

SPEECH-ENABLED INFORMATION PROCESSING

FIELD OF THE INVENTION

The invention relates to telecommunications and more particularly to
5 interactive speech applications.

BACKGROUND OF THE INVENTION

Computer-based speech-processing systems have become widely used for a
variety of purposes. Some speech-processing systems provide Interactive Voice
10 Response (IVR) between the system and a caller/user. Examples of applications
performed by IVR systems include automated attendants for personnel directories,
and customer service applications. Customer service applications may include
systems for assisting a caller to obtain airline flight information or reservations, or
stock quotes.

15 Some customer services are also available through the computer-based global
packet-switched network called the Internet, especially through the world-wide-web
("the web") using world-wide-web pages ("web pages") forming websites. These
websites typically include a "home page" containing some information and links to
other web pages of the website that provide more information and/or services. Web
20 pages of various companies allow users to obtain company information, access
personnel directories, and obtain other information, e.g., stock quotes or flight
information, or services, e.g., purchasing products (e.g., compact discs) or services
(e.g., flight tickets). Many websites contain similar categories of web pages of user
options such as company information, a company directory, current news about the
25 company, and products/services available to the user. The web pages can be
navigated using a web browser, typically with navigation tools such as "back,"
"forward," and "home."

SUMMARY OF THE INVENTION

30 In general, in one aspect, the invention provides an interactive speech system

including a port configured to receive a call from a user and to provide a communication link between the system and the user, memory having personnel directory information stored therein including indicia of a plurality of people and routing information associated with each person for use in routing the call to a
5 selected one of the plurality of people, the memory also having company information stored therein associated with a company associated with the interactive speech system, and a speech element coupled to the port and the memory and configured to convey first audio information to the port to prompt the user to speak to the system, the speech element also being configured to receive speech from the user through the
10 port, to recognize the speech from the user, and to perform an action based on recognized user's speech, the speech element being further configured to convey second audio information to the port in accordance with the company information stored in the memory.

Implementations of the invention may include one or more of the following
15 features.

The speech element is configured to convey speech in at least a partially web-analogous format. The speech element is configured to, in response to a request by the user recognized by the speech element, provide information, stored in the memory, according to the request, and to route the call to a person indicated by the
20 user's request according to the routing information associated with the person. Portions of the company information stored in the memory are associated with each other in pages of information according to a plurality of categories of information including how to contact the company. The speech element is configured to act upon the user's speech if the user's speech is within a vocabulary based upon information
25 of a page most recently accessed by the speech element. The categories of information further include information about the location of the company, and products, if any, and services, if any, offered by the company. The company information stored in the memory includes information available on a website of the company. The memory and the speech element are configured to convey the
30 company information to the user with an organization different than an organization

of the company information provided on the company's website. The speech element is configured to access pages of information in response to spoken commands from the user associated with functions commonly provided by web browsers. The commands include "back," "forward," and "home."

- 5 The speech element is configured to perform transactions indicated by the user's speech.

 The system further includes a speech application monitor configured to monitor activity of the speech element and corresponding incoming speech from the user. The speech element is configured to store conversation data in the memory
10 indicative of at least one of: the user's speech; if the user's speech was accepted as recognized, what action if any the speech element took; and if the user's speech has a confidence below a predetermined threshold; and the speech application monitor is configured to report indicia of the conversation data stored by the speech element. The speech application monitor is coupled to the memory through the Internet.

- 15 The speech element is configured to perform at least one of disambiguating the user's speech and confirming the user's speech.

 The system further includes a control unit coupled to the memory and configured to receive a control signal from outside the system and to modify
information content of the memory in response to the control signal. The control unit
20 is configured to add information to the memory, to delete information from the memory, and to alter information of the memory.

- The speech element is further configured to convey information to the user to prompt the user to provide disambiguating information regarding a person and to use the disambiguating information to disambiguate between which of multiple people the
25 user desires to contact.

 In general, in another aspect, the invention provides a computer program product including computer-readable instructions for causing a computer to establish a communication link with a user in response to receiving a call from the user, to retrieve information from a memory having personnel directory information stored
30 therein including indicia of a plurality of people and routing information associated

with each person for use in routing the call to a selected one of the plurality of people, the memory also having company information stored therein associated with a company associated with the interactive speech system, to convey first audio information to the user to prompt the user to speak, to receive speech from the user, to
5 recognize the speech from the user, to perform an action based on recognized user's speech, and to convey second audio information to the user in accordance with the company information stored in the memory.

Implementations of the invention may include one or more of the following features.

10 The instructions for causing the computer to convey the second audio information cause the computer to convey the second audio information in at least a partially web-analogous format. The instructions for causing the computer to convey the second audio information cause the computer, in response to a request by the user recognized by the computer, to provide information, stored in the memory, according
15 to the request, the computer program product further including instructions for causing the computer to route the call to a person indicated by the request according to the routing information associated with the person. The memory stores information in pages according to a plurality of predetermined categories of information, and the instructions for causing the computer to recognize the user's speech cause the
20 computer to use a vocabulary associated with a current page of speech to recognize the user's speech. The company information stored in the memory includes information available on a website of the company, and the instructions for causing the computer to convey the second audio information to the user cause the computer to convey the second audio information with an organization different than an
25 organization of the company information provided on the company's website. The instructions for causing the computer to retrieve information cause the computer to retrieve information in response to spoken commands from the user associated with functions commonly provided by web browsers. The commands include "back," "forward," and "home."

30 The computer program product further includes instructions for causing the

computer to perform transactions indicated by the user's speech.

The computer program product further includes instructions for causing the computer to store conversation data in the memory indicative of at least one of: the user's speech; if the user's speech was accepted as recognized, what action if any the
5 computer took; and if the user's speech has a confidence below a predetermined threshold, and to report indicia of the stored conversation data.

The computer program product further includes instructions for causing the computer to perform an action based on an attempt to recognize the user's speech.

The computer program product further includes instructions for causing the
10 computer to receive a control signal and to modify information content of the memory in response to the control signal. The instructions for causing the computer to modify information content of the memory include instructions for causing the computer to add information to the memory, to delete information from the memory, and to alter information of the memory.

15 The computer program product further includes instructions for causing the computer to: convey information to the user to prompt the user to provide disambiguating information regarding a person, and to use the disambiguating information to disambiguate between which of multiple people the user desires to contact.

20 In general, in another aspect, the invention provides a method of interfacing with a user through an interactive speech application, the method including receiving an incoming call from the user, establishing a communication link with the user, retrieving a portion of stored data indicative of speech for presentation to the user, and presenting the portion of stored data as speech to the user in a web-analogous form.

25 Implementations of the invention may include one or more of the following features.

The stored data are stored in groups according to associated titles indicative of the content of the data in each corresponding group, and the presenting includes conveying the title of the portion of stored data to the user as speech. The method
30 further includes receiving speech from the user, converting the user's speech into

electrical indicia of the user's speech, retrieving another portion of stored data in accordance with the electrical indicia, presenting the another portion of stored data to the user including conveying a title of the another portion of stored data to the user as speech. The user's speech is the title of the another portion of stored data. The
5 indicia of the user's speech are indicative of the title of the another portion of stored data. The indicia of the speech are indicative of a synonym of the title of the another portion of stored data. The user's speech includes a web-analogous navigation command. The web-analogous navigation command is selected from the group consisting of: "back," "forward," "home," "go to," and "help."

10 The stored data are grouped according to content of the data, and the presenting includes conveying a speech indication to the user of the data content of the portion of stored data, the indication including the word "page."

In general, in another aspect, the invention provides a monitoring system for monitoring at least one speech application system, the monitoring system including a
15 computer network connection, and a monitoring unit coupled to the speech application system and to the computer network connection and configured to receive data from the at least one speech application system through the computer network connection, to process call records of indicia related to calls associated with the speech application system, and to produce reports indicative of the indicia related to the calls.

20 Implementations of the invention may include one or more of the following features.

The monitoring unit is coupled to the speech application system through the computer network connection and the monitoring unit is remotely located from the at least one speech application system. The computer network connection is coupled to
25 the at least one speech application system through the Internet. The monitoring unit is configured to access logs of call records stored in the at least one speech application system. The monitoring unit is coupled through the computer network connection and the Internet to a plurality of distributed speech application systems and is configured to receive data from each of the speech application systems through the network
30 connection, to process records of call events associated with each of the speech

application systems, and to produce reports indicative of the indicia related to the calls for each speech application system.

The monitoring unit is configured to transmit signals to the at least one speech application system to alter operation of the at least one speech application system.

- 5 The signals are adapted to cause malfunctioning communication lines of the at least one speech application system to be effectively rendered busy. The signals are adapted to cause services of the at least one speech application system to be restarted. The signals are adapted to cause configuration file patches to be inserted into configuration files in the at least one speech application system.

- 10 The monitoring unit is configured to produce an indication of a frequency of a selected call event.

The monitoring unit is configured to produce an alert regarding a selected call event. The alert is an indication that a characteristic of a selected call event deviates more than a predetermined amount from a predetermined reference value for that

- 15 characteristic.

The monitoring unit and the speech application system are disposed proximate to each other.

Various aspects of the invention may provide one or more of the following advantages. People can access information about, and obtain services from, a

- 20 company using a telephone or other similar device. Information and/or services can be provided and accessed in an audio form and in a format similar to websites, and can be accessed without a computer. A caller can access information and services through natural language speech. A company can leverage an investment in a website or other information proliferation to provide similar information and/or services in an
25 audio, interactive speech format. Callers can navigate through company information and/or services using commands commonly used by web browsers. Interactive speech performance can be monitored. The monitoring can be performed remotely, such as through the Internet. Multiple interactive voice response systems can be remotely monitored. One or more interactive voice response systems can be remotely
30 controlled. The remote control can include establishing and/or altering data such as

configuration parameters and data used in recognizing speech and/or performing actions in response to speech or other sounds.

These and other advantages of the invention, along with the invention itself, will be more fully understood after a review of the following drawings, detailed
5 description, and claims.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a simplified block diagram of a speech system according to the invention.

10 FIG. 2 is a simplified block diagram of a computer system connected to a server through a network link.

FIG. 3 is a simplified block diagram of an IVR system, shown in FIG. 1, and an analog line, a Single Mail Transfer Protocol server, and a fax server.

15 FIG. 4 is a simplified block diagram of an analysis/reporting service connected to multiple IVR systems.

FIG. 5 is a simplified block flow diagram of an interactive speech process according to the invention.

FIG. 6 is a flow diagram of a call routing process shown in FIG. 5.

FIG. 7 is a flow diagram of an information retrieval process shown in FIG. 5.

20 FIG. 8 is a flow diagram of a transaction processing process shown in FIG. 5.

FIG. 9 is a flow diagram of a process for reporting and analyzing interactive conversations

FIG. 10 is a simplified block diagram of an engine system shown in FIG. 3.

