

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
29 January 2009 (29.01.2009)

PCT

(10) International Publication Number  
WO 2009/014935 A1

- (51) International Patent Classification:  
H04M 3/53 (2006.01) G11B 27/031 (2006.01)  
G10L 15/00 (2006.01)
- (21) International Application Number:  
PCT/US2008/070036
- (22) International Filing Date: 15 July 2008 (15.07.2008)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
11/880,152 20 July 2007 (20.07.2007) US
- (71) Applicant (for all designated States except US): CISCO TECHNOLOGY, INC. [US/US]; 170 West Tasman Drive, SJC/10/2/1, San Jose, California 95134-1706 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): KHOURI, Joseph F. [US/US]; 5925 Foligno Way, San Jose, California 95138 (US). PHILONENKO, Laurent [FR/US]; 3529 Webster Street, San Francisco, California 94123 (US).

JAIN, Mukul [IN/US]; 3604 Wodzenski Drive, San Jose, California 95148 (US). SHAFFER, Shmuel [US/US]; 1211 Cowper Street, Palo Alto, California 94301 (US).

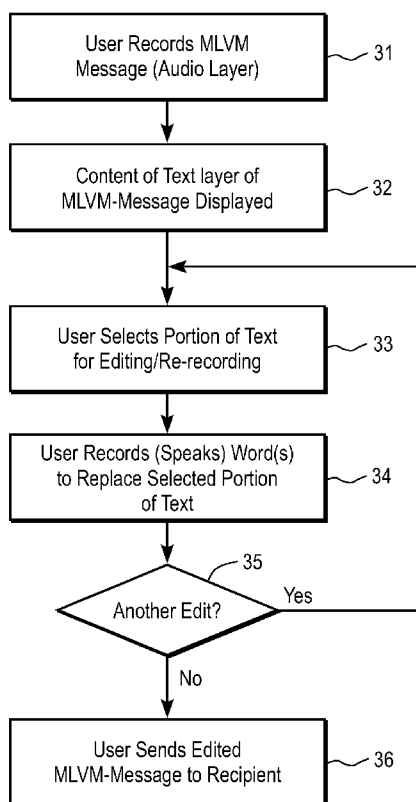
(74) Agents: LAFFERTY, Wm. Brook et al.; Scientific-Atlanta, Inc., Intellectual Property Dept., 5030 Sugarloaf Parkway, Lawrenceville, Georgia 30044 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM,

[Continued on next page]

(54) Title: TEXT-ORIENTED, USER-FRIENDLY EDITING OF A VOICEMAIL MESSAGE



(57) Abstract: A system includes a server associated with a unified messaging system (UMS). The server records speech of a user as an audio data file, translates the audio data file into a text data file, and maps each word within the text data file to a corresponding segment of audio data in the audio data file. A graphical user interface (GUI) of a message editor running on an endpoint associated with the user displays the text data file on the endpoint and allows the user to identify a portion of the text data file for replacement. The server being further operable to record new speech of the user as new audio data and to replace one or more segments of the audio data file corresponding to the portion of the text with the new audio data.

FIG. 3

WO 2009/014935 A1



ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),  
European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI,  
FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL,  
NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG,  
CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**

— *with international search report*

## TEXT-ORIENTED, USER-FRIENDLY EDITING OF A VOICEMAIL MESSAGE

### FIELD OF THE INVENTION

**[0001]** The present disclosure relates generally to the field of electronic messaging, including voice messaging, over telephony networks.

### BACKGROUND OF THE INVENTION

**[0002]** Voice messaging systems that enable users to send and retrieve voice mail messages are known in the communication arts. In a typical prior art voice messaging system a telephone is attached to a private branch exchange (PBX) that utilizes a notification mechanism to notify a message recipient that a new message is waiting for them. Today, many modern communication systems provide messaging services via packet-based networks, i.e., those that operate in accordance with the Internet Protocol (IP). A Unified Messaging / Voicemail System (UMS) handles voice, facsimile, regular text messages, and computer-readable documents as objects in a single mailbox that a user can access either with a regular email client, or by telephone. A UM system usually connects to a PBX to provide automated attendant, audiotext, and voicemail services to subscribers or users. For instance, a personal computer (PC) user with multimedia capabilities typically can open and playback voice messages, either as speech or text.

**[0003]** In a modern UMS, when a user who is recording a voice message makes a mistake he is usually forced to delete the message and re-record the entire voice message using a standard Telephony User Interface (TUI), e.g., to record, delete, re-record, send, etc. This is particularly time consuming and irritating when the voicemail message is lengthy, which also increases the likelihood of making another error during re-recording of the voice message.

**BRIEF DESCRIPTION OF THE DRAWINGS**

**[0004]** The present disclosure will be understood more fully from the detailed description that follows and from the accompanying drawings, which however, should not be taken to limit the invention to the specific embodiments shown, but are for explanation and understanding only.

**[0005]** **Figure 1** illustrates an example communications system with voice messaging capabilities.

**[0006]** **Figure 2A** illustrates an example of a voicemail message displayed as text using a graphical user interface.

**[0007]** **Figure 2B** illustrates the example of Figure 2A after editing of the message text.

**[0008]** **Figure 3** illustrates an example method of operation for the communications system of Figure 1.

**[0009]** **Figure 4** illustrates an example network node or device.

**DETAILED DESCRIPTION**

**[0010]** In the following description specific details are set forth, such as device types, system configurations, protocols, methods, etc., in order to provide a thorough understanding of the disclosure herein. However, persons having ordinary skill in the relevant arts will appreciate that these specific details may not be needed to practice the embodiments described.

