



(72) REID, Clifford A., US

(72) GLAZER, David, US

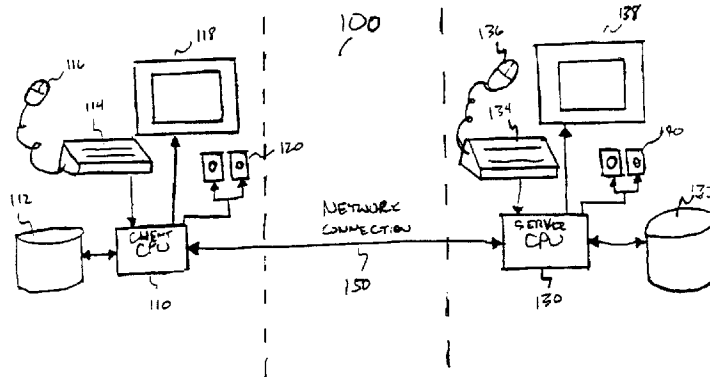
(71) ELOQUENT, INC., US

(51) Int.Cl.<sup>6</sup> H04L 7/10, H04N 7/56, H04N 7/025

(30) 1996/04/26 (08/638,350) US

(54) **PROCEDE ET SYSTEME DE SYNCHRONISATION DE, ET DE NAVIGATION SUR, PLUSIEURS FLUX DE DONNEES ISOCHRONES ET NON ISOCHRONES**

(54) **A METHOD AND SYSTEM FOR SYNCHRONIZING AND NAVIGATING MULTIPLE STREAMS OF ISOCHRONOUS AND NON-ISOCHRONOUS DATA**



(57) L'invention porte sur un procédé et un système de synchronisation de plusieurs flux de données (100) isochrones et non isochrones, et de navigation entre des flux synchronisés en se référant à une base de temps commune (210), et au moyen cadres structurés d'événements conceptuels. Les susdits procédé et système fournissent aux utilisateurs d'ordinateurs un moyen efficace d'interaction entre des représentations sur plusieurs supports d'ordinateurs donnant des conférences (400). Lesdits programmes multisupports consistent en données synchronisées de vidéo, d'audio, de graphiques, de texte, d'hypertexte, et autres enregistrements pouvant être stockées dans un serveur (130). Les utilisateurs peuvent naviguer entre lesdits enregistrement et les lire à partir de l'ordinateur central d'un client (110) par l'intermédiaire d'une liaison de réseau non synchrone (150).

(57) A method and system for synchronizing multiple streams of isochronous and non-isochronous data (100) and navigating through the synchronized streams by reference to a common time base (210) and by means of a structured framework of conceptual events provides computer users with an effective means to interact with multimedia programs of speakers giving presentations (400). The multimedia programs consisting of synchronized video, audio, graphics, text, hypertext, and other data types can be stored on a server (130), and users can navigate and play them from a client CPU (110) over a non-isochronous network connection (150).





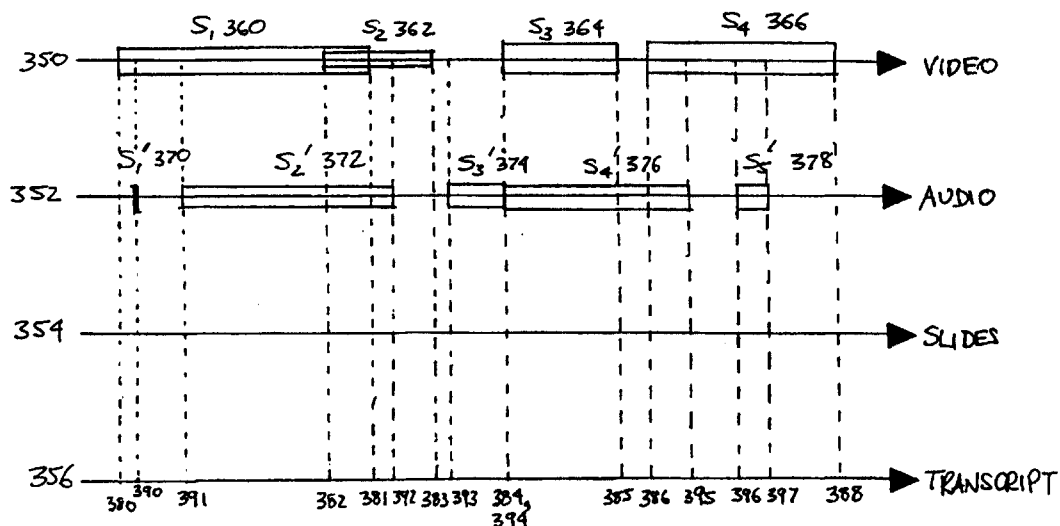
PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification <sup>6</sup> : G06F 3/00, 3/14, 13/00, 13/42</p>	A1	<p>(11) International Publication Number: <b>WO 97/41504</b></p> <p>(43) International Publication Date: 6 November 1997 (06.11.97)</p>
<p>(21) International Application Number: PCT/US97/06982</p> <p>(22) International Filing Date: 24 April 1997 (24.04.97)</p> <p>(30) Priority Data: 08/638,350 26 April 1996 (26.04.96) US</p> <p>(71) Applicant (for all designated States except US): ELOQUENT, INC. [US/US]; Suite 200, 1710 S. Amphlett Boulevard, San Mateo, CA 94402-2703 (US).</p> <p>(72) Inventors; and (75) Inventors/Applicants (for US only): REID, Clifford, A. [US/US]; Apartment 2-710, 2 Townsend Street, San Francisco, CA 94107 (US). GLAZER, David [US/US]; 263 Glenwood Avenue, Woodside, CA 94062 (US).</p> <p>(74) Agents: BARR, Robert et al.; Weil, Gotshal &amp; Manges L.L.P., Suite 280, 2882 Sand Hill Road, Menlo Park, CA 94025 (US).</p>		<p>(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).</p> <p><b>Published</b> <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i></p>

(54) Title: A METHOD AND SYSTEM FOR SYNCHRONIZING AND NAVIGATING MULTIPLE STREAMS OF ISOCHRONOUS AND NON-ISOCHRONOUS DATA



(57) Abstract

A method and system for synchronizing multiple streams of isochronous and non-isochronous data (100) and navigating through the synchronized streams by reference to a common time base (210) and by means of a structured framework of conceptual events provides computer users with an effective means to interact with multimedia programs of speakers giving presentations (400). The multimedia programs consisting of synchronized video, audio, graphics, text, hypertext, and other data types can be stored on a server (130), and users can navigate and play them from a client CPU (110) over a non-isochronous network connection (150).

**A METHOD AND SYSTEM FOR SYNCHRONIZING AND NAVIGATING  
MULTIPLE STREAMS OF ISOCHRONOUS AND NON-ISOCHRONOUS  
DATA**

5 BACKGROUND OF THE INVENTION

1. Field of the Invention

10 The present invention generally relates to the production and delivery of video recordings of speakers giving presentations, and, more particularly, to the production and delivery of digital multimedia programs of speakers giving presentations. These digital multimedia programs consist of multiple synchronized streams of isochronous and non-isochronous data, including video, audio, graphics, text, hypertext, and other data types.

15

2. Description of the Prior Art

20 The recording of speakers giving presentations, at events such as professional conferences, business or government organizations' internal training seminars, or classes conducted by educational institutions, is a common practice. Such recordings provide access to the content of the presentation to individuals who were not able to attend the live event.

25 The most common form of such recordings is analog video taping. A video camera is used to record the event onto a video tape, which is subsequently duplicated to an analog medium suitable for distribution, most commonly a VHS

tape, which can be viewed using a commercially-available VCR and television set. Such video tapes generally contain a video recording of the speaker and a synchronized audio recording of the speaker's words. They may also contain a video recording of any visual aids which the speaker used, such as text or graphics projected in a manner visible to the audience. Such video tapes may also be edited prior to duplication to include a textual transcript of the audio component recording, typically presented on the bottom of the video display as subtitles. Such subtitles are of particular use to the hearing impaired, and if translated into other languages, are of particular use to viewers who prefer to read along in a language other than the language used by the speaker.

Certain characteristics of such analog recordings of speakers giving presentations are unattractive to producers and to viewers. Analog tape players offer limited navigation facilities, generally limited to fast forward and rewind capabilities. In addition, analog tapes have the capacity to store only a few hours of video and audio, resulting in the need to duplicate and distribute a large number of tapes, leading to the accumulation of a large number of such tapes by viewers.

Advancements in computer technology have allowed analog recordings of speakers giving presentations to be converted to digital format, stored on a digital storage medium, such as a CD-ROM, and presented using a computer CPU and display, rather than a VCR and a television set. Such digital recordings generally include both isochronous and non-isochronous data. Isochronous data is data that is time ordered and must be presented at a particular rate. The isochronous data contained in such a digital recording generally includes video and audio. Non-isochronous data may or may not be time ordered, and need not be presented at a particular rate. Non-isochronous data contained in such a digital recording may include graphics, text, and hypertext.

The use of computers to play digital video recordings of speakers giving presentations provides navigational capabilities not available with analog video tapes. Computer-based manipulation of the digital data offers random access to any point in the speech, and if there is a text transcript, allows the users to search  
5 for words in the transcript to locate a particular segment of the speech.

Certain characteristics of state-of-the-art digital storage and presentation of recordings of speakers giving presentations are unattractive to producers and to viewers. There is no easy way to navigate directly to a particular section of a  
10 presentation that discusses a topic of particular interest to the user. In addition, there is no easy way to associate a table of contents with a presentation, and navigate directly to section of the presentation associated with each entry in the table of contents. Finally, like analog tapes, CD-ROMs can store only a view  
15 hours of digital video and audio, resulting in the need to duplicate and distribute a large number of CD-ROMs, leading to the accumulation of a large number of such CD-ROMs by viewers.

#### SUMMARY OF THE INVENTION

20 It is therefore an object of the present invention to provide a mechanism for synchronizing multiple streams of isochronous and non-isochronous digital data in a manner that supports navigating by means of a structured framework of conceptual events.

25 It is another object of the invention to provide a mechanism for navigating through any stream using the navigational approach most appropriate to the structure and content of that stream.

It is another object of the invention to automatically position each of the streams at the position corresponding to the selected position in the navigated stream, and simultaneously display some or all of the streams at that position.

5 It is another object of the invention to provide for the delivery of programs made up of multiple streams of synchronized isochronous and non-isochronous digital data across non-isochronous network connections.