25 DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

Overview

Embodiments of the invention provide speech-based information processing systems that are complementary to existing world-wide-web websites and systems. For example, an enterprise or company that has a web-based securities trading system
30 may configure a speech-based information processing system that permits users to be

connected to a broker or inquire about the status of trades placed through the web using the speech-based system, accessed by telephone, and having a user interface that is consistent with the enterprise's website. As used herein, the term "company" includes any variety of entity that may use the techniques described herein. The entity
5 may or may not be a business, and may or may not be intended to earn profit. As such, the term "company" includes, but is not limited to, companies, corporations, partnerships, and private parties and persons. "Company" is used because websites typically use this term, although not necessarily as it is used herein.

Embodiments of the invention support classes of applications that are
10 presently available using web technologies including, but not limited to, communications applications, information retrieval, and transaction processing. Preferably, all such applications are available through a single, consistent user interface that is analogous to a website, including hyperlinks; the user may speak a command for any of several applications without regard to whether one or more
15 servers or systems actually run the applications.

The user can navigate through information presented in a directed dialogue format. During interactive conversation, the user is presented with a set of options with corresponding commands for the user to speak. For example, the user may hear, "You can say 'Contact Us,' 'Company Information,' or 'Products.'" The user may
20 also be presented with a short description of the function of the command, e.g., "You can say 'Fax It' to receive a facsimile about the information that you just heard." Using a directed dialogue can help limit the recognizable vocabulary and speed speech recognition.

Communications applications may include call routing, in which the caller
25 speaks the name of a party or describes a department to which a call is to be routed.

Transaction processing applications may include non-revenue support processing, for example transferring funds from one bank account to another. Since an enterprise typically does not generate revenue for this type of support function, use of a speech interface and speech-based system as disclosed in this document
30 represents a large potential savings in processing costs.

Transaction processing applications may also include e-commerce or purchase transactions. Consequently embodiments of the invention may provide a speech-based gateway to a general-purpose transaction processing system for carrying out commercial transactions, through an online commerce system or a conventional back-office commerce system.

Transaction processing may also include interactive dialogues for enabling a caller to register for events. The dialogue may include identifying an individual by name, address, fax number, etc. The dialogue may also include obtaining payment information by credit card and the like.

Applications may also include registering users to obtain privileged access to one or more services or sets of information, or registering users to enable users to use personalized menus, affinity buying, or "cookies". Applications may also include linking a speech-processing system to other speech-processing systems by one or more circuit-switched carriers, or by Internet telephony ("Voice Over Internet Protocol") connections. Applications may also include providing pointers on a website to one or more speech-processing systems so that users may know what services are speech-activated and may rapidly move from the website to services provided by the speech-processing systems.

Embodiments of the invention also improve access to legacy servers.

Embodiments of the invention serve as front ends or gateways to back-office data servers in the same way that a web server may sit in front of legacy data.

Embodiments of the invention may be configured in association with a web server to provide a convenient interface and presentation layer for the same information retrieval functions or transactions carried out by the web server.

Accordingly, an enterprise can leverage its web investment. Functions that are commonly found on websites may now be accessed by telephone using a natural language speech interface in which the user speaks, e.g., the name of the desired function, such as Contact Us, Job Opportunities, About Us, etc. A particular enterprise may also speech enable functions that are unique to that enterprise. For example, a courier service may offer a Drop-Off Locator service or a Rate Finder

through its website. The same services may be accessed by telephone using embodiments of the invention, whether or not such services are provided on a website.

The caller simply speaks the name of the desired service in response to a system greeting. The information provided in these services may be provided by real-time
5 links to external content providers, as well as more complex actions.

Information retrieval applications may encompass very simple information updates, such as tracking packages that have been shipped using a courier service, tracking luggage moved by an airline, obtaining the balance of a bank account, etc., as well as more complex actions.

10 Another information retrieval application involves providing driving directions to a caller who is trying to reach a location of an enterprise. The caller calls a speech-based system according to embodiments of the invention, and in response to a greeting, says "company information, directions" or the like. In response, the caller is prompted with a query, "What direction are you coming from?" The caller responds
15 with a direction or a point of identification, such as a major road. In response, the caller hears, "I'll read you the directions". The directions are then presented to the caller. As a result, speech delivery of a useful information retrieval function is provided.

Information retrieval functions may also include retrieval of press releases,
20 data sheets, and other electronic documents, and transmission of the electronic documents by text to speech, fax or other media.

Applications may be configured in a variety of ways. For example, embodiments may include tools, packages, and a configuration wizard that enables an operator to set up a new application that provides different information retrieval and
25 transaction processing functions.

Accordingly, embodiments are disclosed that improve telephone answering, leverage a company's investment in the world-wide web and provide a cornerstone or gateway for a variety of information retrieval and transaction processing functions. Embodiments of the invention provide an interactive speech system following a web-
30 based model for organization, information content, and services. A user can use a

telephone to place a call and access information and/or services by speaking naturally with an interactive voice response (IVR) system. Embodiments of the invention thus allow callers to be routed to employees of a selected company by name and/or department, and provide access to company information and transactions with web-site-like organization, terms, and commands. Embodiments of the invention are implemented using software to control a computer processor.

Exemplary embodiments of the invention include a base platform and tool set, and a collection of configurable, pre-packaged application modules. With the base platform and tool set, the tool set can be used to customize a system to provide an individually-tailored interactive speech application. With the collection of configurable pre-packaged application modules, a customer can essentially purchase a turn-key product and with only minor modifications, and configure the product to meet the customer's needs. As embodiments of the invention provide web-site-like functionality in an IVR system, embodiments of the invention may be referred to as a SpeechSite™ IVR system including a SpeechSite™ IVR interface. Within the SpeechSite™ IVR system, speech pages, that are analogous to web pages, provide information and/or services, with different speech pages providing different groups or categories of information and/or services similar to information and/or services commonly provided by websites and in an organization typical of websites.

The following description assumes that a company buys and uses the described embodiments. Thus, it is assumed that the described embodiments provide information about and products/services of the purchasing company. Entities other than companies, however, are acceptable.

25 STRUCTURAL CONFIGURATION

Overall System

Referring to FIG. 1, an interactive speech system 10 includes a user/caller 12, a Public-Switched Telephone Network (PSTN) 14, an IVR system 16, a Simple Mail Transfer Protocol (SMTP) server 18, a firewall 20, a network, here the Internet, 22, and an Analysis/Reporting (A/R) service 24. As shown, communication between

each of the components of system 10 is bi-directional. The caller 12 has access to a phone 26 and a facsimile machine 28. The caller 12 can communicate with the PSTN 14, or vice versa, through either the phone 26 or the fax 28. The caller 12 communicates through the PSTN 14 with the IVR system 16. The IVR system 16 interacts with the caller 12 by playing prompts to the caller 12 in a directed dialogue format and recognizing (or at least attempting to recognize) speech from the caller 12. The IVR system 16 also communicates with the A/R service 24 via the Internet 22. The SMTP server 18 provides an interface between the IVR system and the Internet 22. The firewall 20 protects communications from the IVR system 16 through the Internet 22 and vice versa using known techniques. The IVR system 16 includes an engine system 30, an administration system 32, and a configuration and log system 34.

The systems 30, 32, 34 process the interactions between the IVR system 16 and the caller 12, configure the engine system 30, and store configuration parameters, prompts and other data, and records of interactions with the caller 12, respectively, as described in more detail below.

THE INTERACTIVE VOICE RESPONSE SYSTEM

Introduction

The IVR system 16 can be implemented using a personal computer. For example, the following components and/or features could be used as part of a computer to implement the IVR system 16: a single-processor work station using an INTEL® PENTIUM® III (NT workstation certified) processor with a clock speed of 450 MHz or higher; 384 Mb RAM or more; 9GB of disk space and a high speed DLT backup system; a 10/100 Ethernet connection and a 56K modem for connectivity; a monitor, a mouse, and a keyboard, for data displaying, entering and manipulating data; D41 ESC and D240 SC-T1 telephony interface cards; an Antares 6000/50 digital signal processor; an operating system of NT 4.0 work station service pack 5; an environment of Artisoft® 5.0 Enterprise; an Access or SQL server; an IIS or Pure Information Services HTTP server and a Microsoft® FTP service for a Windows NT

Server for FTP service or an Apache Software Foundation HTTP server; a one-line license from Lucent for text-to-speech (TTS); PolyPM or PCAnywhere programs for remote (e.g., desktop) management.

Referring to FIG. 2, a computer system 50 for implementing the IVR system
5 16 includes a bus 52 or other communication mechanism for transferring information, and a processor 54, coupled with the bus 52, for processing information. The computer system 50 also includes a main memory 56, such as a random access memory (RAM) or other dynamic storage device, coupled to the bus 52 for storing information and instructions to be executed by the processor 54. The main memory
10 56 also may be used for storing temporary variables or other intermediate information during execution of instructions to be executed by the processor 54. The computer system 50 further includes a read only memory (ROM) 58 or other static storage device coupled to the bus 52 for storing static information and instructions for the processor 54. A storage device 60, such as a magnetic disk or optical disk, is
15 configured for storing information and instructions and is coupled to the bus 52.

The computer system 50 is coupled via the bus 52 to a display 62, such as a cathode ray tube (CRT), for displaying information to a computer user. An input device 64, such as a keyboard including alphanumeric and other keys, is coupled to the bus 52 for communicating information and command selections to the processor
20 54. Another type of user input device included in the system 50 is cursor control 66, such as a mouse, a trackball, or cursor direction keys, for communicating direction information and command selections to the processor 54 and for controlling cursor movement on the display 62. This input device typically has control of the cursor in two axes, a first axis (e.g., x) and a second axis (e.g., y), that allows the device to
25 specify positions on a plane.

According to embodiments of the invention, the computer system 50 can generate a speech recognition application in response to the processor 54 executing one or more sequences of one or more instructions contained in the main memory 56. Such instructions may be read into the main memory 56 from another computer-
30 readable medium, such as the storage device 60. Execution of the sequences of

instructions contained in the main memory 56 causes the processor 54 to perform the processes described herein. In alternative embodiments, hard-wired circuitry, firmware, hardware or combinations of any of these and/or software may be used to implement embodiments of the invention.

5 The term "computer-readable medium" as used herein includes any medium capable of providing instructions to the processor 54 for execution. Such a medium may take many forms including, but not limited to, non-volatile media, volatile media, and transmission media. Non-volatile media includes, for example, optical or magnetic disks, such as the storage device 60. Volatile media includes dynamic
10 memory, such as the main memory 56. Transmission media includes coaxial cables, copper wire and fiber opticables, including the wires that comprise the bus 52. Transmission media can also take the form of acoustic or electromagnetic (e.g., light waves), such as those generated during radio-wave and infra-red data communications.