**[0011]** In the context of the present application, a communications network is a geographically distributed collection of interconnected subnetworks for transporting data between nodes, such as intermediate nodes and end nodes (also referred to as endpoints). A local area network (LAN) is an example of such a subnetwork; a plurality of LANs may be further interconnected by an intermediate network node, such as a router, bridge, or switch, to extend the effective "size" of the computer network and increase the number of communicating nodes. Examples of the devices or nodes include servers, mixers, control units, and personal computers. The nodes typically communicate by exchanging discrete frames or packets of data according to predefined protocols.

**[0012]** A "voicemail system", as that term is used in the present disclosure, refers to a computer system (e.g., a UMS) or network, or more expansive systems that include gateway devices attached to other computer systems, thus enabling users to send / retrieve recorded multimedia messages anywhere in the world. A "voicemail" or multimedia message which comprises voice, video, text, etc., is a recorded (e.g., audio) message stored in such a computerized system. An endpoint is any device, component, element, or object capable of recording, editing, sending or receiving a voicemail message over a network. An endpoint may comprise a personal computer (PCs), workstation, personal digital assistant (PDA), laptop or notebook computer, VoIP phone, or other communications equipment. A Unified Messaging (or UM) system is a computerized system that integrates different

streams of messages (e-mail, fax, voice, video, etc.) into a single “in-box”, accessible from a variety of different endpoint devices.

**[0013]** Speech-to-text (STT) technology is a known method of transforming speech into a text file. Similarly, a text-to-speech (TTS) system converts normal language text into speech. Direct editing of audio files, e.g., a .WAV file, is also a known technology. (WAV or WAVE is short for Waveform audio format, which is a Microsoft and IBM audio file format standard for storing audio on PCs.) Examples these latter types of voice editors can be found at <http://www.catauto.com/ed1000w.html>; <http://www.catauto.com/ed300dxl.html>; <http://www.goldwave.com>; <http://www.snapfiles.com/get/soliton.html>; and <http://www.voicemailspecial.com/>. The difficulty with these voice editors is that the user only sees the audio “wavefront” and must move the cursor in a “hit or miss” trial fashion trying to locate the exact place of the error.

### **Overview**

**[0014]** According to one embodiment of the present invention, a UMS provides a user with a textual view for editing a voicemail message that has been recorded, but not yet sent. The user may edit the voicemail message by highlighting a word or portion of the displayed text and then either typing new text or recording new speech that replaces the audio data corresponding to the highlighted text. That is, the user may edit his voicemail message without having to listen to any portion of the recording. Once editing of the voicemail message has been completed, the user may send the edited voicemail message as one or more audio data files to the intended recipient’s mailbox.

**[0015]** The UMS handles the edited voicemail message just like any other rich-media UMS message. That is, the UMS handles the edited voicemail message as one or more objects in a single mailbox that the receiver can access either with a regular endpoint device, such as a personal computer, (PC), a telephone, or other

suitable communications device. The edited audio file(s) may also be filtered or processed to eliminate or minimize any differences in volume, tone, noise levels, etc., between the original recording and the edited portions. This post-recording processing of the message smoothes out the speech, thereby making the edited portions undetectable to the recipient.

**[0016]** Referring to Figure 1, an example communications system 10 is shown with voice messaging capabilities provided by a UMS 14 (e.g., an application or code running on a server) attached to a digital network 13. A Multi-Layer Voice Messaging Server (MLVM-Server) 12, also attached to network 13, is shown attached to UMS 14. Individual users 11 and 18 may send / retrieve messages to / from UMS 14 (and take responsive actions thereto) utilizing any number of different endpoint devices and communication paths. For example, user 11 is shown using a PC 15 to call into UMS 14 to leave a voicemail message for user (recipient) 18. User 11 is also shown utilizing a Multi-Layer Voice Messaging Editor (MLVM-Editor) 16 that translates his recorded speech into text appearing on the screen of his PC 15 to facilitate editing of his voice message. User 18 may listen to her voicemail messages and initiate responsive actions using telephone (e.g., a VoIP phone), which is shown attached to network 13. Users 11 and 18 may also connect with UMS 14 using other types of endpoint devices that provide two-way communications with UMS 14 over one or more network connections.

**[0017]** In the embodiment shown, MLVM-Server 12 comprises an add-on hardware (firmware) and/or software module to UMS 14 and utilizes standard speech-to-text (STT) capabilities in real-time to create a Multi-Layered VoiceMail Message (MLVM-Message). That is, MLVM-Server 14 may incorporate or be associated with a STT converter for converting a voice message to text for editing by a user, as well as a text-to-speech (TTS) converter for converting a text message to audible speech.

**[0018]** The MLVM-Message has a data structure which comprises an Audio Layer, which is the actual audio data consisting of the user's speech recorded in any one of a number of different data formats (e.g., .wav, G711, etc.), a Text Layer, which is the text data generated as a result of the translation of audio data using an STT converter, and a Text-to-Audio Mapping Layer, which is a layer that maps each unit within the Text Layer (e.g. text word) to its corresponding segment or unit (chunk of audio data) in the Audio Layer, and vice versa. By way of example, when user 11 calls into UMS 14 and starts recording a voicemail message, MLVM-Server 12 creates, in real-time, the translated Text Layer and the Mapping Layer that links the generated text to the corresponding audio data in the Audio Layer of the MLVM-Message.

**[0019]** In the embodiment of Figure 1, MLVM-Editor 16 comprises a "client" application or plug-in module (i.e., executable code) that runs on the user's client device (e.g. PC 15). MLVM-Editor 16 may be invoked by user 11 to operate on the MLVM-Message created by MLVM-Server 12. MLVM-Editor 16 provides user 11 with a textual interface (e.g., a graphical user interface (GUI)), for viewing and editing of the recorded voice message, either in real-time or after the complete message has been recorded. For example, while user 11 is recording his voicemail message for recipient 18 he may utilize MLVM-Editor 16 to view his recorded speech as a corresponding text translation displayed on the screen of PC 15 (either in real-time or at the end of the recording). In other words, MLVM-Editor 16 displays the content of the Text Layer of the MLVM-Message created by MLVM-Server 12 as user 11 speaks.

