(54) Title: SEMIAUTOMATED RELAY METHOD AND APPARATUS
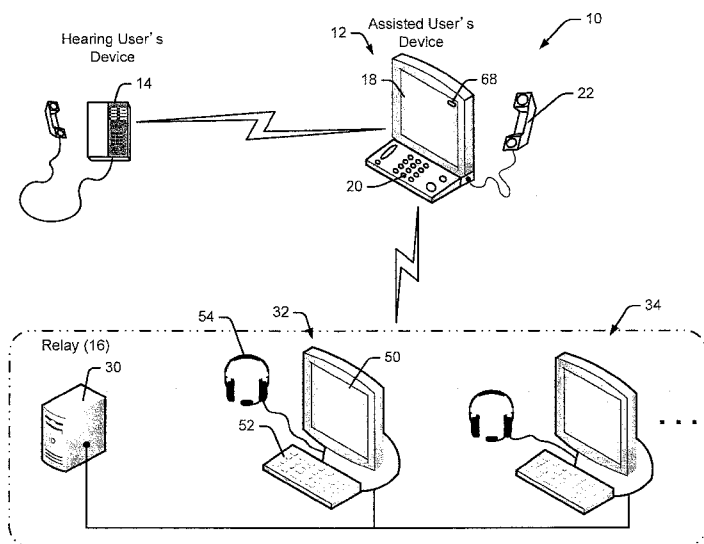
Fig. 1

(57) **Abstract:** A system for providing voice-to-text captioning service comprising a relay processor that receives voice messages generated by a hearing user during a call, the processor programmed to present the voice messages to a first call assistant, a call assistant device used by the first call assistant to generate call assistant generated text corresponding to the hearing user's voice messages, the processor further programmed to run automated voice-to-text transcription software to generate automated text corresponding to the hearing user's voice messages, use the call assistant generated text, the automated text and the hearing user's voice messages to train the voice-to-text transcription software to more accurately transcribe the hearing user's voice messages to text, determine when the accuracy exceeds an accuracy requirement threshold, during the call and prior to the automated text exceeding the accuracy requirement threshold, transmitting the call assistant generated text to the assisted user's device for display to the assisted user and subsequent to the automated text exceeding the accuracy requirement threshold, transmitting the automated text to the assisted user's device for display to the assisted user.

SEMIAUTOMATED RELAY METHOD AND APPARATUS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001]   This application is based on U.S. Provisional Patent Application Serial No. 61/946,072 filed on February 28, 2014 and entitled "SEMIAUTOMATED RELAY METHOD AND APPARATUS."

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

[0002]   Not applicable.

BACKGROUND OF THE INVENTION

[0003]   The present invention relates to relay systems for providing voice-to-text captioning for hearing impaired users and more specifically to a relay system that uses automated voice-to-text captioning software to transcribe voice-to-text.

[0004]   Many people have at least some degree of hearing loss. For instance, in the United states, about 3 out of every 1000 people are functionally deaf and about 17 percent (36 million) of American adults report some degree of hearing loss which typically gets worse as people age. Many people with hearing loss have developed ways to cope with the ways their loss effects their ability to communicate. For instance, many deaf people have learned to use their sight to compensate for hearing loss by either communicating via sign language or by reading another person's lips as they speak.

[0005]   When it comes to remotely communicating using a telephone, unfortunately, there is no way for a hearing impaired person (e.g., an assisted user (AU)) to use sight to compensate for hearing loss as conventional telephones do not enable an assisted user to see a person on the other end of the line (e.g., no lip reading or sign viewing). For persons with only partial hearing impairment, some simply turn up the volume on their telephones to try to compensate for their loss and can make do in

most cases. For others with more severe hearing loss conventional telephones cannot compensate for their loss and telephone communication is a poor option.

[0006]   An industry has evolved for providing communication services to assisted users whereby voice communications from a person linked to an assisted user's communication device are transcribed into text and displayed on an electronic display screen for the assisted user to read during a communication session. In many cases the assisted user's device will also broadcast the linked person's voice substantially simultaneously as the text is displayed so that an assisted user that has some ability to hear can use their hearing sense to discern most phrases and can refer to the text when some part of a communication is not understandable from what was heard.

[0007]   US patent No. 6,603,835 (hereinafter "the '835 patent) titled system for text assisted telephony teaches several different types of relay systems for providing text captioning services to assisted users. One captioning service type is referred to as a single line system where a relay is linked between an AU's device and a telephone used by the person communicating with the AU. Hereinafter, unless indicated otherwise the other person communicating with the assisted user will be referred to as a hearing user (HU) even though the AU may in fact be communicating with another assisted user. In single line systems, one line links the HU to the relay and one line (e.g., the single line) links to the relay to the AU device. Voice from the HU is presented to a relay call assistant (CA) who transcribes the voice-to-text and then the text is transmitted to the AU device to be displayed. The HU's voice is also, in at least some cases, carried or passed through the relay to the AU device to be broadcast to the AU.

[0008]   The other captioning service type described in the '835 patent is a two line system. In a two line system a hearing user's telephone is directly linked to an assisted user's device for voice communications between the AU and the HU. When captioning is required, the AU can select a captioning control button on the AU device to link to the relay and provide the HU's voice to the relay on a first line. Again, a relay CA listens to the HU voice message and transcribes the voice message into text which is transmitted back to the AU device on a second line to be displayed to the AU. One of the primary advantages of the two line system over one line systems is that the AU can add captioning to an on-going call. This is important as many AUs are only partially impaired and may only want captioning when absolutely necessary. The option to not have captioning is also important in cases where an AU device can be used as a normal

telephone and where non-assisted users (e.g., a spouse that has good hearing capability) that do not need captioning may also use the AU device.

[0009]   With any relay system, the primary factors for determining the value of the system are accuracy, speed and cost to provide the service. Regarding accuracy, text should accurately represent voice messages from hearing users so that an AU reading the text has an accurate understanding of the meaning of the message. Erroneous words provide inaccurate messages and also can cause confusion for an AU reading the messages.

[0010]   Regarding speed, ideally text is presented to an AU simultaneously with the voice message corresponding to the text so that an AU sees text associated with a message as the message is heard. In this regard, text that trails a voice message by several seconds can cause confusion. Current systems present captioned text relatively quickly (e.g. 1-3 seconds after the voice message is broadcast) most of the time. However, at times a CA can fall behind when captioning so that longer delays (e.g., 10-15 seconds) occur.

[0011]   Regarding cost, existing systems require a unique and highly trained CA for each communication session. In known cases CAs need to be able to speak clearly and need to be able to type quickly and accurately. CA jobs are also relatively high pressure jobs and therefore turnover is relatively high when compared jobs in many other industries which further increases the costs associated with operating a relay.

[0012]   One innovation that has increased captioning speed appreciably and that has reduced the costs associated with captioning at least somewhat has been the use of voice-to-text transcription software by relay CAs. In this regard, early relay systems required CAs to type all of the text presented via an AU device. To present text as quickly as possible after broadcast of an associated voice message, highly skilled typists were required. During normal conversations people routinely speak at a rate between 110 to 150 words per minute. During a conversation between an AU and an HU, typically only about half the words voiced have to be transcribed (e.g., the AU typically communicates to the HU during half of a session). This means that to keep up with transcribing the HU's portion of a typical conversation a CA has to be able to type at around 55 to 75 words per minute. To this end, most professional typists type at around 50 to 80 words per minute and therefore can keep up with a normal conversation for at least some time. Professional typists are relatively expensive. In

addition, despite being able to keep up with a conversation most of the time, at other times (e.g., during long conversations or during particularly high speed conversations) even professional typists fall behind transcribing real time text and more substantial delays occur.

[0013]   In relay systems that use voice-to-text transcription software trained to a CA's voice, a CA listens to an HU's voice and revoices the HU's voice message to a computer running the trained software. The software, being trained to the CA's voice, transcribes the revoiced message much more quickly than a typist can type text and with only minimal errors. In many respects revoicing techniques for generating text are easier and much faster to learn than high speed typing and therefore training costs and the general costs associated with CA's are reduced appreciably. In addition, because revoicing is much faster than typing in most cases, voice-to-text transcription can be expedited appreciably using revoicing techniques.

[0014]   At least some prior systems have contemplated further reducing costs associated with relay services by replacing CA's with computers running voice-to-text software to automatically convert HU voice messages to text. In the past there have been several problems with this solution which have resulted in no one implementing a workable system. First, most voice messages (e.g., an HU's voice message) delivered over most telephone lines to a relay are not suitable for direct voice-to-text transcription software. In this regard, automated transcription software on the market has been tuned to work well with a voice signal that includes a much larger spectrum of frequencies than the range used in typical phone communications. The frequency range of voice signals on phone lines is typically between 300 and 3000 Hz. Thus, automated transcription software does not work well with voice signals delivered over a telephone line and large numbers of errors occur. Accuracy further suffers where noise exists on a telephone line which is a common occurrence.

[0015]   Second, most automated transcription software has to be trained to the voice of a speaker to be accurate. When a new HU calls an AU's device, there is no way for a relay to have previously trained software to the HU voice and therefore the software cannot accurately generate text using the HU voice messages.

[0016]   Third, many automated transcription software packages use context in order to generate text from a voice message. To this end, the words around each word in a voice message can be used by software as context for determining which word has

been uttered. To use words around a first word to identify the first word, the words around the first word have to be obtained. For this reason, many automated transcription systems wait to present transcribed text until after subsequent words in a voice message have been transcribed so that context can be used to correct prior words before presentation. Systems that hold off on presenting text to correct using subsequent context cause delay in text presentation which is inconsistent with the relay system need for real time or close to real time text delivery.

BRIEF SUMMARY OF THE INVENTION

[0017]   It has been recognized that a hybrid semi-automated system can be provided where, when acceptable accuracy can be achieved using automated transcription software, the system can automatically use the transcription software to transcribe HU voice messages to text and when accuracy is unacceptable, the system can patch in a human CA to transcribe voice messages to text. Here, it is believed that the number of CAs required at a large relay facility may be reduced appreciably (e.g., 30% or more) where software can accomplish a large portion of transcription to text. In this regard, not only is the automated transcription software getting better over time, in at least some cases the software may train to an HU's voice and the vagaries associated with voice messages received over a phone line (e.g., the limited 300 to 3000 Hz range) during a first portion of a call so that during a later portion of the call accuracy is particularly good. Training may occur while and in parallel with a CA manually (e.g., via typing, revoicing, etc.) transcribing voice-to-text and, once accuracy is at an acceptable threshold level, the system may automatically delink from the CA and use the text generated by the software to drive the AU display device.

[0018]   It has been recognized that in a relay system there are at least two processor that may be capable of performing automated voice recognition processes and therefore that can handle the automated voice recognition part of a triage process involving a call assistant. To this end, in most cases either a relay processor or an assisted user's device processor may be able to perform the automated transcription portion of a hybrid process. For instance, in some cases an assisted user's device will perform automated transcription in parallel with a relay assistant generating call assistant generated text where the relay and assisted user's device cooperate provide

text and assess when the call assistant should be cut out of a call with the automated text replacing the call assistant generated text.

[0019]   In other cases where a hearing user's communication device is a computer or includes a processor capable of transcribing voice messages to text, a hearing user's device may generated automated text in parallel with a call assistant generating text and the hearing user's device and the relay may cooperate to provide text and determine when the call assistant should be cut out of the call.

[0020]   Regardless of which device is performing automated captioning, the call assistant generated text may be used to assess accuracy of the automated text for the purpose of determining when the call assistant should be cut out of the call. In addition, regardless of which device is performing automated text captioning, the call assistant generated text may be used to train the automated voice-to-text software or engine on the fly to expedite the process of increasing accuracy until the call assistant can be cut out of the call.

[0021]   It has also been recognized that there are times when a hearing impaired person is listening to a hearing user's voice without an assisted user's device providing simultaneous text when the hearing user is confused and would like transcription of recent voice messages of the hearing user. For instance, where an assisted user uses an assisted user's device to carry on a non-captioned call and the assisted user has difficulty understanding a voice message so that the assisted user initiates a captioning service to obtain text for subsequent voice messages. Here, while text is provided for subsequent messages, the assisted user still cannot obtain an understanding of the voice message that prompted initiation of captioning. As another instance, where call assistant generated text lags appreciably behind a current hearing user's voice message an assisted user may request that the captioning catch up to the current message.

[0022]   To provide captioning of recent voice messages in these cases, in at least some embodiments of this disclosure an assisted user's device stores a hearing user's voice messages and, when captioning is initiated or a catch up request is received, the recorded voice messages are used to either automatically generate text or to have a call assistant generate text corresponding to the recorded voice messages.

[0023]   In at least some cases when automated software is trained to a hearing user's voice, a voice model for the hearing user that can be used subsequently to tune

automated software to transcribe the hearing user's voice may be stored along with a voice profile for the hearing user that can be used to distinguish the hearing user's voice from other hearing users. Thereafter, when the hearing user calls an assisted user's device again, the profile can be used to indentify the hearing user and the voice model can be used to tune the software so that the automated software can immediately start generating highly accurate or at least relatively more accurate text corresponding to the hearing user's voice messages.

[0024]   To the accomplishment of the foregoing and related ends, the invention, then, comprises the features hereinafter fully described. The following description and the annexed drawings set forth in detail certain illustrative aspects of the invention. However, these aspects are indicative of but a few of the various ways in which the principles of the invention can be employed. Other aspects, advantages and novel features of the invention will become apparent from the following detailed description of the invention when considered in conjunction with the drawings.