15 Common forms of computer-readable media include, for example, a floppy disk, hard disk, magnetic tape, or any other magnetic medium, a CD-ROM, any other optical medium, punchcards, papertape, any other physical medium with patterns of holes, a RAM, a PROM, an EPROM, a FLASH-EPROM, any other memory chip or cartridge, (e.g., electrical and/or electromagnetic, including optical) a carrier wave as
20 described hereinafter, or any other medium from which a computer can read.

 Various forms of computer-readable media may be involved in carrying one or more sequences of one or more instructions to the processor 54 for execution. For example, the instructions may initially be carried on a magnetic disk of a remote computer. The remote computer can load the instructions into its dynamic memory
25 and send the instructions over a telephone line using a modem. A modem local to the computer system 50 can receive the data on the telephone line and use an infra-red transmitter to convert the data to an infra-red signal. An infra-red detector can receive the data carried in the infra-red signal and appropriate circuitry can place the data on the bus 52. The bus 52 can carry the data to the main memory 56, from which the
30 processor 54 can retrieve and execute the instructions. The instructions received by

the main memory 56 may optionally be stored on the storage device 60 either before or after execution by the processor 54.

The computer system 50 also includes a communication interface 68 coupled to the bus 52. The communication interface 60 provides a two-way data communication coupling to a network link 70 coupled to the SMTP server 18. For example, the communication interface 68 may be an integrated services digital network (ISDN) card or a modem to provide a data communication connection to a corresponding type of telephone line. As another example, the communication interface 68 may be a local area network (LAN) card to provide a data communication connection to a compatible LAN. Wireless links may also be implemented. The communication interface 68 can send and receive electrical and/or electromagnetic (including optical) signals that carry digital data streams representing various types of information.

The computer system 50 can send messages and receive data, including program code, through the SMTP server 18, the network link 70 and the communication interface 68. For example, code can be downloaded from the Internet 22 (FIG. 1) for generating a speech recognition application as described herein. The received code may be executed by the processor 54 as it is received, and/or stored in the storage device 60, or other non-volatile storage for later execution. In this manner, the computer system 50 may obtain application code in the form of a carrier wave.

Referring to FIG. 3, the IVR system 16 includes the engine system 30, the administration system 32, the configuration and log system 34, a remote control system (RCS) 36, a support system 38, and a monitoring interface system 40. These systems can communicate bi-directionally as indicated by the double-ended arrows in FIG. 3. Additionally, the remote control system 36 can communicate bi-directionally with each of the systems 30, 32, 34, 38, and 40. The administration system 32 is responsible for the configuration, reporting, and monitoring of the IVR system 16. The administration system 32 includes a web application server 42 and application server logic 44. Access to the administration system 32 is provided through the web application server 42, which is, e.g., an HTTP server although other types of servers

are acceptable. The application server 42 is controlled by the application server logic 44, that is implemented with software.

THE ADMINISTRATION SYSTEM

5 The administration system 32 is responsible for configuring other components of the IVR system 16. The administration system 32 is configured to access information stored in the configuration and log system 34 and to provide this information as configuration information to other components of the IVR system 16. For example, the administration system 32 is configured to read configuration data
10 from the configuration and log system 34 and to provide this information to the engine system 30. Included in the configuration data sent to the engine system 30 are data that determine which speech pages are active, and the pages' content including locations of prompts, which grammars to use and which vocabularies to use.

 Speech pages are grouped according to configuration information from the
15 administration system 32 into speech modules. Different speech modules provide different categories of information and services provided by the pages within the modules. Each module contains multiple speech pages. Exemplary modules are SpeechAttendant, Contact Us, and Company Information. The SpeechAttendant module includes pages for a personnel directory for the company. The Contact Us
20 module includes information for contacting the company, e.g., an email address, mailing address, street address and directions, fax number and a telephone number. The Company Information module includes pages describing general company information such as the type of business and/or services provided by the business, news releases, and geographical areas serviced.

25 The administration system 32 is also configured to update/edit information contained in the configuration and log system 34. Thus, for example, the administration system 32 can access and edit employee name and telephone lists, company information such as news, and call flow control data.

THE ENGINE SYSTEM

30 The engine system 30 is responsible for and configured for executing the

speech interface between the IVR system 16 and the caller 12 (FIG. 1) and connecting with the telephony system. Thus, the engine system 30 is connected to the PSTN 14 for bi-directional communication and is adapted to execute a call flow with the caller 12 including recognizing speech, playing prompts, retrieving data, routing calls, 5 disambiguating responses, confirming responses, selecting pages to link to, and connecting to support devices (e.g., TTS and fax). The engine system 30 is configured to recognize speech using known techniques, here parsing the speech into segments and applying acoustic models. Functions of the engine system 30 are implemented using DialogModule™ speech-processing software units available from 10 SpeechWorks® International, Inc. of Boston, MA. For at least each of the functions of call routing, information retrieval, and transaction processing, the speech engine 30 is adapted to attempt to recognize the caller's speech and act accordingly.

The engine system 30 also includes an execution engine 80 configured to control the processing of the engine system 30. This processing includes retrieval and 15 playing of prompts, speech recognition, and monitoring and reporting of interactions between the engine system 30 and the caller 12. The engine system 30 is controlled by instructions and/or data stored in the engine system 30 or the configuration and log system 34 or provided by the caller 12.

Referring also to FIG. 10, the engine system 30 includes the execution engine 20 80, DialogModule™ speech-processing units 300, a SMARTRecognizer™ speech recognizer 302 (available from SpeechWorks® International, Inc.), a prompt unit 304, a record unit 306, all operating in a Service Logic Execution Environment (SLEE) 308 including an Operating System (OS), and operating on hardware 310. The SLEE is the computational environment in which call logic executes. This environment 25 includes any tools provided by the Artisoft® Visual Voice service platform for call and event handling. The execution engine 80 is configured to control operation of the engine system 30.

The engine system 30 is configured to recognize speech and to adapt to new speech. The speech-processing units 300 are configured to process waveforms of received utterances by the caller 12 and SLEE logs (discussed below) and to provide the processed data to the recognizer 302. The recognizer 302 uses acoustic models, semantic models (probabilities of word phrases), pronunciation graphs, and a dictionary, stored in the IVR system 16, to attempt to recognize a word or words in the caller's utterance. Acoustic models represent statistical probabilities that given waveforms relate to associated parts of speech. The recognizer 302 can produce an N-best list of N words or phrases most likely to be what the caller 12 spoke. Each item in the N-best list has a corresponding confidence score representing the likelihood that the identified word or phrase is actually what the caller 12 spoke. The engine system 30 is configured to load the models and parameters that control the engine system 30 each time the engine system 30 runs. These data are stored in the configuration and log system 34 such that they are replicable in an off-line batch processing mode.

The recognizer 302 can build or add to a dictionary and adapt acoustic models, semantic models, and pronunciation graphs to adjust probabilities linking utterance waveforms to speech. Acoustic model retraining can be performed during inactive times of the IVR system 16. The recognizer 302 is configured to build and evaluate semantic models using parsed and raw text and to use the SLEE logs to automatically build the semantic models.

To implement the auto attendant functionality, the speech engine 30 is configured to perform call routing functions. These functions include recognizing names and/or departments and/or positions of employees, including synonyms of the employees' names, departments, and/or positions, and providing prompts to the caller.

To perform the call routing features, the execution engine 80 is configured to retrieve information from the configuration and log system 34 and to compare this with the caller's speech to determine whether the spoken name, department or position corresponds to a person or department of the company or business associated with the IVR system 16. The engine system 30 can disambiguate responses, e.g., by prompting the caller 12 to identify an employee's department in addition to the

employee's name. The call routing functions also include the execution engine 80 connecting the caller to the requested person/department. In particular, the execution engine 80 is configured to transfer calls using blind, flash-hook transfers in accordance with data stored in the configuration and log system 34. The engine system 30 can be configured to perform other types of transfers such as supervised transfers.

For information retrieval functions, the engine system 30 can identify a specific speech page requested by the caller 12 or determine which page will contain the information requested. The engine system 30 is configured to recognize speech from the caller 12 in order to determine what information the caller 12 is requesting. In response to recognized speech from the caller 12, the engine system 30 is configured to access the specified/determined page, and play prompts to the caller regarding the information requested. Thus, for example, the engine system 30 can link to a "Contact Us" page in response to the user/caller 12 saying "Contact Us" when at an appropriate page (one providing for links to the contact us page). Additionally, the engine system 30 can link to the Contact Us page in response to the user, when at an appropriate page (one providing for links to the contact us page), requesting information from the Contact Us page. For example, the caller 12 could say "Directions to Boston" and the engine system 30 will link the caller 12 to the Contact Us page.

The engine system 30 is also configured to perform transactions available to the caller 12 for the specific IVR system 16. To perform transactions, the engine system 30 is configured to identify a specified page or to determine a page containing information or services (e.g., stock trading) requested by the caller 12, to access the specified/determined page, to play prompts for the caller 12, and to perform/initiate functions specified by the caller 12. The engine system 30 can recognize the caller's speech and associate the recognized speech with data stored in the configuration and log system 34. In accordance with instructions stored in the engine system 30 and/or data, including instructions where appropriate, stored in the configuration and log system 34, the engine system 30 can perform the transactions indicated in accordance

with the data provided by the caller 12.

To control the interactive conversation with the caller 12 the engine system 30 can interact with the caller 12 through the PSTN 14 and the configuration and log system 34. The engine system 30 can receive speech from the caller 12 through the
5 PSTN 14. The engine system 30, under control of the execution engine 80, is configured to attempt to recognize the speech from the caller 12. In order to recognize the speech, the engine system 30 can access the configuration and log system 34 for information indicating what speech is recognizable by the IVR system 16. The engine system 30 is configured to, under the control of the execution engine
10 80, manage the conversation between the IVR system 16 and the caller 12. The execution engine can instruct the engine system 30 to output prompts, stored in the configuration and log system 34, to the caller 12 depending on whether the engine system 30 recognized speech from the caller 12. These prompts can be, e.g., to request information from the caller 12 in accordance with the previously-recognized
15 speech, to ask the caller 12 to retry the speech for unrecognized speech or low-confidence recognized speech, or other appropriate error messages for nonspeech information received by the engine system 30.

