, and telephone numbers. By providing web page authors with access to well-designed dialog modules that can be easily deployed through simple-to-use web authoring tools, such as the ASTP protocols, the predominately graphical nature of the web changes to accommodate a speech-only, telephone-based interface.

Finally, the ASTP link 5 is what provides the conduit between the web page and the telephone. This allows web authors to include telephone numbers associated with hotlinks that can be dialed by the speech/web server 8. This capability may change how switching is currently done in the telephone network 12.

Figure 2 shows how a user that happens across the web page containing connection information on the present invention initiates the process of speech enabling his or her web browser using the preferred embodiment. A user 15 using a local web browser 16 initiates a TCP connection 17 with the speech/web site, which is served by the speech/web server 18, by selecting a hotlink such as "surf the web by voice" at the web site.

In Figure 3, user 15 of a local web browser 16 and local telephone 19 uploads 20 to the speech/web server 18 from the local web browser 16 the local telephone number and downloads 21 the ASTP plug-in from the speech/web server 18. In Figure 4, the user 15 of the local web browser 16 and local telephone 19 simultaneously connects by ASTP connection 17 and by telephone connection 22 with the speech/web server 18.

The setup of the preferred embodiment is now completed, as shown in Figure 5. The user 15 of the local web browser 16 and local telephone 19 simultaneously

communicates with the speech/web server 18 via ASTP connection 17 and telephone connection 22. The user 15 is also connected by a TCP link 25 to other web servers 24 simultaneously 26 with the speech/web server 18 connection by a TCP link 23 with those other web servers 24.

As a result of these simultaneous links 26, the user can browse the Internet using voice while looking at the screen of the local web browser 16 and speaking over the phone 19. Typically these links allow a user to speak into the phone using words within the system's capability. These words are recognized and interpreted by the speech/web browser located at the speech/web server 18 and translated into a TCP link 23 command for the speech/web browser at the speech/web server 18. At the same time, the ASTP supplies the same TCP link command 17 on the local web browser 16. Thus, the user 15 speaks to control browsing of the Internet.

A significant advantage of the preferred embodiment is responsiveness. The dual link approach allows time for the speech/web server to generate grammars while the user's browser is busy displaying graphics. A secondary advantage is that neither of the web browsers need to be modified for the system to work.

Variation and Modifications

Two variations on the invention are illustrated in Figures 6 and 7. These approaches differ from the one described in Figure 1 in that they require only a single link into the Internet, rather than the two links described previously.

In the method shown in Figure 6, the local web browser 1 with ASTP plug-in 4 is linked 5 to an ASTP controller 6 located within a speech/web browser 7 housed within a Pentium processor PC-based speech/web server 8. This PC is typically running Windows. This PC also hosts, or a separate PC coupled to the speech/web server 8 hosts, the speech server 9, which is coupled 10 to the ASTP controller 6. The speech server 9 is linked 11 to a telephone network 12. The speech/web browser 7 is also TCP linked 13 to the Internet 2.

The primary difference between this alternative and the earlier embodiment (Figure 1) is that a direct link 13 does not exist between the speech/web browser 6 and

the Internet 2 simultaneous with a link between the local web browser 1 and the Internet 2 (link 3 of Figure 1).

In the method shown in Figure 7, the local web browser 1 with ASTP plug-in 4 is linked 5 to an ASTP controller 6 located within a speech/web browser 7 housed within a Pentium processor-based PC speech/web server 8. This PC also hosts, or a separate PC coupled to the speech/web server 8 hosts, the speech server 9, which is coupled 10 to the ASTP controller 6. The speech server 9 is linked 11 to a telephone network 12. The local web browser 1 is also TCP linked 3 to the Internet 2.

The primary difference between this alternative and the earlier embodiment (Figure 1) is that a direct link does not exist between the speech/web browser 6 and the Internet 2 (link 13 of Figure 1) simultaneous with a link 3 between the local web browser 1 and the Internet 2.

WHAT IS CLAIMED IS:

1. A method for utilizing speech recognition of voice commands by a user to connect with and perform functions on a network via a web site that is served by a voice recognition server, comprising the steps of: accessing said voice recognition server through said web site; accessing said voice recognition server through a voice transmission device; translating voice transmissions into electronic signals; and using said translated voice transmissions to perform functions on the network via voice translation being performed by said server.

2. The method of claim 1 whereby said access to the Internet is through a local web browser.

3. The method of claim 2 whereby said local web browser is a PC.

4. The method of claim 1 further including the step of said user providing a voice connection identification to said voice recognition server.

