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

(11) EP 0 762 384 A2

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
12.03.1997 Bulletin 1997/11

(51) Int Cl.[6]: **G10L 5/04**

(21) Application number: **96306091.8**

(22) Date of filing: **21.08.1996**

(84) Designated Contracting States:
DE FR GB IT

(30) Priority: **01.09.1995 US 522895**

(71) Applicant: **AT&T IPM Corp.**
Coral Gables, Florida 33134 (US)

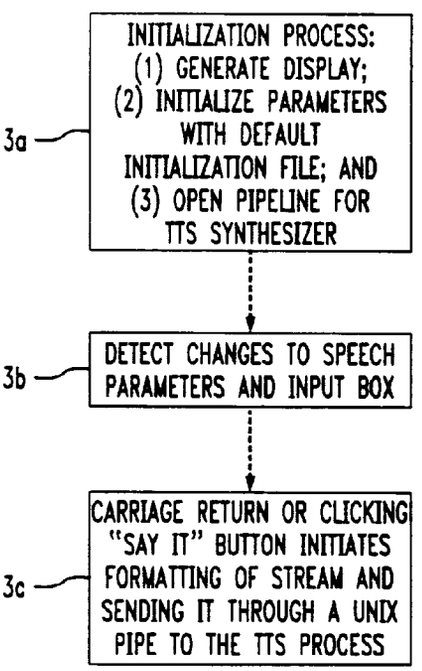
(72) Inventor: **Buntschuh, Bruce Melvin**
Berkeley Heights, New Jersey 07922 (US)

(74) Representative:
Watts, Christopher Malcolm Kelway, Dr.
Lucent Technologies (UK) Ltd,
5 Mornington Road
Woodford Green Essex, IG8 0TU (GB)

(54) Method and apparatus for modifying voice characteristics of synthesized speech

(57) A speech synthesizer system and method is disclosed which facilitates experimentation with speech parameters that control acoustical characteristics of a base synthesized voice. The invention utilizes an interactive graphical user interface for manipulating and transmitting speech parameter values to a text to speech synthesizer. The text to speech synthesizer almost immediately converts test utterances, i.e., sample text to synthesize, to speech using the base synthesized voice altered according to the current speech parameter values received from the graphical user interface.

FIG. 3



EP 0 762 384 A2

Description

BACKGROUND OF THE INVENTION

1. FIELD OF THE INVENTION

The present invention relates to speech synthesizer systems, and more particularly to a speech synthesizer system that utilizes an interactive graphical user interface that controls acoustical characteristics of synthesized speech.

2. BACKGROUND OF THE RELATED ART

Most text-to-speech synthesizer systems provide one base male and female synthesized voice. It is known that by altering the acoustical characteristics of the base synthesized voice, a new voice can be created. This new voice will have simulated voice characteristics as manifesting from a person of age and/or sex different from the base voice. Until recently, the potential for creating new voices offered by this knowledge has not been fully exploited by present day text-to-speech synthesizer systems.

Kun-Shan Lin, Alva E. Henderson and Gene A. Frantz disclosed in U.S. patent number 4,624,012 a method and apparatus for modifying voice characteristics of synthesized speech. Their method relies upon separating selected acoustical characteristics, such as the pitch period, the vocal tract model and the speech rate, into their respective speech parameters. These speech parameters are then varied and recombined with the original voice to create a modified synthesized voice having acoustical characteristics differing from the original voice.

Similar to the invention disclosed in U.S. patent number 4,624,012, Bell Labs text-to-speech synthesizer also permits users to manipulate the speech parameters that control the acoustical characteristics of synthesized speech. In the Bell Labs text-to-speech synthesizer system, users can modify the speech parameters using escape sequences. The escape sequences consist of ASCII codes that indicate to the Bell Labs text-to-speech synthesizer the manner in which to alter one or more speech parameter. At least the following speech parameters are controllable in the Bell Labs text-to-speech synthesizer system: three pitch parameters, rate, the front and back head of the vocal tract, and aspiration.

By manipulating the above mentioned speech parameters, a virtual continuum of new voices can be created from a base synthesized voice. To create specific voices, a user is often required to undergo a time consuming process of experimentating with various combinations of speech parameters before ascertaining which particular combination achieves the desired sound. Experimentation is facilitated if the user is familiar with the text-to-speech synthesizer and the manner in which the speech parameters modify the base voice.

To fully exploit the capability of present day text-to-speech synthesizer systems to create virtually unlimited voices, a facility to explore the effects of various combinations of speech parameters in a simple, efficient manner is needed. It is therefore the object of the present invention to provide such a method and system for facilitating new combinations of speech parameters utilizing an interactive graphical user interface.

SUMMARY OF THE INVENTION

The present invention is directed to a method and system that satisfies the need for a facility to explore new combinations of speech parameters in a simple, efficient manner. The method utilizes a graphical user interface for manipulating the speech parameters that control acoustical characteristics of a base synthesized voice. The method comprises the steps of: (1) generating and displaying the graphical user interface; (2) modifying current speech parameter values through the graphical user interface; (3) forming a text string; and (4) outputting the text string to a text to speech synthesizer. The text string includes the current speech parameter values which indicates to the text to speech synthesizer the change in the corresponding acoustical characteristics of the base synthesized voice. The text string may also include test utterances and escape codes. The test utterances represent text to be converted to speech by the text to speech synthesizer. The escape codes indicate to the text to speech synthesizer the particular acoustical characteristics to alter.

Advantageously, modifying the current speech parameter values may be accomplished by selecting a named voice from a listbox in the graphical user interface or by manipulating any combination of parameter scales in the graphical user interface. The named voices in the listbox have associated speech parameter values which are assigned as the current speech parameter values when selected by a user. The graphical user interface includes the following manipulable parameter scales: three pitches, front and rear head of the vocal tract, rate and aspiration. The position of sliders within the parameter scales determines the current speech parameter values.

The speech synthesizer system for carrying out the above described method has a text to speech synthesizer operative to modify acoustical characteristics of a base synthesized voice. The speech synthesizer system comprises a facilitating means for manipulating speech parameters and an output means. The facilitating means includes a graphical user interface. The graphical user interface includes parameter scales and formation means. The parameter scales are responsive to input from a user for altering current speech parameter values. By manipulating sliders within the parameter scales, the user can modify the current speech parameter values. The values of the current speech pa-

rameter are determined by the positions of the sliders within the parameter scales. The formation means are operative to create a text string which includes the current speech parameter values. These values indicate to the text to speech synthesizer change in corresponding acoustical characteristics of the base synthesized voice. The text string may also include test utterances and escape codes. The output means transmits the text string from the facilitating means to the text to speech synthesizer. Includable within the speech synthesizer system is an opening means for initiating the text to speech synthesizer so the text to speech synthesizer is operative to receive the text string from the output means.