The engine system 30 is configured to communicate with the caller 12 in a directed dialogue manner that guides the caller 12 to speak within a limited
20 vocabulary to achieve the caller's desired results. The engine system 30 presents the caller with possible commands to speak, and uses a recognition vocabulary that includes the possible commands (e.g., "Contact Us"), as well as some synonyms or other words that convey similar intent (e.g., "Directions"). The recognition vocabulary changes with different stages of the interactive dialogue with the caller 12.
25 The caller 12 can also say any of several "global" or "universal" commands that are available at any speech page such as "Help," "Back," or "Forward." The "Back" and "Forward" commands would only work if there are speech pages that are behind or in front of the current speech page in the history of speech pages that the caller 12 has visited. Utterances other than those permitted would result in error messages to the

caller 12. Limiting the available recognizable speech can help improve recognition accuracy and speed, and overall robustness of the IVR system 16.

The engine system 30 is configured to use known techniques to recognize and respond to requests by the caller 12. For example, the speech from the caller 12 can be parsed into units of speech and transformed by a digital signal processor to produce speech unit vectors. These vectors are grouped into speech segments that may be of varying lengths. The segments are converted into feature vectors that are analyzed with respect to linguistic constraints (e.g., the recognition vocabulary) to produce an N-best list of the N word strings with the highest confidence scores.

The engine system 30 interacts with the caller 12 by presenting a user interface that translates into an audio format the visual format provided in typical websites, including functionality provided by typical website browsers. The user interface is accomplished by the prompts accessed by the engine system 30 and presented to the caller 12. Similar to information provided to a person browsing a website, prompts played by the engine system 30 stored in the configuration and log system 34 can, e.g., inform the caller of the current location (e.g., "Home page") or the location being transferred to (e.g., "transferring to the Contact Us page"). These prompts played for the caller 12 use terminology (e.g., for links) associated with websites such as "page", "Contact Us", "Company Information", "going back to page...", "moving forward to ... page". As another example, the caller 12 can be played a prompt for information that says "You can say [Page 1] [Page 2]" The [Page 1] and [Page 2] prompts are text prompts that are configurable, e.g., by storing the custom text in the configuration and log system 34, and are retrieved to play the prompt to the caller 12. This text is substituted in the greeting, e.g., in real time, by the engine system 30 to provide the customized text to the caller 12 based on the information stored in the configuration and log system 34. Exemplary customized pages would be for, e.g., specific products of a company, specific divisions of a company and/or specific services of a company. Still other prompts can present queries to the caller 12, such as yes/no queries. Information presented to the caller 12 by the engine system 30 may be organized

differently than a corresponding website containing essentially the same information as that presented to the caller 12.

As part of the web-analogy user interface, the engine system 30 is configured to respond to website-like commands to navigate through the information in the IVR system 16. A caller can say commands typically provided by web browsers such as "home", "back", "forward," "help," and "go to" and the engine system 30 is configured to recognize such commands and act accordingly. The engine system 30 thus can transfer the caller 12 back to the previous page, or to the next page, of information in the history of pages visited, or back to the home page, in response to the caller 12 saying, respectively, the exemplary commands listed above.

For each speech page, there may be links particular to that page for which the engine system 30 will prompt the caller 12. For example, the home page can have specific links to a Company Information page, a Contact Us page, and a Products/Services page, and prompts informing the caller 12 of the links. For example, the caller 12 can be told "You can say 'Company Information,' 'Contact Us,' or 'Products and Services.'" The engine system 30 can transfer to any of these specific pages as requested by the caller 12.

The engine system 30 is also configured to provide search services and automatic faxing services to the caller 12. From an appropriate point in a dialogue with the IVR system 16, the caller 12 can say "find it." In response to this request/utterance, the engine system 30 can search the stored speech pages for the indicated text and/or information. The caller 12 at anytime may also say "fax it" and the engine system 30 is configured to respond to this request by faxing the content of the current page (associated with the current speech page) to a fax number specified by the caller 12. The fax number may be previously stored and, possibly, confirmed, or may be requested by the engine system 30 in response to the fax it command.

The engine system 30 is configured to record call events and other transactions in the configuration and log system 34. Call events are the various stages of the interactive conversation between the IVR 16 and the caller 12. These events include requests by the caller 12, attempted recognition by the engine system 30, indications

of whether the caller's speech was recognized, whether the speech was rejected as a low-confidence recognition or was not recognized, actions initiated by the engine system 30, and prompts played to the caller 12. The events can include what pages the caller was directed to, what commands the caller 12 requested and in what sequence, and which actions the engine system 30 performed. The engine system 30 is configured to send indicia of the call events to the configuration and log system 34 for storage for future reference. The engine system 30 is configured to send indicia of some call events each time the call event occurs, while transferring indicia of other call events only upon the occurrence of certain conditions, and can be configured not to send indicia of other call events at all. For example, the engine system 30 can send indicia each time that a low-confidence rejection occurred, or that a high-confidence acceptance occurred (e.g., a person was successfully connected to using auto attendant features). The engine system 30 is also configured to produce reports based on call statistics of the call events for storage in the configuration and log system 34 and/or for retrieval by the monitoring interface system 40.

THE CONFIGURATION AND LOG SYSTEM

The configuration and log system 34 includes a log storage area 86, a database storage area 88, and a general storage area 90. The configuration and log system 34 is configured to store the information used by the administration system 32, the engine system 30, the support system 38 and the monitoring interface system 40, and to interact bi-directionally with each of these systems. Thus, each of these systems 30, 32, 38, 40 can retrieve information from, and store information to, the configuration and log system 34.

The database 88 stores the configuration files and content of the speech pages. The content of the speech pages resembles content and format commonly available on websites. The configuration files are used to configure the systems 30, 32, 36, 38, 40. These files store the information necessary to configure each of these systems 30, 32, 36, 38, 40 during configuration and set up as described below. These files can be established and/or modified by a manufacturer and/or purchaser of the IVR system 16

to provide and/or alter custom configurations. The database 88 is also configured to store a variety of information relating to the speech pages. For example, the database 88 is configured to store information related to prompts. Prompt data include the identification, date recorded, name of person recording the prompt, source, and type of the prompt. Additionally, whether the prompt is published, a unique user interface name for the prompt, and the text of the prompt are stored in the database 88. Also, the location of the prompt in the database 88 and the date that the prompt was produced are stored in the database 88. Information is also stored in the database 88 for linking multiple prompts together to form phrases or other segments of prompts.

10 The database 88 also stores information relating to speech modules. This information includes identifying information for speech modules and speech pages, including the contents of each of the modules and pages. This identifying information is adapted to be used by the speech engine 30 to locate, retrieve, and process the various speech modules and speech pages, including the prompts contained therein.

15 The database 88 also stores data relating to speech pages. Links between pages and components of the pages are contained in the database and help the speech engine 30 link to other pages and/or modules to more readily retrieve information and perform other actions. The database 88 also stores information linking DialogModule™ speech-processing units 300 (FIG. 10) to specific speech pages and/or speech modules. The linking information for DialogModule™ speech-processing units 300 (FIG. 10) provides a mapping for determining which DialogModule™ speech-processing units 300 (FIG. 10) to execute when executing a speech page.

25 Data stored in the database 88 also provides links between data serving as synonyms for one another. This helps to increase accuracy of recognition for an item when synonyms are available for that item.

30 The database 88 also stores several other types of information. Included in this information is information to help support navigation of speech pages, including navigation terms and links to navigation functions, and key words used to locate a speech page when executing a "find it" function by the execution engine 80 in the

engine system 30. A user dictionary is also stored within the database 88. The database 88 also contains information related to the operations of the company. For example, the dates and/or hours of operation of the company are stored in the database 88. The database 88 also stores the information for the personnel directory for the auto attendant functionality. Information stored in the database 88 for the personnel directory is stored in fields of data. Exemplary fields include personnel names, nicknames, positions, departments, synonyms of entries in any of these fields, and telephone extensions for persons, rooms, and departments, or other information for transferring/routing callers to persons and/or departments, etc. These fields can be updated to reflect new personnel, changes of departments, changes in names of departments, changes in names and people and additional nicknames or other synonyms.

The stored speech-page content includes all the information, including prompts (e.g., queries and information), layout and links, for each page of the IVR system 16. Speech-page content is divided into fields of data that can be selected/chosen and modified to custom configure the page content before transfer to a purchaser/customer, or by a customer of the IVR system 16. The content of the speech pages can be updated as necessary by modifying the data fields, e.g., to update stock prices, provide current daily news, or to indicate any changes that may have occurred in the company.

The storage area 90 stores all prompts, fax pages, GIFs, and speech models. The prompts are all the audio information to be relayed to the caller 12. For example, the prompts include queries to the caller 12 and statements of information for the caller 12. The fax pages are the data to be transmitted to the fax 28 (FIG. 1) of the caller 12 in response to the caller requesting faxed information, e.g., the caller 12 saying "fax it" or the like. Graphical information in the form of Graphics Interchange Format (GIF) files can be included in the fax pages. Speech models are used by the engine system 30 to recognize portions of speech in order to recognize words and/or phrases spoken by the caller 12.

The log storage area 86 is configured to store logs of call events and other information needed by the system 40, e.g., in Service Logic and Execution Environment (SLEE) logs. Logs of the call events includes statistics regarding e.g.,
5 call time, call length, speech pages requested, success rates of recognition, out of vocabulary recognitions, failed recognitions, and commands used.

THE SUPPORT SYSTEM

The support system 38 is configured to be invoked by the administration
10 system 32 and/or the engine system 30 and to provide support features for these systems 30, 32. The support system 38 includes a text-to-speech (TTS) feature 92, a log converter 94, a fax feature 96, a report generator 98, and a speech adapter 100.

The TTS 92 allows the engine system 30 to output speech or other appropriate audio to the caller 12 based upon text stored in the IVR 16. The TTS 92 can be
15 implemented using known techniques, such as the Lucent TTS engine. The TTS 92 allows the IVR 16 to be updated quickly. For example, a news release can be quickly stored in the configuration and log system 34 as text and can immediately be output as speech to a caller 12 using the TTS 92 and the engine system 30. A recording, such as
20 by a famous personality of that news release can be made later and used in place of the TTS 92 converting the text of the news release to appropriate speech using the engine system 30. Other portions of the IVR 16 can also be updated in this manner, for example, the list of employees in the personnel directory.

The log converter 94 is configured to convert information in the logs stored in the storage area 86 to an appropriate format for processing by the report generator 98.
25 Here, the log converter 94 is configured to access SLEE files stored in the storage area 86 and to convert these files into National Center for Supercomputing Applications (NCSA) standard logs. The log converter 94 can effectively convert indicia of accesses by callers to the IVR 16 into the equivalent of a website page "hit", and stores these hits in a file. Thus, the log converter 94 is configured to store a file
30 containing an ID of the caller 12 (e.g., a phone number), the date and time of a request

by the caller 12, and indicia of a request from the caller 12 for information or an action. The logs stored by the log converter 94 are stored in the configuration and log system 34.

