US 20030083577A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2003/0083577 A1**

Greenberg (43) **Pub. Date: May 1, 2003**

(54) **VOICE-ENHANCED DIAGNOSTIC MEDICAL ULTRASOUND SYSTEM AND REVIEW STATION**

(76) Inventor: **Jeffrey M. Greenberg**, Palo Alto, CA (US)

Correspondence Address:
Siemens Corporation
Attn: Elsa Keller, Legal Administrator
Intellectual Property Department
186 Wood Avenue South
Iselin, NJ 08830 (US)

(21) Appl. No.: **10/306,369**

(22) Filed: **Nov. 26, 2002**

Related U.S. Application Data

(62) Division of application No. 09/239,271, filed on Jan. 29, 1999, now Pat. No. 6,514,201.

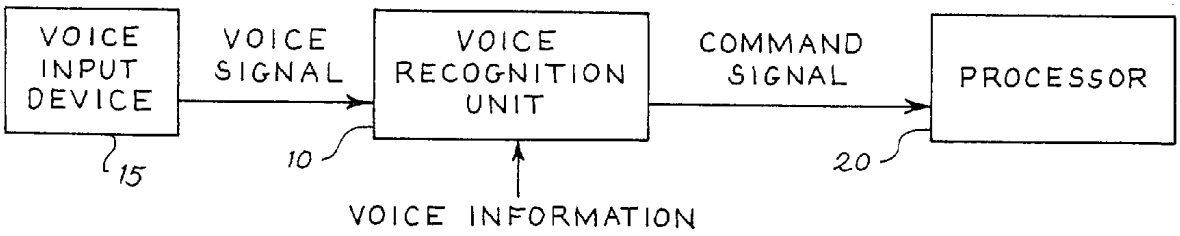
Publication Classification

(51) **Int. Cl.**[7] **A61B 8/00**; A61B 8/12

(52) **U.S. Cl.** **600/437**

(57) **ABSTRACT**

The preferred embodiments described herein relate to voice-enhanced diagnostic medical ultrasound imaging systems and review stations as well as to voice-related user interfaces. With these preferred embodiments, a user can interact with an imaging system or review station by issuing verbal commands instead of using a mouse, keyboard, or other user interface that requires physical manipulation by the user. This provides a very user-friendly interface, especially to those users who have difficulty navigating complex window and menu hierarchies or who have trouble manipulating pointing devices. This also improves patient flow and provides a more efficient report generation system. Voice feedback can also be used to allow the imaging system or review station to better communicate with a user.



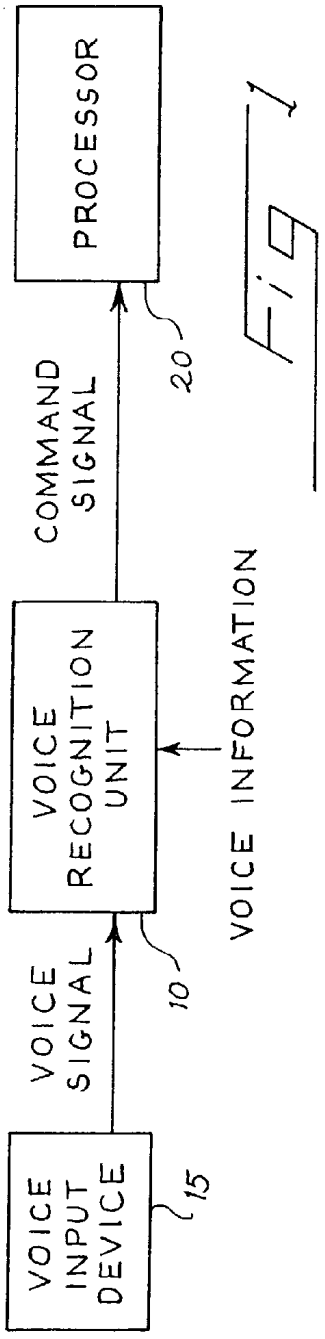


Fig. 1

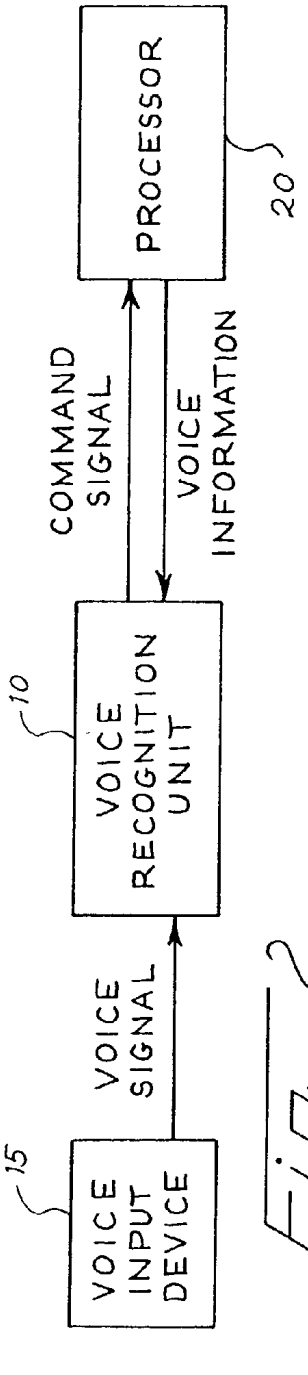


Fig. 2

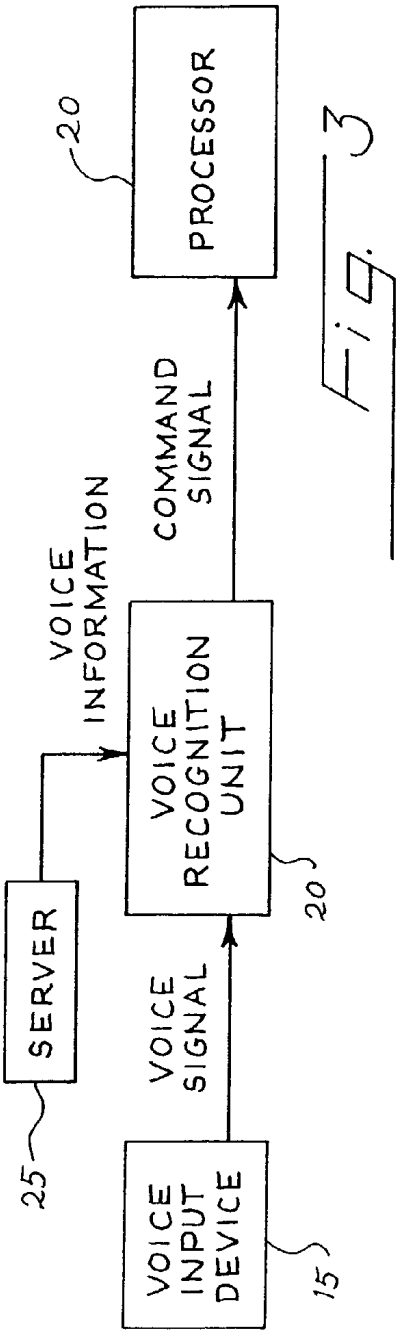


Fig. 3

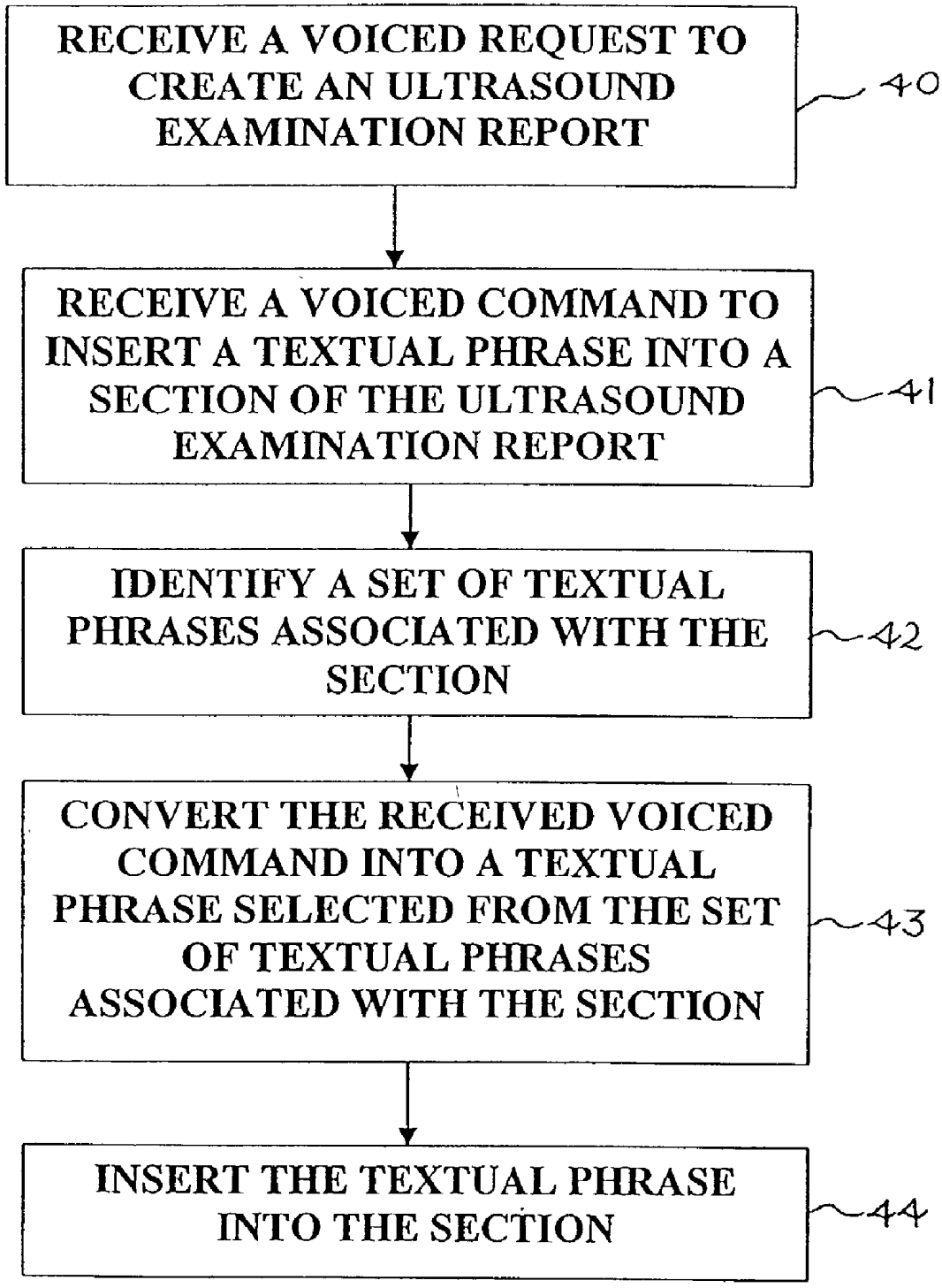

RECEIVE A VOICED REQUEST TO CREATE AN ULTRASOUND EXAMINATION REPORT
40
RECEIVE A VOICED COMMAND TO INSERT A TEXTUAL PHRASE INTO A SECTION OF THE ULTRASOUND EXAMINATION REPORT
41
IDENTIFY A SET OF TEXTUAL PHRASES ASSOCIATED WITH THE SECTION
42
CONVERT THE RECEIVED VOICED COMMAND INTO A TEXTUAL PHRASE SELECTED FROM THE SET OF TEXTUAL PHRASES ASSOCIATED WITH THE SECTION
43
INSERT THE TEXTUAL PHRASE INTO THE SECTION
44

Fig. 4

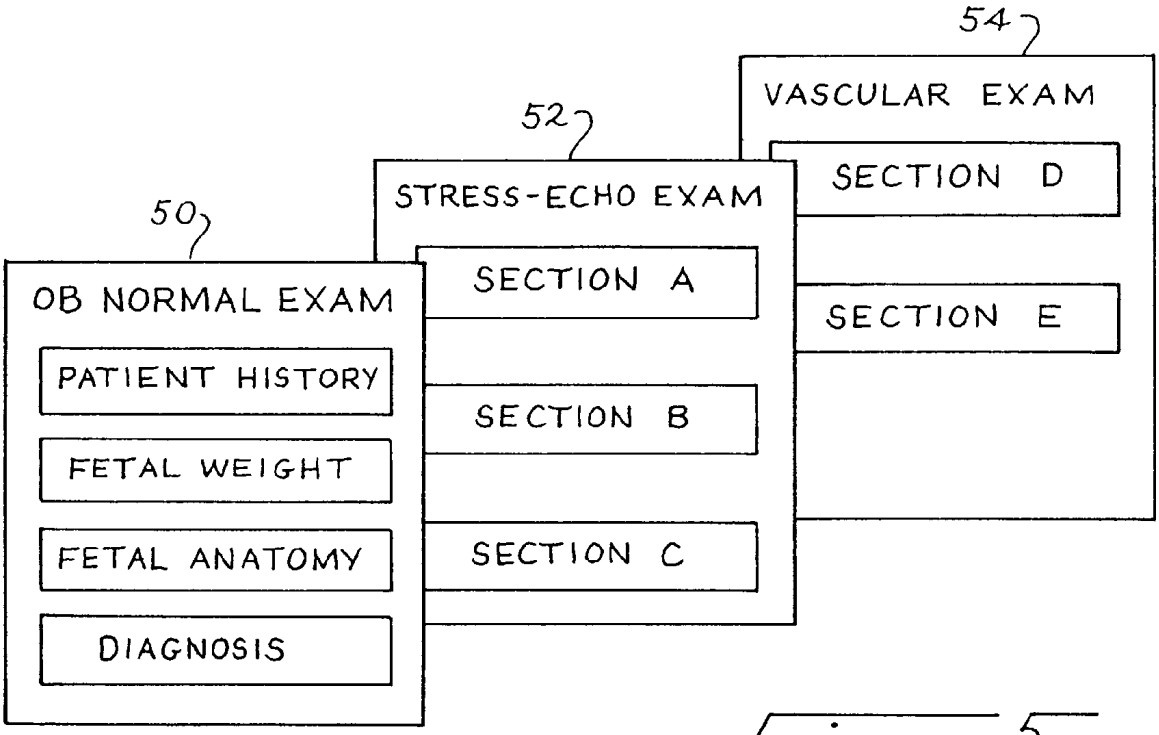


Fig. 5

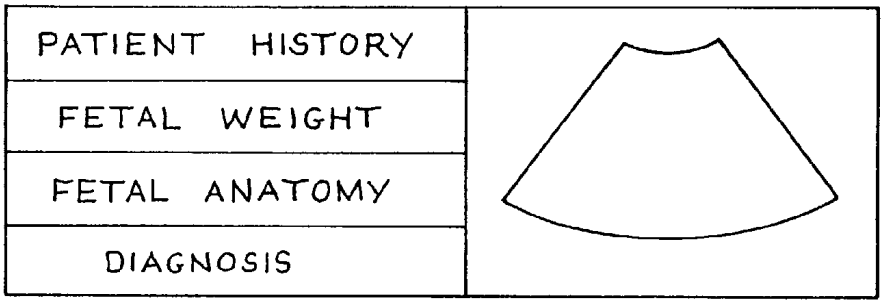


Fig. 6

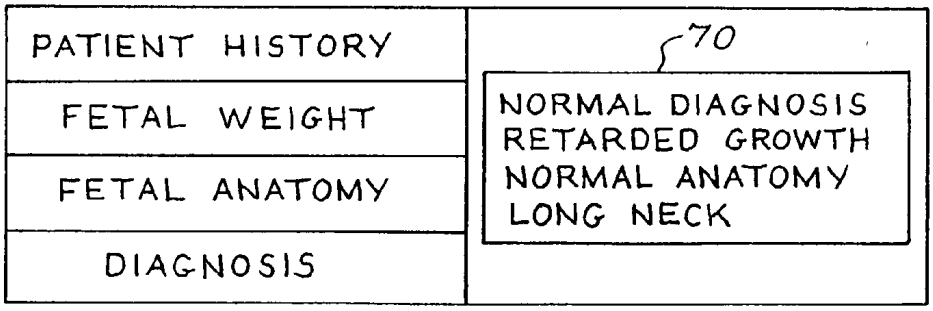


Fig. 7

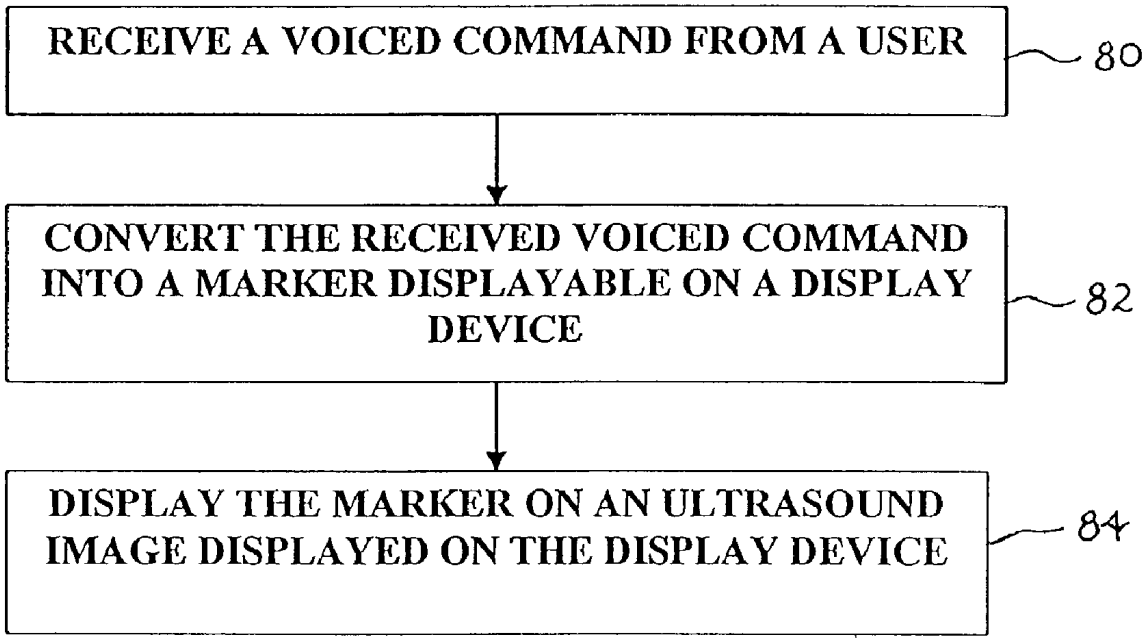

RECEIVE A VOICED COMMAND FROM A USER
80
CONVERT THE RECEIVED VOICED COMMAND INTO A MARKER DISPLAYABLE ON A DISPLAY DEVICE
82
DISPLAY THE MARKER ON AN ULTRASOUND IMAGE DISPLAYED ON THE DISPLAY DEVICE
84

