AUTOMATED TRANSCRIPTION SYSTEM AND METHOD USING TWO SPEECH CONVERTING INSTANCES AND COMPUTER-ASSISTED CORRECTION

Background of the Invention



G00001

1. Field of the Invention

The present invention relates in general to computer speech recognition systems and, in particular, to a system and method for automating the text transcription of voice dictation by various end users.

2. Background Art

Speech recognition programs are well known in the art. While these programs are ultimately useful in automatically converting speech into text, many users are dissuaded from using these programs because they require each user to spend a significant amount of time training the system. Usually this training begins by having each user read a series of pre-selected materials for approximately 20 minutes. Then, as the user continues to use the program, as words are improperly transcribed the user is expected to stop and train the program as to the intended word thus advancing the ultimate accuracy of the acoustic model. Unfortunately, most professionals (doctors, dentists, veterinarians, lawyers) and business executive are unwilling to spend the time developing the necessary acoustic model to truly benefit from the automated transcription.

Accordingly, it is an object of the present invention to provide a system that offers transparent training of the speech recognition program to the end-users.

There are systems for using computers for routing transcription from a group of end users Most often these systems are used in large multi-user settings such as hospitals In those systems, a voice user dictates into a general-purpose computer or other recording device and the resulting file is transferred automatically to a human transcriptionist. The human transcriptionist transcribes the file. which is then returned to the original "author" for review These systems have the perpetual overhead of



159159

employing a sufficient number of human transcriptionist to transcribe all of the dictation files.

Accordingly it is another object of the present invention to provide an automated means of translating speech into text where ever suitable so as to minimize the number of human transcriptionist necessary to transcribe audio files coming into the system.

It is an associated object to provide a simplified means for providing verbatim text files for training a user's acoustic model for the speech recognition portion of the system.

It is another associated object of the present invention to automate a preexisting speech recognition program toward further minimizing the number operators necessary to operate the system.

These and other objects will be apparent to those of ordinary skill in the art having the present drawings, specification and claims before them

Summary of the Disclosure

The present disclosure relates to a system and method for substantially automating transcription services for one or more voice users. In particular, this system involves using two speech converting instances to facilitate the establishment of a verbatim transcription text with minimal human transcription

The system includes means for receiving a voice dictation file from a current user That voice dictation file is fed into first means for automatically converting the voice dictation file into a first written text and second means for automatically converting the voice dictation file into a second written text. The first and second means have first and second sets of conversion variables, respectively These first and second sets of conversion variables have at least one difference.

For instance, where the first and second automatic speech converting means each comprise a preexisting speech recognition program, the programs themselves may be different from each other Various speech recognition programs have inherently different speech-to-text conversion approaches, thus, likely resulting in different conversion on difficult speech utterances, which, in turn, can be used to establish the

verbatim text. Among the available preexisting speech converting means are Dragon Systems' Naturally Speaking, IBM's Via Voice and Philips Corporation's Magic Speech.

In another approach, the first and second sets of conversion variables could each comprise a language model (i.e. a general or a specialized language model), which again would likely result in different conversions on difficult utterances leading to easier establishment of the verbatim text. Alternatively, one or more setting associated with the preexisting speech recognition program(s) being used could be modified.

In yet another approach, the voice dictation file can be pre-processed prior to its input into one or both of the automatic conversion means. In this way, the conversion variables (e.g. digital word size, sampling rate, and removing particular harmonic ranges) can be differed between the speech conversion instances.

The system further includes means for manually editing a copy of said first and second written texts to create the verbatim text of the voice dictation file. In one approach, the first written text is at least temporarily synchronized to the voice dictation file. In this instance, the manual editing means includes means for sequentially comparing a copy of the first and second written texts resulting in a sequential list of unmatched words culled from first written text The manual editing means further includes means for incrementally searching for a current unmatched word contemporaneously within a first buffer associated with the first automatic conversion means containing the first written text and a second buffer associated with the sequential list The manual editing means also includes means for correcting the current unmatched word in the second buffer The correcting means including means for displaying the current unmatched word in a manner substantially visually isolated from other text in the first written text and means for playing a portion of said synchronized voice dictation recording from the first buffer associated with the current unmatched word. In one embodiment, the editing means further includes means for alternatively viewing said current unmatched word in context within the copy of the first written text.