5 The fax feature 96 is configured to process fax requests from the caller 12 to fax the requested information to the fax 28 (FIG. 1) accessible by the caller 12. For example, the fax feature 96 can be implemented using WinFax Pro 9.0. This implementation of the fax feature 96 supports both RightFax servers and Internet servers. The fax feature 96 can fax information to a fax number associated with the fax 28, provided by the caller 12, through a fax server 97.

10 The report generator 98 is configured to access logs and other information stored in the configuration and log system 34 and to manipulate these data to produce various reports. For example, the report generator 98 can manipulate the logs stored by the log converter 94 to produce reports relating to speech page hits. The report generator 98 is configured to produce reports indicating the number of calls per hour, 15 the number of calls per hour in all speech modules, and the number of operator transfers per hour. The report generator 98 may also be able to produce a report indicating the number of calls from a given device as identified by its automatic number identifier (ANI) in a selected day/week/month. These reports are produced in written and graphical formats, and are downloadable and can be imported into a 20 database.

The speech adapter 100 is configured to adapt tools used by the engine system 30 to help improve the speech recognition by the engine system 30. The speech adapter 100 can be implemented with LEARN 6.0 software available from SpeechWorks® International, Inc. of Boston, Massachusetts. The speech adapter 100 25 can access information stored in the configuration and log system 34, analyze this information, and determine how to adapt the acoustic models, pronunciation graphs, and/or semantic models stored in the configuration and log system 34 to improve speech recognition by the engine system 30. The speech adapter 100 is also configured to update/change the acoustic models, pronunciation graphs, and/or 30 semantic models according to these determinations. The new models and graphs are

stored again in the configuration and log system 34 for use by the engine system 30 in recognizing speech from the caller 12.

THE REMOTE CONTROL SYSTEM

5 The remote control system (RCS) 36 is configured to provide remote control of the IVR 16 through an analog communication line 104. The RCS 36 includes a remote access system (RAS) 106 controlled by appropriate software, here, PCAnywhere 108. The RAS 106 communicates with the analog line 104 through a modem 110.

10 The RCS 36 allows arbitrary control of the IVR 16 through an NT window. For example, the RCS 36 is configured to allow start/stop processing, to modify the configurations of the systems 30, 32, 34, 38, 40, including data stored therein, to access the administration system 32 and to enable/disable communication lines connected to the IVR 16.

15

THE MONITORING INTERFACE SYSTEM

 The monitoring interface system 40 provides monitoring functions for the IVR 16 and includes a system monitor 112, a prompt monitor 114, and a tuning monitor 116. Each of these monitors 112, 114, 116 is configured to retrieve information from and store information to, in the form of ulaw files (μ -law files; word waveforms), the configuration and log system 34, and to bi-directionally communicate with the SMTP server 18. The prompt monitor 114 is configured to monitor prompt changes and provide alerts as to the changes.

25 The system monitor 112 is configured to monitor computer functions of the IVR 16, to take appropriate actions in response to the monitored functions and to provide a "base heartbeat" to the A/R service 24 (FIG. 1). The base heartbeat is a message sent to the A/R service 24 to inform the A/R service 24 that the IVR 16 is operational and functioning with normal operational parameters. Alarms and alerts are provided by the system monitor 112 for hardware and telephony errors, resource limits, run time errors, and/or transactional errors. The errors for the resource limits

30

apply to the application software code in the IVR 16. Run time errors are provided for SLEE, speech recognizer, and DialogModule™ speech-processing unit libraries. The SLEE libraries are configured to accept the call from the caller 12 and invoke the engine system 30 including its speech recognizer. Run time and transactional errors
5 in the IVR 16 software code include all varieties of errors encountered in processing the call from the caller 12. The system monitor 112 can report transaction errors by storing indications of these transaction errors in the configuration and log system 34. The system monitor 112 is also configured to perform some remedial action such as restarting selected non-critical services of the IVR 16. Alarms and alerts can be sent
10 by the system monitor 112 to the A/R service 24 (FIG. 1) via the Internet (FIG. 1).

The tuning monitor 116 is configured to monitor and analyze speech performance regarding the interaction between the caller 12 and the IVR 16. The tuning monitor 116 is configured to compute performance statistics from the SLEE logs stored in the configuration and log system 34 and to track the performance
15 statistics. From the performance statistics, the tuning monitor 116 can send alerts about these performance statistics. The tuning monitor 116 can also send, for external monitoring, the SLEE logs and waveforms of portions of speech from the caller 12 that are flagged as potentially problematic waveforms. The tuning monitor 116 can also output status messages regarding conversation statistics. These alerts, logs,
20 waveforms, and messages can be sent by the tuning monitor 116 over the Internet 22 (FIG. 1) to the A/R service 24 (FIG. 1).

The tuning monitor 116 is configured to provide numerous reports regarding performance statistics of the conversation between the caller 12 and the IVR 16. The tuning monitor 116 can analyze the performance statistics according to several criteria
25 such as transaction completion rates for important transactions, DialogModule™ speech-processing unit completion rates, failed calls, caller-perceived response time, personnel names not accessed within predetermined amounts of time, average call time, percentage of short calls, numbers of disconnected versus transferred calls, number of calls transferred to operator, and call volume. Which transactions are
30 designated as important transactions may be configured upon system setup or

modified later. DialogModule™ speech-processing unit completion rates information includes how much confirmation is occurring, and how much and often failures occur.

Information regarding DialogModule™ speech-processing unit completion rates is formatted according to speech pages associated with the DialogModule™ speech-processing units 300 (FIG. 10). Caller-perceived response time can be used to determine whether the IVR 16 is being overloaded. The predetermined time for personnel names not being used can be selected as desired, e.g., 1, 6, and/or 12 weeks. Reports regarding the number of disconnected calls versus transferred calls, and the number of calls transferred to operators may be useful in analyzing the auto attendant performance.

The tuning monitor 116 also can provide several business reports. For example, reports for the number of calls per hour, number of calls per hour in important dialogue legs, the number of operator transfers per hour, and the number of calls from a given ANI in a predetermined time period are provided. The number of calls per hour are provided in a downloadable format as both text and graphs. Important dialogue legs can be defined by the configuration file stored in the configuration and log system 34. The predetermined amount of time for the ANI reports can be, e.g., a day, a week, and/or a month. These reports are provided in an exportable format via a text file for loading into a database for business data mining and other reporting functions.

Alarms can be triggered by the tuning monitor 116 for a wide variety of failures and/or performance conditions. These alarms are sent as structured messages, using, e.g., Single Network Management Protocol (SNMP) or email, to one or more destinations. Alarms may help off-site monitoring of system performance by a customer operations center, an off-site outsourcing monitoring company, and/or the entity selling and/or configuring the IVR 16.

THE ANALYSIS/REPORTING SERVICE

The tuning monitor 116 can send reports over the Internet 22 (FIG. 1) to the A/R service 24 (FIG. 1). Referring again to FIG. 1, the A/R service 24 is configured

to monitor performance of the IVR 16 and to provide alarms to initiate diagnostic action regarding the IVR performance. The diagnostic action can be taken by, for example, the vendor of the IVR 16 such as SpeechWorks® International, Inc. The A/R service 24 is configured to access data in, and retrieve data from, the
5 configuration and log system 34, to analyze these data and to determine appropriate actions, to produce appropriate alarms, and/or to produce appropriate reports. The A/R service 24 is configured to periodically access and/or receive data such as files of recorded utterances and SLEE logs stored in the configuration and log system 34 for use in recognition, tuning, monitoring and report production.

10 One alarm producible by the A/R service 24 is for potentially high OOV rates. A high OOV rate can be caused, e.g., by the list of names in the personnel directory stored in the configuration and log system 34 not being maintained. Thus, when the caller 12 asks to be routed to a particular name, the IVR 16 may reject the requested name as being unrecognized despite the fact that the requested person is an employee
15 of the company serviced by the IVR 16.

Alarms and/or reports can be produced by the A/R service 24 for likely candidates for synonym/nickname identification. When an unrecognized phrase or a low-confidence phrase is accepted as recognized (with high confidence) on retry (e.g., caller: "your CEO"; IVR: "I did not understand. Please say the first and last name.";
20 caller: "Stuart Patterson.") the phrase used by the caller 12 in the first try is a good candidate as an additional synonym for the person recognized in the second try. The A/R service 24 can produce a report indicating the potential synonyms (e.g., CEO) and the recognized speech (e.g., Stuart Patterson).

Alarms can also be produced by the A/R service 24 for repeated bad
25 pronunciations. A high percentage of confirmations for a given phrase spoken by the caller 12 is an indication that the IVR 16 is programmed with a poor pronunciation relative to the phrase. An alarm indicating repeated confirmation of identified words/phrases can be used to initiate action to adjust the pronunciation recognized by the IVR 16 to reduce the number of confirmations required for the particular
30 word/phrase.

The A/R service 24 also is configured to produce alerts, alarms, and/or reports for names no longer being recognized by the IVR 16. A high rate of a name having a low-confidence score that previously had a high-confidence score can indicate any of several problems such as poor pronunciation recognition and/or high noise levels
5 and/or that the person is a former employee and should be listed in the database 88 as having left the company.

The A/R service 24 is configured to monitor confidence distributions to help manage recognition thresholds for confidence scores. The IVR 16 may adapt to improve recognition of the caller's speech to increase recognition accuracy. In so
10 doing, confidence score distributions across all utterances may shift. A fall in the percentage of rejected utterances, however, may indicate a rise in false acceptances (i.e., misrecognitions that are accepted as being valid) due to acceptance thresholds that are too low. Conversely, if the rejection thresholds are too high, rejection rates may be artificially high, inhibiting true recognition accuracy from being realized. The
15 A/R service 24 can monitor the confidence distributions and affect the thresholds to help achieve proper thresholds to help realize true recognition accuracy. The A/R can also produce alarms indicating, or other indicia of, confidence scores for personnel, and rejection rates.

The A/R service 24 can also provide indicia of disambiguation configuration
20 problems. Alarms can be generated if the disambiguation prompt to the caller 12 is unsuccessful in helping the caller 12 to distinguish the information the caller 12 is seeking. For example, if the disambiguation prompt requests the caller 12 to indicate the department of a person being sought, but the caller 12 cannot decide in which department the desired person works, then an indication (such as a time out of the
25 response period) of this failure may be noted. Also, indications may be stored and reported if the disambiguation resulted in the incorrect person being identified. Reports of repeated failures can help detect that improper disambiguation information is provided for a person.

The A/R service 24 can receive, through a secure communication, such as a
30 secure HTTP transfer, or SMTP mail, data representing recorded utterances by callers,

event logs, and other logs, statistics, etc. The recorded utterances and SLEE logs are usable by the A/R service 24 in recognition, tuning and monitoring. The IVR 16 is configured to periodically send the data representing the recorded utterances and SLEE logs to the A/R service 24.

