(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 0 911 808 A1

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
28.04.1999 Bulletin 1999/17

(51) Int. Cl.[6]: **G10L 5/06**, H04L 12/28

(21) Application number: **97118470.0**

(22) Date of filing: **23.10.1997**

(84) Designated Contracting States:
AT BE CH DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE
Designated Extension States:
AL LT LV RO SI

(71) Applicant:
Sony International (Europe) GmbH
50829 Köln (DE)

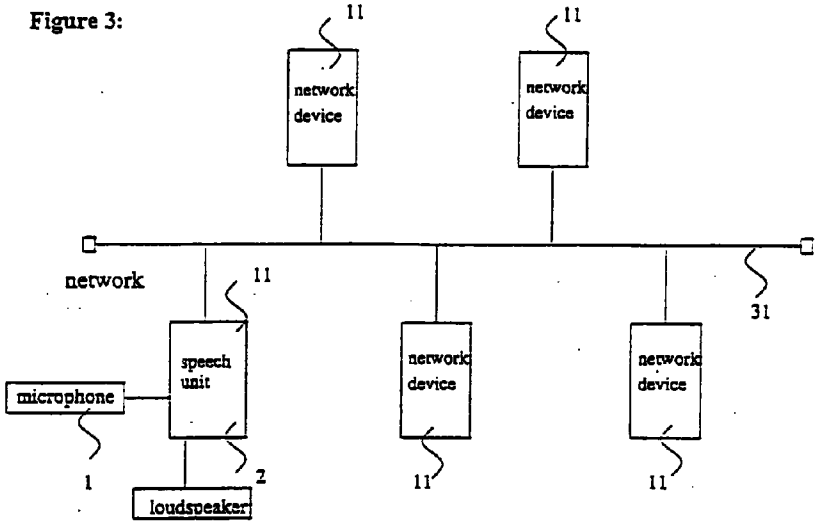
(72) Inventors:
- **Buchner, Peter,**
  c/o Sony International GmbH
  70736 Fellbach (DE)
- **Goronzy, Silke,**
  c/o Sony International GmbH
  70736 Fellbach (DE)
- **Kompe, Ralf,**
  c/o Sony International GmbH
  70736 Fellbach (DE)
- **Rapp, Stefan,**
  c/o Sony International GmbH
  70736 Fellbach (DE)

(74) Representative:
Müller, Frithjof E., Dipl.-Ing.
Patentanwälte
MÜLLER & HOFFMANN,
Innere Wiener Strasse 17
81667 München (DE)

(54) Speech interface in a home network environment

(57) Home networks low-cost digital interfaces are introduced that integrate entertainment, communication and computing electronics into consumer multimedia. Normally, these are low-cost, easy to use systems, since they allow the user to remove or add any kind of network devices with the bus being active. To improve the user interface a speech unit (2) is proposed that enables all devices (11) connected to the bus system (31) to be controlled by a single speech recognition device. The properties of this device, e.g. the vocabulary can be dynamically and actively extended by the consumer devices (11) connected to the bus system (31). The proposed technology is independent from a specific bus standard, e.g. the IEEE 1394 standard, and is well-suited for all kinds of wired or wireless home networks.

Figure 3:



EP 0 911 808 A1

DOCKET ALARM
Find authenticated court documents without watermarks at docketalarm.com.

Description

[0001] This invention relates to a speech interface in a home network environment. In particular, it is concerned with a speech recognition device, a remotely controllable device and a method of self-initialization of a speech recognition device.

[0002] Generally, speech recognizers are known for controlling different consumer devices, i.e. television, radio, car navigation, mobile telephone, camcorder, PC, printer, heating of buildings or rooms. Each of these speech recognizers is built into a specific device to control it. The properties of such a recognizer, such as the vocabulary, the grammar and the corresponding commands, are designed for this particular task.

[0003] On the other hand, technology is now available to connect different of the above listed consumer devices via a home network with dedicated bus systems, e.g. a IEEE 1394 bus. Devices adapted for such systems communicate by sending commands and data to each other. Usually such devices identify themselves when they are connected to the network and get a unique address assigned by a network controller. Thereafter, these addresses can be used by all devices to communicate with each other. All other devices already connected to such a network are informed about address and type of a newly connected device. Such a network will be included in private homes as well as cars.