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWINGS

[0025]   Fig. 1 is a schematic showing various components of a communication system including a relay that may be used to perform various processes and methods according to at least some aspects of the present invention;

[0026]   Fig. 2 is a schematic of the relay server shown in Fig. 1;

[0027]   Fig. 3 is a flow chart showing a process whereby an automated voice-to-text engine is used to generate automated text in parallel with a call assistant generating text where the automated text is used instead of call assistant generated text to provide captioning an assisted user's device once an accuracy threshold has been exceeded;

[0028]   Fig. 4 is a sub-process that maybe substituted for a portion of the process shown in Fig. 3 whereby a control assistant can determine whether or not the automated text takes over the process after the accuracy threshold has been achieved;

[0029]   Fig. 5 is a sub-process that may be added to the process shown in Fig. 3 wherein, upon an assisted user's requesting help, a call is linked to a second call assistant for correcting the automated text;

[0030]   Fig. 6 is a process whereby an automated voice-to-text engine is used to fill in text for a hearing user's voice messages that are skipped over by a call assistant when an assisted user requests instantaneous captioning of a current message;

[0031]   Fig. 7 is a process whereby automated text is automatically used to fill in captioning when transcription by a call assistant lags behind a hearing user's voice messages by a threshold duration;

[0032]   Fig. 8 is a flow chart illustrating a process whereby text is generated for a hearing user's voice messages that precede a request for captioning services;

[0033]   Fig. 9 is a flow chart illustrating a process whereby voice messages prior to a request for captioning service are automatically transcribed to text by an automated voice-to-text engine;

[0034]   Fig. 10 is a flow chart illustrating a process whereby an assisted user's device processor performs transcription processes until a request for captioning is received at which point the assisted user's device presents texts related to hearing user voice messages prior to the request and on going voice messages are transcribed via a relay;

[0035]   Fig. 11 is a flow chart illustrating a process whereby an assisted user's device processor generates automated text for a hear user's voice messages which is presented via a display to an assisted user and also transmits the text to a call assistant at a relay for correction purposes;

[0036]   Fig. 12 is a flow chart illustrating a process whereby high definition digital voice messages and analog voice messages are handled differently at a relay;

[0037]   Fig. 13 is a process similar to Fig. 12, albeit where an assisted user also has the option to link to a call assistant for captioning service regardless of the type of voice message received;

[0038]   Fig. 14 is a flow chart that may be substituted for a portion of the process shown in Fig. 3 whereby voice models and voice profiles are generated for frequent hearing user's that communicate with an assisted user where the models and profiles can be subsequently used to increase accuracy of a transcription process;

[0039]   Fig. 15 is a flow chart illustrating a process similar to the sub-process shown in Fig. 14 where voice profiles and voice models are generated and stored for subsequent use during transcription;

[0040]   Fig. 16 is a flow chart illustrating a sub-process that may be added to the process shown in Fig. 15 where the resulting process calls for training of a voice model at each of an assisted user's device and a relay;

[0041]   Fig. 17 is a schematic illustrating a screen shot that may be presented via an assisted user's device display screen;

[0042]   Fig. 18 is similar to Fig. 17, albeit showing a different screen shot;

[0043]   Fig. 19 is a process that may be performed by the system shown in Fig. 1 where automated text is generated for line check words and is presented to an assisted user immediately upon identification of the words;

[0044]   Fig. 20 is similar to Fig. 17, albeit showing a different screen shot; and

[0045]   Fig. 21 is a flow chart illustrating a method whereby an automated voice-to-text engine is used to identify errors in call assistant generated text which can be highlighted and can be corrected by a call assistant.

[0046]   While the invention is susceptible to various modifications and alternative forms, specific embodiments thereof have been shown by way of example in the drawings and are herein described in detail. It should be understood, however, that the description herein of specific embodiments is not intended to limit the invention to the particular forms disclosed, but on the contrary, the intention is to cover all modifications, equivalents, and alternatives falling within the spirit and scope of the invention as defined by the appended claims.

DETAILED DESCRIPTION OF THE INVENTION

[0047]   The various aspects of the subject invention are now described with reference to the annexed drawings, wherein like reference numerals correspond to similar elements throughout the several views. It should be understood, however, that the drawings and detailed description hereafter relating thereto are not intended to limit the claimed subject matter to the particular form disclosed. Rather, the intention is to cover all modifications, equivalents, and alternatives falling within the spirit and scope of the claimed subject matter.

[0048]   As used herein, the terms "component," "system" and the like are intended to refer to a computer-related entity, either hardware, a combination of hardware and software, software, or software in execution. For example, a component may be, but is

not limited to being, a process running on a processor, a processor, an object, an executable, a thread of execution, a program, and/or a computer. By way of illustration, both an application running on a computer and the computer can be a component. One or more components may reside within a process and/or thread of execution and a component may be localized on one computer and/or distributed between two or more computers or processors.

[0049]   The word "exemplary" is used herein to mean serving as an example, instance, or illustration. Any aspect or design described herein as "exemplary" is not necessarily to be construed as preferred or advantageous over other aspects or designs.

[0050]   Furthermore, the disclosed subject matter may be implemented as a system, method, apparatus, or article of manufacture using standard programming and/or engineering techniques to produce software, firmware, hardware, or any combination thereof to control a computer or processor based device to implement aspects detailed herein. The term "article of manufacture" (or alternatively, "computer program product") as used herein is intended to encompass a computer program accessible from any computer-readable device, carrier, or media. For example, computer readable media can include but are not limited to magnetic storage devices (e.g., hard disk, floppy disk, magnetic strips . . .), optical disks (e.g., compact disk (CD), digital versatile disk (DVD) . . .), smart cards, and flash memory devices (e.g., card, stick). Additionally it should be appreciated that a carrier wave can be employed to carry computer-readable electronic data such as those used in transmitting and receiving electronic mail or in accessing a network such as the Internet or a local area network (LAN). Of course, those skilled in the art will recognize many modifications may be made to this configuration without departing from the scope or spirit of the claimed subject matter.

[0051]   Referring now to the drawings wherein like reference numerals correspond to similar elements throughout the several views and, more specifically, referring to Fig. 1, the present disclosure will be described in the context of an exemplary communication system 10 including an assisted user's communication device 12, a hearing user's telephone or other type communication device 14, and a relay 16. The AU's device 12 is linked to the hearing user's device 14 via any network connection capable of facilitating a voice call between an the AU and the HU. For

instance, the link may be a conventional telephone line, a network connection such as an internet connection or other network, a wireless connection, etc. AU device 12 includes a keyboard 20, a display screen 18 and a handset 22. Keyboard 20 can be used to dial any telephone number to initiate a call and, in at least some cases, includes other keys or may be controlled to present virtual buttons via screen 28 for controlling various functions that will be described in greater detail below. Other identifiers such as IP addresses or the like may also be used in at least some cases to initiate a call. Screen 18 includes a flat panel display screen for displaying, among other things, text transcribed from a voice message generated using HU's device 14, control icons or buttons, caption feedback signals, etc. Handset 22 includes a speaker for broadcasting a hearing user's voice messages to an assisted user and a microphone for receiving a voice message from an assisted user for delivery to the hearing user's device 14. Assisted user device 12 may also include a second loud speaker so that device 12 can operate as a speaker phone type device. Although not shown, device 12 further includes a processor and a memory for storing software run by the processor to perform various functions that are consistent with at least some aspects of the present disclosure. Device 12 is also linked or is linkable to relay 16 via any communication network including a phone network, a wireless network, the internet or some other similar network, etc.

[0052]   Hearing user's device 14 includes a communication device (e.g., a telephone) including a keyboard for dialing phone numbers and a handset including a speaker and a microphone for communication with other devices. In other embodiments device 14 may include a computer, a smart phone, a smart tablet, etc., that can facilitate audio communications with other devices. Devices 12 and 14 may use any of several different communication protocols including analog or digital protocols, a VOIP protocol or others.

[0053]   Referring still to Fig. 1, relay 16 includes, among other things, a relay server 30 and a plurality of call assistant work stations 32, 34, etc. Each of the call assistant work stations 32, 34, etc., is similar and operates in a similar fashion and therefore only station 32 is described here in any detail. Station 32 includes a display screen 50, a keyboard 52 and a headphone/microphone headset 54. Screen 50 may be any type of electronic display screen for presenting information including text transcribed from a hearing user's voice signal or message. In most cases screen 50 will

present a graphical user interface with on screen tools for editing text that appears on the screen. One text editing system is described in US patent No. 7,164,753 which issued on January 16, 2007 which is titled "Real Time Transcription Correction System" and which is incorporated herein in its entirety.

[0054]   Keyboard 52 is a standard text entry QUERTY type keyboard and can be use to type text or to correct text presented on displays screen 50. Headset 54 includes a speaker in an ear piece and a microphone in a mouth piece and is worn by a call assistant. The headset enables a call assistant to listen to the voice of a hearing user and the microphone enables the call assistant to speak voice messages into the relay system such as, for instance, revoiced messages from a hearing user to be transcribed into text. For instance, typically during a call between a hearing user on device 14 and an assisted user on device 12, the hearing user's voice messages are presented to a call assistant via headset 54 and the call assistant revoices the messages into the relay system using headset 54. Software trained to the voice of the call assistant transcribes the assistant's voice messages into text which is presented on display screen 52. The call assistant then uses keyboard 52 and/or headset 54 to make corrections to the text on display 50. The corrected text is then transmitted to the assisted user's device 12 for display on screen 18. In the alternative, the text may be transmitted prior to correction to the assisted user's device 12 for display and corrections may be subsequently transmitted to correct the displayed text via in-line corrections where errors are replaced by corrected text.

[0055]   Although not shown, call assistant work station 32 may also include a foot pedal or other device for controlling the speed with which voice messages are played via headset 54 so that the call assistant can slow or even stop play of the messages while the assistant either catches up on transcription or correction of text.

[0056]   Referring still to Fig. 1 and also to Fig. 2, server 30 is a computer system that includes, among other components, at least a first processor 56 linked to a memory or database 58 where software run by server 56 to facilitate various functions that are consistent with at least some aspects of the present disclosure is stored. The software stored in memory 58 includes pre-trained call assistant voice-to-text transcription software 60 for each call assistant where call assistant specific software is trained to the voice of an associated call assistant thereby increasing the accuracy of transcription activities. For instance, Naturally Speaking continuous speech recognition software by

Dragon, Inc. may be pre-trained to the voice of a specific call assistant and then used to transcribe voice messages voiced by the call assistant into text.

[0057]   In addition to the call assistant trained software, a voice-to- text software program 62 that is not pre-trained to a CA's voice and instead that trains to any voice on the fly as voice messages are received is stored in memory 58. Again, Naturally Speaking software that can train on the fly may be used for this purpose.

[0058]   Moreover, software 64 that automatically performs one of several different types of triage processes to generate text from voice messages accurately, quickly and in a relatively cost effective manner is stored in memory 58. The triage programs are described in detail hereafter.

[0059]   One issue with existing relay systems is that each call is relatively expensive to facilitate. To this end, in order to meet required accuracy standards for text caption calls, each call requires a dedicated call assistant. While automated voice-to-text systems that would not require a call assistant have been contemplated, none has been implemented because of accuracy and speed problems.

[0060]   One aspect of the present disclosure is related to a system that is semi-automated wherein a call assistant is used when accuracy of an automated system is not at required levels and the assistant is cut out of a call automatically or manually when accuracy of the automated system meets or exceeds accuracy standards. For instance, in at least some cases a call assistant will be assigned to every new call linked to a relay and the call assistant will transcribe voice-to-text as in an existing system. Here, however, the difference will be that, during the call, the voice of a hearing user will also be processed by server 30 to automatically transcribe the hearing user's voice messages to text (e.g., into "automated text"). Server 30 compares corrected text generated by the call assistant to the automated text to identify errors in the automated text. Server 30 uses identified errors to train the automated voice-to-text software to the voice of the hearing user. During the beginning of the call the software trains to the hearing user's voice and accuracy increases over time as the software trains. At some point the accuracy increases until required accuracy standards are met. Once accuracy standards are met, server 30 is programmed to automatically cut out the call assistant and start transmitting the automated text to the assisted user's device 12.

[0061]   In at least some cases, when a call assistant is cut out of a call, the system may provide a "Help" or an "Assist" or "Assistance Request" type button (see

68 in Fig. 1) to an assisted user so that, if the assisted user recognizes that the automated text has too many errors for some reason, the assisted user can request a link to a call assistant to increase transcription accuracy (e.g., generated an assistance request). In some cases the help button may be a persistent mechanical button on the assisted user's device 12. In the alternative the help button may be a virtual on screen icon (e.g., see 68 in Fig. 1) and screen 18 may be a touch sensitive screen so that contact with the virtual button can be sensed. Where the help button is virtual, the button may only be presented after the system switches from providing call assistant generated text to an assisted user's device to providing automated text to the assisted user's device to avoid confusion (e.g., avoid a case where an assisted user is already receiving call assistant generated text but thinks, because of a help button, that even better accuracy can be achieved in some fashion). Thus, while call assistant generated text is displayed on an assisted user's device 12, no "help" button is presented and after automated text is presented, the "help" button is presented. After the help button is selected and a call assistant is re-linked to the call, the help button is again removed from the assisted user's device display 18 to avoid confusion.

[0062]   Referring now to Figs. 2 and 3, a method or process 70 is illustrated that may be performed by server 30 to cut out a call assistant when automated text reaches an accuracy level that meets a standard threshold level. Referring also to Fig. 1, at block 72, help and auto flags are each set to a zero value. The help flag indicates that an assisted user has selected a help or assist button via the assisted user's device 12 because of a perception that too many errors are occurring in transcribed text. The auto flag indicates that automated text accuracy has exceeded a standard threshold requirement. Zero values indicate that the help button has not been selected and that the standard requirement has yet to be met and one values indicate that the button has been selected and that the standard requirement has been met.

[0063]   Referring still to Figs. 1 and 2, at block 74, during a phone call between a hearing user using device 14 and an assisted user using device 12, the hearing user's voice messages are transmitted to server 30 at relay 16. Upon receiving the hearing user's voice messages, server 30 checks the auto and help flags at blocks 76 and 84, respectively. At least initially the auto flag will be set to zero at block 76 meaning that automated text has not reached the accuracy standard requirement and therefore control passes down to block 78 where the hearing user's voice messages are provided

to a call assistant. At block 80 the call assistant listens to the hearing user's voice messages and generates text corresponding thereto by either typing the messages, revoicing the messages to voice-to-text transcription software trained to the call assistant's voice or a combination of both. Text generated is presented on screen 50 and the call assistant makes corrections to the text using keyboard 52 and/or headset 54 at block 80. At block 82 the call assistant generated text is transmitted to assisted user device 12 to be displayed for the assisted user on screen 18.

[0064]   Referring again to Figs. 1 and 2, at block 84, at least initially the help flag will be set to zero indicating that the assisted user has not requested additional captioning assistance. In fact, at least initially the "help" button 68 may not be presented to an assisted user as call assistant generated text is initially presented. Where the help flag is zero at block 84, control passes to block 86 where the hearing user's voice messages are fed to voice-to-text software run by server 30 that has not been previously trained to any particular voice. At block 88 the software automatically converts the hearing user's voice-to-text generating automated text. At block 90 server 30 compares the call assistant generated text to the automated text to identify errors in the automated text. At block 92 server 30 uses the errors to train the voice-to-text software for the hearing user's voice. In this regard, for instance, where an error is identified, server 30 modifies the software so that the next time the utterance that resulted in the error occurs, the software will generate the word or words that the call assistant generated for the utterance. Other ways of altering or training the voice-to-text software are well known in the art and any way of training the software may be used at block 92.

[0065]   After block 92 control passes to block 94 where server 30 monitors for a selection of the "help" button 68 by the assisted user. If the help button has not been selected, control passes to block 96 where server 30 compares the accuracy of the automated text to a threshold standard accuracy requirement. For instance, the standard requirement may require that accuracy be great than 96% measured over at least a most recent forty-five second period or a most recent 100 words uttered by a hearing user, whichever is longer. Where accuracy is below the threshold requirement, control passes back up to block 74 where the process described above continues. At block 96, once the accuracy is greater than the threshold requirement, control passes to block 98 where the auto flag is set to one indicating that the system should start using

the automated text and delink the call assistant from the call to free up the assistant to handle a different call. A virtual "help" button may also be presented via the assisted user's display 18 at this time. Next, at block 100, the call assistant is delinked from the call and at block 102 the processor generated automated text is transmitted to the AU device to be presented on display screen 18.