The system may also include training means to improve the accuracy of the speech recognition program

The application also discloses a method for automating transcription services for one or more voice users in a system including at least one speech recognition program. The method includes: (1) receiving a voice dictation file from a current voice user; (2) automatically creating a first written text from the voice dictation file with a speech recognition program using a first set of conversion variables; (3) automatically creating a second written text from the voice dictation file with a speech recognition program using a second set of conversion variables; (4) manually establishing a verbatim file through comparison of the first and second written texts; and (5) returning the verbatim file to the current user. Establishing a verbatim file includes (6) sequentially comparing a copy of the first written text with the second written text resulting in a sequential list of unmatched words culled from the copy of the first written text, the sequential list having a beginning, an end and a current unmatched word, the current unmatched word being successively advanced from the beginning to the end; (7) incrementally searching for the current unmatched word contemporaneously within a first buffer associated with the at least one speech recognition program containing the first written text and a second buffer associated with the sequential list; (8) displaying the current unmatched word in a manner substantially visually isolated from other text in the copy of the first written text and playing a portion of the synchronized voice dictation recording from the first buffer associated with the current unmatched word; and (9) correcting the current unmatched word to be a verbatim representation of the portion of the synchronized voice dictation recording.

Brief Description of the Drawings

Fig. 1 of the drawings is a block diagram of one potential embodiment of the present system for substantially automating transcription services for one or more voice users;

Fig. 1b of the drawings is a block diagram of a general-purpose computer which may be used as a dictation station, a transcription station and the control means within the present system.

Fig. 2a of the drawings is a flow diagram of the main loop of the control means of the present system;

Fig. 2b of the drawings is a flow diagram of the enrollment stage portion of the control means of the present system;

Fig. 2c of the drawings is a flow diagram of the training stage portion of the control means of the present system;

Fig. 2d of the drawings is a flow diagram of the automation stage portion of the control means of the present system;

Fig. 3 of the drawings is a directory structure used by the control means in the present system,

Fig. 4 of the drawings is a block diagram of a portion of a preferred embodiment of the manual editing means;

Fig. 5 of the drawings is an elevation view of the remainder of a preferred embodiment of the manual editing means; and

Fig. 6 of the drawings is an illustration of the arrangement of the system that present automated transcription system and method using two speech converting instances and computer-assisted correction.

Best Modes of Practicing the Invention

While the present invention may be embodied in many different forms, there is shown in the drawings and discussed herein a few specific embodiments with the understanding that the present disclosure is to be considered only as an exemplification of the principles of the invention and is not intended to limit the invention to the embodiments illustrated.

Fig. 1 of the drawings generally shows one potential embodiment of the present system for substantially automating transcription services for one or more voice users. The present system must include some means for receiving a voice dictation file from a current user This voice dictation file receiving means can be a digital audio recorder, an

analog audio recorder, or standard means for receiving computer files on magnetic media or via a data connection.

As shown, in one embodiment, the system 100 includes multiple digital recording stations 10, 11, 12 and 13. Each digital recording station has at least a digital audio recorder and means for identifying the current voice user.

Preferably, each of these digital recording stations is implemented on a general-purpose computer (such as computer 20), although a specialized computer could be developed for this specific purpose. The general-purpose computer, though has the added advantage of being adaptable to varying uses in addition to operating within the present system 100 In general, the general-purpose computer should have, among other elements, a microprocessor (such as the Intel Corporation PENTIUM, Cyrix K6 or Motorola 68000 series); volatile and non-volatile memory; one or more mass storage devices (i.e. HDD (not shown), floppy drive 21, and other removable media devices 22 such as a CD-ROM drive, DITTO, ZIP or JAZ drive (from Iomega Corporation) and the like); various user input devices, such as a mouse 23, a keyboard 24, or a microphone 25; and a video display system 26. In one embodiment, the general-purpose computer is controlled by the WINDOWS 9.x operating system. It is contemplated, however, that the present system would work equally well using a MACINTOSH computer or even another operating system such as a WINDOWS CE, UNIX or a JAVA based operating system, to name a few.

Regardless of the particular computer platform used, in an embodiment utilizing an analog audio input (via microphone 25) the general-purpose computer must include a sound-card (not shown). Of course, in an embodiment with a digital input no sound card would be necessary.

In the embodiment shown in Fig. 1, digital audio recording stations 10, 11, 12 and 13 are loaded and configured to run digital audio recording software on a PENTIUM-based computer system operating under WINDOWS 9 x Such digital recording software is available as a utility in the WINDOWS 9 x operating system or from various third party vendor such as The Programmers' Consortium, Inc of Oakton, Virginia (VOICEDOC), Syntrillium Corporation of Phoenix, Arizona (COOL EDIT) or Dragon Systems Corporation (Dragon Naturally Speaking Professional Edition). These

various software programs produce a voice dictation file in the form of a "WAV" file. However, as would be known to those skilled in the art, other audio file formats, such as MP3 or DSS, could also be used to format the voice dictation file, without departing from the spirit of the present invention. In one embodiment where VOICEDOC software is used that software also automatically assigns a file handle to the WAV file, however, it would be known to those of ordinary skill in the art to save an audio file on a computer system using standard operating system file management methods.