**[0020]** In the event that user 11 makes a mistake while recording his message, he may finish recording the remainder of his voice message and then use MLVM-Editor 16 to easily scroll through the text to locate the portion of the message that needs to be edited. Once that portion of the text has been located, user 11 may



perform normal text editing, such as deleting or moving text from one portion of the message to another. MLVM-Editor 16, by utilizing the Text-to-Audio Mapping Layer of the corresponding MLVM-Message, automatically “mirrors” the exact editing to the corresponding audio data in the Audio Layer of the MLVM-Message. In another example embodiment, rather than using TTS the system identifies and utilizes recorded words with the voice of user 11.

**[0021]** Figure 2A illustrates an example of a voicemail message displayed as text in a GUI window 20 generated by MLVM-Editor 16. Figure 2A shows GUI window 20 comprising a text box or field 21 that includes the translated text of the audio data corresponding to the user’s recorded voice message. An instruction / command field at the bottom of GUI window 20 includes a “Go” command button (icon) 20 that may be clicked-on or otherwise selected by a user to cause the MLVM-Editor 16 to start recording new audio data (speech) to replace the portion of audio data corresponding to the text highlighted by the user (e.g., “Tuesday”). In other words, clicking icon button 22 allows the user to record over a portion of his voicemail message that he wants to change or correct. When the user has finished recording the new speech in replacement of the highlighted portion, he may click-on or otherwise select the “Done” command button (icon) 23.

**[0022]** Figure 2B illustrates the example of Figure 2A after editing of the message text, which, in this case, involved changing the word “Tuesday” to “Monday” in the voicemail message. Note that in the embodiment shown, clicking-on icon 23 (see Figure 2A) has the effect of changing the command field icons. With no highlighted text in field 21, GUI window 20 presents the user with the option of either sending the voice message as it appears in field 21 (i.e., by clicking “Send” button 24), or cancelling the changes previously made (i.e., by clicking “Cancel” button 25). The user may also choose to highlight another portion of the text, which

automatically causes the instruction / command field with buttons 22 & 23 to appear at the bottom of window 20.

**[0023]** Note that clicking on Send button 24 causes the modified audio data file(s) comprising the edited voicemail message to be placed or stored in the intended recipient's mailbox of UMS 14.

**[0024]** In another embodiment, a user of MLVM-Editor 16 may display the Text Layer data (i.e., a text view of the voice message) to find a place in the message where he forgot to say something. For example, if he forgot to leave his administrative assistant's name and phone number he could scroll through or search the text view/message and, using his cursor, highlight the space or word(s) where the new speech is to be inserted. In an alternative implementation, the user may simply position his cursor and hit a "Record" button or icon appearing in GUI window 20 or otherwise invoke recording by another command input (e.g., keypad sequence or combination). The user may then record (by speaking) the missing information (e.g., "Please call my assistant Mary at 408-902-1234 and let her know if you are available"). In this case, MLVM-Editor 16 inserts the newly recorded audio data in the corresponding location of the MLVM-Message Audio Layer, as well as inserting the corresponding text (audio-to-text translated) into the MLVM-Message Text Layer at the appropriate position in text field 21.

**[0025]** Practitioners in the art will understand that MLVM-Editor 16 may run on the user's PC, or, alternative, the GUI shown in Figures 2A & 2B may comprise a collaborative web-based application that is accessed by the browser software running on the user's PC. In other instances, the editor GUI may comprise a downloaded application, or other forms of computer-executable code that may be loaded or accessed by a participant's PC. For instance, the software code for implementing the MLVM-Editor functions may be incorporated into, and executed on, server 12, which may be accessed by callers to UMS 14 who want to utilize the

features provided therein. In another embodiment, instead of running on server 12, MLVM-Editor 16 may run on a different server attached to network 13. It is further appreciated that the editor function described herein may be invoked automatically by the system, or in response to certain user input / action, such as by pressing a certain predefined function key on his endpoint device, or in response to an interactive voice response (IVR) system prompt (e.g., "Do you want to edit any portion of your voice message before sending it?").

**[0026]** In accordance with another embodiment, a user may use the MLVM-Editor to delete / change portions of the recorded message, or insert additional or missing information, by typing new text at appropriate position in the displayed text field. In other words, instead of recording new speech as described above, the system allows the user to simply update the Text layer by typing new text. The system then updates the corresponding audio data using a standard TTS converter. In a specific implementation, the TTS engine may use voice characteristics of the user extracted from the recorded speech (or from stored data files of the user's speech) for converting the text to speech, thereby matching the text-to-speech portion with the rest of recorded message.

**[0027]** In still another embodiment, the user can mark a text segment from the Text Layer of a MLVM-Message (i.e., the textual view of the voicemail message) using the MLVM-Editor and cut / paste that segment to another place in the MLVM-Message, or even into another MLVM-Message. This cut / paste operation has the effect of moving a copy of the audio data corresponding to the cut text data, as well as moving the corresponding and Text-to-Audio Mapping Layer segments.

**[0028]** Additionally, rather than using a TTS module to generate speech that attempts to emulate the voice of user 11, the system searches within existing voice messages for words, sound patterns, or utterances that are similar to those needed

for the editing task. The system may utilize the words, sound patterns, or utterances rather than the TTS generated speech.

**[0029]** Figure 3 illustrates an example method of operation for the communications system of Figure 1. The process begins with the user/caller dialed into the UMS (server) and either in the process of recording his voicemail message or having just finished a recording (block 31). As the user speaks the UMS translates the user's speech into text and creates a MLVM-Message comprising the Audio, Text and Mapping Layers described previously. Either automatically, or in response to a command input of the user (e.g., invoking the MLVM-Editor), the content of the Text Layer of the MLVM-Message is displayed to the user (block 32). This textual view of the voicemail message allows the user to select a word or portion of the text for editing / re-recording (block 33). For example, using the text oriented/user-friendly MLVM-Editor the user can quickly locate the right position within a voicemail message where the user would like to replace an existing portion of the original message or insert newly recorded speech. In the example shown, the user records one or more words in replacement of the selected portion of the displayed text (block 34). Other types of known text editing operations (e.g., insertion, augmentation, drag and drop, copy / paste) may also be performed.