[0004] Speech recognition devices enhance comfort and, if used in a car may improve security, as the operation of consumer devices becomes more and more complicated, e.g. controlling of a car stereo. Also in a home network environment e.g. the programming of a video recorder or the selection of television channels can be simplified when using a speech recognizer. On the other hand, speech recognition devices have a rather complicated structure and need a quite expensive technology when a reliable and flexible operation should be secured, therefore, a speech recognizer will not be affordable for most of the devices listed above.

[0005] Therefore, it is the object of the present invention to provide a generic speech recognizer facilitating the control of several devices. Further, it is the object of the present invention to provide a remotely controllable device that simplifies its network-controllability via speech.

[0006] A further object is to provide a method of self-initialization of the task dependent parts of such a speech recognition device to control such remotely controllable devices.

[0007] These objects are respectively achieved as defined in the independent claims 1, 4, 14, 15 and 18.

[0008] Further preferred embodiments of the invention are defined in the respective subclaims.

[0009] The present invention will become apparent and its numerous modifications and advantages will be better understood from the following detailed description of an embodiment of the invention taken in conjunction with the accompanying drawings, wherein

Fig. 1 shows a block diagram of an example of a speech unit according to an embodiment of the invention;

Fig. 2 shows a block diagram of an example of a network device according to an embodiment of the invention;

Fig. 3 shows an example of a wired 1394 network having a speech unit and several 1394 devices;

Fig. 4 shows an example of a wired 1394 network having a speech unit incorporated in a 1394 device and several normal 1394 devices;

Fig. 5 shows three examples of different types of networks;

Fig. 6 shows an example of a home network in a house having three clusters;

Fig. 7 shows two examples of controlling a network device remotely via a speech recognizer;

Fig. 8 shows an example of a part of a grammar for a user dialogue during a VCR programming;

Fig. 9 shows an example of a protocol of the interaction between a user, a speech recognizer and a network device;

Fig. 10 shows an example of a learning procedure of a connected device, where the name of the device is determined automatically;

Fig. 11 shows an example of a protocol of a notification procedure of a device being newly connected, where the user is asked for the name of the device;

Fig. 12 shows an example of a protocol of the interaction of multiple devices for vocabulary extensions concerning media contents; and

Fig. 13 shows another example of a protocol of the interaction of multiple devices for vocabulary extensions concerning media contents.

[0010] Fig. 1 shows a block diagram of an example of the structure of a speech unit 2 according to the invention. Said speech unit 2 is connected to a microphone 1 and a loudspeaker, which could also be built into said

DOCKET ALARM
Find authenticated court documents without watermarks at docketalarm.com.

speech unit 2. The speech unit 2 comprises a speech synthesizer, a dialogue module, a speech recognizer and a speech interpreter and is connected to an IEEE 1394 bus system 10. It is also possible that the microphone 1 and/or the loudspeaker are connected to the speech unit 2 via said bus system 10. Of course it is then necessary that the microphone 1 and/or the loudspeaker are respectively equipped with a circuitry to communicate with said speech unit 2 via said network, such as A/D and D/A converters and/or command interpreters, so that the microphone 1 can transmit the electric signals corresponding to received spoken utterances to the speech unit 2 and the loudspeaker can output received electric signals from the speech unit 2 as sound.

[0011] IEEE 1394 is an international standard, low-cost digital interface that will integrate entertainment, communication and computing electronics into consumer multimedia. It is a low-cost easy-to-use bus system, since it allows the user to remove or add any kind of 1394 devices with the bus being active. Although the present invention is described in connection with such an IEEE 1394 bus system and IEEE 1394 network devices, the proposed technology is independent from the specific IEEE 1394 standard and is well-suited for all kinds of wired or wireless home networks or other networks.

[0012] As will be shown in detail later, a speech unit 2, as shown in Fig. 1 is connected to the home network 10. This is a general purpose speech recognizer and synthesizer having a generic vocabulary. The same speech unit 2 is used for controlling all of the devices 11 connected to the network 10. The speech unit 2 picks up a spoken-command from a user via the microphone 1, recognizes it and converts it into a corresponding home network control code, henceforth called user-network-command, e.g. specified by the IEEE 1394 standard. This control code is then sent to the appropriate device that performs the action associated with the user-network-command.