5. The method of claim 4 wherein said voice connection identification is a phone number.

6. The method of claim 1 whereby said voice transmission device is a telephone.

7. The method of claim 1 whereby said voice transmission device is a microphone coupled to a computer.

8. The method of claim 1 whereby said access of said voice transmission device is through a telephone network.

9. The method of claim 1 whereby said access is through a telephone line.

10. The method of claim 1 further including the step of said voice recognition server instructing said local browser to perform said function.

11. The method of claim 10 further including the step of communicating said instruction through software loaded on said local browser.

12. The method of claim 11 whereby said software is downloaded from said voice recognition server.

13. The method of claim 10 whereby said instruction is through software loaded on said voice recognition server.

14. A system for utilizing speech recognition of voice commands by a user to connect with and perform functions on a network, comprising:a local web browser coupled to a speech/web server; voice information transmitted from said user to said server via a telecommunication connection; a software program for communicating speech recognition information and executable functions between said local web browser and said speech/web server; a speech server for translating said voice information into electronic signals for use by said speech/web browser; a speech/web browser residing on said speech/web server for executing functions; a coupling between said local web browser and said speech/web server such that said speech/web server executes functions on said local web browser in response to voice commands; and a coupling between said speech/web server and said network.

15. The system of claim 14 in which said speech/web browser includes software for navigating the Internet.

16. The system of claim 14 in which said speech/web server consists of a personal computer.

17. The system of claim 15 in which said speech/web browser runs within said speech/web server.

18. The system of claim 14 in which said software program for communicating said speech recognition information and said executable functions operate within said speech/web browser.

19. The system of claim 14 in which said local web browser includes software for navigating the Internet.

20. The system of claim 19 in which said local web browser runs on a personal computer.

21. The system of claim 14 in which said software program for communicating said speech recognition information and said executable functions operate within said local web browser.

22. The system of claim 14 in which said telecommunication connection includes a telephone network.

23. The system of claim 14 in which said telecommunication connection includes a telephone line.

24. The system of claim 14 in which said speech server consists of a speech recognition engine and hardware for connection to a telephone network.

25. The system of claim 14 in which said coupling is an electronic circuit.

26. The system of claim 14 in which said coupling is a telephone network.

27. The system of claim 14 in which said coupling is a telephone line.

28. The system of claim 14 in which said coupling is a fiber optic line.

29. The system of claim 14 in which said coupling is an electromagnetic wave signal.

30. The system of claim 14 in which said network is the Internet.

31. The system of claim 14 in which said speech/web browser is coupled to said network such that said speech/web browser executes said functions on said network.

