

UNITED STATES PATENT AND TRADEMARK OFFICE

BEFORE THE PATENT TRIAL AND APPEAL BOARD

ZOOM COMMUNICATIONS, INC.
Petitioner

v.

RICOH COMPANY, LTD.
Patent Owner

IPR2026-00461

**PETITION FOR INTER PARTES REVIEW
OF U.S. PATENT NO. 11,289,093**

TABLE OF CONTENTS

	Page
I. INTRODUCTION	1
II. TECHNOLOGY BACKGROUND.....	2
A. Converting Speech to Text and Associating It with Time Information Was Well-Known.....	2
B. A Server Managing Content Data Generated During an Event Was Well-Known.....	5
C. Displaying Time-Ordered Text Alongside a Graphical Playback Control Region Was Well-Known	7
D. Navigating Playback by Selecting Displayed Text Was Well- Known	12
E. Repositioning the Graphical Playback Control in Response to Text Selection Was Known.....	13
III. '093 PATENT.....	15
A. Overview	15
B. Prosecution History	17
IV. POSITA	19
V. CLAIM CONSTRUCTION	20
VI. OVERVIEW OF CHALLENGE.....	20
VII. OVERVIEW OF THE PRIOR ART	21
A. Mahajan (EX1004).....	21
B. Goldberg (EX1005).....	23
C. Wei (EX1006)	26
VIII. THE ASSERTED GROUNDS OF INVALIDITY	29
A. Ground 1: Mahajan in View of Goldberg Renders Obvious Claims 1-22	29
1. Motivation to Combine	29
2. Preambles	32
a) [1.pre].....	32

TABLE OF CONTENTS

(continued)

	Page
b) [10.pre]/[10.a]	33
c) [17.pre].....	37
d) [22.pre].....	38
3. [1.a]/[10.a]/[17.a]/[22.a]	40
4. [1.b]/[10.b]/[17.b]/[22.b].....	42
5. [1.c]/[10.c]/[17.c]/[22.c]	46
6. [1.d]/[17.d]/[22.d]	48
7. [1.e]/[10.d]/[17.e]/[22.e]	48
8. [1.f]/[10.e]/[17.f]/[22.f].....	49
9. Claim 2/18.....	51
10. Claim 3.....	52
11. Claim 4.....	54
12. [5.a]/[20.a].....	54
13. [5.b]/[20.b]	57
14. [6.a]/[14.a]/[21.a].....	59
15. [6.b] /[14.b]/[21.b]	60
16. [6.c]/[14.c]/[21.c]	61
17. Claim 7.....	62
18. [8.a]	64
19. [8.b]	66
20. [8.c]	67
21. Claim 9.....	69
22. [11.a]	70
23. [11.b]	72
24. Claim 12.....	74
25. Claim 13/19.....	75
26. Claim 15.....	76

TABLE OF CONTENTS
(continued)

	Page
27. Claim 16	79
B. Ground 2: Mahajan in View of Goldberg and Further in view of Wei Renders Obvious Claims 1-22	80
1. Motivation to Combine	80
2. [1.e]/[10.d]/[17.e]/[22.e]	83
3. [1.f]/[10.e]/[17.f]/[22.f]	85
4. [5.a]/[20.a]	86
5. [5.b]/[20.b]	89
6. Claim 7	90
7. Claim 12	92
IX. CONCLUSION	94
X. MANDATORY NOTICES PURSUANT TO 37 C.F.R.	94
A. Real Parties-In-Interest under 37 C.F.R.	94
B. Related Matters Under 37 C.F.R.	95
C. Lead and Back-up Counsel under 37 C.F.R.	95
D. Service Information under 37 C.F.R.	95
XI. GROUNDS FOR STANDING	95
XII. PAYMENT OF FEES	96
XIII. STATEMENT OF PRECISE RELIEF REQUESTED	96

TABLE OF AUTHORITIES

Page(s)

Cases

<i>KSR Int'l Co. v. Teleflex Inc.</i> , 550 U.S. 398 (2007).....	31, 82
---	--------

Statutes

35 U.S.C. §102.....	18
35 U.S.C. §102(a)(1).....	20, 21
35 U.S.C. §103.....	20, 96

Other Authorities

37 C.F.R. §§42.6	98
37 C.F.R. §42.8	94
37 C.F.R. §42.8(b)(1).....	94
37 C.F.R. §42.8(b)(2).....	95
37 C.F.R. §42.8(b)(3).....	95
37 C.F.R. §42.8(b)(4).....	95
37 C.F.R. § 42.11	100
37 C.F.R. §§42.15(a).....	96
37 C.F.R. § 42.24(a).....	102
37 C.F.R. §§ 42.24(a)(1).....	102
37 C.F.R. § 42.24(d)	102
37 C.F.R. §§42.103	96
37 C.F.R. §42.104(a).....	95

TABLE OF AUTHORITIES
(continued)

	Page(s)
37 C.F.R. §§42.105	98

Patent No. 11,289,093

Petition for *Inter Partes* Review

PETITIONER's EXHIBIT LIST

No.	Reference	Referred To As
1001	U.S. Patent No. 11,289,093	'093 Patent
1002	Prosecution History of U.S. Patent No. 11,289,093	'093 Patent File History
1003	Declaration of Dr. Kevin Almeroth	Almeroth Decl.
1004	U.S. Patent Application Publication No. 2008/0140385	Mahajan
1005	U.S. Patent Application Publication No. 2008/0319744	Goldberg
1006	U.S. Patent Application Publication No. 2011/0145428	Wei
1007	Japanese Patent Application Publication No. 2012-108631	Sakabe
1008	U.S. Patent Application Publication No. 2012/0245936	Treglia
1009	U.S. Patent Application Publication No. 2017/0083214	Furesjö
1010	U.S. Patent Application Publication No. 2007/0074116	Thomas
1011	A. Knight and K. Almeroth, " Fast Caption Alignment for Automatic Indexing of Audio ," <u>International Journal of Multimedia Data Engineering and Management</u> , vol. 1, num. 2, pp. 1-17, April-June 2010.	Fast Caption Alignment

LISTING OF CHALLENGED CLAIMS

1 [pre.]	A display control apparatus, comprising:
[1.a]	circuitry configured to: receive text data converted from voice data, and time information from a server that manages content data generated during an event;
[1.b]	control a display to: display the text data in an order according to the time information, and
[1.c]	display a graphical control region, at a first display location, that sets playback position in a total playback time of the voice data;
[1.d]	receive selection of first text data from the text data being displayed;
[1.e]	determine, by referring to the time information, a second display location which is different from the first display location, the second display location corresponding to the time information of the first text data; and
[1.f]	control the display to display the graphical control region at the second display location.
2.	The display control apparatus of claim 1, wherein the circuitry is further configured to start playback of the voice data from a time that the first text data is generated.
3.	The display control apparatus of claim 2, wherein the graphical control region further indicates a playback progress of the voice data.
4.	The display control apparatus of claim 3, wherein the graphical control region is a seek bar.
[5.a]	The display control apparatus of claim 1, wherein the circuitry is further configured to: receive, from the server that manages content data generated during the event, screenshot image data associated with the text data; and
[5.b]	control the display to further display the screenshot image data, in the order according to the time information.
[6.a]	The display control apparatus of claim 1, wherein the circuitry is further configured to: in response to selection of the first text data, obtain, from the first text data, a content processing identifier for identifying conversion processing that has been applied to generate the first text data from corresponding voice data;
[6.b]	obtain the time information using the content processing identifier; and
[6.c]	determine the second display location of the graphical control region based on the time information.

7.	The display control apparatus of claim 1, wherein the first text data corresponds to one sentence of speech.
[8.a]	A system, comprising: the display control apparatus of claim 1; and the server which is communicably connected with the display control apparatus,
[8.b]	the server including a memory that stores, for each event, voice data recorded during the event, text data converted from the voice data, and time information; and
[8.c]	another circuitry configured to transmit information on the event stored in the memory to the display control apparatus.
9.	The system of claim 8, further comprising: a communication terminal communicably connected with the server and configured to transmit voice data collected during an event to the server.
10 [pre.]	A system, comprising:
[10.a]	a memory that stores voice data recorded during the event, text data converted from the voice data, and time information for an event; and
[10.b]	first circuitry and second circuitry configured to cooperate together to: control a display to: display text data in an order according to the time information, and
[10.c]	display a graphical control region, at a first display location, that sets playback position in a total playback time of the voice data;
[10.d]	determine, by referring to the time information, a second display location which is different from the first display location, the second display location corresponding to the time information of the first text data; and
[10.e]	control the display to display the graphical control region at the second display location.
[11.a]	The system of claim 10, wherein the first circuitry is further configured to transmit screen data generated based on the information on the event stored in the memory to the second circuitry, and
[11.b]	the second circuitry displays a screen based on the screen data, the screen including the text data arranged in the order according to the time information, and the graphical control region that sets playback position in a total playback time of the voice data.
12.	The system of claim 11, wherein the second circuitry is further configured to: in response to selection of the first text data, obtain, from the first text data, a content processing identifier for identifying conversion

	processing that has been applied to generate the first text data from corresponding voice data; transmit the content processing identifier to the first circuitry to obtain the time information; and determine the second display location of the graphical control region based on the time information.
13.	The system of claim 10, wherein the first circuitry is further configured to transmit the information on the event to the second circuitry, and the second circuitry generates screen data based on the information on the event, and controls the display to display a screen based on the screen data, the screen including the text data arranged in the order according to the time information, and the graphical control region that sets playback position in a total playback time of the voice data.
[14.a]	The system of claim 13, wherein the second circuitry is further configured to: in response to selection of the first text data, obtain, from the first text data, a content processing identifier for identifying conversion processing that has been applied to generate the first text data from corresponding voice data;
[14.b]	obtain the time information using the content processing identifier; and
[14.c]	determine the second display location of the graphical control region based on the time information.
15.	The system of claim 10, further comprising: a server including the first circuitry; and a display control apparatus communicably connected with the server, wherein the display control apparatus includes the second circuitry.
16.	The system of claim 15, wherein the second circuitry of the display control apparatus is configured to: control the display to display the text data in the order according to the time information, and display the graphical control region, at the first display location, determine, by referring to the time information, the second display location; and control the display to display the graphical control region at the second display location.
17 [pre.]	A display control method, comprising:
[17.a]	receiving text data converted from voice data, and time information from a server that manages content data generated during an event;
[17.b]	displaying, on a display, the text data in an order according to the time information;

[17.c]	displaying, on the display at a <u>first display location</u> , a graphical control region that sets playback position in a total playback time of the voice data;
[17.d]	receiving selection of first text data from the text data being displayed;
[17.e]	determining, by referring to the time information, a second display location which is different from the first display location, the second display location corresponding to the time information of the first text data; and
[17.f]	controlling the display to display the graphical control region at the second display location.
18.	The display control method of claim 17, further comprising: starting playback of the voice data from a time that the first text data is generated.
19.	The display control method of claim 17, further comprising: generating screen data based on information on the event, the information on the event including the voice data, the text data converted from the voice data, and the time information, wherein the displaying the text data and the displaying the graphical control region is based on the screen data.
[20.a]	The display control method of claim 17, further comprising: receiving, from the server that manages content data generated during the event, screenshot image data associated with the text data; and
[20.b]	controlling the display to further display the screenshot image data, in the order according to the time information.
[21.a]	The display control method of claim 17, further comprising: in response to selection of the first text data, obtaining, from the first text data, a content processing identifier for identifying conversion processing that has been applied to generate the first text data from corresponding voice data;
[21.b]	obtaining the time information using the content processing identifier; and
[21.c]	determining the second display location of the graphical control region based on the time information.
22 [pre.]	A non-transitory recording medium storing a plurality of instructions which, when executed by one or more processors, cause the processors to perform a display control method, comprising:
[22.a]	receiving text data converted from voice data, and time information from a server that manages content data generated during an event;
[22.b]	displaying, on a display, the text data in an order according to the time information;

[22.c]	displaying, on the display at a first display region, a graphical control region that sets playback position in a total playback time of the voice data;
[22.d]	receiving selection of first text data from the text data being displayed;
[22.e]	determining, by referring to the time information, a second display location which is different from the first display location, the second display location corresponding to the time information of the first text data; and
[22.f]	controlling the display to display the graphical control region at the second display location.

I. INTRODUCTION

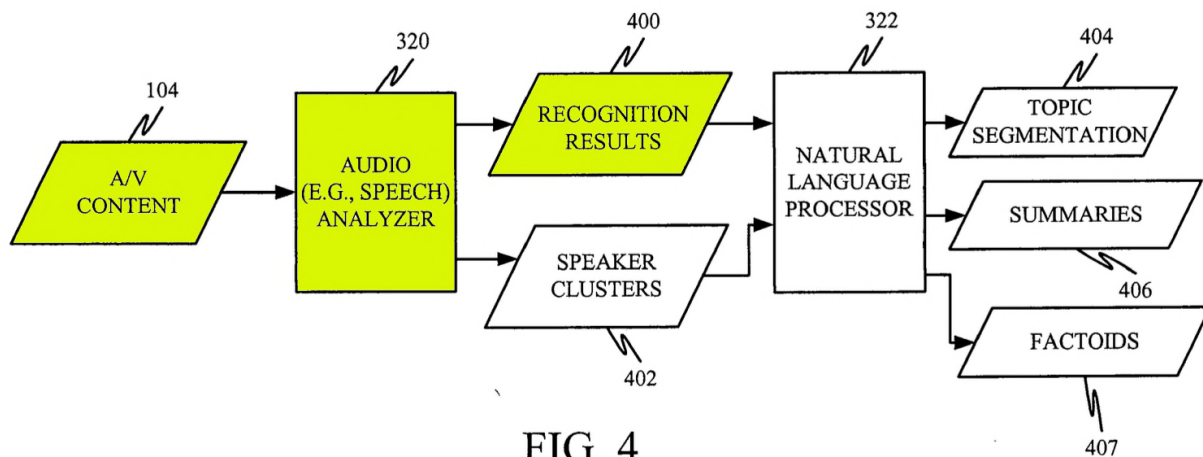
Petitioner requests *inter partes* review and cancellation of claims 1-22 of U.S. Patent 11,289,093 (“the ’093 Patent”) (EX1001). The ’093 Patent is directed to displaying selected text in a timestamped transcript. The challenged claims describe a simple, familiar interaction: a user sees a meeting recording with a searchable, time-stamped transcript, clicks a line of text, and the playback seek bar jumps to the corresponding point in the recording. This merely combine two everyday features: (1) a searchable transcript tied to time, like closed captions; and (2) a standard playback or seek bar, like the one used in nearly every video player that shows and controls playback position. The claimed step is simply: click the transcript and move the seek bar. Every piece of that interaction was already familiar long before the ’093 Patent: speech-to-text and time alignment, transcript-and-player displays, clickable transcript navigation, and movement of the playback indicator were all known, with the core arrangement disclosed more than a decade before November 29, 2018. Mahajan and Goldberg disclose the essential interaction, and Wei confirms the same transcript-synchronized media navigation. The ’093 Patent adds no technological leap; it applies ordinary media-player behavior to ordinary time-stamped text. Petitioner therefore respectfully requests institution of *inter partes* review and cancellation of all claims of the ’093 Patent.

II. TECHNOLOGY BACKGROUND

A. Converting Speech to Text and Associating It with Time Information Was Well-Known

Automatic conversion of recorded speech into text (speech-to-text technology) and aligning the text with multimedia content (like the kind used for closed captioning) was already a mature, well-documented technology before the earliest possible priority date of November 29, 2018. EX1003, ¶[0027].

Mahajan discloses an “automatic content analyzer configured to receive A/V content and analyze the A/V content using speech recognition and natural language processing to generate speech metadata and natural language metadata corresponding to the A/V content.” EX1004, claim 1; *see also* Abstract (“Audio/video (A/V) content is analyzed using speech and language analysis components. Metadata is automatically generated based upon the analysis.”). Mahajan’s audio analyzer 320 “recognizes words in the A/V content 104 and outputs an automatically derived transcript of those words as recognition results 400,” using, “for instance,... a large vocabulary automatic speech recognition system.” EX1004, ¶[0052]. Critically, the analyzer “also aligns the recognized words with the A/V content 104, such that each word in the transcript is aligned with a specific time point in the A/V content 104.” *Id.* EX1003, ¶[0028].



EX1004, FIG4.

Goldberg confirms that automated voice recognition was a recognized technology: “the Rapid Transcript process can also make use of stenography or voice recognition software to produce this initial rough draft.” EX1005, ¶[0069]. Goldberg further discloses that “[t]he rapid transcript includes time coding, speaker identification and summary.” EX1005, ¶[0013]. The association between transcript text and time information is achieved because “the transcript is ... further formatted to convert into a table wherein the time codes and the written words in the transcript are associated with each other in a one-to-one relationship.” EX1005, ¶[0099]; EX1003, ¶[0029].

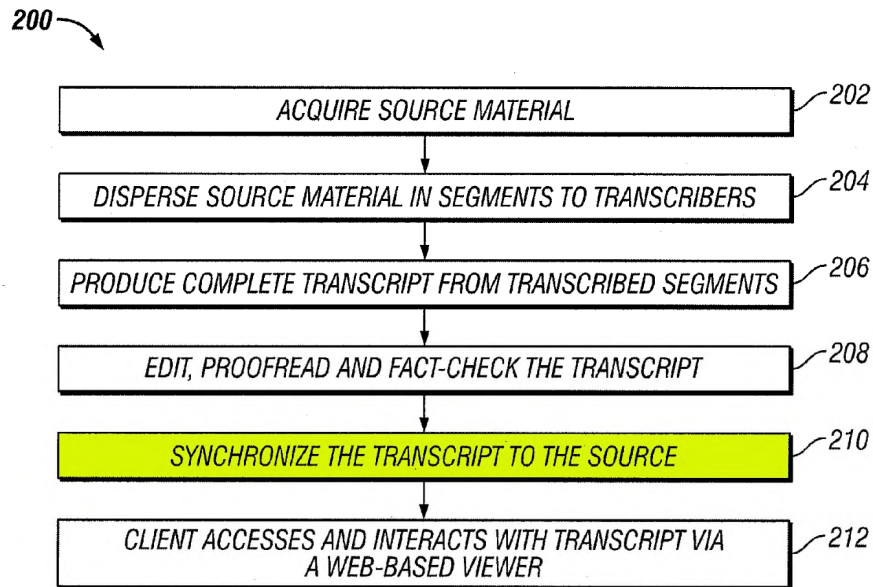


FIG. 2

EX1005, FIG2.

Sakabe discloses a minutes-creation apparatus that “records a speaker’s voice during a meeting, converts the spoken content from the recoded audio into textual information by using a speech recognition function to recognize the spoken content, and displays the textual information on a display device.” EX1007, ¶0002. Treglia likewise discloses a method comprising “transcribing the utterance into a transcribed text, [and] creating a second temporal marker comprising temporal information related to the transcribed text.” EX1008, Abstract. A POSITA would have understood converting recorded event speech to time-stamped text as a well-known technique long before 2018. EX1003, ¶¶ [0030]-[0031].

B. A Server Managing Content Data Generated During an Event Was Well-Known

A server is simply the known place where an application stores recordings and transcripts and sends them to a user's device; that arrangement was well-known long before the '093 Patent. EX1003, ¶[0032].

Goldberg's server receives "spoke source material" from a wide range of events, including, for example "financial conference calls," and "conventions and meetings." EX1005, ¶¶[0003], [0038]-[0053]. Goldberg's server 302 "receives source material 304 in the form of audio input from any source and encodes the source material into audio files." EX1005, ¶[0055]. Once processed, "[a] client begins by logging on to the provider's password protected website," which brings the client to a page where "[t]he web page created for the client consists of the transcript, with time coding, side by side with a column containing a summary." EX1005, ¶¶[0072]-[0073]; EX1003, ¶[0033].

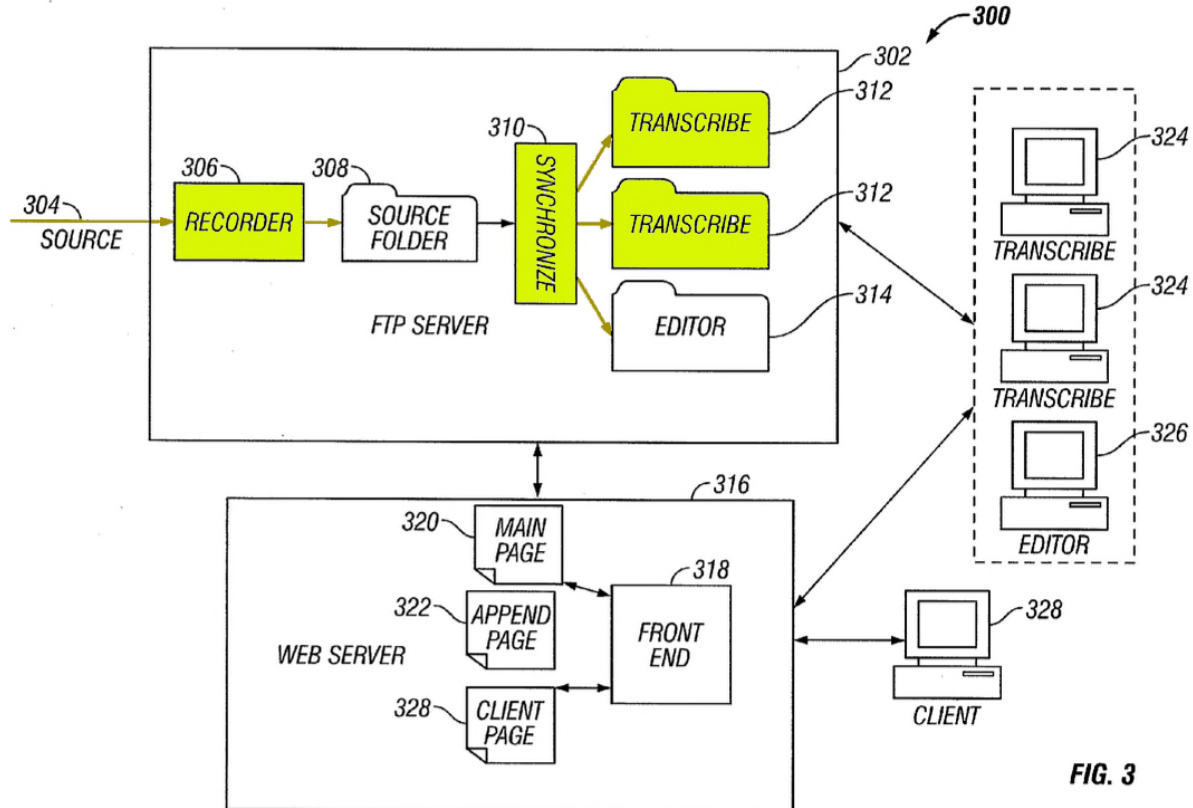


FIG. 3

EX1005, FIG3.

Furesjö confirms that server-mediated management of session data was likewise well-known in the art, disclosing a server “suitable for routing communication data between the user terminals of two or more users” that “can include user-captured audio and/or video data.” EX1009, ¶¶[0002], [0004]; EX1003, ¶[0034].

C. Displaying Time-Ordered Text Alongside a Graphical Playback Control Region Was Well-Known

Putting a time-ordered transcript next to a playback timeline—like closed captions beside a video—was well-known in prior art media and transcript viewers by the time of the '093 Patent. EX1003, ¶[0035].