Advantageously, the present invention includes a dialogue processing means for preparing a dialogue script to be converted to speech. The dialogue processing means is operative (1) to detect speaker names in the dialogue script, (2) to match detected speaker names against named voices in a library of named voices, (3) to modify said dialogue script by replacing the detected speaker names with escape sequences, and (4) to output the modified dialogue script to the text to speech synthesizer. The named voices in the library each have associated speech parameter values. The escape sequences are ASCII codes comprised of escape codes and associated speech parameter values. The escape codes indicate to the text to speech synthesizer particular acoustical characteristics to alter. The associated speech parameter values indicate to the text to speech synthesizer change in acoustical characteristics of the base synthesized voice.

These and other features, aspects, and advantages of the present invention will become better understood with regard to the following description, appended claims, and accompanying drawings.

DESCRIPTION OF THE DRAWINGS

Fig. 1 depicts a speech synthesizer system utilizing a graphical user interface to manipulate parameters that control acoustical characteristics of synthesized speech;

Fig. 2 depicts the communication process between the graphical user interface and a text-to-speech synthesizer;

Fig. 3 depicts a flowchart of the graphical user interface utilized by the present invention for processing data to the text-to-speech synthesizer;

Fig. 4 depicts an exemplary example of a display generated by the graphical user interface for manipulating parameters;

Fig. 5 depicts a flowchart for processing modifications to the parameter scales shown in Fig. 4;

Fig. 6 depicts a flowchart for loading speech parameter values associated with the selected named voice from the scrollable list shown in Fig. 4;

Fig. 7 depicts a flowchart for determining whether the speech parameter value for aspiration have been modified since the last transmission by the graphical user interface to the text-to-speech synthesizer;

Fig. 8 depicts a flowchart for detecting change in the selected base voice;

Fig. 9 depicts a flowchart for a companion preprocessor utilized by the present invention for processing dialogue scripts; and

Fig 10. depicts an example of a dialogue script convertible to speech by the companion preprocessor shown in Fig. 9.

DETAILED DESCRIPTION

As shown in Fig. 1, there is illustrated an exemplary embodiment of a computer-based speech synthesizer system 02 that comprises a processing unit 07, a display screen terminal 08, input devices, e.g., a keyboard 10 and a mouse 12. The processing unit 07 includes a processor 04 and a memory 06. The mouse 12 includes switches 13 having a positive on and a positive off position for generating signals to the speech synthesizer system 02. The screen 08, keyboard 10 and pointing device 12 are collectively known as the display. In the preferred embodiment of the invention, the speech synthesizer system 02 utilizes UNIX® as the computer operating system and X Windows® as the windowing system for providing an interface between the user and a graphical user interface. UNIX and X Windows can be found resident in the memory 06 of the speech synthesizer system 02 or in a memory of a centralized computer, not shown, to which the speech synthesizer system 02 is connected.

X Windows is designed around what is described as client/server architecture. This term denotes a cooperative data processing effort between certain computer programs, called servers, and other computer programs, called clients. X Windows is a display server, which is a program that handles the task of controlling the display. Graphical user interfaces (also referred herein as "GUI") are clients, which are programs that need to gain access to the display in order to receive input from the keyboard 10 and/or mouse 12 and to transmit output to the screen 08. X Windows provides data processing services to the GUI since the GUI cannot perform operations directly on the display. Through X Windows, the GUI is able to interact with the display. X Windows and the GUI communicate with each other by exchanging messages. X Windows uses what is called an event model. The GUI informs X Windows of the events of interest to the GUI, such as information entered via the keyboard 10 or clicking the mouse 12 in a predetermined area, and then waits for any of the events of interest to occur. Upon such occurrence, X Windows notifies the GUI so the GUI can process the data.

The present invention is a graphical user interface and can be found resident in the memory 06 of the

speech synthesizer system 02 or the memory of the centralized computer. The interface provides an interactive means for facilitating experimentation with the speech parameters that control the acoustical characteristics of synthesized speech. The present invention is written in the Tcl-Tk language and operates with the standard windowing shell provided with the Tcl-Tk package. Tcl is a simple scripting language (its name stands for "tool command language") for controlling and extending applications. Tk is an X Windows toolkit which extends the core Tcl facilities with commands for building user interfaces having Motif "look and feel" in Tcl scripts instead of C code. Motif "look and feel" denotes the standard "look and feel" for X Windows as is known in the art and defined by Open Software Foundation®. Tcl and Tk are implemented as a library of C procedures so it can be used in many applications. Tcl and Tk are fully described by John K. Ousterhout in a 1994 publication entitled "Tcl and the Tk Toolkit" from Addison Wesley Publishing Company.

The preferred embodiment of the present invention utilizes UNIX's multitasking and pipe features to create an efficient speech synthesizer system that provides effectively instant feedback for facilitating experimentation with speech parameters. The multitasking feature allows more than one application program to run concurrently on the same computer system. The pipe feature involves multitasking and allows the output of one program to be directly passed as input to another program. The Tcl scripting language utilizes these two UNIX features to provide a mechanism for communicating with other programs. In this embodiment, the Present invention program (written in the Tcl language) communicates with a concurrently running Bell Labs text-to-speech synthesizer program through a UNIX pipe. The Bell Labs text-to-speech synthesizer program can be found resident in the memory 06 of the speech synthesizer system 02 or in the memory of the centralized computer.

As shown in Fig. 2, the present invention uses UNIX pipes to send a text string comprised of a series of escape sequences and test utterances to the Bell Labs text-to-speech synthesizer. The escape sequences are ASCII codes comprised of pairs of escape codes and associated speech parameter values. The escape codes and parameter values identify to the Bell Labs Text-to-speech synthesizer which speech parameters are to be set and the values to be assigned to each of the speech parameters, respectively. The test utterances represent the text to be converted to speech by the Bell Labs text-to-speech synthesizer. Upon receipt of the text string, the Bell Labs text-to-speech synthesizer will convert the test utterances to speech using a base synthesized voice altered according to the escape sequences. Through the present invention interface, users are able to explore combinations of speech parameters that would normally be time consuming if they were to be manually entered into the Bell Labs text-to-speech synthesizer. The fact that the user is actually manipulating the Bell Labs text-to-speech escape sequences is entirely transparent.

