(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
12 September 2019 (12.09.2019)

WIPO | PCT

(10) International Publication Number
WO 2019/173353 A1

(51) International Patent Classification:
G06F 3/00 (2006.01)

(21) International Application Number:
PCT/US2019/020771

(22) International Filing Date:
05 March 2019 (05.03.2019)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/638,809 05 March 2018 (05.03.2018) US
62/803,193 08 February 2019 (08.02.2019) US

(71) Applicant: **NUANCE COMMUNICATIONS, INC.** [US/US]; One Wayside Road, Burlington, Massachusetts 01803 (US).

(72) Inventors: **VOZILA, Paul Joseph**; 17 Grandview Road, Arlington, Massachusetts 02476 (US). **GALLOPYN, Guido Remi Marcel**; 14 Dorothy E. Lucey Drive, Newburyport, Massachusetts 01950 (US). **JOST, Uwe Helmut**; 119 Tavern Road, Groton, Massachusetts 01450 (US). **HELLETZGRUBER, Matthias**; Kravoglgasse 38, 1210 Wien (AT). **JANCSARY, Jeremy Martin**; Arndtstrasse 19/5, 1120 Vienna (AT). **ABHINAV, Kumar**; 5900 Rue Claude-Masson, Montreal, Québec H1K 0H8 (CA). **PINTO, Joel Praveen**; Schloss-Rahe-Winkel, 26, 52072 Aachen (DE). **OWEN, Donald E.**; 9002 Pecky Cypress Way, Orlando, Florida 32836 (US). **OZ, Mehmet Mert**; Josef Kollmannstrasse 29D, 2500 Baden (AT). **SBIHILI, Scott**; 33 Summer Street, Lexington, Massachusetts 02420 (US). **ERSKINE, Garret N.**; 24325 Crenshaw Blvd. #297, Torrance, California 90505 (US). **DELANEY, Brian William**; 75 Bolton Woods Way, Bolton, Massachusetts 01740 (US).

(74) Agent: **ABRAMSON, Michael T.** et al.; Holland & Knight LLP, 10 St. James Avenue, Boston, Massachusetts 02116 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN,

(54) Title: SYSTEM AND METHOD FOR REVIEW OF AUTOMATED CLINICAL DOCUMENTATION

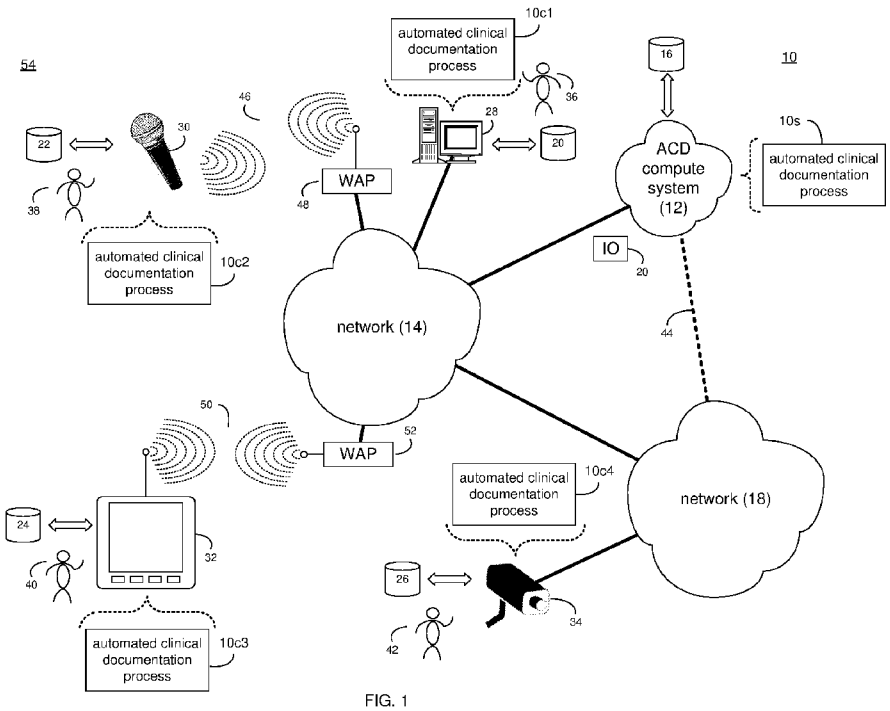

FIG. 1

(57) Abstract: A method, computer program product, and computing system for obtaining, by a computing device, encounter information of a patient encounter, wherein the encounter information may include audio encounter information obtained from at least a first encounter participant. The audio encounter information obtained from at least the first encounter participant may be processed. An alert may be generated to obtain additional encounter information of the patient encounter.