Mahajan's user interface 102 "includes a screen area 200 which can be used to display the A/V content 104" and "a control bar 202 that is used to control the playing of A/V content 104 in screen area 200." EX1004, ¶[0036]. A "text field (or text box) 208 displays speech recognition results corresponding to the A/V content currently being played on screen 200," and "as the A/V content 104 is played on screen 200, the speech recognition results corresponding to the audio portion of A/V content 104 being played are displayed (e.g., scrolled) in text box 208." EX1004, ¶[0038]. Because the text is time-aligned to the content, it is displayed in chronological order. EX1004, ¶[0052]. Mahajan's interface further includes "a time bar 212" having "incremental time displays along the top of bar 212 that indicate a beginning and end, and intermediate points, within the A/V content 104," with "[a] marker 214 ... shown along time bar 212" that "indicates what part of the A/V content stream is currently being displayed on display 200." EX1004, ¶[0044]. The marker 214 can be "manipulated by the user," such that "the user can ... select marker 214 ... and move it along time bar 212 to any desired point," whereupon "A/V player

122 will skip to that time point in the A/V content stream 104.” EX1004, ¶[0045];

EX1003, ¶[0036].

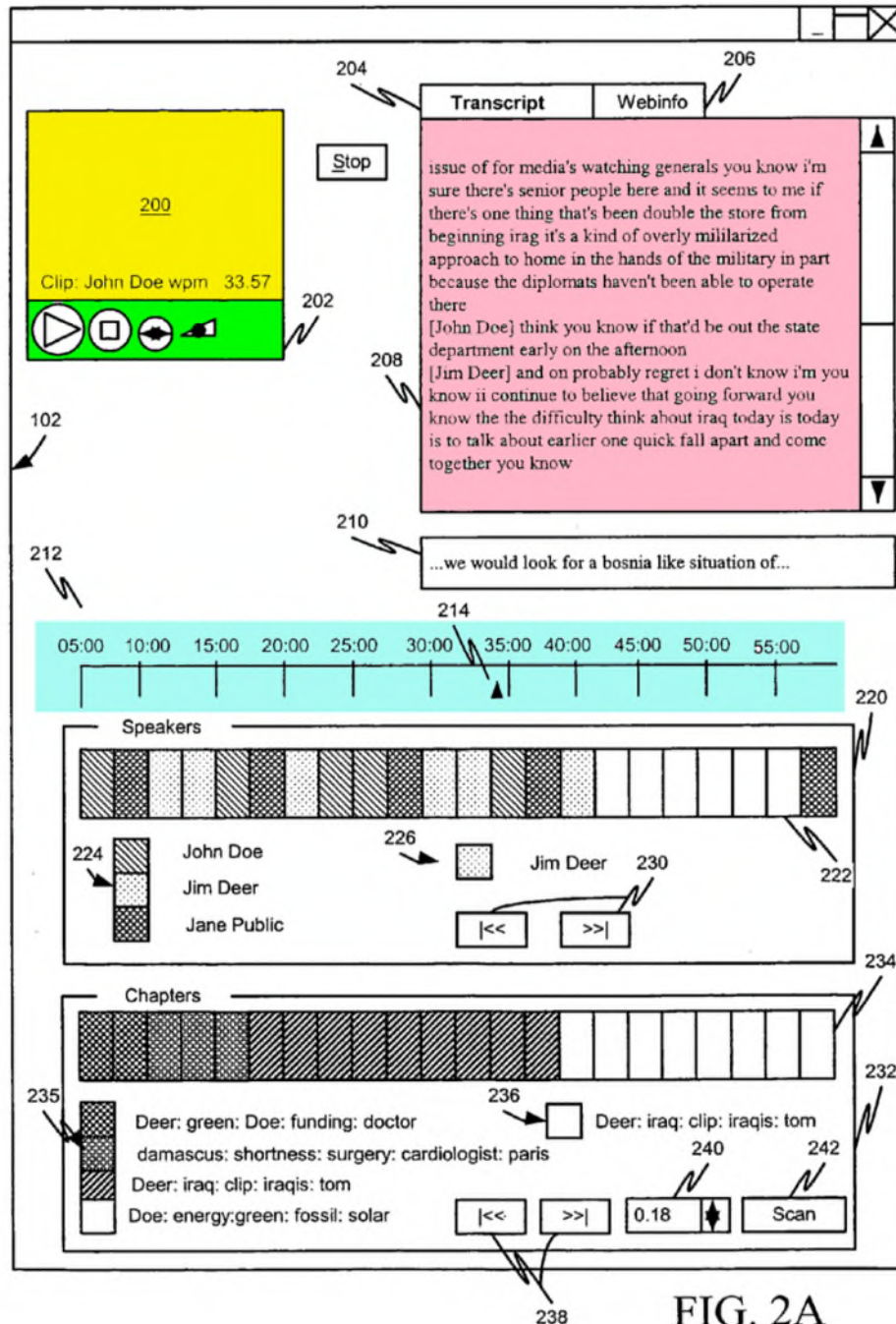


FIG. 2A

EX1004, FIG2A.

Goldberg's synchronized web page similarly "consists of the transcript, with time coding, side by side with a column containing a summary." EX1005, ¶[0073]. Goldberg's viewer uses the QUICKTIME player, and "[t]he client can drag the play bar in the QUICK TIME player to a desired position." EX1005, ¶[0115]; EX1003, ¶[0037].

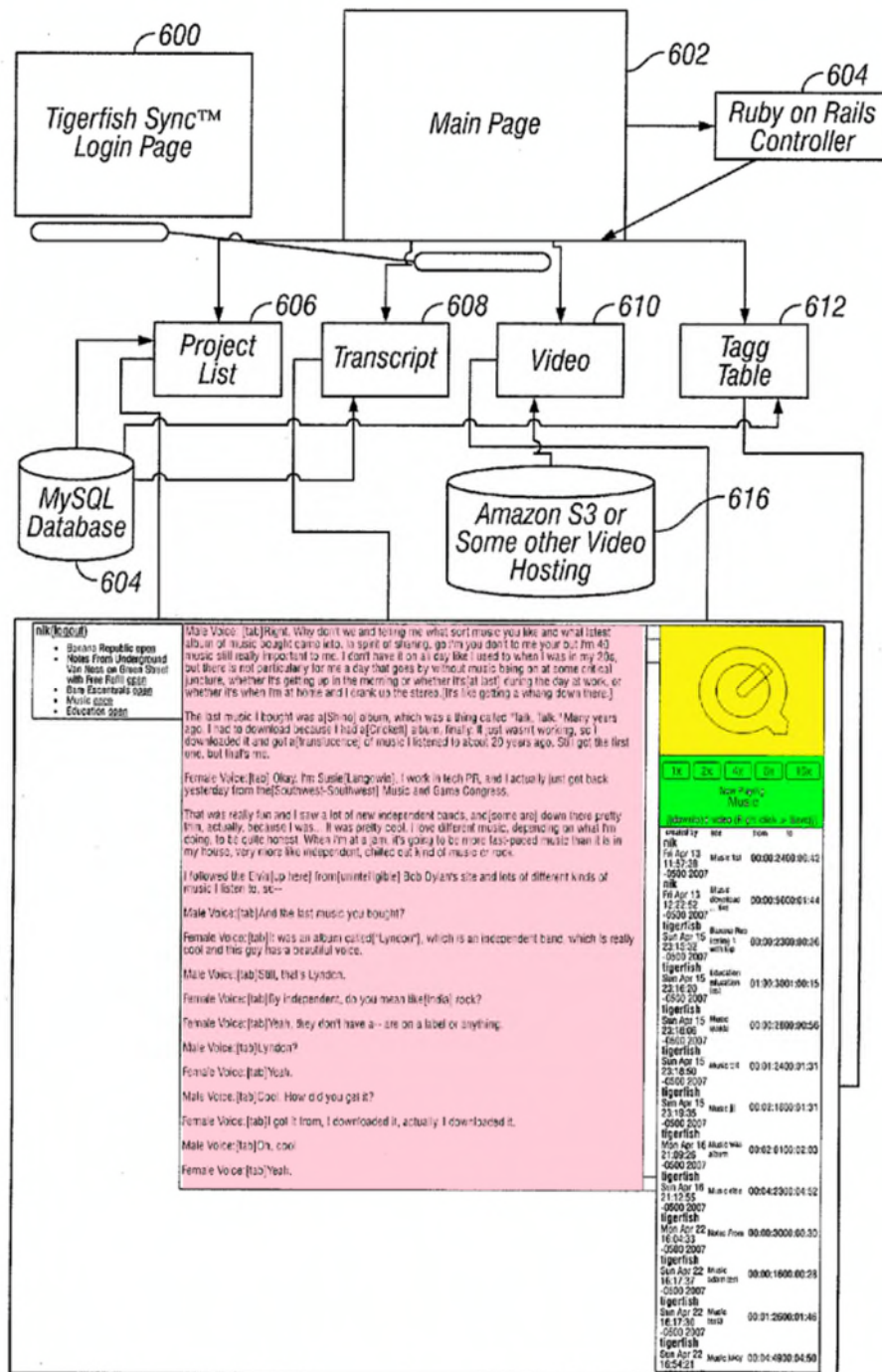


FIG. 6

EX1005, FIG6.

Thomas further discloses this combination of synchronized transcript and playback-position display was a recognized multi-pane interface convention, depicting a playback progress readout alongside synchronized transcript panes. EX1010, FIG17. A POSITA would have found displaying time-ordered transcript text alongside a graphical seek control to be a known user-interface design choice. EX1003, ¶[0038].

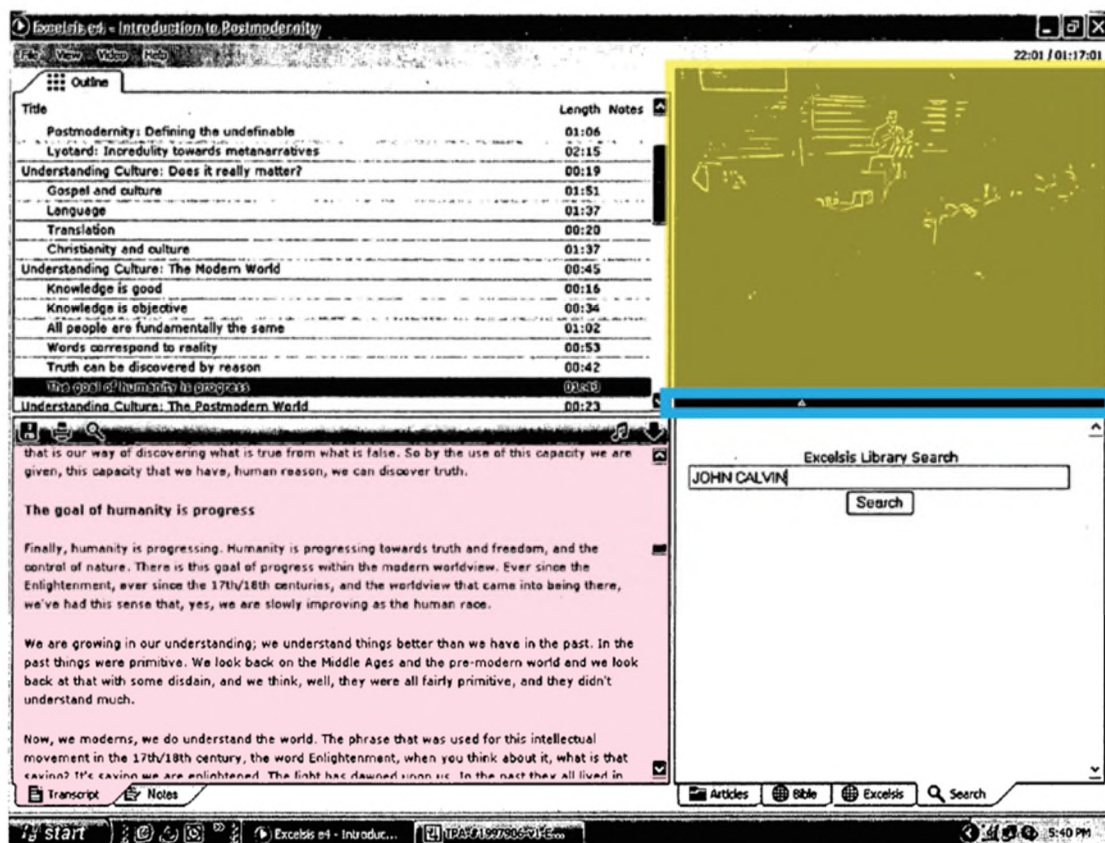


FIG. 17

EX1010, FIG17.

D. Navigating Playback by Selecting Displayed Text Was Well-Known

Clicking a word or line in a transcript to jump to the matching moment in a recording—like the navigation used by podcast apps and video editors—was well-known long before the '093 Patent. EX1003, ¶[0039].

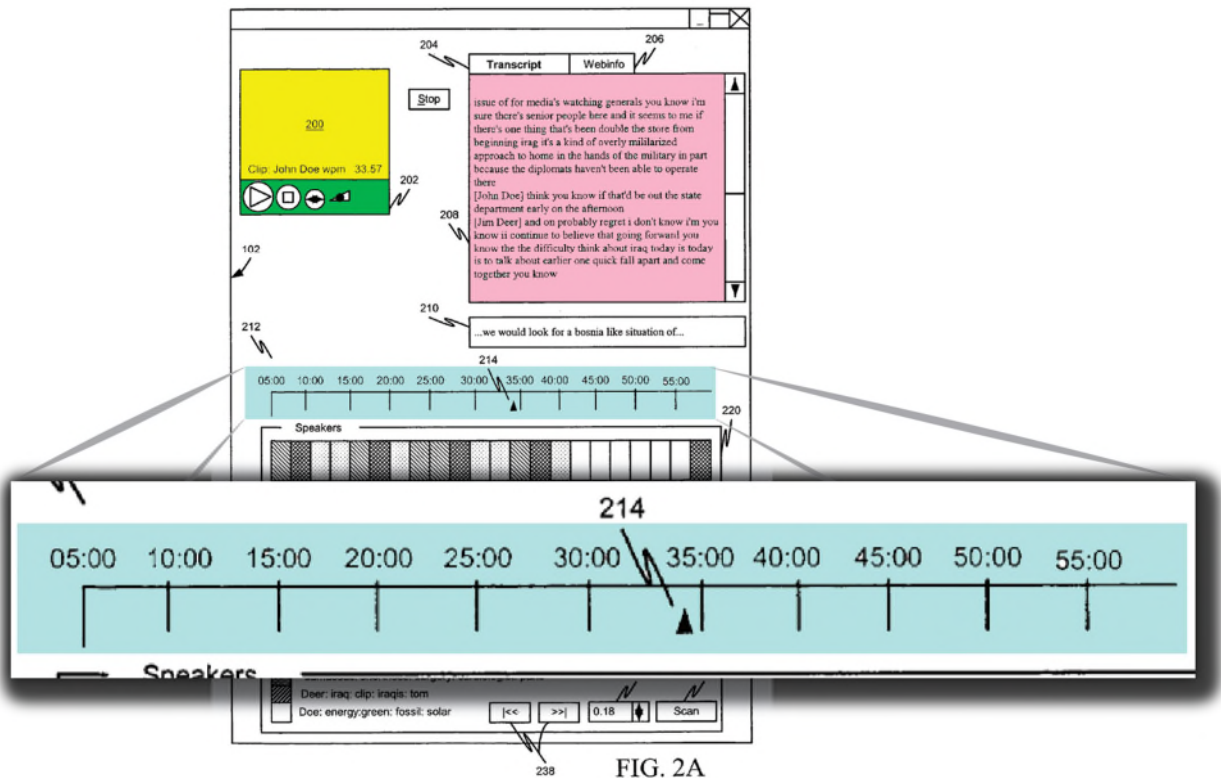
Mahajan teaches that “a user can easily scan the transcript in field 208 and decide on a relative part of the A/V content which the user wishes to view,” after which “a user can ... simply highlight a portion of the transcript which the user finds of interest,” including by “right click[ing] over a word on the transcript, ... highlight[ing] the transcript, or, ... drag[ging] a word from the transcript in field 208 and drop[ping] it on screen 200.” EX1004, ¶¶[0040]–[0041]. “Once the user has done this, player 122 illustratively ... locates the spot in the A/V content stream 104 that corresponds to the word in the recognized transcript in field 208,” because “the speech recognition results contain an identification as to where they were recognized in the A/V content stream,” which “is stored in metadata store 114.” EX1004, ¶[0042]. “Therefore, player 122 finds the particular place in the A/V content stream 104 that corresponds to the selected word, and begins playing A/V content 104 at that place.” EX1004, ¶[0043]. Mahajan also recites “a user interface providing a user actuable input, actuable to select a word in the transcription and to cause the player to begin playing the A/V content at a point in the A/V content that is aligned with the selected word.” EX1004, claim 3; EX1003, ¶[0040].

Goldberg discloses identical functionality: “the client may click on any text in the transcript. Clicking on a point in the ... transcript automatically locates the corresponding spot in the recording and plays the recording from this spot forward.” EX1005, ¶[0110]. This occurs because a “Set Time” event “passes the video ID and the time stamp which are embedded in the text to the QUICKTIME player.” *Id.* A POSITA would have recognized selecting transcript text to reposition audio/video playback was a well-established interaction pattern in transcript viewers. EX1003, ¶[0041].

E. Repositioning the Graphical Playback Control in Response to Text Selection Was Known

When a user clicks transcript text and playback jumps to its timestamp, the progress indicator naturally moves to that same point—ordinary media-player operation, not a new technical feature. EX1003, ¶[0042].

Mahajan discloses marker 214 which indicates “what part of the A/V content stream is currently being displayed on display 200.” EX1004, ¶[0044]. Because the player relocates playback to the time position associated with user-selected transcript text, the marker 214 necessarily moves from its prior position to a new position reflecting the selected text’s time. EX1004, ¶¶[0042]-[0043]; EX1003, ¶[0043].



EX1004, FIG2A

Goldberg's QUICKTIME play bar operates identically: the same play bar that "the client can drag ... to a desired position," also displays the player's current position, so that when the player begins playing "from this spot forward" in response to a text-click, the play bar's displayed position necessarily updates to the new time. EX1005, ¶¶[0115], [0110]; EX1003, ¶[0044]. A POSITA implementing Mahajan's or Goldberg's transcript-selection navigation feature with either reference's playback-position indicator would have understood, as a matter of ordinary skill and routine engineering, that the position indicator must be redrawn at the new playback

location—this is the only way a “current position” indicator can function. EX1003, ¶[0044].

III. ’093 PATENT

A. Overview

The ’093 Patent describes a familiar media-player interaction, not a new kind of technology. A user views a meeting recording with a time-stamped transcript—like closed captions alongside a video—clicks a line of text, and the playback seek bar jumps to the corresponding point in the recording. The patent therefore combines three ordinary features: (1) a transcript tied to timestamps, (2) a standard playback seek bar, and (3) a click that uses the selected text’s timestamp to move the seek bar. The specification implements those features with a server that stores the recording, transcript, and time information and a display apparatus that presents them to the user. EX1001, 1:19-20, 1:37-49, 1:50-54, 4:59-64; EX1003, ¶¶[0045]-[0048].

The patent addresses the difficulty of starting playback at a desired moment. EX1001, 1:24-33; 4:48-64; EX1003, ¶¶[0045]-[0046]. Unlike controls that force users to estimate positions without clear anchors, the alleged invention couples the transcript with the playback control, maps displayed text to its generation time and, upon selection, moves to the correct location on the overall timeline. EX1001, 4:48-5:13; EX1003, ¶[0046]. The graphical control region indicates playback process and

Patent No. 11,289,093

Petition for *Inter Partes* Review

may be a seek bar or equivalent. EX1001, 5: 5-7, 43:30-57, FIGs. 43-44, 47-48; EX1003, ¶[0046].

The specification discloses a sharing assistant server (6) that manages event content records, a schedule management server (8) that manages events, projects, and access control, and communication terminals such as an electronic whiteboard, videoconference terminal, vehicle-mounted device, or PC. EX1001, 5:40-50, 6:23-29, 6:49-55, 15:42-56, FIG8A; EX1003, ¶[0047]. Together, the servers store and organize content data across databases covering user authentication, access management, project membership, event schedules, conducted event records, event management, resource reservation, action items, and related information correlating content items to generation time. EX1001, 1:37-49, 17:8-64; 18:49-21:4, FIGs, 11-15; EX1003, ¶[0047]. A conversion server (9) converts voice received via the communication network into text. EX1001, 6:56-59, 21:56-65; EX1003, ¶[0047]. The sharing assistant server and the schedule management server register content, generate time-organized related information, and transmit time-organized data to communication terminals for display. EX1001, 26:8-28:33, 31:11-64, 34:32-64, 35:54-65; EX1003, ¶[0047].

Overall, the '093 Patent claims a familiar three-step interaction: (1) display a transcript marked with timestamps, (2) display a standard playback seek bar, and (3) when the user clicks text, use that text's timestamp to move the seek bar to the

corresponding point. Those were well-known features of media players and transcript viewers more than a decade before November 29, 2018. The claims dress up this familiar interaction in device, system, and method language—receiving text and time information, identifying the corresponding location, and displaying the control there—but those are the expected steps a media player performs when a user clicks time-stamped text. EX1001, claims 1 and 22; EX1003, ¶[0048]. Claims 8, 10, and 17 merely package the same functions in a client-server arrangement, while dependent claims add ordinary options such as starting playback or showing progress. EX1001, claims 8, 10, and 17; EX1003, ¶[0048].

B. Prosecution History

U.S. Patent Application No. 16/697,190 (the '190 application) was filed on November 27, 2019. It claims priority to two foreign applications: JP2018-223360, filed on November 29, 2018, and JP2019-178481, filed on September 30, 2019. EX1001; EX1003, ¶[0049]. The '190 application was originally filed with twenty claims directed to a display control apparatus and method that receives event voice data with time-stamped transcription, displays the text in chronological order alongside a graphical playback control (e.g., a seek bar), and, upon selection of particular text, positions (and optionally starts) audio playback at the corresponding time, optionally also displaying associated screen shots and using identifiers to obtain timing information from a server. EX1002, 67-71; EX1003, ¶[0049].

In a Non-Final Office Action July 6, 2021, the Examiner rejected the pending claims as anticipated by US 2009/0216531 A1 (“Yanagihara”) under 35 U.S.C. §102. EX1002, 556-567; EX1003, ¶[0050]. On November 8, 2021, the Applicant amended claims 1-20 and added new claims 21-22. EX1002, 587-596; EX1003, ¶[0050]. The amendments required the apparatus/system/method to use the time information to determine a second display location—distinct from the first—that corresponds to the first text data’s time. EX1002, 587-596; EX1003, ¶[0050]. The new claims recited a system comprising a server including the first circuitry and a display control apparatus communicably connected with the server that includes second circuitry (claim 21), the second circuitry being configured to control the display to show text data in chronological order and a graphical control region at a first display location, determine a second display location by reference to the time information, and control the display to show the graphical control region at the second display location (claim 22). EX1002, 594-596; EX1003, ¶[0050]. In the accompanying Remarks, the Applicant contended that no new matter was introduced with claims 21-22 and the rejection is overcome because Yanagihara neither discloses receiving text converted from voice together with time information from a server that manages event-generated content, nor teaches displaying a separate graphical control region that sets audio playback position or determining and relocating that control to a second display location by reference to the time

Patent No. 11,289,093

Petition for *Inter Partes* Review

information, and thus fails to anticipate the claimed features. EX1002, 597-601; EX1003, ¶[0051]. On November 11, 2021, the Examiner issued a Notice of Allowance allowing all pending claims because the prior art did not disclose or suggest selecting displayed text, determining, by reference to time information, a different display location corresponding to that text, and controlling the display to move the graphical control region to that second location. EX1002, 615-617; EX1003, ¶[0052]

IV. POSITA

A person of ordinary skill in the art (“POSITA”) relevant to the ’093 Patent as of November 29, 2018 would have had at least: (1) a bachelor’s degree in electrical engineering, computer science, computer engineering, or equivalent coursework, with at least two years of work experience in display control systems, multimedia communication systems, or voice-to-text processing for conferencing applications; or (2) a master’s degree in electrical engineering, computer science, computer engineering, or equivalent coursework, with coursework or research focused on speech recognition systems, real-time multimedia display control, or networked communication terminal architectures. An individual with additional education could have less practical experience, and vice versa. EX1003, ¶[0053].