Fig. 3, 5-8 are flowcharts illustrating the sequence of steps utilized by the present invention for processing data to the Bell Labs text-to-speech synthesizer. Fig. 3 illustrates the main routine for the present invention graphical user interface and Figs. 5-8 illustrate how changes to the speech parameters are detected and handled by the main routine. The program begins with the initialization process in step 3a, as shown in Fig. 3. A display is generated and a default initialization file in a user's home directory is used to set current values for each speech parameter. Step 3a also creates a pipeline using a Tcl open command and command-line arguments. The pipeline allows the present invention to send data directly to the Bell Labs text-to-speech synthesizer.

An exemplary embodiment of the present invention is written using the Tk toolkit to generate the The present invention graphical user interface display 20 shown in Fig. 4 through X Windows. Tk implements a ready-made set of controls call "widgets" with the Motif "look and feel. " The display 20 comprises the following controls: parameter scales 22, a scrollable list 24 of named voices, a male voice button 26a, a female voice button 26b, an input box 28 for entering test utterances, a "Say It" button 30 and a display box 32. Manipulating any of the controls (except for the display box 32) will cause a change to the current speech parameter values or test utterances. Step 3b in Fig. 3 will detect any of these changes.

The parameter scales 22 are created using the Tk scale widgets and button widgets. The parameter scales 22 provide means to modify the current values for the following speech parameters: pitchT, pitchR, pitchB, rate, front head, back head and aspiration. Each of the parameter scales 22 are manipulable within a range of values set according to acceptable ranges of the Bell Labs text-to-speech synthesizer. Additional scales can be included in the display 20 for manipulating other speech parameters. Each parameter scale 22 has a slider 22a, a "-" button 22b and a "+" button 22c. The parameter scales 22 display a scale value 22d that corresponds to the relative position of the slider 22a within the range of the corresponding parameter scale 22. Each time the sliders 22a are repositioned, the scale widget evaluates a Tcl command that causes the current speech parameter values to be updated with the scale values 22d. Thus repositioning the sliders 22a have the effect of changing the current speech parameter values. The present invention graphical user interface provides three techniques for changing the scale values 22d by repositioning the slider 22a with a mouse 12, joystick or other similar device: clicking on or selecting the "-" or "+" buttons 22b and 22, dragging the slider 22a, or clicking in the scale 22. Any of these actions will trigger the occurrence of an event of interest to the present invention graphical user interface.

The "-" and "+" buttons 22b and 22c are linked to the parameter scales 22 by a Tcl bind command. Clicking on either the "-" button 22b or "+" button 22c in step 5a, as shown in Fig. 5, will cause the corresponding parameter scale 22 to be repositioned left or right a predetermined increment in step 5b. Dragging the sliders 22a or clicking in the parameter scales 22 will also cause the sliders 22a to be repositioned.

Whenever any parameter scale 22 is repositioned, as in steps 5a-c, it becomes necessary to update the current speech parameter value with the current scale value 22d of the repositioned parameter scale 22. This is done by step 5d and is detected by step 3b. Alternately, the present invention can utilize a graphical user interface that has entry boxes for users to change the current speech parameter values by typing in the desired number.

The scrollable list 24 is created with the Tk listbox widget and provides a collection of previously created voices stored as named voices. These named voices are loaded in the list 24 in step 3a from the user's default initialization file or from a system default initialization file. The default initialization file includes named voices and associated speech parameter values. The user can select a named voice from the list 24 by double-clicking on one with the mouse 12. A Tcl bind command is used to link a Tcl script to the double-clicking action. When a named voice is selected, the Tcl script causes the speech parameters values associated with the selected named voice to be assigned as the current speech parameter values, as shown by steps 6a and 6b in Fig. 6. This provides a quick mechanism for recalling previously formed combinations of speech parameter values. The sliders 22a are subsequently repositioned to reflect the current speech parameter values. This change will also be detected by step 3b in Fig. 3.

Like most commercial text-to-speech synthesizers, the Bell Labs text-to-speech synthesizer provides one base male and female speaker. The present invention permits the user to select one of the two speakers as a base voice by clicking on either button 26a or 26b created with the Tk radio button widget. The acoustical characteristics of the selected base voice are altered according to the current speech parameter values. When the user changes the base voice, the current speech parameter value for the sex of the base voice is subsequently updated in step 8b. Step 3b of the main routine will detect this change.

Referring back to Fig. 3, the input box 28 is created with the Tk entry widget to permit the user to enter the test utterances, i.e., the text the user desires to have the Bell Labs text-to-speech synthesizer covert to speech. Any change to the input box 28 (or the test utterances) is detected in step 3b.

When the user is ready to listen to the modified synthesized voice, he or she either presses the carriage return on the keyboard 10 when the focus is on the input box 28 or uses the mouse 12 to click on the "Say It" button 30. Any of these actions will trigger an event and cause another Tcl script to be executed. This Tcl script forms and transmits to the Bell Labs text-to-speech synthesizer via the UNIX pipe, as shown in Figs. 2 and 3 by step 3c, the text string comprised of the series of escape sequences followed by the test utterances from the input box 28. The Tcl script first pairs the escape codes with their associated speech parameter values and then strings them together to form the series of escape sequences. When the text string is received by the Bell Labs text-to-speech synthesizer, the test utterances are converted to speech providing users with effectively instant feedback regarding the effects of the new combination of speech parameters on the selected base voice.

The display box 32 shows the series of escape sequences that were ultimately transmitted by the present invention graphical user interface to the Bell Labs text-to-speech synthesizer. An escape sequence for the base voice and each speech parameter except for aspiration is included in the series of escape sequences. The current Bell Labs text-to-speech synthesizer does not allow for aspiration to be controlled by an escape code. Changes to aspiration are handled by the present invention through a command-line argument that opens the Bell Labs text-to-speech synthesizer. Normally only one Bell Labs text-to-speech synthesizer is opened per session unless the current speech parameter value for aspiration had been changed. When this occurs, another pipeline for a Bell Labs text-to-speech synthesizer using the current aspiration value is opened. As shown in Fig. 7, when the user in step 7a double clicks on the "Say It" button 30 or presses the carriage return, the present invention determines in step 7b whether the parameter value for aspiration had changed since the last transmission of the text string to the Bell Labs text-to-speech synthesizer. If it did not change, then the present invention proceeds to step 7d and passes the text string to the current Bell Labs text-to-speech synthesizer. If it did change, then the present invention proceeds to step 7c before proceeding to step 7d. In step 7c, the present invention closes the pipeline for the current Bell Labs text-to-speech synthesizer and opens another one using the current parameter value for aspiration.