Another means for receiving a voice dictation file is dedicated digital recorder 14, such as the Olympus Digital Voice Recorder D-1000 manufactured by the Olympus Corporation. Thus, if the current voice user is more comfortable with a more conventional type of dictation device, they can continue to use a dedicated digital recorder 14. In order to harvest the digital audio text file, upon completion of a recording, dedicated digital recorder 14 would be operably connected to one of the digital audio recording stations, such as 13, toward downloading the digital audio file into that general-purpose computer. With this approach, for instance, no audio card would be required.

Another alternative for receiving the voice dictation file may consist of using one form or another of removable magnetic media containing a pre-recorded audio file. With this alternative an operator would input the removable magnetic media into one of the digital audio recording stations toward uploading the audio file into the system.

In some cases it may be necessary to pre-process the audio files to make them acceptable for processing by the speech recognition software. For instance, a DSS file format may have to be changed to a WAV file format, or the sampling rate of a digital audio file may have to be upsampled or downsampled For instance, in use the Olympus Digital Voice Recorder with Dragon Naturally Speaking, Olympus' 8MHz rate needs to be upsampled to 11 MHz. Software to accomplish such pre-processing is available from a variety of sources including Syntrillium Corporation and Olympus Corporation.

The other aspect of the digital audio recording stations is some means for identifying the current voice user. The identifying means may include keyboard 24 upon which the user (or a separate operator) can input the current user's unique identification code Of course, the user identification can be input using a myriad of computer input

devices such as pointing devices (e.g. mouse 23), a touch screen (not shown), a light pen (not shown), bar-code reader (not shown) or audio cues via microphone 25, to name a few

In the case of a first time user the identifying means may also assign that user an identification number after receiving potentially identifying information from that user, including (1) name; (2) address; (3) occupation; (4) vocal dialect or accent; etc. As discussed in association with the control means, based upon this input information, a voice user profile and a sub-directory within the control means are established. Thus, regardless of the particular identification means used, a user identification must be established for each voice user and subsequently provided with a corresponding digital audio file for each use such that the control means can appropriately route and the system ultimately transcribe the audio.

In one embodiment of the present invention, the identifying means may also seek the manual selection of a specialty vocabulary It is contemplated that the specialty vocabulary sets may be general for various users such as medical (i.e. Radiology, Orthopedic Surgery, Gynecology) and legal (i.e. corporate, patent, litigation) or highly specific such that within each specialty the vocabulary parameters could be further limited based on the particular circumstances of a particular dictation file. For instance, if the current voice user is a Radiologist dictating the reading of a abdominal CAT scan the nomenclature is highly specialized and different from the nomenclature for a renal ultrasound. By narrowly segmenting each selectable vocabulary set an increase in the accuracy of the automatic speech converter is likely.

As shown in Fig. 1, the digital audio recording stations may be operably connected to system 100 as part of computer network 30 or, alternatively, they may be operably connected to the system via internet host 15 As shown in Fig. 1b, the general-purpose computer can be connected to both network jack 27 and telephone jack. With the use of an internet host, connection may be accomplished by e-mailing the audio file via the Internet. Another method for completing such connection is by way of direct modem connection via remote control software, such as PC ANYWHERE, which is available from Symantec Corporation of Cupertino, California It is also possible, if the IP address of digital audio recording station 10 or internet host 15 is known, to transfer the audio file using basic file transfer protocol. Thus, as can be seen from the foregoing,

the present system allows great flexibility for voice users to provide audio input into the system.

Control means 200 controls the flow of voice dictation file based upon the training status of the current voice user. As shown in Figs. 2a, 2b, 2c, 2d, control means 200 comprises a software program operating on general purpose computer 40. In particular, the program is initialized in step 201 where variable are set, buffers cleared and the particular configuration for this particular installation of the control means is loaded. Control means continually monitors a target directory (such as "current" (shown in Fig. 3)) to determine whether a new file has been moved into the target, step 202. Once a new file is found (such as "6723.id" (shown in Fig. 3)), a determination is made as to whether or not the current user 5 (shown in Fig. 1) is a new user, step 203.

For each new user (as indicated by the existence of a ".pro" file in the "current" subdirectory), a new subdirectory is established, step 204 (such as the "usern" subdirectory (shown in Fig. 3)) This subdirectory is used to store all of the audio files ("xxxx.wav"), written text ("xxxx.wrt"), verbatim text ("xxxx.vb"), transcription text ("xxxx txt") and user profile ("usern.pro") for that particular user. Each particular job is assigned a unique number "xxxx" such that all of the files associated with a job can be associated by that number. With this directory structure, the number of users is practically limited only by storage space within general-purpose computer 40.