**[0030]** Note that according to the method shown, the edited voice message (using textual view) is recorded in user's own voice. In other embodiments, a computer-generated voice, or a computer-synthesized version of the user's voice (based on the user's detected or stored voice characteristics) may be utilized for edits to the audio data.

**[0031]** It is appreciated that the MLVM-Editor may also provides a user-friendly way to locate a voice segment (through the textual view), which segment can be copy/cut and pasted into another place in that same multi-layer message or into another multi-layer voicemail message.

**[0032]** Once the user has finished inserting newly recorded speech into the MLVM-Message, or deleting / replacing portions of the message, he is given the option of making another edit (block 35). If the user elects to make further edits to his message, the process returns to block 33. On the other hand, once editing has been completed the user may then send the voice portion (i.e., audio data) of the message to the intended recipient (block 36). This latter step may be initiated simply by closing the GUI window provided by the MLVM-Editor on the user's PC display screen. The recipient may retrieve the voicemail message by accessing her mailbox within the UMS any time thereafter.

**[0033]** Figure 4 is a generalized block diagram showing an example network device or node 50, such as may comprise any of the systems, devices, or nodes (e.g., a PC or server) shown or described above. Node 50 includes a processor subsystem 51 coupled with a memory unit (main memory) 52 via a memory bus 53, one or more hardware (and/or software / firmware) modules 54, and an input/output (I/O) system 55 via a system bus 56. Modules 54 may include software, firmware, or logic embedded in hardware for implementing any of the functions described herein, e.g., those functions associated with editing a voicemail message recorded by a user, etc.

**[0034]** It is appreciated that any unified messaging / voicemail system utilized by or in conjunction with node 50 may comprise separate hardware devices coupled to the system bus 56, or, alternatively, implemented as software programs or modules 54 that run on one or more processors of subsystem 51. In other words, the composing, sending, storing, and receiving of electronic messages, as well as other associated functions may be implemented as separate hardware devices, memory locations (storing executable code), firmware devices, software modules, or other machine-readable devices. (In the context of the present application, therefore, the term "module" is to be understood as being synonymous with both hardware devices

and computer-executable software code, programs or routines.) Additionally, although voice messaging has been specifically described in the above examples, it should be understood that the concepts presented are equally applicable to all types of multimedia messaging.

**[0035]** It should be further understood that elements of the present invention may also be provided as a computer program product which may include a machine-readable medium having stored thereon instructions which may be used to program a computer (e.g., a processor or other electronic device) to perform a sequence of operations. Alternatively, the operations may be performed by a combination of hardware and software. The machine-readable medium may include, but is not limited to, floppy diskettes, optical disks, CD-ROMs, and magneto-optical disks, ROMs, RAMs, EPROMs, EEPROMs, magnet or optical cards, propagation media or other type of media/machine-readable medium suitable for storing electronic instructions. For example, elements of the present invention may be downloaded as a computer program product, wherein the program may be transferred from a remote computer or telephonic device to a requesting process by way of data signals embodied in a carrier wave or other propagation medium via a communication link (e.g., a modem or network connection).

**[0036]** Additionally, although the present invention has been described in conjunction with specific embodiments, numerous modifications and alterations are well within the scope of the present invention. For instance, although functions and methods of various embodiments have been described as being embodied on a UMS coupled to a server via an IP network, it is appreciated that these same methods and functions may be embodied on an endpoint device of a user, another server or client, a messaging application, or a third party service application program accessible via a web-based browser. Accordingly, the specification and drawings are to be regarded in an illustrative rather than a restrictive sense.

**CLAIMS**

We claim:

1. A method comprising:
  - recording speech of a user as an audio data file, the speech comprising a voicemail message intended for a recipient;
  - translating the audio data file into a text data file;
  - mapping each word within the text data file to a corresponding segment of audio data in the audio data file;
  - editing one or more segments of the audio data file to produce an edited voicemail message, the editing being based on input commands received from the user, the input commands including operations on one or more words in the text data file which correspond to the one or more segments in the audio data file; and
  - sending the edited voicemail message to the recipient.
  
2. The method of claim 1 wherein the editing comprises replacing a segment in the audio data file corresponding to a selected word in the text data file with a new segment which comprises newly recorded speech of the user.
  
3. The method of claim 1 wherein the editing comprises replacing a segment in the audio data file corresponding to a selected word in the text data file with a new segment which comprises synthesized speech of the user, the synthesized speech being generated in response to newly typed text by the user.
  
4. The method of claim 1 wherein the editing comprises:
  - receiving input from the user which identifies a location in the text data file;

inserting one or more new segments at a position in the audio data file corresponding to the location, the one or more segments comprising new speech.

5. The method of claim 1 further comprising displaying the text data file on a display screen of an endpoint utilized by the user.

6. The method of claim 1 wherein the editing comprises:  
receiving input from the user which identifies the one or more words in the text data file;  
recording new speech of the user as new audio data; and  
replacing one or more segments in the audio data file which correspond to the one or more words in the text data file with the new audio data.

7. The method of claim 6 further comprising prompting the user to make edits to the voicemail message.

8. The method of claim 1 further comprising storing the edited voicemail message as audio data in a mailbox of the recipient in a unified messaging system.