10 In order to accomplish these and other objects of the invention, a method and system for manipulating multiple streams of isochronous and non-isochronous digital data is provided, including synchronizing multiple streams of isochronous and non-isochronous data by reference to a common time base, supporting navigation through each stream in the manner most appropriate to that stream, defining a framework of conceptual events and allowing a user to navigate though  
15 the streams using this structured framework, identifying the position in each stream corresponding to the position selected in the navigated stream, and simultaneously displaying to the user some or all of the streams at the position corresponding to the position selected in the navigated stream. Further, a method and system of efficiently supporting sequential and random access into streams of isochronous and  
20 non-isochronous data across non-isochronous networks is provided, including reading the isochronous and non-isochronous data from the storage medium into memory of the server CPU, transmitting the data from the memory of the server CPU to the memory of the client CPU, and caching the different types of data in the memory of the client CPU in a manner that ensures continuous display of the  
25 isochronous data on the client CPU display device.

BRIEF DESCRIPTION OF THE DRAWINGS

The foregoing and other objectives, aspects, and advantages of the present invention will be better understood from the following detailed description of  
5 embodiments thereof with reference to the following drawings.

FIG. 1 is a schematic diagram of the organization of a data processing system incorporating an embodiment of the present invention.

10 FIGS. 2 and 3 are schematic diagrams of the organization of the data in an embodiment of the present invention.

FIG. 4 is a diagram showing how two different sets of "conceptual events" may be associated with the same presentation in an embodiment of the present  
15 invention.

FIGS. 5, 6 and 9 are exemplary screens produced in accordance with an embodiment of the present invention.

20 FIGS. 7, 8, 10, and 11 are flow charts indicating the operation of an embodiment of the present invention.

DETAILED DESCRIPTION OF AN EMBODIMENT OF THE INVENTION

25

Referring now to the drawings, and more particularly to FIG. 1, there is shown, in schematic representation, a data processing system 100 incorporating the invention. Conventional elements of the system include a client central processing unit 110 which includes high-speed memory, a local storage device 112 such as a  
30 hard disk or CD-ROM, input devices such as keyboard 114 and pointing device

116 such as a mouse, and a visual data presentation device 118, such as a computer display screen, capable of presenting visual data perceptible to the senses of a user, and an audio data presentation device 120, such as speakers or headphones, capable of presenting audio data to the senses of a user. Other conventional elements of  
5 the system include a server central processing unit 130 which includes high-speed memory, a local storage device 132 such as a hard disk or CD-ROM, input devices such as keyboard 134 and pointing device 136, and a visual data presentation device 138, and an audio data presentation device 140. The client CPU is connected to the server CPU by means of a network connection 150.

10

The invention includes three basic aspects: (1) synchronizing multiple streams of isochronous and non-isochronous data, (2) navigating through the synchronized streams of data by means of a structured framework of conceptual events, or by means of the navigational method most appropriate to each stream,  
15 and (3) delivering the multiple synchronized streams of isochronous and non-isochronous data over a non-isochronous network connecting the client CPU and the server CPU.

An exemplary form of the organization of the data embodied in the  
20 invention is shown in FIG. 2 and FIG. 3. Beginning with FIG. 2, the video/audio stream 200 is of a type known in the art capable of being played on a standard computer equipped with the appropriate video and audio subsystems, such as shown in FIG. 1. An example of such a video/audio stream is Microsoft Corporation's AVI™ format, which stands for "audio/video interleaved." AVI™ and other such  
25 video/audio formats consist of a series of digital images, each referred to as a "frame" of the video, and a series of samples that make up the digital audio. The frames are spaced equally in time, so that displaying consecutive frames on a display device at a sufficiently high and constant rate produces the sensation of continuous motion to the human perceptual system. The rate of displaying frames  
30 typically must exceed ten to fifteen frames per second to achieve the effect of



continuous motion. The audio samples are synchronized with the video frames, so that the associated audio can be played in synchronization with the displayed video images. Both the digital images and digital audio samples may be compressed to reduce the amount of data that must be stored or transmitted.

5

A time base **210** associates a time code with each video frame. The time base is used to associate other data with each frame of video. The audio data, which for the purposes of this invention consists primarily of spoken words, is transcribed into a textual format, called the Transcript **220**. The transcript is synchronized to the audio data stream by assigning a time code to each word, producing the Time-Coded Transcript **225**. The time codes (shown in angle-brackets) preceding each word in the Time-Coded Transcript correspond to the time at which the speaker begins pronouncing that word. For example, the time code **230** of 22.51 s is associated with the word **235** "the." The Time-Coded Transcript may be created manually or by means of an automatic procedure. Manual time-coding requires a person to associate a time code with each word in the transcript. Automatic time coding, for example, uses a speech recognition system of a type well-known in the art to automatically assign a time code to each word as it is recognized and recorded. The current state of the art of speech recognition systems renders automatic time coding of the transcript less economical than manual time coding.

Referring now to FIG. 3, the set **310** of Slides S1 **311**, S2 **312**, ... that the speaker used as part of the presentation may be stored in an electronic format of any of the types well-known in the art. Each slide may consist of graphics, text, and other data that can be rendered on a computer display. A Slide Index **315** assigns a time code to each Slide. For example, Slide S1 **311** would have a time code **316** of 0 s, S2 **312** having a time code **317** of 20.40 s, and so on. The time code corresponds to the time during the presentation at which the speaker caused the specified Slide to be presented. In one embodiment, all of the Slides are

30

contained in the same disk file, and the Slide Index contains pointers to the locations of each Slide in the disk file. Alternatively, each Slide may be stored in a separate disk file, and the Slide Index contains pointers to the files containing the Slides.

5

An Outline **320** of the presentation is stored as a separate text data object. The Outline is a hierarchy of topics **321**, **322**, .. that describe the organization of the presentation, analogous to the manner in which a table of contents describes the organization of a book. The outline may consist of an arbitrary number of entries, and an arbitrary number of levels in the hierarchy. An Outline Index **325** assigns a time code to each entry in the Outline. The time code corresponds to the time during the presentation at which the speaker begins discussing the topic represented by the entry in the Outline. For example, topic **321**, "Introduction" has entry name "01" and time code **326** of 0 s, topic **322** "The First Manned Flight" has entry name "02" and time code **327** of 20.50 s, "The Wright Brothers" **323** has entry name "021" (and hence is a subtopic of topic **322**) with time code **328** of 120.05 s, and so on. The Outline and the Outline Index may be created by means of a manual or an automatic procedure. Manual creation is accomplished by a person viewing the presentation, authoring the Outline, and assigning a time code to each element in the outline. Automatic creation may be accomplished by automatically constructing the outline consisting of the titles of each of the Slides, and associating with each entry on the Outline the time code of the corresponding Slide. Note that manual and automatic creation may produce different Outlines.

10  
15  
20  
25

The set **330** of Hypertext Objects **331**, **332**, ... relating to the subject of the presentation may be stored in an electronic formats of various types well-known in the art. Each Hypertext Object may consist of graphics, text, and other data that can be rendered on a computer display, or pointers to other software applications, as spreadsheets, word processors, and electronic mail systems, as well as more

specialized applications such as proficiency testing applications or computer-based training applications.

5 A Hypertext Index table **335** is used to assign two time codes and a display location to each Hypertext Object. The first time code **336** corresponds to the earliest time during the presentation at which the Hypertext Object relates to the content of the presentation. The second time code **337** corresponds to the latest time during the presentation at which the Hypertext Object relates to the content of the presentation. The Object Name **338**, as the name suggests, denotes the  
10 Hypertext Object's name. The display location **339** denotes how the connection to the Hypertext Object, referred to as the Hypertext Link, is to be displayed on the computer screen. Hypertext Links may be displayed as highlighted words in the Transcript or the Slides, as buttons or menu items on the end-user interface, or in other visual presentation that may be selected by the user.

15

It may be appreciated by one of ordinary skill in the art that other data types may be synchronized to the common time base in a manner similar to the approaches used to synchronize the video/audio stream with the Transcript, the Slides, and the Hypertext Objects. Examples of such other data types include  
20 animations, series of computer screen images, and other specialty video streams.

An Outline represents an example of what is termed here a set of "conceptual events." A conceptual event is an association one makes with a segment of a data stream, having a beginning and end (though the beginning and end may be the points), that represents something of interest. These data segments  
25 delineating a set of conceptual events may overlap each other, and furthermore, need not cover the entire data stream. An Outline represents a set of conceptual events that does cover the entire data stream and, if arranged hierarchically, such as with sections and subsections, has sections covering subsections. In the Outline  
30 **320** of FIG. 3, one has the sections 01:"Introduction" **321**, 02:"The First Manned

Flight" 322 , and so on, covering the entire presentation. The subsections 021:"The Wright Brothers" 324, 022:"Failed Attempts" 324 and so on, represents another coverage of the same segment as 02:"The First Manned Flight" 322. In accordance with the principles of the present invention, multiple Outlines, created manually or automatically, may be associated with the same presentation, thereby allowing different users with different purposes in viewing the presentation to use the Outline most suitable for their purposes. These Outlines have been described from the perspective of having been created beforehand, but there is no reason, under the principles of the present invention, for this to be so. It should be readily understood by one of ordinary skill in the art that a similar approach would allow a user to create a set of "bookmarks" that denote particular segments, or user-chosen "conceptual events" within presentations. The bookmarks allow the user, for example, to return quickly to interesting parts of the presentation, or to pick up at the previous stopping point.

With reference to FIG. 4, the implementation of sets of conceptual events may be understood. There are time lines representing the various data streams, as for example, video 350, audio 352, slides 354 and transcript 356. There are two sets of conceptual events or data segments of these time lines shown,  $S_1$  360,  $S_2$  362,  $S_3$  364,  $S_4$  366, ... and  $S'_1$  370,  $S'_2$  372,  $S'_3$  374,  $S'_4$  376 ,  $S'_5$  378, ..., the first set indexed into the video 350 stream and second set indexed into the audio 352 stream. Thus, the first set  $S_1$  360,  $S_2$  362,  $S_3$  364, etc., would respectively invoke time codes 380 and 381, 382 and 383, 384 and 385, etc., not only for the video 350 data stream, but for the audio 352 , slides 354 and transcript 356 streams. Similarly, the second set  $S'_1$  370,  $S'_2$  372,  $S'_3$  384, etc., would invoke respectively time codes 390 (a point), 391 and 392, 393 and 394 (394 shown collinear with 384, whether by choice or accident), etc., respectively, not only on the audio 352 data stream, but on the video 350, slides 354 and transcript 356 streams. Consider the following example of a presentation of ice skating performed to music, with voice-over commentaries and slides showing the relative standings of the ice

skaters. A first Outline might list each skater and be broken down further into the individual moves of each skater's program. A second Outline might track the musical portion of the audio stream, following the music piece to piece, even movement to movement. Thus, one user might be interested in how a skater performed a particular move, while another user might wish to study how a particular passage of music inspired a skater to make a particular move. Note that there is no requirement that two sets of conceptual events track each other in any way, they represent two different ways of studying the same presentation. Furthermore, the examples showed sets of conceptual events indexed into isochronous data streams; it may be appreciated by someone of ordinary skill in the art that sets of conceptual events may be indexed into non-isochronous data streams as well. As was stated earlier, an Outline for a presentation may be indexed to the slide stream.