- 5 The A/R service 24 can also monitor the performance of the recognizer 302 contained in the engine system 30. For example, the A/R service 24 can perform off-line recognition testing using a known test sequence.

 The A/R service 24 is also configured to update information in the administration system 32. The A/R system 24 can both add and remove word
10 pronunciations, and add names or words to the vocabularies. Also, the service 24 can modify Backus-Naur Form (BNF) grammars used in the IVR 16. This may help process utterances such as "Mike Phillips, please." The service 24 can also add or update acoustic models, recognizer parameters, and semantic models (e.g., prior probabilities of names). Run time system upgrades and updates can be performed by
15 the service 24. Also, the A/R service 24 is configured to control the amount of waveform and configuration logging through the interface 40. This control includes turning waveform logging on and off and switching from logging a sampling of waveforms, to logging all waveforms, to logging only error waveforms.

 The A/R service 24 is configured to take appropriate support actions based
20 upon various alarms and alerts stemming from the monitoring of the IVR system 16. The A/R service 24 is configured to put bad communication lines into a constant state of busy. The A/R service 24 can also restart portions of the IVR system 16, collect long files for debugging, and insert configuration file patches into damaged configuration files.

25 Referring to FIG. 4, the A/R service 24 is configured to service multiple, distributed IVR systems. As shown, the A/R service 24 can service the IVR system 16 through the SMTP server 18 and firewall 20 via the Internet 22 as well as IVR systems 120, 122 and 124. The systems 120, 122, 124 may be configured differently for individual companies. The A/R service 24 services the IVR systems 120, 122, 124
30 via the Internet 22, and a firewall 126, and an SMTP server 128. Thus, as shown, the

A/R service 24 can service multiple IVR systems 16, 120, 122, 124 via multiple SMTP servers 18, 128 and also multiple IVR systems 120, 122, 124 through a single SMTP server 128. The IVR systems 16, 120, 122, 124 may be geographically distributed remotely from each other.

- 5 Alarms such as emails or SNMP traps, can be sent by the A/R service 24 to an entity such as SWI, or to another vendor of the IVR 16, or other entity, for analysis to determine potential corrective action. The A/R service 24 is configured to send the alarms when performance deviates more than a predetermined amount from an expected behavior (e.g., an expected value such as a frequency or quantity) of the
- 10 monitored statistic. Included in the A/R service 24 is a configuring entity including transcribers for transcribing stored caller utterances in accordance with the alarms and alerts or other notifications by the A/R service 24. People at the configuring entity are provided for reviewing the transcribed utterances, comparing these utterances with the vocabularies, or otherwise analyzing the transcribed utterances to determine
- 15 appropriate corrective action, if any. Such action includes using the RCS 36 to adapt/reconfigure the IVR 16, e.g., to reduce OOVs, to update pronunciations or other information, and/or to correct information stored in the IVR 16.

CONFIGURATION AND SETUP

- 20 How the system 10 is configured and set up depends on the type of system chosen by the customer. The customer can choose a base platform and configuration tools or a collection of configurable models. If the customer chooses the base platform and tools, then the customer can configure and customize the product. The customer can provide configuration/customization data to the vendor and/or
- 25 configuration entity, e.g., SpeechWorks® International, Inc. for the configuration entity to configure the system 10.

- If the customer chooses the base platform and tools, then the customer inputs the data for the desired functionality and any customization parameters. The customer would need to input, either through a database download, or individual entries,
- 30 relevant information for an auto attendant, such as personnel names, nicknames,

departments, and extensions, as well as any appropriate synonyms such as job titles/positions. Additionally, the customer would provide information for the content of the speech pages and instructions for any links to other pages, and instructions for transactions to be supported by the speech pages. Much of the content and

5 functionality, including for transactions, is provided in the base platform, but the customer would need to supply customized data and instructions. The customer can select configuration parameters to customize the performance of the system, for example whether disambiguation is possible for an auto attendant. As another example, for an event registration tool the customer would record the date, event title,

10 and prompts for information needed from a caller to register for the upcoming event. The customer could, in a similar manner, modify/update the initial configuration/setup as necessary to accommodate for changing information, such as recent events, additional or fewer personnel, changes of name, delays or other alterations in events, etc.

15 If the customer selects the collection of configurable models for configuration by the vendor or other entity, then the customer would provide the relevant information to the configuring entity, e.g., SpeechWorks® International, Inc. The customer would provide content information for the speech pages, relevant personnel directory information for an auto attendant as discussed above, and desired options for

20 configuration parameters. The configuring entity uses this information, and its expertise to configure the system for the customer. Additionally, the configuring entity updates the configuration/setup as required by the customer after the initial configuration/setup.

No matter whether the customer or other entity performs the configuration, the

25 configuration files are written to and/or amended by the administration system 32 (FIG. 1) and read by the engine system 30 (FIG. 1) for execution.

OPERATION

In operation, the IVR system 16 interacts with the user/caller 12 in accordance

30 with the user interface that guides the caller 12 through a web-model speech

recognition process that includes performing operations indicated by the caller 12. In accordance with the web model, the caller 12 is typically first presented with the home speech page (unless the caller 12 accesses a different speech page directly) that provides the caller 12 with a variety of options. The caller 12 can select from the options presented by saying any of the specified words/phrases, or in a natural language manner saying what information and/or service the caller 12 desires. Terminology analogous to typical websites is used to help the caller 12 navigate through the various speech pages by the caller 12 saying appropriate utterances. For each utterance by the caller 12, disambiguation and/or retrials of recognition can be preformed, where the system is configured to do so. At each stage of the conversation between the IVR system 16 and the caller 12, the caller 12 is informed of what page is being loaded (e.g. "loading Contact Us page"), and when the speech page has been loaded, what is the page title for the information about to be presented to the caller 12 (e.g., "Contact Us page. To call us toll free...."). The A/R service 24 analyzes and monitors information regarding conversations between the IVR system 16 and the caller 12 and provides appropriate reports, alerts, and/or alarms for use by the customer and/or a configuring entity (e.g., SpeechWorks® International, Inc.) to determine whether updates to the system 10 are warranted.

Referring to FIGS. 1, 2, and 5, an interactive conversation process 200 begins at stage 202 when the caller 12 dials a telephone number associated with the IVR system 16. The caller 12 is connected from the caller's phone 26 through the PSTN 14 to the IVR system 16. Connection is established between the IVR system 16 and the caller 12 through the PSTN 14 for bi-directional communication between the caller 12 and the IVR system 16.

At stage 204 the IVR system 16 plays prompts to the user 12 indicating that the user 12 has reached the home page of the SpeechSite™ IVR system 16. For example, a prompt such as "home page" or "you have reached the [company x] SpeechSite™ voice recognition system home page" is played to the caller 12. Alternatively, if the user 12 dialed a number associated with a specific speech page other than the home page, then the information for the dialed page can be

prompted/played to the caller 12. The information prompted to the caller 12 includes various options as to what other pages the user 12 can access and/or general information contained on the home page. The prompts can inform the caller 12 about speech modules of the SpeechSite™ IVR system. In this example, the prompts

5 include "You can link to a personnel directory by saying 'Company Directory'; you can find out how to contact us by saying 'Contact Us'; you can learn about the company by saying 'Company Information'; to perform [transaction x] say [perform transaction x]." The transaction can be, e.g., to buy stock or other goods. Thus, [perform transaction x] and [transaction x] can both be "buy stock" in this example.

10 Thus, the prompts give the caller 12 instructions how to initiate call routing through a Company Directory (Auto Attendant), information retrieval (Contact Us and Company Information), and transaction processing, respectively. The information also includes instructions as to how to navigate through the various speech pages prompted to the caller 12 including that the caller 12 can navigate the SpeechSite™

15 IVR system by speaking terms associated with website analogous functions, such as "back," "forward," and "home" as well as other functions such as "find it," "fax it," and "where am I?"

At stage 206, the caller speaks into the phone 26 to provide speech 208 to navigate through the speech pages. The speech 208 may be in response to prompts

20 played by the IVR system 16, such as requesting a specific speech page, or can be a natural language request for information or other actions. The speech 208 represents speech-related information, but not necessarily an analog or a digitized speech utterance. For example, the speech 208 may represent the set of N-best word strings generated as output by the recognizer 302.

25 At stage 210, the engine system 30 discriminates between available sub-processes to determine which sub-process is the appropriate one for processing the request indicated by the speech 208. To discriminate between sub-processes (each of which may contain one or more speech modules) the engine system 30 compares the speech 208 with sub-process titles prompted to the caller 12, and/or multiple

30 vocabularies, each of which is associated with at least one of the available sub-

processes. In the latter case the vocabularies contain synonyms of the titles presented to the caller 12. For example, if the caller 12 says "Directions to Boston" the process 200 will proceed to stage 214 for call information retrieval from the Contact Us page.

- If the speech 208 is matched with a sub-process title (e.g., if the speech 208 is
- 5 "Company Information"), then the engine system 30 directs the appropriate corresponding sub-process to process the speech 208 at stage 212 for call routing, at 214 for information retrieval, and/or at stage 216 for transaction processing.

- The various sub-processes 212, 214, and 216 process the speech 208 as described in more detail below. Appropriate prompts are played to the caller 12
- 10 indicating to which sub-process the caller 12 is being directed. For example, the engine system 30 plays a prompt "Transferring to the Company Directory Page" (or "Transferring to the Personnel Directory Page" or "transferring to the Call Routing Page") if the caller 12 is being routed to the call routing sub-process 212. Prompts
- 15 "Transferring to the [information retrieval] page" and "Transferring to the [transaction processing] page" are played to the caller 12 if the caller 12 is being transferred to these sub-processes 214 or 216, respectively. "Contact Us" or "Company Information" replace [information retrieval] and "Stock Buying" replaces [transaction processing] in this example. Alternatively or in addition, prompts are played to the caller 12 indicating that the appropriate page is being loaded, e.g., "Loading the
- 20 Company Directory Page."

The sub-processes 212, 214, and 216 interact with the caller 12 to determine specific responses or actions, such as providing information or performing actions appropriate for the speech 208 or further speech from the caller 12.

- At stage 218, the engine system 30 provides the appropriate response or
- 25 performs the appropriate action as determined by the sub-processes 212, 214, and 216.

- Referring to FIGS. 1, 3, 5, and 6, at stage 220 of the call routing process 212, the caller 12 is presented with the call routing page. The engine system 30 plays prompts for the caller 12 to indicate the information associated with the call routing
- 30 page and links to other pages from the call routing page 220. The engine system 30

plays a personnel directory prompt to have the caller 12 to speak the name or department of a person to whom the caller 12 wishes to speak. The IVR system 16 receives the caller's speech 208 in response to the prompt.