Advantageously, the present invention allows users to save the current speech parameter values as a newly created named voices. The present invention provides an entry box 34 entitled "Name of new voice:," as shown in Fig. 4, to record new combinations of speech parameter values as a named voice. This new named voice will be subsequently added to the list 24 and stored in the default initialization file with its associated speech parameter values.

Another embodiment of the present invention includes a companion preprocessor. This embodiment takes advantage of the named voices created with the present invention interface. Once some voices have been created and stored, they can be used to process

dialogue scripts or other applications. An example of a dialogue script having speaker names and utterances is shown in Fig. 10. The preprocessor accesses data in step 9a from a .voice file, as shown in Fig. 9, which contains a list of named voices and their associated speech parameter values. In steps 9b and 9c, the preprocessor filters out the bracket-enclosed speaker names and then replaces them with escape sequences formed using the speech parameter values associated with the named voices matching the speaker names. The escape sequences and the utterances are output in step 9d to the Bell Labs text-to-speech synthesizer to be converted to speech. The result is a spoken colloquy with different voices. If the voice file does not have a named voice matching the speaker name, a default substitute named voice may be used or the program can prompt the user for an alternate named voice.

Provided herein is a facility to explore the effects of various combinations of speech parameters in a simple, efficient manner. Although the present invention has been described in considerable detail with reference to Bell Labs text-to-speech synthesizer system, the above described invention can be used with similar text-to-speech synthesizers that utilizes escape codes (or similar means) to manipulate speech parameters that control the acoustical characteristics of synthesized speech.

Claims

1. A method for creating new combinations of speech parameters that control acoustical characteristics of a base synthesized voice utilizing a graphical user interface comprising the steps of:

   generating and displaying said graphical user interface;
   modifying current speech parameter values through said graphical user interface, wherein modification of said present speech parameter values is indicative of change to a text to speech synthesizer in corresponding acoustical characteristics of said base synthesized voice;
   forming a text string, wherein said text string includes said current speech parameter values; and
   outputting said text string to said text to speech synthesizer.

2. The method as recited in claim 1 wherein said text string further includes test utterances, said test utterances representing text to be converted to speech by said text to speech synthesizer.

3. The method as recited in claim 1 wherein said step of modifying said current speech parameter values further includes the step of manipulating any combination of parameter scales in said graphical user interface, wherein a position of an adjustment means within said parameter scales determines said current speech parameter values.

4. The method as recited in claim 1 wherein modifying said current speech parameter values further includes the step of

   manipulating an adjustment means in said graphical user interface, said adjustment means manipulable within ranges of scale values,
   determining scale values based on a position of said adjustment means within said ranges of scale values, and
   assigning said scale vales as said current speech parameter values.

5. The method as recited in claim 1, wherein said step of modifying said current speech parameter values further includes manipulating any combination of parameter scales in said graphical user interface for controlling pitch, front head, rear head, rate and aspiration, wherein a position of an adjustment means within said parameter scales determines said current speech parameter values.

6. The method as recited in claim 1, wherein modifying said current speech parameter values further includes the step of selecting a named voice having associated speech parameter values from a listbox in said graphical user interface and assigning said associated speech parameter values as said current speech parameter values.

7. The method as recited in claim 1, wherein said text string further includes escape codes, said escape codes paired with corresponding current speech parameter values, said escape codes indicative of particular acoustical characteristics to said text to speech synthesizer in which to alter.

8. The method as recited in claim 1 comprising the additional step of:
   opening said text to speech synthesizer for receiving said text string from said graphical user interface.

9. The method as recited in claim 1 comprising the additional step of:
   recording as a named voice said current speech parameter values.

10. A method for creating new combinations of speech parameters that control acoustical characteristics of a base synthesized voice utilizing a graphical user

interface comprising the steps of:

modifying a first class speech parameter values and a second class speech parameter values, wherein said first class of speech parameter values and said second class of speech parameter values are indicative of change to a text to speech synthesizer in corresponding acoustical characteristics of said base synthesized voice;
opening a text to speech synthesizer with a command string containing command-line arguments, wherein said command-line arguments include current present ones of first class speech parameter values;
forming a text string, wherein said text string includes present ones of said second class speech parameter values; and
outputting said text string to said text to speech synthesizer.

11. The method as recited in claim 10, wherein opening said text to speech synthesizer further includes detecting change between present ones of said first class speech parameter values and previous ones of said first class speech parameter values.

12. The method as recited in claim 11, wherein another text to speech synthesizer is opened with a command string containing command-line arguments formed using current said first class speech parameter values, if change is detected between current said first class speech parameter values and previous said first class speech parameter values.

13. The method as recited in claim 11, wherein said text to speech synthesizer is closed if change is detected between current first class speech parameter values and previous said first class speech parameter values.

14. The method as recited in claim 10, wherein said text string further includes test utterances, said test utterances representing text to be converted to speech by said text to speech synthesizer.

15. The method as recited in claim 10, wherein said step of modifying said current speech parameter values further includes the step of manipulating any combination of parameter scales in said graphical user interface, wherein a position of sliders within said parameter scales determines present ones of said first class speech parameter values and said second class speech parameter values.

16. A speech synthesizer system having a text to speech synthesizer operative to modify acoustical characteristics of a base synthesized voice comprising:

means for manipulating speech parameters that control said acoustical characteristics of said base synthesized voice, wherein said means for manipulating includes a graphical user interface, said graphical user interface including,
parameters scales responsive to input from a user for altering current speech parameter values, wherein a position of adjustment means within said parameter scales determines said current speech parameter values, and
formation means operative to create a text string utilizing said current speech parameter values indicative of change to said text to speech synthesizer in corresponding acoustical characteristics of said base synthesized voice; and
output means for transmitting said text string from said manipulating means.

17. The speech synthesizer system recited in claim 16, wherein said formation means is further operative to create said text string utilizing escape codes, said text string comprising pairs of said escape codes and corresponding said current speech parameter values, said escape codes indicative of particular acoustical characteristics to said text to speech synthesizer in which to alter.

18. The speech synthesizer system recited in claim 16, wherein said graphical user interface further includes input means for entering test utterances, said test utterances representing text to be converted to speech by said text to speech synthesizer, said formation means further operative to create said text string utilizing said test utterances.

19. The speech synthesizer system recited in claim 16 further comprising:

means for initiating said text to speech synthesizer, wherein said text to speech synthesizer is operative to receive said text string from said output means.