V. CLAIM CONSTRUCTION

Petitioner contends that no express constructions are necessary for determining whether the prior art teaches or renders obvious every limitation of the Challenged Claims. EX1003, ¶[0054].

VI. OVERVIEW OF CHALLENGE

Petitioner requests cancellation of claims 1-22 as set forth below:

Ground	Claims	Statutory Basis	References
1	1-22	§103	Mahajan in view of Goldberg
2	1-22	§103	Mahajan in view of Goldberg and further in view of Wei

Petitioner’s challenge is based on the following references, each of which is prior art under 35 U.S.C. §102(a)(1) because it was publicly available more than a year before the earliest possible priority date of November 29, 2018. None of these references were considered or cited during prosecution of the ’093 Patent.

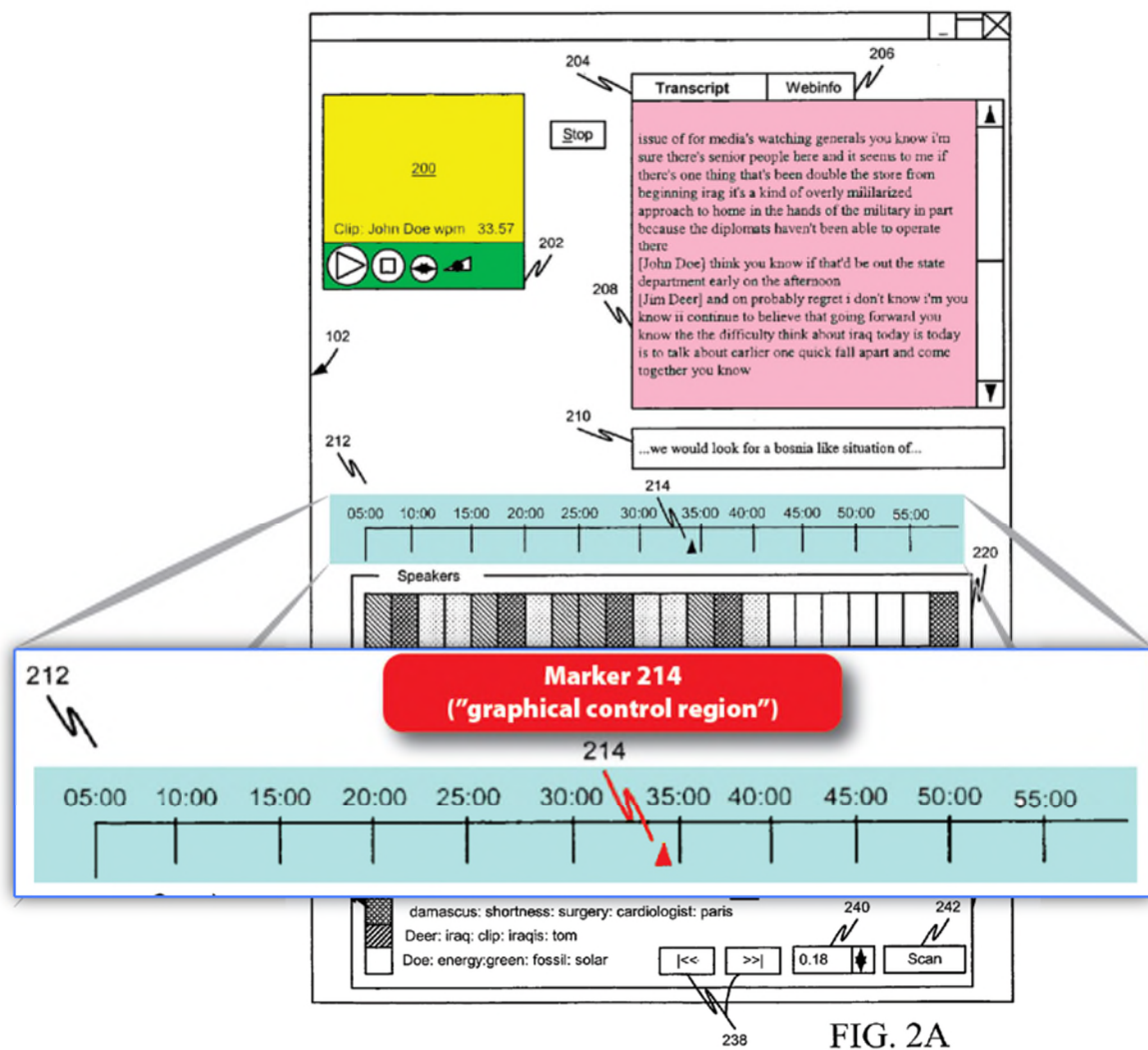
Exhibit	Reference	Date(s)	Availability as Prior Art
EX1004	U.S. Patent Application Publication No. 2008/0140385 (“Mahajan”)	Filed Dec. 7, 2006; Published Jun. 12, 2008	§102(a)(1)
EX1005	U.S. Patent Application Publication No. 2008/0319744 (“Goldberg”)	Filed May 27, 2008; Published Dec. 25, 2008	§102(a)(1)

Exhibit	Reference	Date(s)	Availability as Prior Art
EX1006	U.S. Patent Application Publication No. 2011/0145428 (“Wei”)	Published Jun. 16, 2011	§102(a)(1)

VII. OVERVIEW OF THE PRIOR ART

A. Mahajan (EX1004)

Mahajan discloses an audio/video processing system that performs automatic speech recognition on A/V content, time-aligns the resulting transcript with the content, displays the transcript alongside a graphical playback position indicator, and repositions playback when the user selects transcript text. EX1004, Abstract, ¶¶[0038], [0040]–[0045], [0052]; FIGs. 2A, 3, 4, claims 1–3; EX1003, ¶[0055].



EX1004, FIG2A.

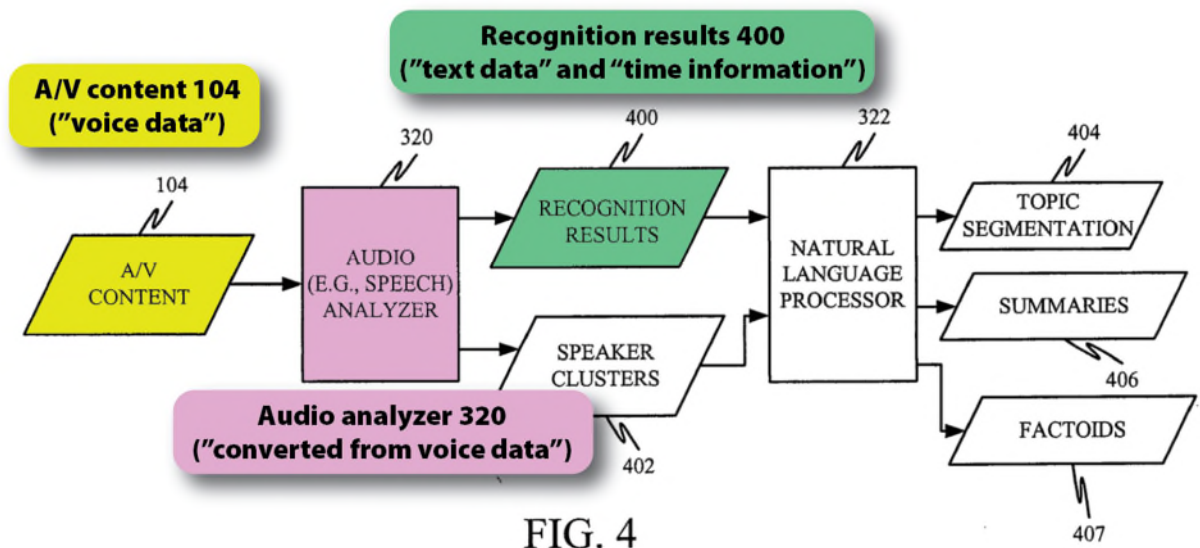


FIG. 4

EX1004, FIG4.

B. Goldberg (EX1005)

Goldberg teaches a server architecture (FTP server 302 and web server 316) that receives source material from live events, stores and organizes the resulting recordings and time-coded transcripts, and serves that content to a client through a web-based viewer. EX1005, ¶¶[0013], [0038]-[0053], [0055], [0072]-[0073], [0099], FIGs. 3, 6; EX1003, ¶[0056].

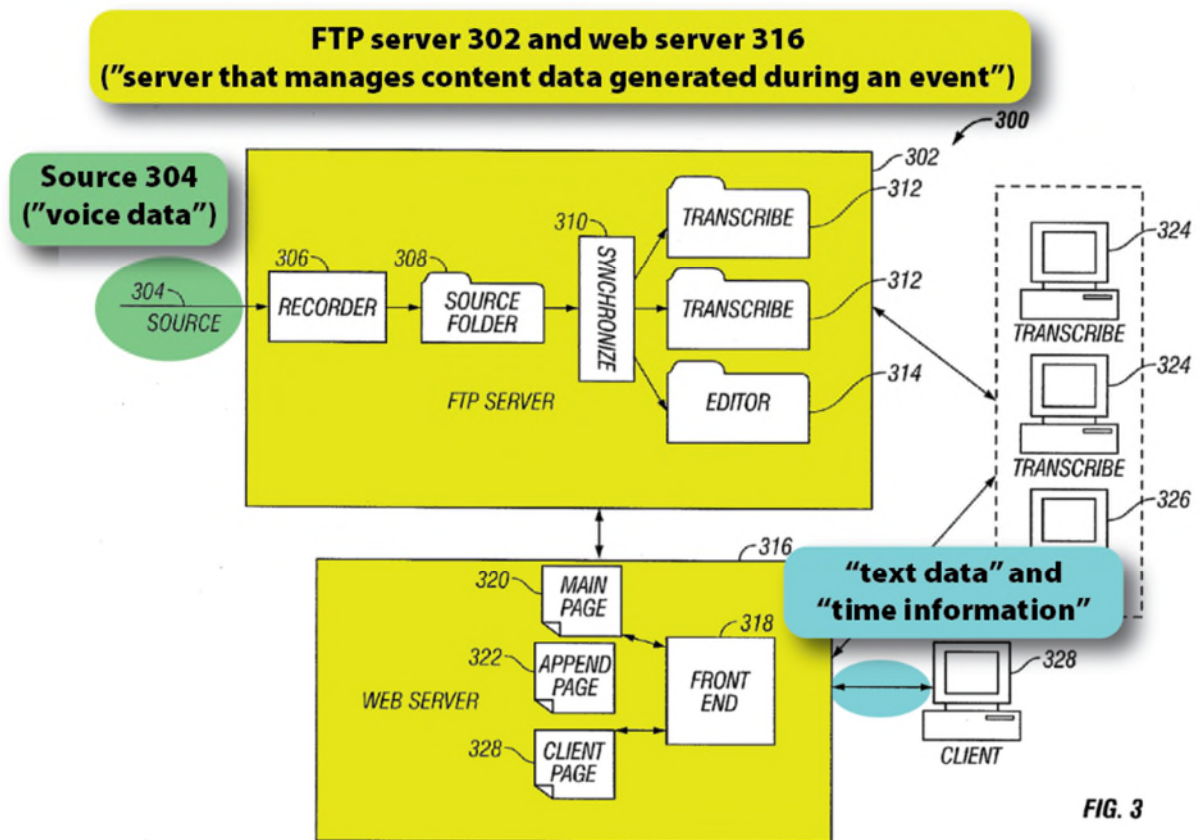


FIG. 3

EX1005, FIG3.

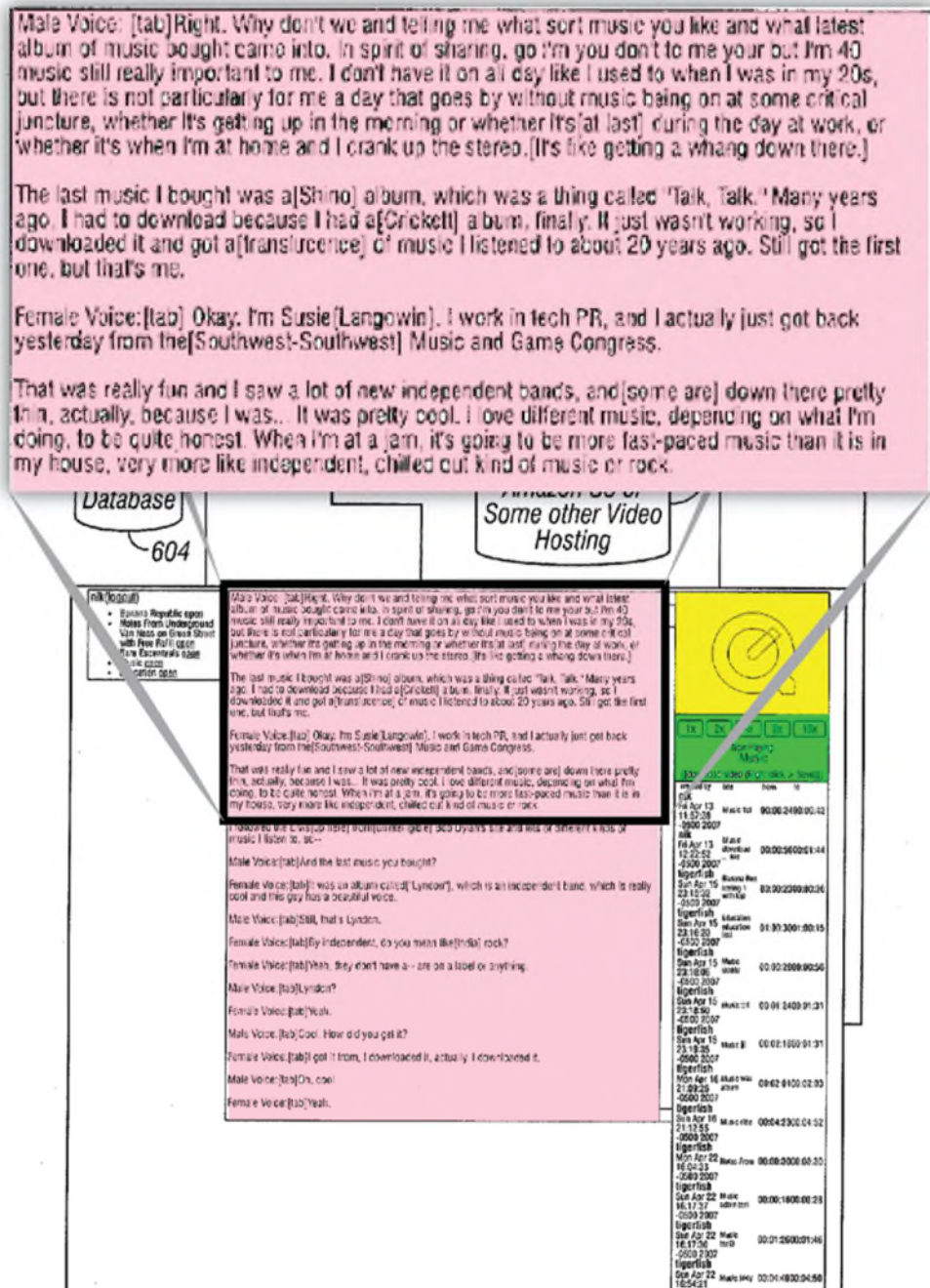
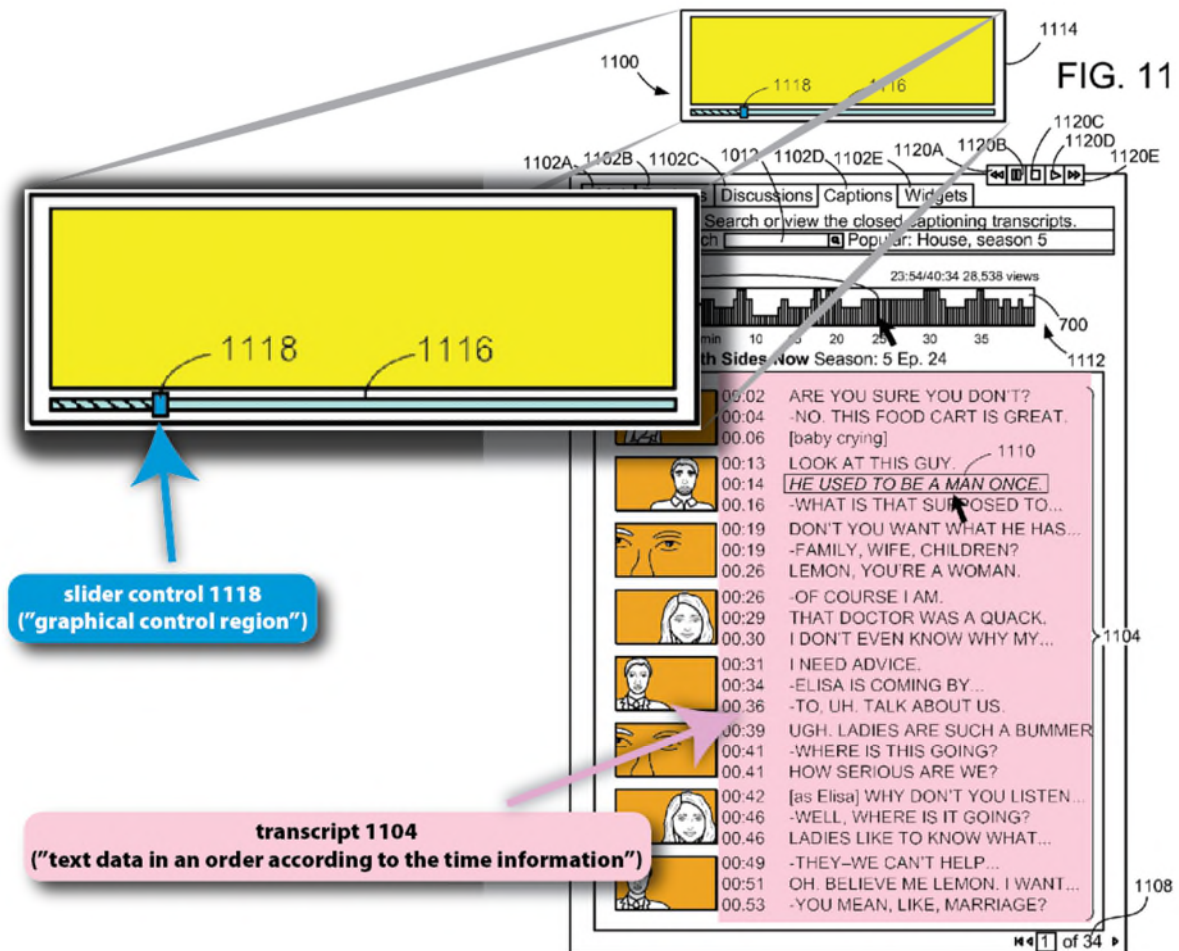


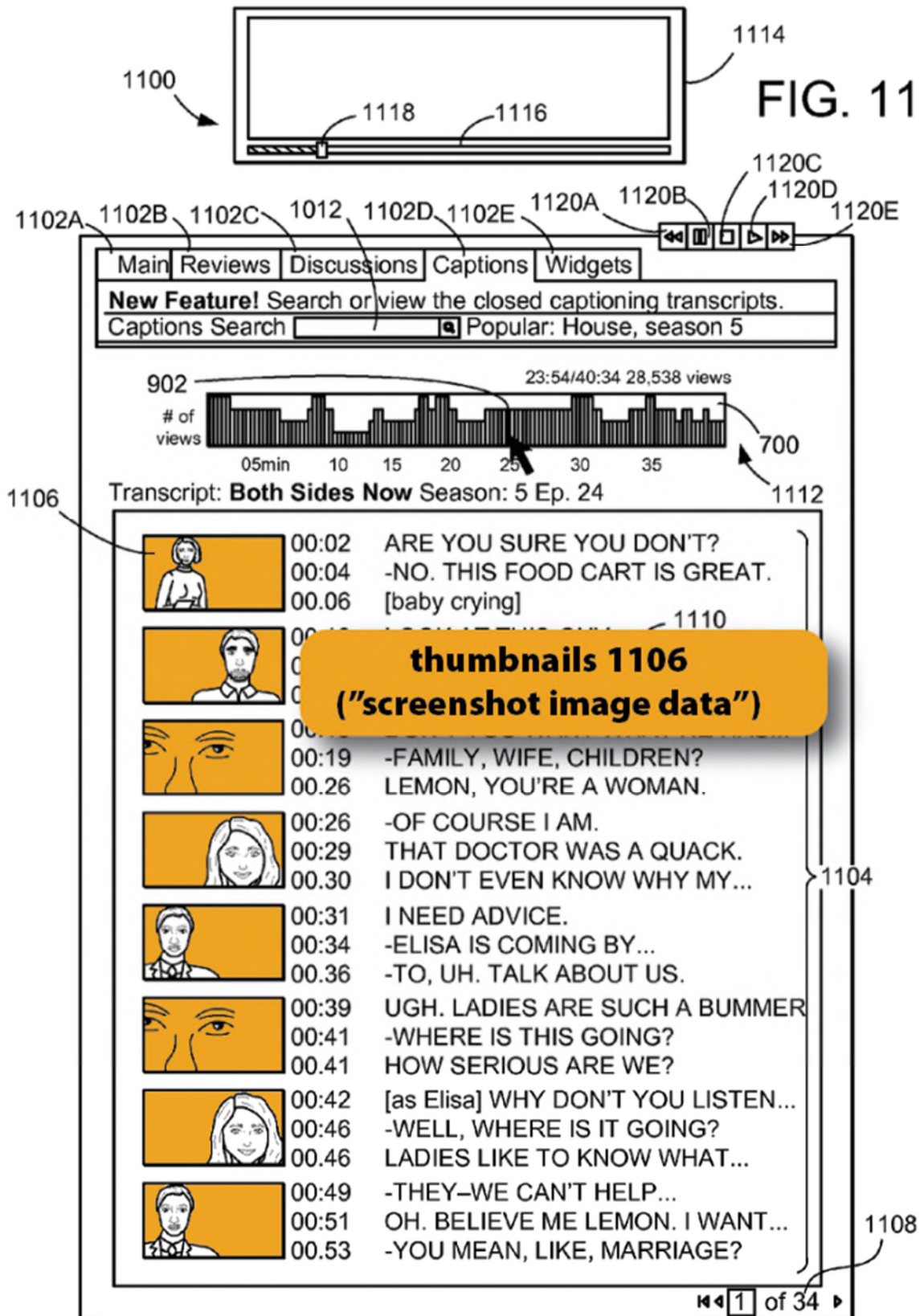
FIG. 6

EX1005, FIG6.

C. Wei (EX1006)

Wei discloses “a method and apparatus for navigating a media program” using a synchronized, searchable transcript displayed alongside a media player. EX1006, ¶[0021], Abstract. Wei’s system streams media program segments from a media server to a client computer. EX1006, ¶[0098]. Wei presents a user interface that includes a media player window, which displays the media program with a progress bar indicating which portion of the media program is being played, and a slider control that the user may move left and right to begin play at another point in the media program. EX1006, ¶[0127]. Wei’s transcript 1104 “includes the phrases of the dialog and an adjacent time stamp, indicating when the phrase was uttered in the media program,” together with “thumbnails 1106 associated with the phrases,” each “generated ... by software which captures a frame temporally associated with the phrase of the transcript.” EX1006, ¶[0129]. Wei further discloses that selecting a transcript phrase or histogram feature synchronizes the display, and that “[t]he portion 1110 of the transcript 1104 associated with that particular portion of the media program is highlighted.” EX1006, ¶[0130]. Wei’s transcript portions are defined, in one embodiment, “by closed captioning phrases,” EX1006, ¶[0124], FIG11, claim 10; EX1003, ¶[0057].