Referring now to the exemplary screen shown in FIG. 5, the exemplary screen 400 shows five windows 410, 420, 430, 440, 450 contained within the display. The Video Window 410 is used to display the video stream. The Slide Window 420 is used to display the slides used in the presentation. The Transcript Window 430 is used to display the transcribed audio of the speech. The Outline Window 440 is used to display the Outline of the presentation. The Control Panel 450 is used to control the display in each of the other four windows. The Transcript Window 430 includes a Transcript Slider Bar 432 that allows the user to scroll through the transcript, and Next 433 and Previous 434 Phrase Buttons that allow the user to step through the transcript a phrase at a time, where a phrase consists of a single line of the transcript. It also includes a Hypertext Link 436, as illustrated here in the form of the highlighted words, "Robert Jones", in the transcript. The Outline Window 440 includes an Outline Slider Bar 442 that allows the user to scroll through the outline, and Next 443 and Previous Entry buttons 444 that allow the user to jump directly to the next or previous topic. The Control Panel 450 includes a Video Slider Bar 452 used to select a position in the video

stream, and a Play Button **454** used to play the program. It also includes a Slider Bar **456** used to position the program at a Slide, and Previous **457** and Next **458** Slide Buttons used to display the next and previous Slides in the Slide Window **420**. It also includes a Search Box **460** used to search for text strings (e.g., words) in the Transcript.

FIG. 5 shows the beginning of a presentation, corresponding to a time code of zero. The speaker's first slide is displayed in the Slide Window **410**, the speaker's first words are displayed in the Transcript Window **430**, and the beginning of the outline is displayed in the Outline Window **440**. The user can press the play button **454** to begin playing the presentation, which will cause the video and audio data to begin streaming, the transcript and outline scroll in synchronization with the video and audio, and the slides to advance at the appropriate times.

Alternatively, the user can jump directly to a point of interest. FIG. 6 shows the result of the user selecting the second entry in the Outline from Outline Window **440'**, entitled "The First Manned Flight" (recall entry **322** of Outline **320** in FIG. 3). From the Outline Index **327** in FIG. 3, the system determines that the time code **327** of "The First Manned Flight" is 20.50 s. The system looks in the Slide Index **315** (also in FIG. 3) and determines that the second slide S2 begins at time code **317** of 20.40 s, and thus the second slide should be displayed in the Slide Window **420'**. The system looks at the Time-Coded Transcript **215** (shown in FIG. 2), locates the word "the" **235** that begins on or immediately after time code of 20.50 s, and displays that word and the appropriate number of subsequent words to fill up the Transcript Window **430'**. The effect of this operation is that the user is able to jump directly to a point in the presentation, and the system positions each of the synchronized data streams to that point, including the video in Video Window **410'**. The user may then begin playing the presentation at this

point, or upon scanning the newly displayed slide and transcript jump directly to another point in the presentation.

Referring now to FIG. 7, the flowchart starting at **600** indicates the operation of an embodiment of the present invention. When the user slides the video slider bar **452** in FIG. 5, the Event Handler **601** in FIG. 7 receives a Move Video Slider Event **610**. The Move Video Slider Event **610** causes the invention to calculate the video frame of the new position of the slider **452**. The position of the video slider **452** is translated into the position in the video data stream in a proportional fashion. For example, if the new position of the video slider **452** is positioned half-way along its associated slider bar, and the video stream consist of 10,000 frames of video, then the 5,000<sup>th</sup> frame of video is displayed on the Video Window **420**. The invention displays the new video frame **611**, and computes the time code of the new video frame **612**. Using this new time code, the system looks up the Slide associated with the displayed video frame, and displays **613** the new Slide in the Slide Window **410**. Again using this new time code, the system looks up the Phrase associated with the displayed video frame, and displays the new Phrase **614** in the Transcript Window **430**. Again using this new time code, the system looks up the Outline Entry associated with the displayed video frame, and displays the new Outline Entry **615** in the Outline Window **440**. Finally, using this new time code, the system looks up the Hypertext Links associated with the displayed video frame, and displays them **616** in the appropriate place in the Transcript Window **430**.

Referring back to FIG. 5, when the user moves the Slide Slider Bar **456** or presses the Previous **457** and Next **458** Slide Buttons, the Event Handler **601** in FIG. 7 receives a New Slide Event **620**. The New Slide Event causes the system to display the selected new Slide **621** in the Slide Window **420**, and to look up the time code of the new Slide in the Slide Index **622**. Using the time code of the new Slide as the new time code, the system computes the video frame associated with

the new time code and displays the indicated video frame **623** in the Video Window. Again using the new time code, the system looks up the Phrase associated with the displayed Slide, and displays the new Phrase **624** in the Transcript Window **430**. Again using the new time code, the invention looks up the  
5 Outline Entry associated with the displayed Slide, and displays the new Outline Entry **625** in the Outline Window **440**. Finally, using the new time code, the system looks up the Hypertext Links associated with the displayed Slide, and displays them **626** in the appropriate place in the Transcript Window **430**.

10 Referring again back to FIG. 5, when the user moves the Transcript Slider Bar **432** or presses the Next **433** or Previous **434** Phrase Buttons, the Event Handler **601** in FIG. 7 receives a New Phrase Event **630**. The New Phrase Event causes the system to display the selected new Phrase **631** in the Transcript Window **430**, and to look up the time code of the new Phrase in the Transcript Index **632**.

15 Using the time code of the new Phrase as the new time code, the invention computes the video frame associated with the new time code and displays the indicated video frame **633** in the Video Window **410**. Again using the new time code, the invention looks up the Slide associated with the displayed Phrase, and displays the new Slide **634** in the Slide Window. Again using the new time code,  
20 the invention looks up the Outline Entry associated with the displayed Phrase, and displays the new Outline Entry **635** in the Outline Window **440**. Finally, using the new time code, the invention looks up the Hypertext Links associated with the displayed Phrase, and displays them **636** in the appropriate place in the Transcript Window **430**.

25 Referring yet again to FIG. 5, when the user types a search string into the Search Box **460** and initiates a search, the Event Handler **601** in FIG. 7 receives a Search Transcript Event **640**. The Search Transcript event causes the system to employ a string matching algorithm of a type well-known in the art to scan the  
30 Transcript and locate the first occurrence of the search string **641**. The system uses



the Transcript Index to determine which Phrase contains the matched string in the Transcript **642**. The system displays the selected new Phrase **631** in the Transcript Window, and looks up the time code of the new Phrase in the Transcript Index **632**. Using the time code of the new Phrase as the new time code, the system  
5 computes the video frame associated with the new time code and displays the indicated video frame **633** in the Video Window **410**. Again using the new time code, the system looks up the Slide associated with the displayed Phrase, and displays the new Slide **634** in the Slide Window **420**. Again using the new time code, the system looks up the Outline Entry associated with the displayed Phrase,  
10 and displays the new Outline Entry **635** in the Outline Window **440**. Finally, using the new time code, the system looks up the Hypertext Links associated with the displayed Phrase, and displays them **636** in the appropriate place.

Referring to FIG. 5, when the user moves the Outline Slider Bar **442** or  
15 presses the Next **443** or Previous **444** Outline Entry Buttons, the Event Handler **601** in FIG. 7 receives a New Outline Entry Event **650**. The New Outline Entry Event causes the system to display the selected new Outline Entry **651** in the Outline Window **440**, and to look up the time code of the new Outline Entry in the Outline Index **652**. Using the time code of the new Outline Entry as the new time code,  
20 the system computes the video frame associated with the new time code and displays the indicated video frame **653** in the Video Window **410**. Again using the new time code, the system looks up the Slide associated with the displayed Outline Entry, and displays the new Slide **654** in the Slide Window **420**. Again using the new time code, the system looks up the Phrase associated with the displayed  
25 Outline Entry, and displays the new Phrase **655** in the Transcript Window **430**. Finally, using the new time code, the system looks up the Hypertext Links associated with the displayed Outline Entry, and displays them **656** in the appropriate place in the Transcript Window **430**.

Referring again to FIG. 5, when the user selects a Hypertext Link 436, the Event Handler 601 in FIG. 7 receives a Display Hypertext Object 660. The system displays the data object pointed to by the selected Hypertext Link 661.

5 Whenever the system is in a stationary state, that is, when no video/audio stream is being played, the system maintains a record of the current time code. The data displayed in FIGS. 4 and 5 always correspond to the current time code. When the user presses the Play Button 454, the Event Handler 601 in FIG. 5 receives a Play Program Event 670. The system begin playing the video and audio  
10 streams, starting at the current time code. Referring now to FIG. 8, as each new video frame is displayed 700, the system uses the time code of the displayed video frame to check the Transcript Index, the Slide Index, the Outline Index, and Hypertext Index and determine if the data displayed in the Slide Window 420, Transcript Window 430, or Outline Window 440 must be updated, or if new  
15 Hypertext Links must be displayed in the Transcript Window 430. If the time code of the new video frame corresponds to the time code of the next Phrase 710, the system displays the next Phrase 711 in the Transcript Window 430. If the time code of the new video frame corresponds to the time code of the next Slide 720, the system displays the next Slide 721 in the Slide Window 420. If the time code  
20 of the new video frame corresponds to the time code of the next Outline Entry 730, the system displays the next Outline Entry 731 in the Outline Window 440. Finally, if the time code of the new video frame corresponds to the time codes of a different set of Hypertext Links than are currently displayed 740, the system displays the new set of Hypertext Links 741 at the appropriate places on the  
25 display in the Transcript Window 430.