Fig. 8

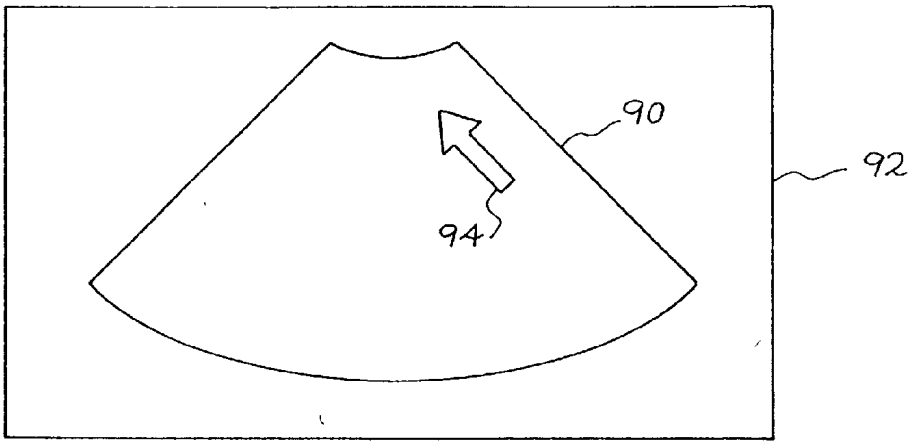


Fig 9

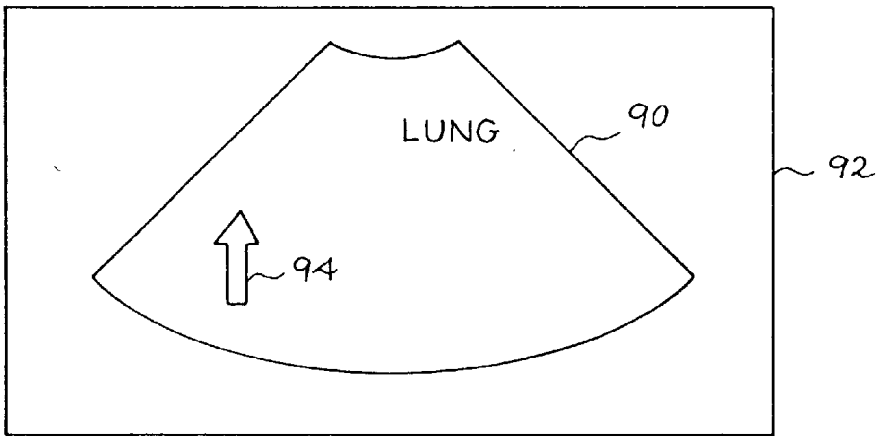


Fig. 10

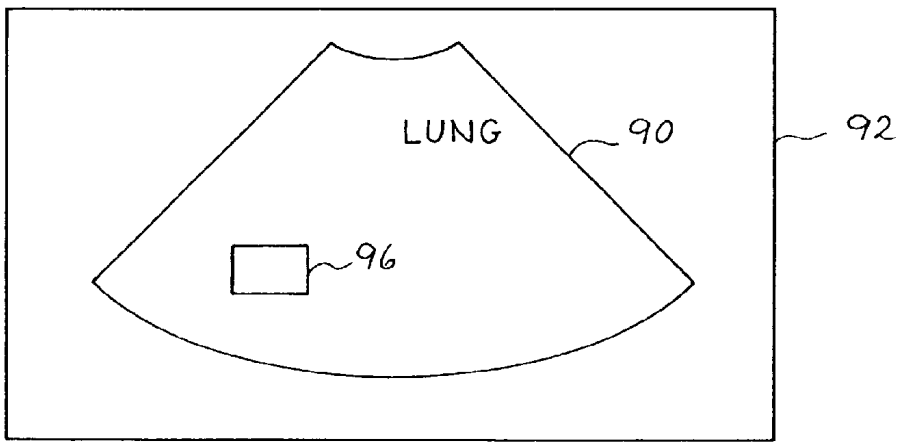


Fig. 11

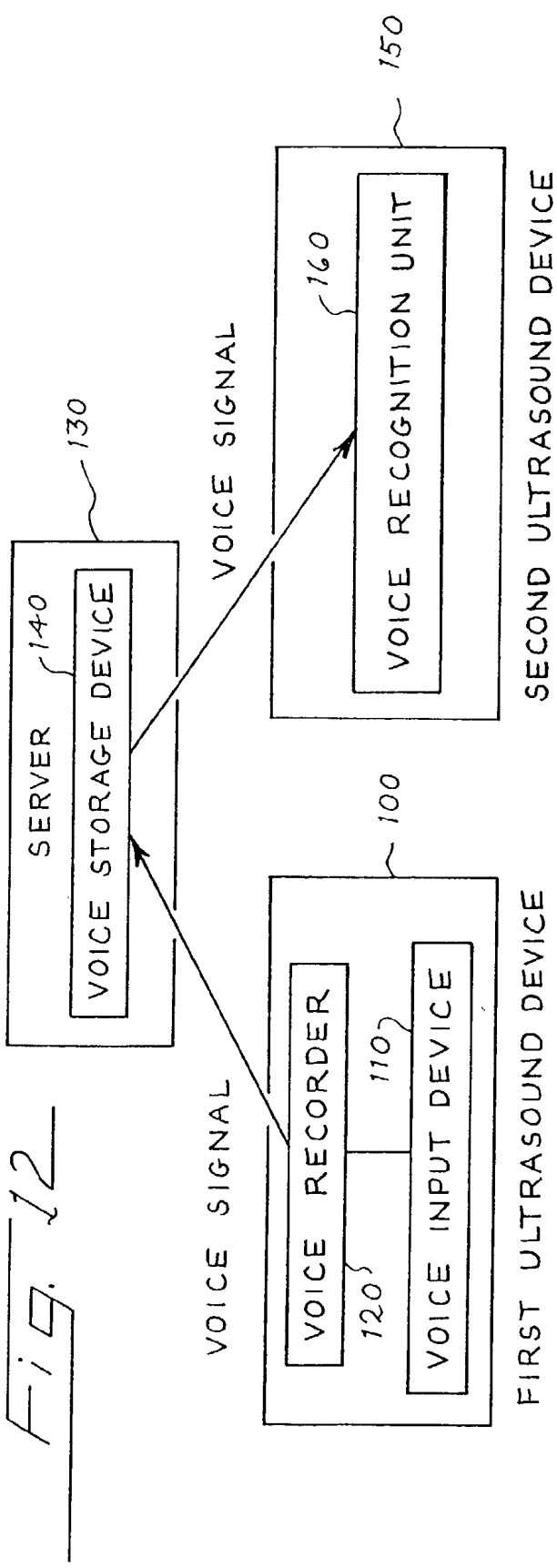

SERVER
130
VOICE STORAGE DEVICE
140
VOICE SIGNAL
VOICE SIGNAL
SECOND ULTRASOUND DEVICE
150
VOICE RECOGNITION UNIT
160
FIRST ULTRASOUND DEVICE
100
VOICE RECORDER
120
VOICE INPUT DEVICE
110

Fig. 12

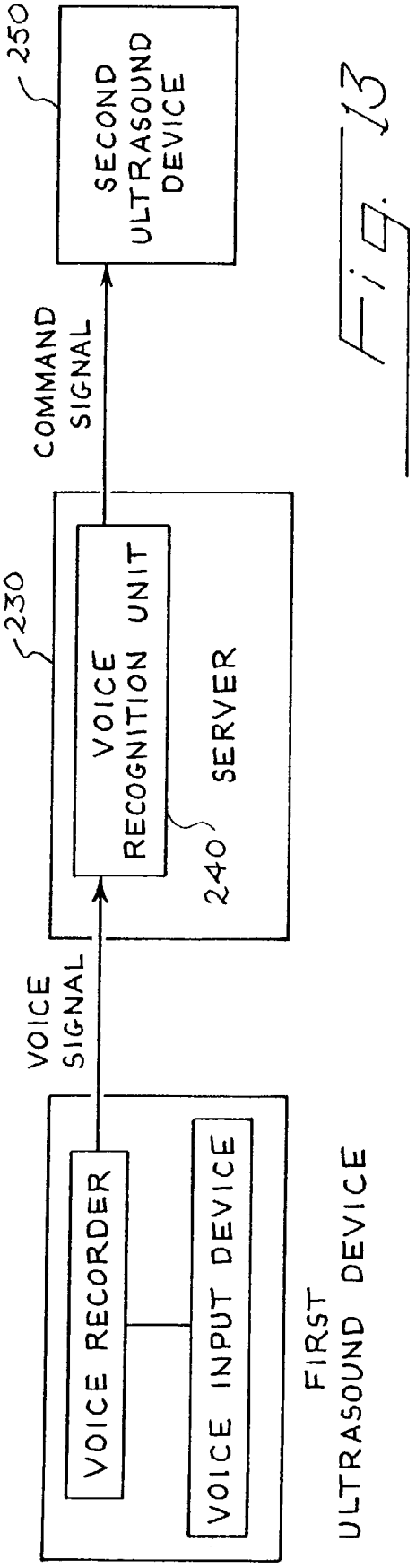

SECOND ULTRASOUND DEVICE
250
COMMAND SIGNAL
SERVER
230
VOICE RECOGNITION UNIT
240
VOICE SIGNAL
FIRST ULTRASOUND DEVICE
VOICE RECORDER
VOICE INPUT DEVICE

Fig. 13

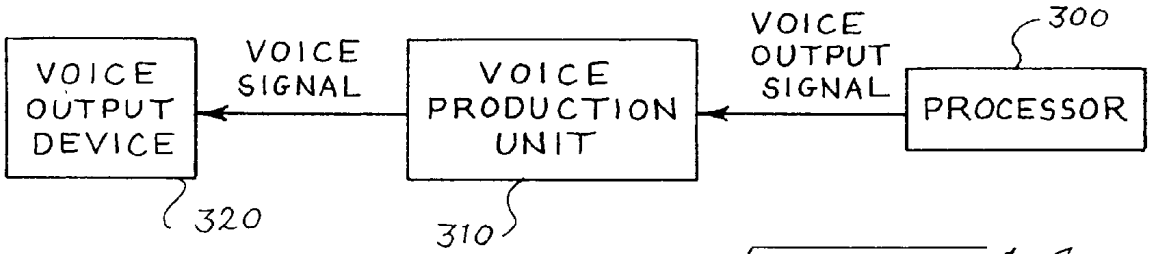


Fig. 14

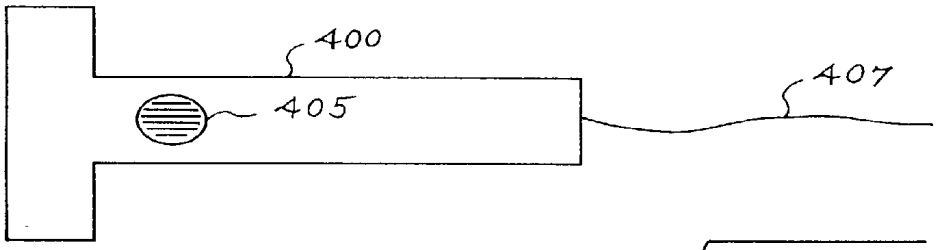


Fig. 15

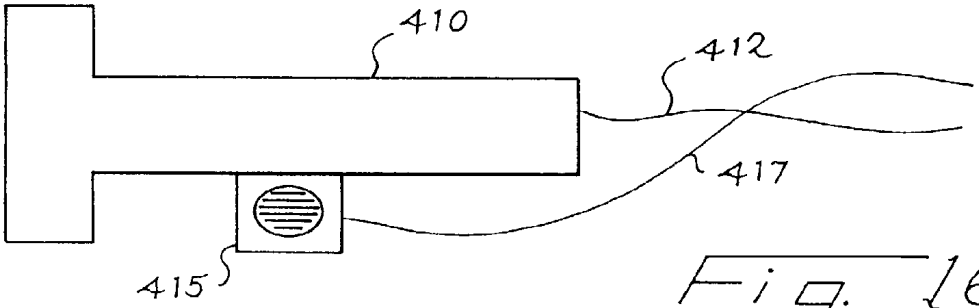


Fig. 16

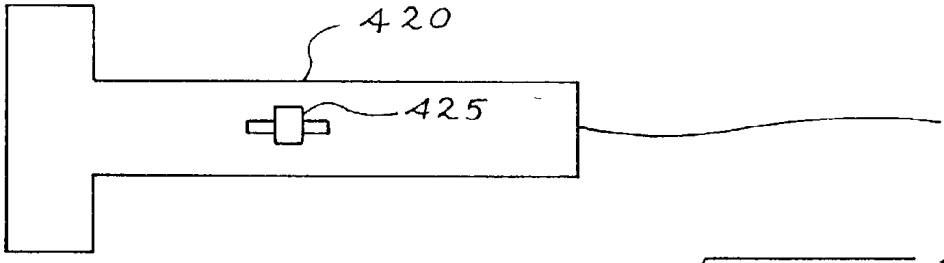


Fig. 17

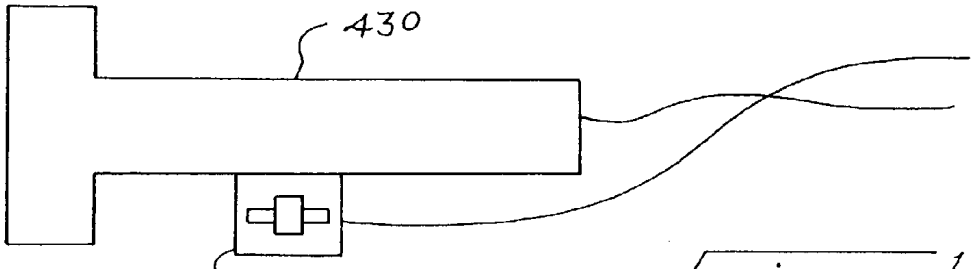


Fig. 18

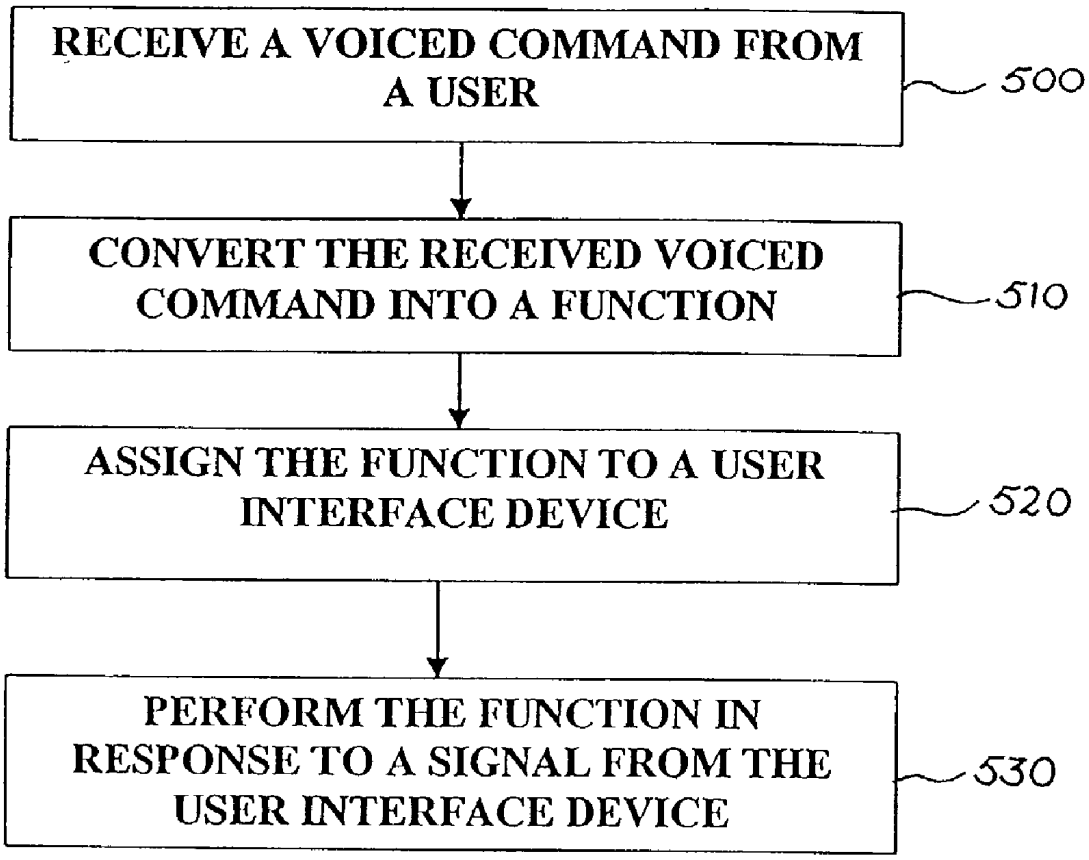

RECEIVE A VOICED COMMAND FROM A USER
500
CONVERT THE RECEIVED VOICED COMMAND INTO A FUNCTION
510
ASSIGN THE FUNCTION TO A USER INTERFACE DEVICE
520
PERFORM THE FUNCTION IN RESPONSE TO A SIGNAL FROM THE USER INTERFACE DEVICE
530

Fig. 19

VOICE-ENHANCED DIAGNOSTIC MEDICAL ULTRASOUND SYSTEM AND REVIEW STATION

BACKGROUND

[0001] There are several steps involved in providing a diagnosis of a patient based on an ultrasound examination. First, the ultrasound examination is performed on an ultrasound imaging system. The images generated from this examination can then be digitally stored and reviewed by a physician on an ultrasound review station, which is typically coupled with an ultrasound imaging system though a network. The ultrasound review station can display images, text, and measurement and calculation data and can also be used to facilitate the production of ultrasound examination reports. Based on his analysis at the review station, the physician generates an ultrasound examination report to provide a diagnosis. Often, a physician will dictate his diagnosis onto an audio tape or recording system, and the diagnosis is later transcribed and entered into an ultrasound examination report. Alternatively, the diagnosis can be typed into the ultrasound imaging system.