VIII. THE ASSERTED GROUNDS OF INVALIDITY

A. Ground 1: Mahajan in View of Goldberg Renders Obvious Claims 1-22

Mahajan in view of Goldberg renders obvious claims 1–22. Mahajan supplies an Automated Speech Recognition (ASR) system that converts recorded event audio into a time-aligned transcript, a display that presents that transcript in chronological order alongside a seek bar, and a text-selection mechanism that relocates the seek bar—and playback—to the position associated with user-selected transcript text. EX1004, ¶¶[0038], [0041]–[0045], [0052]. Goldberg supplies a server (FTP server 302 and web server 316) that manages event data: a “provider’s” server that receives, stores, and manages voice recordings and time-coded transcripts generated during live events, and that transmits that data to a client for synchronized display and navigation. EX1005, ¶¶[0013], [0055], [0072]–[0073], [0110]. As explained below, a POSITA would have been motivated to combine these teachings with a reasonable expectation of success. EX1003, ¶[0059].

1. Motivation to Combine

A POSITA would have been motivated to incorporate Goldberg’s server-managed event content architecture into Mahajan’s automatic speech recognition and transcript-navigation system. A POSITA combining these references would be doing nothing more than what any software developer would do: take an existing

transcript-navigation tool and host its recording, transcript, and timing data on a server. That is the standard client-server architecture used by ordinary web applications. Mahajan already supplies the transcript, timestamps, player, and text-selection navigation; Goldberg supplies the server that stores and serves event content. Combining the two references simply lets Mahajan's existing interface retrieve its data from Goldberg's server. EX1004, ¶¶[0023], [0033]; EX1005, ¶¶[0055], [0072]–[0073]; EX1003, ¶¶[0060]–[0065].

Same Field of Endeavor. Mahajan and Goldberg both address synchronizing time-coded text with recorded speech to enable transcript-driven navigation of the recording. EX1003, ¶[0061]. Mahajan discloses an “automatic content analyzer configured to receive A/V content and analyze the A/V content using speech recognition and natural language processing to generate speech metadata.” EX1004, claim 1. Goldberg discloses a viewer in which “[t]he rapid transcript includes time coding, speaker identification and summary” and “[a] viewer application allows a client to view a video recording side-by-side with a transcript.” EX1005, ¶[0013]. A POSITA designing a system for reviewing recorded events would have looked to both references. EX1003, ¶[0061].

A POSITA Would Have Looked At Goldberg To Make Mahajan's Content Available to Users. Mahajan analyzes “A/V content” generally without describing an architecture for storing, processing and distributing that content to end

users . EX1004, ¶¶[0023]–[0025], [0033]. A POSITA seeking to implement Mahajan’s functionality would naturally consider how to make the content available to others. Goldberg addresses this with its server that receives, stores, and serves recordings and time-coded transcripts of live events such as “financial conference calls,” “basic conference calls,” and “conventions and meetings.” EX1005, ¶¶[0013], [0038]–[0053], [0055], [0072]–[0073]. A POSITA implementing Mahajan’s speech-recognition and navigation functionality would have been motivated to adopt Goldberg’s server architecture to efficiently process and store content, and would have adopted Goldberg’s event-focused organization to allow easy access and retrieval of content by users. EX1003, ¶[0062].

Predictable Combination. The combination is the predictable use of known elements according to known methods to yield predictable results. *KSR Int’l Co. v. Teleflex Inc.*, 550 U.S. 398, 415–421 (2007). Mahajan supplies voice transcription, time alignment, transcript display, the player, and text-selection navigation, EX1004, ¶¶[0038], [0040]–[0045], [0052], while Goldberg supplies the server that stores and transmits event recordings and time-coded transcripts. EX1005, ¶¶[0013], [0055], [0072]–[0073], [0099]. The combination requires no new algorithm, unusual hardware, or change to either reference’s operation: Goldberg’s server provides the event data, and Mahajan’s player displays and navigates that data. The result is predictable—a familiar transcript-driven media player using server-hosted data. A

POSITA would therefore have had a reasonable expectation of success because Mahajan's architecture already contemplates retrieving content from a networked store, (EX1004, ¶[0023]), and Goldberg's server is expressly designed to store and serve that type of content to clients. EX1003, ¶¶[0063]-[0064].

No Disruption to the Core Navigation Functionality. The combination leaves Mahajan's transcript-selection and marker-repositioning functionality unchanged: Goldberg's server supplies event-based voice, text, and time information, while Mahajan's player, transcript display, time bar, and marker operate on that content exactly as Mahajan describes—precisely the architecture recited in the '093 Patent Challenged Claims, which require only that the display control apparatus receive text data and time information “from a server that manages content data generated during an event.” EX1001, claim 1; EX1003, ¶[0065].

2. Preambles

a) [1.pre]

Mahajan discloses this element. Mahajan's **A/V player 122** operates on a computing system and “accesses the A/V content and its associated metadata and presents the A/V content to the user by generating user interface 102” for use by a user. EX1004, ¶¶[0033], [0023]. The A/V player 122 generates and controls the display of the A/V content together with metadata-driven displays, including the text box 208 transcript display and the time bar 212. EX1004, ¶¶[0036], [0038], [0044].

A POSITA would have understood that the A/V player 122, which generates and controls the content of user interface 102 on a display, constitutes a “display control apparatus.” EX1003, ¶[0066].

Goldberg’s **computational device, which runs the viewer application** that displays the video recording side-by-side with the synchronized transcript, likewise constitutes a display control apparatus. EX1005, ¶¶[0013], [0024]-[0026]; EX1003 ¶[0067].

b) [10.pre]/[10.a]

Mahajan discloses this element in combination with Goldberg. Mahajan’s system includes “store 114,” from which “A/V player 122 ... accesses the A/V content and its associated metadata.” EX1004, ¶¶ [0023], [0033]. Mahajan’s content analyzer 320 “recognizes words in the A/V content 104 and outputs an automatically derived transcript of those words as recognition results 400,” while “align[ing] the recognized words with the A/V content 104, such that each word in the transcript is aligned with a specific time point in the A/V content 104.” EX1004, ¶[0052]. A POSITA would have understood that store 114, which retains the A/V content, the ASR-derived transcript, and the time-alignment metadata, constitutes “a memory that stores voice data recorded during the event, text data converted from the voice data, and time information for an event.” EX1003, ¶[0068].

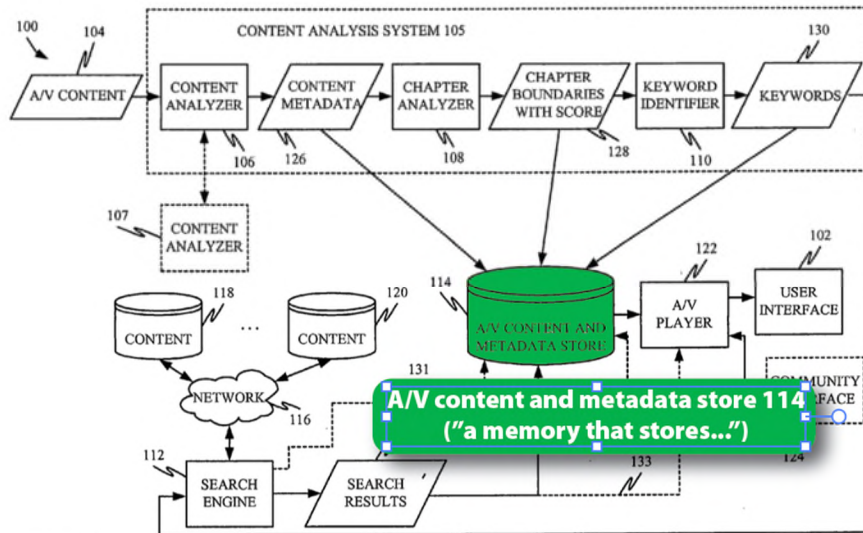


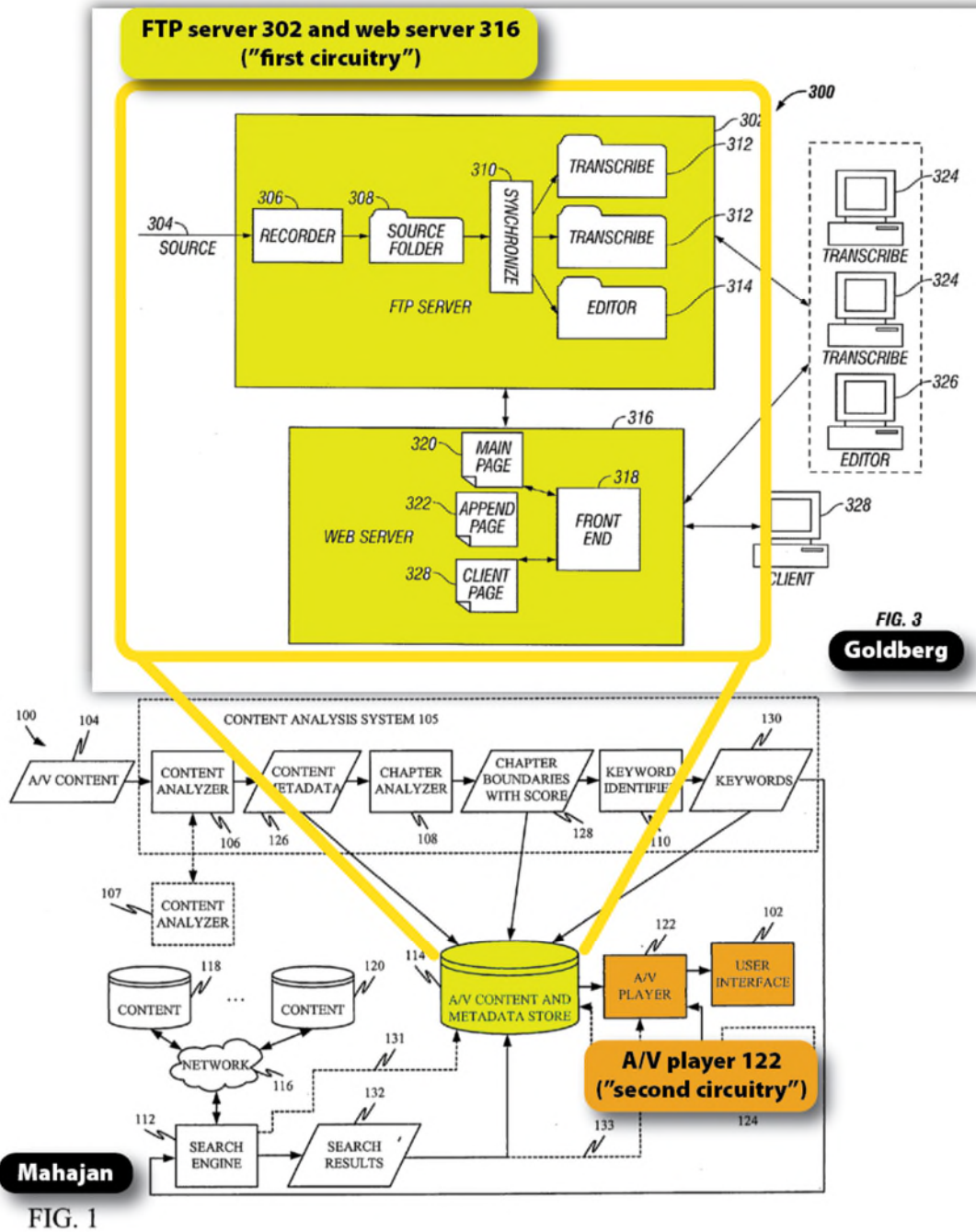
FIG. 1

EX1004, FIG1.

Goldberg supplies additional support for this limitation in the specific context of a live event. Goldberg’s server stores source material received from live events such as “financial conference calls,” “basic conference calls,” and “conventions and meetings,” together with the resulting transcript, which is presented to the client “with time coding, side by side with a column containing a summary.” EX1005, ¶¶[0037]-[0055], [0013], [0073], Abstract. A POSITA would have found it obvious to implement Mahajan’s store 114 as a server-hosted memory of the type disclosed in Goldberg, storing the voice, converted text, and time information for a specific conducted event. EX1003, ¶[0069].

As to the “first circuitry and second circuitry configured to cooperate together,” in the combination, Goldberg’s server (comprising the FTP server 302

and web server 316) that receives, stores, and organizes the event's voice data, converted text, and time information, (EX1005, ¶¶[0055], [0061]–[0062], [0072]–[0073]) constitutes the claimed “first circuitry.” EX1003, ¶[0070]. Mahajan's A/V player 122, which “accesses the A/V content and metadata in store 114 and generates user interface 102,” and performs the display and playback-repositioning functions described below (EX1004, ¶[0023], [0033]), constitutes the claimed “second circuitry.” EX1003, ¶[0070]. The first and second circuitries “cooperate together” in that Goldberg's server-side circuitry manages and supplies the event content while Mahajan's client-side player circuitry controls display of that content in a manner consistent with the functions recited in the remainder of the claim. EX1003, ¶[0070].



EX1005, FIG3; EX1004, FIG1.

c) [17.pre]

Mahajan in view of Goldberg discloses performing the method. Mahajan's A/V player 122 and content analyzer 320 perform the operations of receiving text data converted from voice data and associated time information, displaying that text data in time order, displaying a graphical control region (marker 214) at a first display location, receiving selection of text data, determining a new display location by reference to the time information of the selected text, and controlling the display to show the graphical control region at that new location. EX1004, ¶¶[0038], [0040]–[0045], [0052]; claims 1, 3. A POSITA would have understood that a processor-based system performing these functional operations through software execution necessarily performs the steps of a method. EX1003, ¶[0071]. Mahajan's claim 16 also recites performing a similar method. EX1004, claim 16.

In addition, Goldberg recites “[a] computer implemented method for rapidly producing and working with written transcripts of spoken source material comprising the steps of: receiving source material at a provider server; ... producing from said transcripts a single transcript of said source material, synchronized to said source material; and accessing said single transcript synchronized to said source material through said viewer.” EX1005, claim 1. Goldberg also recites method steps of “displaying a source material player and synchronized transcript for a selected project in a single viewer window[,]... using said transcript to navigate through said

source material[, and] using said source material to navigate through said transcript.”

EX1005, claim 23. In the combination, Goldberg’s server performs the method steps of receiving, storing, and serving the event’s voice, converted text, and time data, (EX1005, ¶¶[0055], [0072]–[0073], [0099]), while Mahajan’s A/V player 122 performs the remaining steps of displaying the text in time order, displaying the graphical control region, receiving text selection, determining the new display location, and repositioning the graphical control region, (EX1004, ¶¶[0038], [0045]), for the same reasons discussed below with respect to the corresponding limitations of claim 1. EX1003, ¶[0072].

d) [22.pre]

Mahajan discloses a non-transitory computer-readable medium storing computer-executable instructions. Mahajan’s system is implemented through software components, including “content analyzer 106,” “chapter analyzer 108,” and “A/V player 122,” that execute on a computing environment to perform the disclosed content-analysis, display, and navigation functions. EX1004, ¶¶[0022]–[0023]. Mahajan further discloses that “computer readable media” typically “embodies computer readable instructions.” EX1004, ¶[0085]; EX1003, ¶[0073].

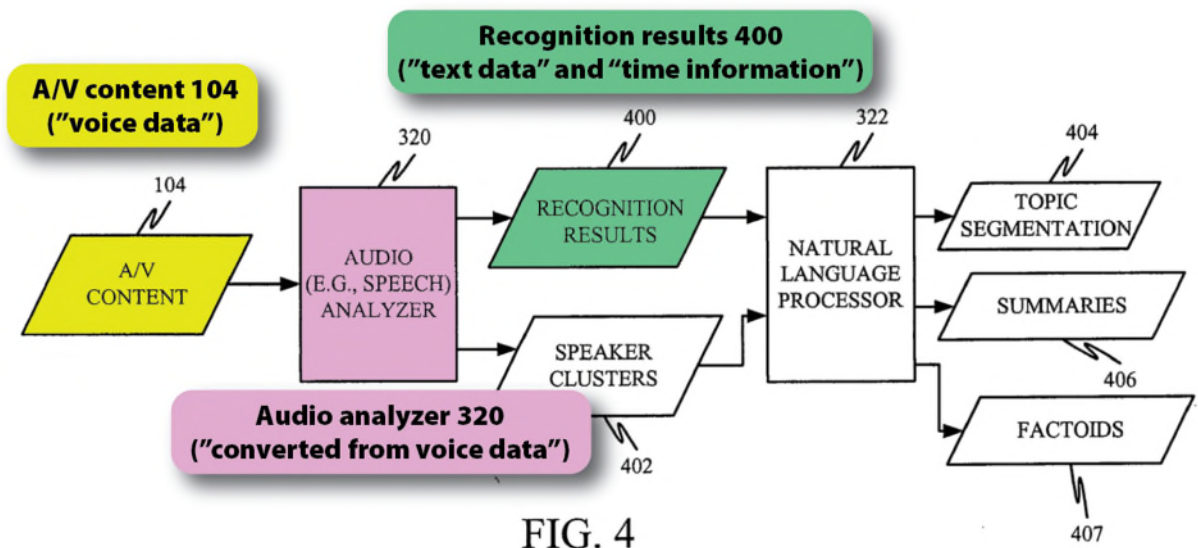
Goldberg likewise discloses stored program instructions on a non-transitory medium. Goldberg recites “[a] computer medium encoded with instructions, which when executed by a computational device performs a method for rapidly producing

and working with written transcripts of speech,” comprising the same steps of “receiving source material at a provider server; ... producing from said transcripts a single transcript of said source material, synchronized to said source material; and accessing said single transcript synchronized to said source material through said viewer.” EX1005, claim 16. Goldberg thus embodies its event-content server and synchronized-transcript-viewer functionality as instructions stored on a non-transitory medium. EX1003, ¶[0074].

In the combination, the instructions stored on Goldberg’s non-transitory medium (implementing the server’s reception, storage, and serving of the event’s voice data, text data, and time information, EX1005, ¶¶[0055], [0072]–[0073], [0099], claim 16) would operate alongside the instructions stored on Mahajan’s non-transitory medium (implementing content analyzer 106’s speech-to-text conversion and time-alignment, and A/V player 122’s text-order display, graphical control region display, text-selection receipt, time-information-based determination, and control-region repositioning EX1004, ¶¶[0038], [0045], [0052]). EX1003, ¶[0075]. When executed by the respective processors of the server and the client player, these combined stored instructions cause the processors to perform the same operations corresponding to claim 1. EX1003, ¶[0075].

3. [1.a]/[10.a]/[17.a]/[22.a]

Mahajan in view of Goldberg teaches this element. Mahajan discloses that its content analyzer is “configured to receive A/V content and analyze the A/V content using speech recognition and natural language processing to generate speech metadata.” EX1004, claim 1. More specifically, Mahajan discloses that “[a]nalyzer 320 ... recognizes words in the A/V content 104 and outputs an automatically derived transcript of those words as recognition results 400 ... using ... a large vocabulary automatic speech recognition system.” EX1004, ¶[0052]. Mahajan’s transcript constitutes “text data converted from voice data.” EX1003, ¶¶[0076]-[0077]. Mahajan further discloses that “[a]nalyzer 320 also aligns the recognized words with the A/V content 104, such that each word in the transcript is aligned with a specific time point in the A/V content 104,” which constitutes the claimed “time information.” EX1004, ¶[0052]; EX1003, ¶[0077]. Mahajan’s A/V player 122 receives this text and time information from metadata store 114. EX1004, ¶[0023]. Mahajan also explains that stores such as content stores 118 and 120 can be accessed via network 116 and further adds that “a wide variety of additional content stores may be available through network 116 as well.” EX1004, ¶[0023], FIG1.



EX1004, FIG4.

Goldberg discloses that the claimed data is received from “a server that manages content data generated during an event.” Goldberg’s server (FTP server 302 and web server 316) receives “source material” recorded during live events such as “financial conference calls,” “basic conference calls,” and “conventions and meetings,” converts that material into a time-coded transcript, and serves the recording and transcript to the client through a web page “with time coding, side by side with a column containing a summary.” EX1005, Abstract, ¶¶[0013], [0037]-[0055], [0060], [0073], [0099], FIG3. In the combination, Goldberg’s server performs the role of the “server that manages content data generated during an event,” supplying Mahajan’s content analyzer and A/V player with the voice data,

the ASR-converted text data, and the time information that Mahajan's system displays and navigates. EX1003, ¶¶[0078]-[0079].

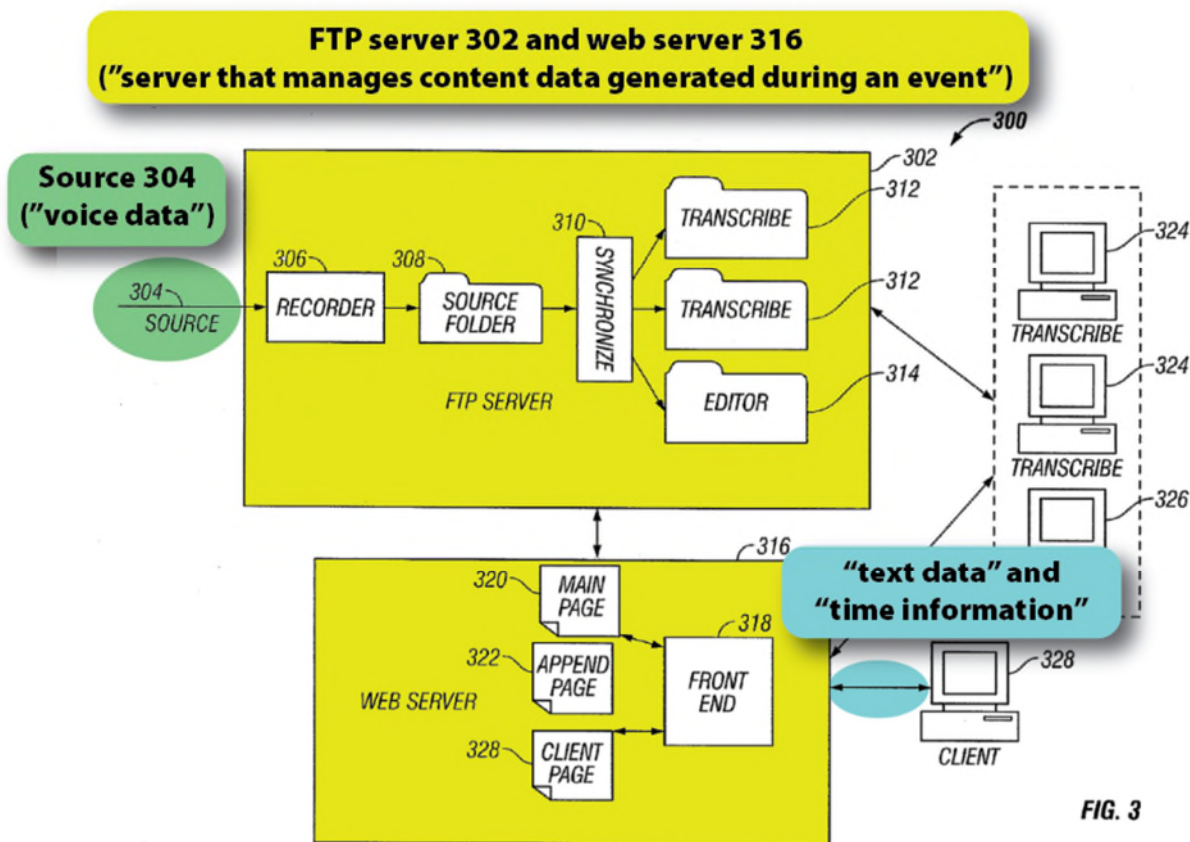


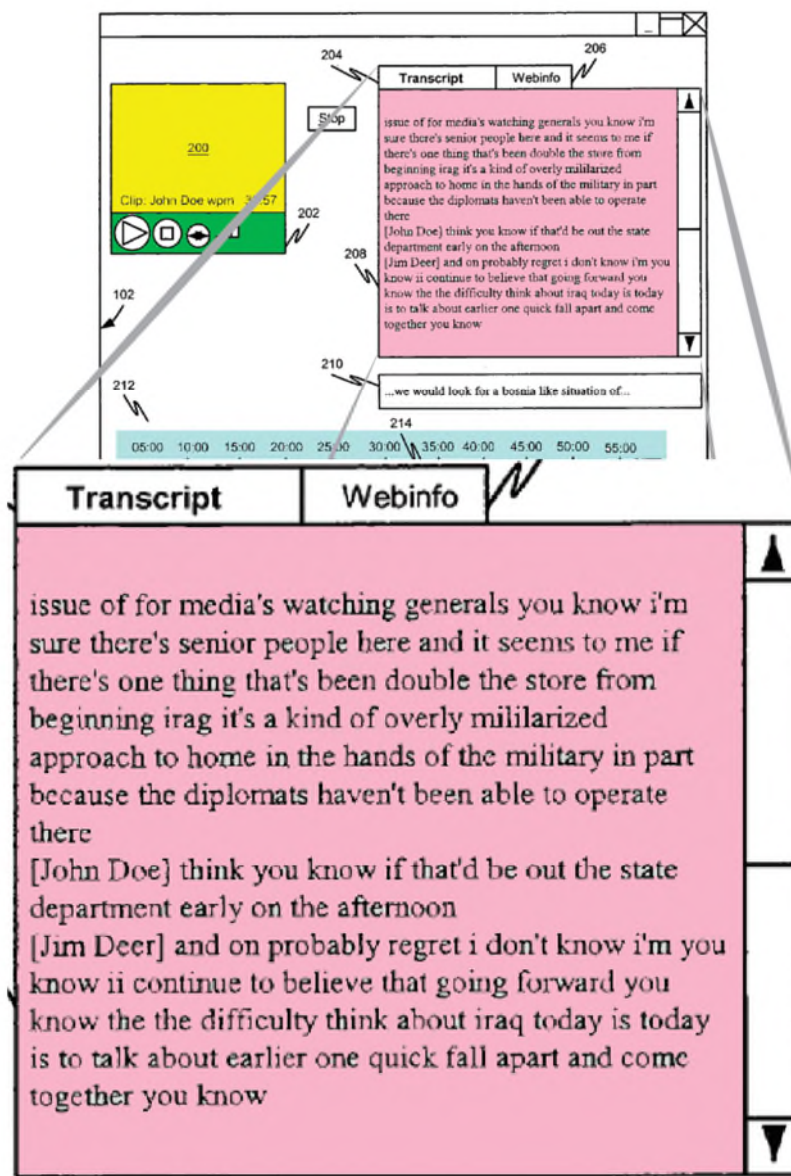
FIG. 3

EX1005, FIG3.