It may be appreciated by one of ordinary skill in the art that the textual transcript may be translated into other languages. Multiple transcripts, corresponding to multiple languages, may be synchronized to the same time base,  
30 corresponding to a single video/audio stream. Users may choose which transcript

language to view, and may switch among different transcripts in different languages during the operation of the invention.

Furthermore, multiple synchronized streams of each data type may be  
5 incorporated into a single multimedia program. Multiple video/audio streams, each  
corresponding to different video resolution, audio sampling rate, or data  
compression technology, may be included in a single program. Multiple sets of  
slides, hypertext links, and other streams of isochronous data types may also be  
included in a single program. One or more of each data type may be displayed on  
10 the computer screen, and users may switch among the different streams of data  
available in the program.

The present invention is compatible with operating with a collection of  
many presentations, and to assist users in locating the particular portion of the  
15 particular presentation that most interests them. The presentations are stored in a  
data base of a type well-known in the art, which may range from a simple non-  
relational data base that stores data in disk files to a complex relational or object-  
oriented data base that stores data in a specialized format. Referring to the  
exemplary screen **800** depicted in FIG. **9**, users can issue structured queries or full  
20 text queries to identify programs they wish to view. The user types in a query in  
the query type-in box **810**. The titles of the programs that match the query are  
displayed in the results box **820**. Structured queries are queries that allow the user  
to select programs on the basis of structured information associated with each  
program, such as title, author, or date. Using any of the structured query engines  
25 well-known in the art, the user can specify a particular title, author, range of dates,  
or other structured query, and select only those programs which have associated  
structured information that matches the query. Full-text queries are queries that  
allow the user to select programs on the basis of text associated with each program,  
such as the abstract, transcript, slides, or ancillary materials connected via  
30 hypertext. Using any of the full-text search engines known in the art, the user can

specify a particular combination of words and phrases, and select only those programs which have associated text that matches the full-text query. Users can also select which of the associated text elements to search. For example, the user can specify to search only the transcript, only the slides, or a combination of both.

5 When the text associated with a program matches the user's query, the user can jump directly to the matched text, and display all of the other synchronized multimedia data types at that point in the program.

Full-text queries can be manually constructed by the users, or they can be automatically constructed by the invention. Such automatically-constructed queries are referred to as "agents." FIG. 10 presents a flow chart of the agent mechanism starting at 900. When the user displays a program 910, the system constructs a summary of the program 920. The summary of the program may be constructed in multiple alternative ways. Each program may have associated with it a list of

15 keywords that describe the major subjects discussed in the program. In this case, constructing the summary simply involves accessing this predefined list of keywords. Alternatively, any text summarization engine well-known in the art may be run across the text associated with program, including the abstract, the transcript, and the slides, to generate a list of keywords that describe the major

20 subjects discussed in the program. This summary is added to the user's profile 930. The user's profile is a list of keywords that collectively describe the programs that the user has viewed in the past. Each time the user views a new program, the keywords that describe that program are added to the user's profile. In this manner, the agent "learns" which subjects are most interesting to the user, and

25 continues to learn about the user's changing interests as the user uses the system. The agent mechanism also incorporates the concept of memory. Each keyword that is added to the user's profile is labeled with the date at which its associated program was viewed. Whenever the agent mechanism is initiated, the difference between the current date and the date label on each keyword is used to assess the

30 relative importance of that keyword. Keywords that entered the profile more

recently are treated as more important than keywords that entered the profile in the distant past. On specified events, such as the user logging into the system, the agents mechanism is initiated 901. The system creates a query from the current user's profile 940. The list of keywords in the profile are reorganized into the query syntax required by the full-text search engine. The ages of the keywords are converted into the relative importance measure required by the full-text search engine. The query is run against all of the programs on the server 950, and the resulting list of programs are presented to the user 960. This list of programs constitutes the programs which the system has determined may be of interest to the user, based on the user's past viewing behavior.

In addition, users can create their own agents by manually constructing a query that describe their ongoing interest. Each time the agents' mechanism is initiated, the user's manually-constructed agents are executed along with the system's automatically-constructed agent, and the selected programs are presented to the user.

The user can create "virtual conferences" that consist of user-defined aggregations of programs. To create a virtual conference, a user composes and executes a query that selects a set of programs that share a common attribute, such as author, or discuss a common subject. This thematic aggregation of programs can be named, saved, and distributed to other users interested in the same theme.

The user can construct "synthetic programs" by sequencing together segments of programs from multiple different programs. To create a synthetic program, the user composes and executes a query, specifying that the invention should select only those portions of the programs that match the query. The user can then view the concatenated portions of multiple programs in a continuous manner. The synthetic program can be named, saved, and distributed to other users interesting in the synthetic program content.

Referring now to FIG. 11, which will be used to describe the operation of an embodiment of the present invention across a non-isochronous network connection. This embodiment incorporates a cooperative processing data distribution and caching model that enables the isochronous data streams to play  
5 continuously immediately following a navigational event, such as moving to the next slide or searching to a particular word in the transcript.

After the process starts **1000**, when the user first selects a program to play **1001**, the system downloads the selected portions of the non-isochronous data from  
10 the server to the client. The downloaded non-isochronous data includes the Slide Index, the Slides, the Transcript Index, the Transcript, and the Hypertext Index. The downloaded non-isochronous data is stored in a disk cache **1010** on the client. The purpose of pre-downloading this non-isochronous data is to avoid having to transmit it over the network connection simultaneously with the transmission of the  
15 isochronous data, thereby interrupting the transmission of the isochronous data. The Hypertext Objects are not pre-downloaded to the client; rather, the system is designed to pause the transmission of the isochronous data to accommodate the downloading of any Hypertext Objects. At the end of playing a program, the client disk cache is emptied in preparation for use with another program.

20

In addition to downloading portions of the non-isochronous data, the system downloads a segment of the isochronous data from the server to a memory cache on the client. The downloaded isochronous data includes the initial segment of the video data and the corresponding initial segment of the audio data. The amount of  
25 isochronous data downloaded typically ranges from 5 to 60 seconds, but may be more or less. The downloaded isochronous data is stored in a memory cache **1020** on the client.

When the user presses the Play Button, the Event Handler **1030** receives a  
30 Play Program Event **1040**. The system begins the continuous delivery of the

isochronous data to the display devices **1041**. Based on the time code of the currently displayed video frame, it also displays the associated non-isochronous data **1042**, including the Transcript, the Slides, and the Hypertext Links. As the system streams the isochronous data to the display devices, it depletes the memory cache. When the amount of isochronous data in the memory cache falls below a specified threshold, the system causes the client CPU to send a request to the server CPU for the next contiguous segment of isochronous data **1043**. This threshold typically works out to be on the order of 5-10 seconds, with a worst-case scenario of 60 seconds. It should be appreciated by one of ordinary skill in art that factors such as network capacity and usage should affect the choice of threshold. Upon receiving this data, the client CPU repopulates the isochronous data memory cache. If, as anticipated, the client CPU experiences a delay in receiving the requested data, caused by the non-isochronous network connection, the client CPU continues to deliver isochronous data remaining in its memory cache in a continuous stream to the display device, until that cache is exhausted.

The method for repopulating the client's memory cache is a critical element in supporting efficient random access into isochronous data streams over a non-isochronous network. The method for downloading the isochronous data from the server to the memory cache on the client is designed to balance two competing requirements. The first requirement is for continuous, uninterrupted delivery of the isochronous data to the video display device and speakers attached to the client CPU. The network connection between the client and server is typically non-isochronous, and may introduce significant delays in the transmission of data from the client to the server. In practice, if the memory cache on the client becomes empty, requiring client to send a request across the network to the server for additional isochronous data, the amount of time needed to send and receive the request will cause the interruption of play of the isochronous data. The requirement for continuous delivery thus encourages the caching of as much data as possible on the client. The second requirement is to minimize the amount of data

that is transmitted across the network. In practice, multiple users share a fixed amount of network bandwidth, and transmitting video and audio data across a network consumes a substantial portion of this limited resource. It is anticipated that a common user behavior will be to use the random access navigation capabilities to reposition the program. But the act of repositioning the program invalidates all or part of the data stored in the memory cache in the client. The larger the amount of data that is stored in the memory cache on the client, the more data is wasted upon repositioning the program, and thus the more network bandwidth was wasted in sending this unused data from the server to the client. Thus the requirement for minimizing the amount of data transmitted across the network encourages the caching of as little data as possible on the client.

The present invention balances the need for continuous delivery of isochronous data to the display devices with the need to avoid wasting network bandwidth by implementing a novel cooperative processing data distribution and caching model. The memory cache on the client is designed specifically for compressed isochronous data, and more specifically for compressed digital video data. The caching strategy differs markedly from traditional caching strategies. Traditional caching strategies measure the number of bytes of data in the cache, and repopulate the cache when the number of bytes falls below a specified threshold. By contrast, one embodiment of the present invention measures the number of *seconds* of isochronous data in the memory cache, and repopulates the cache when the number of seconds falls below a specified threshold. Due to the inherent inhomogeneities in video compression, a fixed number of seconds of compressed video data does not correspond to a fixed number of bytes of data. For video data streams that compress into a smaller than average number of bytes per second, the cooperative distribution and caching model reduces the amount of data sent across the network compared to a traditional caching scheme. For video data streams that compress into a larger than average number of bytes per second, the cooperative distribution and caching model guarantees a certain number of seconds



of video data cached on the server, reducing the likelihood of interrupted play of the video data stream compared to a traditional caching scheme.

In addition to designing the memory cache to contain a range of a number  
5 of seconds of isochronous data, the memory cache employs a policy of *unbalanced  
look ahead and look behind*. *Look ahead* refers to caching the isochronous data  
corresponding to "N" seconds into the future. This isochronous data will be  
delivered to the display device under the normal operation of playing the program.  
*Look behind* refers to caching the isochronous data corresponding to "M" seconds  
10 into the past. This isochronous data will be delivered to the display device under  
the frequent operation of replaying the previously played few seconds of the  
program. *Unbalanced* refers to the policy of caching a different amount (that is, a  
different number of seconds) of look ahead and look behind data. Generally, more  
look ahead data is cached than look behind data, typically in the approximate ratio  
15 of 7:1. It can be appreciated by one of ordinary skill in the art that different  
caching policies can be employed in anticipation of different common user  
behaviors. For example, the use of a circular data structure, a structure well-known  
in the art, may effect this operation.

20 During program play **1040**, the server sends data to the client at the nominal  
rate of one second of isochronous data each second. The server adapts to the  
characteristics of the network, bursting data if the network supports a high burst  
rate, or steadily transmitting data if the network does not support a high burst rate.  
The client monitors its memory cache, and sends requests to the server to speed up  
25 or slow down. The client also sends requests to the server to stop, restart at a new  
place in the program, or start playing a different program.