[0013] To be capable of enabling all connected network devices to be controlled by speech, the speech unit has to "know" the commands that are needed to provide operability of all individual devices 11. Initially, the speech unit "knows" a basic set of commands, e.g., commands that are the same for various devices. There can be a many-to-one mapping between spoken-commands from a user and user-network-commands generated therefrom. Such spoken-commands can e.g. be *play, search for radio station YXZ* or (sequences of) numbers such as phone numbers. These commands can be spoken in isolation or they can be explicitely or implicitely embedded within full sentences. Full sentences will henceforth as well be called spoken-command.

[0014] In general, speech recognizers and technologies for speech recognition, interpretation, and dialogues are well-known and will not be explained in detail in connection with this invention. Basically, a speech recognizer comprises a set of vocabulary and a set of knowledge-bases (henceforth grammars) according to which a spoken-command from a user is converted into a user-network-command that can be carried out by a device. The speech recognizer also may use a set of alternative pronunciations associated with each vocabulary word. The dialogue with the user will be conducted according to some dialogue model.

[0015] The speech unit 2 according to an embodiment of the invention comprises a digital signal processor 3 connected to the microphone 1. The digital signal processor 3 receives the electric signals corresponding to the spoken-command from the microphone 1 and performs a first processing to convert these electric signals into digital words recognizable by a central processing unit 4. To be able to perform this first processing, the digital signal processor 3 is bidirectionally coupled to a memory 8 holding information about the process to be carried out by the digital signal processor 3 and a speech recognition section 3a included therein. Further, the digital signal processor 3 is connected to a feature extraction section 7e of a memory 7 wherein information is stored of how to convert electric signals corresponding to spoken-commands into digital words corresponding thereto. In other words, the digital signal processor 3 converts the spoken-command from a user input via the microphone 1 into a computer recognizable form, e.g. text code.

[0016] The digital signal processor 3 sends the generated digital words to the central processing unit 4. The central processing unit 4 converts these digital words into user-network-commands sent to the home network system 10. Therefore, the digital signal processor 3 and the central processing unit 4 can be seen as speech recognizer, dialogue module and speech interpreter.

[0017] It is also possible that the digital signal processor 3 only performs a spectrum analysis of the spoken-command from a user input via the microphone 1 and the word recognition itself is conducted in the central processing unit 4 together with the convertion into user-network-commands. Depending on the capacity of the central processing unit 4, it can also perform the spectrum analysis and the digital signal processor 3 can be omitted.

[0018] Further, the central processing unit 4 provides a learning function for the speech unit 2 so that the speech unit 2 can learn new vocabulary, grammar and user-network-commands to be sent to a network device 11 corresponding thereto. To be able to perform these tasks the central processing unit 4 is bidirectionally coupled to the memory 8 that is also holding information about the processes to be performed by the central processing unit 4. Further, the central processing unit 4 is bidirectionally coupled to an initial vocabulary section 7a, an extended vocabulary section 7b, an initial grammar section 7c, an extended grammar section 7d and a software section 7f that comprises a recognition section

DOCKET ALARM
Find authenticated court documents without watermarks at docketalarm.com.

and a grapheme-phoneme conversion section of the memory 7. Further, the central processing unit 4 is bidirectionally coupled to the home network system 10 and can also send messages to a digital signal processor 9 included in the speech unit 2 comprising a speech generation section 9a that serves to synthesize messages into speech and outputs this speech to a loudspeaker.

[0019] The central processing unit 4 is bidirectionally coupled to the home network 10 via a link layer control unit 5 and an I/F physical layer unit 6. These units serve to filter out network-commands from bus 10 directed to the speech unit 2 and to address network-commands to selected devices connected to the network 10.