At stage 222, the engine system 30 obtains a call-routing vocabulary from the configuration and log system 34. This information can be obtained before, after, or during stage 220. The call routing vocabulary in this example includes data related to a personnel directory. Other information, however, is possible for other examples such as information related to an airline flight scheduling system.

At stage 224, the engine system 30 determines N word strings that may possibly correspond to the intended words/phrases from the caller 12. These N-best word strings (the N-best list) are compared by the engine system 30 to the call routing vocabulary obtained at stage 222. For example, if the confidence score of the highest-confidence word string in the N-best list exceeds an upper threshold, then the word is considered to be recognized and accepted. Low-confidence word strings with confidence scores below a lower threshold are rejected. Word strings with confidence scores between the lower and upper thresholds are queued for disambiguation, as may also be the case if confidence scores for multiple word strings exceed the upper threshold.

To help uniquely identify the word strings spoken by the caller 12, at stage 225, disambiguation is performed by the engine system 30 if needed. For example, if there are two employees with the name spoken by the caller 12, the engine system 30 attempts to distinguish between them by having the caller 12 identify the department of the desired employee. The engine system 30 thus prompts the caller 12 to "say the name of the department of the person that you are trying to contact." Depending on the caller's response to the disambiguation prompt, the engine system 30 chooses one of the N-best word strings as the word string spoken by the caller 12.

At stage 226, the engine system 30 determines appropriate responsive action according to the comparing of the speech with the call-routing vocabulary at stage 224. This determination can result in the call being routed to an identified person, or performing a requested action.

At stage 228 the caller's call is routed to a person identified by the caller's speech. The engine system 30 uses the call-routing information, such as an extension, associated with the person identified at stage 226 as being the person to whom the caller 12 wishes to speak to connect the caller 12 and the desired person. For example, the caller 12 is routed to John Doe's extension if the speech was "John Doe," or to another speech page or an operator, e.g., if the speech was "flight schedule", such that the caller 12 is routed to an operator for scheduling flights.

At stage 230, the engine system 30 carries out an action other than call routing, as indicated by the speech such as playing or faxing information.

Referring to FIGS. 1, 3, 5, and 7, an information retrieval process 214 is shown in FIG. 7. For the following description an example of obtaining information from a Company Information page is described. This is not intended to be a limiting example; other possibilities for information to be retrieved, including other pages to retrieve information from, are within the scope of the invention. At stage 232, an information retrieval page is presented to the caller 12. The engine system 30 plays a prompt "Loading Company Information page" and when this page has been loaded plays an additional prompt "Company Information page." Following these prompts, the engine system 30 plays prompts indicating information on the Company Information page such as links to other speech pages, and general company information. This general information can include the general nature of the company, including the technology of the company, and the company's products and/or services.

At stage 234 the engine system 30 obtains an information retrieval vocabulary for use in recognizing the caller's speech. The engine system 30 obtains this information from the configuration and log system 34. This information is based on the information contained on the Company Information page and those pages identified as links from the Company Information page. The engine system 30 plays a prompt such as "You can be linked to pages with the following information by saying the names of the pages: 'Company History', 'News and Press Releases', or 'Current Events'."

At stage 236, the engine system 30 matches the N-best word strings for the caller's spoken response to the prompts to the information retrieval vocabulary. The engine system 30 develops several word strings that may represent what the caller 12 spoke. The N-best of these word strings are selected by the engine system 30 for
5 comparison with the information retrieval vocabulary to determine which of the word strings the caller 12 spoke.

To help uniquely identify the word strings spoken by the caller 12, at stage 238, disambiguation can be performed by the engine system 30. The engine system 30 can play appropriate prompts for the caller 12 such as "I think you said 'company
10 history' and not 'current events'. Is that correct?" Depending on the caller's response, the engine system 30 chooses one of the N-best word strings as the word string spoken by the caller 12.

At stage 240, the engine system 30 retrieves the resource requested by the caller 12. In response to the uniquely-identified word string determined at stage 236,
15 and possibly 238, the engine system 30 uses the information from the identified word string to access the configuration and log system 34 to retrieve the information associated with the caller's request. For example, if the caller 12 responded to the disambiguation question above with a "yes" answer, then the engine system 30 would retrieve information relating to the company history, such as a Company History
20 speech page stored in the configuration and log system 34. The speech engine 30 plays a prompt "Loading Company History page."

At stage 242, the engine system 30 delivers the requested resource to the caller 12. In this example, the engine system 30 prompts the caller 12 with the associated information of the Company History page. For example, the prompt could be
25 "Company History page. You may link to the following speech pages...."

Referring to FIGS. 1, 3, 5, and 8, a transaction processing process 216 is shown in FIG. 8. For the following description an example of booking an airline flight is used. This is not intended to be a limiting example, other possibilities for transactions to be processed including purchasing products or commodities are within
30 the scope of the invention. At stage 244, a Flight Reservation page is presented to the

caller 12. The engine system 30 plays a prompt "Loading Flight Reservation page" and when this page has been loaded plays an additional prompt "Flight Reservation page." Following these prompts, the engine system 30 plays prompts indicating information on the Flight Reservation page such as links to other speech pages, and
5 general flight reservation information contained. This general information can include information about pricing, inflight services, and/or travel procedures such as check-in time and luggage limits. At stage 246 the engine system 30 obtains a flight reservation vocabulary for use in recognizing the caller's speech. The engine system 30 obtains this information from the configuration and log system 34. This
10 information is based on the information contained on the company Flight Reservation page and those pages identified as links from the company information page. The engine system 30 plays a prompt such as "You can be linked to pages with the following information by saying the names of the pages: Domestic Flights, on International Flights."

15 At stage 248, the engine system 30 matches the N-best word strings to the flight reservation vocabulary. The engine system 30 develops several word strings that may represent what the caller 12 spoke. The N-best of these word strings are selected by the engine system 30 for comparison with the flight reservation vocabulary to determine which of the word strings the caller 12 spoke.

20 To help uniquely identify the word strings spoken by the caller 12, at stage 250, disambiguation can be performed by the engine system 30. The engine system 30 can play appropriate prompts for the caller 12 such as "If you said 'Northwest,' say 'One,' If you said 'Southwest,' say 'Two,' Otherwise say 'Neither.'" Depending on the caller's response, the engine system 30 chooses one of the N-best word strings as
25 the word string spoken by the caller 12.

At stage 252, the engine system 30 produces one or more transaction requests in response to the identified request by the caller 12. In response to the uniquely-identified word string determined at stage 248, and possibly 250, the engine system 30 uses the information from the identified word string to access the configuration and
30 log system 34 to retrieve the information for the transaction associated with the

caller's request. The transaction request will initiate the requested transaction or instruct appropriate hardware and/or software to perform the requested transaction. The transaction requests can be retrieved from storage in the configuration and log system 34 and/or custom produced by inserting values for variables in the information
5 retrieved from the configuration and log system 34 or by completely producing a custom-built request. For example, if the caller 12 responded to the disambiguation question above by saying "One", then the engine system 30 would produce a transaction request relating to Northwest Airlines, such as "Book a roundtrip flight from Washington, D.C., to Detroit, leaving March 1 at 8 a.m. and returning on March
10 2 with a 10 p.m. departure time."

At stage 254, the engine system 30 delivers the transaction request to the appropriate portion of the engine system 30, or to another appropriate location, such as a website server for Northwest Airlines. In this example, the engine system 30 executes the transaction according to the transaction request. Alternatively, this
15 execution can include such actions as transmitting an order for stock, or transmitting a request over the fax server 97 to fax information to the caller's fax machine 28, depending on the caller's request.

At stage 258, the engine system 30 produces a response to the executed transaction. Here, the response indicates the success or failure of booking or
20 purchasing tickets for the requested flights, and if successful, pertinent information such as flight numbers, times, seats, and prices etc. Alternatively, the response could indicate the sale or purchase price of a stock, or the success or failure of an attempt to fax information to the caller 12, if these transactions were requested.

At stage 260, the engine system 30 prompts the caller 12 regarding the
25 response. For example, the engine system 30 can play prompts such as "You have been booked on flight 123 leaving Washington, D.C., at 8:12 a.m. on March 1, arriving Detroit at 10:48 a.m.; and on flight 456 leaving Detroit at 9:47 p.m., arriving in Washington, D.C., at 12:13 a.m. on March 3," or, e.g., "The requested information has been successfully faxed to (617) 555-1212."

The caller 12 is returned to the transaction processing page 244 so that the caller 12 can initiate another transaction if desired.

Referring to FIGS. 1, 3, 5, and 9, a process 270 for reporting and analyzing interactive conversations is shown in FIG. 9. At stage 272 the caller 12 and has an
5 interactive conversation with other parts of the system 10. Data from this conversation, including utterances by the caller 12 and/or actions taken by the system 10 are stored/logged. The storing/logging can occur during or after the conversation.

At stage 274, stored data from the interactive conversation are monitored and/or reported. The reporting can be in the form of alarms or alerts, or in formal
10 reports organizing the data for analysis. Alarms can highlight potential causes of errors in the system 10, or at least areas for improvement of the system 10. Reports can show the performance of the system 10. The performance characteristics reported are organized to help make the analysis, especially of correctable features of the system 10, easy. Performance characteristics are also organized to facilitate
15 performance reporting to the IVR customer, to indicate how well the customer's purchase is operating.

At stage 276, the monitored/reported data are analyzed. People at the configuring entity, or other analysis entity, review the reports and/or alarms regarding performance characteristics/statistics of interest. For example, the people can analyze
20 the characteristics to determine whether too many calls are failing to be routed to employees, or that too many calls are being routed to an operator or disconnected. The people can also compare transcribed utterances that failed to result in the caller 12 being connected to an employee with recognition vocabularies to determine OOV utterances. A wide range of analysis can be performed at stage 276, of which the
25 above analyses are examples.

From the analyses at stage 276, the people can determine what, if any, corrective action can and/or should be taken. For example, it can be determined that an alternative pronunciation should be added to a recognition vocabulary, or that a person's name or transaction's title were mistakenly not added to appropriate
30 recognition vocabularies, to reduce OOVs. Also, it could be determined that a

disambiguation feature should be added to one or more portions of the interactive conversation process 200, e.g., to reduce the frequency of misdirected calls. The corrective action can be to use the RCS 36 to add, delete, or alter information, prompts, links, configuration parameters, etc. of the IVR 16 to help improve operation
5 of the system 10. The corrective action determined in stage 276 is taken in stage 278.

Other embodiments are within the scope and spirit of the claims. For example, the A/R service 24, or one or more portions thereof, may be provided at the location of, or within, the IVR system 16. Also, portions of the system 10 may have different configurations than as described above. For example, environments other than
10 Artisoft® 5.0 Visual Voice Enterprise may be used.