32. The system of claim 31 in which said coupling is an electronic circuit.

33. The system of claim 31 in which said coupling is a telephone network.

34. The system of claim 31 in which said coupling is a telephone line.

35. The system of claim 31 in which said coupling is a fiber optic line.

36. The system of claim 31 in which said coupling is an electromagnetic wave signal.

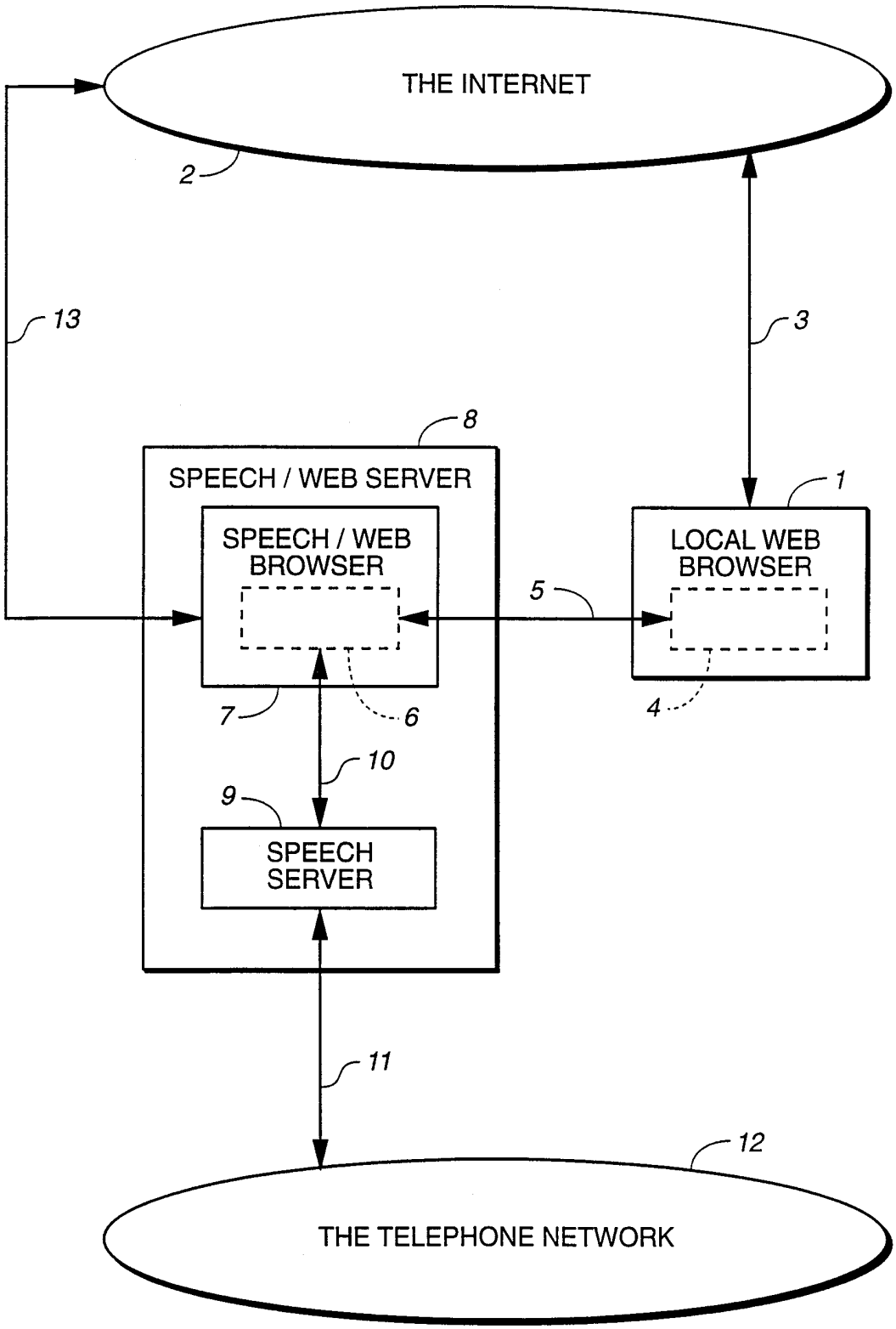

THE INTERNET
2
13
3
8
SPEECH / WEB SERVER
SPEECH / WEB BROWSER
1
LOCAL WEB BROWSER
5
7
6
4
10
9
SPEECH SERVER
11
12
THE TELEPHONE NETWORK