[0066]   Referring again to block 74, the hearing user's voice is continually received during a call and at block 76, once the auto flag has been set to one, the lower portion of the left hand loop including blocks 78, 80 and 82 is cut out of the process as control loops back up to block 74.

[0067]   Referring again to Block 94, if, during an automated portion of a call when automated text is being presented to the assisted user, the assisted user decides that there are too many errors in the transcription presented via display 18 and the assisted user selects the "help" button 68 (see again Fig. 1), control passes to block 104 where the help flag is set to one indicating that the assisted user has requested the assistance of a call assistant and the auto flag is reset to zero indicating that call assistant generated text will be used to drive the assisted user's display 18 instead of the automated text. Thereafter control passes back up to block 74. Again, at block 76, with the auto flag set to zero the next time through decision block 76, control passes back down to block 78 where the call is again linked to a call assistant for transcription as described above. In addition, the next time through block 84, because the help flag is set to one, control passes back up to block 74 and the automated text loop including blocks 86 through 104 is effectively cut out of the rest of the call.

[0068]   In at least some embodiments, there will be a short delay (e.g., 5 to 10 seconds in most cases) between setting the flags at block 104 and stopping use of the automated text so that a new call assistant can be linked up to the call and start generating call assistant generated text prior to halting the automated text. In these cases, until the call assistant is linked and generating text for at least a few seconds (e.g., 3 seconds), the automated text will still be used to drive the assisted user's display 18. The delay may either be a pre-defined delay or may have a case specific duration that is determined by server 30 monitoring call assistant generated text and switching over to the call assistant generated text once the call assistant is up to speed.

[0069]   In some embodiments, prior to delinking a call assistant from a call at block 100, server 30 may store a call assistant identifier along with a call identifier for

the call. Thereafter, if an assisted user requests help at block 94, server 30 may be programmed to identify if the call assistant previously associated with the call is available (e.g. not handling another call) and, if so, may re-link to the call assistant at block 78. In this manner, if possible, a call assistant that has at least some context for the call can be linked up to restart transcription services.

[0070] In some embodiments it is contemplated that after an assisted user has selected a help button to receive call assistance, the call will be completed with a call assistant on the line. In other cases it is contemplated that server 30 may, when a call assistant is re-linked to a call, start a second triage process to attempt to delink the call assistant a second time if a threshold accuracy level is again achieved. For instance, in some cases, midstream during a call, a second hearing user may start communicating with the assisted user via the hearing user's device. For instance, a child may yield the hearing user's device 14 to a grandchild that has a different voice profile causing the assisted user to request help from a call assistant because of perceived text errors. Here, after the hand back to the call assistant, server 30 may start training on the grandchild's voice and may eventually achieve the threshold level required. Once the threshold again occurs, the call assistant may be delinked a second time so that automated text is again fed to the assisted user's device.

[0071] As another example text errors in automated text may be caused by temporary noise in one or more of the lines carrying the hearing user's voice messages to relay 16. Here, once the noise clears up, automated text may again be a suitable option. Thus, here, after an assisted user requests call assistant help, the triage process may again commence and if the threshold accuracy level is again exceeded, the call assistant may be delinked and the automated text may again be used to drive the assisted user's device 12. While the threshold accuracy level may be the same each time through the triage process, in at least some embodiments the accuracy level may be changed each time through the process. For instance, the first time through the triage process the accuracy threshold may be 96%. The second time through the triage process the accuracy threshold may be raised to 98%.

[0072] In at least some embodiments, when the automated text accuracy exceeds the standard accuracy threshold, there may be a short transition time during which a call assistant on a call observes automated text while listening to a hearing user's voice message to manually confirm that the handover from call assistant

generated text to automated text is smooth. During this short transition time, for instance, the call assistant may watch the automated text on her workstation screen 50 and may correct any errors that occur during the transition. In at least some cases, if the call assistant perceives that the handoff does not work or the quality of the automated text is poor for some reason, the call assistant may opt to retake control of the transcription process.

[0073]   One sub-process 120 that may be added to the process shown in Fig. 3 for managing a call assistant to automated text handoff is illustrated in Fig. 4. Referring also to Figs. 1 and 2, at block 96 in Fig. 3, if the accuracy of the automated text exceeds the accuracy standard threshold level, control may pass to block 122 in Fig. 4. At block 122, a short duration transition timer (e.g. 10-15 seconds) is started. At block 124 automated text (e.g., text generated by feeding the hearing user's voice messages directly to voice-to-text software) is presented on the call assistant's display 50. At block 126 an on screen a "Retain Control" icon or virtual button is provided to the call assistant via the assistant's display screen 50 which can be selected by the call assistant to forego the handoff to the automated voice-to-text software. At block 128, if the "Retain Control" icon is selected, control passes to block 132 where the help flag is set to one and then control passes back up to block 76 in Fig. 3 where the call assistant process for generating text continues as described above. At block 128, if the call assistant does not select the "Retain Control" icon, control passes to block 130 where the transition timer is checked. If the transition timer has not timed out control passes back up to block 124. Once the timer times out at block 130, control passes back to block 98 in Fig. 3 where the auto flag is set to one and the call assistant is delinked from the call.

[0074]   In at least some embodiments it is contemplated that after voice-to-text software takes over the transcription task and the call assistant is delinked from a call, server 30 itself may be programmed to sense when transcription accuracy has degraded substantially and the server 30 may cause a re-link to a call assistant to increase accuracy of the text transcription. For instance, server 30 may assign a confidence factor to each word in the automated text based on how confident the server is that the word has been accurately transcribed. The confidence factors over a most recent number of words (e.g., 100) or a most recent period (e.g., 45 seconds) may be averaged and the average used to assess an overall confidence factor for transcription

accuracy. Where the confidence factor is below a threshold level, server 30 may re-link to a call assistant to increase transcription accuracy. The automated process for re-linking to a call assistant may be used instead of or in addition to the process described above whereby an assisted user selects the "help" button to re-link to a call assistant.

[0075]   In at least some cases when an assisted user selects a "help" button to re-link to a call assistant, partial call assistance may be provided instead of full call assistant service. For instance, instead of adding a call assistant that transcribes a hearing user's voice messages and then corrects errors, a call assistant may be linked only for correction purposes. The idea here is that while software trained to a hearing user's voice may generate some errors, the number of errors after training will still be relatively small in most cases even if objectionable to an assisted user. In at least some cases call assistants may be trained to have different skill sets where highly skilled and relatively more expensive to retain call assistants are trained to re-voice hearing user voice messages and correct the resulting text and less skilled call assistants are trained to simply make corrections to automated text. Here, initially all calls may be routed to highly skilled revoicing or "transcribing" call assistants and all re-linked calls may be routed to less skilled "corrector" call assistants.

[0076]   A sub-process 134 that may be added to the process of Fig. 3 for routing re-linked calls to a corrector call assistant is shown in Fig. 5. Referring also to Figs. 1 and 3, at decision block 94, if an assisted user selects the help button, control may pass to block 136 in Fig. 3 where the call is linked to a second corrector call assistant. At block 138 the automated text is presented to the second call assistant via the call assistant's display 50. At block 140 the second call assistant listens to the voice of the hearing user and observes the automated text and makes corrections to errors perceived in the text. At block 142 server 30 transmits the corrected automated text to the assisted user's device for display via screen 18. After block 142 control passes back up to block 76 in Fig. 2.

[0077]   In some cases where a call assistant generates text that drives an assisted user's display screen 18 (see again Fig. 1), for one reason or another the call assistant's transcription to text may fall behind the hearing user's voice message stream by a substantial amount. For instance, where a hearing user is speaking quickly, is using odd vocabulary, and/or has an unusual accent that is hard to understand, call

assistant transcription may fall behind a voice message stream by 20 seconds, 40 seconds or more.

[0078]   In many cases when captioning falls behind, an assisted user can perceive that presented text has fallen far behind broadcast voice messages from a hearing user based on memory of recently broadcast voice message content and observed text. For instance, an assisted user may recognize that currently displayed text corresponds to a portion of the broadcast voice message that occurred thirty seconds ago. In other cases some captioning delay indicator may be presented via an assisted user's device display 18. For instance, see Fig. 17 where captioning delay is indicated in two different ways on a display screen 18. First, text 212 indicates an estimated delay in seconds (e.g., 24 second delay). Second, at the end of already transcribed text 214, blanks 216 for words already voiced but yet to be transcribed may be presented to give an assisted user a sense of how delayed the captioning process has become.

[0079]   When an assisted user perceives that captioning is too far behind or when the user cannot understand a recently broadcast voice message, the assisted user may want the text captioning to skip ahead to the currently broadcast voice message. For instance, if an assisted user had difficulty hearing the most recent five seconds of a hearing user's voice message and continues to have difficulty hearing but generally understood the preceding 25 seconds, the assisted user may want the captioning process to be re-synced with the current hearing user's voice message so that the assisted user's understanding of current words is accurate.

[0080]   Here, however, because the assisted user could not understand the most recent 5 seconds of broadcast voice message, a re-sync with the current voice message would leave the assisted user with at least some void in understanding the conversation (e.g., at least 5 the most recent 5 seconds of misunderstood voice message would be lost). To deal with this issue, in at least some embodiments, it is contemplated that server 30 may run automated voice-to-text software on a hearing user's voice message simultaneously with a call assistant generating text from the voice message and, when an assisted user requests a "catch-up" or "re-sync" of the transcription process to the current voice message, server 30 may provide "fill in" automated text corresponding to the portion of the voice message between the most recent call assistant generated text and the instantaneous voice message which may be

provided to the assisted user's device for display and also, optionally, to the call assistant's display screen to maintain context for the call assistant. In this case, while the fill in automated text may have some errors, the fill in text will be better than no text for the associated period and can be referred to by the assisted user to better understand the voice messages.

[0081]   In cases where the fill in text is presented on the call assistant's display screen, the call assistant may correct any errors in the fill in text. This correction and any error correction by a call assistant for that matter may be made prior to transmitting text to the assisted user's device or subsequent thereto. Where corrected text is transmitted to an assisted user's device subsequent to transmission of the original error prone text, the assisted user's device corrects the errors by replacing the erroneous text with the corrected text.

[0082]   Because it is often the case that assisted users will request a re-sync only when they have difficulty understanding words, server 30 may only present automated fill in text to an assisted user corresponding to a pre-defined duration period (e.g., 8 seconds) that precedes the time when the re-sync request occurs. For instance, consistent with the example above where call assistant captioning falls behind by thirty seconds, an assisted user may only request re-sync at the end of the most recent five seconds as inability to understand the voice message may only be an issue during those five seconds. By presenting the most recent eight seconds of automated text to the assisted user, the user will have the chance to read text corresponding to the misunderstood voice message without being inundated with a large segment of automated text to view. Where automated fill in text is provided to an assisted user for only a pre-defined duration period, the same text may be provided for correction to the call assistant.

[0083]   Referring now to Fig. 6, a method 190 by which an assisted user requests a re-sync of the transcription process to current voice messages when call assistant generated text falls behind current voice messages is illustrated. Referring also to Fig. 1, at block 192 a hearing user's voice messages are received at relay 16. After block 192, control passes down to each of blocks 194 and 200 where two simultaneous sub-processes occur in parallel. At block 194, the hearing user's voice messages are stored in a rolling buffer. The rolling buffer may, for instance, have a two minute duration so that the most recent two minutes of a hearing user's voice messages are always stored.

At block 196, a call assistant listens to the hearing user's voice message and transcribes text corresponding to the messages via re-voicing to software trained to the call assistant's voice, typing, etc. At block 198 the call assistant generated text is transmitted to assisted user's device 12 to be presented on display screen 18 after which control passes back up to block 192. Text correction may occur at block 196 or after block 198.

[0084]   Referring again to Fig. 6, at process block 200, the hearing user's voice is fed directly to voice-to-text software run by server 30 which generates automated text at block 202. Although not shown in Fig. 5, after block 202, server 30 may compare the automated text to the call assistant generated text to identify errors and may use those errors to train the software to the hearing user's voice so that the automated text continues to get more accurate as a call proceeds.

[0085]   Referring still to Figs. 1 and 6, at decision block 204, controller 30 monitors for a catch up or re-sync command received via the assisted user's device 12 (e.g., via selection of an on-screen virtual "catch up" button 220, see again Fig. 17). Where no catch up or re-sync command has been received, control passes back up to block 192 where the process described above continues to cycle. At block 204, once a re-sync command has been received, control passes to block 206 where the buffered voice messages are skipped and a current voice message is presented to the ear of the call assistant to be transcribed. At block 208 the automated text corresponding to the skipped voice message segment is filled in to the text on the call assistant's screen for context and at block 210 the fill in text is transmitted to the assisted user's device for display.

[0086]   Where automated text is filled in upon the occurrence of a catch up process, the fill in text may be visually distinguished on the assisted user's screen and/or on the call assistant's screen. For instance, fill in text may be highlighted, underlined, bolded, shown in a distinct font, etc. for example, see Fig. 18 that shows fill in text 222 that is underlined to visually distinguish. See also that the captioning delay 212 has been updated. In some cases, fill in text corresponding to voice messages that occur after or within some pre-defined period prior to a re-sync request may be distinguished in yet a third way to point out the text corresponding to the portion of a voice message that the assisted user most likely found interesting (e.g., the portion that prompted selection of the re-sync button). For instance, where 24 previous seconds of

text are filled in when a re-sync request is initiated, all 24 seconds of fill in text may be underlined and the 8 seconds of text prior to the re-sync request may also be highlighted in yellow. See in Fig. 18 that some of the fill in text is shown in a phantom box 226 to indicate highlighting.

[0087]   In at least some cases it is contemplated that server 30 may be programmed to automatically determine when call assistant generated text substantially lags a current voice message from a hearing user and server 30 may automatically skip ahead to re-sync a call assistant with a current message while providing automated fill in text corresponding to intervening voice messages. For instance, server 30 may recognize when call assistant generated text is more than thirty seconds behind a current voice message and may skip the voice messages ahead to the current message while filling in automated text to fill the gap. In at least some cases this automated skip ahead process may only occur after at least some (e.g., 2 minutes) training to a hearing user's voice so ensure that minimal errors are generated in the fill in text.

[0088]   A method 150 for automatically skipping to a current voice message in a buffer when a call assistant falls to far behind is shown in Fig. 7. Referring also to Fig. 1, at block 152, a hearing user's voice messages are received at relay 16. After block 152, control passes down to each of blocks 154 and 162 where two simultaneous sub-processes occur in parallel. At block 154, the hearing user's voice messages are stored in a rolling buffer. At block 156, a call assistant listens to the hearing user's voice message and transcribes text corresponding to the messages via re-voicing to software trained to the call assistant's voice, typing, etc., after which control passes to block 170.