4. [1.b]/[10.b]/[17.b]/[22.b]

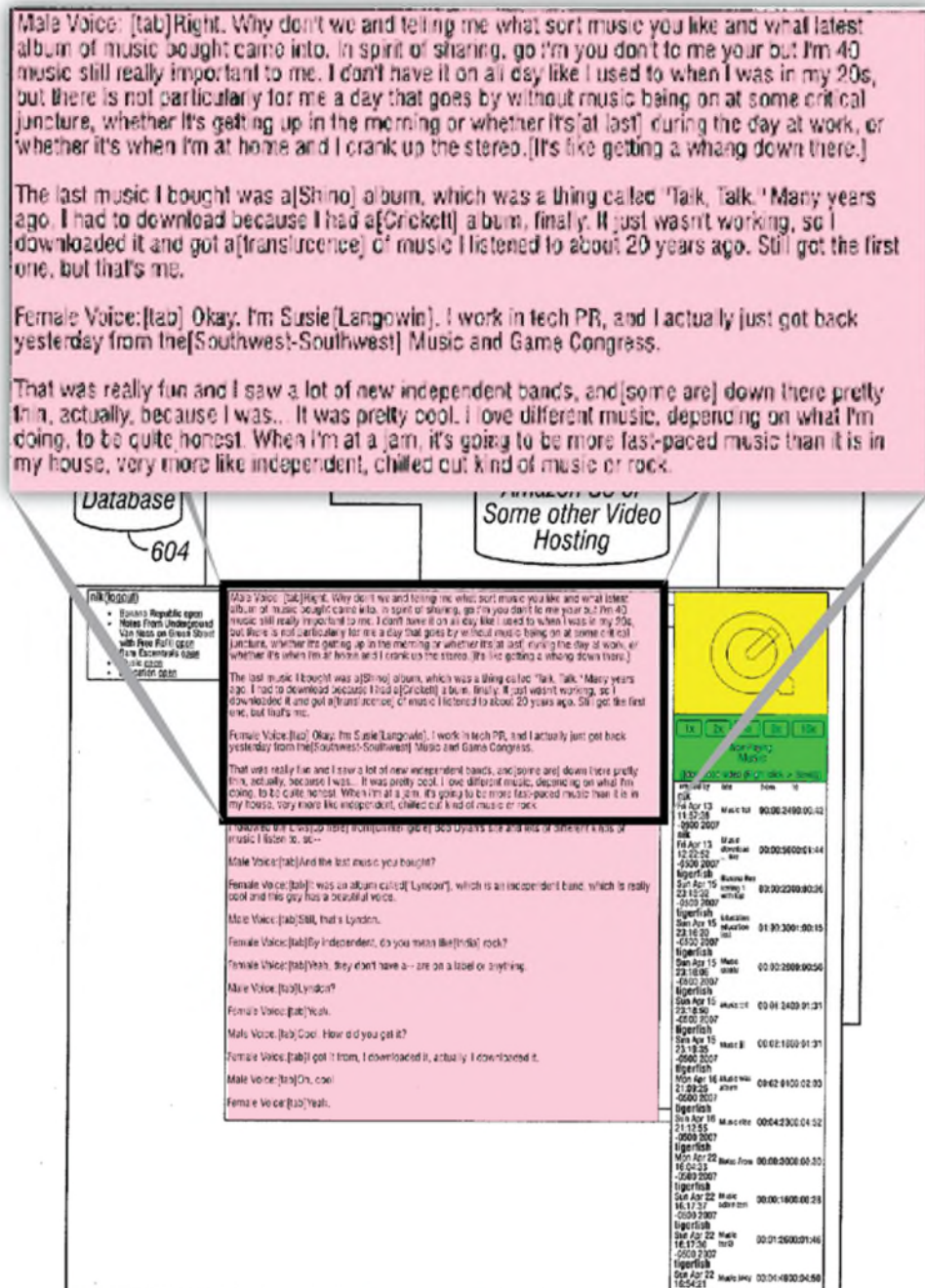
Mahajan discloses this element. Mahajan's text box 208 "displays speech recognition results corresponding to the A/V content currently being played on screen 200," and "as the A/V content 104 is played on screen 200, the speech recognition results corresponding to the audio portion of A/V content 104 being

played are displayed (e.g., scrolled) in text box 208.” EX1004, ¶[0038]. Because each word in the transcript “is aligned with a specific time point in the A/V content 104,” and is scrolled as corresponding A/V content is displayed, the transcript is displayed in chronological order according to the time information. EX1004, ¶[0052]; EX1003, ¶[0080].



Patent No. 11,289,093
Petition for *Inter Partes* Review
EX1004, FIG2A.

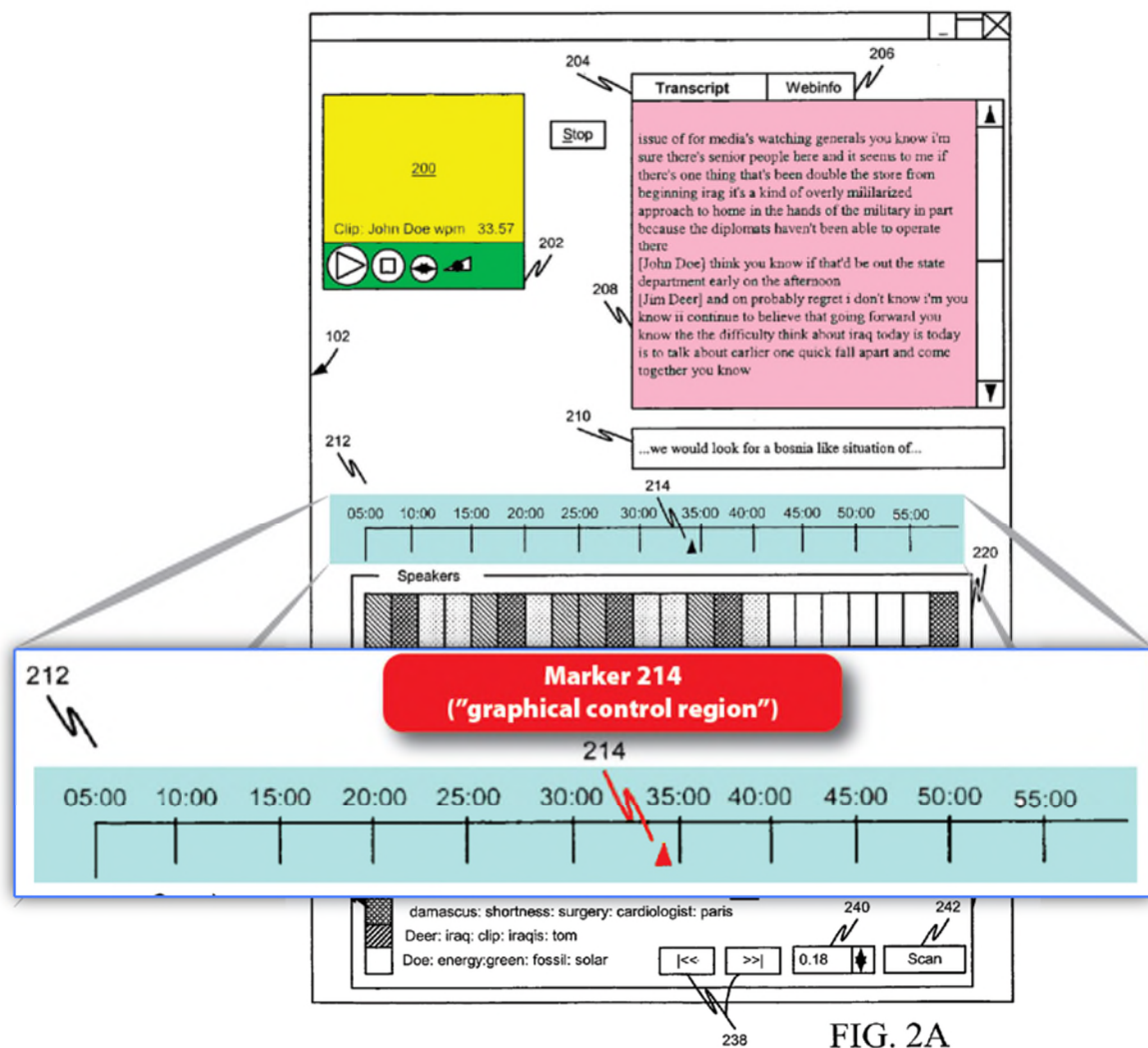
Goldberg's transcript is likewise displayed "with time coding, side by side with a column containing a summary," and the time codes and the transcript words are "associated with each other in a one-to-one relationship" such that the transcript reflects the chronological order of the underlying recording. EX1005, ¶¶[0073], [0099]; EX1003, ¶[0081].



EX1005, FIG.6.

5. [1.c]/[10.c]/[17.c]/[22.c]

Mahajan discloses this element. Mahajan's "user interface 102 includes a time bar 212," which "has incremental time displays ... along the top of bar 212 that indicate a beginning and end, and intermediate points, within the A/V content 104." EX1004, ¶[0044]. "A marker 214 is also shown along time bar 212," and "[m]arker 214 indicates what part of the A/V content stream is currently being displayed on display 200." *Id.* Additionally, "time bar 212 can be manipulated by the user," such that "the user can illustratively ... select marker 214 ... and move it along time bar 212 to any desired point," whereupon "A/V player 122 will skip to that time point in the A/V content stream 104." EX1004, ¶[0045]. **Marker 214 on time bar 212** constitutes the claimed "graphical control region ... that sets playback position in a total playback time," and its position when first displayed (for example, at "approximately 34 minutes into the A/V content stream 104" as shown in Mahajan's FIG2A, EX1004, ¶[0044]) constitutes the "first display location." EX1003, ¶¶[0084]-[0085].



EX1004, FIG2A.

Goldberg's QUICKTIME player play bar performs the identical function:
 "[t]he client can drag the play bar in the QUICK TIME player to a desired position."

EX1005, ¶[0115]; EX1003, ¶[0086]-[0087].

6. [1.d]/[17.d]/[22.d]

Mahajan discloses this element. Mahajan teaches that “a user can then simply highlight a portion of the transcript which the user finds of interest,” and that “the user can right click over a word on the transcript, the user can highlight the transcript, or,... the user [] can drag a word from the transcript in field 208 and drop it on screen 200.” EX1004, ¶[0041]. This results in “receiv[ing] selection of first text data from the text data being displayed.” EX1003, ¶¶ [0088]-[0089].

Goldberg discloses the same functionality: “the client may click on any text in the transcript.” EX1005, ¶[0110]; EX1003, ¶¶[0090]-[0091].

7. [1.e]/[10.d]/[17.e]/[22.e]

Mahajan discloses this element. Mahajan teaches that “[o]nce the user has done this [i.e., selected a transcript word], player 122 illustratively locates the spot in the A/V content stream 104 that corresponds to the word in the recognized transcript in field 208,” and that “the speech recognition results contain an identification as to where they were recognized in the A/V content stream,” which “is stored in metadata store 114.” EX1004, ¶[0042]. “Therefore, player 122 finds the particular place in the A/V content stream 104 that corresponds to the selected word, and begins playing A/V content 104 at that place.” EX1004, ¶[0043]. Because Mahajan expressly discloses that marker 214 indicates “what part of the A/V content stream is currently being displayed on display 200,” locating and beginning playback

at a new time position requires the system to determine, using the time information associated with the selected word, a new position on time bar 212 for marker 214 that is different from marker 214's prior position and that corresponds to the time information of the selected word. EX1004, ¶[0044]; EX1003, ¶¶[0092]-[0093].

Goldberg discloses the same behavior: “[c]licking on a point in the transcript automatically locates the corresponding spot in the recording and plays the recording from this spot forward. This occurs because an event called ST (Set Time) 704 in a JavaScript library 702[] passes the video ID and the time stamp which are embedded in the text to the QUICKTIME player 710.” EX1005, ¶[0110]; EX1003, ¶[0094].

A POSITA would have understood that determining the new playback position from the selected text's embedded time stamp, as taught by Mahajan and Goldberg, includes determining a new location for the position indicator (Mahajan's marker 214 or Goldberg's play bar position) that indicates playback position, because such indicators only serve their purpose by continuously reflecting the current playback position within the total playback time. EX1003, ¶¶[0095]-[0096].

8. [1.f]/[10.e]/[17.f]/[22.f]

Mahajan discloses this element for the reasons discussed above. Because marker 214 “indicates what part of the A/V content stream is currently being displayed on display 200,” and selecting text causes the player to “begin[] playing A/V content 104” at the new time position associated with the selected word, the

system necessarily controls the display to show marker 214 at the new position on time bar 212 corresponding to that new playback position. EX1004, ¶¶[0044], [0043]; EX1003, ¶¶[0097]-[0098]. This is confirmed by Mahajan's disclosure that the user can manually "move [marker 214] along time bar 212 to any desired point," after which "A/V player 122 will skip to that time point in the A/V content stream 104 and being[sic] playing the A/V content stream 104 at that point," demonstrating that marker 214's displayed position and the player's playback position are synchronized; changing one necessarily changes the other. EX1004, ¶[0045]; EX1003, ¶¶[0098].

Goldberg's play bar likewise displays the current playback position and would be displayed at the new position after the QUICKTIME player begins playing "from this spot forward" in response to a text-click. EX1005, ¶¶[0110], [0115]; EX1003, ¶[0099].

A POSITA implementing Mahajan's transcript-navigation feature, whether or not combined with Goldberg's event-content server, would have understood that redisplaying the graphical position indicator at the new playback location is the only way such an indicator can perform its stated function of showing "what part of the A/V content stream is currently being displayed on display 200." EX1004, ¶[0044]; EX1003, ¶¶[0100]-[0101].

9. Claim 2/18

Mahajan discloses this element. As discussed at [1.e]/[1.f], once the user selects transcript text, “player 122 finds the particular place in the A/V content stream 104 that corresponds to the selected word, and begins playing A/V content 104 at that place.” EX1004, ¶¶[0042]–[0043]. Because the selected text is time-aligned with a specific point in the content, and playback begins at that point, Mahajan discloses starting playback from the time the selected text was generated. EX1004, ¶¶[0052], [0043]; EX1003, ¶¶[0102]–[0103]. Mahajan’s claim 3 also recites this functionality. EX1004, claim 3.

Goldberg discloses the same: clicking transcript text “automatically locates the corresponding spot in the recording and plays the recording from this spot forward,” because the “Set Time” event “passes the video ID and the time stamp which are embedded in the text to the QUICKTIME player,” which “plays the video, beginning at the spot indicated by the time stamp 706.” EX1005, ¶[0110]. Because that time stamp corresponds to when the underlying voice was recorded, Goldberg discloses starting playback from the time the selected text was generated. EX1003, ¶[0104].

In the combination, Goldberg’s server supplies the event’s voice, converted text, and time information as discussed at [1.a], while Mahajan’s A/V player 122 performs the playback-start function using that time information exactly as Mahajan

discloses. EX1003, ¶[0105]. A POSITA would have understood that starting playback from the selected text's time is taught by each reference individually, such that combining them yields nothing more than the predictable result of starting playback from that time. EX1003, ¶¶[0105]-[0106].

10. Claim 3

Mahajan discloses this element. As discussed above at [1.c], marker 214 on time bar 212 “indicates what part of the A/V content stream is currently being displayed,” and its position, for example, “approximately 34 minutes into” a “57 minutes long” stream, continuously reflects how far playback has advanced, (i.e., the current playback progress). EX1004, ¶[0044]; EX1003, ¶¶[0107]-[0108]. Time bar 212 is further depicted as showing a time legend that indicates the total length of the recording, and intermediate marks that show the time corresponding to marker 214. EX1003, ¶[0108]. Mahajan confirms that as the user moves marker 214, playback skips “to that time point,” demonstrating that the marker's position and playback progress are one and the same. EX1004, ¶[0045]; EX1003, ¶[0108].

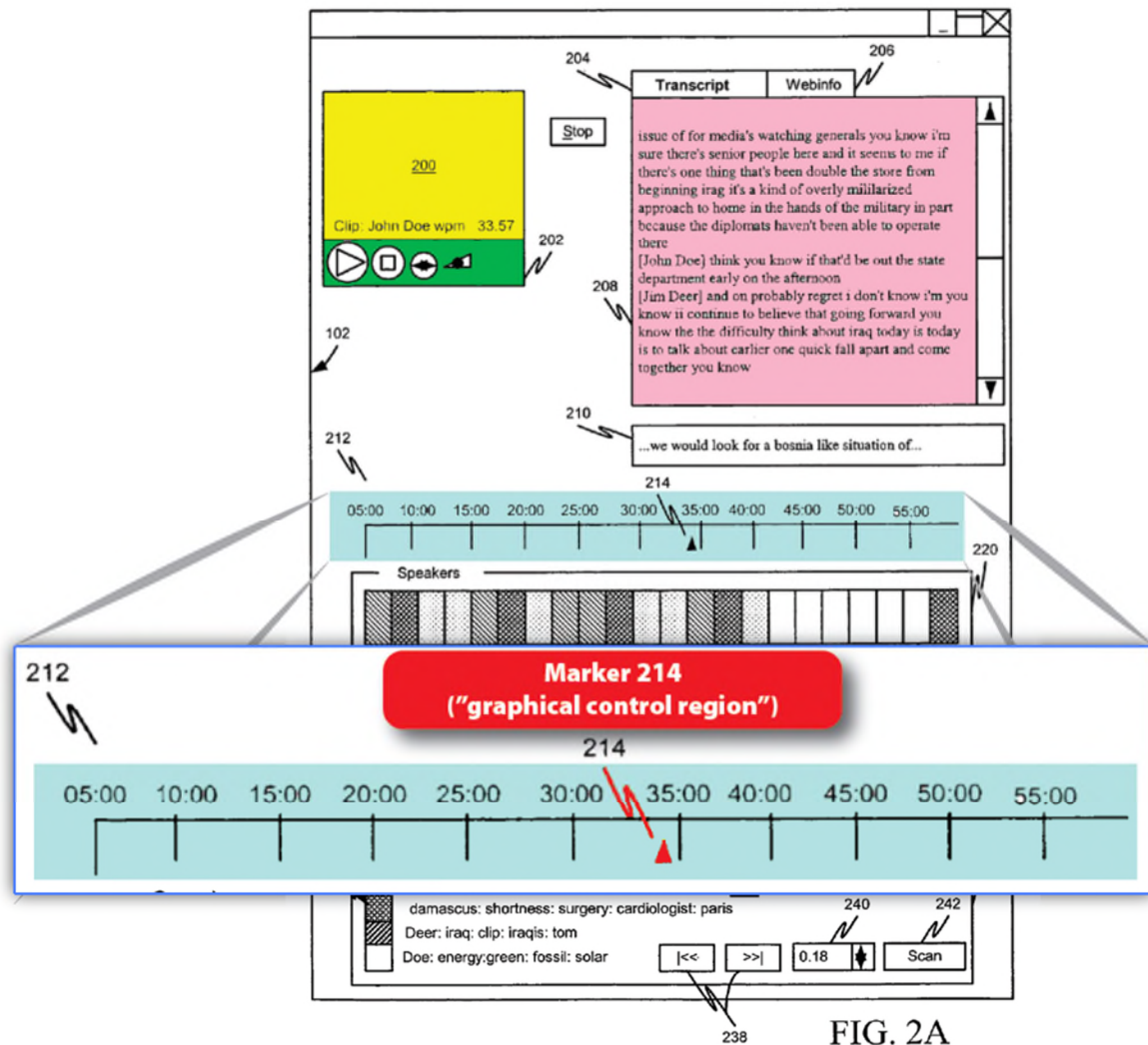


FIG. 2A

EX1004, FIG2A.

Goldberg confirms this functionality through its play bar, which the client can “drag ... to a desired position,” and its “Triple Mode Seekbar,” which lets the user “navigate through the video” while displaying the current timecode. EX1005, ¶[0015], FIG10; EX1003, ¶[0109]. A POSITA would have understood that any such graphical playback-position indicator indicates playback progress as a feature of its

function, and that this functionality, applied to marker 214 as populated with Goldberg's event data, renders obvious claim 3. EX1003, ¶¶[0109]-[0110].

11. Claim 4

Mahajan discloses this element. Time bar 212, together with marker 214, allows the user to “select marker 214 ... and move it along time bar 212 to any desired point,” causing playback to “skip to that time point.” EX1004, ¶[0045]. A bar with a draggable position indicator/control is a “seek bar.” EX1003, ¶¶[0111]-[0112].

Goldberg confirms this. Goldberg's “Triple Mode Seekbar” allows the user to “navigate through the video with a fundamental change by the minute,” second, or millisecond, and operates together with the same play bar the “client can drag ... to a desired position.” EX1005, FIG10, ¶[0115]; EX1003, ¶[0113]. A POSITA would have understood Goldberg's express use of “Seekbar” to confirm that Mahajan's draggable marker 214/time bar 212 was known in the art under that name, such that implementing it as a seek bar is the predictable application of a known GUI element to functionality Mahajan already discloses. EX1003, ¶[0113].