FIG._1

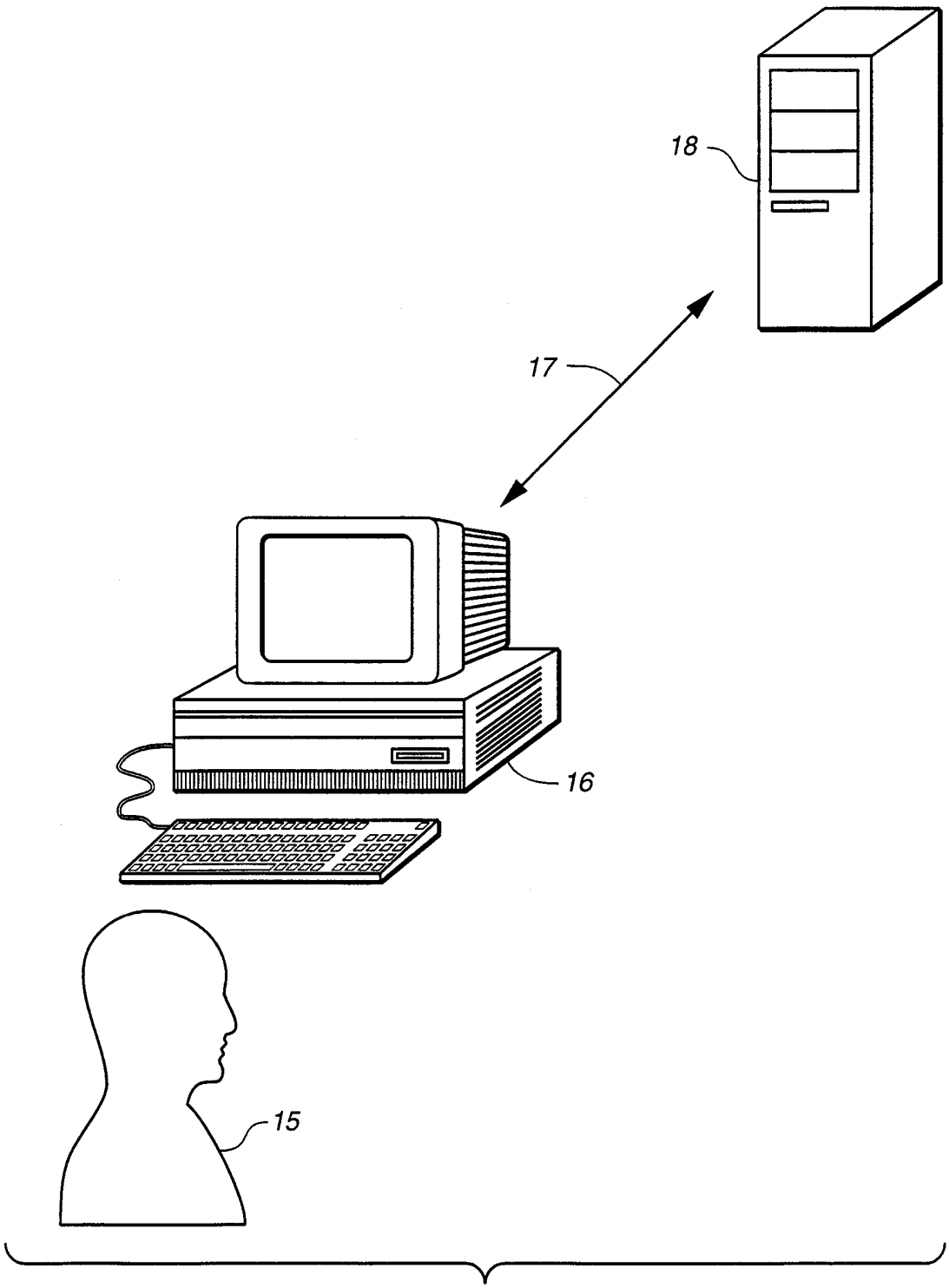


FIG._2

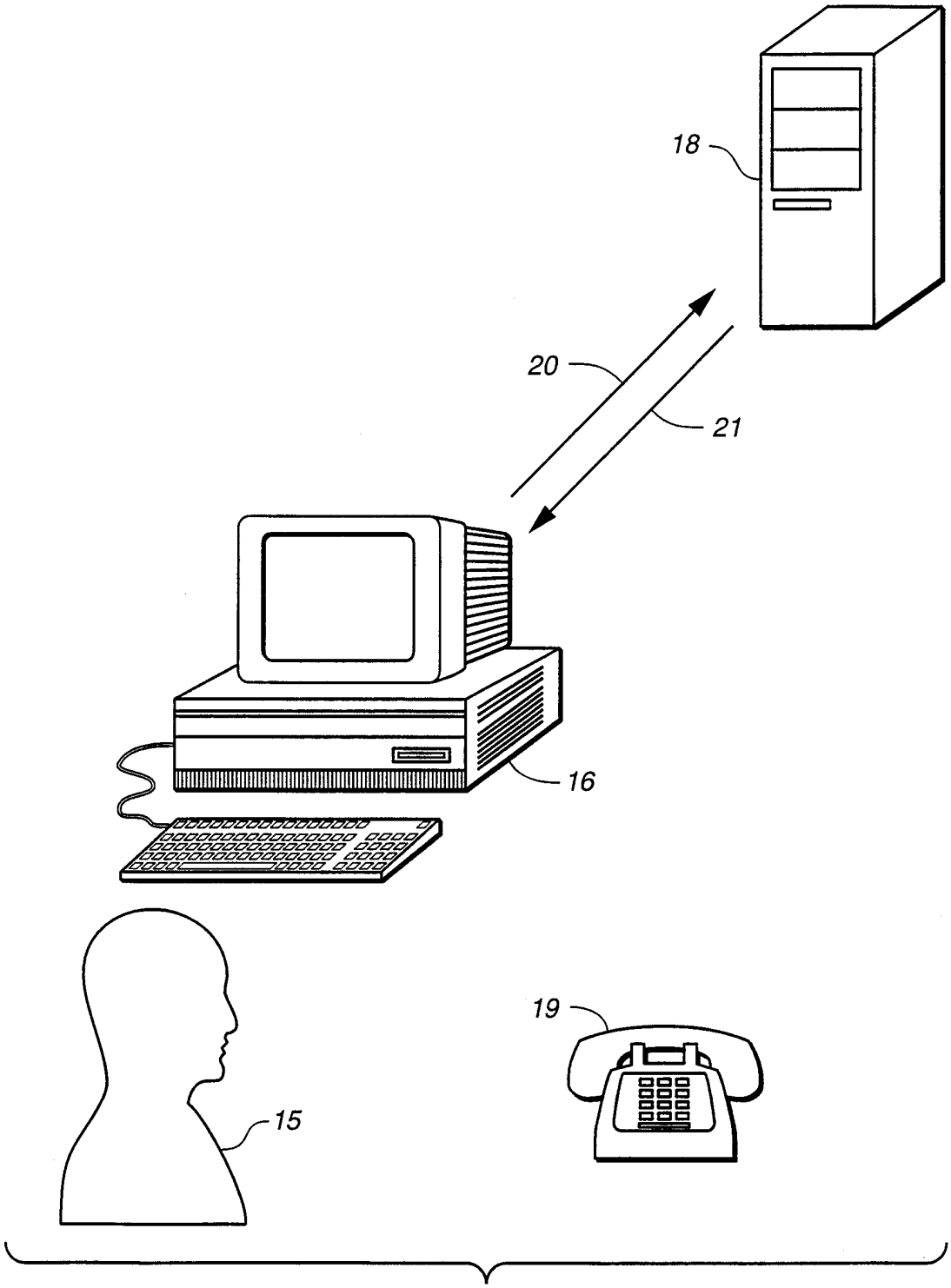