[0089]   Referring still to Fig. 7, at process block 162, the hearing user's voice is fed directly to voice-to-text software run by server 30 which generates automated text at block 164. Although not shown in Fig. 7, after block 164, server 30 may compare the automated text to the call assistant generated text to identify errors and may use those errors to train the software to the hearing user's voice so that the automated text continues to get more accurate as a call proceeds.

[0090]   Referring still to Figs. 1 and 7, at decision block 166, controller 30 monitors how far call assistant text transcription is behind the current voice message and compares that value to a threshold value. If the delay is less than the threshold value, control passes down to block 170. If the delay exceeds the threshold value, control passes to block 168 where server 30 uses automated text from block 164 to fill

in the call assistant generated text and skips the call assistant up to the current voice message. After block 168 control passes to block 170. At block 170, the text including the call assistant generated text and the fill in text is presented to the call assistant via display screen 50 and the call assistant makes any corrections to observed errors. At block 172, the text is transmitted to assisted user's device 12 and is displayed on screen 18. Again, uncorrected text may be transmitted to and displayed on device 12 and corrected text may be subsequently transmitted and used to correct errors in the prior text in line on device 12. After block 172 control passes back up to block 152 where the process described above continues to cycle.

[0091]   Many assisted user's devices can be used as conventional telephones without captioning service or as assisted user devices where captioning is presented and voice messages are broadcast to an assisted user. The idea here is that one device can be used by hearing impaired persons and persons that have no hearing impairment and that the overall costs associated with providing captioning service can be minimized by only using captioning when necessary. In many cases even a hearing impaired person may not need captioning service all of the time. For instance, a hearing impaired person may be able to hear the voice of a person that speaks loudly fairly well but may not be able to hear the voice of another person that speaks more softly. In this case, captioning would be required when speaking to the person with the soft voice but may not be required when speaking to the person with the loud voice. As another instance, an impaired person may hear better when well rested but hear relatively more poorly when tired so captioning is required only when the person is tired. As still another instance, an impaired person may hear well when there is minimal noise on a line but may hear poorly if line noise exceeds some threshold. Again, the impaired person would only need captioning some of the time.

[0092]   To minimize captioning service costs and still enable an impaired person to obtain captioning service whenever needed and even during an ongoing call, some systems start out all calls with a default setting where an assisted user's device 12 is used like a normal telephone without captioning. At any time during an ongoing call, an assisted user can select either a mechanical or virtual "Caption" icon or button (see again 68 in Fig. 1) to link the call to a relay, provide a hearing user's voice messages to the relay and commence captioning service. One problem with starting captioning only after an assisted user experiences problems hearing words is that at least some words

(e.g., words that prompted the assisted user to select the caption button in the first place) typically go unrecognized and therefore the assisted user is left with a void in their understanding of a conversation.

[0093]   One solution to the problem of lost meaning when words are not understood just prior to selection of a caption button is to store a rolling recordation of a hearing user's voice messages that can be transcribed subsequently when the caption button is selected to generate "fill in" text. For instance, the most recent 20 seconds of a hearing user's voice messages may be recorded and then transcribed only if the caption button is selected. The relay generates text for the recorded message either automatically via software or via revoicing or typing by a call assistant or a combination of both. In addition, the call assistant starts transcribing current voice messages. The text from the recording and the real time messages is transmitted to and presented via assisted user's device 12 which should enable the assisted user to determine the meaning of the previously misunderstood words. In at least some embodiments the rolling recordation of hearing user's voice messages may be maintained by the assisted user's device 12 (see again Fig. 1) and that recordation may be sent to the relay for immediate transcription upon selection of the caption button.

[0094]   Referring now to Fig. 8, a process 230 that may be performed by the system of Fig. 1 to provide captioning for voice messages that occur prior to a request for captioning service is illustrated. Referring also to Fig. 1, at block 232 a hearing user's voice messages are received during a call with an assisted user at the assisted user's device 12. At block 234 the assisted user's device 12 stores a most recent 20 seconds of the hearing user's voice messages on a rolling basis. The 20 seconds of voice messages are stored without captioning initially. At decision block 236, the assisted user's device monitors for selection of a captioning button (not shown). If the captioning button has not been selected, control passes back up to block 232 where blocks 232, 234 and 236 continue to cycle.

[0095]   Once the caption button has been selected, control passes to block 238 where assisted user's device 12 establishes a communication link to relay 16. At block 240 assisted user's device 12 transmits the stored 20 seconds of the hearing user's voice messages along with current ongoing voice messages from the hearing user to relay 16. At this point a call assistant and/or software at the relay transcribes the voice-to-text, corrections are made (or not), and the text is transmitted back to device 12 to be

displayed. At block 242 assisted user's device 12 receives the captioned text from the relay 16 and at block 244 the received text is displayed or presented on the assisted user's device display 18. At block 246, in at least some embodiments, text corresponding to the 20 seconds of hearing user voice messages prior to selection of the caption button may be visually distinguished (e.g., highlighted, bold, underlined, etc.) from other text in some fashion. After block 246 control passes back up to block 232 where the process described above continues to cycle.

[0096]   Referring to Fig. 9, a relay server process 270 whereby automated software transcribes voice messages that occur prior to selection of a caption button and a call assistant at least initially captions current voice messages is illustrated. At block 272, after an assisted user requests captioning service by selecting a caption button, server 30 receives a hearing user's voice messages including current ongoing messages as well as the most recent 20 seconds of voice messages that had been stored by assisted user's device 12 (see again Fig. 1). After block 27, control passes to each of blocks 274 and 278 where two simultaneous processes commence in parallel. At block 274 the stored 20 seconds of voice messages are provided to voice-to-text software run by server 30 to generate automated text and at block 276 the automated text is transmitted to the assisted user's device 12 for display. At block 278 the current or real time hearing user's voice messages are provided to a call assistant and at block 280 the call assistant transcribes the current voice messages to text. The call assistant generated text is transmitted to an assisted user's device at block 282 where the text is displayed along with the text transmitted at block 276. Thus, here, the assisted user receives text corresponding to misunderstood voice messages that occur just prior to the assisted user requesting captioning. One other advantage of this system is that when captioning starts, the call assistant is not starting captioning with an already existing backlog of words to transcribe and instead automated software is used to provide the prior text.

[0097]   In addition to using a service provided by relay 16 to transcribe stored rolling text, other resources may be used to transcribe the stored rolling text. For instance, in at least some embodiments an assisted user's device may link via the Internet or the like to a third party provider that can receive voice messages and transcribe those messages, at least somewhat accurately, to text. In these cases it is contemplated that real time transcription where accuracy needs to meet a high accuracy

standard would still be performed by a call assistant or software trained to a specific voice while less accuracy sensitive text may be generated by the third party provider, at least some of the time for free, and transmitted back to the assisted user's device for display.

[0098]   In other cases, it is contemplated that the assisted user's device 12 itself may run voice-to-text software that could be used to at least somewhat accurately transcribe voice messages to text where the text generated by the assisted user's device would only be provided in cases where accuracy sensitivity is less than normal such as where rolling voice messages prior to selection of a caption icon to initiate captioning are to be transcribed.

[0099]   Fig. 10 shows another method 300 for providing text for voice messages that occurred prior to a caption request, albeit where the an assisted user's device generates the pre-request text as opposed to a relay. Referring also to Fig. 1, at block 310 a hearing user's voice messages are received at an assisted user's device 12. At block 312, the assisted user's device 12 runs voice-to-text software that, in at least some embodiments, trains on the fly to the voice of a linked hearing user and generates caption text. In this embodiment, at least initially, the caption text generated by the assisted user's device 12 is not displayed to the assisted user. At block 314, until the assisted user requests captioning, control simply routes back up to block 310. Once captioning is requested by an assisted user, control passes to block 316 where the text corresponding to the last 20 seconds generated by the assisted user's device is presented on the assisted user's device display 18. Here, while there may be some errors in the displayed text, at least some text associated with the most recent voice message can be quickly presented and give the assisted user the opportunity to attempt to understand the voice messages associated therewith. At block 318 the assisted user's device links to a relay and at block 320 the hearing user's ongoing voice messages are transmitted to the relay. At block 322, after call assistant transcription at the relay, the assisted user's device receives the transcribed text from the relay and at block 324 the text is displayed. After block 324 control passes back up to block 320 where the sub-loop including blocks 320, 322 and 324 continues to cycle.

[00100]   In at least some cases it is contemplated that voice-to-text software run outside control of the relay may be used to generate at least initial text for a hearing user's voice and that the initial text may be presented via an assisted user's device.

Here, because known software still may include more errors than allowed given standard accuracy requirements, a relay correction service may be provided. For instance, in addition to presenting text transcribed by the assisted user's device via a device display 50, the text transcribed by the assisted user's device may also be transmitted to a relay 16 for correction. In addition to transmitting the text to the relay, the hearing user's voice messages may also be transmitted to the relay so that a call assistant can compare the text generated by the assisted user's device to the voice messages. At the relay, the call assistant can listen to the voice of the hearing person and can observe the text. Any errors in the text can be corrected and corrected text blocks can be transmitted back to the assisted user's device and used for in line correction on the assisted user's display screen. One advantage to this type of system is that relatively less skilled call assistants may be retained at a lesser cost to perform the call assistant tasks. A related advantage is that the stress level on call assistants may be reduced appreciably by eliminating the need to both transcribe and correct at high speeds and therefore call assistant turnover at relays may be appreciably reduced which ultimately reduces costs associated with providing relay services.

[00101]   A similar system may include an assisted user's device that links to some other third party provider transcription/caption server to obtain initial captioned text which is immediately displayed to an assisted user and which is also transmitted to the relay for call assistant correction. Here, again, the call assistant corrections may be used by the third party provider to train the software on the fly to the hearing user's voice. In this case, the assisted user's device may have three separate links, one to the hearing user, a second link to a third party provider server, and a third link to the relay.

[00102]   Referring to Fig. 11, a method 360 whereby an assisted user's device transcribes hearing user's voice-to-text and where corrections are made to the text at a relay is illustrated. At block 362 a hearing user's voice messages are received at an assisted user's device 12 (see also again Fig. 1). At block 364 the assisted user's device runs voice-to-text software to generate text from the received voice messages and at block 366 the generated text is presented to the assisted user via display 18. At block 370 the transcribed text is transmitted to the relay 16 and at block 372 the text is presented to a call assistant via the call assistant's display 50. At block 374 the call assistant corrects the text and at block 376 corrected blocks of text are transmitted to the assisted user's device 12. At block 378 the assisted user's device 12 uses the

corrected blocks to correct the text errors via in line correction. At block 380, the assisted user's device uses the errors, the corrected text and the voice messages to train the captioning software to the hearing user's voice.

[00103]　　　　In some cases instead of having a relay or an assisted user's device run automated voice-to-text transcription software, a hearing user's device may include a processor that runs transcription software to generate text corresponding to the hearing user's voice messages. To this end, device 14 may, instead of including a simple telephone, include a computer that can run various applications including a voice-to-text program or may link to some third party real time transcription software program to obtain an initial text transcription substantially in real time. Here, as in the case where an assisted user's device runs the transcription software, the text will often have more errors than allowed by the standard accuracy requirements. Again, to correct the errors, the text and the hearing user's voice messages are transmitted to relay 16 where a call assistant listens to the voice messages, observes the text on screen 18 and makes corrections to eliminate transcription errors. The corrected blocks of text are transmitted to the assisted user's device for display. The corrected blocks may also be transmitted back to the hearing user's device for training the captioning software to the hearing user's voice. In these cases the text transcribed by the hearing user's device and the hearing user's voice messages may either be transmitted directly from the hearing user's device to the relay or may be transmitted to the assisted user's device 12 and then on to the relay. Where the hearing user's voice messages and text are transmitted directly to the relay 16, the voice messages and text may also be transmitted directly to the assisted user's device for immediate broadcast and display and the corrected text blocks may be subsequently used for in line correction.

[00104]　　　　In these cases the caption request option may be supported so that an assisted user can initiate captioning during an on-going call at any time by simply transmitting a signal to the hearing user's device instructing the hearing user's device to start the captioning process. Similarly, in these cases the help request option may be supported. Where the help option is facilitated, the automated text may be presented via the assisted user's device and, if the assisted user perceives that too many text errors are being generated, the help button may be selected to cause the hearing user's device or the assisted user's device to transmit the automated text to the relay for call assistant correction.

[00105]   One advantage to having a hearing user's device manage or perform voice-to-text transcription is that the voice signal being transcribed can be a relatively high quality voice signal. To this end, a standard phone voice signal has a range of frequencies between 300 and about 3000 Hertz which is only a fraction of the frequency range used by most voice-to-text transcription programs and therefore, in many cases, automated transcription software does only a poor job of transcribing voice signals that have passed through a telephone connection. Where transcription can occur within a digital signal portion of an overall system, the frequency range of voice messages can be optimized for automated transcription. Thus, where a hearing user's computer that is all digital receives and transcribes voice messages, the frequency range of the messages is relatively large and accuracy can be increased appreciably. Similarly, where a hearing user's computer can send digital voice messages to a third party transcription server accuracy can be increased appreciably.

[00106]   In at least some configurations it is contemplated that the link between an assisted user's device 12 and a hearing user's device 14 may be either a standard analog phone type connection or may be a digital connection depending on the capabilities of the hearing user's device that links to the assisted user's device. Thus, for instance, a first call may be analog and a second call may be digital. Because digital voice messages have a greater frequency range and therefore can be automatically transcribed more accurately than analog voice messages in many cases, it has been recognized that a system where automated voice-to-text program use is implemented on a case by case basis depending upon the type of voice message received (e.g., digital or analog) would be advantageous. For instance, in at least some embodiments, where a relay receives an analog voice message for transcription, the relay may automatically link to a call assistant for full call assistant transcription service where the call assistant transcribes and corrects text via revoicing and keyboard manipulation and where the relay receives a high definition digital voice message for transcription, the relay may run an automated voice-to-text transcription program to generate automated text. The automated text may either be immediately corrected by a call assistant or may only be corrected by an assistant after a help feature is selected by an assisted user as described above.

[00107]   Referring to Fig. 12, one process 400 for treating high definition digital messages differently than analog voice messages is illustrated. Referring also to Fig. 1,

at block 402 a hearing user's voice messages are received at a relay 16. At decision block 404, relay server 30 determines if the received voice message is a high definition digital message or is an analog message. Where a high definition message has been received, control passes to block 406 where server 30 runs an automated voice-to-text program on the voice messages to generate automated text. At block 408 the automated text is transmitted to the assisted user's device 12 for display. Referring again to block 404, where the hearing user's voice messages are in analog, control passes to block 412 where a link to a call assistant is established so that the hearing user's voice messages are provided to a call assistant. At block 414 the call assistant listens to the voice messages and transcribes the messages into text. Error correction may also be performed at block 414. After block 414, control passes to block 408 where the call assistant generated text is transmitted to the assisted user's device 12. Again, in some cases, when automated text is presented to an assisted user, a help button may be presented that, when selected causes automated text to be presented to a call assistant for correction. In other cases automated text may be automatically presented to a call assistant for correction.