12. [5.a]/[20.a]

Mahajan discloses receiving screenshots associated with the text data. Mahajan discloses that “[u]ser interface 800 also includes a thumbnail section 802 which illustratively includes a plurality of different thumbnail photographs that represent, in some way, a portion of the A/V content 104.” EX1004, ¶[0071].

Mahajan further discloses that “[u]ser interface 800 also includes a subject matter section 812,” which “has a corresponding topic index 814 that can be used to identify the topic, or subject matter, being discussed during the corresponding portion of the A/V content.” EX1004, ¶[0073]. Mahajan expressly associates thumbnail images with specific text-identified subject matter: “during the portions of the A/V content corresponding to the first two thumbnails in section 802, intelligence failures are the primary topic of interest. During the portion of the A/V content associated with the second and third thumbnails in section 802, the war in Iraq is the primary topic of interest, etc.” EX1004, ¶[0073]. Because a POSITA would have understood that a “thumbnail” is a reduced-sized version of a larger image, and because Mahajan discloses that each thumbnail represents a captured portion of the on-screen A/V content, a POSITA would have equated thumbnails with screenshots. EX1003, ¶[0114].

Because each thumbnail is expressly linked to the corresponding subject matter being discussed, a POSITA would further have understood Mahajan’s thumbnail images to be “screenshot image data associated with the text data.” EX1004, ¶[0073]; EX1003, ¶[0115].

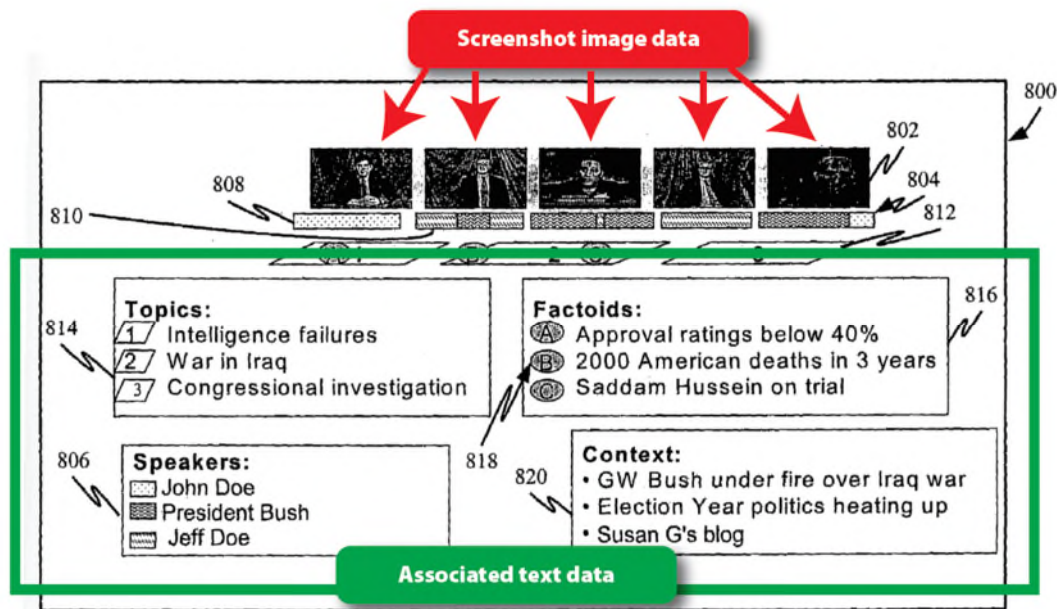


FIG. 8

EX1004, FIG8.

Adding Goldberg results in the associated image data being received “from the server that manages content data generated during the event.” EX1003, ¶[0116]. Goldberg’s server stores the recorded video of the live event, and that video is displayed to the client “side-by-side” with its synchronized, time-coded transcript. EX1005, ¶¶[0013], [0055], [0060], [0061]; EX1003, ¶[0116]. Because Goldberg’s video comprises a temporal sequence of individual image frames each corresponding to a specific point in time and to a specific portion of the transcript, a POSITA would have understood that images captured from that video feed in the manner Mahajan captures thumbnails would be received from Goldberg’s server in association with

the event's transcript text. EX1005, ¶[0097] (“[e]ach point in the recording is synchronized to the corresponding text in the written transcript”); EX1003, ¶[0116]. In the combination, Goldberg's server manages the event's voice data, transcript text, and associated image content, and supplies that content, including Mahajan's thumbnails, to the client display apparatus. EX1005, ¶¶[0013], [0055], [0073]; EX1003, ¶¶[0116]-[0117].

13. [5.b]/[20.b]

Mahajan discloses displaying such image data in temporal order. Mahajan's thumbnail section 802 displays “a plurality of different thumbnail photographs that represent a portion of the A/V content 104,” and Mahajan describes the thumbnails as sequenced, referring to “the first two,” then “the second and third” thumbnails, confirming that the thumbnails are presented in the order corresponding to their position within the A/V content stream. EX1004, ¶¶[0071], [0073]; EX1003, ¶¶[0118]-[0119]. Because each thumbnail is tied to the topic and time period during which its corresponding subject matter was discussed, and because Mahajan's user interface is, more generally, “updated based on a point in the A/V content stream that is currently being played” as content plays back, a POSITA would have understood Mahajan to display its image content in the order determined by the time information associated with the underlying A/V content. EX1004, ¶¶[0073], [0069]; EX1003, ¶[0119].

Goldberg reinforces this ordering requirement. Goldberg's transcript is displayed "with time coding," and the video recording is displayed "side-by-side" with that time-coded transcript such that "[e]ach point in the recording is synchronized to the corresponding text in the written transcript." EX1005, ¶¶[0073], [0097]; EX1003, ¶[0120]. Because the video itself advances chronologically as it plays, any screenshot captured from that video and displayed to the user appears in the same time order as the synchronized transcript text. EX1003, ¶[0120].

In the combination, Mahajan's display of thumbnail/screenshot images in temporal order alongside subject-matter text (EX1004, ¶¶[0071]-[0073]) would be applied to the event-specific voice, text, and associated image data supplied by Goldberg's server (EX1005, ¶¶[0013], [0055], [0060]-[0061], [0072]-[0073], [0097]), yielding a display in which screenshot image data associated with the text data is displayed in the order according to the time information. EX1003, ¶[0121].

A POSITA would have had a reasonable expectation of success in combining these teachings because both references already display metadata-driven visual content (Mahajan's thumbnails and chapter/subject-matter indicators; Goldberg's synchronized video frames) in temporal alignment with transcript text, and applying one reference's display-ordering technique to the other's captured image content is no more than the predictable application of a known technique to achieve a known and expected result. EX1003, ¶¶[0122]-[0123].

14. [6.a]/[14.a]/[21.a]

Mahajan discloses this element. Mahajan teaches that “the speech recognition results contain an identification as to where they were recognized in the A/V content stream,” and that “[a]ll this is stored in metadata store 114.” EX1004, ¶[0042]. Because Mahajan’s recognition results 400 are the “automatically derived transcript” produced by “analyzer 320” performing “automatic speech recognition” on the A/V content, the “identification” associated with the selected transcript word is an identifier associated with the speech-recognition processing that produced that word from the corresponding underlying audio. EX1004, ¶[0052]; EX1003, ¶¶[0124]-[0125]. When “a user can[] simply highlight a portion of the transcript” and “drag a word from the transcript in field 208 and drop it on screen 200,” the system necessarily obtains, from that selected word (the “first text data”), the identification information that identifies that word in the content stream. EX1004, ¶¶[0041]-[0042]; EX1003, ¶[0125].

Goldberg reinforces this identifier-based mechanism in the context of a synchronized transcript-selection interface. Goldberg discloses that when a client highlights transcript text, “‘OnMouseDown’ passes a **Word ID**, transcript ID, and a timestamp of the first word highlighted to a hidden html form 812,” and “‘OnMouseUp’ passes a Word ID, transcript ID, and a timestamp of the last word highlighted to the hidden html form 812.” EX1005, ¶¶[0124]-[0125] (emphasis

added). Goldberg’s “Word ID” is an identifier associated with the specific transcribed word (i.e., text data generated from voice data through Goldberg’s transcription process, obtained in response to the user’s selection action). EX1005, ¶¶[0115], [0069]; EX1003, ¶[0126]. A POSITA would have understood Goldberg’s Word ID, like Mahajan’s identification information, to be a “content processing identifier for identifying conversion processing that has been applied to generate the first text data from corresponding voice data” because it identifies the specific unit of the transcript—and hence the specific transcription/conversion operation—that produced the selected text from the underlying recorded speech. EX1003, ¶¶[0126]-[0127].

15. [6.b] / [14.b] / [21.b]

Mahajan discloses this element. As set forth above, the “identification” obtained from the selected word is stored in “metadata store 114” together with the corresponding time-alignment information, because “[a]nalyzer 320 also aligns the recognized words with the A/V content 104, such that each word in the transcript is aligned with a specific time point in the A/V content 104.” EX1004, ¶[0052]; EX1003, ¶¶[0128]-[0129]. Mahajan discloses that “player 122 ... locates the spot in the A/V content stream 104 that corresponds to the word in the recognized transcript in field 208” precisely because “the speech recognition results contain an identification as to where they were recognized in the A/V content stream,” and this

identification-to-time correspondence is what allows the player to “find[] the particular place in the A/V content stream 104 that corresponds to the selected word.” EX1004, ¶¶[0042]–[0043]. A POSITA would have understood that the player necessarily uses the selected word’s identification information to find, in metadata store 114, time information associated with that identification, so as to locate the corresponding place in the A/V content stream. EX1003, ¶¶ [0129]–[0130].

16. [6.c]/[14.c]/[21.c]

Mahajan discloses this element for the same reasons as [1.e]. Mahajan’s marker 214 indicates “what part of the A/V content stream is currently being displayed on display 200.” EX1004, ¶[0044]. Because the player, using the time information obtained as discussed in [6.b], “finds the particular place in the A/V content stream 104 that corresponds to the selected word, and begins playing A/V content 104 at that place,” the system necessarily determines a new position for marker 214 on time bar 212 corresponding to that time information. EX1004,¶[0043]; EX1003, ¶¶[0131]–[0132].

Goldberg discloses the same: after obtaining the timestamp, “the JavaScript library 702[] passes the video ID and the time stamp which are embedded in the text to the QUICKTIME player 710,” which then “plays the video, beginning at the spot indicated by the time stamp 706.” EX1005, ¶[0110]. Because Goldberg’s play bar

displays the current playback position (“[t]he client can drag the play bar in the QUICK TIME player to a desired position”) the play bar updates to a position corresponding to the time information once playback moves to the new spot. EX1005, ¶[0115]; EX1003, ¶[0133].

A POSITA implementing Mahajan’s identifier-and-metadata-store lookup mechanism, as confirmed by Goldberg’s own handling of time information, would have understood that the graphical control region’s new display location is determined from the time information obtained via the content processing identifier. EX1003, ¶¶[0134]-[0135].

17. Claim 7

Mahajan discloses that the user’s selection may extend to an arbitrary portion of the displayed transcript, that “a user can then simply highlight a portion of the transcript which the user finds of interest,” and that “[t]his can be done in a number of ways,” including that “the user can right click over a word on the transcript, the user can highlight the transcript, or,... drag a word from the transcript in field 208 and drop it on screen 200.” EX1004, ¶[0041]. The reference to highlighting “a portion of the transcript” confirms that Mahajan’s selection mechanism accommodates a user-chosen span of text which may correspond to a single sentence. EX1003, ¶¶[0136]-[0137].

A POSITA implementing Mahajan's transcript-selection feature would have found it obvious to allow the user to select a complete sentence, the most natural grammatical unit of spoken content, rather than merely a single word, particularly because Mahajan already teaches that the highlighted "portion" is not fixed to any particular length. EX1004, ¶[0041]; EX1003, ¶[0138]. Allowing selection of sentences also prevents accidental selection of the wrong portion of the transcript, as a given word may appear many times during the recording of an event, but a sentence is far less likely to occur at different portions of the transcript. EX1003, ¶[0138].

Goldberg, consistent with Mahajan, discloses that "the client may click on any text in the transcript," and that "[c]licking on a point in the ... transcript automatically locates the corresponding spot in the recording." EX1005, ¶[0110]. The phrase "any text in the transcript" encompasses any selected span of text. EX1003, ¶[0139]. Goldberg's "tagg" feature likewise confirms that Goldberg's system supports selection of arbitrary spans of text: "[t]aggs are created by highlighting a desired portion of text," and "[t]aggs may be of any length." EX1005, ¶[0120]; EX1003, ¶[0139]. A tagg (created by highlighting "a desired portion of text" of "any length,") is associated with "the beginning and ending time codes of the segment," and Goldberg's tagg-creation and navigation-selection transcript-selection mechanisms are materially the same. EX1005, ¶¶[0110], [0120]-[0121]; EX1003, ¶[0139]. A

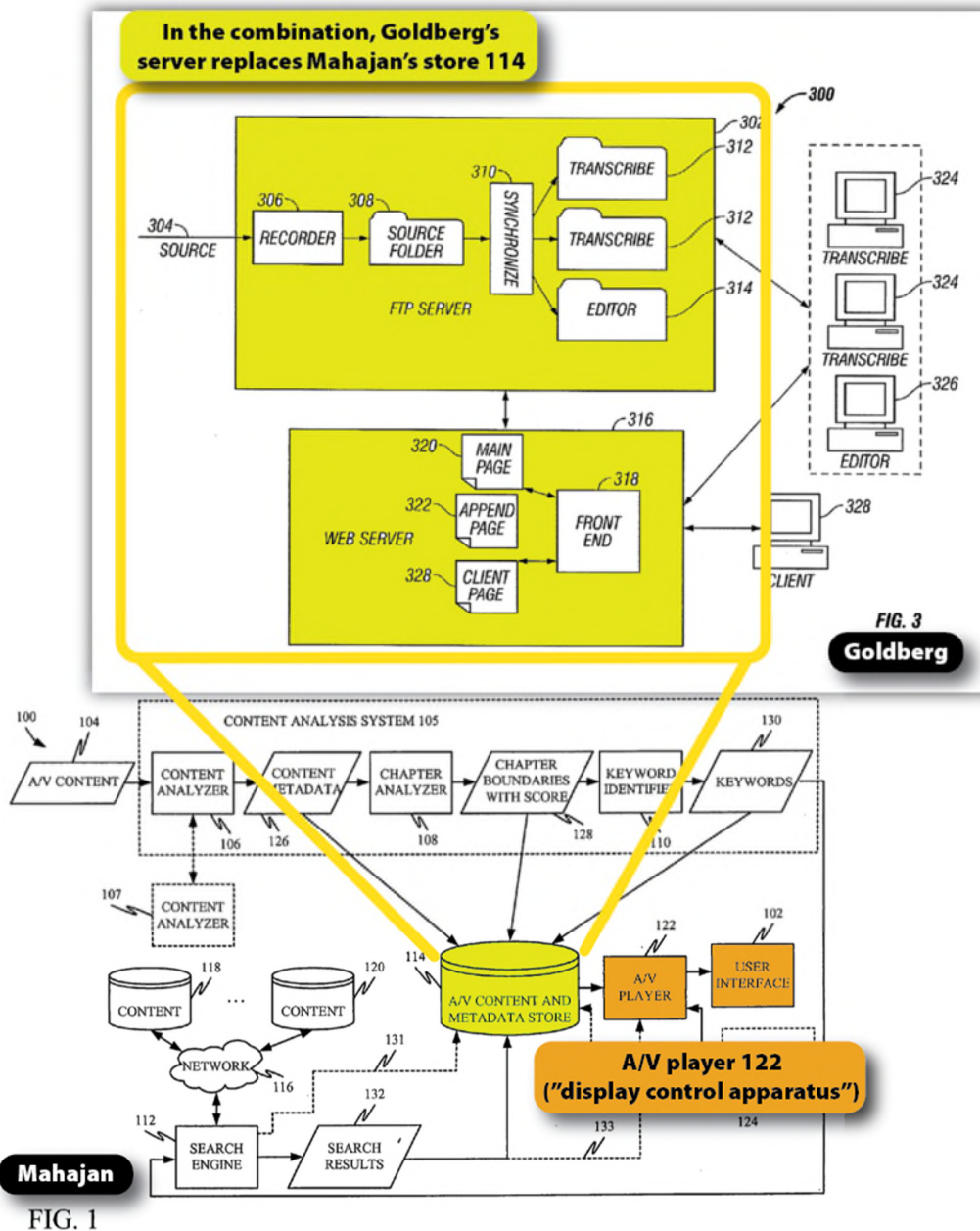
POSITA would have understood that this mechanism supports selection of any span of text—including a single complete sentence of speech. EX1003, ¶[0139].

A POSITA would have had ample motivation and a reasonable expectation of success in implementing sentence-level selection granularity in the combined Mahajan-Goldberg system. EX1003, ¶[0140]. Selecting a complete sentence, rather than an isolated word, provides the user with more linguistically meaningful context when deciding which portion of a transcript to select for navigation purposes, where Goldberg’s server supplies event-specific voice, converted text (as taught by Mahajan), and time information for a live event such as a meeting or conference call. EX1005, ¶¶[0013], [0055], [0060]-[0061]; EX1004, ¶[0052]; EX1003, ¶[0140]. Because both references already disclose arbitrary-length text selection (Mahajan’s “portion of the transcript,” EX1004, ¶0041, and Goldberg’s “any text in the transcript” and tags “of any length,” EX1005, ¶¶[0110], [0120]), allowing selection of one sentence is the predictable application of a known selection mechanism to a known linguistic unit, yielding the predictable result of allowing sentence-level navigation. EX1003, ¶¶[0140]-[0141].

18. [8.a]

Mahajan and Goldberg disclose this element for the reasons discussed above at [1.pre] and [10.pre]/[10.a]: Mahajan’s A/V player 122 constitutes the display control apparatus of claim 1. EX1004, ¶[0023]. The A/V player 122 receives text

and time information from metadata store 114, which the combination implements as Goldberg's server. EX1003, ¶¶[0142]-[0143]. Notably, Mahajan also discloses that content stores 118 and 120 can be accessed via network 116 and that "a wide variety of additional content stores may be available through network 116 as well," which would have provided motivation for modifying Mahajan to use a network connection between metadata store 114 and A/V player 122. EX1004, ¶[0023]; EX1003, ¶[0143]. Goldberg's claims confirm that its server is "communicably connected" to client-side viewing devices: "at least one computational device controlled by a client of said provider [is] communicatively coupled to said provider server." EX1005, claim 18; EX1003, ¶[0143]. A POSITA would have found it obvious to implement Mahajan's store 114 as a server of the type Goldberg discloses. EX1003, ¶¶[0143]-[0144].



EX1005, FIG3, EX1004, FIG1.

19. [8.b]

Goldberg discloses this element. Goldberg's server receives and stores voice data ("source material") recorded during live events together with the resulting

transcript. EX1005, ¶¶[0038]-[0055]; EX1003, ¶¶[0145]-[0146]. This is later “reconstituted by displaying the words stored in the database in the correct order” using “time codes and the written words in the transcript[] associated with each other in a one-to-one relationship.” EX1005, ¶[0099]. Goldberg’s server memory thus stores, for each event, the voice data, the transcript, and the time information linking them. EX1003, ¶[0146]].

Mahajan reinforces this element: “A/V content 104” and its transcript metadata are stored together in “metadata store 114,” with each word “aligned with a specific time point.” EX1004, ¶¶[0023], [0033], [0052]; EX1003, ¶[0147].

A POSITA would have found it obvious to combine Goldberg’s event-based storage architecture with Mahajan’s voice-to-transcript functionality, because both address the same problem of managing time-synchronized voice and text data for later review, and Goldberg’s event-focused architecture improves useability by enabling a user of Mahajan’s system to easily located and retrieve content from a specified event. EX1003, ¶¶[0148]-[0149].

20. [8.c]

Goldberg discloses this element. Goldberg’s “web server 316” transmits the event’s stored content to the client: “[t]he web page created for the client consists of the transcript, with time coding, side by side with a column containing a summary.” EX1005, ¶¶[0061], [0072]–[0073]; EX1003, ¶¶[0150]-[0151]. The server-side

circuitry that generates “page 328 wherein the client is able to view the transcript” constitutes “another circuitry” that transmits the stored event information to the display control apparatus. EX1005, ¶[0100]; EX1003, ¶[0151].

Mahajan similarly discloses circuitry that transmits event information to the display control apparatus. Mahajan discloses that “A/V player 122 [] accesses the A/V content and its associated metadata” from “store 114” over “network 116,” necessarily requiring transmission circuitry associated with store 114 (or the server hosting it) that sends the stored A/V content and metadata to player 122 for display. EX1004, ¶¶[0023], [0033]; EX1003, ¶[0152].

In the combination, Goldberg’s web server 316/client page 328, which transmits the event’s time-coded voice data and transcript text to the client’s viewer upon request, performs this transmission, sending the stored event information to Mahajan’s player 122 (the display control apparatus). EX1005, ¶¶[0061], [0072]-[0073]; EX1003, ¶[0153].

A POSITA would have had a reasonable expectation of success because both references already contemplate an architecture in which stored event content is transmitted to a client-side viewer, making this combination a predictable application of known elements according to their established functions. EX1003, ¶¶[0154]-[0155].

21. Claim 9

Goldberg discloses this element. Goldberg's server receives voice data collected during a live event from a communication terminal. EX1003, ¶¶[0156]-[0157]. Goldberg discloses that "an FTP (file transfer protocol) server 302 receives source material 304 in the form of audio input from any source," and that "[a]udio may arrive via the Internet using VoiceOverIP (VOIP), various streaming media, or SHOUTCAST ... software," or "may also arrive via landline telephone or cell phone connection." EX1005, ¶[0157]. A telephone, cell phone, VOIP device, or streaming-media source that captures and transmits, via FTP, the audio of a live event to the server is a "communication terminal communicably connected with the server and configured to transmit voice data collected during an event to the server." EX1003, ¶[0157]. Goldberg also recites that its server receives, "a transmission of a speech signal from a live event." EX1005, claim 4. Goldberg lists events such as "financial conference calls," "basic conference calls," "conventions and meetings," "panel discussion[s]," and "press conferences," confirming that the communication terminal transmitting the voice data is collected during and at the event itself, and is transmitted to the server. EX1005, ¶¶[0039]-[0052]; EX1003, ¶[0157].

As combined, this communication terminal supplies the event voice data to Goldberg's server, which (as established with respect to claim 8) is combined with Mahajan's content analyzer 320 and A/V player 122 to convert that voice data into

time-aligned text which is displayed alongside a repositionable graphical control region. EX1005, ¶[0055]; EX1004, ¶¶[0043]–[0045], [0052]; EX1003, ¶[0158]. A POSITA implementing the combined system would have found it obvious to retain Goldberg’s communication terminal (such as a telephone, cell phone, or VOIP-enabled device present at the meeting or conference) as the source of the voice data received by the server, as voice data must come from somewhere, and in a digital system a communication terminal is the natural source. EX1003, ¶¶[0158]–[0159].