[0020] Therefore, it is also possible that new user-network-commands together with corresponding vocabulary and grammars can be learned by the speech unit 2 directly from other network devices. To perform such a learning, the speech unit 2 can send control commands stored in the memory 8 to control the network devices, henceforth called control-network-commands, to request their user-network-commands and corresponding vocabulary and grammars according to which they can be controlled by a user. The memory 7 comprises an extended vocabulary section 7b and an extended grammar section 7d to store newly input vocabulary or grammars. These sections are respectively designed like the initial vocabulary section 7a and the initial grammar section 7c, but newly input user-network-commands together with information needed to identify these user-network-commands can be stored in the extended vocabulary section 7b and the extended grammar section 7d by the central processing unit 4. In this way, the speech unit 2 can learn user-network-commands and corresponding vocabulary and grammars built into an arbitrary network device. New network devices have then no need to have a built-in speech recognition device, but only the user-network-commands and corresponding vocabulary and grammars that should be controllable via a speech recognition system. Further, there has to be a facility to transfer these data to the speech unit 2. The speech unit 2 according to the invention learns said user-network-commands and corresponding vocabulary and grammar and the respective device can be voice-controlled via the speech unit 2.

[0021] The initial vocabulary section 7a and the initial grammar section 7c store a basic set of user-network-commands that can be used for various devices, like user-network-commands corresponding to the spoken-commands *switch on*, *switch off*, *pause*, *louder*, *etc.*, these user-network-commands are stored in connection with vocabulary and grammars needed by the central processing unit 4 to identify them out of the digital words produced by the speech recognition section via the digital signal processor 3. Further, questions or messages are stored in a memory. These can be output from the speech unit 2 to a user. Such questions or messages may be used in a dialogue in-between the speech unit 2 and the user to complete commands spoken by the user into proper user-network-commands, examples are *please repeat, which device, do you really want to switch off?*, *etc.* All such messages or questions are stored together with speech data needed by the central processing unit 4 to generate digital words to be output to the speech generation and synthesis section 9a of the digital signal processor 9 to generate spoken utterances output to the user via the loudspeaker. Through the microphone 1, the digital signal processors 3 and 9 and the loudspeaker a "bidirectional coupling" of the central processing unit 4 with a user is possible. Therefore, it is possible that the speech unit 2 can communicate with a user and learn from him or her. Like in the case of the communication with a network device 11, the speech unit 2 can access a set of control-network-commands stored in the memory 8 to instruct the user to give certain information to the speech unit 2.

[0022] As stated above, also user-network-commands and the corresponding vocabulary and grammars can be input by a user via the microphone 1 and the digital signal processor 3 to the central processing unit 4 on demand of control-network-commands output as messages by the speech unit 2 to the user. After the user has uttered a spoken-command to set the speech unit 2 into learning state with him, the central processing unit 4 performs a dialogue with the user on the basis of control-network-commands stored in the memory 8 to generate new user-network-commands and corresponding vocabulary to be stored in the respective sections of the memory 7.

[0023] It is also possible that the process of learning new user-network-commands is done half-automatically by the communication in-between the speech unit 2 and an arbitrary network device and half-dialogue controlled between the speech unit 2 and a user. In this way, user-dependent user-network-commands for selected network devices can be generated.

[0024] As stated above, the speech unit 2 processes three kinds of commands. i.e. spoken-commands uttered by a user, user-network-commands. i.e. digital signals corresponding to the spoken-commands, and control-network-commands to perform a communication with other devices or with a user to learn new user-network-commands from other devices 11 and to assign certain functionalities thereto so that a user can input new spoken-commands or to assign a new functionality to user-network-commands already included.

[0025] Output of the speech unit directed to the user are either synthesized speech or pre-recorded utterances. A mixture of both might be useful, e.g. pre-recorded utterances for the most frequent messages and synthesized speech for other messages. Any network device can send messages to the speech unit. These messages are either directly in orthographic form or they encode or identify in some way an orthographic message. Then these orthographic messages are output via a loudspeaker. e.g. included in the speech unit 2. Messages can contain all kinds of information usually

DOCKET ALARM
Find authenticated court documents without watermarks at docketalarm.com.

presented on a display of a consumer device. Furthermore, there can be questions put forward to the user in course of a dialogue. As stated above, such a dialogue can also be produced by the speech unit 2 itself to verify or confirm spoken-commands or it can be generated by the speech unit 2 according to control-network-commands to learn new user-network-commands and corresponding vocabulary and grammars.

[0026] The speech input and/or output facility, i.e. the microphone 1 and the loud-speaker, can also be one or more separate device(s). In this case messages can be communicated in orthographic form in-between the speech unit and the respective speech input and/or output facility.