[00108]   Another system is contemplated where all incoming calls to a relay are initially assigned to a call assistant for at least initial captioning where the option to switch to automated software generated text is only available when the call includes high definition audio and after accuracy standards have been exceeded. Here, all analog hearing user's voice messages would be captioned by a call assistant from start to finish and any high definition calls would cut out the call assistant when the standard is exceeded.

[00109]   In at least some cases where an assisted user's device is capable of running automated voice-to-text transcription software, the assisted user's device 12 may be programmed to select either automated transcription when a high definition digital voice message is received or a relay with a call assistant when an analog voice message is received. Again, where device 12 runs an automated text program, call assistant correction may be automatic or may only start when a help button is selected.

[00110]   Fig. 13 shows a process 430 whereby an assisted user's device 12 selects either automated voice-to-text software or a call assistant to transcribe based on the type of voice messages received. At block 432 a hearing user's voice messages are received by an assisted user's device 12. At decision block 434, a processor in device

12 determines if the assisted user has selected a help button. Initially no help button is selected as no text has been presented so at least initially control passes to block 436. At decision block 436, the device processor determines if a hearing user's voice signal that is received is high definition digital or is analog. Where the received signal is high definition digital, control passes to block 438 where the assisted user's device processor runs automated voice-to-text software to generate automated text which is then displayed on the assisted user device display 18 at block 440. Referring still to Fig. 13, if the help button has been selected at block 434 or if the received voice messages are in analog, control passes to block 442 where a link to a call assistant at relay 16 is established and the hearing user's voice messages are transmitted to the relay. At block 444 the call assistant listens to the voice messages and generates text and at block 446 the text is transmitted to the assisted user's device 12 where the text is displayed at block 440.

[00111]   In has been recognized that in many cases most calls facilitated using an assisted user's device will be with a small group of other hearing or non-hearing users. For instance, in many cases as much as 70 to 80 percent of all calls to an assisted user's device will be with one of five or fewer hearing user's devices (e.g., family, close friends, a primary care physician, etc.). For this reason it has been recognized that it would be useful to store voice-to-text models for at least routine callers that link to an assisted user's device so that the automated voice-to-text training process can either be eliminated or substantially expedited. For instance, when an assisted user initiates a captioning service, if a previously developed voice model for a hearing user can be identified quickly, that model can be used without a new training process and the switchover from a full service call assistant to automated captioning may be expedited (e.g., instead of taking a minute or more the switchover may be accomplished in 15 seconds or less, in the time required to recognize or distinguish the hearing user's voice form other voices).

[00112]   Fig. 14 shows a sub-process 460 that may be substituted for a portion of the process shown in Fig. 3 wherein voice-to-text templates along with related voice recognition profiles for callers are stored and used to expedite the handoff to automated transcription. Prior to running sub-process 460, referring again to Fig. 1, server 30 is used to create a voice recognition database for storing hearing user device identifiers along with associated voice recognition profiles and associated voice-to-text models. A

voice recognition profile is a data construct that can be used to distinguish one voice from others. In the context of the Fig. 1 system, voice recognition profiles are useful because more than one person may use a hearing user's device to call an assisted user. For instance in an exemplary case, an assisted user's son or daughter-in-law or one of any of three grandchildren may use device 14 to call an assisted user and therefore, to access the correct voice-to-text model, server 30 needs to distinguish which caller's voice is being received. Thus, in many cases, the voice recognition database will include several voice recognition profiles for each hearing user device identifier (e.g., each hearing user phone number). A voice-to-text model includes parameters that are used to customize voice-to-text software for transcribing the voice of an associated hearing user to text. The voice recognition database will include at least one voice model for each voice profile to be used by server 30 to automate transcription whenever a voice assocated with the specific profile is identified. Data in the voice recognition database will be generated on the fly as an assisted user uses device 12. Thus, initially the voice recognition database will include a simple construct with no device identifiers, profiles or voice models.

[00113]   Referring still to Figs. 1 and 14 and now also to Fig. 3, at decision block 84 in Fig. 3, if the help flag is still zero (e.g., an assisted user has not requested call assistant help to correct automated text errors) control may pass to block 464 in Fig 13 where the hearing user's device identifier (e.g., a phone number, an IP address, a serial number of a hearing user's device, etc.) is received by server 30. At block 468 server 30 determines if the hearing user's device identifier has already been added to the voice recognition database. If the hearing user's device identifier does not appear in the database (e.g., the first time the hearing user's device is used to connect to the assisted user's device) control passes to block 482 where server 30 uses a general voice-to-text program to convert the hearing user's voice messages to text after which control passes to block 476. At block 476 the server 30 trains a voice-to-text model using transcription errors. Again, the training will include comparing call assistant generated text to automated text to identify errors and using the errors to adjust model parameters so that the next time the word associated with the error is uttered by the hearing user, the software will identify the correct word. At block 478 server 30 trains a voice profile for the hearing user's voice so that the next time the hearing user calls, a voice profile will exist for the specific hearing user that can be used to identify the hearing user. At block

480 the server 30 stores the voice profile and voice model for the hearing user along
with the hearing user device identifier for future use after which control passes back up
to block 94 in Fig. 3.

[00114]   Referring still to Figs. 1 and 14, at block 468 if the hearing user's device is
already represented in the voice recognition database, control passes to block 470
where server 30 runs voice recognition software on the hearing user's voice messages
in an attempt to identify a voice profile associated with the specific hearing user. At
decision block 472, if the hearing user's voice does not match one of the previously
stored voice profiles associated with the device identifier, control passes to block 482
where the process described above continues. At block 472, if the hearing user's voice
matches a previously stored profile, control passes to block 474 where the voice model
associated with the matching profile is used to tune the voice-to-text software to be
used to generate automated text.

[00115]   Referring still to Fig. 14, at blocks 476 and 478, the voice model and voice
profile for the hearing user are continually trained. Continual training enables the
system to constantly adjust the model for changes in a hearing user's voice that may
occur over time or when the hearing user experiences some physical condition (e.g., a
cold, a raspy voice) that affects the sound of their voice. At block 480, the voice profile
and voice model are stored with the HU device identifier for future use.

[00116]   In at least some embodiments server 30 may adaptively change the order
of voice profiles applied to a hearing user's voice during the voice recognition process.
For instance, while server 30 may store five different voice profiles for five different
hearing users that routinely connect to an assisted user's device, a first of the profiles
may be used 80 percent of the time. In this case, when captioning is commenced,
server 30 may start by using the first profile to analyze a hearing user's voice at block
472 and may cycle through the profiles from the most matched to the least matched.

[00117]   To avoid server 30 having to store a different voice profile and voice
model for every hearing person that communicates with an assisted user via device 12,
in at least some embodiments it is contemplated that server 30 may only store models
and profiles for a limited number (e.g., 5) of frequent callers. To this end, in at least
some cases server 30 will track calls and automatically identify the most frequent
hearing user devices used to link to the assisted user's device 12 over some rolling
period (e.g., 1 month) and may only store models and profiles for the most frequent

callers. Here, a separate counter may be maintained for each hearing user device used to link to the assisted user's device over the rolling period and different models and profiles may be swapped in and out of the stored set based on frequency of calls.

[00118]   In other embodiments server 30 may query an assisted user for some indication that a specific hearing user is or will be a frequent contact and may add that person to a list for which a model and a profile should be stored for a total of up to five persons.

[00119]   While the system described above with respect to Fig. 14 assumes that the relay 16 stores and uses voice models and voice profiles that are trained to hearing user's voices for subsequent use, in at least some embodiments it is contemplated that an assisted user's device 12 processor may maintain and use or at least have access to and use the voice recognition database to generate automated text without linking to a relay. In this case, because the assisted user's device runs the software to generate the automated text, the software for generating text can be trained any time the user's device receives a hearing user's voice messages without linking to a relay. For example, during a call between a hearing user and an assisted user on devices 14 and 12, respectively, in Fig. 1, and prior to an assisted user requesting captioning service, the voice messages of even a new hearing user can be used by the assisted user's device to train a voice-to-text model and a voice profile for the user. In addition, prior to a caption request, as the model is trained and gets better and better, the model can be used to generate text that can be used as fill in text (e.g., text corresponding to voice messages that precede initiation of the captioning function) when captioning is selected.

[00120]   Fig. 15 shows a process 500 that may be performed by an assisted user's device to train voce models and voice profiles and use those models and profiles to automate text transcription until a help button is selected. Referring also to Fig. 1, at block 502, an assisted user's device 12 processor receives a hearing user's voice messages as well as an identifier (e.g. a phone number) of the hearing user's device 14. At block 504 the processor determines if the assisted user has selected the help button (e.g., indicating that current captioning includes too many errors). If an assisted user selects the help button at block 504, control passes to block 522 where the assisted user's device is linked to a call assistant at relay 16 and the hearing user's voice is presented to the call assistant. At block 524 the assisted user's device receives text

back from the relay and at block 534 the call assistant generated text is displayed on the assisted user's device display 18.

[00121]   Where the help button has not been selected, control passes to block 505 where the processor uses the device identifier to determine if the hearing user's device is represented in the voice recognition database. Where the hearing user's device is not represented in the database control passes to block 528 where the processor uses a general voice-to-text program to convert the hearing user's voice messages to text after which control passes to block 512.

[00122]   Referring again to Figs. 1 and 15, at block 512 the processor adaptively trains the voice model using perceived errors in the automated text. To this end, one way to train the voice model is to generate text phonetically and thereafter perform a context analysis of each text word by looking at other words proximate the word to identify errors. Another example of using context to identify errors is to look at several generated text words as a phrase and compare the phrase to similar prior phrases that are consistent with how the specific hearing user strings words together and identify any discrepancies as possible errors. At block 514 a voice profile for the hearing user is generated from the hearing user's voice messages so that the hearing user's voice can be recognized in the future. At block 516 the voice model and voice profile for the hearing user are stored for future use during subsequent calls and then control passes to block 518 where the process described above continues. Thus, blocks 528, 512, 514 and 516 enable the assisted user's device to train voice models and voice profiles for hearing users that call in anew where a new voice model can be used during an ongoing call and during future calls to provide generally accurate transcription.

[00123]   Referring still to Figs. 1 and 15, if the hearing user's device is already represented in the voice recognition database at block 505, control passes to block 506 where the processor runs voice recognition software on the hearing user's voice messages in an attempt to identify one of the voice profiles associated with the device identifier. At block 508, where no voice profile is recognized, control passes to block 528.

[00124]   At block 508, if the hearing user's voice matches one of the stored voice profiles, control passes to block 510 where the voice-to0text model associated with the matching profile is used to generate automated text from the hearing user's voice messages. Next, at block 518, the assisted user's device processor determine if the

caption button on the assisted user's device has been selected. If captioning has not bee selected control passes to block 502 where the process continues to cycle. Once captioning has been requested, control passes to block 520 where assisted user's device 12 displays the most recent 10 seconds of automated text and continuing automated text on display 18.

[00125]   In at least some embodiments it is contemplated that different types of voice model training may be performed by different processors within the overall Fig. 1 system. For instance, while an assisted user's device is not linked to a relay, the assisted user's device cannot use any errors identified by a call assistance at the relay to train a voice model as no call assistant is generating errors. Nevertheless, the assisted user's device can use context to identify errors and train a model. Once an assisted user's device is linked to a relay where a call assistant corrects errors, the relay server can use the call assistant identified errors and corrections to train a voice model which can, once sufficiently accurate, be transmitted to the assisted user's device where the new model is substituted for the old content based model or where the two models are combined into a single robust model in some fashion. In other cases when an assisted user's device links to a relay for call assistant captioning, a context based voice model generated by the assisted user's device for the hearing user may be transmitted to the relay server and used as an initial model to be further trained using call assistant identified errors and corrections. In still other cases call assistant errors may be provided to the assisted user's device and used by that device to further train a context based voice model for the hearing user.

[00126]   Referring now to Fig. 16, a sub-process 550 that may be added to the process shown in Fig. 15 whereby an assisted user's device trains a voice model for a hearing user using voice message content and a relay server further trains the voice model generated by the assisted user's device using call assistant identified errors is illustrated. Referring also to Fig. 15, sub-process 550 is intended to be performed in parallel with block 524 and 534 in Fig. 15. Thus, after block 522, in addition to block 524, control also passes to block 552 in Fig. 16. At block 552 the voice model for a hearing user that has been generated by an assisted user's device 12 is transmitted to relay 16 and at block 553 the voice model is used to modify a voice-to-text program at the relay. At block 554 the modified voice-to-text program is used to convert the hearing user's voice messages to automated text. At block 556 the call assistant

generated text is compared to the automated text to identify errors. At block 558 the errors are used to further train the voice model. At block 560, if the voice model has an accuracy below the required standard, control passes back to block 502 in Fig. 15 where the process described above continues to cycle. At block 560, once the accuracy exceeds the standard requirement, control passes to block 562 wherein server 30 transmits the trained voice model to the assisted user's device for handling subsequent calls from the hearing user for which the model was trained. At block 564 the new model is stored in the database maintained by the assisted user's device.

[00127]   Referring still to Fig. 16, in addition to transmitting the trained model to the assisted user's device at block 562, once the model is accurate enough to meet the standard requirements, server 30 may perform an automated process to cut out the call assistant and instead transmit automated text to the assisted user's device as described above in Fig. 1. In the alternative, once the model has been transmitted to the assisted user's device at block 562, the relay may be programmed to hand off control to the assisted user's device which would then use the newly trained and relatively more accurate model to perform automated transcription so that the relay could be disconnected.

[00128]   Several different concepts and aspects of the present disclosure have been described above. It should be understood that many of the concepts and aspects may be combined in different ways to configure other triage systems that are more complex. For instance, one exemplary system may include an assisted user's device that attempts automated captioning with on the fly training first and, when automated captioning by the assisted user's device fails (e.g., a help icon is selected by an assisted user), the assisted user's device may link to a third party captioning system via the internet or the like where another more sophisticated voice-to-text captioning software is applied to generate automated text. Here, if the help button is selected again, the assisted user's device may link to a call assistant at the relay for call assistant captioning with simultaneous voice-to-text software transcription where errors in the automated text are used to train the software until a threshold accuracy requirement is met. Here, once the accuracy requirement is exceeded, the system may automatically cut out the call assistant and switch to the automated text from the relay until the help button is again selected. In each of the transcription hand offs, any learning or model

training performed by one of the processors in the system may be provided to the next processor in the system to be used to expedite the training process.