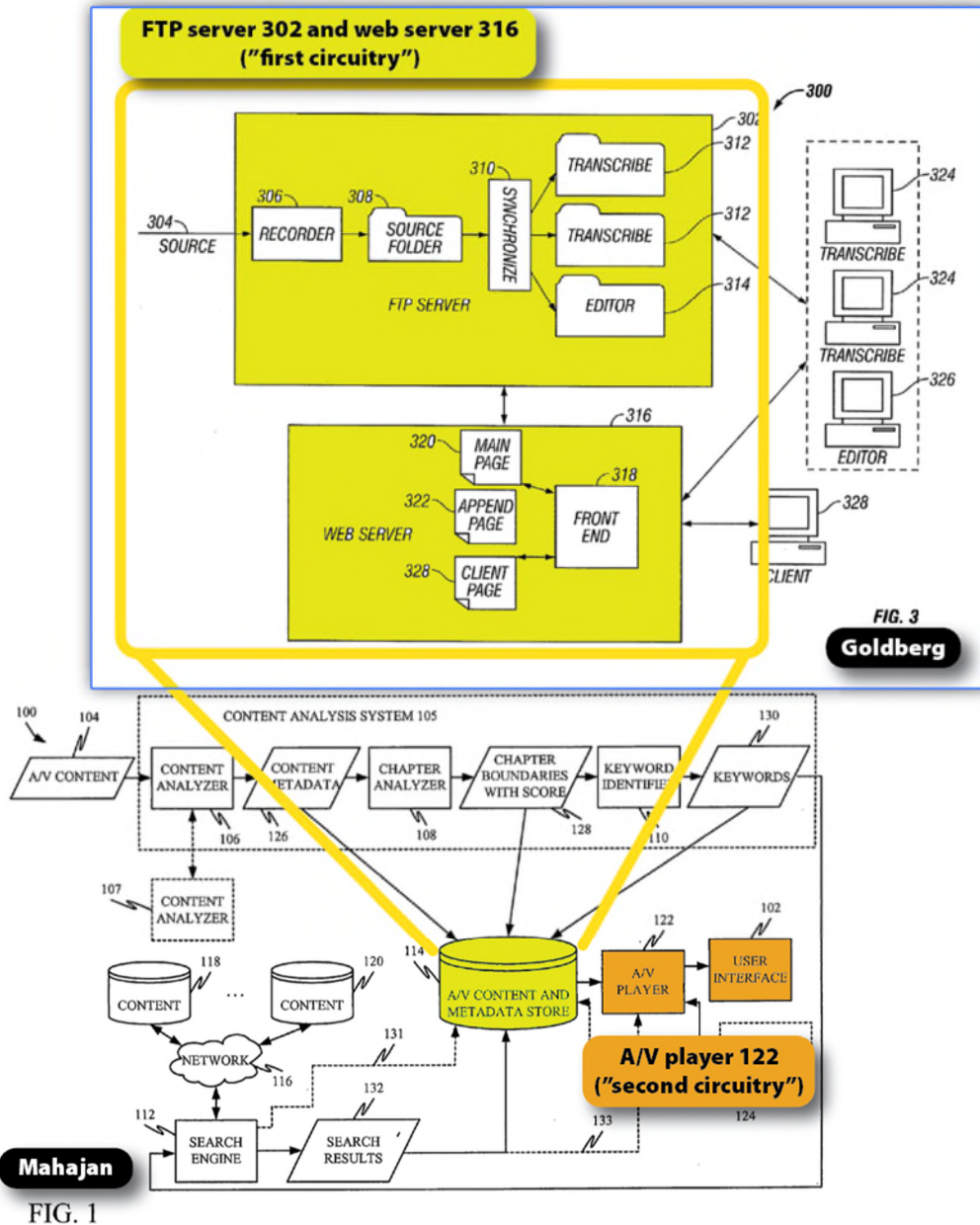
22. [11.a]

Goldberg discloses this element. As established at [8.c], Goldberg’s server (first circuitry) generates and transmits to the client (second circuitry) the data needed to render the client’s screen: its web server 316 delivers “a dedicated web page containing a list of the client’s projects,” where “[t]he web page created for the client consists of the transcript, with time coding, side by side with a column containing a summary.” EX1005, ¶¶[0061], [0072]–[0073]; EX1003, ¶¶[0160]–[0161]. The contents of this web page, the stored voice data, transcript, and summary data, are “screen data generated based on the information on the event stored in the memory” and transmitted to the client. EX1005, ¶¶[0055], [0099], EX1003, ¶[0161].

Mahajan likewise discloses that A/V player 122 receives, from a store, the content and metadata used to generate “metadata driven displays based on the

automatically generated metadata.” EX1004, ¶[0033]. That content—used to generate user interface 102, including text box 208 and time bar 212, is transmitted from the source (store 114, or in the combination, Goldberg’s server) to player 122 for rendering. EX1004, ¶¶[0038], [0044]; EX1003, ¶[0162].

In the combination, Goldberg’s server (first circuitry) transmits this screen-rendering information to Mahajan’s player 122 (second circuitry). EX1003, ¶¶[0163]-[0164].



EX1005, FIG3; EX1004, FIG1.

23. [11.b]

Mahajan discloses this element. Using content received from store 114 (or, in the combination, Goldberg's server), player 122 generates "user interface 102" as

discussed at [1.b] and [1.c]: text box 208 displays the transcript “arranged in the order according to the time information,” and time bar 212 with marker 214 constitutes “the graphical control region that sets playback position in a total playback time of the voice data.” EX1004, ¶¶[0036]–[0045], [0052]; EX1003, ¶¶[0165]–[0166].

Goldberg’s client’s synchronized-viewer screen, including both the time-coded transcript (“reconstituted by displaying the words stored in the database in the correct order,”) and the QUICKTIME play bar (which sets and displays playback position within the total duration of the content), is similar to Mahajan’s arrangement. EX1005, ¶¶[0073], [0099], [0115]; EX1003, ¶[0167].

In the combination, player 122 (second circuitry) displays a screen based on screen data transmitted from Goldberg’s server (first circuitry), including both elements. EX1003, ¶[0168].

A POSITA would have had a reasonable expectation of success because both references already disclose displaying time-ordered transcript text alongside a graphical playback-position indicator; combining Goldberg’s server-side screen-generation architecture with Mahajan’s client-side rendering functionality is the predictable application of known techniques to the expected result. EX1003, ¶¶[0169]–[0170].

24. Claim 12

The first two sub-steps of claim 12—obtaining a content processing identifier from the selected first text data—are disclosed by Mahajan and Goldberg for the reasons discussed in [6.a]. EX1003, ¶¶[0171]-[0172]. Mahajan discloses that “the speech recognition results contain an identification as to where they were recognized in the A/V content stream,” EX1004, ¶0042; EX1003, ¶[0172].

Claim 12 requires that the second circuitry “transmit the content processing identifier to the first circuitry to obtain the time information.” EX1001, claim 12. Mahajan discloses that the “identification” of location (i.e., time offset) within the A/V content stream is stored in store 114. EX1004, ¶0042; EX1003, ¶[0173]. Mahajan describes the storage location of this identification and then recites “[t]herefore, player 122 finds the particular place in the A/V content stream 104 that corresponds to the selected word.” EX1004, ¶¶0042, 0043. A POSITA would thus understand that Mahajan discloses that player 122 (the “second circuitry”) must query the store 114 (or, in the combined system, Goldberg’s server) using the identification to obtain the time information so that it may find the corresponding A/V content stream location. EX1003, ¶[0173].

In the combination, Goldberg’s server (“first circuitry”) stores the event’s voice, converted text, identifier, and time information in its database, (EX1005, ¶¶[0055], [0099]), and Mahajan’s player 122 (“second circuitry”) displays the

transcript, receives the user's text selection, and transmits the corresponding identifier to server in order to retrieve the corresponding time information (EX1004, ¶¶[0038], [0041]-[0044]). EX1003, ¶[0174].

Once the time information is obtained, Mahajan's player 122 determines the second display location of marker 214 based on that time information, for the same reasons discussed in [1.e]/[6.c]. EX1004, ¶¶[0042]-[0044]; EX1003, ¶¶[0175]-[0176].

25. Claim 13/19

Mahajan discloses this element. Mahajan's player 122 (the "second circuitry") receives raw "A/V content" and its "associated metadata" (i.e., the information on the event, including the time-coded transcript text) and itself generates ("generating user interface 102") the display screen from that information—including text box 208 displaying the transcript in time order, as discussed at [1.b], and marker 214 on time bar 212, as discussed at [1.c]. EX1004, ¶¶[0023], [0033], [0038], [0044], [0052]; EX1003, ¶¶[0177]-[0178].

Goldberg uses a similar approach: the client's browser executes JavaScript that dynamically updates the displayed screen from server-supplied information, rather than displaying an image rendered by the server. EX1003, ¶[0179]. Goldberg's "Video-Text JSQT Engine" "interprets playtime and selects ... texts ... corresponding to that time in video" and highlights and focuses that text on screen,

and the client replaces displayed project elements “without the browser being reloaded” using client-side scripting. EX1005, ¶[0108], FIG9; EX1003, ¶[0179].

In the combination, Goldberg’s server (first circuitry) transmits raw event information—the voice, converted text, and time information—to Mahajan’s player 122 (second circuitry), which generates the display screen from that information, exactly as Mahajan discloses. EX1004, ¶¶[0023], [0033], [0052], [0055]; EX1003, ¶[0180].

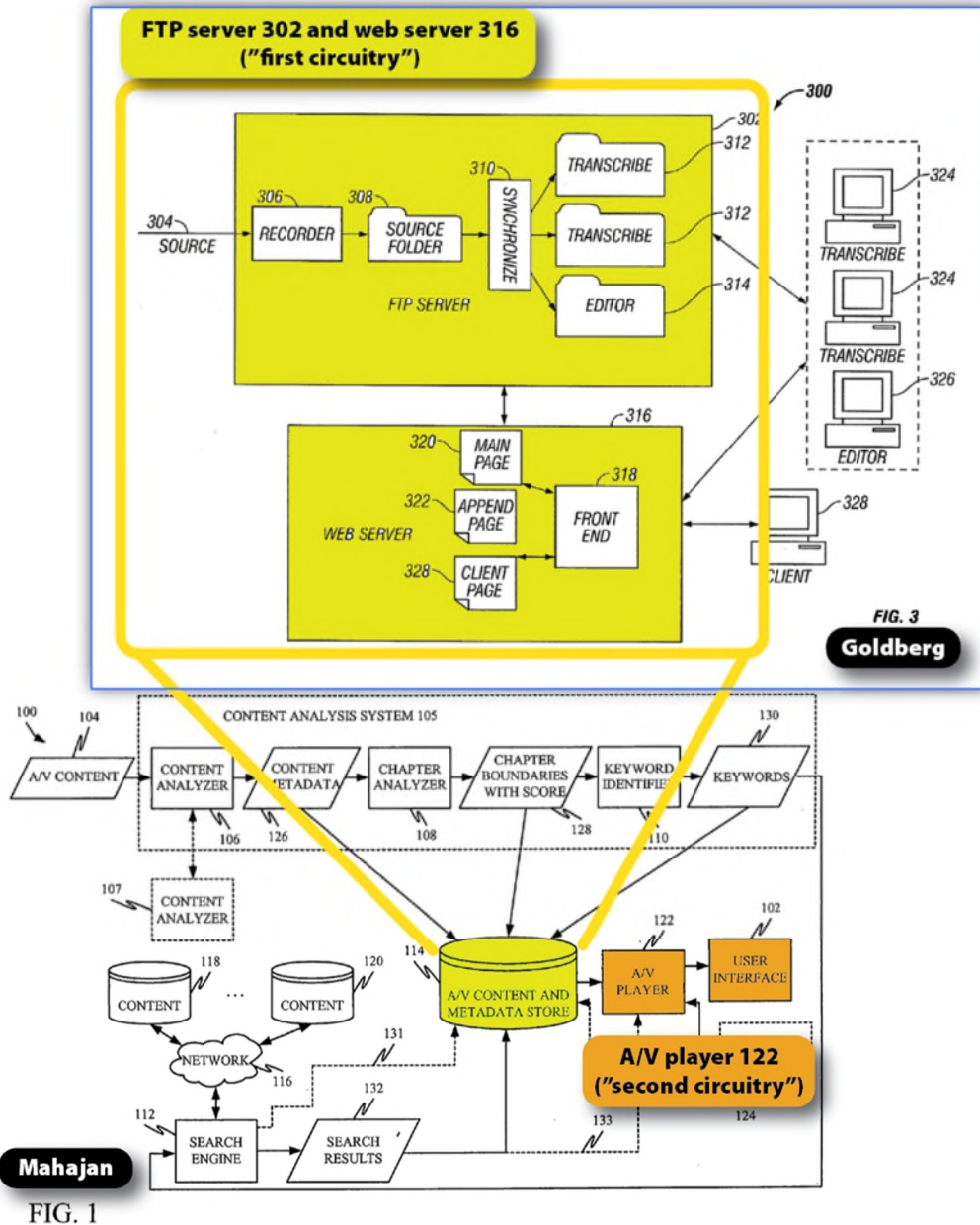
A POSITA would have had a reasonable expectation of success because client-side screen rendering was a well-understood alternative to server-side rendering in multimedia display systems, and Mahajan’s own disclosure already teaches this division of responsibility. EX1004, ¶[0023]; EX1003, ¶¶[0181]-[0182].

26. Claim 15

Claim 15’s architectural arrangement is rendered obvious by the same combination of Mahajan and Goldberg discussed in [8.a]. EX1003, ¶¶[0183]-[0184].

As discussed at [8.a], Goldberg’s server (FTP server 302 and web server 316), which transmit the event’s stored transcript and voice data, is the claimed “server” and houses the “first circuitry” responsible for storing (per [8.b]) and transmitting (per [8.c]) the event’s voice data, text data, and time information. EX1005, ¶¶[0055], [0061], [0072]–[0073], claim 18; EX1003, ¶[0185].

Mahajan discloses “a display control apparatus communicably connected with the server, wherein the display control apparatus includes the second circuitry.” As discussed at [1.pre] and Claim 16 below, Mahajan’s player 122 “accesses the A/V content and its associated metadata and presents the A/V content to the user by generating user interface 102,” constituting the display control apparatus that is “communicably connected” to “store 114”—which the combination implements as Goldberg’s server. EX1004, ¶¶[0023], [0033]; EX1003, ¶[0186]. Player 122 (“second circuitry”) performs the display, determination, and repositioning functions discussed at [1.b]–[1.f]. EX1004, ¶¶[0038], [0042]–[0045], [0052]; EX1003, ¶[0186].



EX1005, FIG3; EX1004, FIG1.

A POSITA combining Mahajan and Goldberg would have arrived at this two-part architecture as the natural and predictable result of implementing Mahajan's client-side display functionality on top of Goldberg's server-side event-management

functionality, for the same reasons discussed in the Motivation to Combine section and claims 8/10. EX1003, ¶¶[0187]-[0188].

27. Claim 16

Mahajan discloses that A/V player 122 (using the “second circuitry”) performs each claimed function, as discussed at [1.b]–[1.f]: text box 208 displays the transcript in time order; marker 214 on time bar 212 constitutes the graphical control region at the first display location; and upon text selection, player 122 locates and begins playing from the corresponding time, determining and displaying a new (second) display location for marker 214. EX1004, ¶¶[0038], [0042]–[0045]; EX1003, ¶¶[0189]-[0191]. Because A/V player 122 performs these functions using processing circuitry on the client, a POSITA would have understood this circuitry to be the “second circuitry of the display control apparatus,” distinct from the server-side circuitry (Goldberg’s server) that manages and transmits the event content. EX1004, ¶[0023]; EX1003, ¶[0191].

Goldberg confirms the same allocation of circuitry: its server (first circuitry) stores and transmits the event’s voice data and transcript, while the client (second circuitry) displays the synchronized video and transcript and, upon a text click, “locates the corresponding spot in the recording and plays the recording from this spot forward.” EX1005, ¶¶[0055], [0072]-[0073], [0110]; EX1003, ¶[0192]. Because Goldberg’s play bar displays the current playback position, this same client-

side circuitry necessarily determines and displays the play bar's new position.

EX1005, ¶[0115]; EX1003, ¶[0192].

A POSITA combining Mahajan and Goldberg would have implemented Mahajan's display, text-ordering, and repositioning functions using the second circuitry, operating on the voice data, text data, and time information supplied by Goldberg's first circuitry—the predictable combination of known client-side and server-side functionality. EX1003, ¶¶[0193]-[0194].

B. Ground 2: Mahajan in View of Goldberg and Further in view of Wei Renders Obvious Claims 1-22

Mahajan in view of Goldberg renders obvious every limitation of claims 1–22 for the reasons in Ground 1. Wei is added only to expressly confirm what Mahajan and Goldberg already teach—particularly the transcript-synchronized navigation, progress-bar repositioning, phrase-level selection, thumbnail association, and client/server data exchange relevant to limitations [1.e]/[1.f] and their dependents, claim 7, claim 12, and claims 5/20. EX1003, ¶¶[0195]-[0196]. Wei does not change the basic combination; it provides an additional, direct confirmation of the same known features. All other limitations are obvious for the reasons stated in Ground 1.

1. Motivation to Combine

Same Field of Endeavor. Wei, like Mahajan and Goldberg, is directed to synchronizing a time-coded transcript of speech with a media player interface to

enable transcript-driven navigation of recorded audio/video content. EX1003, ¶[0197]. Wei discloses “a searchable transcript of the dialog of the media program” presented “concurrently with a media program player,” such that “the media program player, the histogram 700, and the transcript may be synchronized so the selection of a navigation control of one of the elements is reflected in the other elements.” EX1006, ¶¶[0097], [0125]. A POSITA designing a transcript-navigation system would have looked to Wei alongside Mahajan and Goldberg, as all three address the identical problem of using a time-coded transcript to navigate the underlying recording. EX1003, ¶[0197].

Wei Confirms and Reinforces Features That Mahajan and Goldberg Teach. Wei discloses features similar to Mahajan and Goldberg: a slider control that the user can “grab ... and move ... to begin play at an earlier or later point in the media program,” per-phrase transcript segmentation and thumbnail generation, EX1006, ¶¶[0127], [0124], [0129], claim 10. A POSITA would have understood Wei’s explicit teachings to confirm that the sentence/phrase-level selection, thumbnail association, and progress-bar repositioning already taught by Mahajan and Goldberg were well-understood implementations in the same field. EX1003, ¶[0198].

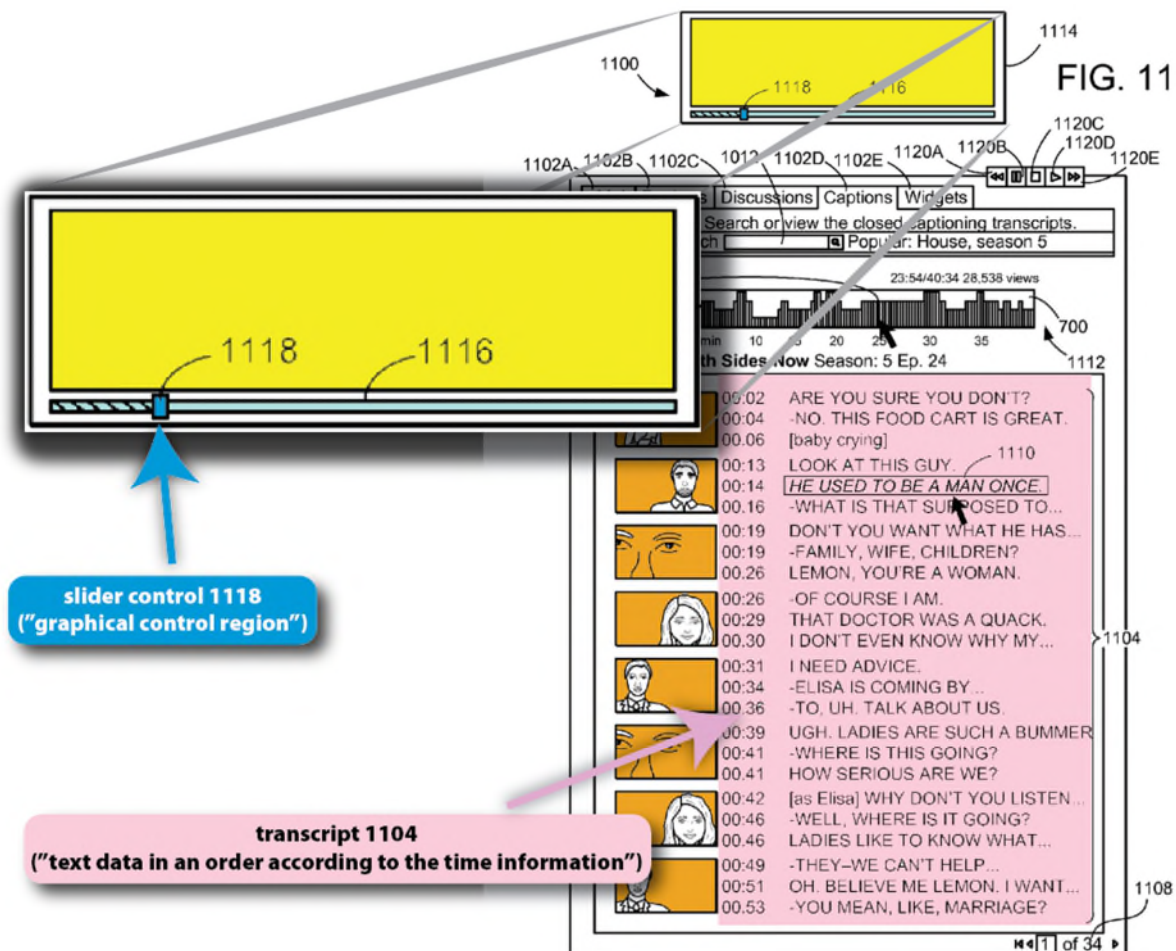
Wei Reinforces Data Transmissions Between Client and Server. Wei discloses that transcript- and navigation-related information can be transmitted

between the client and server in either direction, stating that “the client computer 102 may store only sufficient user interface data to permit display of the user interface elements to the user, and rely on requests to the web server to obtain data in response to user input,” or alternatively that “the user interface data transmitted to the client computer 102 may include sufficient data to allow synchronization between all histogram and transcript for all portions of the media program without the need for further data from the web server.” EX1006, ¶[0139]; EX1003, ¶[0199].

Predictable Combination. The combination is the predictable use of known prior art elements according to known methods to yield predictable results. *KSR Int'l Co. v. Teleflex Inc.*, 550 U.S. 398, 415–421 (2007). Wei’s explicit slider-control jumping, phrase-segmentation, thumbnail, and client/server data-allocation teachings would be applied to the Mahajan/Goldberg combination’s converted text and event-content-server architecture as a straightforward confirmation and refinement of features already rendered obvious in Ground 1. EX1003, ¶[0200]. A POSITA would have had a reasonable expectation of success because all three references implement the same fundamental transcript-to-media synchronization concept using conventional, interchangeable client-server design choices. EX1003, ¶[0201].

2. [1.e]/[10.d]/[17.e]/[22.e]

Wei discloses a graphical playback-position control—the “slider control 1118”—that is positioned on a “progress bar 1116 which indicates which portion of the media program is currently being played.” EX1006, ¶[0127]. Wei discloses that this control is repositioned in direct response to navigating the transcript. EX1003, ¶¶[0202]-[0203].



EX1006, FIG11.

Wei's user interface presents "a transcript 1104 of the media program," which "includes the phrases of the dialog and an adjacent time stamp, indicating when the phrase was uttered in the media program." EX1006, ¶[0129]. Wei further discloses that the transcript is synchronized with, and can be used to navigate, the media program: The user can navigate the transcript using controls 1108, in which case "the histogram 700 and the transcript are synchronized so that the same time interval of the media program is presented on both," EX1006, ¶¶[0130], [0131]; EX1003, ¶[0204]. The histogram and the transcript are designed to "allow the user 132 to find a known portion of the media program quickly." EX1006, ¶[097].