[0027] Spoken messages sent from the speech unit 2 itself to the user, like *which device should be switched on?*, could also be asked back to the speech unit 2, e.g. *which network device do you know?*, and first this question could be answered by the speech unit 2 via speech, before the user answers the initial spoken message sent from the speech unit.

[0028] Fig. 2 shows a block diagram of an example of the structure of remotely controllable devices according to an embodiment of this invention, here a network device 11. This block diagram shows only those function blocks necessary for the speech controllability. A central processing unit 12 of such a network device 11 is connected via a link layer control unit 17 and an I/F physical layer unit 16 to the home network bus 10. Like in the speech unit 2, the connection in-between the central processing unit 12 and the home network bus 10 is bidirectional so that the central processing unit 12 can receive user-network-commands and control-network-commands and other information data from the bus 10 and send control-network-commands, messages and other information data to other network devices or a speech unit 2 via the bus 10. Depending on the device, it might also be possible that it will also send user-network-commands. The central processing unit 12 is bidirectionally coupled to a memory 14 where all information necessary for the processing of the central processing unit 12 including a list of control-network-commands needed to communicate with other network devices is stored. Further, the central processing unit 12 is bidirectionally coupled to a device control unit 15 controlling the overall processing of the network device 11. A memory 13 holding all user-network-commands to control the network device 11 and the corresponding vocabulary and grammars is also bidirectionally coupled to the central processing unit 12. These user-network-commands and corresponding vocabularies and grammars stored in the memory 13 can be down-loaded into the extended vocabulary section 7b and the extended grammar section 7d of the memory 7 included in the speech unit 2 in connection with a device name for a respective network device 11 via the central processing unit 12 of the network device 11, the link layer control unit 17 and the I/F physical layer unit 16 of the network device 11, the home network bus system 10, the I/F physical layer unit 6 and the link layer control unit 5 of the speech unit 2 and the central processing unit 4 of the speech unit 2. In this way all user-network-commands necessary to control a network device 11 and corresponding vocabulary and grammars are learned by the speech unit 2 according to the present invention and therefore, a network device according to the present invention needs no built-in device dependent speech recognizer to be controllable via speech, but just a memory holding all device dependent user-network-commands with associated vocabulary and grammars to be down-loaded into the speech unit 2. It is to be understood that a basic control of a network device by the speech unit 2 is also given without vocabulary update information. i.e. the basic control of a network device without its device dependent user-network-commands with associated vocabulary and grammars is possible. Basic control means here to have the possibility to give commands generally defined in some standard, like switch-on, switch-off, louder, switch channel, play, stop, etc..

[0029] Fig. 3 shows an example of a network architecture having an IEEE 1394 bus and connected thereto one speech unit 2 with microphone 1 and loudspeaker and four network devices 11.

[0030] Fig. 4 shows another example of a network architecture having four network devices 11 connected to an IEEE 1394 bus. Further, a network device 4 having a built-in speech unit with microphone 1 and loudspeaker is connected to the bus 31. Such a network device 41 with a built-in speech unit has the same functionality as a network device 11 and a speech unit 2. Here, the speech unit controls the network device 11 and the network device 41 which it is built-in.

[0031] Fig. 5 shows further three examples for network architectures. Network A is a network similar to that shown in Fig. 3, but six network devices 11 are connected to the bus 31. In regard to the speech unit 2 that is also connected to the bus 31, there is no limitation of network devices 11 controllable via said speech unit 2. Every device connected to the bus 31 that is controllable via said bus 31 can also be controlled via the speech unit 2.

[0032] Network B shows a different type of network. Here, five network devices 11 and one speech unit 2 are connected to a bus system 51. The bus system 51 is organized so that a connection is only necessary in-between two devices. Network devices not directly connected to each other can communicate via other third network devices. Regarding the functionality, network B has no restrictions in comparison to network A.

[0033] The third network shown in Fig. 5 is a wireless network. Here, all devices can directly communicate with each other via a transmitter and a receiver built into each device. This example shows also that several speech units 2 can be connected to one network. Those speech units 2 can have both the same functionality or

DOCKET ALARM
Find authenticated court documents without watermarks at docketalarm.com.

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.