9. A method comprising:  
creating a Multi-Layered Voicemail (MLVM) message in a unified messaging system (UMS) mailbox of a recipient, the voicemail message being left by a caller from an endpoint, the MLVM message including an audio layer comprising audio data corresponding to speech of the caller, a text layer comprising text data generated from a translation of the audio data, and a mapping layer that maps each word in the text layer to a corresponding segment of audio data in the audio layer;



receiving input from a user interface running on the endpoint, the input comprising edits to one or more words of the text layer;

changing one or more segments of audio data in the audio layer in correspondence with the edits to the one or more words of the text layer;

storing the MLVM message with the changed one or more segments of audio data in the UMS mailbox of the recipient.

10. The method of claim 9 wherein the receiving of the input from the user interface comprises receiving new speech from the caller in replacement of the audio data corresponding to the one or more words of the text layer.

11. The method of claim 9 wherein the changing of the one or more segments of the audio data comprises:

generating synthesized speech from voice characteristics of the caller; and

replacing audio data corresponding to the one or more words of the text layer with the synthesized speech.

12. Logic encoded in one or more media for execution and when executed is operable to:

record speech of a caller as a voicemail message for a recipient;

translating the speech into text for display on a graphical user interface (GUI) running on an endpoint associated with the caller;

map each word of the text to a corresponding segment of audio data of the recorded speech;

receive input from the GUI that identifies one or more words of the text;

record new speech from the caller;

replace one or more segments of audio data of the recorded speech corresponding to the one or more words of the text with the new speech.

13. The logic of claim 12 wherein the logic, when executed, is further operable to store the voicemail message with the new speech in a voicemail mailbox of the recipient.

14. The logic of claim 12 wherein the logic, when executed, is further operable to:

receive additional input from the GUI that identifies a portion of the text for deletion; and

delete one or more segments of audio data of the recorded speech corresponding to the portion of the text.

15. The logic of claim 12 wherein the logic, when executed, is further operable to:

receive additional input from the GUI that identifies a portion of the text; and  
copy one or more segments of audio data of the recorded speech corresponding to the portion of the text.

16. The logic of claim 15 wherein the logic, when executed, is further operable to paste the one or more segments of audio data in another voicemail message at a location identified by a cursor in a text file translation of an audio data file of the another voicemail message.

17. Logic encoded in one or more media for execution and when executed is operable to:

record speech of a caller as an audio data file, the speech comprising a media message for a recipient;

translating the speech into a text file for display on a graphical user interface (GUI) running on an endpoint associated with the caller;

map each word of the text file to a corresponding segment of audio data of the recorded speech;

receive input from the GUI that identifies a location within the text file;

insert one or more segments of audio data at a position in the audio data file corresponding to the location within the text file, the one or more segments comprising new speech.

18. The logic of claim 17 wherein the new speech comprises spoken words of the caller.

19. A system comprising:

one or more network nodes running one or more application programs that implement a unified messaging system (UMS), the one of the nodes including:

one or more processors; and

a memory comprising one or more instructions executable at the processors, the one or more processors being operable, when executing the instructions, to:

create a voicemail message from speech left by a caller associated with an endpoint, the voicemail message including an audio layer comprising audio data corresponding to the speech, a text layer comprising text data generated from a

translation of the audio data, and a mapping layer that maps each word in the text layer to a corresponding segment of audio data in the audio layer;

receive input from a user interface running on the endpoint, the input comprising one or more edits to the text layer;

replace or insert one or more segments of audio data in the audio layer in correspondence with the one or more edits to the text layer; and

store the MLVM message with the one or more segments of audio data in the UMS mailbox of the recipient.

20. The system of claim 19 wherein the one or more processors are further operable, when executing the instructions, to receive new speech from the caller in replacement or insertion of the one or more segments of audio data corresponding to the one or more edits to the text layer.

21. The system of claim 19 wherein the one or more processors are further operable, when executing the instructions, to:

generate synthesized speech from voice characteristics of the caller, the synthesized speech comprising the one or more segments of audio data.

22. A system comprising:

a server associated with a unified messaging system (UMS), the server being operable to:

record speech of a user as an audio data file, the speech comprising a voicemail message intended for a recipient;

translate the audio data file into a text data file;

map each word within the text data file to a corresponding segment of audio data in the audio data file; and

a message editor coupled with the server, the message editor comprising a graphical user interface (GUI) running on an endpoint associated with the user, the GUI being operable to display the text data file on a display of the endpoint and allow the user to identify a portion of the text data file for replacement or deletion, the GUI being further operable to identify a position in the text data file for insertion of new text, the server being further operable to record new speech of the user as new audio data and to either replace one or more segments of the audio data file corresponding to the portion of the text with the new audio data, or insert the new audio data at a location of the audio data file corresponding to the position in the text data file.

23. The system of claim 22 wherein the server is further operable to store the audio data file, including the new audio data in a voicemail mailbox in the UMS associated with the recipient.