20. The speech synthesizer system recited in claim 16 further comprising:

dialogue processing means for preparing a dialogue script to be converted to speech, said dialogue processing including,
means for detecting speaker names in said dialogue script,
means for matching detected speaker names against named voices in a library of named voices, said named voices in said library of named voices having associated speech pa-

rameter values, said associated speech parameter values indicative of change to said text to speech synthesizer in acoustical characteristics of said base synthesized voice,

means for modifying said dialogue script by replacing said detected speaker names with escape sequences, said escape sequences comprising escape codes and said associated speech parameter values, said escape codes indicative of particular acoustical characteristics to said text to speech synthesizer in which to alter, and

means for outputing modified postions of said dialogue script to said text to speech synthesizer.

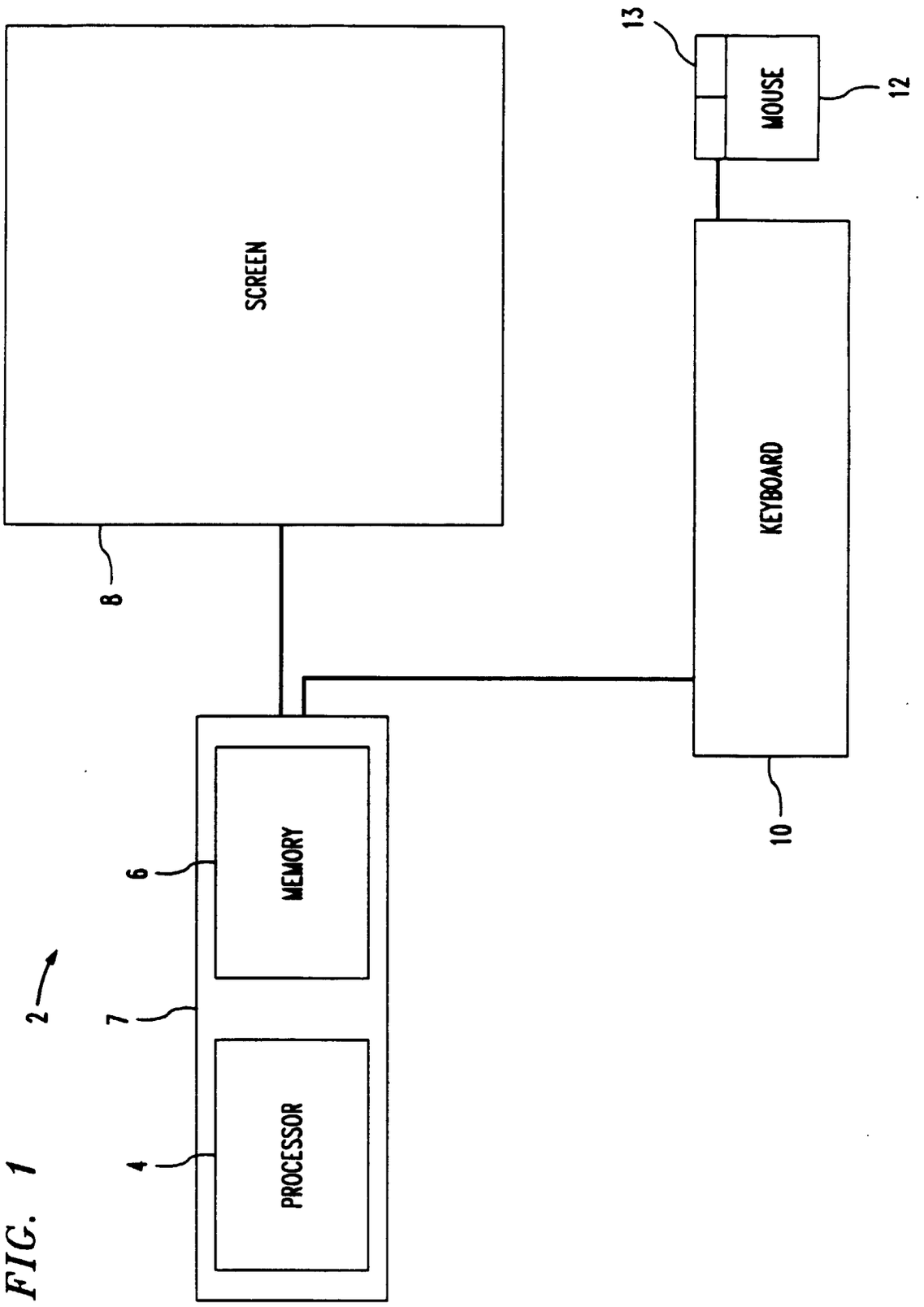


FIG. 1

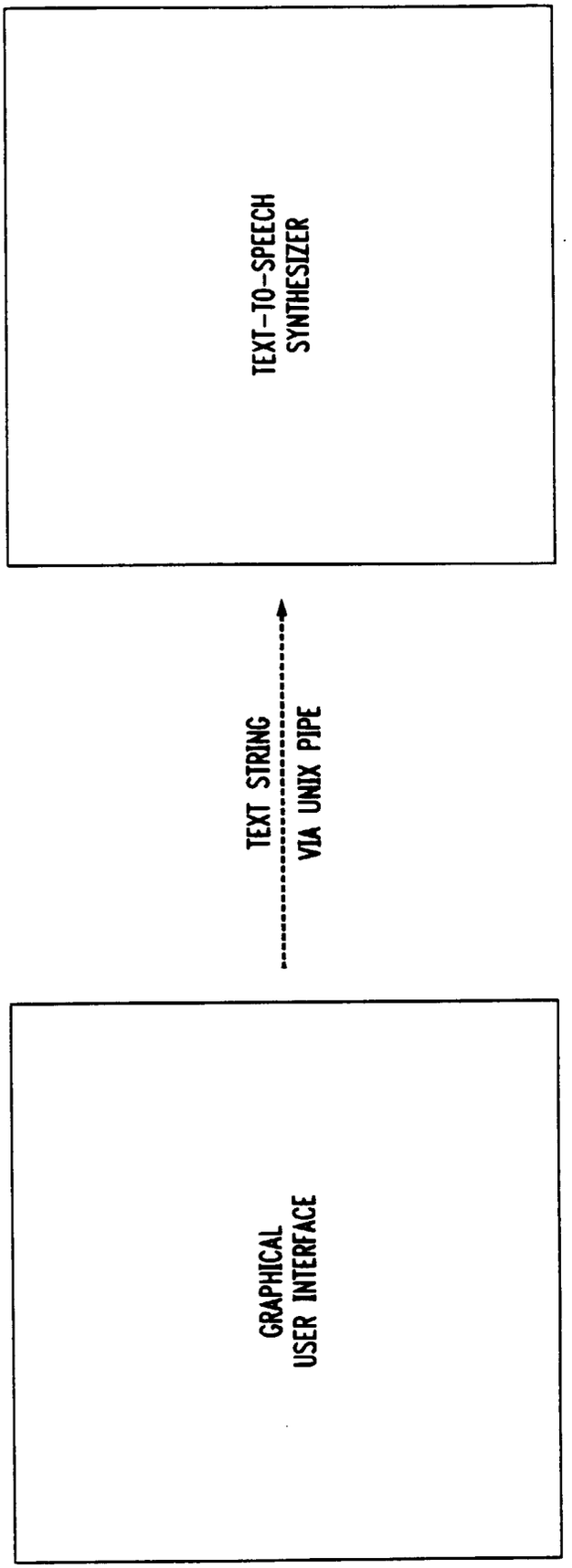
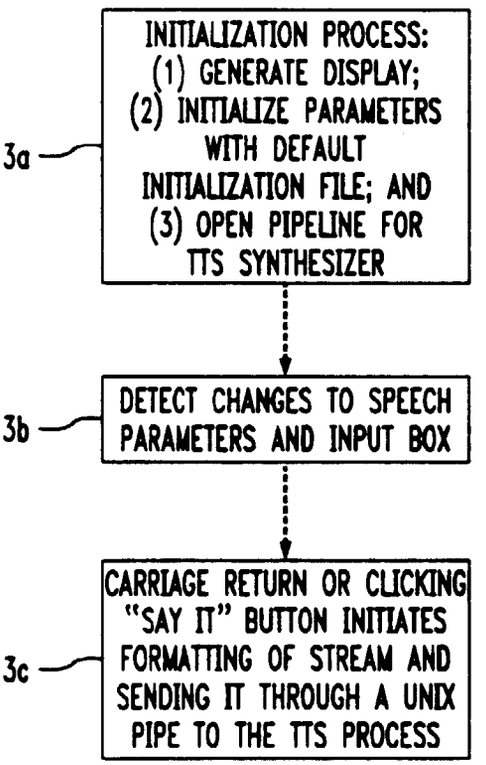

TEXT-TO-SPEECH
SYNTHESIZER
TEXT STRING
VIA UNIX PIPE
GRAPHICAL
USER INTERFACE

FIG. 2

FIG. 3


INITIALIZATION PROCESS:
(1) GENERATE DISPLAY;
(2) INITIALIZE PARAMETERS WITH DEFAULT INITIALIZATION FILE; AND
(3) OPEN PIPELINE FOR TTS SYNTHESIZER
3a
DETECT CHANGES TO SPEECH PARAMETERS AND INPUT BOX
3b
CARRIAGE RETURN OR CLICKING "SAY IT" BUTTON INITIATES FORMATTING OF STREAM AND SENDING IT THROUGH A UNIX PIPE TO THE TTS PROCESS
3c