FIG._3

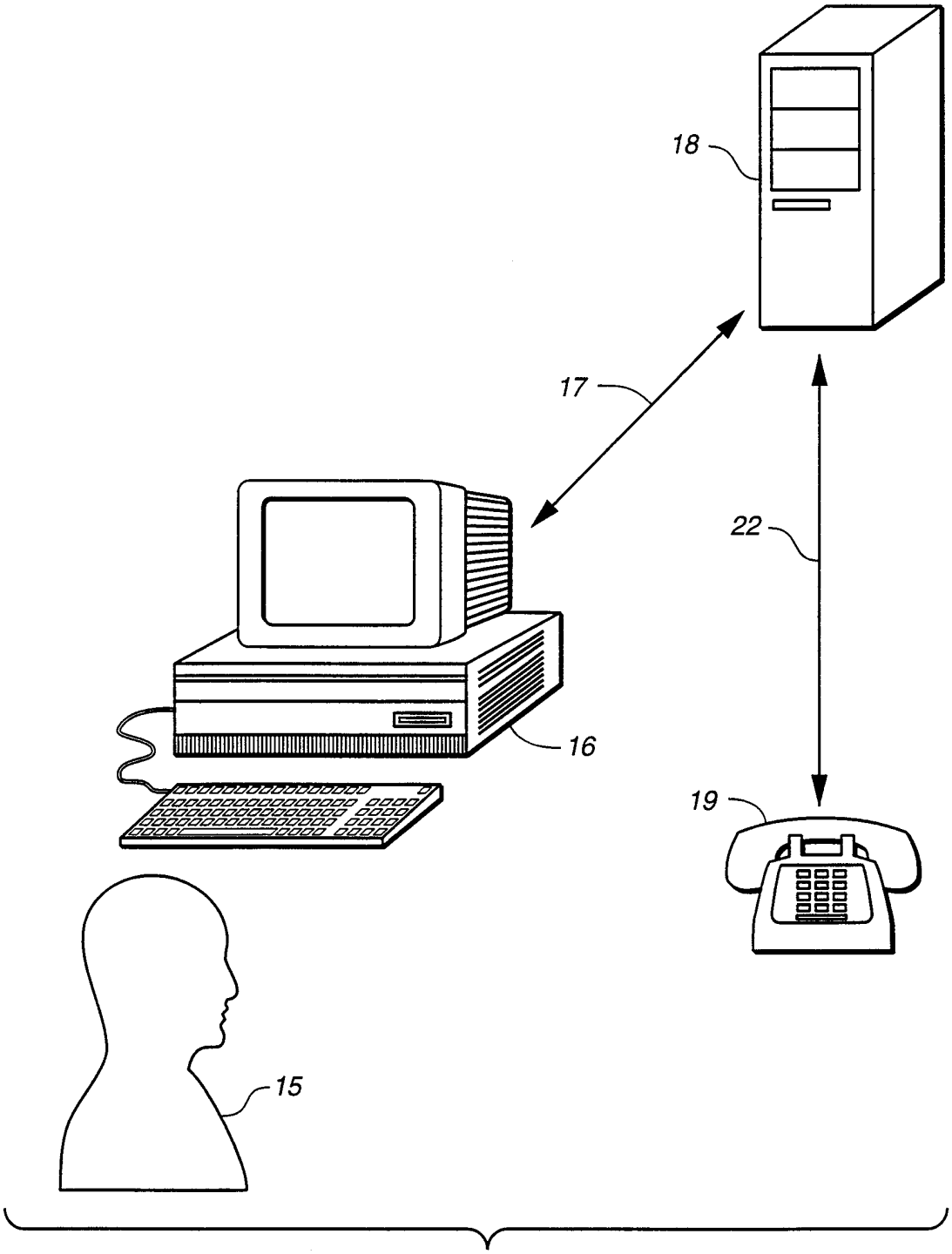


FIG._4

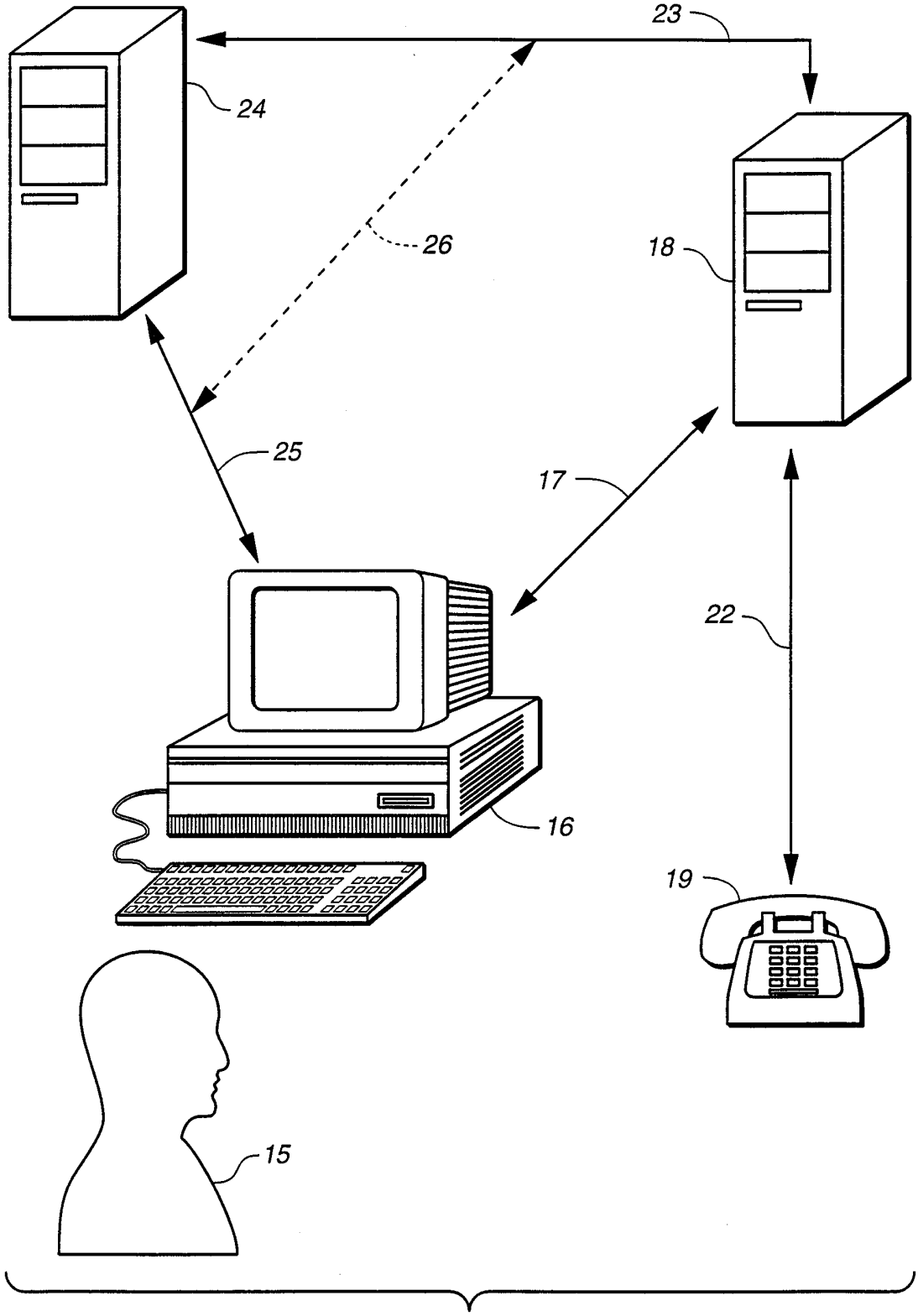