The system administrator can specify how much network bandwidth is  
available to the system, for each individual program, and collectively across all  
30 programs. The system automatically tunes its memory caching scheme to reflect

these limits. If the transmitted data would exceed the specified limits, the system automatically drops video frames as necessary.

5 When the user performs a navigational activity, such as moving to the next slide or searching to a particular word in the transcript, the Event Handler **1030** receives a Navigational Event **1050**. The system computes the time base value of the new position **1051**. It then downloads a new segment of the isochronous data from the server to the memory cache on the client **1052**. The downloaded isochronous data includes a segment of the video data and a corresponding segment  
10 of the audio data. The system then displays the video frame corresponding to the current time base value, and the non-isochronous data corresponding to the displayed video frame **1053**.

15 When the user selects a hypertext link, the Event Handler **1030** receives a Display Hypertext Object Event **1060**. The system pauses the play of the program **1061**. The client CPU requests that the server CPU send the Hypertext Object across the network connection **1062**, and upon receiving the Hypertext Object, causes it to be displayed **1063**.

20 Referring back to FIG. 1, the server **130** records the actions of each user, including not only which programs each user viewed, but also which portions of the programs each user viewed. This record can be used for usage analysis, billing, or report generation. The user can ask the server **130** for a usage summary, which contains an historical record of that particular user's usage. A manager or system  
25 administrator can ask the server **130** for a summary across some or all users, thereby developing an understanding of the patterns of usage. One might use any of the data mining tools as is known in the art for assisting in this purpose.

30 The usage record may serve as a guide to restructure old programs or to structure new ones, having learned what works from a presentation perspective and

what does not, for example. The usage record furthermore enables the system to notify users of changing data. The list of users who have viewed a program can be determined from the usage records. If a program is updated, the system reviews the usage record to determine which users have viewed the program, and notifies  
5 them that the program that they previously viewed has changed.

While the present invention has been described in terms of a few embodiments, the disclosure of the particular embodiment disclosed herein is for the purposes of teaching the present invention and should not be construed to limit  
10 the scope of the present invention which is solely defined by the scope and spirit of the appended claims.

EQUIS 97/06982  
IPEAUS 24 APR 1998

Having thus described our invention, what we claim as new and desire to secure by Letters Patent is as follows:

- 1           1.       A method of manipulating a plurality of streams of isochronous and non-  
2 isochronous digital data comprising the steps of:  
3               synchronizing the plurality of streams of isochronous and non-isochronous  
4 data by reference to a common time base;  
5               navigating to a position in a first stream of the plurality of streams using at  
6 least one of a sequential access approach and a random access approach available  
7 for and adapted to the first stream;  
8               identifying positions for each of the plurality of streams corresponding to  
9 the position in the first stream; and  
10              simultaneously displaying at least some of the plurality of streams at the  
11 positions corresponding to the position in the first stream.
  
- 1           2.       The method of claim 1, further comprising the step of delivering the  
2 plurality of streams of synchronized isochronous and non-isochronous data from a  
3 server to a client over a non-isochronous network.
  
- 1           3.       The method of claim 1, further comprising the step of caching isochronous  
2 data on the client, and modulating the delivery of the isochronous data over a  
3 network on which the client resides in a manner that maintains a predetermined  
4 range of time's worth of data cached on the client.
  
- 1           4.       The method of claim 1, further comprising the step of translating a  
2 transcript stream of the plurality of streams into one or more foreign languages, the  
3 transcript stream including a plurality of transcripts, each transcript synchronized to  
4 a common time base and each transcript independently navigable.
  
- 1           5.       A system for interacting with a computerized presentation comprising:

ATTORNEY DOCKET NO.: 44025.4011

PCTUS 97/06982  
IPEAUS 24 APR 1998

2 a plurality of isochronous and non-isochronous data streams, wherein each  
3 of the plurality of data streams are synchronized together by reference to a  
4 common time base;

5 for each of the plurality of data streams, means for at least one of sequential  
6 access navigation and random access navigation of such data stream, and means for  
7 display of such data stream; and

8 identification means, coupled to each of the means for at least one of  
9 sequential access navigation and random access navigation, wherein, given a  
10 position in one of the plurality of data streams as pointed to by one of the means  
11 for at least one of sequential access navigation and random access navigation, the  
12 identification means provides, via the common time base, the corresponding  
13 positions for each of the plurality of data streams.

1 6. The system of claim 5 further comprising:

2 a server for storing the plurality of isochronous and non-isochronous data  
3 streams;

4 a client for containing the means for display and the means for at least one  
5 of sequential access navigation and random access navigation of such data streams;  
6 and

7 a non-isochronous network for delivery of such data streams from the server  
8 to the client device;

9 the client further including a data cache and a modulation means both  
10 coupled to the network, wherein one or more of the data streams delivered by the  
11 network are stored in the data cache, and further wherein the modulation means  
12 maintains a predetermined range of time's worth of data within the data cache.

1 7. The system of claim 5, wherein one or more of the plurality of isochronous  
2 and non-isochronous data streams corresponds to a speaker giving an informational  
3 or educational presentation.

PCT/US 97/06982  
IP/EA/US 24 APR 1998

1 8. The system of claim 5, wherein at least one of the isochronous data streams  
2 includes digital video.

1 9. The system of claim 5, wherein at least one of the isochronous data streams  
2 includes digital audio.

1 10. The system of claim 5, wherein at least one of the non-isochronous data  
2 streams includes slides.

1 11. The system of claim 5, wherein at least one of the non-isochronous data  
2 streams includes hypertext links to related data objects.

1 12. The system of claim 5, wherein at least one of the non-isochronous data  
2 streams includes an outline of the computerized presentation.

1 13. The system of claim 5, wherein at least one of the non-isochronous data  
2 streams includes a transcript of spoken words in the computerized presentation.

1 14. The system of claim 13, wherein the means for at least one of sequential  
2 access navigation and random access navigation further includes a string matching  
algorithm for use upon the transcript of spoken words.

1 15. The system of claim 14, further comprising:  
2 a plurality of computerized presentations which are capable of being  
3 selected by a user, at least some of the presentations including one or more  
4 keywords associated therewith; and  
5 a profiling means which maintains a user profile on each user, the user  
6 profile including an aggregation of at least some of the keywords of the  
7 presentations selected by the user.

CONFIDENTIAL

ATTORNEY DOCKET NO.: 44025.4011

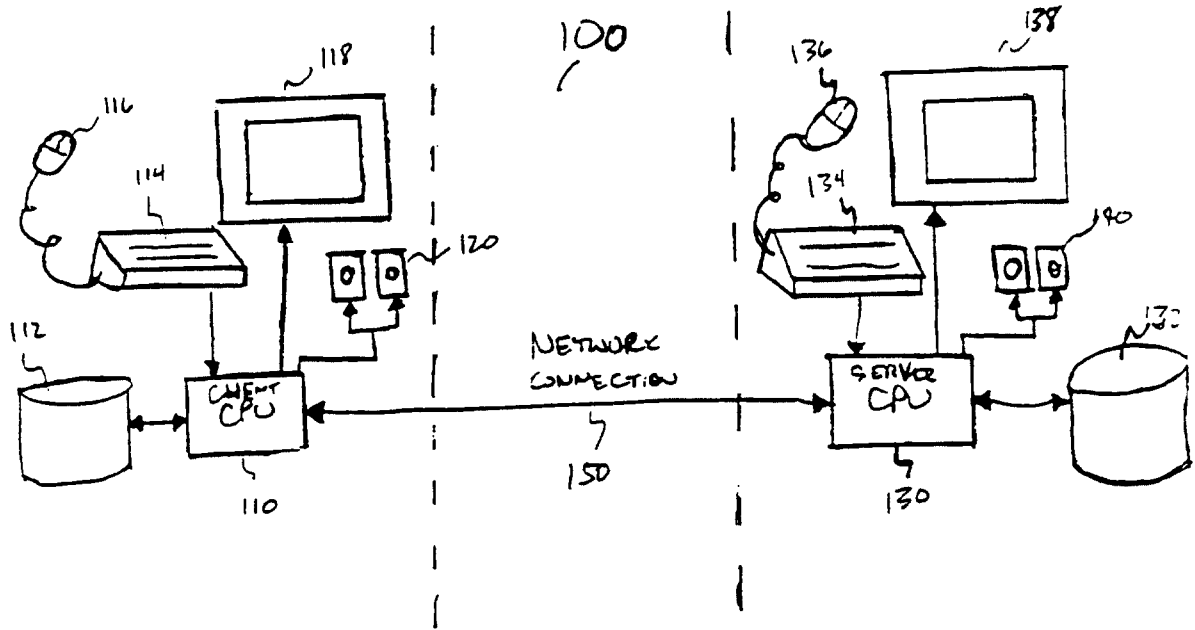
**FOCUS 97/06982**  
**IPEA/US 24 APR 1998**

1           16.    A system for interacting with a computerized presentation comprising:  
2                    a plurality of isochronous and non-isochronous data streams;  
3                    two or more sets of conceptual events, each set indexed into an indexed  
4           stream of the plurality of isochronous and non-isochronous data streams;  
5                    for each of the plurality of isochronous and non-isochronous data streams, a  
6           means for navigation and a means for display of such data stream, and for the  
7           indexed streams for each set of conceptual events, the means for navigation  
8           including a means for selection of a conceptual event; and  
9                    an identification means, coupled to each navigation means, wherein, given a  
10          selected conceptual event, the identification means provides a position in each of  
11          the plurality of isochronous and non-isochronous data streams corresponding to the  
12          conceptual event.

1           17.    The system of claim 16 wherein a first set of conceptual events is indexed  
2           into an isochronous data stream and a second set of conceptual events is indexed  
3           into a non-isochronous data stream.

1           18.    The system of claim 16 further comprising a bookmarking means for  
2           user-defined creation of conceptual events.