[0002] To assist in the performance of an ultrasound examination, some ultrasound imaging systems allow voice control of some of the operations of the system. Typically, a voice recognition unit, which is either part of or separate from the ultrasound imaging system's processor, converts an incoming voice signal to a control signal using voice information stored in the voice recognition unit. To enhance recognition performance, U.S. Pat. No. 5,544,654, which is assigned to the assignee of the present invention, describes an ultrasound imaging system in which a subset of voice information is used based on the operating state of the ultrasound imaging system. Specifically, the ultrasound imaging system's processor provides the voice recognition unit with an indication of its operating state, and the voice recognition unit selects only the portions the voice information that are relevant to the operating state. Because the voice recognition unit makes the selection based on the provided indication of operating state, the voice recognition unit and processor must be synchronized to ensure proper selection, especially when the processor is shipped separately from the voice recognition unit and when the processor is updated without updating the voice recognition unit.

[0003] To assist the physician review ultrasound images at a review station, graphical user interfaces have been used to provide a more user-friendly environment for the physician. Typically, these graphical user interfaces have windows, menus, and buttons, and a visual focus manipulated by a pointing device such as a mouse, keyboard, or trackball. Ultrasound review stations often have so many functions that applications are divided into hierarchies of menus and sub-menus, dialogs and sub-dialogs, and windows and sub-windows. Although graphical user interfaces were intended to facilitate interaction with the review station, some users have difficulty finding the desired functionality in the complex windows and menu hierarchies. Some users also find it difficult to fluidly manipulate pointing devices that require click and double-click actions.

[0004] Finally, to reduce the time needed to produce an ultrasound examination report and to improve the overall diagnostic workflow for a patient, automatic transcription systems have been used, such as Medspeak from IBM, Clinical Reporter from Lernout & Hauspie, and Powerscribe from the MRC Group. These systems are stand-alone devices with specialized vocabularies and are not incorporated with the ultrasound imaging system or review station. Some transcription systems attempt to transcribe every word voiced by the physician. Because of limitations in current transcription technology, these systems often produce inaccurate transcriptions. To overcome this problem, some systems reduce the amount of automatic dictation that is needed by creating macros, which, when spoken, trigger a longer text to be inserted into the report. Although typically more accurate than automatic dictation systems, these systems also encounter recognition problems. To further enhance accuracy, some systems use inline-style macro displays, in which a proposed textual phrase is displayed to a user for acceptance. Because these systems only need to recognize the command to accept or reject the proposed textual phrase, recognition accuracy is increased. However, presenting proposed textual phrases to a user can be a time consuming process, especially if the user rejects several proposed phrases before reaching an acceptable phrase.

[0005] There is, therefore, a need for an improved diagnostic medical ultrasound imaging system and review station to overcome the problems described above.

SUMMARY

[0006] The present invention is defined by the following claims, and nothing in this section should be taken as a limitation on those claims.

[0007] By way of introduction, the preferred embodiments described below relate to voice-enhanced diagnostic medical ultrasound imaging systems and review stations as well as to voice-related user interfaces. With these preferred embodiments, a user can interact with an imaging system or review station by issuing verbal commands instead of using a mouse, keyboard, or other user interface that requires physical manipulation by the user. This provides a very user-friendly interface, especially to those users who have difficulty navigating complex window and menu hierarchies or who have trouble manipulating pointing devices. This also improves patient flow and provides a more efficient report generation system. Voice feedback can also be used to allow the imaging system or review station to better communicate with a user.

[0008] The preferred embodiments will now be described with reference to the attached drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0009] **FIG. 1** is a block diagram illustrating an operation of a voice recognition unit of a preferred embodiment.

[0010] **FIG. 2** is a block diagram illustrating a voice recognition unit of a preferred embodiment receiving voice information from an ultrasound imaging system or review station processor.

[0011] **FIG. 3** is a block diagram illustrating a voice recognition unit of a preferred embodiment receiving voice information from a server.

[0012] **FIG. 4** is a flow chart of a preferred method for using voice commands to insert a textual phrase into a section of an ultrasound examination report.

[0013] **FIG. 5** is an illustration of several preferred ultrasound examination reports.

[0014] **FIG. 6** is an illustration of a displayed ultrasound examination report of a preferred embodiment.

[0015] **FIG. 7** is an illustration of a displayed ultrasound examination report of another preferred embodiment in which a menu of available commands is displayed along with the report.

[0016] **FIG. 8** is a flow chart of a preferred method for using voice commands to display a marker on an ultrasound image.

[0017] FIGS. **9-11** are illustrations of preferred displayed ultrasound images used in conjunction with the preferred method of the flow chart of **FIG. 8**.

[0018] **FIG. 12** is a block diagram of a network arrangement of a preferred embodiment.

[0019] **FIG. 13** is a block diagram of another network arrangement of a preferred embodiment.

[0020] **FIG. 14** is a block diagram illustrating an operation of a voice production unit of a preferred embodiment.

[0021] **FIG. 15** is an illustration of a preferred transducer having a built-in microphone.

[0022] **FIG. 16** is an illustration of a preferred transducer having an attached microphone.

[0023] **FIG. 17** is an illustration of a preferred transducer having a built-in slider.

[0024] **FIG. 18** is an illustration of a preferred transducer having an attached slider.

[0025] **FIG. 19** is a flow chart of a preferred method for using voice commands to assign a function to an ultrasound imaging system or ultrasound review station user interface.

DETAILED DESCRIPTION OF THE PRESENTLY PREFERRED EMBODIMENTS

[0026] Voice Recognition Unit

[0027] A voice recognition unit can be used to provide voiced commands to an ultrasound imaging system or review station, and **FIG. 1** provides a block diagram of a voice recognition unit **10** of a preferred embodiment. As shown in **FIG. 1**, the voice recognition unit **10** is coupled with a voice input device **15** (such as a microphone) and a processor **20** of an ultrasound imaging system or review station. As used herein, the term "coupled with" means directly coupled with or indirectly coupled with through one or more components. It is preferred that the voice recognition unit **10** comprise a recognition engine (such as one available from Dragon Systems, IBM, Lernout & Hauspie, or Entropic Ltd.) and a voice technology software interface (such as SAPI 4.x from Microsoft Corp., SRAPI, JSAPI java from Sun Microsystems Corp., or SVAPI).

[0028] In regard to architecture, the voice recognition unit **10** can be separate from the processor **20**, such as when the voice recognition unit **10** takes the form of software running on a separate processor. In one embodiment, the separate processor is a general-purpose computer directly coupled with the ultrasound imaging system or review station. For example, a general-purpose computer can be directly connected to an ultrasound imaging system and carried on the system cart, thereby appearing to a user to be integrated with the system. As described below, a separate processor can also be located in a server coupled with the ultrasound imaging system or review station through a network. In another preferred embodiment, some or all of the functionality of the voice control unit **10** is implemented with the ultrasound imaging system's or review station's processor **20**.

[0029] For simplicity, the term "voice recognition unit" is used in the specification and claims to broadly refer to hardware and/or software components that use voice information to recognize an incoming voice signal from a voice input device **15** to generate and provide a command signal to a processor **20** of an ultrasound imaging system or review station. As used herein, the term "voice information" refers to data that correlates a set of voice signals (e.g., voiced commands from a user) with an associated set of respective command signals recognizable by the processor of the ultrasound imaging system or review station. The term "set" in the specification and claims refers to one or more than one element. In addition to providing this recognition profile, voice information can include engine usage information (e.g., percent of the CPU dedicated to recognition), user-adjustable recognition parameters (e.g., minimum volumes, timeouts to recognized complete and incomplete phrases), and a list of voice input devices and their capabilities that the user trained with and that the recognition profile is appropriate for. Voice information can also include user-specific voice commands and non-GUI user-specific voice parameters, such as preferred speak-back voice, dictation parameters, and dialog parameters. The recognition engine of the voice recognition unit **10** compares the incoming voice signal with the recognition profile of the voice information to determine which command signal should be sent to the processor **20**. To enhance recognition performance, it is preferred that a finite-state-machine language description (e.g., Backus-Naurer) be used to provide the voice recognition unit **10** with various forms of legal speech and that a dictionary of synonyms be used to recognize equivalent voiced commands.

[0030] Also as used herein, the term "command signal" is used to refer to any signal that is recognizable by the processor **20** as an instruction to perform an operation or function performable by the processor **20** (e.g., the selection of a field, window, or monitor). In addition to controlling some aspect of the ultrasound imaging system or review station, a command signal can be a signal that provides the processor **20** with a text or other message: For example, the command signal can comprise a textual phrase that will be inserted into an ultrasound examination report.

[0031] Voice Information

[0032] As described above, the voice recognition unit **10** uses voice information to generate a command signal in response to a voiced command from a user. Voice information can be provided to the voice recognition unit **10** by any appropriate source. In the preferred embodiment shown in **FIG. 2, a** server coupled with and integral with the ultrasound imaging system or review station processor **20** provides the voice information to the voice recognition unit **10**. With this preferred embodiment, the processor **20** can create voice information containing only those voice and command

signals that are relevant to the current operating state of the processor **20**, thereby enhancing voice recognition. For example, if an ultrasound review station is currently being used to generate an ultrasound examination report for an obstetrics examination, the processor **20** can provide the voice recognition unit **10** with voice information containing only those voice and command signals commonly used with an obstetrics examination. Because the processor **20** provides voice information and not merely an indication of operating state, the processor **20** and voice recognition unit **10** can operate independently, thereby eliminating the problems encountered in prior system when the processor and voice recognition unit were not synchronized.

[**0033**] The voice information can also be provided to the voice recognition unit **20** by a server **25** externally coupled with the voice recognition unit **20**, as shown in **FIG. 3** This preferred embodiment is particularly useful in a network environment in which the server **25** is coupled with several ultrasound imaging systems or review stations. In such an environment, voice information customized for a particular user is centrally located in the server **25**. When that user identifies himself to a device on the network, the user's customized voice information is provided to the voice recognition unit of that device. This provides a coherent network environment in which commands are consistently recognized. If the user changes the voice information when using the device (such as when the user corrects a misidentified word), the centrally-stored voice information is updated as well. In this way, the user is not only provided with greater access to his customized voice information, but he is also given more opportunities to update the voice information to enhance recognition performance. If a user is using multiple ultrasound devices on the network simultaneously, the voice information that is accessed first or, alternatively, the voice information that is the most up-to-date can be used.

[**0034**] To select customized voice information (stored in an external server **25** or in a server integrated with the ultrasound imaging system or review station), the user can identify himself by providing non-verbal identification information, such as by typing his user name and password into a log-in screen. If the voice recognition unit **20** comprises a speaker identification engine (such as Keyware by Keyware Technologies or SpeakEZ Voice Print by T-Netix Inc.), the user can also identify himself by providing verbal identification information. For example, with a speaker identification engine, the voice recognition unit **20** can identify the user when he voices a command to the voice input device **15** or when he provides a voice sample to gain access to the ultrasound system or review station, as described in U.S. patent application Ser. No. _____ (Attorney Docket No. 5050/292), which is hereby incorporated by reference and assigned to the assignee of the present application.

[**0035**] Applications

[**0036**] As described above, the voice recognition unit **10** can be implemented in either an ultrasound imaging system or an ultrasound review station to control the operation of the system or station. For example, by using the voice recognition unit **10** with an ultrasound review station, instructions to the review station can be given using voice commands instead of or in combination with using a mouse, keyboard, or other user interface that requires physical manipulation by the user. A verbal interface provides a user with a much more user-friendly interface, especially for those users who have difficulty finding the desired function in complex window and menu hierarchies or who have trouble manipulating pointing devices. With the voice recognition unit **10**, a user can instruct the review station to view a desired report, worksheet, study list, or image. The user can also use voice commands to navigate through display information (e.g., "page up"), respond to visual requests (e.g., "press ok"), and perform operations (e.g., "zoom image **5**"). The voice command can also be associated with multiple operations. For example, the command "use equation A to calculate birth weight" can trigger the ultrasound review system to perform the requested calculation and place the result into a particular section of an ultrasound examination report. Such operations can reference imaging, post-processing, and computational and calculation data.

[**0037**] The following is a description of specific implementations of methods that can be performed with a voice recognition unit. For simplicity, the following embodiments will be described in terms of an ultrasound review station. It is important to note that all of these implementation can also be implemented on an ultrasound imaging system and that one or more of these applications can be used in combination.

[**0038**] In one preferred embodiment, the voice recognition unit **10** implements a method for using voiced commands to insert a textual phrase into a section of an ultrasound examination report. As used herein, the term "phrase" refers to a string of one or more characters, such as one or more letters, words, numbers, or symbols, and the term "insert" refers to inserting, overwriting, replacing, and/or adding at a specified location. Such a method is illustrated in the flow chart of **FIG. 4**. The first step in this method is to receive a voice request to create an ultrasound examination report (**40**). As shown in **FIG. 5**, there are typically several types of ultrasound examination reports available to the user. Each report is associated with a particular type of ultrasound examination. For example, an ultrasound examination report can be used to provide information about a normal obstetrics examination **50**, a stress-echo examination **52**, or a vascular examination **54**. Each report also has at least one section that will accept information from a user. For example, the normal obstetrics examination report **50** includes patient history, fetal weight, fetal anatomy, and diagnosis sections.

[**0039**] In response to the voiced request to create a particular ultrasound examination report, the ultrasound review station displays the report template to the user. For example, if the user requests a normal obstetrics examination report, the template with the sections associated with such a report is displayed to the user, as shown in **FIG. 6**. The user is also informed which section in the report is active (i.e., which section will received input from the user). For example, a cursor positioned in the diagnosis section would inform the user that the diagnosis section will receive input.

[**0040**] Next, a voiced command is received from the user to insert a textual phrase into the active section (**41**). Preferably, the voiced command is a single word or short phrase that triggers a macro for the insertion of a longer textual phrase into the report. For example, the command

"normal diagnosis" can be associated with the textual phrase "There is no evidence of abnormal development. The fetus is normal" To assist the user in selecting a command, the available voice commands for the active section can be provided to the user, for example, by displaying a menu of available commands for that section. In the example shown in **FIG. 7**, the menu **70** informs the user that the commands "normal diagnosis,""retarded growth,""normal anatomy," and "long neck" are available for the diagnosis section. In one preferred embodiment, the menu is automatically displayed to the user, while in another preferred embodiment, the menu is provided and/or removed in response to a verbal request from the user.

[**0041**] To increase recognition performance, each section of the report can be associated with a respective set of textual phrases. To convert the voiced command into a textual phrase, first the set of textual phrases associated with the active section is identified (**42**), and then the voiced command is converted into one of the textual phrases from the set (**43**). By basing the conversion on only those textual phrases associated with the active section instead of all available textual phrases, recognition performance is enhanced. Finally, the textual phrase is inserted into the section (**44**). This method can be used instead of or in conjunction with a pure transcription service. Additionally, a user can voice commands to perform an editing operation (e.g., cut-and-paste) to the ultrasound examination report or to insert an ultrasound image into the report.

[**0042**] In another preferred embodiment, the voice recognition unit is used to place a marker on an ultrasound image displayed on a display device of an ultrasound review station. As used herein, the term "marker" is intended to broadly refer to any textual word or phase or any graphic that can be displayed on an ultrasound image displayed on a display device of an ultrasound imaging system or review station. A marker includes, but is not limited to, a word or phase used to identify anatomy and a geometric shape (such as a square or circle) used to identify a region of interest.

[**0043**] **FIG. 8** is a flow chart of a preferred method for using voice commands to display a marker on an ultrasound image, which will be illustrated in conjunction with the ultrasound images shown in FIGS. **9-11**. **FIG. 9** is an illustration of an ultrasound image **90** displayed on a display device **92** of an ultrasound review station. When the user desires to place a marker on the ultrasound image **90**, he positions a pointer **94** over the location on the ultrasound image **90** where he wishes to place the marker After the pointer is positioned, the user issues a voice command to display the marker on the image (**80**). For example, if the user wishes to label the image **90** under the pointer **94** with the word "lung," he would issue a voiced command such as "lung."When voice commands are used for a plurality of applications, it is preferred to provide the system with a voiced command indicating that a subsequent voiced command will be associated with a marker. For example, the voiced command "label" can be given before every voiced command for a marker to inform the system that it should display the marker associated with the first subsequent voiced command. Alternatively, the voiced command "label on" can be given once, and all subsequent voiced commands can be converted to markers until the voiced command "label off" is given or until another command is issued that implicitly ends labeling.

[**0044**] Next, the received voiced command is converted into a marker displayable on the display device (**82**). In one embodiment, the received voiced command is compared to the system's entire voice information to determine which marker is associated with the voiced command. To improve recognition, it is preferred that only a subset of the voice information be used to convert the voiced command into a marker. Preferably, ultrasound images are classified by a study type, and only those markers associated with that study type are used to convert the voiced command into a marker. An indication of the study type can be provided with the image, or the system can analyze the image to determine the set of anatomical regions and the corresponding vocabulary.

[**0045**] After the voiced command is converted to a marker, the marker is displayed on the ultrasound image (**84**). As shown in **FIG. 10**, the word "lung" is displayed on the ultrasound image **90** at the location identified by the pointer **94** in **FIG. 9**. In this way, a user can rapidly name structure and apply markers identifying that structure to the relevant parts of an image. As mentioned above, a geometric shape can be used as a marker. This finds particular application when a sonographer notices a diagnostic region of interest that a physician should review but does not wish to alarm the patient In this case, the sonographer can position the pointer **94** over a location of the ultrasound image (as shown in **FIG. 10**) and issue a voiced command, such as "rectangle." This voiced command results in a rectangle **96** displayed on the image **90**, as shown in **FIG. 11**.

[**0046**] In the embodiment describe above, a user first positioned a pointer over an ultrasound image, and the marker was placed on the image at the location identified by the pointer. In an alternate embodiment, the marker that is displayed on the ultrasound image can be positioned by the user with a pointing device such as a mouse. Pressing the mouse button would pin the marker in place. Further, a voiced command can be converted into a plurality of markers that are each positioned by the user. For example, if the user says "tumors," multiple "tumor" markers are displayed. The user then can use a pointing device to drag and drop each marker at the appropriate location on the image. Additionally, the user can position the pointer over a displayed marker and voice a command to copy or delete the marker.

[**0047**] In another preferred embodiment shown in **FIG. 12, a** first ultrasound device **100** comprising a voice input device **110** and a voice recorder **120** is coupled with a server **130** comprising a voice storage device **140**. The server **130** is also coupled with a second ultrasound device **150** comprising a voice recognition unit **160**. As used in this preferred embodiment and those described below, the term "ultrasound device" is intended to define either an ultrasound imaging system or an ultrasound review station. For example, in the embodiment shown in **FIG. 12**, the first ultrasound device **100** can be an ultrasound imaging system, and the second ultrasound device **150** can be an ultrasound review station.