FIG._5

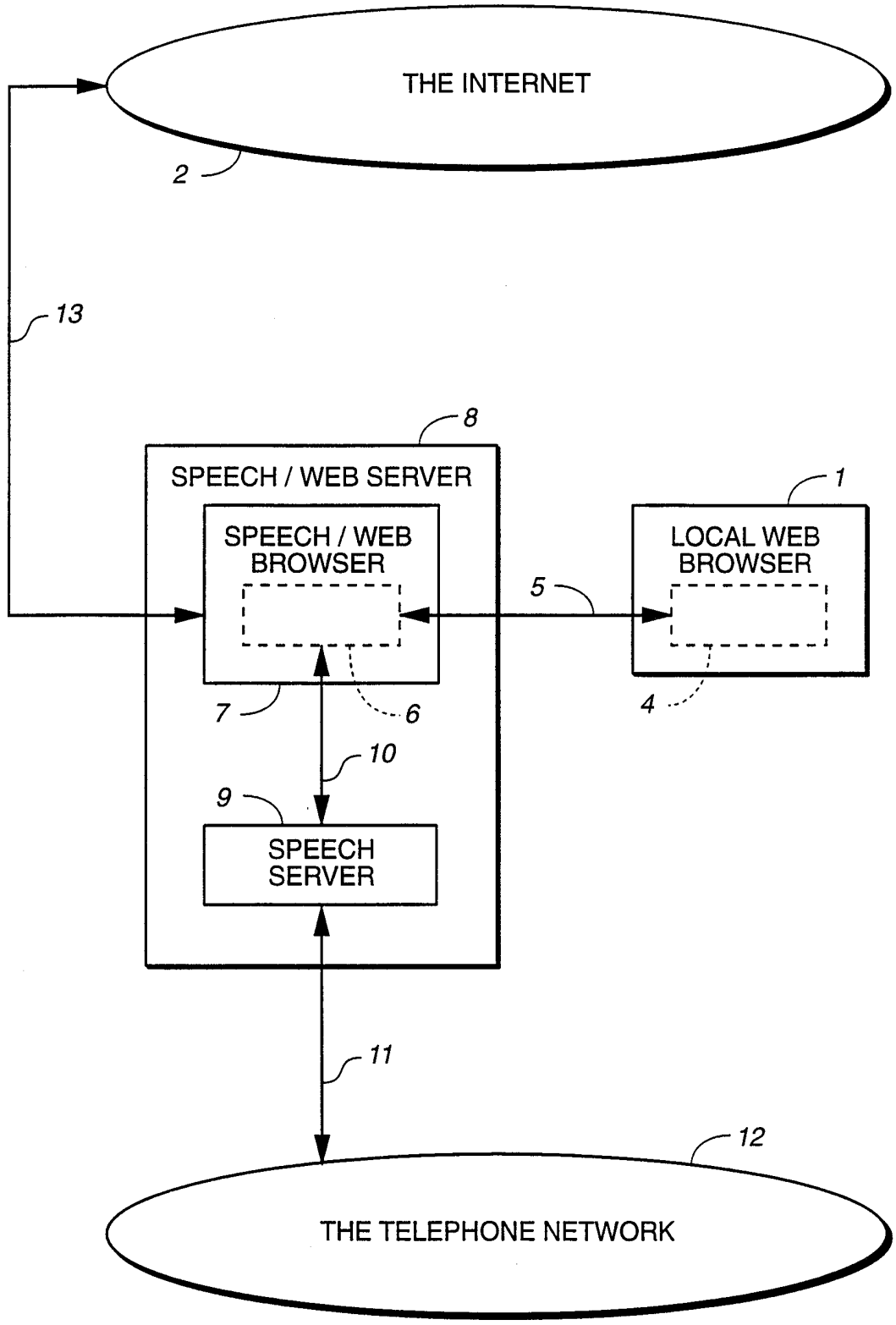

THE INTERNET
2
13
8
SPEECH / WEB SERVER
SPEECH / WEB BROWSER
1
LOCAL WEB BROWSER
5
7
6
4
10
9
SPEECH SERVER
11
12
THE TELEPHONE NETWORK