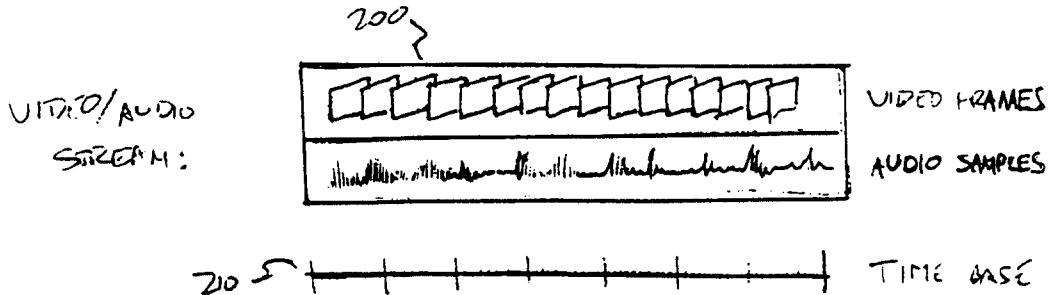
FIG. 1





2/11

FIG. 2



TRANSCRIPT:

220

Hello, I'm Robert Jones, and I'm going to tell you about the history of manned flight. This talk will take about 20 minutes. The first manned flight was the culmination of centuries of research and development.

TIME-CODED TRANSCRIPT:

225

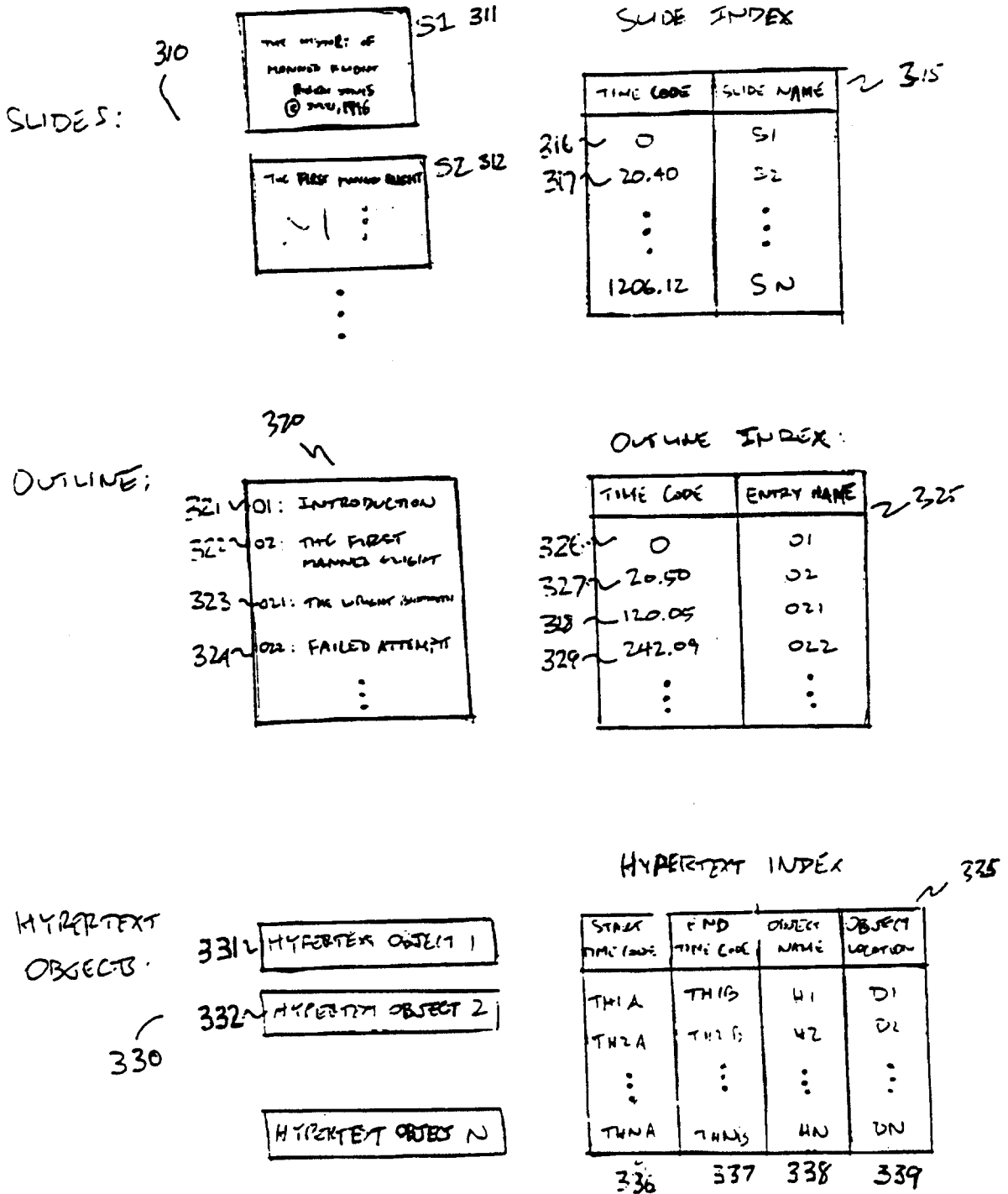
230

235

<0> Hello <0.2> I'm <0.4> Robert <0.7> Jones <1.0> and I'm <1.4> I'm  
 <2.2> going <2.9> to <3.5> tell <4.0> you <4.5> about <4.6> the  
 <5.0> history <5.2> of <5.7> manned <7.4> flight <11.0> This <11.5> talk  
 <12.1> will <12.6> take <13.3> about <14.1> 20 <15.1> minutes <20.5> to  
 <20.8> first <21.2> manned <21.6> flight <22.0> was <22.5> the  
 <23.3> culmination <23.9> of <24.5> centuries <25.0> of  
 <25.4> research <26.2> and <27.0> development.

3/11

FIG. 3



4/11

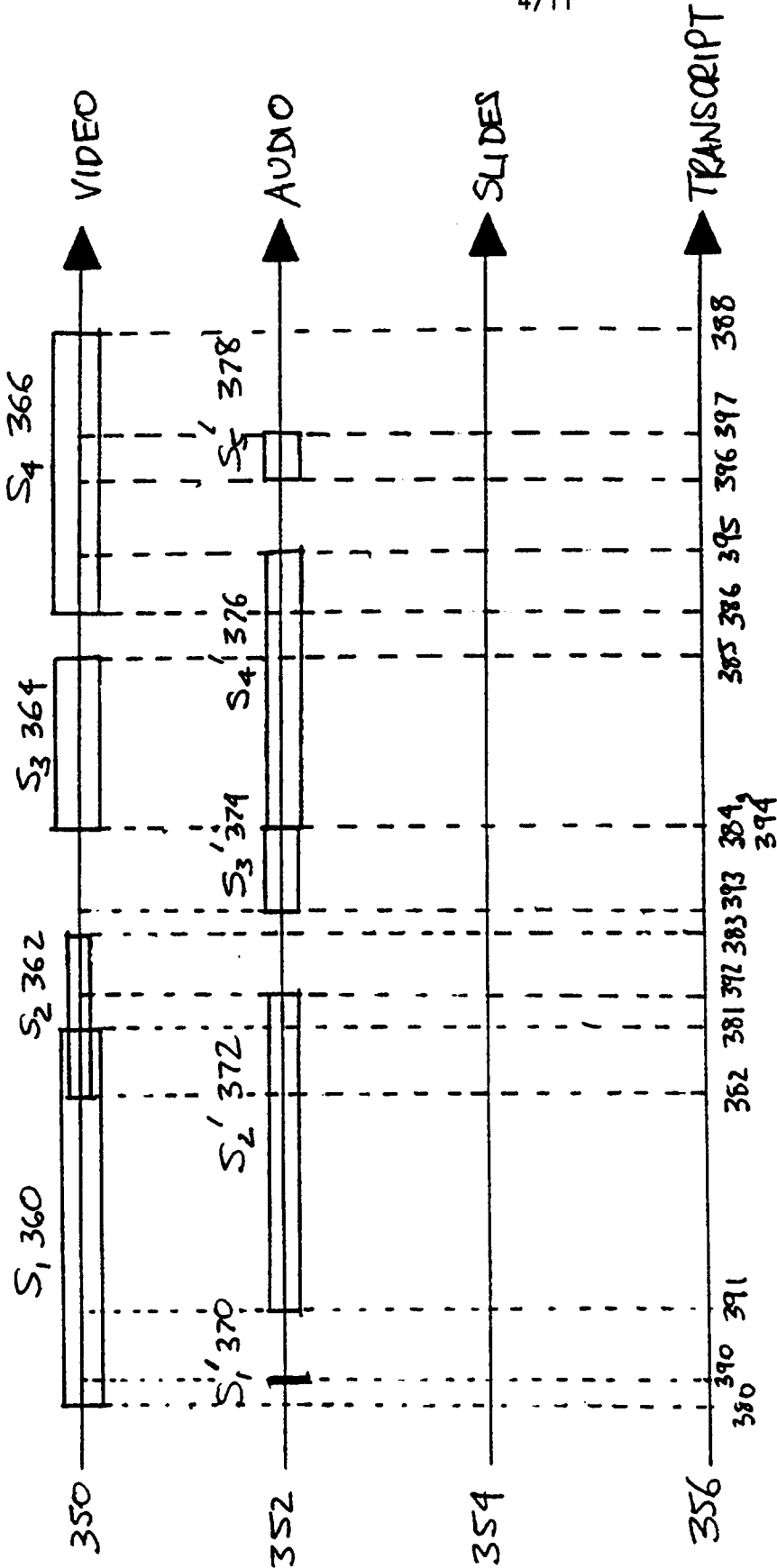
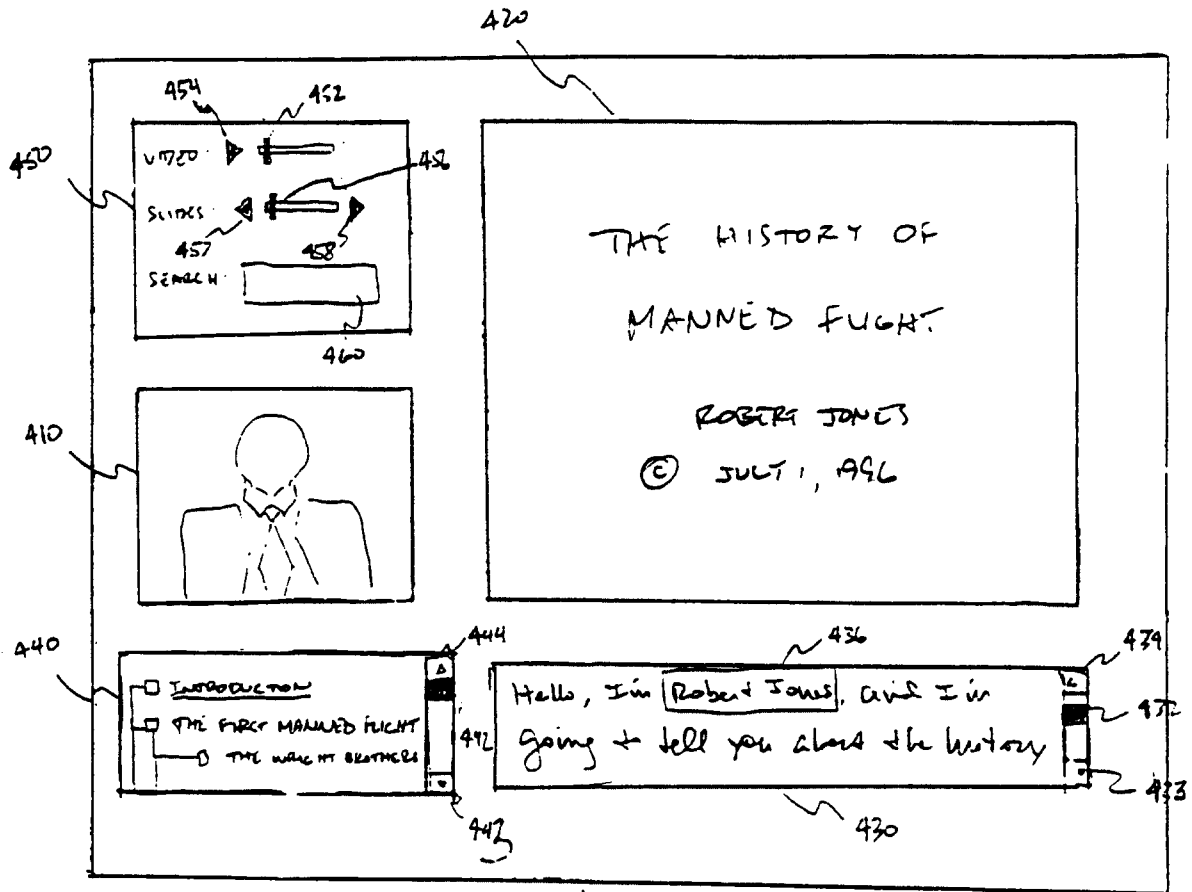