[**0048**] With the network relationship shown in **FIG. 12**, voice signals provided to the voice input device **110** of the first ultrasound device **100** are converted into digital form by the voice recorder **120** and are transmitted for storage in the voice storage device **140** of the server **130**. These stored

voice signals can then be transmitted to the second ultrasound device **150** for conversion by its voice recognition unit **160**. In one alternate embodiment, instead of using a network relationship, voice signals can be transferred from the first ultrasound device **100** via a portable storage device, such as a magneto-optical disc, or via a communication link (such as a cable) directly coupling the first and second ultrasound devices **100**, **150**. In another alternate embodiment shown in **FIG. 13**, the server **230** has a voice recognition unit **240** and supplies the second ultrasound device **250** with a command signal instead of a voice signal. In yet another alternate embodiment, the server **230** provide several ultrasound devices in the ultrasound network with the command signal. This is particularly useful in teaching environments, in which the command signal can be used to control several devices on the network. Also, the first and second ultrasound devices can be the same device, such as when an ultrasound imaging system provides voiced commands to the server, and the server returns command signals to the ultrasound imaging system.

[**0049**] These preferred embodiments are particularly useful when the first ultrasound device is an ultrasound imaging system and the second ultrasound device is an ultrasound review station. After an ultrasound examination is performed with the ultrasound imaging system, a user can dictate information into the voice input device for a report to be generated for the examination on the review station. The dictated information is recorded by a voice recorder of the ultrasound examination system, and the voice data is provided to the ultrasound imaging system via the server or portable storage device. The voice recognition unit of the server or ultrasound review station transcribes the voice data and inserts the transcription into an ultrasound examination report. The above-described embodiments associated with generating an ultrasound examination report can be used to enhance recognition. For example, in dictating a Normal OB report, the user can utter: "Diagnosis section: Normal diagnosis". On conversion, the command "normal diagnosis" would be trigger a macro to provide a textual phrase in the diagnosis section of a Normal OB report.

[**0050**] Voice Feedback

[**0051**] A voice production unit can be used with to provide voice feedback with an ultrasound imaging system or review station, and **FIG. 14** provides a block diagram of a voice production unit **310** of a preferred embodiment. As shown in **FIG. 14**, the voice production unit **310** is coupled with a voice output device **320** (such a speaker) and a processor **300** of an ultrasound imaging system or review station. As with the voice recognition unit, the voice production unit **310** can comprise hardware and/or software components. A suitable voice production unit **310** can be purchased from Dragon Systems, Elan Informatique, IBM, Microsoft, Eloquent, Lernout & Hasupie, and AcuVoice. The voice production unit **310** can generate a synthesized or prerecorded voice. In operation, the processor **300** provides the voice production unit **310** with a voice output signal. The voice production unit **310** converts this signal to a voice signal and provides the voice signal to the voice output device **320**.

[**0052**] With a voice production unit **310**, an ultrasound imaging system or review station can bring information to the user's attention using voice feedback instead of or along with displaying such information visually. In this way, the ultrasound imaging system or review station can communicate with a user without cluttering a display screen. Additionally, voice feedback can be used to provide the user with information that is of interest but not important enough to merit distracting the user by presenting it visually. In this way the information can be provided to the user as background audio, which the user can choose to ignore. The voice production unit can be used with or integrated with a voice recognition unit, such as when the voice production unit and the voice recognition unit share some or all or its hardware and/or software components. An ultrasound imaging system or review station using both a voice recognition unit and a voice production unit can provide a fluid voice environment. For example, voice feedback can be used to confirm an action ("Are-you-sure?") and can also be used to reply to a voiced request (e.g., User asks, "Is the study in room **10** done? and the system responds in voice, "The study is complete and prior studies are being obtained"). The voice production unit can also be used to provide verbal alerts, the status of a voice recognition unit, and an indication of a completion-of an activity. To avoid interruptions, the user can also command the system not to provide voice feedback.

[**0053**] Voice-Related User Interfaces

[**0054**] There are several voice-related user interfaces that can be used with an ultrasound imaging system or review station. A user can issue voiced commands to an ultrasound imaging system or review station via a headset, a wireless microphone (such as a microphone manufactured by Shure), an attached microphone, or an array microphone. To allow a user to work closer to a patient, the ultrasound transducer shown in **FIGS. 15 and 16** can be used. **FIG. 15** is an illustration of a preferred ultrasound transducer **400**. This ultrasound transducer **400** has at least one transducer element and a microphone **405**. Both the transducer element(s) and microphone **405** provide electrical signals to an ultrasound imaging system coupled with the ultrasound transducer **400**. In the transducer **400** shown in **FIG. 15**, the microphone **405** is built-into the transducer **400**, and a single electrical connector **407** carries the electrical signals from the transducer element(s) and the microphone **405** to the ultrasound imaging system. In **FIG. 16**, the microphone **415** is added onto the transducer **410** as a removably-coupled component, which has an electrical connector **417** that is separate from the connector **412** used to carry electrical signals from the transducer element(s), although any appropriate number of connectors (including a single connector) can be used.

[**0055**] In another preferred embodiment, voice commands can be used to assign a function to a user interface device built into or attached to an ultrasound imaging system or ultrasound review station. The user interface device can take any suitable form (such as, but not limited to, a wheel, button, trackball, slider, and knob) attached directly or indirectly to the ultrasound imaging system or review station. For example, the user interface can be a pre-existing button on an ultrasound review station keyboard or a specially-designed knob added to an ultrasound imaging system. The user interface device can also be part of or attached to an ultrasound transducer. For example, **FIG. 17** is an illustration of an ultrasound transducer **420** with a built-in slider **425**, and **FIG. 18** is an illustration of an ultrasound transducer **430** with a removably-attached slider **435**. Although two separate electrical connectors are shown in

FIG. 17, a single connector can be used. With the user interface device on the transducer, continuous or button-type controls previously located at the ultrasound imaging system can be manipulated at a distance.

[**0056**] **FIG. 19** is a flow chart of a preferred method for using voice commands to assign a function to an ultrasound imaging system or ultrasound review station user interface. First, a voiced command from the user is received (**500**). This voiced command is converted into a function (**510**), and the function is assigned to a user interface device (**520**). When the user manipulates the user interface device causing it to transmit a signal to the processor of the ultrasound imaging system or review station, the assigned function is performed (**530**).

[**0057**] With these preferred embodiments, the function of a user interface device can be easily changed by voice, such as when an ultrasound transducer has both a built-in depressible wheel and an attached microphone. As a user is performing an ultrasound examination, he speaks the word "gain" into the microphone of the ultrasound transducer. The ultrasound imaging system would then assign the gain function to the wheel. When the user scrolls the wheel forward, the ultrasound imaging system would increase the gain, and when the user scrolls the wheel backwards, the ultrasound imaging system would decrease the gain. If the user then says "image," an image would be generated when the user presses the depressible wheel. In this way, by issuing simple voice commands, the user can control a variety of functions with a single user interface device. Of course multiple user interface devices can be used in combination.

[**0058**] Appendix A provides further details of a presently preferred embodiment. Additionally, while the above-preferred embodiments were described above with respect to an ultrasound review station, review stations for use with other imaging modalities can be used.

[**0059**] The foregoing detailed description has described only a few of the many forms that this invention can take. Of course, many changes and modifications are possible to the preferred embodiments described above. For this reason it is intended that this detailed description be regarded as an illustration and not as a limitation of the invention. It is only the following claims, including all equivalents, that are intended to define the scope of this invention.

Appendix A

Table of Contents

1 Objective:

A set of prototypes that investigated voice technology and structured reporting interfaces is reported on. Field research is also underway and the results noted. The implications for the Vista review station effort are detailed and recommendations made. Generalization to the ultrasound machines is made as well.

2 General Considerations Concerning Speech Recognition

2.1 Strategies for Incorporating Speech Recognition

2.1.1 Speaking Freely

One proposal for the interface is to allow the user to enter text by speaking freely (herein refereed to as Free-Dictation). When a user speaks freely, they often emit umms, clicks, exhalations and other noises. There is not software available that can handle these sounds and still achieve a high recognition accuracy. What is available is known as continuous and discrete speech recognition. Discrete speech requires the user to pause between words so that the software can distinguish when one word ends and another starts. Users need to learn to speak this way, and it is not natural. Typically users will stumble on words with many syllables, pausing inappropriately between syllables rather than between words -- it isn't natural. On the other hand, discrete recognition has a slightly higher recognition rate. Continuous speech recognition on the other hand does away with the pauses and is much more natural. But when it goes wrong, it's a mess.

All the products expect the user to experience recognition errors. This will always happen, no matter where the technology goes. Even in human-to-human speech, speakers engage in repetitions, clarifications, and simple "what did you say". Any interface will also have to provide a strong way to rapidly correct errors.

2.1.2 Correcting Errors

The current state of the technology is that after short time of training, most users will experience approximately 95% recognition. While this is impressive and sounds good, that means that 1 in 20 words will be wrong -- very wrong. The user can then go about a number of options depending on the product and implementation to fix the errors. They can do this by the mouse and keyboard, which we are trying to

avoid in the first place, or by speech commands, subject to further and compounding errors. In addition, the product makers all claim to have difficulties with people with accents and with foreign speakers.

2.1.3 Smoothness

The interface should be fluid and smooth. One should not have to switch modes between using the keyboard, the mouse, and a microphone. One should be able to navigate and use the application solely using any of these input modes. At the same time, moving from mouse control to speech and back should swift and fluid. The modes should interoperate. Speech is global (see below) while the mouse and keyboard are very precise and specific. Each will be more efficient for certain operations.

2.1.4 Command & Control of the Interface

Display of what commands.is being executed
Display of recognition accuracy
Display of recognition state (listening or not)
Confirmation of important commands before execution

2.1.5 Speech is Global

Voice is global, it is not pointed or visual. People have vast vocabularies and can use them to penetrate right to the heart of things, as compared to using a mouse to sift through piles of nested, overlapping menus of choices by clicking here and there on screen. Voice interfaces have the opportunity to jump right in and act, finding context from the speech. However, there are things that are best pointed at rather than described or spoken: For example when there is more than one of something, one must describe perhaps at length which one is being talked about, while one could more quickly point at the one you wanted. Selecting images or drawing outlines in an image are best done with the hands via the mouse, while selecting functions and modes can best be done with the voice. Switching from one mode of entry to the other and back should be smooth and fluid. (See Smoothness above).

2.1.6 Sound can be Peripheral

While one can focus on a few things, you can monitor many other things in the background. [Wieser] If you are driving and talking, you will nonetheless immediately notice if the engine should make a strange noise. This attention to many sounds could be utilized to monitor the completion of an activity, as if whispered in the background "The report on Lester Bowie has been sent". Users are more inclined to monitor than to go to some other window and scroll through a list of completed activities. Of course, they should be able to fall back on such a list.

Other sort of sound cues can be used to indicate that something has happened.

Of course there can be too much sound.

2.1.7 Recitation of state

The interface should be able to talk back to describe *system state*, and *study state*.
How many images? Is the study read? Who signed the study? The system could be interrogated, rather than approached and physically manipulated. The system would be 'easier' to use if it didn't demand that the user continually stare at it's ever-glowing screen to monitor state. It could announce if an error has occurred, without pushing an annoying message infront of the user. (see Sound can be Peripheral). The user could also enquire if there are any errors.

2.1.8 Listening All the Time

The computer will be listening all the time, even when the users words are not directed at it. With the current technology, the computer has no listening context – doesn't know when not to listen -- so it must be directed to not listen, or the microphone must easily be turned off.

2.2 The Physical Interface

Speech technology carries with it a set of input/output devices. We discuss the human factors characteristics and opportunities in what follows:

2.2.1 Input Devices: Headsets vs. Standing Mikes

Having a headphone on the head can be annoying and ruin your hair. This concern is reflected in user feedback from our field research. However they improve the efficacy of recognition significantly due to the proximity of the microphone to the mouth. Noise is highly reduced and it is noise that is a significant problem in current voice recognition technology [see literature]. Headsets have other conveniences as well, such as, one can move around without losing contact with the mike. Standing mikes should also be able to provide excellent sound, but possibly not as good as the user must always be near it, trapped by it. Lavaliercs allow freedom of movement but their quality is worse.

In addtion there are array microphones that can track a voice and eliminate surrounding sound. These might prove to be ideal for our applications.

2.2.2 Output Devices: Speakers vs. Headset

Headphones over the cars, especially for long periods of time, can be annoying. Yet forcing everyone to hear the sound over speakers can be problematic. In addition, interference between the speakers and the mike can make configuration difficult or impossible.

2.2.3 Combination Input / Output Devices

There are some newer physical devices that are promising. There are handheld microphone with a trackball and additional buttons plus a barcode scanner. Such a device would be very good for us, excepting that a wireless device can be easily removed, stolen, lost.

3 Field Research: Workflows, User Scenarios

We called and visited a number of sites using voice technology and forms of structured reporting. There two kinds of workflow: the flow of a report through an institution, and the flow of work on the review station itself.

3.1 Report Workflows

Report workflow can be divided into three *phases*:

1. **Report Generation**: once images and worksheets are obtained
2. **Report Correction & Initial Signing**: A resident's report is initial until approved by faculty
3. **Final Signing**: Once a faculty or senior physician has approved a report

These phases are combined into three kinds of *workflows*:

1. **Structured Workflow**
2. **Correctionist Workflow**
3. **Dictation Workflow**

These are diagrammed in the following figures:

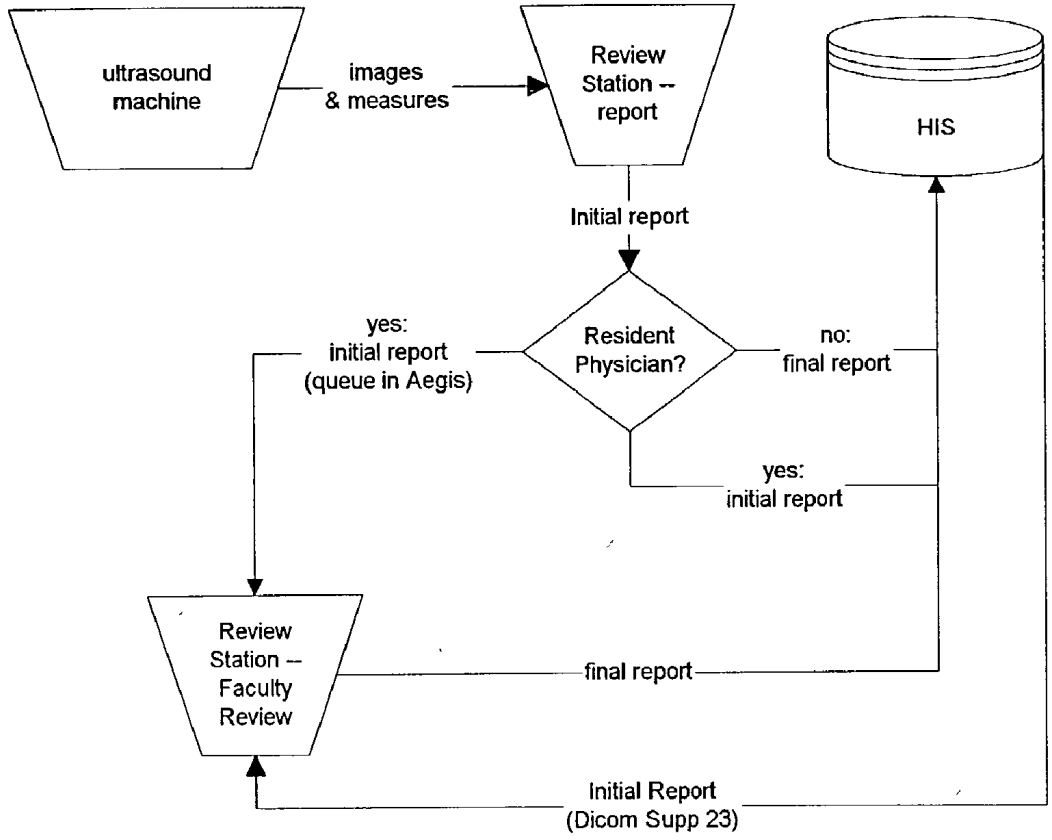


Figure 1: "Structured" Workflow.