Now that the user subdirectory has been established, the user profile is moved to the subdirectory, step 205. The contents of this user profile may vary between systems. The contents of one potential user profile is shown in Fig. 3 as containing the user name, address, occupation and training status. Aside from the training status variable, which is necessary, the other data is useful in routing and transcribing the audio files.

The control means, having selected one set of files by the handle, determines the identity of the current user by comparing the " id" file with its "user.tbl," step 206 Now that the user is known the user profile may be parsed from that user's subdirectory and the current training status determined, step 207. Steps 208-211 are the triage of the current training status is one of enrollment, training, automate, and stop automation.

Enrollment is the first stage in automating transcription services. As shown in Fig 2b, the audio file is sent to transcription, step 301. In particular, the "xxxx.wav" file is transferred to transcriptionist stations 50 and 51. In a preferred embodiment, both stations are general-purpose computers, which run both an audio player and manual input means. The audio player is likely to be a digital audio player, although it is possible that an analog audio file could be transferred to the stations. Various audio players are commonly available including a utility in the WINDOWS 9.x operating system and various other third parties such from The Programmers' Consortium, Inc. of Oakton, Virginia (VOICESCRIBE). Regardless of the audio player used to play the audio file, manual input means is running on the computer at the same time. This manual input means may comprise any of text editor or word processor (such as MS WORD, WordPerfect, AmiPro or Word Pad) in combination with a keyboard, mouse, or other user-interface device. In one embodiment of the present invention, this manual input means may, itself, also be speech recognition software, such as Naturally Speaking from Dragon Systems of Newton, Massachusetts, Via Voice from IBM Corporation of Armonk, New York, or Speech Magic from Philips Corporation of Atlanta, Georgia. Human transcriptionist 6 listens to the audio file created by current user 5 and as is known, manually inputs the perceived contents of that recorded text, thus establishing the transcribed file, step 302. Being human, human transcriptionist 6 is likely to impose experience, education and biases on the text and thus not input a verbatim transcript of the audio file Upon completion of the human transcription, the human transcriptionist 6 saves the file and indicates that it is ready for transfer to the current users subdirectory as "xxxx.txt", step 303.

Inasmuch as this current user is only at the enrollment stage, a human operator will have to listen to the audio file and manually compare it to the transcribed file and create a verbatim file, step 304 That verbatim file "xxxx vb" is also transferred to the current user's subdirectory, step 305. Now that verbatim text is available, control means 200 starts the automatic speech conversion means, step 306 This automatic speech conversion means may be a preexisting program, such as Dragon System's Naturally Speaking, IBM's Via Voice or Philips' Speech Magic, to name a few. Alternatively, it could be a unique program that is designed to specifically perform automated speech recognition

In a preferred embodiment, Dragon Systems' Naturally Speaking has been used by running an executable simultaneously with Naturally Speaking that feeds phantom keystrokes and mousing operations through the WIN32API, such that Naturally Speaking believes that it is interacting with a human being, when in fact it is being controlled by control means 200. Such techniques are well known in the computer software testing art and, thus, will not be discussed in detail. It should suffice to say that by watching the application flow of any speech recognition program, an executable to mimic the interactive manual steps can be created.

If the current user is a new user, the speech recognition program will need to establish the new user, step 307 Control means provides the necessary information from the user profile found in the current user's subdirectory. All speech recognition require significant training to establish an acoustic model of a particular user. In the case of Dragon, initially the program seeks approximately 20 minutes of audio usually obtained by the user reading a canned text provided by Dragon Systems. There is also functionality built into Dragon that allows "mobile training." Using this feature, the verbatim file and audio file are fed into the speech recognition program to beginning training the acoustic model for that user, step 308. Regardless of the length of that audio file, control means 200 closes the speech recognition program at the completion of the file, step 309

As the enrollment step is too soon to use the automatically created text, a copy of the transcribed file is sent to the current user using the address information contained in the user profile, step 310. This address can be a street address or an e-mail address. Following that transmission, the program returns to the main loop on Fig. 2a.