Critically, Wei discloses that the position of the slider control 1118 (the graphical control region) on progress bar 1116 is determined based on time information corresponding to the user's selected media program segment. EX1003, ¶[0205]. Wei discloses that "the current playback point is visualized as a cursor such as the slider control 1118 depicted in FIG. 11 that can be moved forward and backwards to navigate through different locations on the progress bar 1116, each of which represent a different temporal location in the media program." EX1006, ¶[0104]. Wei discloses that navigating causes the media program player to begin play at the corresponding temporal location, such that the media player identifies "a location that is temporally behind [or ahead of] the current playback point," exactly as is required to determine a new display location for the slider control corresponding

to the selected content's time information. EX1006, ¶[0104], [0130], [0131]; EX1003, ¶[0205]. A POSITA would have understood that Wei's disclosure of synchronizing "the media program player, the histogram ..., and the transcript ... so [that] the selection of a navigation control of one of the elements is reflected in the other elements," requires the system to determine—by reference to the time stamp associated with the selected transcript phrase—a new position along progress bar 1116 for slider control 1118 that is different from its prior position and that corresponds to that time stamp. EX1006, ¶¶[0125], [0129]; EX1003, ¶¶[0205]-[0206].

3. [1.f]/[10.e]/[17.f]/[22.f]

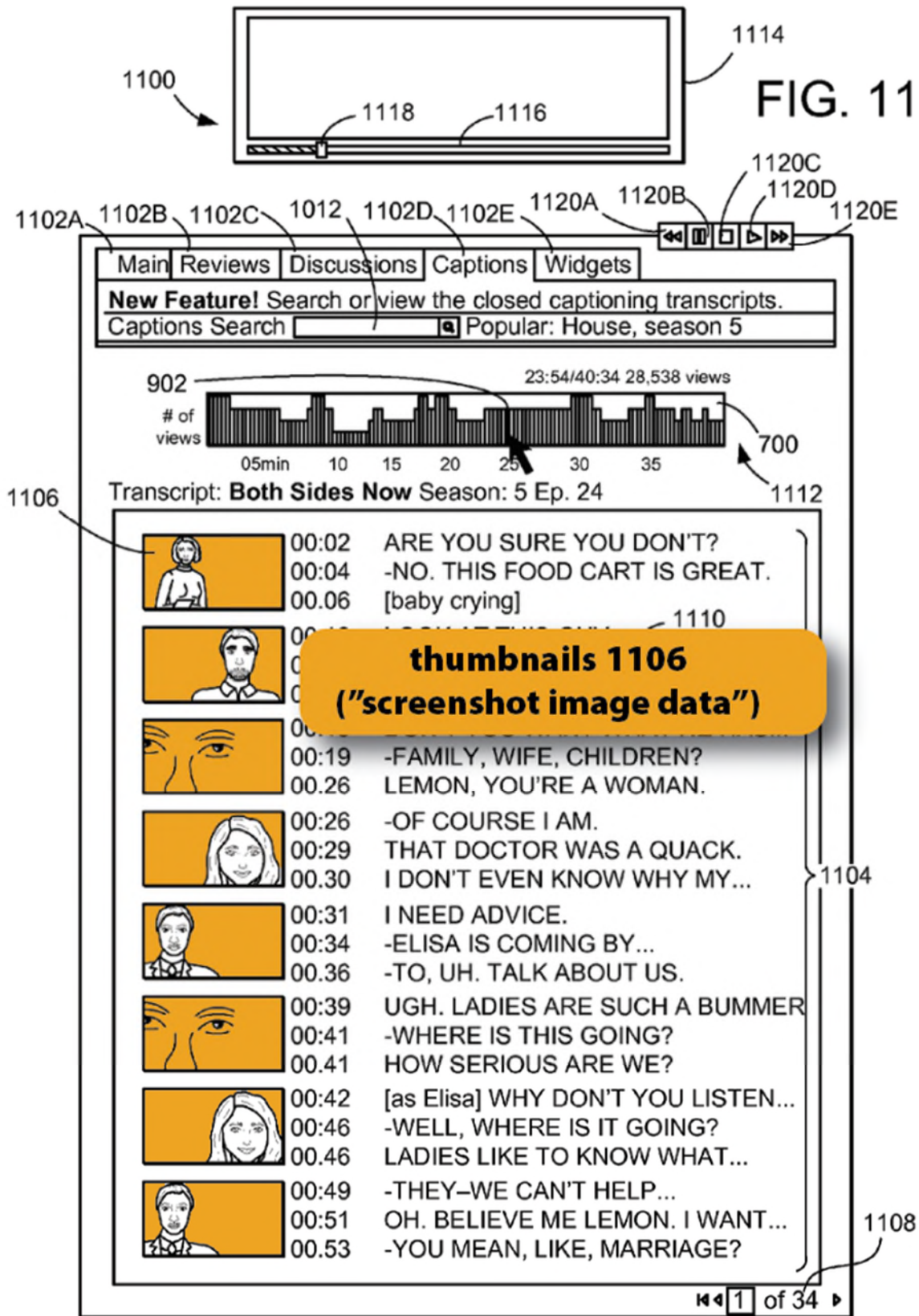
Wei discloses this element by disclosing the mechanism described above. Wei recites that "user interface elements such as the media program player, the histogram 700, and the transcript may be synchronized so the selection of a navigation control of one of the elements is reflected in the other elements." EX1006, ¶[0125]. The slider control 1118 is part of the media player window 1114. EX1006, ¶[0127], FIG13; EX1003, ¶¶[0207]-[0208]. Wei therefore discloses synchronizing three location-indicating controls (slider control, histogram, and transcript) so that a change in one is reflected in the others. EX1003, ¶[0208].

A POSITA would further have understood that if multiple controls control a single parameter, they should be kept in synch in order to prevent user confusion.

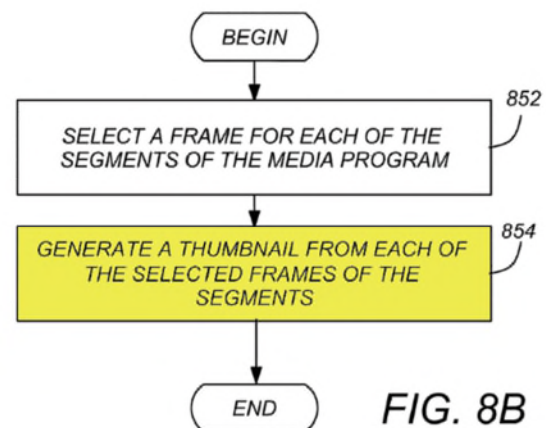
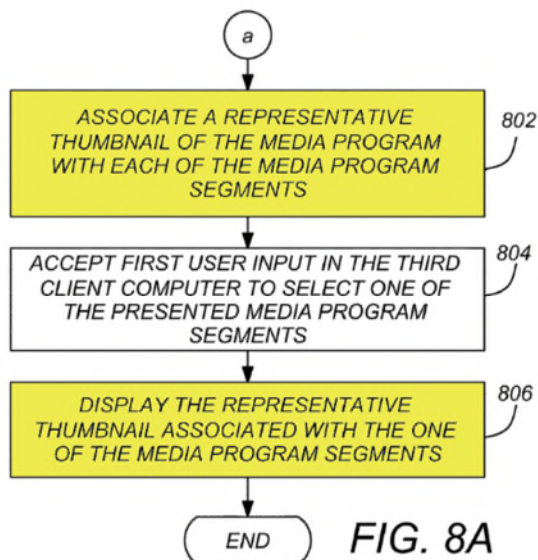
EX1003, ¶[0209]. But even were this not the case, Wei discloses that repositioning a graphical playback-position control to a location corresponding to selected transcript text's time information was known in this field. EX1006, ¶¶[0104], [0122], [0124]–[0131]; EX1003, ¶[0209]. Applying this teaching to Mahajan's marker 214/time bar 212 or Goldberg's play bar, each of which performs the identical underlying function, would have been the predictable application of a known technique to a substantially identical element. EX1004, ¶[0044]; EX1005, ¶[0115]; EX1003, ¶[0209]. The combination thus renders obvious limitations [1.e]–[1.f]. EX1003, ¶¶[0209]–[0210].

4. [5.a]/[20.a]

Wei's transcript display presents “thumbnails 1106” that are “associated with the phrases” of the transcript and displayed “adjacent the phrase.” EX1006, ¶[0129]. Wei discloses these thumbnails are screenshots: “[t]he thumbnails 1106 may be generated manually, or by software which **captures a frame** temporally **associated with the phrase of the transcript** 1104.” EX1006, ¶[0129] (emphasis added); EX1003, ¶¶[0211]–[0212]. This discloses a “screenshot image data associated with the text data,” because Wei's thumbnail is generated precisely by capturing a video frame—i.e., taking a screenshot— that is associated with a specific phrase of the displayed transcript text. EX1003, ¶[0212].



Wei recites this association: “[a]ssociat[ing] a representative thumbnail of the media program with each of the media program segments,” (EX1006, FIG8A step 802), and generating that thumbnail by “select[ing] a frame for each of the segments of the media program” and “generat[ing] a thumbnail from each of the selected frames of the segments” (EX1006, FIG8B step 854). EX1003, ¶[0213]. Because each transcript phrase corresponds to a segment of the media program, and each segment has an associated thumbnail generated from a captured frame, Wei discloses screenshot image data that is directly associated with specific text data. EX1006, ¶[0129], FIG. 8A/8B; EX1003, ¶[0213].



Regarding the requirement that this image data be received “from the server that manages content data generated during the event,” Wei’s architecture discloses that the web server generates and supplies this thumbnail data to the client. EX1003, ¶[0214]. Wei discloses that “the web server 406 identifies a thumbnail that represents each of the segments of the media program,” and that “the user interface data is transmitted to the client computer 102 where it is displayed,” including “the requested portions of the textural transcript” together with the associated thumbnails. EX1006, ¶¶[0118], [0125].

In the combination, Goldberg’s server (Ground 1, EX1005, ¶¶[0013], [0055]) supplies this same type of frame-captured thumbnail image data, generated using Wei’s disclosed technique, in association with the event’s transcript text. EX1003, ¶¶[0215]-[0216].

5. [5.b]/[20.b]

Wei’s thumbnails are displayed “adjacent the phrase” of the transcript to which each is temporally associated, and because the transcript itself is displayed with “the phrases of the dialog and an adjacent time stamp,” the thumbnails necessarily appear in the same chronological sequence as the transcript text they accompany. EX1006, ¶[0129]; EX1003, ¶[0217]-[0218].

A POSITA would have looked to Wei’s express disclosure of frame-captured thumbnail generation and phrase-level association as direct confirmation that

screenshot image data tied to transcript text and displayed in chronological order was a known, documented technique in this field. EX1006, ¶¶[0118]–[0121], [0129], FIGs. 8A/8B; EX1003, ¶[0219]. Applying Wei’s frame-capture technique to the Mahajan/Goldberg combination would have been the predictable substitution of a known technique for Mahajan’s more generalized thumbnail disclosure, yielding the expected result. EX1003, ¶[0219]. The combination thus renders obvious claims 5/20. EX1003, ¶[0220].

6. Claim 7

Wei’s claim 10 recites: “[t]he method of claim 1, wherein the portion comprises a phrase.” EX1006, claim 10. Wei discloses that this “phrase” is the fundamental unit into which Wei’s transcript is divided: “a textual transcript of the dialog is generated. The transcript includes a plurality of portions, each of which is associated with a segment of the media program. In one embodiment, the portions are defined by closed captioning phrases.” EX1006, ¶[0124]. Wei further discloses that “the transcript 1104 includes the phrases of the dialog and an adjacent time stamp, indicating when the phrase was uttered in the media program,” and that selecting a segment of the histogram causes “[t]he portion 1110 of the transcript 1104 associated with that particular portion of the media program” to be “highlighted,” where “[i]f the selected bar 902 represents multiple phrases of the transcript 1104, those multiple phrases may be highlighted.” EX1006, ¶¶[0129],

[0130]; EX1003, ¶¶[0221]-[0222]. A POSITA would have understood a “closed captioning phrase,” as used in Wei, to correspond to a complete grammatical utterance—typically a full sentence or an independent clause of speech—because closed captioning conventions segment spoken dialog into discrete caption units that are generally coextensive with sentences or clauses of speech in order to remain readable and synchronized with the corresponding audio. EX1003, ¶[0222]. Wei’s disclosure that its transcript portions/phrases are “associated with a segment of the media program,” and that each such portion carries “an adjacent time stamp” corresponding to when it “was uttered,” confirms that Wei’s selectable transcript unit is a complete spoken utterance—i.e., a sentence—rather than an isolated word. EX1006, ¶¶[0124], [0129]; EX1003, ¶[0222].

Wei’s disclosure of user interaction with these phrase-level units confirms that they are independently selectable for navigation purposes, consistent with the “first text data.” EX1003, ¶[0223]. Wei discloses that “the user 132 can navigate to different portions of the transcript 1104 by selecting other pages, using controls 1108. Or, the user 132 can navigate to other portions of the transcript 1104 by selecting a bar 902 in the histogram 700.” EX1006, ¶[0130]. Because each bar of Wei’s histogram represents one or more phrases of the transcript, and selecting that bar highlights the “portion” (phrase) of the transcript associated with the selected segment, a POSITA would have understood that the transcript itself is likewise

navigable and selectable at the phrase level directly, exactly as claim 7 requires with respect to “first text data” corresponding to “one sentence of speech.” EX1006, ¶[0130]; EX1003, ¶[0223]. Wei’s claim 8 further recites direct phrase-level user selection. EX1006, claim 8.

A POSITA implementing the Mahajan/Goldberg combination’s transcript-selection functionality would have looked to Wei’s express disclosure of phrase-based transcript segmentation as direct confirmation that segmenting and selecting transcript text at the level of a complete spoken utterance—rather than at the level of an individual word—was a known and expressly claimed technique in this field. EX1006, ¶¶[0124], [0129]–[0130], claims 8, 10; EX1003, ¶[0224]. Applying Wei’s phrase-based segmentation to the text-highlighting mechanisms already disclosed by Mahajan (“highlight a portion of the transcript,” EX1004, ¶[0041]) and Goldberg (“click on any text in the transcript,” and tags “of any length,” EX1005, ¶¶[0110], [0120]) would have been the predictable application of a known convention to the known selection mechanisms of the combination, yielding the expected result that the selectable “first text data” corresponds to one sentence of speech. EX1003, ¶[0224]. The combination thus renders obvious claim 7. EX1003, ¶[0225].

7. Claim 12

Wei expressly discloses that “the client computer 102 may store only sufficient user interface data to permit display of the user interface elements to the

user, and rely on requests to the web server to obtain data in response to user input.”

EX1006, ¶[0139]. This discloses a client that, upon a user selection, transmits a request to the server and obtains data—including time information needed for navigation—in response. EX1003, ¶¶[0226]-[0227]. Wei’s alternative disclosure that “the user interface data transmitted to the client computer 102 may include sufficient data to allow synchronization ... without the need for further data from the web server,” confirms that a POSITA would have recognized the client-to-server round trip of claim 12 as one of two known, interchangeable design choices. EX1006, ¶[0139]; EX1003, ¶[0227]. Wei’s “beacon service,” which transmits client-side selection information to “a module in the web server 406” that then “transmits the histogram 700 to the client computer 102 for presentation,” further confirms this send-a-request/receive-a-response pattern for obtaining segment- and time-related information. EX1006, ¶¶[0109], [0117]; EX1003, ¶[0227].

The remaining sub-elements—obtaining a content processing identifier from the selected text and determining the second display location from the resulting time information—are disclosed for the reasons given in Ground 1’s claim 6 analysis. This is also reinforced by Wei’s disclosure that each transcript phrase carries “an adjacent time stamp, indicating when the phrase was uttered,” and that the obtained time information is used to reposition “the current playback point, ... visualized as ... the slider control 1118 ... [that] can be moved forwards and backwards to navigate

Patent No. 11,289,093

Petition for *Inter Partes* Review

through different locations on the progress bar 1116, each of which represent a different temporal location in the media program.” EX1006, ¶¶[0129], [0104]; EX1003, ¶[0228].

A POSITA implementing the Mahajan/Goldberg identifier-based lookup would have looked to Wei’s explicit disclosure of a client that “rel[ies] on requests to the web server to obtain data in response to user input,” as confirmation that transmitting a selection-derived identifier to the server and receiving corresponding time information in return was a known, conventional architecture in this field. EX1006, ¶[0139]; EX1003, ¶[0229]. Applying Wei’s request-and-response mechanism to Goldberg’s server (first circuitry) and Mahajan’s player (second circuitry) would have been the predictable use of a known technique to achieve the expected result. EX1003, ¶[0229]. The combination of Mahajan, Goldberg, and Wei thus renders obvious claim 12. EX1003, ¶[0230].

IX. CONCLUSION

Petitioner respectfully requests institution of IPR and cancellation of all claims of the ’093 Patent.

X. MANDATORY NOTICES PURSUANT TO 37 C.F.R. §42.8

A. Real Parties-In-Interest under 37 C.F.R. §42.8(b)(1).

The real party-in-interest is Zoom Communications, Inc. No other parties exercised or could have exercised control over this Petition. No other parties funded or directed the filing of this Petition.

Patent No. 11,289,093
Petition for *Inter Partes* Review

B. Related Matters Under 37 C.F.R. §42.8(b)(2).

The '093 Patent is asserted in *Ricoh Company, Ltd. v. Zoom Communications, Inc. f/k/a Zoom Video Communications, Inc.*, No. 25-2095 (JCG) (D. Del.).

C. Lead and Back-up Counsel under 37 C.F.R. §42.8(b)(3).

Amanda S. Bonner (Lead Counsel) Registration No. 65,224 MAYER BROWN LLP 71 South Wacker Drive Chicago, IL 60606 Telephone: (312) 701-8645 Facsimile: 312-701-7711 asbonner@mayerbrown.com	B. Clayton McCraw (Back-up Counsel) Registration No. 68,258 MAYER BROWN LLP 1221 Avenue of the Americas New York, NY 10020-1001 Telephone: (212) 506-2649 CMcCraw@mayerbrown.com
Cliff A. Maier (Back-up Counsel) Registration No. 59,526 Mayer Brown LLP 3000 El Camino Real Palo Alto, CA 94206-2112 Telephone: 650-331-2026 CMaier@mayerbrown.com	Nicholas Ciulla (Back-up Counsel) Registration No. 78,409 Mayer Brown LLP 71 South Wacker Drive Chicago, IL 60606 Telephone: (312) 701-7779 Facsimile: 312-701-7711 nciulla@mayerbrown.com

D. Service Information under 37 C.F.R. §42.8(b)(4).

Please direct all correspondence to lead counsel at the address identified above. Petitioner consents to electronic service on counsel of record.

XI. GROUNDS FOR STANDING

Petitioner certifies that the '093 Patent is available for *inter partes* review and that Petitioner is not barred or estopped from requesting an *inter partes* review of the '093 Patent. 37 C.F.R. §42.104(a).

XII. PAYMENT OF FEES

A fee of \$54,695 charged to Deposit Account 130019 was paid with the filing this petition. 37 C.F.R. §§42.15(a), 42.103. Should any further fees be required by the present Petition, the Board is hereby authorized to charge the above-referenced Deposit Account.

XIII. STATEMENT OF PRECISE RELIEF REQUESTED

Petitioner requests IPR and cancellation of claims 1-22 because they are unpatentable under 35 U.S.C. §103 based on the foregoing grounds.

Date: August 24, 2026

Respectfully submitted,

By:/ B. Amanda Streff Bonner /
Amanda Streff Bonner (65,224)
Nicholas Ciulla (78,409)
MAYER BROWN LLP
71 S. Wacker Drive
Chicago, IL 60606
Telephone: 312-701-8645
Facsimile: 312-701-7711
asbonner@mayerbrown.com

B. Clayton McCraw
Registration No. 68,258
MAYER BROWN LLP
1221 Avenue of the Americas
New York, NY 10020-1001
Telephone: (212) 506-2649
CMcCraw@mayerbrown.com

Cliff A. Maier
Registration No. 59,526
Mayer Brown LLP
3000 El Camino Real

Patent No. 11,289,093
Petition for *Inter Partes* Review

Palo Alto, CA 94206-2112
Telephone: 650-331-2026
CMaier@mayerbrown.com

Counsel for Petitioner

CERTIFICATE OF SERVICE

Pursuant to 37 C.F.R. §§42.6 and 42.105, I hereby certify that on August 24, 2026, a copy of the attached PETITION FOR INTER PARTES REVIEW OF U.S. PATENT NO. 11,289,093, together with all exhibits, and all other papers filed therewith, including a Power of Attorney on behalf of Petitioners, was served via Priority Mail Express on the attorneys of record for the patent owner at the following address:

Attorney of Record for Patent Owner

154930 - Xsensus
8521 Leesburg Pike
Suite 310
Vienna, VA 22182
UNITED STATES

A courtesy copy was also sent via email to counsel of record in the corresponding litigation.

Date: August 24, 2026

Respectfully submitted,

By:/ Amanda Streff Bonner/
Amanda Streff Bonner (65,224)
Nicholas Ciulla (78,409)
MAYER BROWN LLP
71 S. Wacker Drive
Chicago, IL 60606
Telephone: 312-701-8645
Facsimile: 312-701-7711
asbonner@mayerbrown.com

B. Clayton McCraw
Registration No. 68,258
MAYER BROWN LLP

Patent No. 11,289,093
Petition for *Inter Partes* Review

1221 Avenue of the Americas
New York, NY 10020-1001
Telephone: (212) 506-2649
CMcCraw@mayerbrown.com

Cliff A. Maier
Registration No. 59,526
Mayer Brown LLP
3000 El Camino Real
Palo Alto, CA 94206-2112
Telephone: 650-331-2026
CMaier@mayerbrown.com

Counsel for Petitioner

CERTIFICATION PURSUANT TO 37 C.F.R. § 42.11

Pursuant 37 C.F.R. § 42.11, the undersigned certifies that this PETITION FOR INTER PARTES REVIEW OF U.S. PATENT NO. 11,289,093, and accompanying evidence are not being presented for an improper purpose and that all legal contentions, allegations, and denials are warranted and have evidentiary support.

August 24, 2026

Respectfully submitted,

By:/ Amanda Streff Bonner/
Amanda Streff Bonner (65,224)
Nicholas Ciulla (78,409)
MAYER BROWN LLP
71 S. Wacker Drive
Chicago, IL 60606
Telephone: 312-701-8645
Facsimile: 312-701-7711
asbonner@mayerbrown.com

B. Clayton McCraw
Registration No. 68,258
MAYER BROWN LLP
1221 Avenue of the Americas
New York, NY 10020-1001
Telephone: (212) 506-2649
CMcCraw@mayerbrown.com

Cliff A. Maier
Registration No. 59,526
Mayer Brown LLP
3000 El Camino Real
Palo Alto, CA 94206-2112
Telephone: 650-331-2026
CMaier@mayerbrown.com

Counsel for Petitioner

CERTIFICATION PURSUANT TO 37 C.F.R. § 42.24(d)

Pursuant to 37 C.F.R. § 42.24(d), the undersigned certifies that this PETITION FOR INTER PARTES REVIEW OF U.S. PATENT NO. 11,289,093 complies with the type-volume limitation of 37 C.F.R. § 42.24(a). The word count application of the word processing program used to prepare this paper states that the paper contains 13,999 words, excluding the parts of the paper exempted by 37 C.F.R. §§ 42.24(a)(1).

Date: August 24, 2026

Respectfully submitted,

By:/ Amanda Streff Bonner/
Amanda Streff Bonner (65,224)
Nicholas Ciulla (78,409)
MAYER BROWN LLP
71 S. Wacker Drive
Chicago, IL 60606
Telephone: 312-701-8645
Facsimile: 312-701-7711
asbonner@mayerbrown.com

B. Clayton McCraw
Registration No. 68,258
MAYER BROWN LLP
1221 Avenue of the Americas
New York, NY 10020-1001
Telephone: (212) 506-2649
CMcCraw@mayerbrown.com

Cliff A. Maier
Registration No. 59,526
Mayer Brown LLP
3000 El Camino Real
Palo Alto, CA 94206-2112

Telephone: 650-331-2026
CMaier@mayerbrown.com

Counsel for Petitioner