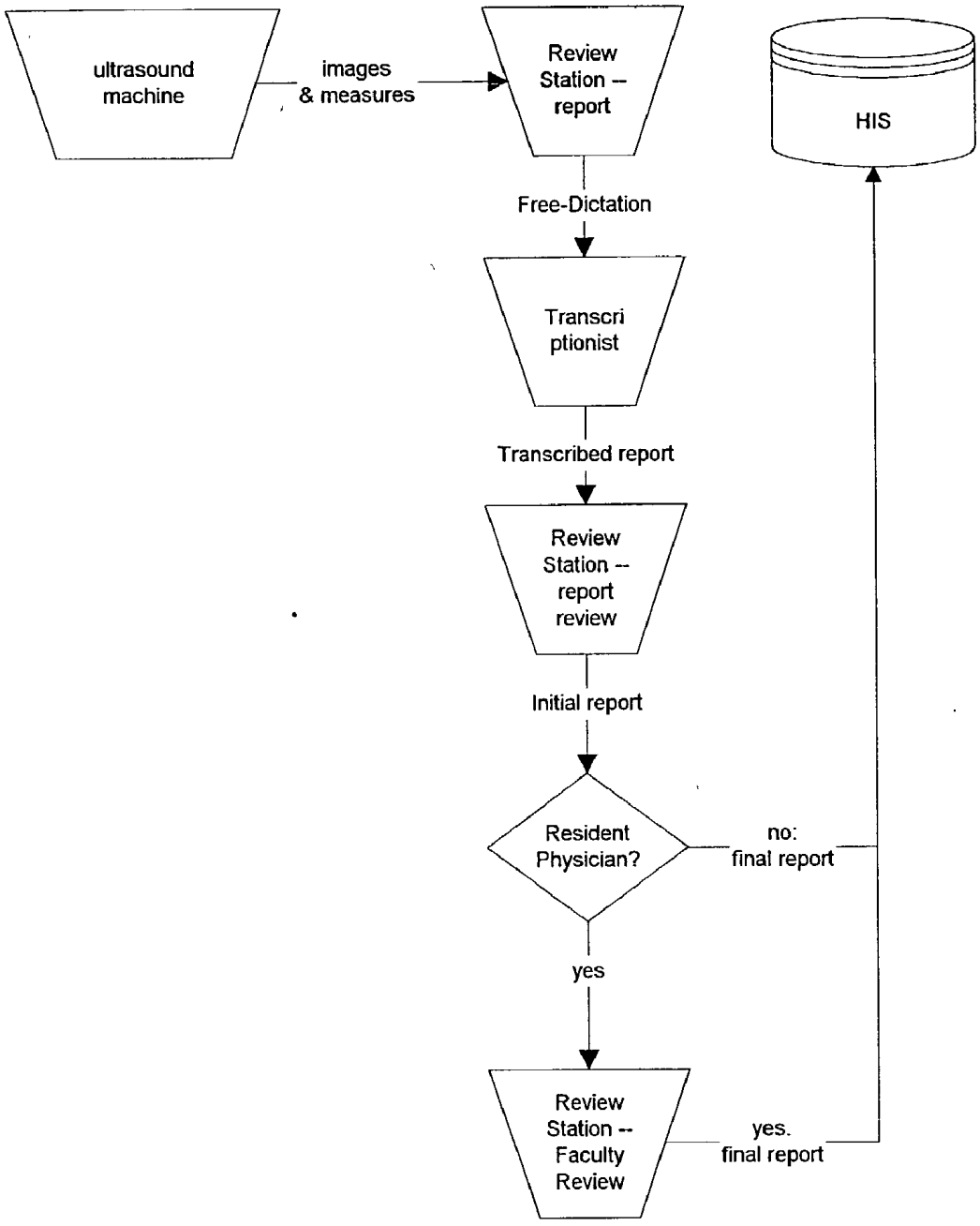


Figure 2: Current "Dictated" Report Workflow

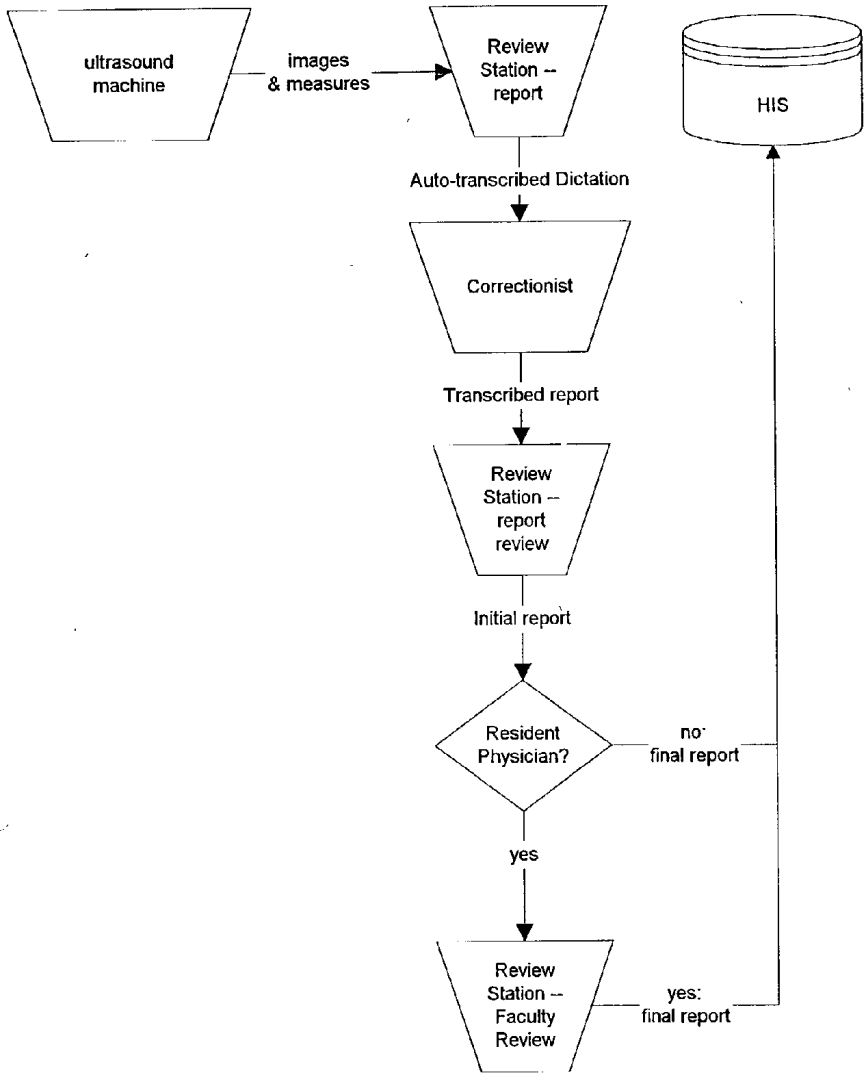


Figure 3: Alternative "Correctionist" Report Workflow

3.2 Discussion

Based on our field research, here are the scenarios and a measure of their efficacy in generating final reports:

Process	Average Time to Complete
Structured: Ultrasound Machine → Review Station to Read Study → Automated Dictation & Correction → Final Report to HIS.	<4hrs
Structured: Ultrasound Machine → Resident Review Station to Read Study → Automated Dictation & Correction → Initial Report to HIS → Faculty MD Final Correction → Final Report to HIS	<24 hrs

Correctionist: Ultrasound Machine → Review Station to Read Study → Automated Dictation → Correctionist → MD → Completed Report to HIS.	36hrs
Correctionist: Ultrasound Machine → Resident Review Station to Read Study → Automated Dictation → Correctionist → Resident MD → Initial Report to HIS → Faculty MD Final Correction → Final Report to HIS.	36hrs
Dictated: Ultrasound Machine → Review Station to Read Study → Transcriptionist → MD → Completed Report to HIS.	36hrs
Dictated: Ultrasound Machine → Resident Review Station to Read → Transcriptionist → Resident MD → Initial Report to HIS → Faculty MD Final Correction → Final Report to HIS.	36hrs

Note that in both the Dictated and Correctionist method, a resident releases an initial report which is later reviewed by the faculty doctor. Usually, this report is correct, so that further processing of the patient can take place. But sometimes, it is not correct. While this practice can be continued with the Structured method, it could be disallowed for a hospital going forward, since the reviewed, final report will be available within 24hrs on average.

Note that the Correctionist model benefits from the pre-transcription done by automated means. The use of this scenario comes when the users simply will not put up with automated dictation.

Clearly, the Structured model is far more efficient and the market will go to it sooner or later. The difficulty is that voice-dictation and structured reporting in general is not as easy as free-dictation.

Even when an institution jumps to this technology, they still will need to keep some facilities in place should the automated transcription facility fails. The cost will be time to patient and institution, on top of the backup costs.

Note that a Correctionist facility can be supplied as an additional system component i.e. another item to sell the customer. Alternatively, this can be left to the hospital.

Note that voice technology records the speech digitally. Thus it is available for human transcription if so desired.

3.3 Intra-Review Station Workflows

The preceding discussion covers the report workflow. But what about the flow of a study or a report on the review station?

3.3.1 Read-Report Styles

Field research has identified two working styles:

- ***Read-Report-Read Style***
- ***Read-Read-Report-Report Style***

Read-Report-Read style means that the physician reviews the images and worksheets and then generates a report all the way to completion. While Read-Read-Report-Report style means that the physician is reviewing images and worksheets and perhaps dictating a report but not correcting it, and then going on to the next study. After a number of studies have been partially completed, the physician then goes through the reports and corrects any dictation and factual errors as needed. Thus there is a "stack" of unread studies, a stack of read but uncorrected reports, and a stack of final reports for a user operating in this style.

We may want to change the interface to allow for the queing up of uncorrected or incomplete reports, and to support the searching of reports, rather than the idea of "read" and "unread" studies. This would facilitate either style.

3.3.2 Report States

Additionally, the field indicates that reports have states partially independent of studies. A report may be:

- ***Begun***
- ***Initial***
- ***Final***

Once a study is being read, a report might have been started but not yet completed: the reporting process has *begun*. Once the report is completed and corrected, it can be signed off into the *Initial* state or the *Final* state depending on whether the physician is junior or senior, resident or faculty, adjunct or partner, etc.

We may want to provide users with state information and allow operations accordingly. For example, selecting Initial reports for review and final approval.

4 Interface Recommendations

4.1 Requirements Driven From Speech Technology Limitations

Here follows requirements that are driven from limitations in the technology and interfaces.

4.1.1 User Education

We have found that users must learn to use the technology. If a user varies her speech a great deal, or does not train it enough, or the room is too loud, recognition results will be poor. But if the user works to get accuracy to a high enough level, then users can dictate at up to 160 words-per-minute – far faster than typing.

4.1.2 Improving Recognition Accuracy

Because speech technology is not yet up near 99.5% accuracy, a level that most people would find convenient, a number of requirements arise to improve the accuracy of recognition and the usefulness of this technology.

4.1.2.1 User Training Files Follow the User

In order to achieve higher recognition rates, users train the engines with their voice. This is done via explicit training and enrollment sessions, and also when correcting their dictations. This is not a 1 minute process, it can take months of use. Therefore, training files need to follow the user from review station to review station (or ultrasound machine to ultrasound machine).

4.1.2.2 Improve Recognition by Focusing the Vocabulary and Domain Model on Cardiology and Radiology

Continuous speech technology is just not ideal because:

1. It isn't natural (gets confused with umm, hmm, etc)
2. The recognition accuracy is too low, and will not work with all speakers.

What we can do is focus the vocabulary and the domain model very tightly on radiology and cardiology. And perhaps even focus on sub-specialties within these domains.

Note that this effort goes to supporting a free-speech interface.

4.1.2.3 Improve Recognition by Reducing the Number of Choices

In a prompted system, we can turn the speech recognition problem into a simpler problem of choosing from a short list of choices. This can be expected to have a much higher accuracy than allowing for free speech.

The effort here is to put together a workflow and an interface that would support selection from a short list of items. For example, as the user works her way down a hierarchy of report choices to fill out a particular form and field, the user is prompted with a small set of choices and is asked to select item 1 thorugh 10. Thus, she speaks only a set of number that might become an easy habit: 6 5 2 8 4 as a way of filling out an

uneventful ob report. The difficulty here is finding ways to reduce the very large vocabularies of medicine down to a few choices. I note that some researchers have gone the other way, seeking to find (visual) interfaces that can elegantly support choosing from a large set of choices [Poon 96]

Note that this effort goes to supporting prompted reporting.

4.1.2.4 Custom Language Models and Grammars

A language model that is tuned to a particular language domain aids voice recognition: for example, there are voice models for medical specialties such as radiology. These significantly improve recognition far more than simply adding words to an existing model. We should obtain a radiology language model.

In addition, a custom grammar can be applied to:

- help constrain the words to be spoken to a focused sub-domain,
- to aid the recognizer in looking for words,
- to spot certain key words out of a phrase.

We should supply these grammars as needed.

4.1.2.5 Inverse Text Normalization

The software should normalize dictated speech such that telephone numbers appear as such without having to say "5 5 5 dash 1 2 1 2", such that numbers have commas, such that spacing rules are applied with uniformity, etc. This feature is known as inverse text normalization.

4.2 Microphones & Headsets

Ideally, a combination device that can be secured to the workstation and that can be used as a mouse, as a microphone and speaker, and as a barcode reader would be advantageous.

4.2.1 Noise

The quieter it is, the better the results. Background noise will significantly affect performance of dictation And, it will cause commands to be issued randomly. The latter can be minimized with command confirmation.
For dictation, noise-canceling microphones are needed, preferably attached to a headset.

4.2.2 Bad Hair

Some users object to having their hair mussed up by a headset microphone. A convenient alternative are the all-in-one microphone/speaker/mouse/(barcode) controllers c.f. Phillips SpeechMike.

4.3 Sound Cards & Drivers

Full duplex audio is required. Many cards or card/driver combinations do not support this; SoundBlaster 64 does not, for example, while SoundBlaster 16 does.

4.4 Report Workflow Related Changes

Allow for the queuing up of uncorrected or incomplete reports
Support the searching of reports by state, in addition to categorizing studies as "read" and "unread". For example, selecting Initial reports for review and final approval.

4.5 Interface Changes

4.5.1 Logon via Speaker Recognition

Currently, the user logs on with a user name and password. This can be overridden by the user simply stating who they are; for example, stating "I am Dr. Greenberg" , the system can validate whether it is or isn't give access accordingly.

4.5.2 Command and Control Integrated into the Interface

4.5.2.1 Numbering Name Items

Items with names need to be numbered so that they can be recognized. It is unlikely that all names will be in the voice database. While some can be recognized, it will be more effective to number all such items so that they can be referred to. This especially pertains to names on the login page and in the study list and work list.

4.5.2.2 What Can I Say?

The display must let the user find out what they can say. In some sense, the screen indicates this by the text on the buttons, for example, but there are "hidden" menus that are not visible. Fundamentally, the user needs to be able to ask "what can I say". While repeated use will minimize this, it is still required.

4.5.2.3 Devices with Named Function

A knob, wheel, trackball, button etc function can be controlled by saying what it does. Saying the mouse controls a particular image's brightness and leveling, obviates the need to actually move the mouse onto that image, select the 'tools menu', then the 'brightness & leveling' sub-menu, and **then** manipulate the mouse to adjust brightness & leveling.
If such a device is attached to the transducer, then it would better facilitate one-hand operation and allow the user to focus or at least face the patient while doing their imaging. One could say "gain" to control gain, and so on.

4.5.2.4 Command Confirmation

1. Command confirmation is required when the action might be destructive. There is an issue of duplicating a mouse/keyboard/visual confirmation versus confirmation that can be done in the speech engine itself. That is, the speech api has a confirmation capability that makes it possible to prompt the user about what the engine heard, in advance of the command being attempted. An example of this is in deleting a study: the prompt could be done in the voice engine or in the command code. Currently, it is done in the Vista command code.

4.5.2.5 Audio and Recognition State

It must be able to report the state of the audio and recognizing engine. The following information needs to be displayed:

- No Microphone
- Speaking too quietly
- Speaking too loudly
- Too much noise
- Another application is using the audio
- The audio is turned off
- The engine is listening
- The command was unrecognized

4.5.3 Speaking Errors and Other State Changes

The system can speak as well as hear. The user can be notified of activity occurring in the background, or the user can ask a question of the system and receive a response. These can go on while the user is engaged in other verbal activity; i.e. the system is full-duplex audio.

4.5.3.1 Unprompted Spoken Notifications

These items will be spoken to the user without being prompted to do so:

- "When launching a study, "There are 57 images".
- "The system is short of disk space. Please do xyz"
- "The study of Margaret chan has started" (in progress)

- "The study of Margaret Chan is ready for review. (no longer in progress)
- "The images have been printed"
- "The printer is out of film"
- "The study has been sent to the PACs"

4.5.3.2 Spoken Question & Answer / Dialoging

The user can ask about the state of the system:

Question	Answer
"How many images?"	"57"
"What time is it?"	"5:55 PM"
"What is the date?"	"May 20, 1998"
"Who else is viewing this study?"	"Dr. Kienholz, Dr. Duchamp, Dr. Cage"

We should use available NLP technology here, and leverage keyword detection and finite-state-machine language descriptions to accomplish such dialoging as naturally as possible.

4.5.3.3 Spoken Command, Visual Response

Most of the commands are of this sort.

Command	Visual Response
"Show images"	Images are displayed
etc	Etc

4.5.4 Dictation

4.5.4.1 Verbal Correction of text

Lots of correction information & capability e.g. select this word, delete this line, delete up to 'parenchyma', next line, end-of-line, etc. <<detail this>>

4.5.4.2 Aiding Recognition

Showing alternative phrases (already handled by SAPI, but not neccesarily in the best way; i.e. it's a function of the recognition engine we use).
Training the engine by reading texts<<interface to enrollment & custom vocabulary>>
Training the engine by entering troublesome phrases <<interface for that>>
Adjusting recognition quality

4.5.4.3 Voice Macros

Macros are pre-built phrases are spoken phrases that cause a longer text to be generated. Abbreviations are one form of macros, another is saying "Normal OB" and having a bunch of text that is the boilerplate for a normal ob exam.
Macros can have substitution phrases in them so that a user can say "OB Normal 15" to give a normal OB report with HC of 15 for example.
The generated text can have database fields in them so that they can be populated on creation.

4.5.4.4 Macro Generation interface

These should be a part of the phrase generation customization screens.

5 Architectural Implications

5.1 Voice-Enabling Windows and Controls

To fully implement voice navigation in an application, users will want to be able to speak the name of a field in order to fill it out, rather than moving the focus by saying "next field" or "previous field". This is

true for any control: push buttons, radio buttons, etc. In addition, in Vista, when comparing studies, there are right and left side windows with the same controls and different data. To navigate between these, the user should be able to say which window they are referring to, and then subsequent commands should pertain to that window. In general, the user needs to name the item they are speaking about, and then speak the data that goes into that item. Furthermore, it is desirable to attach a grammar to a control or window so that recognition can be focused on the text that pertains to that control, control set, or window. An architectural solution to this needs to be found that is otherwise not obstrusive to the Vista architecture.

Such an architecture would

1. Sub-class all controls and windows such that they can be referred to by name
2. Each control is available in a window such that the window, given a spoken text string, can locate the control and set focus to it. Perhaps each control 'registers' itself with the window it is in.
3. Each window, as a collection of controls, can have a grammar associated with it that is enabled and disabled as the window is activated or deactivated.
4. Each control can have a grammar such that when the control has focus, the grammar is active.
5. Each control's grammar is available early in the window's life, or better still in advance of the window coming into existence, so that the grammars can be loaded, compiled and deactivated in advance of their being needed. (Compiling grammars and a time expensive operation.)

5.2 Separation of Voice Engines from Application

In general, the voice engine infrastructure should be isolated from the rest of the application. The interfaces to these components should require the application to supply whatever they need to operate. This is so that the application can be changed without the need to modify the voice portion -- it serves only as a voice recogntion, identification, and synthesis service. This can be accomplished if we rely on a standardized voice api such as Microsoft SAPI which promises the ability to plug & play voice engines into our applications without requiring application changes.

6 Appendix: Voice Commands in Vista

Depending on the screen or capability that is active, the following commands are enabled:

6.1 Global Commands - (Always active)

"Stop Listening"	- Turns Audio Off (must be turned on manually)
"Go To Sleep"	- pause listening
"Wake Up"	- start listening
"Attention"	- start listening
"Start Listening"	- start listening
"Logout"	
"I am Dr Greenberg"	- speaker recognition for login over capability?

6.2 Logon Commands

"I am Dr. Science"	- speaker recognition to logon
"12"	- each user name is displayed and numbered so that the number can be called out. Then password check performed. This will preclude pronunciation limitations with the voice technology especially for names that are foreign to the speaker.

6.3 Study List Commands

"Show Images"	- display selected study
"Show Report"	- display selected study report
"Show Worklist"	- display selected study report
"Show Report"	- display selected study report
"Select Item n"	- select study in list (numbered numerically)
"Select Study n"	- select study in list (numbered numerically)

6.4 Image Mode Commands

6.4.1 Image Selection

NOTE: Drawing, Windowing and Leveling are not done via voice. These are best left to the hands. Nonethese, it is desirable for user interface continuity that nearly all operations can be accomplished fluidly and interchangably with the hand and the voice.