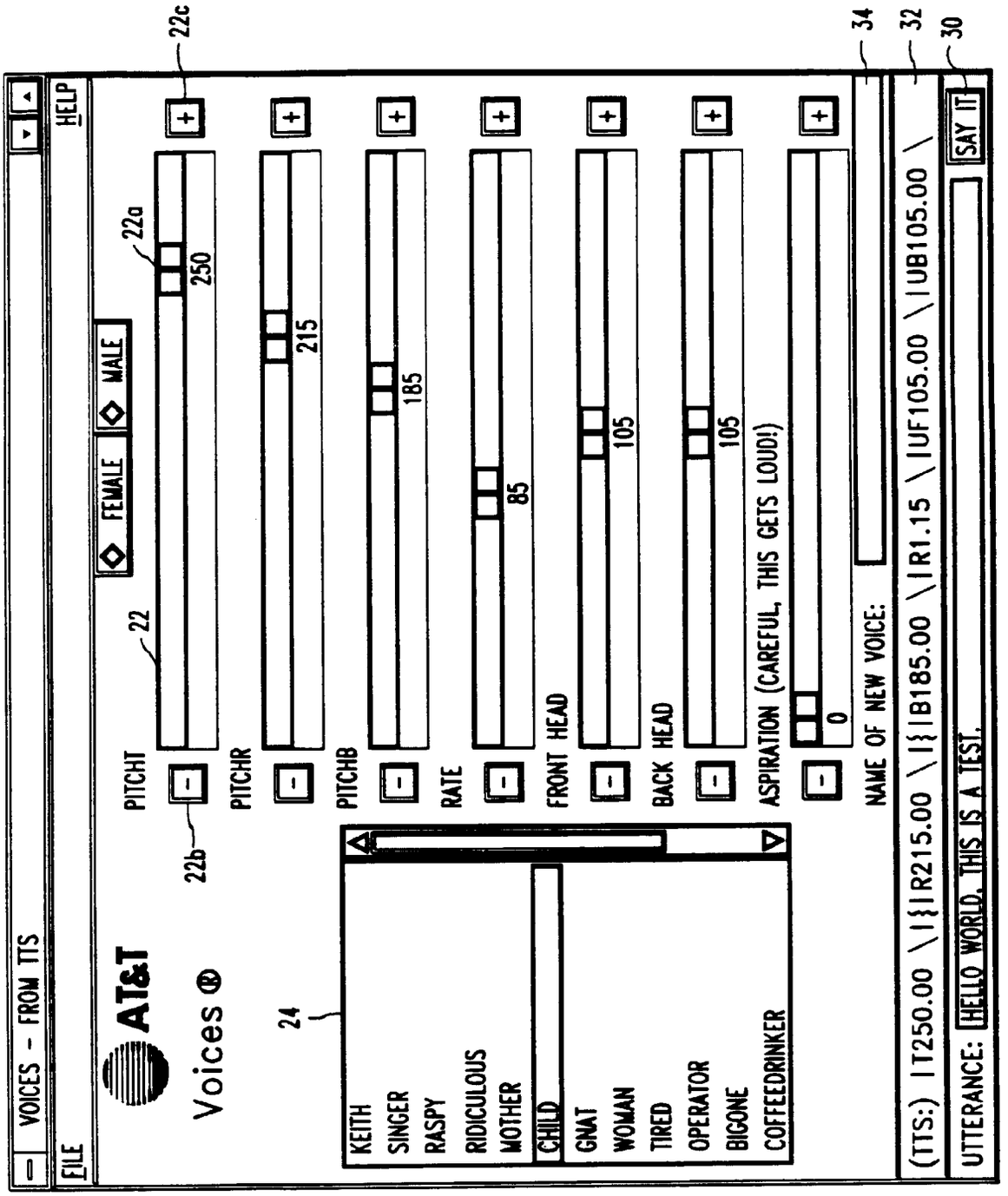

VOICES - FROM TTS
FILE
HELP
AT&T
Voices®
FEMALE
MALE
PITCHT
PITCHR
PITCHB
RATE
FRONT HEAD
BACK HEAD
ASPIRATION (CAREFUL, THIS GETS LOUD!)
NAME OF NEW VOICE:
250
215
185
85
105
105
0
KEITH
SINGER
RASPY
RIDICULOUS
MOTHER
CHILD
GNAT
WOMAN
TIRED
OPERATOR
BIGONE
COFFEEDRINKER
(TTS:) \T250.00 \R215.00 \B185.00 \R1.15 \UF105.00 \UB105.00 \
UTTERANCE: HELLO WORLD. THIS IS A TEST.
SAY IT
20
22
22a
22b
22c
24
30
32
34

FIG. 4

FIG. 5

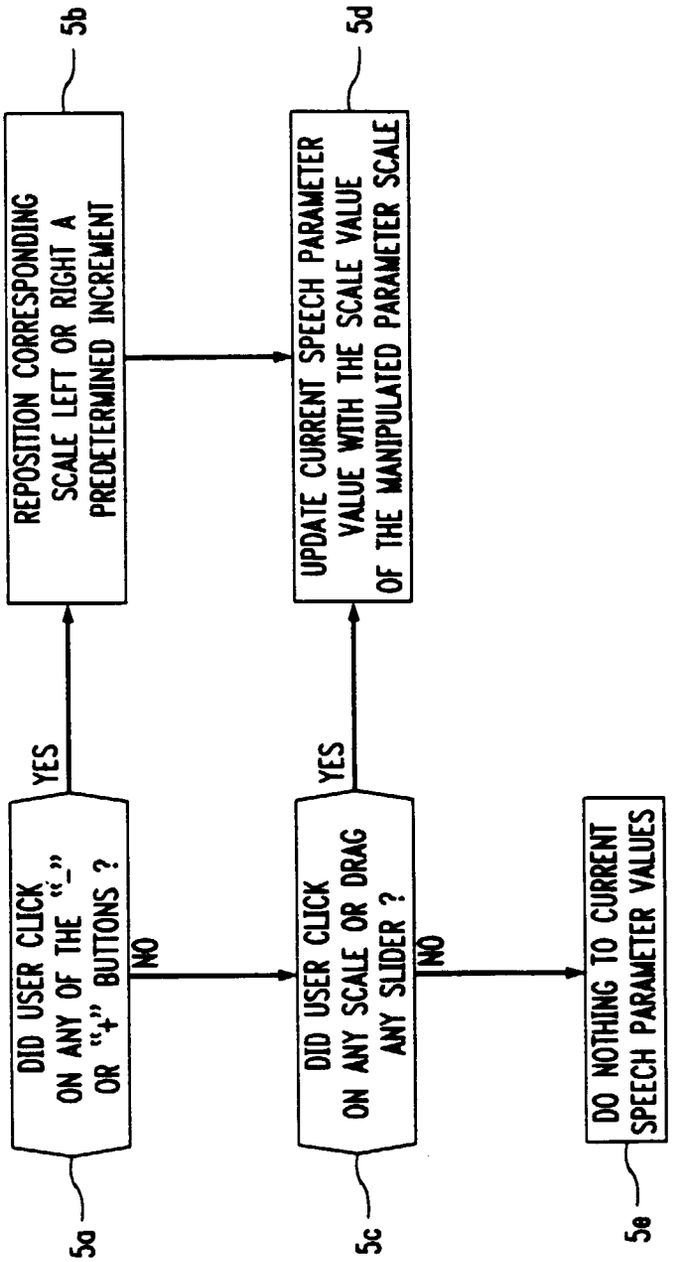

DID USER CLICK ON ANY OF THE "-" OR "+" BUTTONS ?
YES
NO