FIG._6

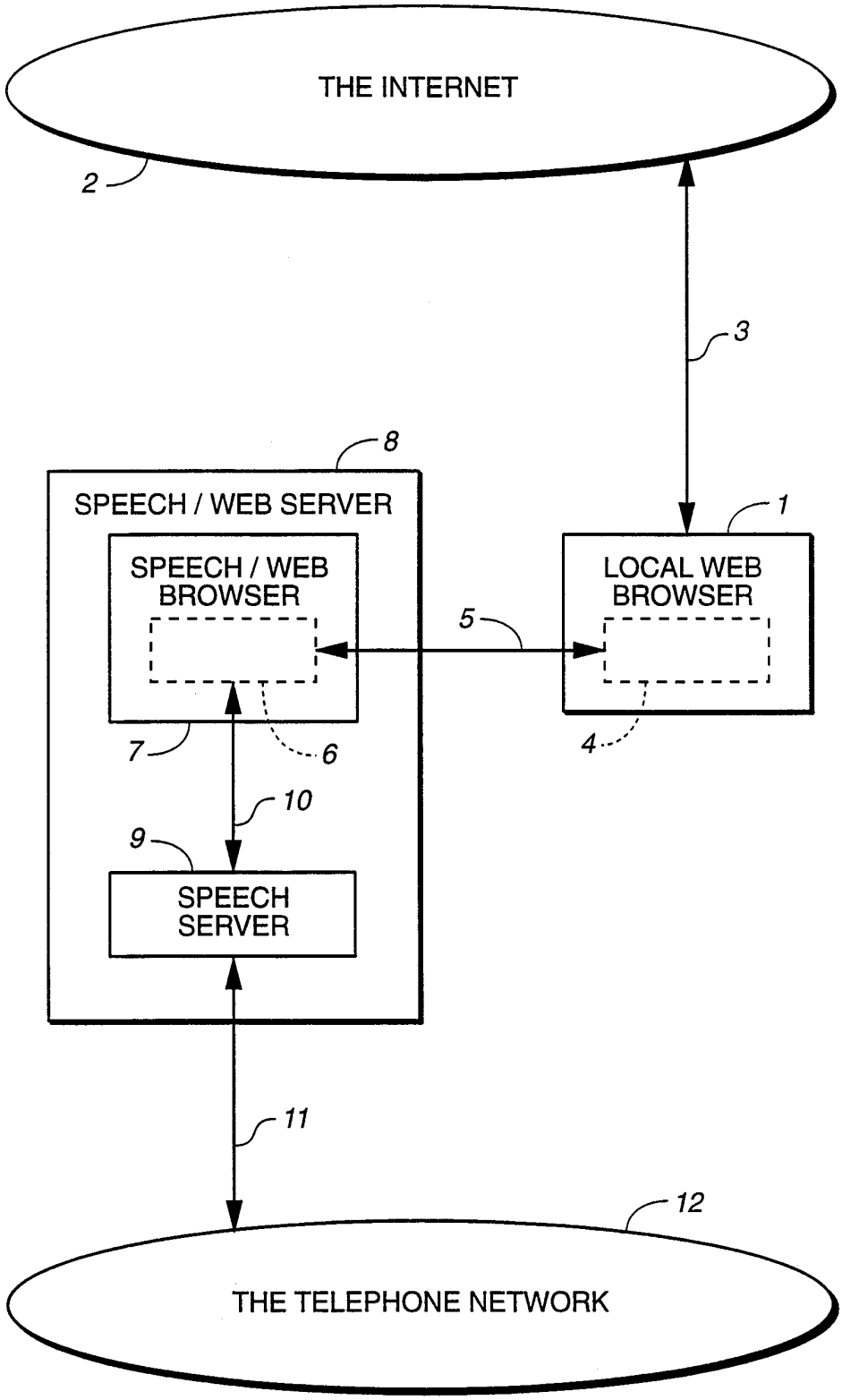

THE INTERNET
2
3
8
SPEECH / WEB SERVER
SPEECH / WEB BROWSER
LOCAL WEB BROWSER
1
5
7
6
4
10
9
SPEECH SERVER
11
12
THE TELEPHONE NETWORK

FIG._7