[00129]   In at least some embodiments an automated voice-to-text engine may be utilized in other ways to further enhance calls handled by a relay. For instance, in cases where transcription by a call assistant lags behind a hearing user's voice messages, automated transcription software may be programmed to transcribe text all the time and identify specific words in a hearing user's voice messages to be presented via an assisted user's display immediately when identified to help the assisted user determine when a hearing user is confused by a communication delay. For instance, assume that transcription by a call assistant lags a hearing user's most current voice message by 20 seconds and that an assisted user is relying on the call assistant generated text to communicate with the hearing user. In this case, because the call assistant generated text lag is substantial, the hearing user may be confused when the assisted user's response also lags a similar period and may generate a voice message questioning the status of the call. For instance, the hearing user may utter "Are you there?" or "Did you hear me?" or "Hello" or "What did you say?". These phrases and others like them querying call status are referred to herein as "line check words" (LCWs) as the hearing user is checking the status of the call on the line.

[00130]   If the line check words are not presented until they occurred sequentially in the hearing user's voice messages, they would be delayed for 20 or more seconds in the above example. In at least some embodiments it is contemplated that the automated voice engine may search for line check words in a hearing user's voice messages and present the line check words immediately via the assisted user's device during a call regardless of which words have been transcribed and presented to an assisted user. The assisted user, seeing line check words or a phrase can verbally respond that the captioning service is lagging but catching up so that the parties can avoid or at least minimize confusion.

[00131]   When line check words are presented to an assisted user the words may be presented in-line within text being generated by a call assistant with intermediate blanks representing words yet to be transcribed by the call assistant. To this end, see again Fig. 17 that shows line check words "Are you still there?" in a highlighting box 590 at the end of intermediate blanks 216 representing words yet to be transcribed by the call assistant. Line check words will, in at least some embodiments, be highlighted on

the display or otherwise visually distinguished. In other embodiments the line check words may be located at some prominent location on the assisted user's display screen (e.g., in a line check box or field at the top or bottom of the display screen).

[00132]   One advantage of using an automated voice engine to only search for specific words and phrases is that the engine can be tuned for those words and will be relatively more accurate than a general purpose engine that transcribes all words uttered by a hearing user. In at least some embodiments the automated voice engine will be run by an assisted user's device processor while in other embodiments the automated voice engine may be run by the relay server with the line check words transmitted to the assisted user's device immediately upon generation and identification.

[00133]   Referring now to Fig. 19, a process 600 that may be performed by an assisted user's device 12 and a relay to transcribe hearing user's voice messages and provide line check words immediately to an assisted user when transcription by a call assistant lags in illustrated. At block 602 a hearing user's voice messages are received by an assisted user's device 12. After block 602 control continues along parallel sub-processes to blocks 604 and 612. At block 604 the assisted user's device processor uses an automated voice engine to transcribe the hearing user's voice messages to text. Here, it is assumed that the voice engine may generate several errors and therefore likely would be insufficient for the purposes of providing captioning to the assisted user. The engine, however, is optimized and trained to caption a set (e.g., 10 to 100) line check words and/or phrases which the engine can do extremely accurately. At block 606 the assisted user's device processor searches for line check words in the automated text. At block 608, if a line check word or phrase is not identified control passes back up to block 602 where the process continues to cycle. At block 608, if a line check word or phrase is identified, control passes to block 610 where the line check word/phrase is immediately presented to the assisted user via display 18 either in-line or in a special location and, in at least some cases, in a visually distinct manner.

[00134]   Referring still to Fig. 19, at block 612 the hearing user's voice messages are sent to a relay for transcription. At block 614 transcribed text is received at the assisted user's device back from the relay. At block 616 the text from the relay is used to fill in the intermediate blanks (see again Fig. 17 and also Fig. 18 where text has been filled in) on the assisted user's display.

[00135]   In at least some embodiments it is contemplated that an automated voice-to-text engine may operate all the time and may check for and indicate any potential errors in call assistant generated text so that the call assistant can determine if the errors should be corrected. For instance, in at least some cases, the automated voice engine may highlight potential errors in call assistant generated text on the call assistant's display screen inviting the call assistant to correct the potential errors. In these cases the call assistant would have the final say regarding whether or not a potential error should be altered.

[00136]   Consistent with the above comments, see Fig. 20 that shows a screen shot of a call assistant's display screen where potential errors have been highlighted to distinguish the errors from other text. Exemplary call assistant generated text is shown at 650 with errors shown in phantom boxes 652, 654 and 656 that representing highlighting. In the illustrated example exemplary words generated by an automated voice-to-text engine are also presented to the call assistant in hovering fields above the potentially erroneous text as shown at 658, 660 and 662. Here, a call assistant can simply touch a suggested correction in a hovering field to make a correction and replace the erroneous word with the automated text suggested in the hovering field. If a call assistant instead touches an error, the call assistant can manually change the word to another word. If a call assistant does not touch an error or an associated corrected word, the word remains as originally transcribed by the call assistant. An "Accept All" icon is presented at 669 that can be selected to accept all of the suggestions presented on a call assistant's display. All corrected words are transmitted to an assisted user's device to be displayed.

[00137]   Referring to Fig. 21, a method 700 by which a voice engine generates text to be compared to call assistant generated text and for providing a correction interface as in Fig. 20 for the call assistant is illustrated. At block 702 the hearing user's voice messages are provided to a relay. After block 702 control follows to two parallel paths to blocks 704 and 716. At block 704 the hearing user's voice messages are transcribed into text by an automated voice-to-text engine run by the relay server before control passes to block 706. At block 716 a call assistant transcribes the hearing user's voice messages to call assistant generated text. At block 718 the call assistant generated text is transmitted to the assisted user's device to be displayed. At block 720 the call

assistant generated text is displayed on the call assistant's display screen 50 for correction after which control passes to block 706.

[00138]   Referring still to Fig., 21, at block 706 the relay server compares the call assistant generated text to the automated text to identify any discrepancies. Where the automated text matches the call assistant generated text at block 708, control passes back up to block 702 where the process continues. Where the automated text does not match the call assistant generated text at block 708, control passes to block 710 where the server visually distinguishes the mismatched text on the call assistant's display screen 50 and also presents suggested correct text (e.g., the automated text). Next, at block 712 the server monitors for any error corrections by the call assistant and at block 714 if an error has been corrected, the corrected text is transmitted to the assisted user's device for in-line correction.

[00139]   In at least some embodiments the relay server may be able to generate some type of probability factor related to how likely a discrepancy between automated and call assistant generated text is related to a call assistant error and may only indicate errors and present suggestions for probable errors or discrepancies likely to be related to errors. For instance, where an automated text segment that is different than an associated call assistant generated text segment but the automated segment makes no sense contextually in a sentence, the server may not indicate the discrepancy or show the automated text segment as an option for correction. The same discrepancy may be shown as a potential error at a different time if the automated segment makes contextual sense.

[00140]   In still other embodiments automated voice-to-text software that operates at the same time as a call assistant to generate text may be trained to recognize words often missed by a call assistant such as articles, for instance, and to ignore other words that a call assistant is trained to transcribe.

[00141]   The particular embodiments disclosed above are illustrative only, as the invention may be modified and practiced in different but equivalent manners apparent to those skilled in the art having the benefit of the teachings herein. Furthermore, no limitations are intended to the details of construction or design herein shown, other than as described in the claims below. It is therefore evident that the particular embodiments disclosed above may be altered or modified and all such variations are considered

within the scope and spirit of the invention. Accordingly, the protection sought herein is as set forth in the claims below.

[00142]   Thus, the invention is to cover all modifications, equivalents, and alternatives falling within the spirit and scope of the invention as defined by the following appended claims. For example, while the methods above are described as being performed by specific system processors, in at least some cases various method steps may be performed by other system processors. For instance, where a hearing user's voice is recognized and then a voice model for the recognized hearing user is employed for voice-to-text transcription, the voice recognition process may be performed by an assisted user's device and the identified voice may be indicated to a relay 16 which then identifies a related voice model to be used. As another instance, a hearing user's device may identify a hearing user's voice and indicate the identity of the hearing user to the assisted user's device and/or the relay.

[00143]   As another example, while the system is described above in the context of a two line captioning system where one line links an assisted user's device to a hearing user's device and a second line links the assisted user's device to a relay, the concepts and features described above may be used in any transcription system including a system where the hearing user's voice is transmitted directly to a relay and the relay then transmits transcribed text and the hearing user's voice to the assisted user's device.

[00144]   As still one other example, while inputs to an assisted user's device may include mechanical or virtual on screen buttons/icons, in some embodiments other inputs arrangements may be supported. For instance, in some cases help or captioning may be indicated via a voice input (e.g., verbal a request for assistance or for captioning).

[00145]   As another example, in at least some cases where a relay includes first and second differently trained call assistants where first call assistants are trained to capable of transcribing and correcting text and second call assistants are only trained to or capable of correcting text, a call assistant may always be on a call but the automated voice-to-text software may aid in the transcription process whenever possible to minimize overall costs. For instance, when a call is initially linked to a relay so that a hearing user's voice is received at the relay, the hearing user's voice may be provided to a first call assistant fully trained to transcribe and correct text. Here, voice-to-text

software may train to the hearing user's voice while the first call assistant transcribes the text and after the voice-to-text software accuracy exceeds a threshold, instead of completely cutting out the relay or call assistant, the automated text may be provided to a second call assistant that is only trained to correct errors. Here, after training the automated text should have minimal errors and therefore even a minimally trained call assistant should be able to make corrections to the errors in a timely fashion.

[00146]   In other systems an assisted user's device processor may run automated voice-to-text software to transcribe hearing user's voice messages and may also generate a confidence factor to each word in the automated text based on how confident the processor is that the word has been accurately transcribed. The confidence factors over a most recent number of words (e.g., 100) or a most recent period (e.g., 45 seconds) may be averaged and the average used to assess an overall confidence factor for transcription accuracy. Where the confidence factor is below a threshold level, the device processor may link to relay for more accurate transcription either via more sophisticated automated voice-to-text software or via a call assistant. The automated process for linking to a relay may be used instead of or in addition to the process described above whereby an assisted user selects a "caption" button to link to a relay.

[00147]   In addition to storing hearing user voice models, a system may also store other information that could be used when an assisted user is communicating with specific hearing user's to increase accuracy of automated voice-to-text software when used. For instance, a specific hearing user may routinely use complex words from a specific industry when conversing with an assisted user. The system software can recognize when a complex word is corrected by a call assistant or contextually by automated software and can store the word and the pronunciation of the word by the specific hearing user in a hearing user word list for subsequent use. Then, when the specific hearing user subsequently links to the assisted user's device to communicate with the assisted user, the stored word list for the hearing user may be accessed and used to automate transcription. The hearing user's word list may be stored at a relay, by an assisted user's device or even by a hearing user's device where the hearing user's device has data storing capability.

[00148]   In other cases a word list specific to an assisted user's device (i.e., to an assisted user) that includes complex or common words routinely used to communicate

with the assisted user may be generated, stored and updated by the system. This list may include words used on a regular basis by any hearing user that communicates with an assisted user. In at least some cases this list or the hearing user's word lists may be stored on an internet accessible database so that the assisted user has the ability to access the list(s) and edit words on the list via an internet portal or some other network interface.

[00149]   In still other embodiments various aspects of a hearing user's voice messages may be used to select different voice-to-text software programs that optimized for voices having different characteristic sets. For instance, there may be different voice-to-text programs optimized for male and female voices or for voices having different dialects. Here, system software may be able to distinguish one dialect from others and select an optimized voice engine/software program to increase transcription accuracy. Similarly, a system may be able to distinguish a high pitched voice from a low pitched voice and select a voice engine accordingly.

[00150]   In some cases a voice engine may be selected for transcribing a hearing user's voice based on the region of a country in which a hearing user's device resides. For instance, where a hearing user's device is located in the southern part of the United States, an engine optimized for a southern dialect may be used while a device in New England may cause the system to select an engine optimized for another dialect. Different word lists may also be used based on region of a country in which a hearing user's device resides.

[00151]   To apprise the public of the scope of this invention, the following claims are made:

CLAIMS

WHAT IS CLAIMED IS:

1.   A system for providing voice-to-text captioning service for a hearing impaired assisted user using an assisted user's device to communicate with a hearing user that uses a hearing user's device, the system comprising:

a relay processor that receives voice messages generated by the hearing user during a call between the assisted user and the hearing user, the relay processor programmed to present the voice messages to a first call assistant;

a call assistant device used by the first call assistant to generate call assistant generated text corresponding to the hearing user's voice messages;

the relay processor further programmed to:

run automated voice-to-text transcription software to generate automated text corresponding to the hearing user's voice messages;

use the call assistant generated text, the automated text and the hearing user's voice messages to train the voice-to-text transcription software to more accurately transcribe the hearing user's voice messages to text;

determine when the accuracy of the automated text exceeds an accuracy requirement threshold;

during the call and prior to the automated text exceeding the accuracy requirement threshold, transmitting the call assistant generated text to the assisted user's device for display to the assisted user; and

subsequent to the automated text exceeding the accuracy requirement threshold, transmitting the automated text to the assisted user's device for display to the assisted user.

2.   The system of claim 1 wherein, subsequent to the automated text exceeding the accuracy requirement threshold, the relay processor stops presenting the hearing user's voice messages to the first call assistant so that the first call assistant can transcribe a different hearing user's voice messages.

3.   The system of claim 2 further including presenting the text received at the assisted user's device to the assisted user via a device display screen, when automated text from the relay is presented via the display screen, monitoring for an assistance request via the assisted users device and, when an assistance request is received, presenting the hearing user's voice messages to another call assistant to again generate call assistant generated text, transmitting the call assistant generated text to the assisted user's device for display and halting transmission of the automated text to the assisted user's device.

4.   The system of claim 3 wherein the step of presenting the hearing user's voice messages to a call assistant after the assistance request has been received includes, determining if the first call assistant is available to handle the hearing user's voice messages and, if the first call assistant is available, presenting the hearing user's voice messages to the first call assistant to be transcribed.

5.   The system of claim 3 wherein an assistance request is initiated by selecting a help button of the assisted user's device.

6.   The system of claim 1 further including the steps of, subsequent to the automated text exceeding the accuracy requirement threshold, presenting the automated text and the hearing user's voice messages to a second call assistant for correction and, after the second call assistant corrects errors in the automated text, transmitting the corrections to the assisted user's device.

7.   The system of claim 6 wherein the assisted user's device presents the corrections on the assisted user's device display screen.

8.   The system of claim 7 wherein the assisted user's device presents the corrections by replacing errors in the text on the assisted user's device display screen with corrected blocks of text.

9.   The system of claim 1 further including the steps of, after at least some training of the voice-to-text software, storing a voice model for the hearing user in a database to be used during subsequent communications between the hearing user and the assisted user.

10.   The system of claim 9 wherein each stored voice model is associated with a hearing user's device identifier in the database so that the next time the hearing user's device is linked to the assisted user's device to facilitate a communication, the device identifier can be used to identify the voice model associated with the hearing user.

11.   The system of claim 10 wherein more than one voice model is associated with at least one of the device identifiers in the database.

12.   The system of claim 11 wherein a voice profile for identifying a specific voice is stored in the database for each of the more than one voice models associated with the at least one of the device identifiers.