REPOSITION CORRESPONDING SCALE LEFT OR RIGHT A PREDETERMINED INCREMENT
DID USER CLICK ON ANY SCALE OR DRAG ANY SLIDER ?
YES
NO
UPDATE CURRENT SPEECH PARAMETER VALUE WITH THE SCALE VALUE OF THE MANIPULATED PARAMETER SCALE
DO NOTHING TO CURRENT SPEECH PARAMETER VALUES
5a
5b
5c
5d
5e

FIG. 6

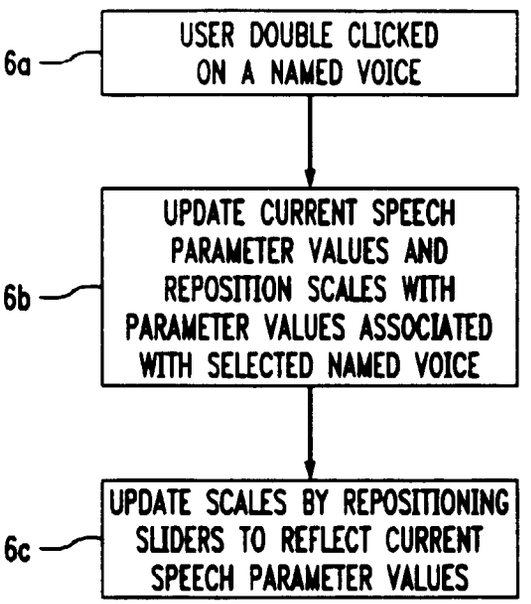

6a
USER DOUBLE CLICKED ON A NAMED VOICE
6b
UPDATE CURRENT SPEECH PARAMETER VALUES AND REPOSITION SCALES WITH PARAMETER VALUES ASSOCIATED WITH SELECTED NAMED VOICE
6c
UPDATE SCALES BY REPOSITIONING SLIDERS TO REFLECT CURRENT SPEECH PARAMETER VALUES