"Select image n"	- n'th image is selected
"Select image x to y"	- contiguous range of images is selected
"Select x, y, z, 11 to 15, end"	- complex selection
"Select All"	- all images selected
"Select None"	- deselect all images
"Deselect All"	- deselect all images
"Deselect n"	- nth image is deselected
"Invert Selection"	- selected images are de-selected and vica-versa.

6.4.2 Actions on a Selection

"Delete Selected"	- selected images are deleted
"Mark as key"	- images become key images
:Save selected"	- selected images are saved
"Print selected"	- selected images are printed
"Copy select"	- selected images are copied to clipboard.

6.4.3 Image Manipulation

"Restore contrast settings"	- undo all <or selected?> window/leveling changes

6.4.4 Series Manipulation

"Delete Series"	- the imaging series is deleted
"Show Series"	- display list of series available
"Choose N"	- nth series in list is chosen (only if series is shown)

6.4.5 Worklist Commands

"Select Study n"	- select study in list (numbered numerically)
"Mark as read"	- study is marked as read.
"Unmark study"	- unmark the study

6.4.6 Navigation

"Show Worklist"	- drop down the worklist
"Show Report"	- show report
"Show Study List"	- switch to the study list
"Toggle Two Studies"	- turn on/off 'compare two studies' mode

6.4.7 Study Manipulation

Delete Study
Backup study
Last study
Route study
Merge study
Split study
Set bookmark
Goto bookmark
Mark as read

6.5 Compare Mode Commands

"Scroll Lock"	- both sides act as one
"Sync Lock"	- both sides act as one
"Left-side"	- subsequent commands go left-side
"Right-side"	- subsequent commands go right-side
"Other side"	- toggle sides

6.6 Report Mode Commands

"Report"	- Goes to the left report
"Left Report"	- Goes to the left report
"Right Report"	- Goes to the right report

6.6.1 Navigation

"Next page"
"Previous Page"
"Next Section"
"Previous Section"
"Closing Section" - goes to the closing section or whatever named section

6.6.2 Dictation

"Begin Dictation"	- begin free-text dictation
"Stop Dictation"	- stops dictation.

6.6.3 Correction: Playback

"Playback Selection"	speaker's recorded voice is used to read the text (and coressponding words highlighted.
"Playback Document"	
"Playback Section"	
"Stop Playback"	
"Read Selection"	synthetic voice is used to read the text
"Read Document"	
"Read Section"	
"Stop Reading"	

6.6.4 Correction: Text selection and editing

"Scratch that"
"Cross-out"
"Select ..."
"Go back n words"
"Go forward n words"
Etc....

6.6.5 Text

"Cut"
"Copy"
"Paste"

6.7 Worksheet Mode Commands

"Worksheet"	- Goes to the left worksheet
"Left Worksheet"	- Goes to the left worksheet
"Right Worksheet"	- Goes to the right worksheet

6.7.1 Navigation / Data Entry

"HC 15" - set the HC field to 15. I.e. jump to a field by name and give it focus.

6.7.2 Dictation

"Begin Dictation" - begin free-text dictation
"Stop Dictation" - stops dictation.

6.7.3 Text

"Cut"
"Copy"
"Paste"

7 Prototype

7.1 Voice Information

When the system loads, it obtains:

- the users voice recognition profile for the particular engine (voice training information)
- engine usage information (e.g. percent of CPU to dedicate to recognition)
- user adjustable recognition parameters (e.g. minimum volumes, timeouts to recognize complete and incomplete phrases...)
- list of microphones and their capabilities that the user trained with and that the profile is appropriate for (or this can be supplied so as obtain the proper profile) e.g. has echo cancellation or not, gain, description of bandwidth...

7.2 Voice User Interface (VUI)

Each screen that contains voice input needs to indicate this to the user.

A voice control should be present as shown below, containing

- Status of the recognizer and voice synthesis engines,
- the detected microphone volume to help the user speak at the proper level,
- the recognized command or dictation phrase.
- a button for manually turning on and off the recognizer or starting and stopping dictation (when it is enabled by the context of the application).

When the microphone is on and the application is listening the following is displayed:

If the user doesn't speak loud enough, or if the spoken command is not recognized, the display will indicate the problem on the lower line:

Here the user has spoken their name, which has been recognized, and the volume is being displayed to aid speaking properly:

7.3 Location of Voice Control

After the user has logged on and located an initial study to review, the user might say "Show Images": Note that the previously recognized command is shown in the voice control located in the upper right of the screen. On any window for voice is accepted, this control appears, and always in this location. The only exception is the logon screen, for which it is positioned elsewhere based on the design of that screen.

7.4 Report Types, Sections, and Section Phrases

One can bring up the report by uttering "Show Report". One can select alternative report types by saying the report type's name e.g. "O B Normal Report". Uttering the name of the section causes focus to go to that section and the list of macros that can be uttered to be displayed automatically (this popping up of the prompting menu can be turned off).

In the above, "Show Report" has been uttered, and the OB Normal report has been displayed in correspondence to the examination report type. Then "Diagnosis Section" has been uttered and the diagnosis macros are listed. In addition, the macro "Normal Diagnosis" has been spoken, and the corresponding phrase has been inserted into the text.

The sections visible in this OB Normal Report are:

- Patient information section
- Patient history section
- Gestational age section
- Fetal Weight Section
- Fetal Anatomy Section
- Diagnois Section
- Referral Section
- Closing section

Note that one can remove and add sections as well.

7.4.1 Equivalent Voice & Pointer Macro Menus

In addition, note that pointer can be used to select the phrases just as if they had been done by voice (This is in distinction to other systems in which one can say "What can I say" and be given a list of items of possible utterances but be unable to select them via the pointer.) The visual menu and the verbal menus are identical in this implementation.

7.4.2 Navigation by Section Name

Besides being able to navigate by pages as in, "Next page" or "Page 3", the user can jump to particular section in the report wherever it may be by uttering its name.

7.4.3 Inserting Images into Reports

In the following "Images section" has been uttered and verbal capability of putting images into the report is shown in the phrases menu. Saying "Insert image 10" or "Insert selected images".

7.5 Multiple Non-Overlapping Command Vocabularies

It is important to note that the vocabularies are such that macros and commands that manipulate reports are not confused with the vocabularies of measurements & calculation worksheets or with images manipulation, because all of which may be displayed together. For example, one can "page down" the report pages, and "scroll down" the images. One can jump to sections in the report by saying "Fetal anatomy section" and differentiate that from the worksheet with "Fetal anatomy worksheet".

In the preceding picture the images have been scrolled down, selected and inserted into the report. As each kind of ultrasound imaging output is displayed (i.e. images, worksheets, and reports), the corresponding vocabulary is enabled. This optimizes recognition.

8 Specific examples

Image Review Station Access to Functions

The user can utter particular commands to manipulate the medical imaging review station. There are two kinds of verbal manipulations controlled by the system:

- What kinds of medical information they are going to view: reports, worksheets, study lists, and images – perhaps all at once.
- What to do within the view.

These particular commands controlled by the system are discussed below:

1) **Bringing up and controlling multiple screen arenas by voice:**
   a) On a RS, multiple arenas of information are present at the same time. For example, images and reports can be presented together.
   b) Creating visual arenas.
      A set of verbal commands that uniquely identify each arena uniquely are created. For example:
      i) Show Images (images are presented)
      ii) Show Report (report and images are presented)
      iii) Show Worksheet (worksheet is added to screen)
      iv) Show Preferences (preferences windows are shown)
   c) Navigation within the arena
      A set of verbal command that move the focus to a particular area within the visual arena:
      i) Press OK (that particular button is pressed).
      ii) Next page, previous page (next/prev page in report presented)
      iii) Page n (numbered page in report is presented)
      iv) Next section, previous section (section in a report is presented)
      v) Patient history section, etc (named section in a report is presented)
      vi) Gestational Age Worksheet (named worksheet is presented)
      vii) Scroll down, scroll up (next page of images are presented)
      viii) Next row, previous row (next row of images are presented)
   d) Alternative commands
      The system allows for alternative phrases to be spoken for the given actions. The system prevents phrases from more than one function.
2) **Manipulating a particular medical function by voice:**
   Each particular arena has a set of capabilities which can be unambiguously controlled through voice commands:
   a) Images
      i) Zoom 5 (nth image is Zoomed or, if already zoomed, is returned to normal)
      ii) Select 14 (nth image is selected/deselected for further manipulation.
      iii) Played (nth image video is played)
      iv) Etc
   b) Selecting Studies for review
      i) Select by name
      ii) Ask for precanned searches e.g. "sort by date", or "studies left to read"
   c) Worksheets & Calculations
      Naming a particular equation to recalc that from:
      i) "Use Hadlock", etc
   d) Creating Reports and text
      A verbal shortcut can be uttered which triggers a longer phrase possibly containing references to data in worksheets and other areas of the exam. For example:

i) Uttering "Normal liver" would cause the following set of sentences to be inserted at the cursor: "There is no evidence for chlelithiasis or bilary dilitation. The visualized liver and right are unremarkable. The pancreas is normal."
ii) Editing is a combination of uttering voice macros, selecting macros by mouse and keyboard, and by directly typing and editing with the mouse and keyboard.
iii) The system determines when a particular section of a report has been entered and presents to the user a list of appropriate verbal shortcuts. Once the user is familiar with these shortcuts, she tells the system to stop presenting them.
(1) In addition, if the user says at any time, "What Can I say here" they are presented with the appropriate commands and shortcuts that pertain.

Addenda: List of Verbal Commands

Items with an ampersand (&) require confirmation by user.

```
// These can always be uttered:
        MENUITEM "What Can I Say",              ID_VCMD_WHATCANISAY
        MENUITEM "Go To Sleep",                 ID_VCMD_GOTOSLEEP
        MENUITEM "&Stop Listening",             ID_VCMD_STOPLISTENING
        MENUITEM "Which Engine",                ID_VCMD_WHICHENGINE
        MENUITEM "Control Engine",              ID_VCMD_CONTROLENGINE
        MENUITEM "Customize Speech",            ID_VCMD_CUSTOMIZESPEECH
        MENUITEM "Train Words",                 ID_VCMD_TRAINWORDS
        MENUITEM "Train Microphone",            ID_VCMD_TRAINMICROPHONE
        MENUITEM "Cancel",                      ID_VCMD_CANCEL
        MENUITEM "Ok",                          ID_VCMD_OK
        MENUITEM "Yes",                         ID_VCMD_YES
        MENUITEM "No",                          ID_VCMD_NO
        MENUITEM "Enter",                       ID_VCMD_ENTER
        MENUITEM "Next field",                  ID_VCMD_NEXTFIELD
        MENUITEM "Previous Field",              ID_VCMD_PREVFIELD

// Available when listening but asleep
        MENUITEM "Vista",                       ID_VCMD_VISTA
        MENUITEM "Wakeup",                      ID_VCMD_WAKEUP
        MENUITEM "&Stop Listening",             ID_VCMD_STOPLISTENING
        MENUITEM "Stop Prompting",              ID_VCMD_PROMPTMODE

// Available when logging on
MENUITEM "I am <name>",         ID_VCMD_LOGON
        MENUITEM "&Exit Vista",                 ID_VCMD_EXIT

// Available when images are visible on the screen
        MENUITEM "Show Report",                 ID_VCMD_SHOWREPORT
        MENUITEM "Show Worksheet",              ID_VCMD_SHOWWORKSHEET
        MENUITEM "Show Images",                 ID_VCMD_SHOWIMAGES
        MENUITEM "Previous row",                ID_VCMD_NEXTROWUP
        MENUITEM "Next row",                    ID_VCMD_NEXTROWDOWN
        MENUITEM "Scroll Up",                   ID_VCMD_SCROLLUP
        MENUITEM "Scroll Down",                 ID_VCMD_SCROLLDOWN
        MENUITEM "Show First Image",            ID_VCMD_SHOWFIRSTIMAGE
        MENUITEM "Show Last Image",             ID_VCMD_SHOWLASTIMAGES
        MENUITEM "Next Worksheet",              ID_VCMD_NEXTWORKSHEET
        MENUITEM "Previous Worksheet",          ID_VCMD_PREVWORKSHEET
        MENUITEM "Fetal Anatomy Worksheet",     ID_VCMD_WS_FETALANATOMY
        MENUITEM "Patient Info Worksheet",      ID_VCMD_WS_PATIENTINFO
        MENUITEM "Gestational Age Worksheet", ID_VCMD_WS_GESTATIONALAGE
```

```
        MENUITEM "Doppler Worksheet",              ID_VCMD_WS_DOPPLER
        MENUITEM "Measurements Worksheet",         ID_VCMD_WS_MEASUREMENTS
        MENUITEM "Pregnancy Worksheet",            ID_VCMD_WS_SACINFO
        MENUITEM "Growth Chart Worksheet",         ID_VCMD_WS_GROWTHCHART
        MENUITEM "Show Study List",                ID_VCMD_SHOWSTUDYLIST
        MENUITEM "Toggle Two Studies",           ID_VCMD_TOGGLETWOSTUDIES
        MENUITEM "Toggle Sync",                    ID_VCMD_SCOLLLOCK
        MENUITEM "Mark Study Read",                ID_VCMD_MARKSTUDY
        MENUITEM "Toggle Inbox",                   ID_VCMD_TOGGLEINBOX
        MENUITEM "&Logout",                        ID_VCMD_LOGOUT
        MENUITEM "&Exit Vista",                    ID_VCMD_EXIT
        MENUITEM "Zoom 5",                         ID_VCMD_ZOOM5
        MENUITEM "Zoom 10",                        ID_VCMD_ZOOM10
        MENUITEM "How many images",                ID_VCMD_HOWMANYIMAGES
        MENUITEM "How many studies left",       ID_VCMD_HOWMANYSTUDIESLEFT
        MENUITEM "Smith study ready",              ID_VCMD_ISSTUDYREADY

    // available when a report is visible
        MENUITEM "Begin Dictation",                ID_VCMD_BEGINDICTATION
        MENUITEM "Stop Dictation",                 ID_VCMD_STOPDICTATION
        MENUITEM "Next Section",                   ID_VCMD_SHOWNEXTSECTION
        MENUITEM "Previous Section",            ID_VCMD_SHOWPREVIOUSSECTION
        MENUITEM "Heading Section",                ID_VCMD_RPT_HEADING
        MENUITEM "Patient Info Section",           ID_VCMD_RPT_PATIENTINFO
        MENUITEM "Patient History Section",     ID_VCMD_RPT_PATIENTHISTORY
        MENUITEM "Gestational Age Section",     ID_VCMD_RPT_GESTATIONALAGE
        MENUITEM "Fetal Weight Section",           ID_VCMD_RPT_FETALWEIGHT
        MENUITEM "Fetal Anatomy Section",        ID_VCMD_RPT_FETALANATOMY
        MENUITEM "Fetal Doppler Section",        ID_VCMD_RPT_FETALDOPPLER
        MENUITEM "Diagnosis Section",              ID_VCMD_RPT_DIAGNOSIS
        MENUITEM "Amniotic Fluid Section",      ID_VCMD_RPT_AMNIOTICFLUID
        MENUITEM "Thank You Section",              ID_VCMD_RPT_THANKYOU
        MENUITEM "Closing Section",                ID_VCMD_RPT_CLOSING
        MENUITEM "Addendum Section",               ID_VCMD_RPT_ADDENDUM
        MENUITEM "Fetal Growth Section",           ID_VCMD_RPT_FETALGROWTH
        MENUITEM "Images Section",                 ID_VCMD_RPT_IMAGES
        MENUITEM "Study Summary Section",        ID_VCMD_RPT_STUDYSUMMARY
        MENUITEM "Indications Section",            ID_VCMD_RPT_INDICATIONS
        MENUITEM "Fetus Summary Section",        ID_VCMD_RPT_FETUSSUMMARY
        MENUITEM "Clinical Info Section",        ID_VCMD_RPT_CLINICALINFO
        MENUITEM "Next Page",                      ID_VCMD_SHOWNEXTPAGE
        MENUITEM "Previous Page",                ID_VCMD_SHOWPREVIOUSPAGE
        MENUITEM "Begin Dictation",                ID_VCMD_BEGINDICTATION
        MENUITEM "Stop Dictation",                 ID_VCMD_STOPDICTATION
        MENUITEM "baby pictures report",        ID_VCMD_REPORTBABYPICTURES
        MENUITEM "o b normal report",              ID_VCMD_REPORTOBNORMAL
        MENUITEM "general report",                 ID_VCMD_REPORTGENERAL

    // available when Worksheet visible
        MENUITEM "Begin Dictation",                ID_VCMD_BEGINDICTATION
        MENUITEM "Stop Dictation",                 ID_VCMD_STOPDICTATION
        MENUITEM "Next Worksheet",                 ID_VCMD_NEXTWORKSHEET
        MENUITEM "Previous Worksheet",             ID_VCMD_PREVWORKSHEET
        MENUITEM "Fetal Anatomy Worksheet",        ID_VCMD_WS_FETALANATOMY
        MENUITEM "Patient Info Worksheet",         ID_VCMD_WS_PATIENTINFO
        MENUITEM "Gestational Age Worksheet", ID_VCMD_WS_GESTATIONALAGE
        MENUITEM "Doppler Worksheet",              ID_VCMD_WS_DOPPLER
```