13.   The system of claim 1 wherein the hearing user's voice messages received at the relay processor are received from the assisted user's device.

14.   The system of claim 1 wherein the relay processor trains the voice-to-text transcription software by comparing the call assistant generated text to the automated text to identify errors in the automated text and then adjusts the software to avoid the same error subsequently.

15. The system of claim 1 wherein the relay processor determines when the accuracy of the automated text exceeds an accuracy requirement threshold by comparing the automated text to the call assistant generated text to generate an accuracy value for the automated text and then compares the accuracy value to the accuracy requirement.

16. The system of claim 1 wherein the call assistant generates text by revoicing the hearing user's voice messages to voice-to-text transcription software trained to the voice of the call assistant, the call assistant generated text transmitted to the assisted user's device and simultaneously presented to the call assistant for manual correction on a display screen, corrections by the call assistant to the call assistant generated text transmitted to the assisted user's device for presentation in line in the text presented to the assisted user, the processor using the corrected text to train the software to the voice of the hearing user.

17. A system for providing voice-to-text captioning service for a hearing impaired assisted user using an assisted user's device to communicate with a hearing user that uses a hearing user's device, the system comprising:

a call assistant device used by a call assistant to generate call assistant generated text corresponding to the hearing user's voice messages during a call between the assisted user and the hearing user wherein the call assistant revoices the hearing user's voice messages to voice-to-text software trained to the voice of the call assistant to generate the text; and

a processor that receives the hearing user's voice messages, the processor running a voice-to-text program to generate automated text corresponding to the hearing user's voice messages, the processor also programmed to compare the automated text to the call assistant generated text to identify errors and using the identified errors to train the voice-to-text program on the fly as the hearing user generates additional voice messages.

18.   The system of claim 17 including a relay remote from each of the assisted user's device and the hearing user's device wherein the call assistant receives the hearing user's voice messages at the relay and the call assistant generated text is transmitted to the assisted user's device from the relay.

19.   The system of claim 18 wherein the processor that runs the voice-to-text program to generate the automated text receives the hearing user's voice messages at the relay.

20.   The system of claim 17 wherein the call assistant generated text is transmitted to the assisted user's device to be presented to the assisted user via a display screen, the processor further programmed to compare the call assistant generated text to the automated text to determine the accuracy of the automated text and, when the accuracy of the automated text exceeds a threshold accuracy requirement, transmit the automated text to the assisted user's device and delink the call from the call assistant.

21.   A method for providing voice-to-text captioning service for a hearing impaired assisted user using an assisted user's device to communicate during a call with a hearing user that uses a hearing user's device, the method comprising the steps of:

initiating a voice call between the assisted user's device and the hearing user's device during which voice messages are transmitted between the assisted user's device and the hearing user's device;

while the call is on-going, receiving a request via the assisted user's device for a captioning service;

upon receiving the request for a captioning service, providing the hearing user's voice messages to a call assistant at the relay to be transcribed into text by the call assistant, the relay transmitting the call assistant generated text to the assisted user's device to be presented on an assisted user's device display;

using automated voice-to-text software to transcribe the hearing user's voice messages into automated text; and

transmitting the automated text instead of the call assistant generated text to the assisted user's device once the automated text exceeds a threshold accuracy requirement.

22.   The method of claim 21 further including the step of, while the automated text is being transmitted to the assisted user's device, receiving a request for captioning assistance from the assisted user's device and thereafter transmitting call assistant generated text instead of the automated text to the assisted user's device.

23.   The method of claim 21 further including the step of, while the automated text is being transmitted to the assisted user's device, receiving a request for captioning assistance from the assisted user's device and thereafter providing the automated text to a call assistant for correction wherein the corrected text is presented to the assisted user via a display screen on the assisted user's device.

24.   The method of claim 23 wherein the automated text is initially presented on the assisted user's device without the corrections made by the call assistant and is then in line corrected once the corrected text is subsequently received from the call assistant.

25.   The method of claim 21 further including the steps of comparing the automated text to the call assistant generated text to identify errors in the automated text and using the errors to train the voice-to-text software to more accurately transcribe the voice messages of the hearing user as the call progresses.

26.   The method of claim 25 further including the step of using the errors to identify an accuracy level of the automated text.

27.   A method for providing voice-to-text captioning service for a hearing impaired assisted user using an assisted user's device to communicate with a hearing user that uses a hearing user's device during a call, the assisted user's device including a display screen, the method comprising the steps of:

        initiating a voice call between the assisted user's device and the hearing user's device during which voice messages are transmitted between the assisted user's device and the hearing user's device;

        while the call is on-going, the assisted user's device storing the most recent X seconds of the hearing user's voice messages;

        receiving a request via the assisted user's device for a captioning service;

        upon receiving the request for a captioning service, transcribing the stored hearing user's voice messages and the hearing user's on-going messages to text; and

        presenting the transcribed text to the assisted user via the display screen wherein the transcribed text includes text corresponding to the stored hearing user's voice messages and the on-going hearing user's voice messages.

28.   The method of claim 27 wherein X is at least 8.

29.   The method of claim 27 wherein the text corresponding to the stored hearing user's voice messages is visually distinguished from the other transcribed text on the display screen.

30.   The method of claim 29 wherein the text corresponding to the Y seconds of hearing user's voice messages immediately prior to the request for captioning services is visually distinguished from other text on the display screen.

31.   The method of claim 30 wherein text is visually distinguished by highlighting the text in a fashion different from other text on the display screen.

32.   The method of claim 30 wherein Y is at least 5.

33.   The method of claim 27 wherein the step of transcribing the voice messages to text includes transmitting the stored hearing user's voice messages and on-going hearing user's voice messages to a remote relay for transcription into text and receiving the transcribed text from the relay at the assisted user's device.

34.   The method of claim 33 wherein a call assistant at the relay transcribes the hearing user's on-going voice messages.

35.   The method of claim 34 wherein the stored hearing user's voice messages are transcribed by automated voice-to-text software at the relay.

36.   The method of claim 34 wherein the call assistant also transcribes the hearing user's stored voice messages.

37.   The method of claim 27 wherein the hearing user's stored voice messages are transcribed by automated voice-to-text software at the assisted user's device.

38.   A method for providing voice-to-text captioning service for a hearing impaired assisted user using an assisted user's device to communicate with a hearing user that uses a hearing user's device during a call, the assisted user's device including a display screen, the method comprising the steps of:

initiating a voice call between the assisted user's device and the hearing user's device during which voice messages are transmitted between the assisted user's device and the hearing user's device;

while the call is on-going, receiving a request via the assisted user's device for a captioning service at a first time; and

presenting text corresponding to the hearing user's voice messages to the assisted user via the display screen including text corresponding to hearing user's voice messages that occurred prior to the first time as well as text corresponding to on-going voice messages from the hearing user that occur after the first time.

39.   The method of claim 38 further including the step of storing at least portions of the hearing user's voice messages during the call to be transcribed subsequently when the request for captioning service is received.

40.   The method of claim 39 wherein the assisted user's device stores the at least portions of the hearing user's voice messages.

41.   The method of claim 40 further including the steps of, upon receiving the request for captioning service, providing the hearing user's voice messages including the stored messages to a remote relay, the relay transcribing the voice messages to text and the relay transmitting the text to the assisted user's device.

42.   The method of claim 38 wherein the step of receiving a request includes selecting an assistance request button on the assisted user's device.

43.   A method for providing voice-to-text captioning service for a hearing impaired assisted user using an assisted user's device to communicate with a hearing user that uses a hearing user's device during a call, the assisted user's device including a display screen, the method comprising the steps of:

initiating a voice call between the assisted user's device and the hearing user's device during which voice messages are transmitted between the assisted user's device and the hearing user's device;

storing the hearing user's voice messages;

presenting the hearing user's voice messages synchronously to a call assistant at a relay, the call assistant transcribing received hearing user's voice messages into call assistant generated text;

transmitting the call assistant generated text to the assisted user's device for display;

while the call is on-going, receiving a catch up request via the assisted user's device requesting that the captioning process skip ahead to a hearing user's current voice message;

upon receiving a catch up request, presenting the current hearing user's voice message to the call assistant for transcription to text, the portion of the hearing user's stored voice messages skipped being a skipped message segment;

using voice-to-text software to transcribe at least a portion of the skipped segment into automated text; and

transmitting the automated text to the assisted user's device for display.

44.   The method of claim 43 further including the steps of transmitting the call assistant generated text and the automated text to the assisted user's device and displaying the call assistant generated text and the automated text to assisted user via the display screen.

45.   The method of claim 44 wherein the step of display the automated text includes displaying the automated text in line within the call assistant generated text when the segment of the hearing user's voice message occurred in time.

46.   The method of claim 45 wherein the automated text is visually distinguished from the call assistant generated text on the assisted user's device display screen.

47.   The method of claim 46 wherein the automated text is visually distinguished by highlighting the automated text.

48.   The method of claim 44 wherein the voice-to-text software is run by the relay.

49.   The method of claim 44 wherein the voice-to-text software is run by the assisted user's device.

50.   The method of claim 44 wherein the portion of the skipped segment includes a predefined duration of the segment.

51.   The method of claim 44 wherein the voice-to-text software runs prior to the catch up request.

52.   A method for providing voice-to-text captioning service for a hearing impaired assisted user using an assisted user's device to communicate with a hearing user that uses a hearing user's device, the assisted user's device including a display screen, the method comprising the steps of:

initiating a voice call between the assisted user's device and the hearing user's device during which voice messages are transmitted between the assisted user's device and the hearing user's device;

while the call is on-going, the assisted user's device running voice-to-text software to transcribe the hearing user's voice messages into automated text;

presenting the automated text to the assisted user via the assisted user's device display;

transmitting the automated text and the hearing user's voice messages to a remote relay;

a call assistant at the relay listening to the hearing user's voice messages and observing the automated text on a display screen and correcting errors in the automated text;

transmitting the corrected text to the assisted user's device; and

presenting the corrected text on the assisted user's device display.

53.   The method of claim 52 further including the step of using the identified errors to train the automated voice-to-text software to the voice of the hearing user during the call.

54.   The method of claim 53 further including storing a voice model for the hearing user after the automated voice-to-text software has been at least somewhat trained for use during subsequent calls from the hearing user to the assisted user's device.

55.   The method of claim 54 wherein the assisted user's device stores the voice model.

56.   A method for providing voice-to-text captioning service for a hearing impaired assisted user using an assisted user's device to communicate with a hearing user that uses a hearing user's device during a call, the assisted user's device including a display screen, the method comprising the steps of:

initiating a voice call between the assisted user's device and the hearing user's device during which voice messages are transmitted between the assisted user's device and the hearing user's device;

determining if the voice messages are digital or analog;

where the voice messages are digital, running automated voice-to-text software on the hearing user's voice messages to generate automated text; and

where the voice messages are analog, presenting the hearing user's voice messages to a first call assistant for transcription into call assistant generated text and presenting the call assistant generated text via the assisted user's device display.

57.   The method of claim 56 further including the step of presenting the automated text via the assisted user's device display.

58.   The method of claim 57 further including the step of presenting the automated text to a second call assistant for manual correction and then presenting the corrected automated text via the assisted user's device display.

59.   The method of claim 56 wherein the call assistant is located at a remote relay, the method further including the steps of linking the call to the relay and transmitting the hearing user's voice messages to the relay.

60.   The method of claim 59 wherein a processor at the relay runs the automated voice-to-text software and where the relay transmits text back to the assisted user's device.

Fig. 1

Fig. 2

Fig. 3

From 96 ("Yes")
In Fig. 3

← 120

122 ～ Start Transition
Timer

124

Present Processor Generated
Automated Text On CA Display

126 ～ Present Retain Control
Icon On CA Display

128 ～ CA
Retain Control
Selected?

Yes

132

Help
Flag = 1

No

130 ～ Transition
Timer Timed
Out?

No

Yes

To 76 In
Fig. 3

To 98 In
Fig. 3

Fig. 4

From 104 in
Fig. 3 "Yes"

136 — Link to Second
CA

134

138 — Present Processor
Generated Automated Text
To Second CA Display

140 — Receive Corrections
From CA To
Automated Text

142 — Transmit
Corrected
Automated Text
To AU Device

To 76 In
Fig. 3

Fig. 5

Fig. 6

```
                                          ┌─ 150

              ┌─────────┐
              │  Start  │
              └────┬────┘
                   │
                   ▼
         ┌──────────────────┐
   152 ─ │   Receive HU     │
         │ Voice At Relay   │────────────────────────┐
         └────────┬─────────┘                         │
                  │                                    │
                  ▼                                    ▼
         ┌──────────────────┐              ┌────────────────────────┐
   154 ─ │ Buffer HU Voice And│      162 ─ │ Provide HU Voice To     │
         │ Feed Out To CA At │             │ Voice-To-Text Processor │
         │ Controlled Rate   │             └───────────┬─────────────┘
         └────────┬─────────┘                          │
                  │                                     ▼
                  ▼                          ┌────────────────────────┐
         ┌──────────────────┐          164 ─ │ Processor Uses Voice Model│
         │  CA Listens To    │               │ To Convert HU Voice To Text│
         │   Voice And       │               └───────────┬─────────────┘
   156 ─ │ Generates Text    │                           │
         │  (e.g., typing,   │                           ▼
         │   revoicing,      │                      ╱─────────────╲
         │ correcting, etc.) │              166 ─  ╱   Processor    ╲
         └────────┬─────────┘                     ╱  Ahead Of CA By  ╲  No
                  │                               ╲    Threshold      ╱─────┐
                  │                                ╲   Amount?       ╱      │
                  │                                 ╲───────────────╱       │
                  │                                        │ Yes            │
                  │                                        ▼                │
                  │                            ┌────────────────────────┐   │
                  │                      168 ─ │ Use Processor Text To Fill In│
                  │                            │  CA Text On CA Screen   │   │
                  │                            └───────────┬─────────────┘   │
                  │                                        │                 │
                  │                                        ▼◄────────────────┘
                  │                            ┌────────────────────┐
                  │                      170 ─ │  CA Correct Text    │
                  │                            │    On Display       │
                  │                            └──────────┬──────────┘
                  │                                       │
                  │                                       ▼
                  │                            ┌────────────────────┐
                  │                      172 ─ │  Transmit Text      │
                  │                            │   To AU Device      │
                  │                            └──────────┬──────────┘
                  │                                       │
                  └───────────────────────────────────────┘
```