After a certain number of minutes of training have been conducted for a particular user, that user's training status may be changed from enrollment to training. The border for this change is subjective, but perhaps a good rule of thumb is once Dragon appears to be creating written text with 80% accuracy or more, the switch between states can be made. Thus, for such a user the next transcription event will prompt control means 200 into the training state. As shown in Fig 2c, steps 401-403 are the same human transcription steps as steps 301-303 in the enrollment phase Once the transcribed file is established, control means 200 starts the automatic speech conversion means (or speech recognition program) and selects the current user, step 404. The audio

file is fed into the speech recognition program and a written text is established within the program buffer, step 405. In the case of Dragon, this buffer is given the same file handle on very instance of the program. Thus, that buffer can be easily copied using standard operating system commands and manual editing can begin, step 406

In one particular embodiment utilizing the VOICEWARE system from The Programmers' Consortium, Inc. of Oakton, Virginia, the user inputs audio into the VOICEWARE system's VOICEDOC program, thus, creating a ".wav" file. In addition, before releasing this ".wav" file to the VOICEWARE server, the user selects a "transcriptionist." This "transcriptionist" may be a particular human transcriptionist or may be the "computerized transcriptionist " If the user selects a "computerized transcriptionist" they may also select whether that transcription is handled locally or remotely This file is assigned a job number by the VOICEWARE server, which routes the job to the VOICESCRIBE portion of the system. Normally, VOICESCRIBE is used by the human transcriptionist to receive and playback the job's audio (".wav") file. In addition, the audio file is grabbed by the automatic speech conversion means. In this VOICEWARE system embodiment, by placing VOICESCRIBE in "auto mode" new jobs (i.e. an audio file newly created by VOICEDOC) are automatically downloaded from the VOICEWARE server and a VOICESCRIBE window having a window title formed by the job number of the current ".wav" file. An executable file, running in the background "sees" the VOICESCRIBE window open and using the WIN32API determines the job number from the VOICESCRIBE window title The executable file then launches the automatic speech conversion means. In Dragon System's Naturally Speaking, for instance, there is a built in function for performing speech recognition on a preexisting ".wav" file. The executable program feeds phantom keystrokes to Naturally Speaking to open the " wav" file from the "current" directory (see Fig. 3) having the job number of the current job.

In this embodiment, after Naturally Speaking has completed automatically transcribing the contexts of the " wav" file, the executable file resumes operation by selecting all of the text in the open Naturally Speaking window and copying it to the WINDOWS 9.x operating system clipboard. Then, using the clipboard utility, save the clipboard as a text file using the current job number with a "dmt" suffix. The executable file then "clicks" the "complete" button in VOICESCRIBE to return the "dmt" file to the

VOICEWARE server As would be understood by those of ordinary skill in the art, the foregoing procedure can be done utilizing other digital recording software and other automatic speech conversion means. Additionally, functionality analogous to the WINDOWS clipboard exists in other operating systems. It is also possible to require human intervention to activate or prompt one or more of the foregoing steps. Further, although, the various programs executing various steps of this could be running on a number of interconnected computers (via a LAN, WAN, internet connectivity, email and the like), it is also contemplated that all of the necessary software can be running on a single computer

Another alternative approach is also contemplated wherein the user dictates directly into the automatic speech conversion means and the VOICEWARE server picks up a copy in the reverse direction. This approach works as follows, without actually recording any voice, the user clicks on the "complete" button in VOICEDOC, thus, creating an empty ".wav" file. This empty file is nevertheless assigned a unique job number by the VOICEWARE server. The user (or an executable file running in the background) then launches the automatic speech conversion means and the user dictates directly into that program, in the same manner previously used in association with such automatic speech conversion means. Upon completion of the dictation, the user presses a button labeled "return" (generated by a background executable file), which executable then commences a macro that gets the current job number from VOICEWARE (in the manner describe above), selects all of the text in the document and copies it to the clipboard. The clipboard is then saved to the file "<jobnumber>.dmt," as discussed above The executable then "clicks" the "complete" button (via the WIN32API) in VOICESCRIBE, which effectively returns the automatically transcribed text file back to the VOICEWARE server, which, in turn, returns the completed transcription to the VOICESCRIBE user. Notably, although, the various programs executing various steps of this could be running on a number of interconnected computers (via a LAN, WAN, internet connectivity, email and the like), it is also contemplated that all of the necessary software can be running on a single computer As would be understood by those of ordinary skill in the art, the foregoing procedure can be done utilizing other digital recording software and other automatic speech conversion means Additionally, functionality analogous to the WINDOWS clipboard exists in other operating systems It

is also possible to require human intervention to activate or prompt one or more of the foregoing steps.

Manual editing is not an easy task. Human beings are prone to errors. Thus, the present invention also includes means for improving on that task. As shown in Fig. 4, the transcribed file ("3333.txt") and the copy of the written text ("3333.wrt") are sequentially compared word by word 406a toward establishing sequential list of unmatched words 406b that are culled from the copy of the written text. This list has a beginning and an end and pointer 406c to the current unmatched word. Underlying the sequential list is another list of objects which contains the original unmatched words, as well as the words immediately before and after that unmatched word, the starting location in memory of each unmatched word in the sequential list of unmatched words 406b and the length of the unmatched word

As shown in Fig. 5, the unmatched word pointed at by pointer 406c from list 406b is displayed in substantial visual isolation from the other text in the copy of the written text on a standard computer monitor 500 in an active window 501. As shown in Fig. 5, the context of the unmatched word can be selected by the operator to be shown within the sentence it resides, word by word or in phrase context, by clicking on buttons 514, 515, and 516, respectively.