FIG. 7

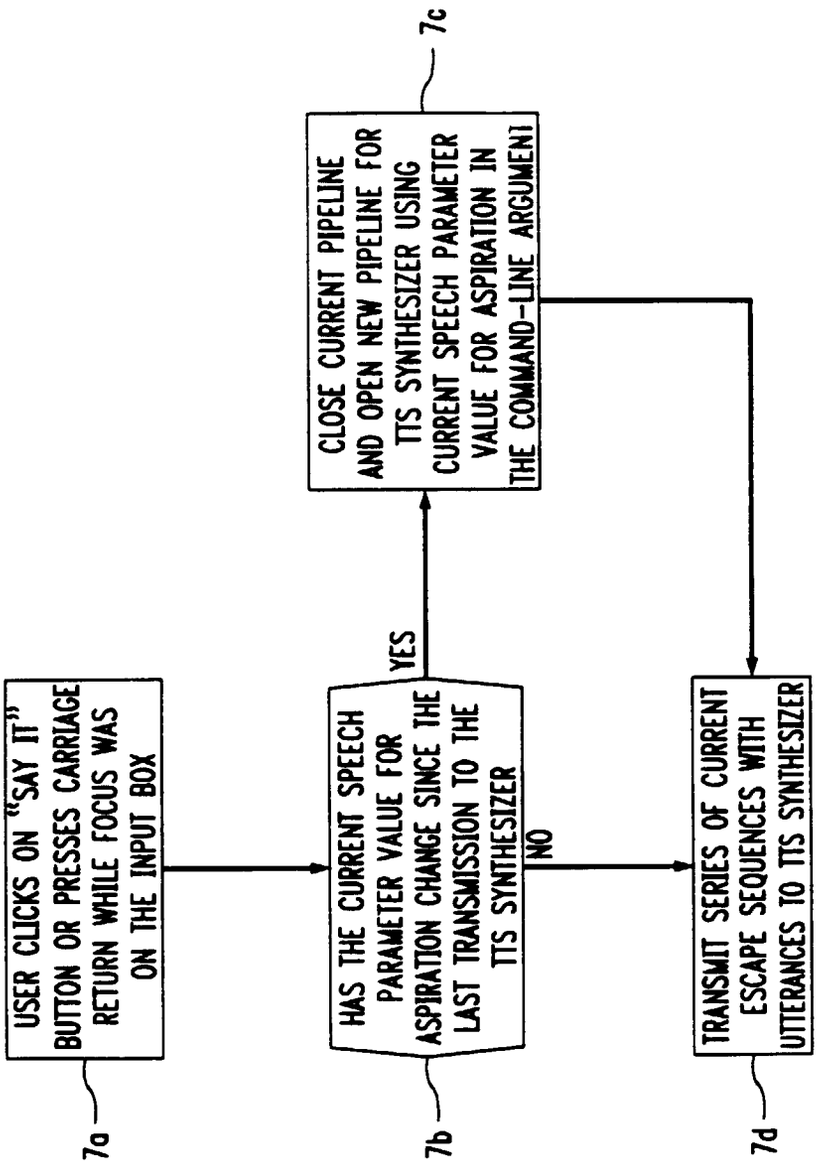

USER CLICKS ON "SAY IT" BUTTON OR PRESSES CARRIAGE RETURN WHILE FOCUS WAS ON THE INPUT BOX
7a
HAS THE CURRENT SPEECH PARAMETER VALUE FOR ASPIRATION CHANGE SINCE THE LAST TRANSMISSION TO THE TTS SYNTHESIZER
7b
YES
NO
CLOSE CURRENT PIPELINE AND OPEN NEW PIPELINE FOR TTS SYNTHESIZER USING CURRENT SPEECH PARAMETER VALUE FOR ASPIRATION IN THE COMMAND-LINE ARGUMENT
7c
TRANSMIT SERIES OF CURRENT ESCAPE SEQUENCES WITH UTTERANCES TO TTS SYNTHESIZER
7d

FIG. 8

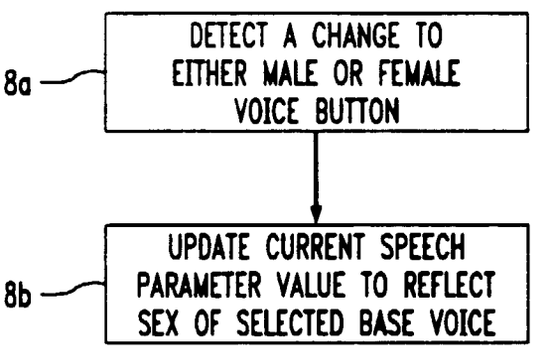

8a
DETECT A CHANGE TO EITHER MALE OR FEMALE VOICE BUTTON
8b
UPDATE CURRENT SPEECH PARAMETER VALUE TO REFLECT SEX OF SELECTED BASE VOICE

FIG. 9

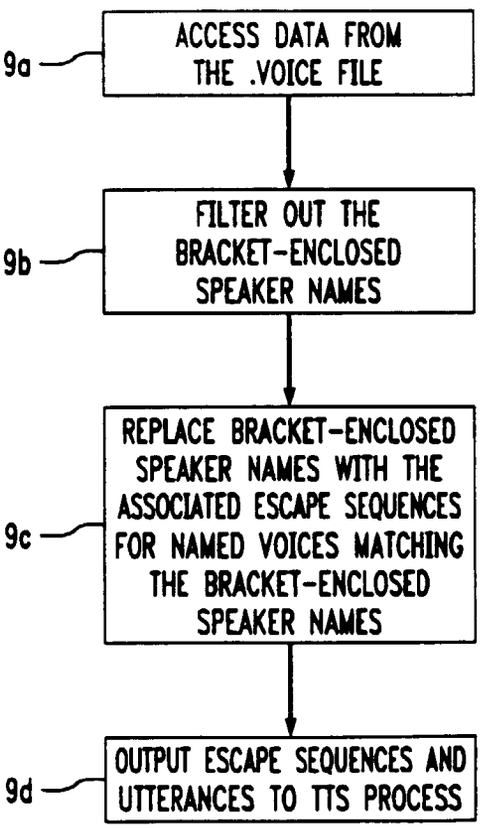

9a
ACCESS DATA FROM THE .VOICE FILE
9b
FILTER OUT THE BRACKET-ENCLOSED SPEAKER NAMES
9c
REPLACE BRACKET-ENCLOSED SPEAKER NAMES WITH THE ASSOCIATED ESCAPE SEQUENCES FOR NAMED VOICES MATCHING THE BRACKET-ENCLOSED SPEAKER NAMES
9d
OUTPUT ESCAPE SEQUENCES AND UTTERANCES TO TTS PROCESS

FIG. 10

[SINGER] HELLO WORLD, I'M A SINGER.

[GNAT] AND I'M A BIT OF A FAST TALKER.

[CHILD] AND I'M JUST A KID.