FIG. 4

5/11

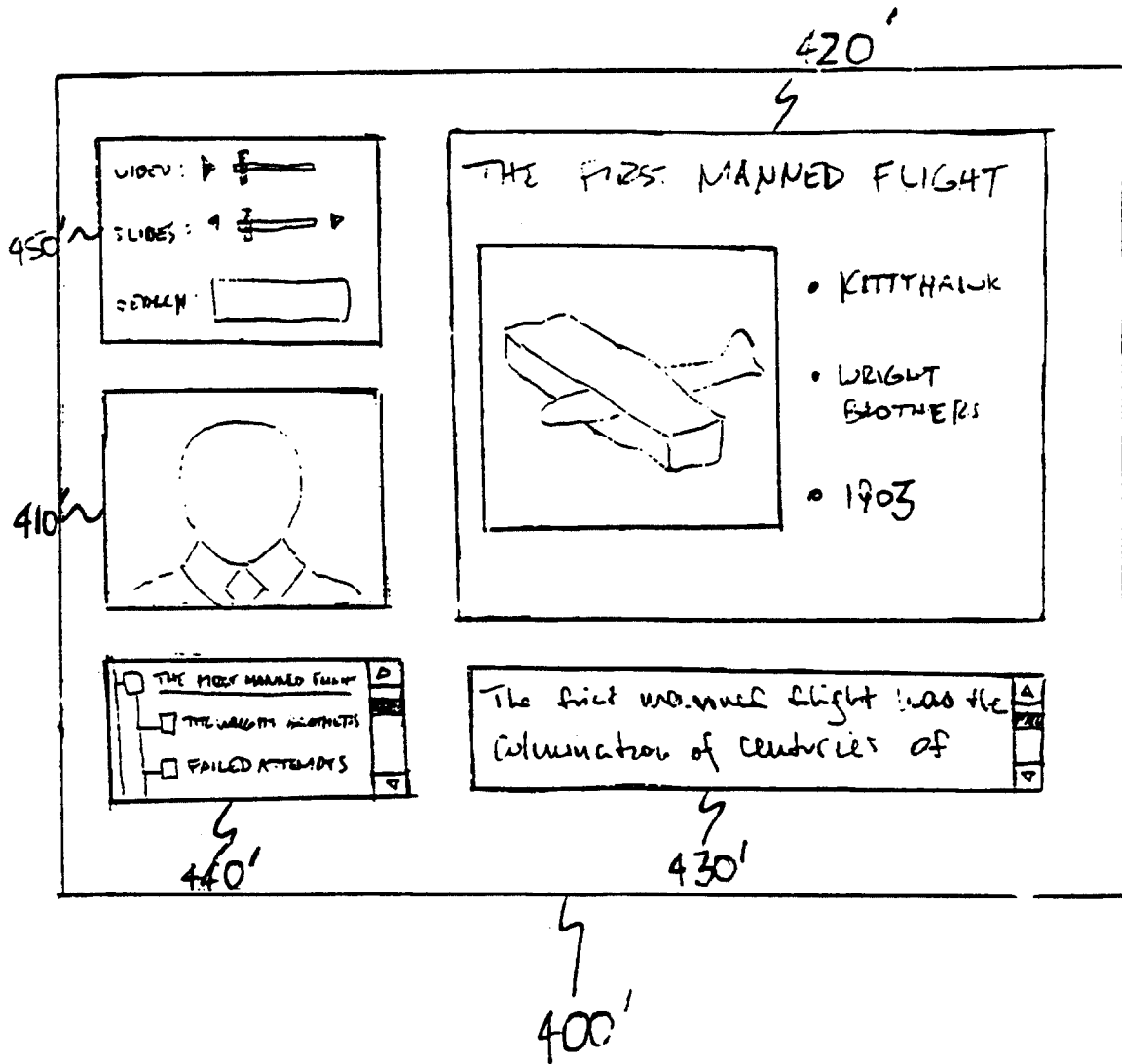
FIG. 5



400

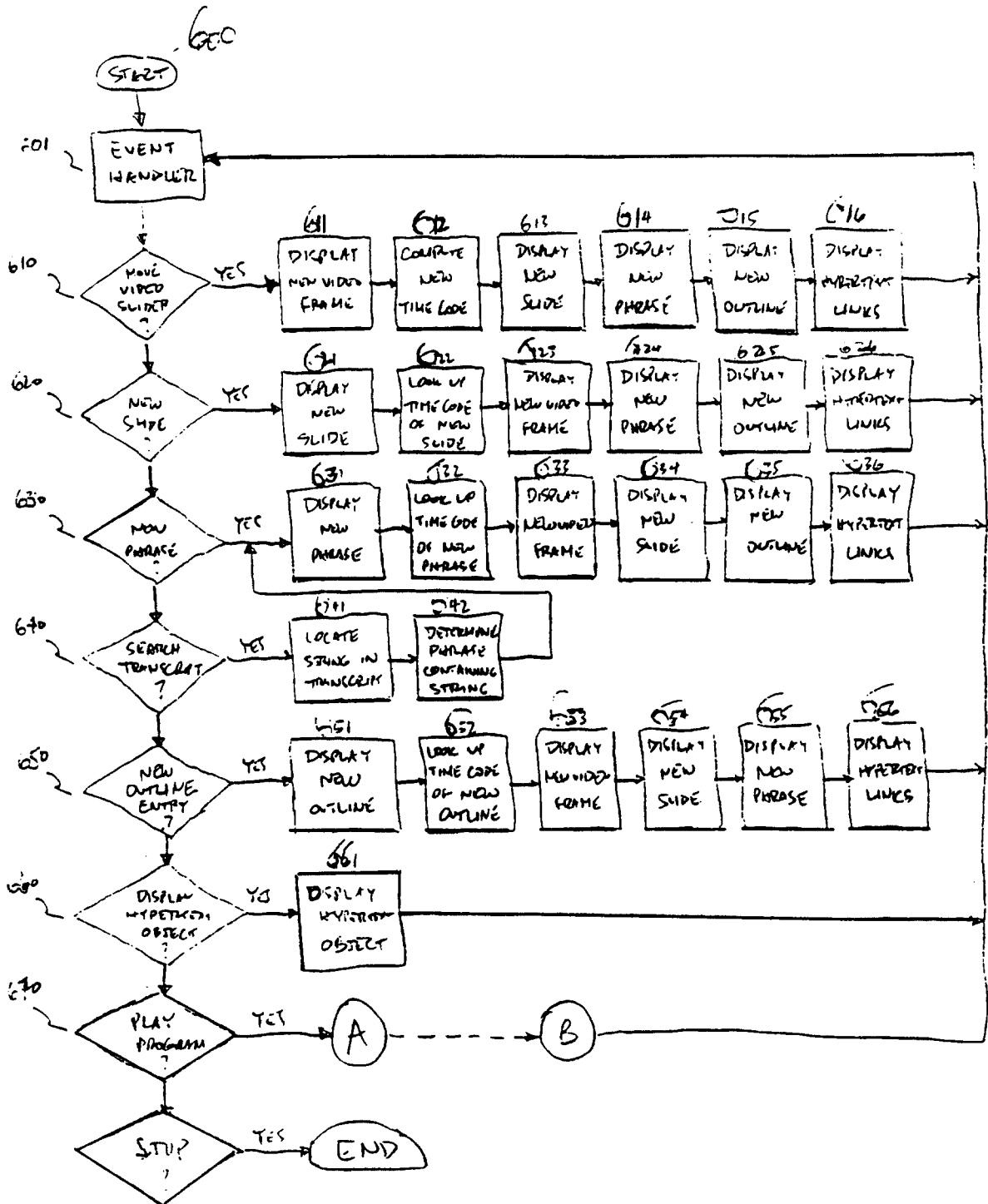
6/11

FIG. 6



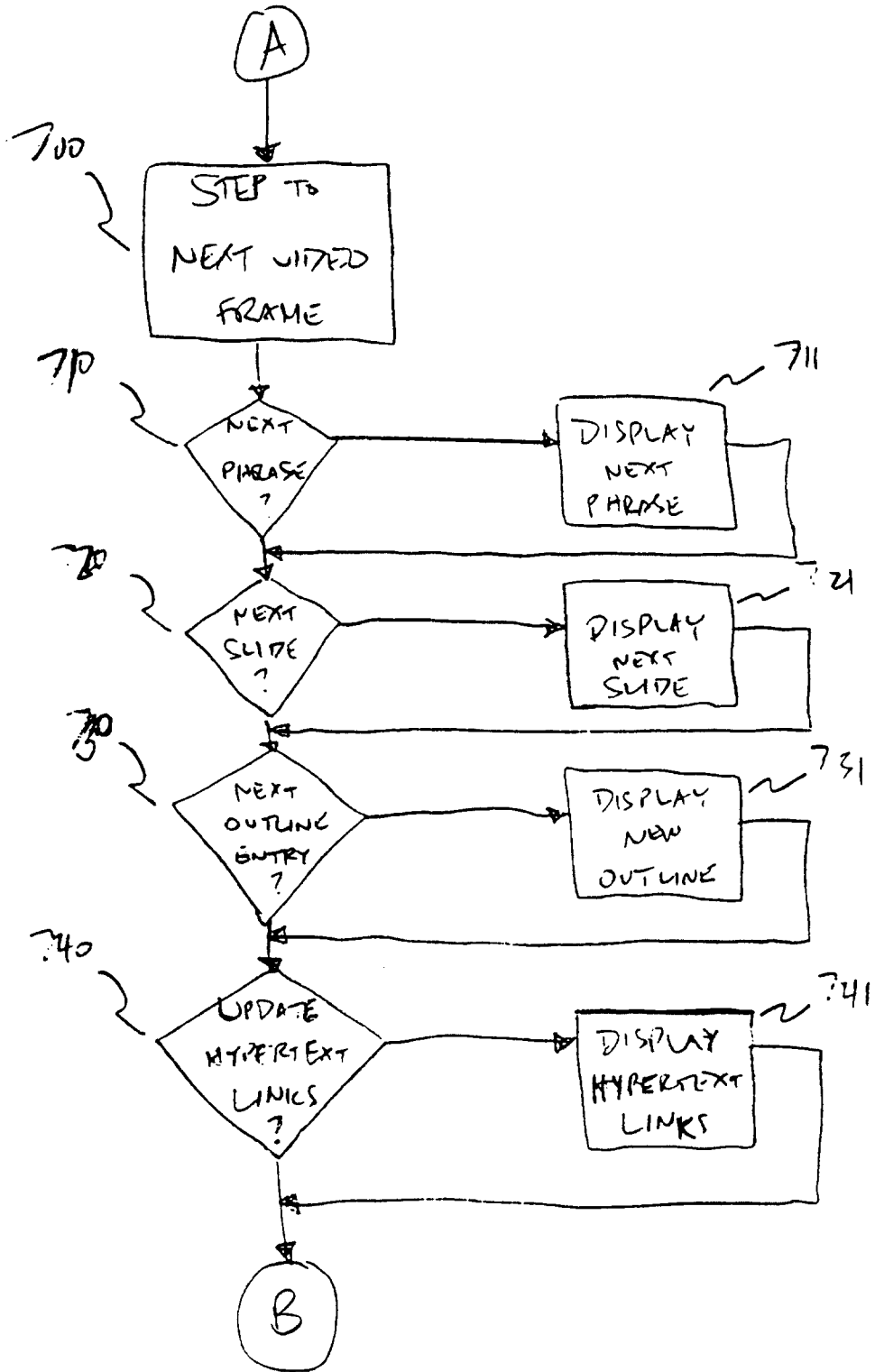