Associated with active window 501 is background window 502, which contains the copy of the written text file. As shown in background window 502, a incremental search has located (see pointer 503) the next occurrence of the current unmatched word "cash." Contemporaneously therewith, within window 505 containing the buffer from the speech recognition program, the same incremental search has located (see pointer 506) the next occurrence of the current unmatched word A human user will likely only being viewing active window 501 activate the audio replay from the speech recognition program by clicking on "play" button 510, which plays the audio synchronized to the text at pointer 506 Based on that snippet of speech, which can be played over and over by clicking on the play button, the human user can manually input the correction to the current unmatched word via keyboard, mousing actions, or possibly even audible cues to another speech recognition program running within this window

In the present example, even if the choice of isolated context offered by buttons 514, 515 and 516, it may still be difficult to determine the correct verbatim word out-of-context, accordingly there is a switch window button 513 that will move background window 502 to the foreground with visible pointer 503 indicating the current location within the copy of the written text. The user can then return to the active window and input the correct word, "trash." This change will only effect the copy of the written text displayed in background window 502.

When the operator is ready for the next unmatched word, the operator clicks on the advance button 511, which advances pointer 406c down the list of unmatched words and activates the incremental search in both window 502 and 505. This unmatched word is now displayed in isolation and the operator can play the synchronized speech from the speech recognition program and correct this word as well. If at any point in the operation, the operator would like to return to a previous unmatched word, the operator clicks on the reverse button 512, which moves pointer 406c back a word in the list and causes a backward incremental search to occur This is accomplished by using the underlying list of objects which contains the original unmatched words. This list is traversed in object by object fashion, but alternatively each of the records could be padded such that each item has the same word size to assist in bi-directional traversing of the list. As the unmatched words in this underlying list are read only it is possible to return to the original unmatched word such that the operator can determine if a different correction should have been made.

Ultimately, the copy of the written text is finally corrected resulting in a verbatim copy, which is saved to the user's subdirectory. The verbatim file is also passed to the speech recognition program for training, step 407. The new (and improved) acoustic model is saved, step 408, and the speech recognition program is closed, step 409. As the system is still in training, the transcribed file is returned to the user, as in step 310 from the enrollment phase.

As shown in Fig 4, the system may also include means for determining the accuracy rate from the output of the sequential comparing means. Specifically, by counting the number of words in the written text and the number of words in list 406b the ratio of words in said sequential list to words in said written text can be determined, thus providing an accuracy percentage. As before, it is a matter of choice when to

advance users from one stage to another. Once that goal is reached, the user's profile is changed to the next stage, step 211.

One potential enhancement or derivative functionality is provided by the determination of the accuracy percentage. In one embodiment, this percentage could be used to evaluate a human transcriptionist's skills. In particular, by using either a known verbatim file or a well-established user, the associated ".wav" file would be played for the human transcriptionist and the foregoing comparison would be performed on the transcribed text versus the verbatim file created by the foregoing process. In this manner, additional functionality can be provided by the present system.

As understood, currently, manufacturers of speech recognition programs use recording of foreign languages, dictions, etc. with manually established verbatim files to program speech models. It should be readily apparent that the foregoing manner of establishing verbatim text could be used in the initial development of these speech files simplifying this process greatly.

Once the user has reached the automation stage, the greatest benefits of the present system can be achieved. The speech recognition software is started, step 600, and the current user selected, step 601. If desired, a particularized vocabulary may be selected, step 602. Then automatic conversion of the digital audio file recorded by the current user may commence, step 603. When completed, the written file is transmitted to the user based on the information contained in the user profile, step 604 and the program is returned to the main loop.

Unfortunately, there may be instances where the voice users cannot use automated transcription for a period of time (during an illness, after dental work, etc.) because their acoustic model has been temporarily (or even permanently) altered. In that case, the system administrator may set the training status variable to a stop automation state in which steps 301, 302, 303, 305 and 310 (see Fig. 2b) are the only steps performed.

Fig 6 of the drawings depicts another potential arrangement of various elements associated with the present invention. In this arrangement, as before, a user verbally dictates a document that they desire to have transcribed, which is saved as a voice

dictation file 700 in one of the manners described above. In this embodiment -- rather than have a human transcriptionist ever produce a transcribed file – the voice dictation file is automatically converted into written text at least twice.