Also, different processes of analyzing performance data are possible. For example, frequently occurring OOVs of the same utterance can be analyzed while ignoring less common OOV utterances. OOV utterances with similar features can be grouped such that a person only listens to enough utterances from the group to
15 identify the OOV utterance. This can be accomplished by collecting utterance waveforms (in the form of ulaws) from all recognition failures, or low-confidence recognitions. Each ulaw is converted into a sequence of feature vectors (e.g., Mel-Frequency Cersmal Coefficients (MFCCs)) using a standard recognizer front end. An MFCC vector is produced for each frame (e.g., 1ms) of speech. Similar utterances are
20 clustered together using dynamic alignment of the feature vectors, or clustering techniques such as k-means. Each cluster represents a collection of example utterances of an OOV, plus some noise. A human transcriber listens to a few of the utterances from a cluster to determine the primary OOV from the cluster. The clustering helps the transcriber avoid listening to all the utterances to identify an
25 OOV.

Still further, automatic techniques for transcribing utterances may be used. Instead of being transcribed by humans, utterances may be transcribed, e.g., by a phone-loop recognizer to produce phonetic representations. A few utterances from each cluster of utterances can be transcribed in this manner. The phonetic
30 representations can be cross-referenced into a phonetic dictionary, or passed to a

human to verify the OOV utterance. The OOV utterances can be flagged for consideration for corrective action. Alternatively, utterances or may be compared against a large dictionary (e.g., of names).

What is claimed is:

CLAIMS



1. An interactive speech system comprising:
a port configured to receive a call from a user and to provide a communication link between the system and the user;
5 memory having personnel directory information stored therein including indicia of a plurality of people and routing information associated with each person for use in routing the call to a selected one of the plurality of people, the memory also having company information stored therein associated with a company associated with the interactive speech system; and
10 a speech element coupled to the port and the memory and configured to convey first audio information to the port to prompt the user to speak to the system, the speech element also being configured to receive speech from the user through the port, to recognize the speech from the user, and to perform an action based on recognized user's speech, the speech element being further configured to convey
15 second audio information to the port in accordance with the company information stored in the memory.
2. The system of claim 1 wherein the speech element is configured to convey speech in at least a partially web-analogous format.
- 20 3. The system of claim 2 wherein the speech element is configured to, in response to a request by the user recognized by the speech element, provide information, stored in the memory, according to the request, and to route the call to a person indicated by the user's request according to the routing information associated
25 with the person.
4. The system of claim 3 wherein portions of the company information stored in the memory are associated with each other in pages of information according to a plurality of categories of information including how to contact the company.



5. The system of claim 4 wherein the speech element is configured to act upon the user's speech if the user's speech is within a vocabulary based upon information of a page most recently accessed by the speech element.

5 6. The system of claim 4 wherein the categories of information further include information about the location of the company, and products, if any, and services, if any, offered by the company.

7. The system of claim 4 wherein the company information stored in the
10 memory includes information available on a website of the company.

8. The system of claim 7 wherein the memory and the speech element are configured to convey the company information to the user with an organization different than an organization of the company information provided on the company's
15 website.

9. The system of claim 4 wherein the speech element is configured to access pages of information in response to spoken commands from the user associated with functions commonly provided by web browsers.

20

10. The system of claim 9 wherein the commands include "back," "forward," and "home."

11. The system of claim 1 wherein the speech element is configured to
25 perform transactions indicated by the user's speech.

12. The system of claim 1 further comprising a speech application monitor configured to monitor activity of the speech element and corresponding incoming speech from the user.

30



13. The system of claim 12 wherein the speech element is configured to store conversation data in the memory indicative of at least one of: the user's speech; if the user's speech was accepted as recognized, what action if any the speech element took; and if the user's speech has a confidence below a predetermined threshold; and
5 wherein the speech application monitor is configured to report indicia of the conversation data stored by the speech element.

14. The system of claim 12 wherein the speech application monitor is coupled to the memory through the Internet.
10

15. The system of claim 1 wherein the speech element is configured to perform at least one of disambiguating the user's speech and confirming the user's speech.

15 16. The system of claim 1 further comprising a control unit coupled to the memory and configured to receive a control signal from outside the system and to modify information content of the memory in response to the control signal.

17. The system of claim 16 wherein the control unit is configured to add
20 information to the memory, to delete information from the memory, and to alter information of the memory.

18. The system of claim 1 wherein the speech element is further configured to convey information to the user to prompt the user to provide
25 disambiguating information regarding a person and to use the disambiguating information to disambiguate between which of multiple people the user desires to contact.

19. A computer program product comprising computer-readable
30 instructions for causing a computer to:



establish a communication link with a user in response to receiving a call from the user;

retrieve information from a memory having personnel directory information stored therein including indicia of a plurality of people and routing information

5 associated with each person for use in routing the call to a selected one of the plurality of people, the memory also having company information stored therein associated with a company associated with the interactive speech system;

convey first audio information to the user to prompt the user to speak;

receive speech from the user;

10 recognize the speech from the user;

perform an action based on recognized user's speech; and

convey second audio information to the user in accordance with the company information stored in the memory.

15 20. The computer program product of claim 19 wherein the instructions for causing the computer to convey the second audio information cause the computer to convey the second audio information in at least a partially web-analogous format.

20 21. The computer program product of claim 20 wherein the instructions for causing the computer to convey the second audio information cause the computer, in response to a request by the user recognized by the computer, to provide information, stored in the memory, according to the request, the computer program product further comprising instructions for causing the computer to route the call to a person indicated by the request according to the routing information associated with the
25 person.

22. The computer program product of claim 21 wherein the memory stores information in pages according to a plurality of predetermined categories of information, and wherein the instructions for causing the computer to recognize the
30 user's speech cause the computer to use a vocabulary associated with a current page of

speech to recognize the user's speech.



23. The computer program product of claim 22 wherein the company information stored in the memory includes information available on a website of the company, and wherein the instructions for causing the computer to convey the second audio information to the user cause the computer to convey the second audio information with an organization different than an organization of the company information provided on the company's website.

24. The computer program product of claim 22 wherein the instructions for causing the computer to retrieve information cause the computer to retrieve information in response to spoken commands from the user associated with functions commonly provided by web browsers.

25. The computer program product of claim 24 wherein the commands include "back," "forward," and "home."

26. The computer program product of claim 19 further comprising instructions for causing the computer to perform transactions indicated by the user's speech.

27. The computer program product of claim 19 further comprising instructions for causing the computer to:

store conversation data in the memory indicative of at least one of: the user's speech; if the user's speech was accepted as recognized, what action if any the computer took; and if the user's speech has a confidence below a predetermined threshold; and

report indicia of the stored conversation data.

28. The computer program product of claim 19 further comprising



instructions for causing the computer to perform an action based on an attempt to recognize the user's speech.

29. The computer program product of claim 19 further comprising
5 instructions for causing the computer to receive a control signal and to modify information content of the memory in response to the control signal.

30. The computer program product of claim 29 wherein the instructions for causing the computer to modify information content of the memory include
10 instructions for causing the computer to add information to the memory, to delete information from the memory, and to alter information of the memory.

31. The computer program product of claim 19 further comprising instructions for causing the computer to: convey information to the user to prompt the
15 user to provide disambiguating information regarding a person; and use the disambiguating information to disambiguate between which of multiple people the user desires to contact.

32. A method of interfacing with a user through an interactive speech
20 application, the method comprising:
receiving an incoming call from the user;
establishing a communication link with the user;
retrieving a portion of stored data indicative of speech for presentation to the user; and
25 presenting the portion of stored data as speech to the user in a web-analogous form.

33. The method of claim 32 wherein the stored data are stored in groups according to associated titles indicative of the content of the data in each
30 corresponding group, and wherein the presenting includes conveying the title of the

portion of stored data to the user as speech.



34. The method of claim 33 further comprising:
receiving speech from the user;
5 converting the user's speech into electrical indicia of the user's speech;
retrieving another portion of stored data in accordance with the electrical
indicia; and
presenting the another portion of stored data to the user including conveying a
title of the another portion of stored data to the user as speech.

10

35. The method of claim 34 wherein the user's speech is the title of the
another portion of stored data.

36. The method of claim 34 wherein the indicia of the user's speech are
15 indicative of the title of the another portion of stored data.

37. The method of claim 36 wherein the indicia of the speech are
indicative of a synonym of the title of the another portion of stored data.

20 38. The method of claim 34 wherein the user's speech includes a web-
analogous navigation command.

39. The method of claim 38 wherein the web-analogous navigation
command is selected from the group consisting of: "back," "forward," "home," "go
25 to," and "help."

40. The method of claim 32 wherein the stored data are grouped according
to content of the data, and wherein the presenting includes conveying a speech
indication to the user of the data content of the portion of stored data, the indication
30 including the word "page."



41. A monitoring system for monitoring at least one speech application system, the monitoring system comprising:

a computer network connection; and

5 a monitoring unit coupled to the speech application system and to the computer network connection and configured to receive data from the at least one speech application system through the computer network connection, to process call records of indicia related to calls associated with the speech application system, and to produce reports indicative of the indicia related to the calls.

10

42. The system of claim 41 wherein the monitoring unit is coupled to the speech application system through the computer network connection and wherein the monitoring unit is remotely located from the at least one speech application system.

15

43. The system of claim 42 wherein the computer network connection is coupled to the at least one speech application system through the Internet.

44. The system of claim 43 wherein the monitoring unit is configured to access logs of call records stored in the at least one speech application system.

20

45. The system of claim 43 wherein the monitoring unit is coupled through the computer network connection and the Internet to a plurality of distributed speech application systems and is configured to receive data from each of the speech application systems through the network connection, to process records of call events
25 associated with each of the speech application systems, and to produce reports indicative of the indicia related to the calls for each speech application system.

46. The system of claim 41 wherein the monitoring unit is configured to transmit signals to the at least one speech application system to alter operation of the
30 at least one speech application system.

47. The system of claim 46 wherein the signals are adapted to cause malfunctioning communication lines of the at least one speech application system to be effectively rendered busy.

5

48. The system of claim 46 wherein the signals are adapted to cause services of the at least one speech application system to be restarted.

49. The system of claim 46 wherein the signals are adapted to cause configuration file patches to be inserted into configuration files in the at least one speech application system.

10

50. The system of claim 41 wherein the monitoring unit is configured to produce an indication of a frequency of a selected call event.

15

51. The system of claim 41 wherein the monitoring unit is configured to produce an alert regarding a selected call event.

52. The system of claim 51 wherein the alert is an indication that a characteristic of a selected call event deviates more than a predetermined amount from a predetermined reference value for that characteristic.

20

53. The system of claim 41 wherein the monitoring unit and the speech application system are disposed proximate to each other.