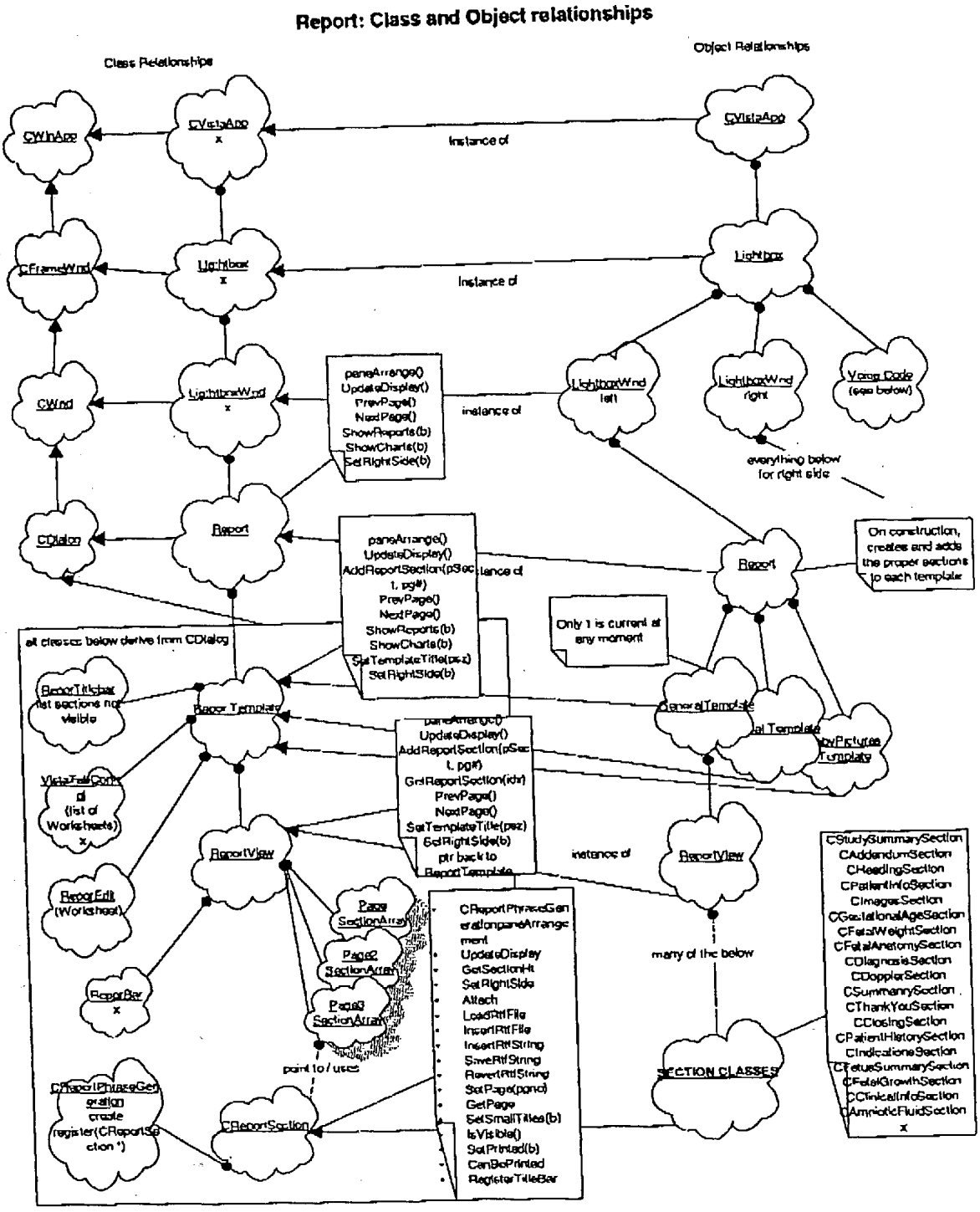
```
        MENUITEM "Study Review Worksheet",      ID_VCMD_WS_MEASUREMENTS
        MENUITEM "Pregnancy Worksheet",         ID_VCMD_WS_SACINFO
        MENUITEM "Growth Chart Worksheet",      ID_VCMD_WS_GROWTHCHART

// Phrase triggers that insert text into reports
        MENUITEM "emergency broadcast",         ID_VPHRASE
        MENUITEM "demonstration text",          ID_VPHRASE
        MENUITEM "signature",                   ID_VPHRASE
        MENUITEM "normal growth",               ID_VPHRASE
        MENUITEM "decreased growth",            ID_VPHRASE
        MENUITEM "retarded growth",             ID_VPHRASE
        MENUITEM "normal anatomy",              ID_VPHRASE
        MENUITEM "abnormal anatomy",            ID_VPHRASE
        MENUITEM "long neck",                   ID_VPHRASE
        MENUITEM "normal visualization",        ID_VPHRASE
        MENUITEM "choroid plexus",              ID_VPHRASE
        MENUITEM "strawberry",                  ID_VPHRASE
        MENUITEM "normal amniotic fluid",       ID_VPHRASE
        MENUITEM "no previa",                   ID_VPHRASE
        MENUITEM "normal left lateral",         ID_VPHRASE
        MENUITEM "normal anterior",             ID_VPHRASE
        MENUITEM "normal diagnosis",            ID_VPHRASE
        MENUITEM "normal abdomen",              ID_VPHRASE
        MENUITEM "normal aorta",                ID_VPHRASE
        MENUITEM "normal gallbladder",          ID_VPHRASE
        MENUITEM "normal liver",                ID_VPHRASE
        MENUITEM "normal no prior",             ID_VPHRASE
        MENUITEM "normal kidneys",              ID_VPHRASE
        MENUITEM "normal scrotum",              ID_VPHRASE
        MENUITEM "normal carotid",              ID_VPHRASE
        MENUITEM "mild plaque",                 ID_VPHRASE
        MENUITEM "normal unilateral",           ID_VPHRASE
        MENUITEM "normal bilateral",            ID_VPHRASE
        MENUITEM "normal o b",                  ID_VPHRASE
        MENUITEM "normal pelvis",               ID_VPHRASE
        MENUITEM "normal general diagnosis",    ID_VPHRASE
```

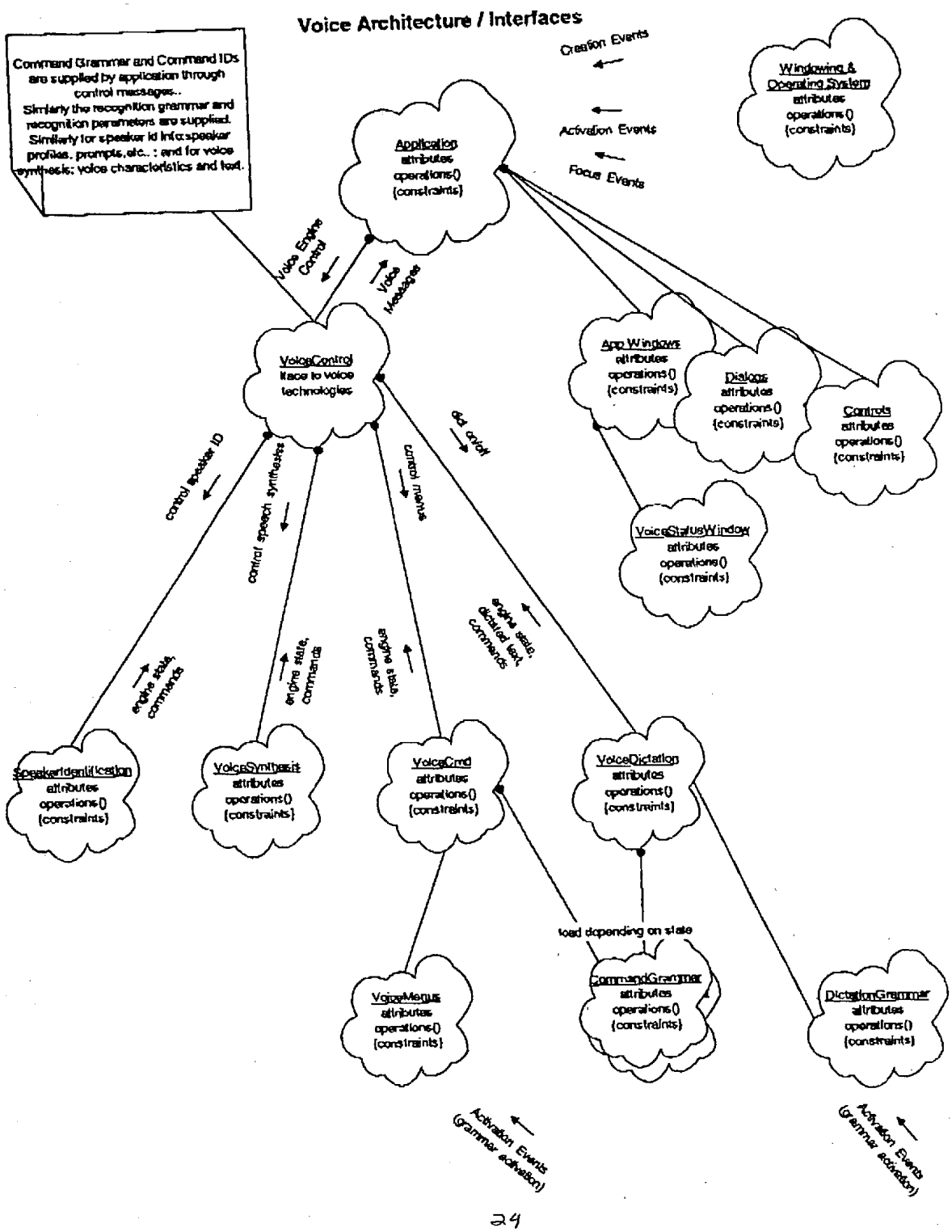
The following three diagrams are provided for further description. "Vista" refers to an implementation of the imaging review workstation prototype. "Lightbox" and so on refer to classes contained in it. "Application" is meant rather generically and can refer to the Ultrasound machine (command & control, and reporting infrastructure perhaps more specifically) or the imaging review station.
Classes beginning with the letter "C", tend to refer to classes made available under Microsoft Windows MFC version 4.2. The remaining classes with "C"-prepended are part of our work: CVistaApp, CReport...

The diagram entitled "Report: Class and Object Relationships" shows how the reporting hierarchy and infrastructure sits within the application.. The CreportPhraseGeneration classes are the custom phrases that apply to each section. These are triggered to be inserted into the section by using the mouse or the voice by naming them.

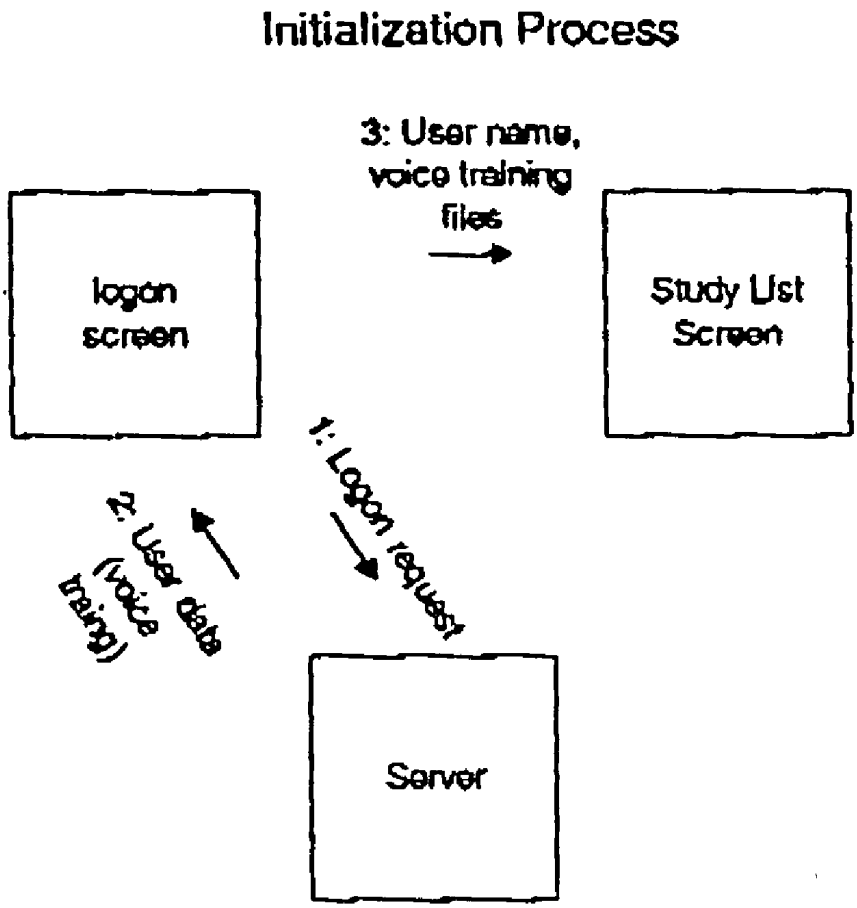
Activation and focus events in the diagram entitled "Voice Architecture/Interfaces" are coming from messages from the GUI/Operating system. The means for doing this on the ultrasound machine when the voice infrastructure is running on an external processor, is through a simple protocol ontop of the standard communications techniques: i.e. tcp/ip over ethernet, serial port, parallel port, etc.

As the diagram entitled "Initialization Process" shows, once the application begins, it initializes its VoiceControl object with the grammars and commands that are specific to it. A user can begin the validation process (perhaps doing so using voice commands). Once the user has been validated by a security means, user specific voice files are obtained from the network server. These are used to facilitate recognition in a manner suitable to the recognition engine contained by the VoiceControl, and user preferences such as the style of voice synthesis preferred (e.g. male or female, slow or fast speech, etc.), and custom commands and recognition shortcuts/macros/abbreviations.


Report: Class and Object relationships
Class Relationships
Object Relationships
CWinApp
CVistaApp
x
Instance of
CVistaApp
CFrameWnd
Lightbox
x
Instance of
Lightbox
CWnd
LightboxWnd
x
paneArrange()
UpdateDisplay()
PrevPage()
NextPage()
ShowReports(b)
ShowCharts(b)
SetRightSide(b)
instance of
LightboxWnd
left
LightboxWnd
right
Varag Code
(see below)
everything below
for right side
CDialog
Report
paneArrange()
UpdateDisplay()
AddReportSection(pSec
t, pg#)
PrevPage()
NextPage()
ShowReports(b)
ShowCharts(b)
SetTemplateTitle(psz)
SetRightSide(b)
Report
On construction,
creates and adds
the proper sections
to each template
all classes below derive from CDialog
Only 1 is current at
any moment
ReportTitlebar
(list sections not
visible
Report Template
paneArrange()
UpdateDisplay()
AddReportSection(pSec
t, pg#)
GetReportSection(idx)
PrevPage()
NextPage()
SetTemplateTitle(psz)
SetRightSide(b)
ptr back to
ReportTemplate
GeneralTemplate
Template
PicturesTemplate
VistaFieldCont
rol
(list of
Worksheets)
x
ReportView
instance of
ReportView
ReportEdit
(Worksheet)
Page
SectionArray
Page2
SectionArray
Page3
SectionArray
CReportPhraseGen
erationpaneArrange
ment
UpdateDisplay
GetSectionHt
SetRightSide
Attach
LoadRtfFile
InsertRtfFile
InsertRtfString
SaveRtfString
RevertRtfString
SetPage(pane)
GetPage
SetSmallTitles(b)
IsVisible()
SetPrinted(b)
CanBePrinted
RegisterTitleBar
many of the below
ReportBar
x
point to / uses
CReportPhraseGen
eration
create
register(CReportSe
ction *)
CReportSection
SECTION CLASSES
CStudySummarySection
CAddendumSection
CHeadingSection
CPatientInfoSection
CImagesSection
CGestationalAgeSection
CFetalWeightSection
CFetalAnatomySection
CDiagnosisSection
CDopplerSection
CSummarySection
CThankYouSection
CClosingSection
CPatientHistorySection
CIndicationsSection
CFetusSummarySection
CFetalGrowthSection
CClinicalInfoSection
CAmnioticFluidSection
x


Voice Architecture / Interfaces
Command Grammar and Command IDs are supplied by application through control messages..
Similarly the recognition grammar and recognition parameters are supplied.
Similarly for speaker id info:speaker profiles, prompts,etc.. : and for voice synthesis: voice characteristics and text.
Application
attributes
operations()
{constraints}
Creation Events
Activation Events
Focus Events
Windowing & Operating System
attributes
operations()
{constraints}
Voice Engine Control
Voice Messages
VoiceControl
Ilace to voice technologies
App Windows
attributes
operations()
{constraints}
Dialogs
attributes
operations()
{constraints}
Controls
attributes
operations()
{constraints}
VoiceStatusWindow
attributes
operations()
{constraints}
control speaker ID
control speech synthesis
control menus
dictated text
engine state, commands
SpeakerIdentification
attributes
operations()
{constraints}
VoiceSynthesis
attributes
operations()
{constraints}
VoiceCmd
attributes
operations()
{constraints}
VoiceDictation
attributes
operations()
{constraints}
load depending on state
VoiceMenus
attributes
operations()
{constraints}
CommandGrammar
attributes
operations()
{constraints}
DictationGrammar
attributes
operations()
{constraints}
Activation Events (Grammar activation)
Activation Events (Grammar activation)

24


Initialization Process
3: User name, voice training files
logon screen
Study List Screen
1: Logon request
2: User data (voice training)
Server

What is claimed is:

1. An ultrasound review station comprising:

a voice input device;

a processor; and

a voice recognition unit coupled with the voice input device and processor, the voice recognition unit being operative to convert a voice signal received from the voice input device into a command signal recognizable by the processor.

2. The ultrasound review station of claim 1, wherein the command signal instructs the processor to display an ultrasound image.

3. The ultrasound review station of claim 1, wherein the command signal instructs the processor to display an ultrasound examination report.

4. The ultrasound review station of claim 1, wherein the command signal instructs the processor to insert a textual phrase into an ultrasound examination report.

5. The ultrasound review station of claim 1, wherein the command signal instructs the processor to insert an ultrasound image into an ultrasound examination report.

6. The ultrasound review station of claim 1, wherein the command signal instructs the processor to perform an editing operation within an ultrasound examination report.

7. The ultrasound review station of claim 1, wherein the command signal instructs the processor to navigate through displayed information.

8. The ultrasound review station of claim 1, wherein the command signal instructs the processor to display a marker on an ultrasound image.

9. The ultrasound review station of claim 1 further comprising a display device coupled with the processor, and wherein the command signal instructs the processor to display available voice commands on the display device.

10. The ultrasound review station of claim 1, wherein the command signal selects a focus.

11. The ultrasound review station of claim 1 further comprising a plurality of display devices coupled with the processor, and wherein the command signal instructs the processor to select one of said plurality of display devices.

12. The ultrasound review station of claim 1 further comprising:

a voice output device; and

a voice production unit coupled with the processor and the voice output device, the voice production unit being operative to convert a voice output signal from the processor into a voice reproducible by the voice output device.

13. The ultrasound review station of claim 12, wherein the voice recognition unit and the voice production unit are integrated.

14. An ultrasound review station comprising:

a voice output device;

a processor; and

a voice production unit coupled with the processor and the voice output device, the voice production unit being operative to convert a voice output signal from the processor into a voice reproducible by the voice output device.

15. The ultrasound review station of claim 14, wherein the voice reproducible by the voice output device asks the user to confirm a command.

16. The ultrasound review station of claim 14, wherein the voice reproducible by the voice output device provides the user with a verbal alert.

17. The ultrasound review station of claim 14, wherein the voice reproducible by the voice output device provides an answer to a query presented to the processor by a user.