After that double automatic text conversation, the resulting first and second written text files are compared one to another using manual copy editing means (as described above in association with Figs 4 and 5) facilitating a human operator in expeditiously and manually correcting the second written text file.

In this manner, it is believed that transcription service can be provided with far less human transcriptionist effort. The key to obtaining a sufficiently accurate written text for delivery to the end user is to differ the speech-to-text conversion in some way between the first and second runs. In particular, between the first and second conversion step the system may change one or more of the following:

(1) speech recognition programs (e.g. Dragon Systems' Naturally Speaking, IBM's Via Voice or Philips Corporation's Magic Speech);

(2) language models within a particular speech recognition program (e.g. general English versus a specialized vocabulary (e.g. medical, legal));

(3) settings within a particular speech recognition program (e.g. "most accurate" versus "speed"); and/or

(4) the voice dictation file by pre-processing same with a digital signal processor (such as Cool Edit by Syntrillium Corporation of Phoenix, Arizona or a programmed DSP56000 IC from Motorola, Inc.) by changing the digital word size, .sampling rate, removing particular harmonic ranges and other potential modifications

By changing one or more of the foregoing "conversion variables" it is believed that the second speech-to-text conversion will produce a slightly different written text than the first speech-to-text conversion and that by comparing the two resulting written texts using the novel manual editing means disclosed herein, a human operator can review the differences in the manner noted above and quickly produce a verbatim text for delivery to the end user Thus, in this manner, it is believed that fully automated transcription can be achieved with less human intervention that in the other approaches disclosed.

This system and the underlying method is illustrated in Fig. 6. It should be noted that while two automatic speech conversion means 702 and 703 are depicted, there may be only a single instance of a speech recognition program running on a single computer, but using different conversion variables between iterations of conversion of the voice dictation file. Of course, it is equally possible to have multiple instances of a speech recognition program running on a single machine or even on separate machines interconnected by a computerized network (LAN, WAN, peer-to-peer, or the like) as would be known to those of ordinary skill in the art.

Similarly, while manual editing means 705 is depicted as being separate from the automated speech conversion means, it too may be implemented on the same computer as one or both of the instances of the automatic speech conversion means. Likewise, the manual editing means may also be implemented on a separate computer, as well interconnected with the other computers along a computerized network.

Finally, Digital Signal Processor 701 is shown to illustrate that one approach to changing the conversion variables is to alter the voice dictation file input to one or both of the instances of the automatic speech conversion means. Again, this digital signal processor can be implemented on the same computer as any one or all of the foregoing functional blocks or on a separate computer interconnected with the other computers via a computerized network

It is contemplated that the foregoing case in which two iterations of speech-to-text conversion is used could be extrapolated to a case where even more conversion iterations are performed each using various sets of conversion variables with text comparison being performed between unique pairs of written text outputs and thereafter between each other with a resulting increase in the accuracy of the automatic transcription leaving fewer words to be considered in manual editing.

The foregoing description and drawings merely explain and illustrate the invention and the invention is not limited thereto. Those of the skill in the art who have the disclosure before them will be able to make modifications and variations therein without departing from the scope of the present invention. For instance, it is possible to implement all of the elements of the present system on a single general-purpose computer by essentially time sharing the machine between the voice user, transcriptionist

and the speech recognition program. The resulting cost saving makes this system accessible to more types of office situations not simply large medical clinics, hospital, law firms or other large entities.

WHAT IS CLAIMED IS.

1. A system for substantially automating transcription services for one or more voice users, said system comprising.

- means for receiving a voice dictation file from a current user, said current user being one of said one or more voice users;

- first means for automatically converting said voice dictation file into a first written text, said first automatic conversion means having a first set of conversion variables,

- second means for automatically converting said voice dictation file into a second written text, said second automatic converting means having a second set of conversion variables, said first and second sets of conversion variables having at least one difference; and

- means for manually editing a copy of said first and second written texts to create a verbatim text of said voice dictation file.

2. The invention according to Claim 1 wherein said first written text is at least temporarily synchronized to said voice dictation file, said manual editing means comprises.

- means for sequentially comparing a copy of said first written text with said second written text resulting in a sequential list of unmatched words culled from said copy of said first written text, said sequential list having a beginning, an end and a current unmatched word, said current unmatched word being successively advanced from said beginning to said end;

- means for incrementally searching for said current unmatched word contemporaneously within a first buffer associated with said first automatic conversion means containing said first written text and a second buffer associated with said sequential list: and

- means for correcting said current unmatched word in said second buffer, said correcting means including means for displaying said current unmatched word in a

manner substantially visually isolated from other text in said copy of said first written text and means for playing a portion of said synchronized voice dictation recording from said first buffer associated with said current unmatched word.