Fig. 6

190

```
              ┌─────────┐
              │  Start  │
              └────┬────┘
                   │
    192 ─┐  ┌───────▼────────┐
         │  │  Receive HU    │
         │  │  Voice At Relay│
         │  └───────┬────────┘
         │          │
         │          ├──────────────────────┐
  194 ─┐ │          │              200 ─┐   ▼
       │ │          ▼                    ┌──────────────────┐
       │ │ ┌────────────────┐            │ Provide HU Voice To │
       │ │ │ Buffer HU Voice And│        │Voice-To-Text Processor│
       │ │ │ Feed Out To CA At │         └────────┬─────────┘
       │ │ │ Controlled Rate  │      202 ─┐        │
       │ │ └────────┬─────────┘           ┌────────▼─────────┐
       │ │          │                     │Processor Uses Voice Model│
       │ │          ▼                     │To Convert HU Voice To Text│
       │ │ ┌────────────────┐             └────────┬─────────┘
       │ │ │  CA Listens To │       204 ─┐         │
       │ │ │  Voice And     │            ◇─────────▼─────────◇
  196 ─┤ │ │ Generates Text │            │   Re-Sync         │
       │ │ │ (e.g., typing, │            │   Command    No   │
       │ │ │  revoicing,    │            │Received From AU├──────┐
       │ │ │ correcting, etc.)│          │   Device?       │     │
       │ │ └────────┬─────────┘          ◇───────┬─────────◇     │
       │ │          │                          Yes│               │
       │ │          ▼              206 ─┐  ┌───────▼─────────┐    │
  198 ─┐ │ ┌────────────────┐          │  │Skip Buffer To Real Time So CA│ │
       │ │ │  Transmit Text │          │  │Transcribes Current HU Voice │  │
       │ │ │  To AU Device  │          └──┴────────┬─────────┘    │
       │ │ └────────────────┘     208 ─┐  ┌────────▼─────────┐    │
       │ │                             │  │Use Processor Text To Fill In│  │
       │ │                             │  │Text Corresponding To Skipped│  │
       │ │                             │  │HU Voice On CA Screen│        │
       │ │                             └──┴────────┬─────────┘    │
       │ │                        210 ─┐  ┌────────▼─────────┐    │
       │ │                             │  │Transmit Processor Text│    │
       │ │                             │  │Corresponding To Skipped HU│  │
       │ │                             │  │Voice To AU Device For Display│ │
       │ │                             └──┴──────────────────┘    │
       │ │                                                         │
       └─┴─────────────────────────────────────────────────────────┘
```

Fig. 7

```
                    ┌─────────┐
                    │  Start  │              ╱─ 230
                    └─────────┘          ╱
                         │        ╱
        ┌────────────────┼──────────────────────────┐
        │                ▼                           │
        │         ┌──────────────┐                   │
        │         │ Receive HU   │                   │
   232 ─┼─────────│ Voice At AU  │                   │
        │         │   Device     │                   │
        │         └──────────────┘                   │
        │                │                           │
        │                ▼                           │
        │         ┌──────────────┐                   │
        │         │ Store Rolling│                   │
   234 ─┼─────────│ 20 Seconds Of│                   │
        │         │   HU Voice   │                   │
        │         └──────────────┘                   │
        │                │                           │
        │                ▼                           │
        │            ╱───────╲                       │
        │          ╱           ╲                     │
   236 ─┼────────╱     AU        ╲    No             │
        │        ╲   Request     ╱───────────────────┘
        │          ╲ Captioning?╱
        │            ╲───────╱
        │                │
        │              Yes
        │                ▼
        │         ┌──────────────┐
   238 ─┼─────────│  AU Device   │
        │         │ Link To Relay│
        │         └──────────────┘
        │                │
        │                ▼
        │         ┌──────────────┐
        │         │Transmit Most │
   240 ─┼─────────│Recent 20     │
        │         │Seconds Of HU │
        │         │Voice To Relay│
        │         └──────────────┘
        │                │
        │                ▼
        │         ┌──────────────┐
        │         │  AU Device   │
   242 ─┼─────────│Receive       │
        │         │Captioned Text│
        │         │  From Relay  │
        │         └──────────────┘
        │                │
        │                ▼
        │         ┌──────────────┐
   244 ─┼─────────│Display       │
        │         │Received Text │
        │         │On AU Device  │
        │         │   Display    │
        │         └──────────────┘
        │                │
        │                ▼
        │         ┌──────────────┐
        │         │Visually      │
   246 ─┼─────────│Distinguish 20│
        │         │Most Recent   │
        │         │Seconds Of    │
        │         │Text From     │
        │         │Subsequent    │
        │         │   Text       │
        │         └──────────────┘
        │                │
        └────────────────┘
```

Fig. 8

Fig. 9

300 ⟍

Start

310 ⟍ Receive HU
Voice At AU
Device

312 ⟍ AU Device Perform
Automated Transcription

314 ⟍ AU
Request
Captioning?   No

Yes

316 ⟍ Present Text
Corresponding To Last
20 Seconds Of Voice
Via AU Device Display

318 ⟍ AU Device
Link To Relay

320 ⟍ Transmit Real Time
HU Voice To Relay

322 ⟍ AU Device
Receive Captioned
Text From Relay

324 ⟍ Display Received Text
On AU Device Display

Fig. 10

```
                        ┌─────────┐
                        │  Start  │
                        └────┬────┘
   ┌─────────────────────────┼                           ╲── 360
   │                    ┌─────▼──────┐
   │          362 ─────│ Receive HU Voice │
   │                   │  At AU Device    │
   │                    └─────┬──────┘
   │                    ┌─────▼──────┐
   │          364 ─────│ Auto-Transcribe To │
   │                   │  Text At AU Device │
   │                    └─────┬──────┘
   │      366 ─┐        ┌─────▼──────┐
   │           │       │ Present Automated │
   │           └──────│ Text To AU Device  │
   │                   │     Display        │
   │                    └─────┬──────┘
   │                    ┌─────▼──────┐
   │                   │ Transmit Text To │── 370
   │                   │     Relay        │
   │                    └─────┬──────┘
   │                    ┌─────▼──────┐
   │                   │ Present Text To  │
   │                   │ CA Via Screen    │── 372
   │                   │  And Audio       │
   │                    └─────┬──────┘
   │                    ┌─────▼──────┐
   │                   │ CA Correct Text  │── 374
   │                    └─────┬──────┘
   │                    ┌─────▼──────┐
   │                   │Transmit Corrections│── 276
   │                   │  To AU Device    │
   │                    └─────┬──────┘
   │                    ┌─────▼──────┐
   │                   │In-line Correction On│── 378
   │                   │ AU Device Display │
   │                    └─────┬──────┘
   │                    ┌─────▼──────┐
   │                   │ Use Errors To Train  │
   │                   │Transcription Software│── 380
   │                   │   To HU Voice       │
   │                    └─────┬──────┘
   └─────────────────────────┘
```

Fig. 11

```
                    ┌─────────┐
                    │  Start  │
                    └─────────┘
                         │
                         ▼
        402 ─┐   ┌──────────────┐           ╱─ 400
             └───│  Receive HU  │
                 │ Voice At Relay│
                 └──────────────┘
                         │
                         ▼
    404 ─┐         ◇─────────────◇
         └────────╱     HU        ╲         No
                 ◇  Voice In High   ◇─────────────────────┐
                  ╲  Definition?   ╱                       │
                   ◇─────────────◇                         │
                         │                                 ▼
                        Yes               412 ─┐   ┌──────────────┐
                         │                      └──│  Link To CA And│
                         ▼                         │  Provide HU    │
    406 ─┐   ┌──────────────┐                      │  Voice To CA   │
         └───│  Run Voice-To-│                     └──────────────┘
             │ Text Processor│                            │
             │  Program To   │                            ▼
             │   Generate    │         414 ─┐   ┌──────────────┐
             │   Automated   │              └──│  CA Listens To │
             │     Text      │                 │   Voice And    │
             └──────────────┘                  │ Generates Text │
                         │                      │ (e.g., typing, │
                         │                      │   revoicing,   │
                         │                      │ correcting, etc.)│
                         │                      └──────────────┘
                         │                            │
                         └──────────────┬─────────────┘
                                        │
                                        ▼
            408 ─┐   ┌──────────────┐
                 └───│ Transmit  Text│
                     │  To AU Device │
                     └──────────────┘
```

Fig. 12

Fig. 13

From 84 "No"
In Fig. 3

464 — Provide HU Device
Identifier To Server

460

468 — HU
Device Represented In
Voice Recognition
Database?

No

Yes

470 — Run Voice Recognition Software On
HU Voice To Identify Voice Profile

472 — HU
Voice Match
Profile?

No

Yes

474 — Use Voice Model To
Generate Automated Text

482 — Use General Voice-To-Text
Program To Convert Voice To Text

476 — Train Voice Model Using
Transcription Errors

478 — Train Voice Profile

480 — Store Voice Profile And Voice
Model Along With HU Device
Identifier For Future Use

To 94 In Fig. 3

Fig. 14

Fig. 15

550

From 522
In Fig. 15

Transamit AU
Generated Voice Model
To Relay Server — 552

Use Voice Model To Tune Voice-
To-Text Program — 553

Use Voice-To-Text
Program To Convert
Voice To Automated Text — 554

556 — Compare CA Generated
Text To Automated Text
To Identify Errors

558 — Use Errors To Train
Voice Model

560 — Accuracy
Great Than
Threshold? — No

562 — Yes

Transmit Trained Voice
Model to AU Device
For Subsequent Calls

564 — Store New Model For
Hearing User In Database

To 502 In
Fig. 15

Fig. 16

18

214 ┐   216 ┐

┌───┐
│   Tom and I are going out for pizza at Pistol Pals. Do │
│ you want to join us at 7 tonight? (Pause)   Hello?   │
│ Hello? _____ __ _____ _____ _____ __ _____ ___ │
│ ___ _____ _____ _____ _____ ___ │
│ ___ _____ _____ _____ _____ _____ _____. │
│ ⟨Are you still there?⟩   │
│   │
│   │
│   │
│ ─── │
│ Current Captioning Delay: 24 seconds   ⟨Catch Up⟩ │
└───┘

590 ┘

212 ┐   220

Fig. 17

Fig. 18

Fig. 19

654

658   50

650   652   660   656   662

[going]

Tom and I are going out for pizza at Pistol Pals. Do
You want to join us at 7 tonight? (Pause) The
restaurant is on the corner of 36th and Walnut Avenue.

[joining]   [eat]

I think Jane and Sue will be jobs us and after we eel we
may go out and catch a movie at Highland. Do you
want to join us or not?

Accept
all

Fig. 20

Start

702 — Receive HU Voice
Messages At Relay

700

704 — Transcribe HU Voice Messages
To Automated Text At Relay

716 — CA Transcribes HU Voice
Messages To CA Generated Text

718 — Transmit CA Generated Text TO
AU Device For DIsplay

720 — Present CA Generated Text On
CA Device Display

706 — Compare CA Generate Text To
Automated Text

708
No — Potential
Errors?

Yes

710 — Visually Distinguish Potential Errors
In Text On CA Device Display And
Present Suggested Corrections

712
No — Error
Corrected By
CA?

Yes

714 — Transmit Correction To AU
Device For In-Line Correction

Fig. 21

INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER

INV. G10L15/06   G10L15/22   H04M1/247   H04M1/725   H04M3/42
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G10L H04M

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 practicable, search terms used)

EPO-Internal, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2010/323728 A1 (GOULD ADAM [US] ET AL) 23 December 2010 (2010-12-23) abstract; claim 24; figure 1 page 1, paragraph 0004 - paragraph 0006 page 2, paragraph 0021 page 3, paragraph 0034 page 9, paragraph 0090 page 13, paragraph 0147 -----	1-16
Y	US 6 374 221 B1 (HAIMI-COHEN RAZIEL [US]) 16 April 2002 (2002-04-16) column 3, line 19 - line 38 ----- -/--	1-16

[X] Further documents are listed in the continuation of Box C.   [X] 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 application or patent 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	Date of mailing of the international search report
11 June 2015	17/08/2015

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 Perez Molina, E

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2015/017954

Box No. II   Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III   Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

        see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

        1-16

Remark on Protest   ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

Form PCT/ISA/210 (continuation of first sheet (2)) (April 2005)

C(Continuation).	DOCUMENTS CONSIDERED TO BE RELEVANT	
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 02/25910 A2 (ULTRATEC INC [US]) 28 March 2002 (2002-03-28) page 6, paragraph 21 - page 9, paragraph 26 page 11, paragraph 33 - page 13, paragraph 39 -----	1-16
A	US 2011/123003 A1 (ROMRIELL JOSEPH N [US] ET AL) 26 May 2011 (2011-05-26) abstract -----	1-16
A	US 2011/128953 A1 (WOZNIAK LOUIS JOSEPH [US] ET AL) 2 June 2011 (2011-06-02) page 2, paragraph 17 - paragraph 21 -----	1-16
A	US 2007/208570 A1 (BHARDWAJ VINOD K [US] ET AL) 6 September 2007 (2007-09-06) claim 1; figures 1-3 -----	1-16

5

INTERNATIONAL SEARCH REPORT
Information on patent family members

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
US 2010323728	A1	23-12-2010	US 2010323728 A1		23-12-2010
			US 2012302269 A1		29-11-2012
			US 2013244705 A1		19-09-2013
US 6374221	B1	16-04-2002	NONE		
WO 0225910	A2	28-03-2002	GB 2382744 A		04-06-2003
			US 2002080926 A1		27-06-2002
			US 2003108165 A1		12-06-2003
			WO 0225910 A2		28-03-2002
US 2011123003	A1	26-05-2011	US 2011123003 A1		26-05-2011
			US 2013158995 A1		20-06-2013
			US 2015051908 A1		19-02-2015
US 2011128953	A1	02-06-2011	US 2011128953 A1		02-06-2011
			US 2013128881 A1		23-05-2013
			US 2015003445 A1		01-01-2015
US 2007208570	A1	06-09-2007	US 2007208570 A1		06-09-2007
			US 2010198596 A1		05-08-2010
			WO 2007103299 A2		13-09-2007

FURTHER INFORMATION CONTINUED FROM **PCT/ISA/** 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-16

   Voice captioning service for a hearing impaired user of a telephone call assisted by a relay for voice-to-text transcription using a computer generated text or a call (human) assistant generated text.
   ---

2. claims: 17-20

   Voice captioning service for a hearing impaired user of a telephone call assisted by a relay for voice-to-text transcription using a (human) assistant to re-voicing the messages to a voice-to-text software.
   ---

3. claims: 21-26

   Voice captioning service for a hearing impaired user of a telephone call assisted by a relay for voice-to-text transcription with capabilities for requesting for transcription while the call is on-going.
   ---

4. claims: 27-51

   Voice captioning service for a hearing impaired user of a telephone call assisted by a relay for voice-to-text transcription with capabilities for storing the voice messages to be transcribed.
   ---

5. claims: 52-55

   Voice captioning service for a hearing impaired user of a telephone call assisted by a relay for voice-to-text transcription with a call (human) assistant for correcting computer generated text.
   ---

6. claims: 56-60

   Voice captioning service for a hearing impaired user of a telephone call assisted by a relay for voice-to-text transcription having capabilities for determining whether the voice messages to be captioned are digital or analog, and selecting a computer generated transcription or a (human) assistant generated transcription depending on the determined (digital or analog) nature of the voice messages.