18. The ultrasound review station of claim 14, wherein the voice reproducible by the voice output device provides the user a status of a voice recognition unit.

19. The ultrasound review station of claim 14, wherein the voice reproducible by the voice output device provides an indication of a completion of an activity.

20. The ultrasound review station of claim 14 further comprising:

a voice input device; and

a voice recognition unit coupled with the voice input device and processor, the voice recognition unit being operative to convert a voice signal received from the voice input device into a command signal recognizable by the processor.

21. The ultrasound review station of claim 20, wherein the voice recognition unit and the voice production unit are integrated.

22. An ultrasound imaging system comprising;

a processor;

a voice output device; and

a voice production unit coupled with the processor and the voice output device, the voice production unit being operative to convert a voice output signal from the processor into a voice reproducible by the voice output device.

23. The ultrasound imaging system of claim 22, wherein the voice reproducible by the voice output device asks the user to confirm a command.

24. The ultrasound imaging system of claim 22, wherein the voice reproducible by the voice output device provides the user with a verbal alert.

25. The ultrasound imaging system of claim 22, wherein the voice reproducible by the voice output device provides an answer to a query presented to the processor by a user.

26. The ultrasound imaging system of claim 22, wherein the voice reproducible by the voice output device provides the user with a status of a voice recognition unit.

27. The ultrasound imaging system of claim 22, wherein the voice reproducible by the voice output device provides an indication of a completion of an activity.

28. The ultrasound imaging system of claim 22 further comprising:

a voice input device; and

a voice recognition unit coupled with the voice input device and processor, the voice recognition unit being operative to convert a voice signal received from the voice input device into a command signal recognizable by the processor.

29. The ultrasound imaging system of claim 28, wherein the voice recognition unit and the voice production unit are integrated.

30. A method for converting a received voiced command into a command signal recognizable by an ultrasound imaging system processor, the method comprising:

(a) with an ultrasound imaging system processor, providing a voice recognition unit with voice information correlating a set of command signals recognizable by the ultrasound imaging system processor with an associated set of respective voiced commands;

(b) with the voice recognition unit, receiving a voiced command; and

(c) converting the received voiced command into a command signal using the voice information provided by the ultrasound imaging system processor.

31. The method of claim 30 further comprising:

(d) with the voice recognition unit, providing the command signal to the ultrasound imaging system processor.

32. A method for converting a received voiced command into a command signal recognizable by an ultrasound review station processor, the method comprising:

(a) with an ultrasound review station processor, providing a voice recognition unit with voice information correlating a set of command signals recognizable by the ultrasound review station processor with an associated set of respective voiced commands;

(b) with the voice recognition unit, receiving a voiced command; and

(c) converting the received voiced command into a command signal using the voice information provided by the ultrasound review station processor.

33. The method of claim 32 further comprising:

(d) with the voice recognition unit, providing the command signal to the ultrasound review station processor.

34. A method for providing an ultrasound imaging system with voice information, the method comprising:

(a) identifying a user of an ultrasound imaging system;

(b) providing a server coupled with the ultrasound imaging system with the identity of the user; and

(c) with the server, providing the ultrasound imaging system with voice information associated with the user.

35. The method of claim 34, wherein (b) comprises providing a server in an ultrasound network coupled with the ultrasound imaging system with the identity of the user.

36. The method of claim 34, wherein (b) comprises providing a server integral with the ultrasound imaging system with the identity of the user.

37. The method of claim 34, wherein (c) comprises with the server, providing the ultrasound imaging system with voice information associated with the user, said voice information correlating a set of command signals recognizable by the ultrasound imaging system processor with an associated set of respective voiced commands.

38. The method of claim 34, wherein (c) comprises with the server, providing the ultrasound imaging system with voice information associated with the user, said voice information correlating a set of voice output signals transmitted by the ultrasound imaging system processor with an associated set of voice phrases reproducible by a voice output device of the ultrasound imaging system.

39. The method of claim 34 further comprising (d) providing the server with a change to the voice information.

40. The method of claim 34, wherein (a) comprises:

(a1) receiving identification information from the user; and

(a2) identifying the user based on the received identification information.

41. The method of claim 40, wherein (a1) comprises receiving verbal identification information from the user.

42. The method of claim 41, wherein (a1) comprises receiving verbal identification information from the user in response to an identification request from the ultrasound imaging system.

43. The method of claim 40, wherein (a1) comprises receiving non-verbal identification information from the user.

44. A method for providing an ultrasound review station with voice information, the method comprising:

(a) identifying a user of an ultrasound review station;

(b) providing a server coupled with the ultrasound review station with the identity of the user; and

(c) with the server, providing the ultrasound review station with voice information associated with the user.

45. The method of claim 44, wherein (b) comprises providing a server in an ultrasound network coupled with the ultrasound review station with the identity of the user.

46. The method of claim 44, wherein (b) comprises providing a server integral with the ultrasound review station with the identity of the user.

47. The method of claim 44, wherein (c) comprises with the server, providing the ultrasound review station with voice information associated with the user, said voice information correlating a set of command signals recognizable by the ultrasound review station processor with an associated set of respective voiced commands.

48. The method of claim 44, wherein (c) comprises with the server, providing the ultrasound review station with voice information associated with the user, said voice information correlating a set of voice output signals transmitted by the ultrasound review station processor with an associated set of voice phrases reproducible by a voice output device of the ultrasound review station.

49. The method of claim 44 further comprising (d) providing the server with a correction to the voice information.

50. The method of claim 44, wherein (a) comprises:

(a1) receiving identification information from the user; and

(a2) identifying the user based on the received identification information.

51. The method of claim 50, wherein (a1) comprises receiving verbal identification information from the user.

52. The method of claim 51, wherein (a1) comprises receiving verbal identification information from the user in response to an identification request from the ultrasound review station.

53. The method of claim 50, wherein (a1) comprises receiving non-verbal identification information from the user.

54. A method for using voice commands to insert a textual phrase into a section of an ultrasound examination report, the method comprising:

(a) receiving a voiced command from a voice input device of an ultrasound imaging system to insert a textual phrase into a section of an ultrasound examination report;

(b) identifying a set of textual phrases associated with the section;

(c) converting the received voiced command into a textual phrase selected from the set of textual phrases identified in (b); and

(d) inserting the textual phrase selected in (c) into the section.

55. The method of claim 54 further comprising receiving a voiced request to create an ultrasound examination report, the ultrasound examination report comprising a plurality of sections associated with a respective set of textual phrases.

56. The method of claim 54 further comprising automatically displaying a set of available voice commands associated with the section.

57. The method of claim 54 further comprising:

(e) receiving a voiced request for a display of a set of available voice commands associated with the section; and

(f) displaying the set of available voice commands associated with the section in response to the voiced request received in (e).

58. The method of claim 54, wherein (a) comprises receiving a voiced command to insert a textual phrase into a section of an ultrasound examination report displayed on an ultrasound imaging system.

59. The method of claim 54, wherein (a) comprises receiving a voiced command to insert a textual phrase into a section of an ultrasound examination report displayed on an ultrasound review station.

60. The method of claim 54 further comprising:

(e) receiving a voiced request to insert an image into a section of the ultrasound examination report; and

(f) inserting the image into the ultrasound examination report in response to the voiced request received in (e).

61. The method of claim 54 further comprising:

(e) receiving a voiced request to edit a section of the ultrasound examination report; and

(f) editing the section of the ultrasound examination report in response to the voiced request received in (e).

62. A method for using voice commands to insert a textual phrase into a section of an ultrasound examination report, the method comprising:

(a) receiving a voiced command from a voice input device of an ultrasound review station to insert a textual phrase into a section of an ultrasound examination report;

(b) identifying a set of textual phrases associated with the section;

(c) converting the received voiced command into a textual phrase selected from the set of textual phrases identified in (b); and

(d) inserting the textual phrase selected in (c) into the section.

63. The method of claim 62 further comprising receiving a voiced request to create an ultrasound examination report, the ultrasound examination report comprising a plurality of sections associated with a respective set of textual phrases.

64. The method of claim 62 further comprising automatically displaying a set of available voice commands associated with the section.

65. The method of claim 62 further comprising:

(e) receiving a voiced request for a display of a set of available voice commands associated with the section; and

(f) displaying the set of available voice commands associated with the section in response to the voiced request received in (e).

66. The method of claim 62, wherein (a) comprises receiving a voiced command to insert a textual phrase into a section of an ultrasound examination report displayed on an ultrasound imaging system.

67. The method of claim 62, wherein (a) comprises receiving a voiced command to insert a textual phrase into a section of an ultrasound examination report displayed on an ultrasound review station.

68. The method of claim 62 further comprising:

(e) receiving a voiced request to insert an image into a section of the ultrasound examination report; and

(f) inserting the image into the ultrasound examination report in response to the voiced request received in (e).

69. The method of claim 62 further comprising:

(e) receiving a voiced request to edit a section of the ultrasound examination report; and

(f) editing the section of the ultrasound examination report in response to the voiced request received in (e).

70. A method for using voice commands to display a marker on an ultrasound image, the method comprising:

(a) receiving a voiced command from a user;

(b) converting the received voiced command into a marker displayable on a display device; and

(c) displaying the marker on an ultrasound image displayed on the display device.

71. The method of claim 70, wherein (b) comprises converting the received voiced command into at least one word, and wherein (c) comprises displaying said at least one word on an ultrasound image displayed on the display device.

72. The method of claim 71, wherein (b) comprises converting the received voiced command into at least one anatomical-identifying word, and wherein (c) comprises displaying said at least one anatomical-identifying word on an ultrasound image displayed on the display device.

73. The method of claim 70, wherein (b) comprises converting the received voiced command into a geometric shape, and wherein (c) comprises displaying the geometric shape on an ultrasound image displayed on the display device.

74. The method of claim 70, wherein an ultrasound image is associated with one of a plurality of study types, wherein each of said plurality of study types is associated with a respective set of markers, and wherein (b) comprises:

(b1) identifying a study type of the ultrasound image;

(b1) identifying a set of markers associated with the study type; and

(b2) converting the received voiced command into a marker selected from the set of markers identified in (b2).

75. The method of claim 70 further comprising receiving a voiced command from a user indicating that a subsequent voiced command will be associated with a marker, said receiving occurring before (a).

76. The method of claim 70, wherein (c) comprises displaying the marker on the ultrasound image at a position indicated by a pointer displayed on the ultrasound image.

77. The method of claim 70, wherein (c) comprises selecting a location on the ultrasound image to position the marker.

78. The method of claim 70, wherein (b) comprises converting the received voiced command into a plurality of markers, and wherein (c) comprises selecting a plurality of locations on the ultrasound image to respectively position each of the plurality of markers.

79. The method of claim 70 further comprising:

(d) positioning a pointer over a marker displayed on the ultrasound image;

(e) receiving a verbal delete command from a user; and

(f) removing the marker from the ultrasound image in response to the received verbal delete command.

80. The method of claim 70 further comprising:

(d) positioning a pointer displayed on the ultrasound image over the marker;

(e) receiving a verbal copy command from the user;

(f) creating a copy of the marker in response to the received verbal copy command; and

(g) positioning the copy of the marker on another location of the ultrasound image.

81. A method for providing voice data from a first to a second ultrasound device, the method comprising:

(a) receiving voice data with a first ultrasound device;

(b) providing a second ultrasound device with the voice data received in (a); and

(c) with the second ultrasound device, converting the voice data into a command signal recognizable by a processor of the second ultrasound device, the first and second ultrasound devices being devices selected from the group of ultrasound devices consisting of an ultrasound imaging system and an ultrasound review station.

82. The method of claim 81, wherein (b) comprises:

(b1) transmitting the voice data received in (a) to a server coupled with the first and second ultrasound devices; and

(b2) with the server, transmitting the voice data received in (b1) to the second ultrasound device.

83. The method of claim 81, wherein (b) comprises:

(b1) storing the voice data received in (a) on a portable storage device; and

(b2) providing the portable storage device to the second ultrasound device.

84. The method of claim 81, wherein (b) comprises transmitting the voice data received in (a) to the second ultrasound device via a communication link directly coupling the first and second ultrasound devices.

85. A method for providing a command signal from a first to a second ultrasound device, the method comprising:

(a) receiving voice data with a first ultrasound device;

(b) transmitting the voice data received in (a) to a server coupled with the first ultrasound device; and

(c) with the server, converting the voice data transmitted in (b) to a command signal recognizable by a processor of a second ultrasound device; and

(d) with the server, providing the command signal to the second ultrasound device, the first and second ultrasound devices being devices selected from the group of ultrasound devices consisting of an ultrasound imaging system and an ultrasound review station.

86. The method of claim 85 further comprising:

(e) with the server, providing the command signal to the third ultrasound device, the third ultrasound device being a device selected from the group of ultrasound devices consisting of an ultrasound imaging system and an ultrasound review station.

87. The method of claim 85, wherein the first ultrasound device comprises the second ultrasound device.

88. An ultrasound transducer comprising:

at least one transducer element; and

a microphone;

said at least one transducer element and said microphone providing electrical signals to an ultrasound imaging system coupled with the ultrasound transducer.

89. The ultrasound transducer of claim 88, wherein the microphone is built-into the ultrasound transducer.

90. The ultrasound transducer of claim 88, wherein the microphone is removeably coupled with the ultrasound transducer.

91. A method for using voice commands to assign a function to an ultrasound imaging system or ultrasound review station user interface, the method comprising:

(a) receiving a voiced command from a user;

(b) converting the received voiced command into a function;

(c) assigning the function to a user interface device; and

(d) performing the function in response to a signal from the user interface device.

92. The method of claim 91, wherein (c) comprises assigning the function to a user interface device, the user interface device being depressible.

93. The method of claim 91, wherein (c) comprises assigning the function to a user interface device, the user interface device being movable.

94. The method of claim 91, wherein (c) comprises assigning the function to a user interface device, the user interface device being depressible and movable.

95. The method of claim 91, wherein (c) comprises assigning the function to a user interface device on an

ultrasound imaging system, and wherein (d) comprises performing the function in response to a signal from the user interface device on the ultrasound imaging system.

96. The method of claim 91, wherein (c) comprises assigning the function to a user interface device built into an ultrasound transducer, and wherein (d) comprises performing the function in response to a signal from the user interface device built into the ultrasound transducer.

97. The method of claim 91, wherein (c) comprises assigning the function to a user interface device removably coupled with an ultrasound transducer, and wherein (d) comprises performing the function in response to a signal from the user interface device removably coupled with the ultrasound transducer.

98. The method of claim 91, wherein (c) comprises assigning the function to a user interface device on an ultrasound review station, and wherein (d) comprises performing the function in response to a signal from the user interface device on the ultrasound review station.

99. The method of claim 91, wherein (c) comprises assigning the function to a user interface device selected from a group of user interface devices consisting of a wheel, a ball, a knob, a button, and a slider.

100. A method for inserting text into an ultrasound examination report on an ultrasound imaging system, the method comprising:

(a) receiving a voice signal from a voice input device of an ultrasound imaging system;

(b) with the ultrasound imaging system, converting the voice signal into text; and

(c) with the ultrasound imaging system, inserting the text into an ultrasound examination report.

101. The method of claim 100, wherein (b) comprises converting the voice signal into text using a voice recognition unit of the ultrasound imaging system.

102. An ultrasound imaging system comprising:

a voice input device; and

a voice recognition unit coupled with the voice input device, the voice recognition unit being operative to convert a voice signal received from the voice input device into text.

103. The invention of claim 102, wherein the voice recognition unit is further operative to provide the text to an ultrasound examination report.

104. The invention of claim 102 further comprising a processor coupled with the voice input device, wherein the voice recognition unit is implemented with the processor.

105. The invention of claim 102 further comprising a processor coupled with the voice input device and the voice recognition unit.

106. A method for inserting text into an ultrasound examination report on an ultrasound review station, the method comprising:

(a) receiving a voice signal from a voice input device of an ultrasound review station;

(b) with the ultrasound review station, converting the voice signal into text; and

(c) with the ultrasound review station, inserting the text into an ultrasound examination report.

107. The method of claim 106, wherein (b) comprises converting the voice signal into text using a voice recognition unit of the ultrasound review station.

108. An ultrasound review station comprising:

a voice input device; and

a voice recognition unit coupled with the voice input device, the voice recognition unit being operative to convert a voice signal received from the voice input device into text.

109. The invention of claim 108, wherein the voice recognition unit is further operative to provide the text to an ultrasound examination report.

110. The invention of claim 108 further comprising a processor coupled with the voice input device, wherein the voice recognition unit is implemented with the processor.

111. The invention of claim 108 further comprising a processor coupled with the voice input device and the voice recognition unit.

112. A method for providing voice data from a first ultrasound device to a second ultrasound device, the method comprising:

(a) receiving voice data from a voice input device of a first ultrasound device;

(b) digitizing the voice data receiving in (a); and

(c) transmitting the digitized voice data from the first ultrasound device to a second ultrasound device, the first and second ultrasound devices being selected from the group of ultrasound devices consisting of an ultrasound imaging system and an ultrasound review station.

113. The invention of claim 112, wherein (c) further comprises transmitting an ultrasound image with the digitized voice data.

114. The invention of claim 112, wherein (c) further comprises transmitting an ultrasound examination report with the digitized voice data.

* * * * *

专利名称(译)	语音增强诊断医学超声系统和检查站		
公开(公告)号	US20030083577A1	公开(公告)日	2003-05-01
申请号	US10/306369	申请日	2002-11-26
[标]申请(专利权)人(译)	格林伯格JEFFREY中号		
申请(专利权)人(译)	格林伯格杰弗里·		
当前申请(专利权)人(译)	格林伯格杰弗里·		
[标]发明人	GREENBERG JEFFREY M		
发明人	GREENBERG, JEFFREY M.		
IPC分类号	A61B8/00 G01S7/52 G06F19/00 G06N3/063 A61B8/12		
CPC分类号	A61B5/7475 A61B8/00 A61B8/468 G06N3/063 G06F19/321 G06F19/3406 G06F19/3487 G01S7/52023 G16H15/00 G16H30/40 G16H40/63		
其他公开文献	US6743175		
外部链接	Espacenet USPTO		

摘要(译)

这里描述的优选实施例涉及语音增强诊断医学超声成像系统和检查站以及语音相关用户界面。利用这些优选实施例，用户可以通过发出口头命令而不是使用需要用户进行物理操纵的鼠标，键盘或其他用户界面来与成像系统或检查站交互。这提供了一个非常用户友好的界面，特别是那些难以导航复杂窗口和菜单层次结构或操作指点设备有困难的用户。这也改善了患者流量并提供了更有效的报告生成系统。语音反馈还可用于允许成像系统或检查站更好地与用户通信。