3 The invention according to Claim 2 wherein said editing means further includes means for alternatively viewing said current unmatched word in context within said copy of said first written text.

4 The invention according to Claim 1 wherein said first and second automatic speech converting means each comprises a preexisting speech recognition program intended for human interactive use, each of said first and second automatic speech converting means includes means for automating responses to a series of interactive inquiries from said preexisting speech recognition program

5 The invention according to Claim 4 wherein said difference between said first and second sets of conversion variables is said preexisting speech recognition program comprising said first and second automatic speech converting means.

6 The invention according to Claim 5 wherein said automatic speech converting means is selected from the group consisting essentially of Dragon Systems' Naturally Speaking, IBM's Via Voice and Philips Corporation's Magic Speech.

7. The invention according to Claim 4 wherein said difference between said first and second sets of conversion variables comprises a language model used in association with said preexisting speech recognition program.

8. The invention according to Claim 7 wherein a generalized language model is used in said first set of conversion variables and a specialized language model is used in said second set of conversion variables.

9. The invention according to Claim 4 wherein said difference between said first and second sets of conversion variables comprises at least one setting associated with said preexisting speech recognition program.

10 The invention according to Claim 4 wherein said difference between said first and second sets of conversion variables comprises means for pre-processing audio prior to its input to said first automatic conversion means.

11 The invention according to Claim 10 wherein said difference between said first and second sets of conversion variables comprises means for pre-processing audio prior to its input to said second automatic conversion means, wherein said first and second pre-processing variable is different.

12. The invention according to Claim 11 wherein said pre-processing variables is selected from the group consisting essentially of digital word size, sampling rate, and removing particular harmonic ranges.

13. The invention according to Claim 1 wherein said difference between said first and second sets of conversion variables comprises a language model used in association with said preexisting speech recognition program.

14. The invention according to Claim 13 wherein a generalized language model is used in said first set of conversion variables and a specialized language model is used in said second set of conversion variables.

15. The invention according to Claim 1 wherein said difference between said first and second sets of conversion variables comprises means for pre-processing audio prior to its input to said first automatic conversion means

16. The invention according to Claim 16 wherein said difference between said first and second sets of conversion variables comprises means for pre-processing audio prior to its input to said second automatic conversion means, wherein said first and second pre-processing variable is different.

17. The invention according to Claim 1 further including means for training said automatic speech converting means to achieve higher accuracy with said voice dictation file of current user.

18 The invention according to Claim 17 wherein said training means comprises a preexisting training portion of a preexisting speech recognition program intended for human interactive use, said training means includes means for automating responses to a series of interactive inquiries from said preexisting training portion of said preexisting speech recognition program

19. A method for automating transcription services for one or more voice users in a system including at least one speech recognition program, said method comprising the steps of:

- receiving a voice dictation file from a current voice user;

- automatically creating a first written text from the voice dictation file with a speech recognition program using a first set of conversion variables;

- automatically creating a second written text from the voice dictation file with a speech recognition program using a second set of conversion variables;

- manually establishing a verbatim file through comparison of the first and second written texts; and

- returning the verbatim file to the current user.

20. The invention according to Claim 19 wherein said step of manually establishing a verbatim file includes the sub-steps of:

- sequentially comparing a copy of the first written text with the second written text resulting in a sequential list of unmatched words culled from the copy of the first written text, the sequential list having a beginning, an end and a current unmatched word, the current unmatched word being successively advanced from the beginning to the end;

- incrementally searching for the current unmatched word contemporaneously within a first buffer associated with the at least one speech recognition program containing the first written text and a second buffer associated with the sequential list; and

- displaying the current unmatched word in a manner substantially visually isolated from other text in the copy of the first written text and playing a portion of the synchronized voice dictation recording from the first buffer associated with the current unmatched word, and

- correcting the current unmatched word to be a verbatim representation of the portion of the synchronized voice dictation recording.

21. The invention according to Claim 19 further comprising.

- selecting the first set of conversion variables from available preexisting speech recognition programs; and

- differently selecting the second set of conversion variables from available preexisting speech recognition programs.

22. The invention according to Claim 19 further comprising:

- selecting the first set of conversion variables from available language models, and

- differently selecting the second set of conversion variables from available language models.

23. The invention according to Claim 19 further comprising preprocessing the voice dictation file before automatically creating a first written text, the preprocessing forming at least a part of the first set of conversion variables

24. The invention according to Claim 23 further comprising preprocessing the voice dictation file differently than the first set of preprocessing conversion variables before automatically creating a second written text, the preprocessing forming at least a part of the second set of conversion variables.