7/11

FIG 7.



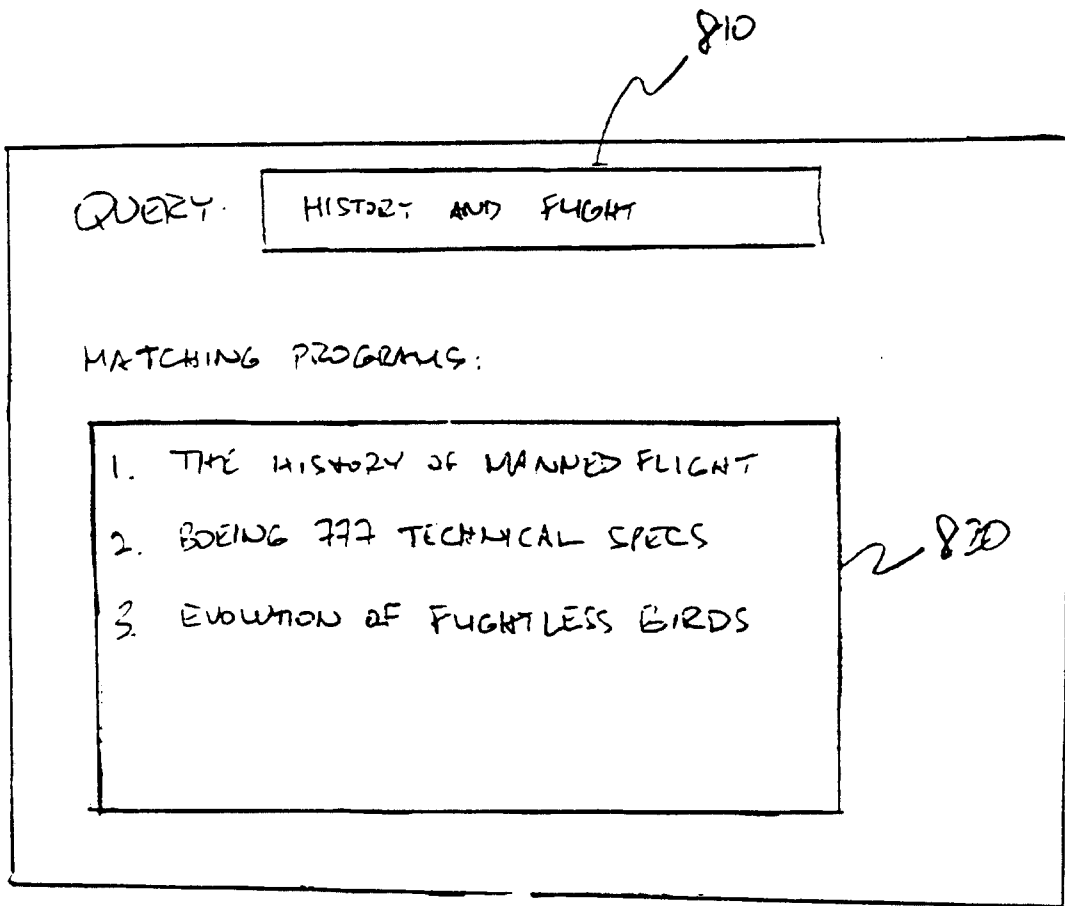
8/11

FIG. 8



9/11

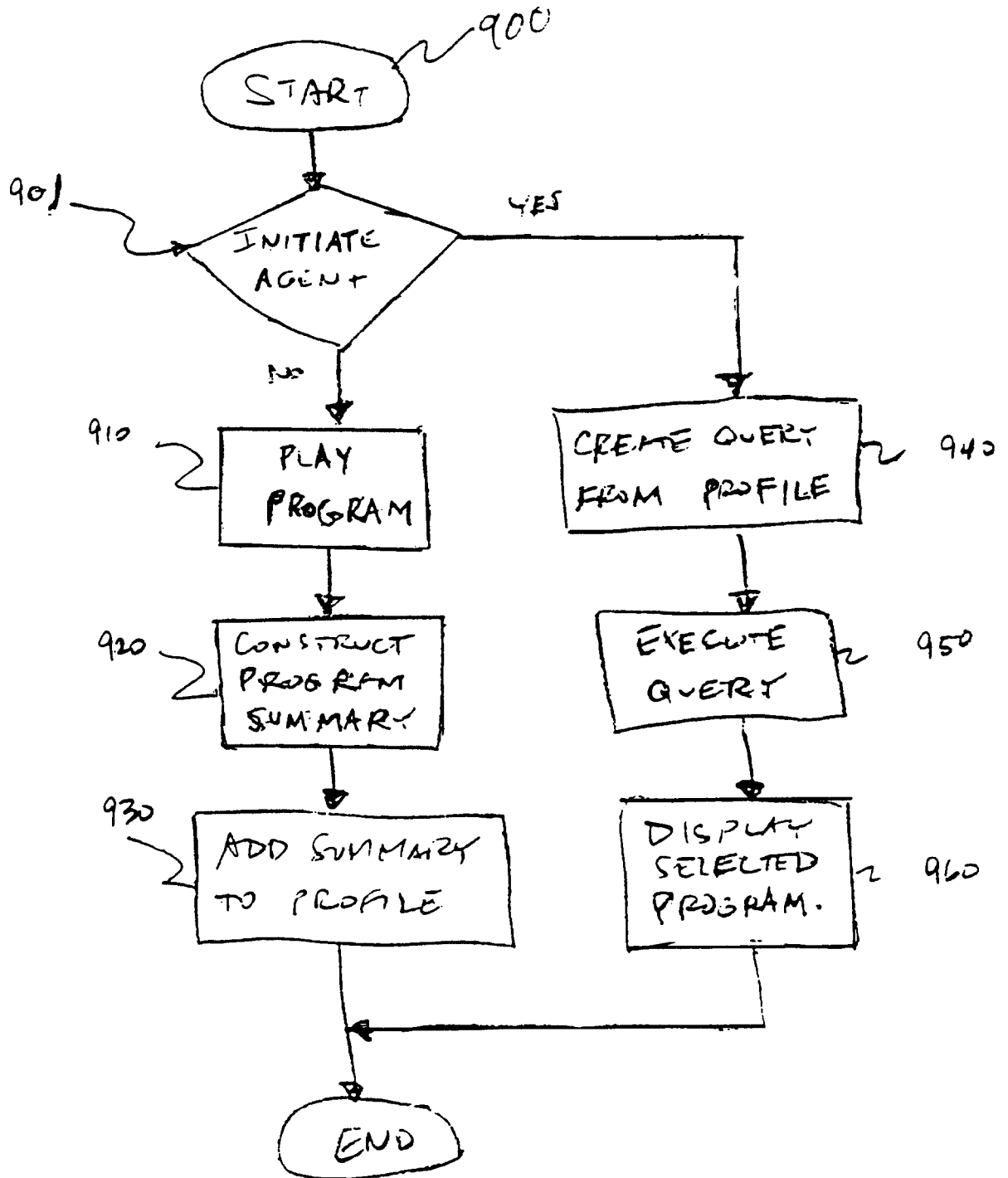
FIG. 9





10/11

FIG. 10



11/11

FIG. 11