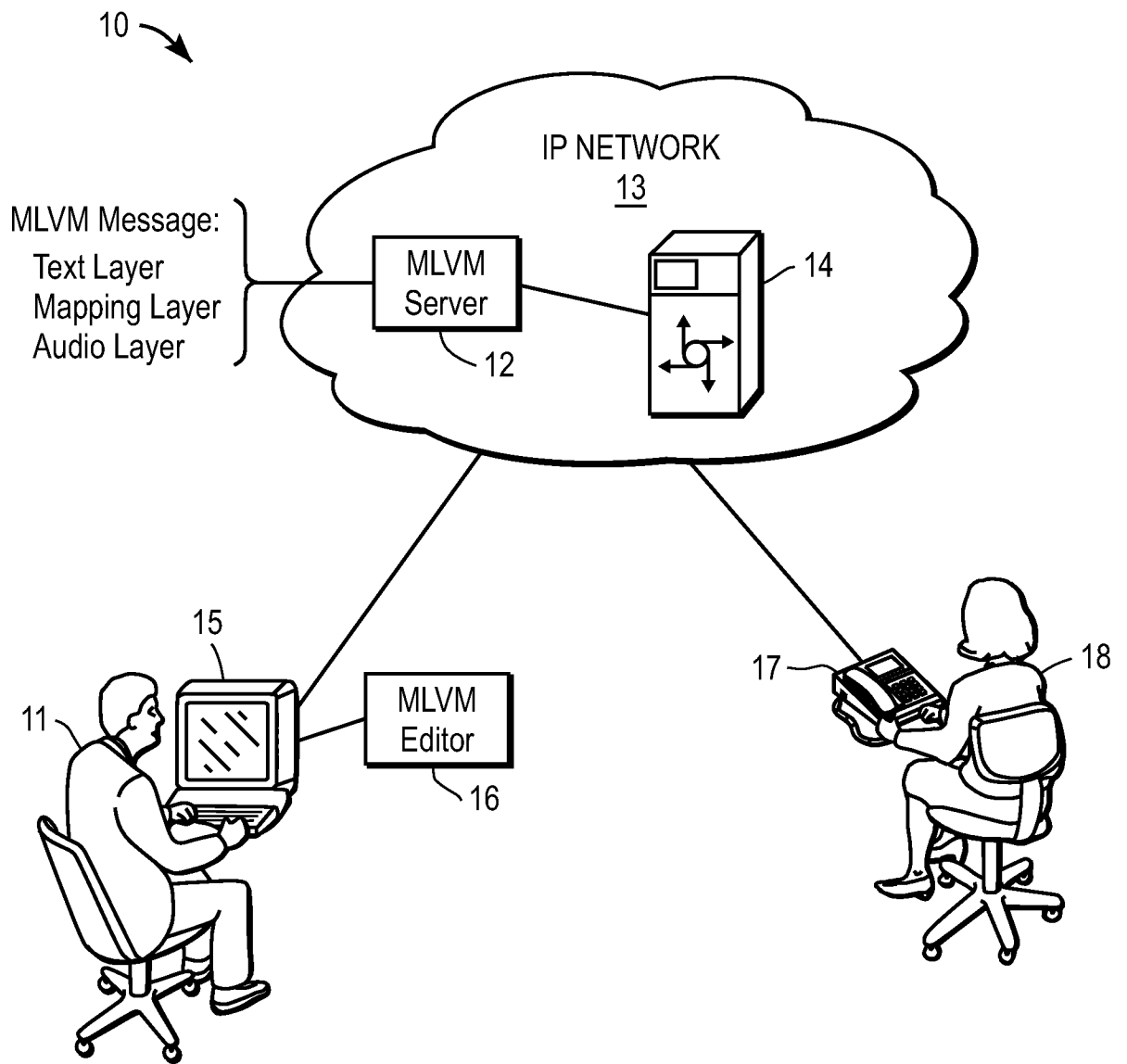


FIG. 1

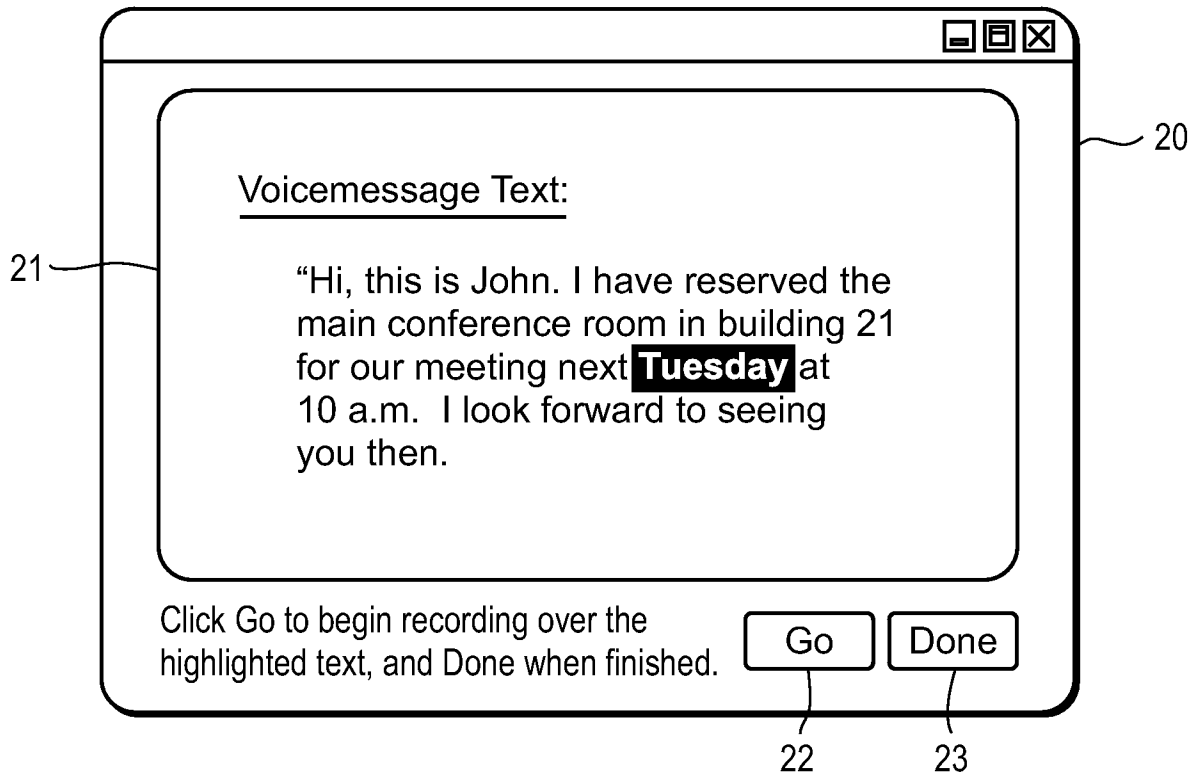


FIG. 2A

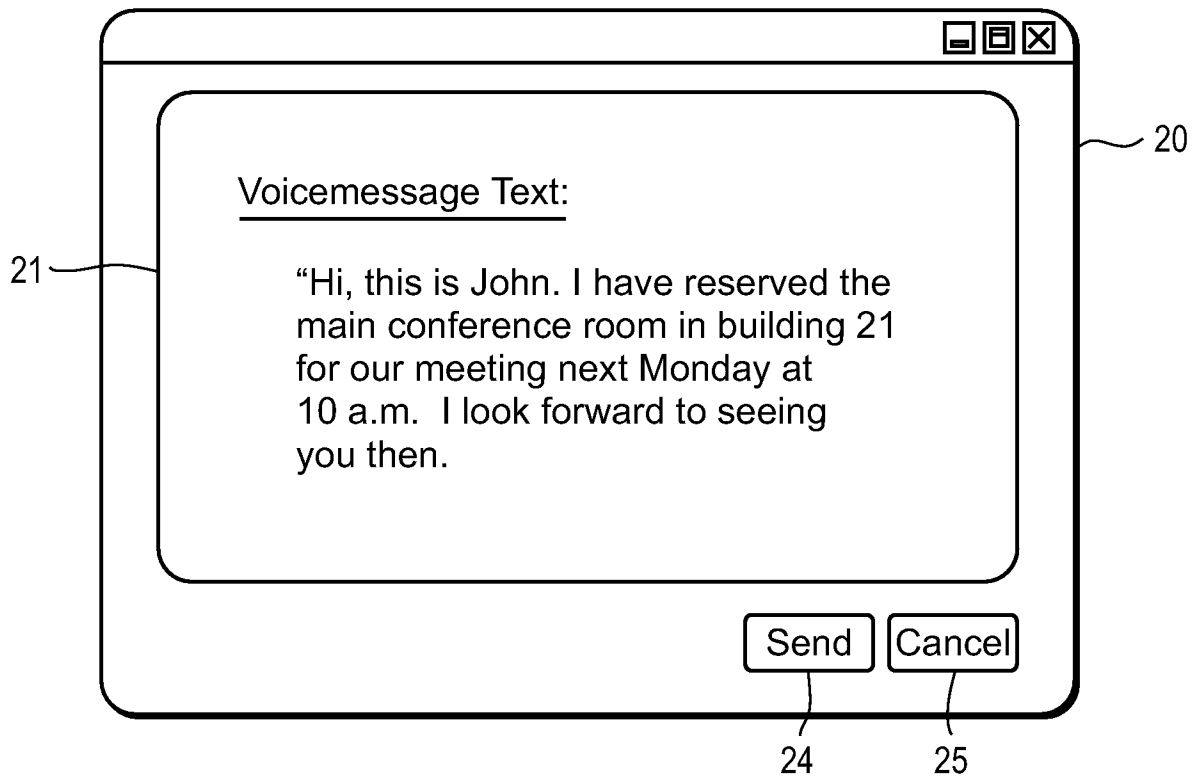


FIG. 2B



4/5

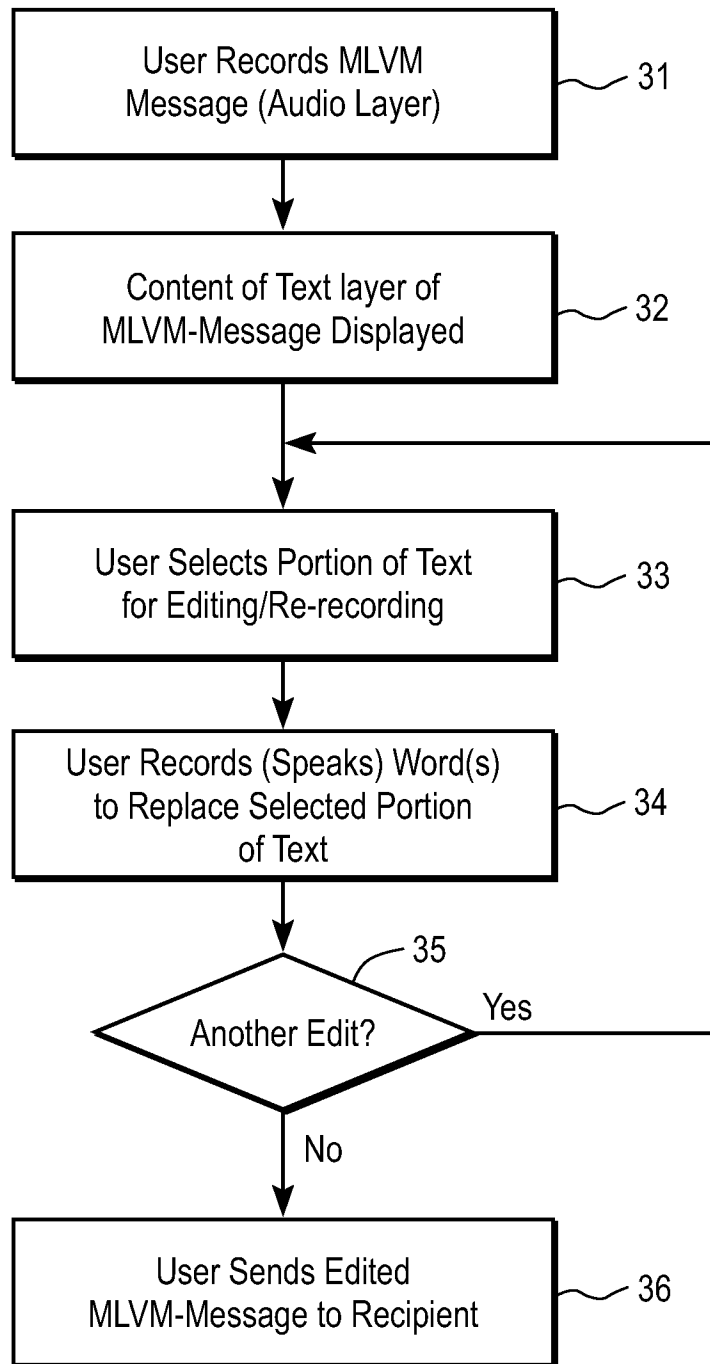


FIG. 3

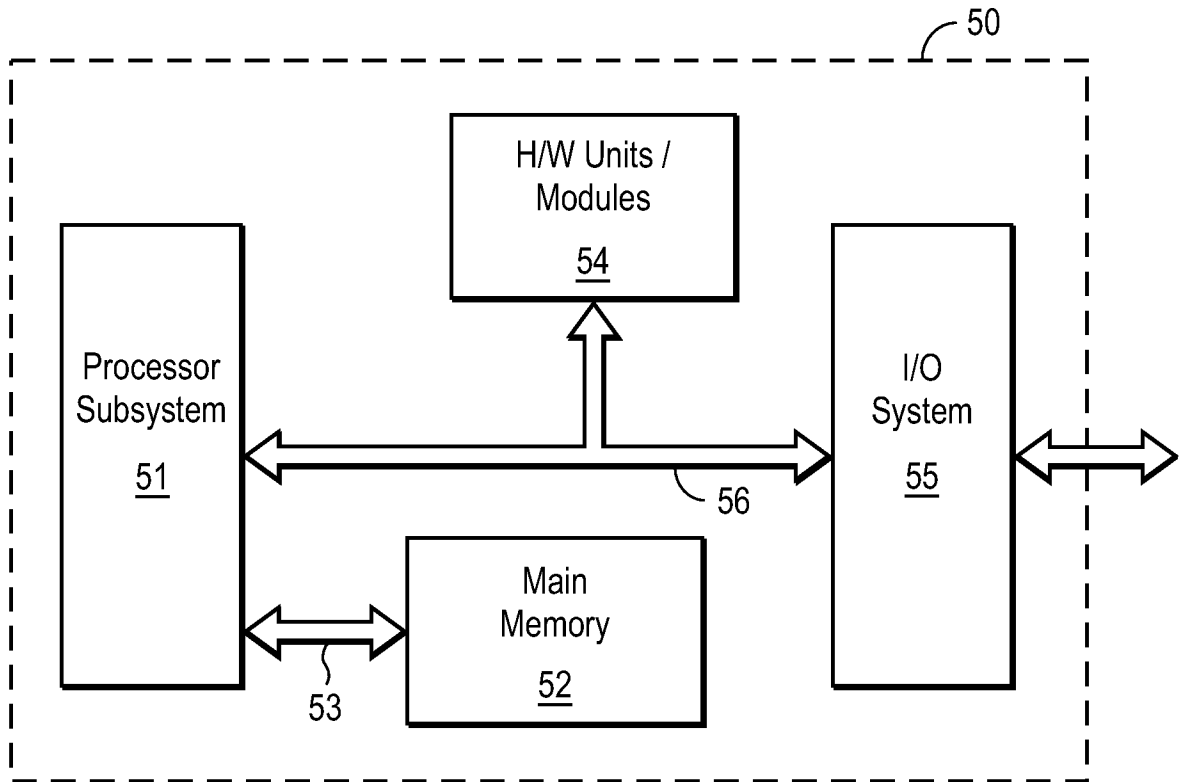


FIG. 4

# INTERNATIONAL SEARCH REPORT

International application No  
PCT/US2008/070036

<b>A. CLASSIFICATION OF SUBJECT MATTER</b> INV. H04M3/53      G10L15/00      G11B27/031		
According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b>		
Minimum documentation searched (classification system followed by classification symbols) H04M G10L G11B		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practical, search terms used) EPO-Internal		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2005/037739 A1 (ZHONG JIALIN [US]) 17 February 2005 (2005-02-17) abstract; figure 6 paragraph [0008]	1-23
Y	US 6 604 078 B1 (SHIMAZAKI IZUMI [JP]) 5 August 2003 (2003-08-05) abstract; figure 1 columns 1,2,7	1-23
A	US 5 675 778 A (JONES CAMERON [US]) 7 October 1997 (1997-10-07) the whole document	1-23
A	EP 1 170 726 A (IBM [US]) 9 January 2002 (2002-01-09) the whole document	1-23
----- -/--		
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.		
* Special categories of cited documents :		
*A* document defining the general state of the art which is not considered to be of particular relevance *E* earlier document but published on or after the international filing date *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) *O* document referring to an oral disclosure, use, exhibition or other means *P* document published prior to the international filing date but later than the priority date claimed	*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. *&* document member of the same patent family	
Date of the actual completion of the international search  <p style="text-align: center; font-weight: bold;">30 October 2008</p>	Date of mailing of the international search report  <p style="text-align: center; font-weight: bold;">06/11/2008</p>	
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Authorized officer  <p style="text-align: center; font-weight: bold;">Domingos, Luis</p>	

## INTERNATIONAL SEARCH REPORT

International application No

PCT/US2008/070036

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages.	Relevant to claim No.
A	US 5 855 000 A (WAIBEL ALEXANDER [US] ET AL) 29 December 1998 (1998-12-29) -----	1-23

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No <b>PCT/US2008/070036</b>
--

Patent document cited in search report	A1	Publication date	Patent family member(s)	Publication date
US 2005037739	A1	17-02-2005	NONE	
US 6604078	B1	05-08-2003	JP 3417355 B2 JP 2001060097 A	16-06-2003 06-03-2001
US 5675778	A	07-10-1997	NONE	
EP 1170726	A	09-01-2002	CN 1356628 A US 7200555 B1	03-07-2002 03-04-2007
US 5855000	A	29-12-1998	US 5712957 A	27-01-1998