WO 2019/173353 A1

[Continued on next page]

HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

System and Method for Review of Automated Clinical Documentation

Related Application(s)

[001] This application claims the benefit of the following U.S. Provisional Application Nos.: 62/803,193 filed on 08 February 2019 and 62/638,809 filed on 05 March 2018, the contents of which are all incorporated herein by reference.

Background

[002] Automated Clinical Documentation (ACD) may be used, e.g., to turn transcribed conversational (e.g., physician, patient, and/or other participants such as patient's family members, nurses, physician assistants, etc.) speech into formatted (e.g., medical) reports. Such reports may be reviewed, e.g., to assure accuracy of the reports by the physician, scribe, etc.

Summary of Disclosure

[003] In one implementation, a computer-implemented method executed by a computer may include but is not limited to obtaining, by a computing device, encounter information of a patient encounter, wherein the encounter information may include audio encounter information obtained from at least a first encounter participant. The audio encounter information obtained from at least the first encounter participant may be processed. An alert may be generated to obtain additional encounter information of the patient encounter.

[004] One or more of the following features may be included. The alert may be at least one of an audio alert and a visual alert. Processing the first audio encounter information may include defining linkages between each of the plurality of layers associated with the audio encounter information. A user interface may be generated displaying a plurality of layers associated with the audio encounter information obtained from at least the first encounter participant. The first layer of the plurality of layers may be an audio signal associated with the audio encounter information, wherein the second layer of the plurality of layers may be a transcript associated with the audio

encounter information, and wherein the third layer of the plurality of layers may be a medical report associated with the audio encounter information. Processing the first audio encounter information may include comparing at least one layer of the plurality of layers with at least one data source. The at least one data source may include at least one of physical exam finding information, lab order information, medical condition checklist information, and compliance information.

[005] In another implementation, a computer program product resides on a computer readable medium and has a plurality of instructions stored on it. When executed by a processor, the instructions cause the processor to perform operations including but not limited to obtaining encounter information of a patient encounter, wherein the encounter information may include audio encounter information obtained from at least a first encounter participant. The audio encounter information obtained from at least the first encounter participant may be processed. An alert may be generated to obtain additional encounter information of the patient encounter.

[006] One or more of the following features may be included. The alert may be at least one of an audio alert and a visual alert. Processing the first audio encounter information may include defining linkages between each of the plurality of layers associated with the audio encounter information. A user interface may be generated displaying a plurality of layers associated with the audio encounter information obtained from at least the first encounter participant. The first layer of the plurality of layers may be an audio signal associated with the audio encounter information, wherein the second layer of the plurality of layers may be a transcript associated with the audio encounter information, and wherein the third layer of the plurality of layers may be a medical report associated with the audio encounter information. Processing the first audio encounter information may include comparing at least one layer of the plurality of layers with at least one data source. The at least one data source may include at least one of physical exam finding information, lab order information, medical condition checklist information, and compliance information.

[007] In another implementation, a computing system includes a processor and memory is configured to perform operations including but not limited to obtaining encounter information of a patient encounter, wherein the encounter information may include audio encounter information obtained from at least a first encounter participant. The audio encounter information obtained from at least the first encounter participant may be processed. An alert may be generated to obtain additional encounter information of the patient encounter.

[008] One or more of the following features may be included. The alert may be at least one of an audio alert and a visual alert. Processing the first audio encounter information may include defining linkages between each of the plurality of layers associated with the audio encounter information. A user interface may be generated displaying a plurality of layers associated with the audio encounter information obtained from at least the first encounter participant. The first layer of the plurality of layers may be an audio signal associated with the audio encounter information, wherein the second layer of the plurality of layers may be a transcript associated with the audio encounter information, and wherein the third layer of the plurality of layers may be a medical report associated with the audio encounter information. Processing the first audio encounter information may include comparing at least one layer of the plurality of layers with at least one data source. The at least one data source may include at least one of physical exam finding information, lab order information, medical condition checklist information, and compliance information.

[009] The details of one or more implementations are set forth in the accompanying drawings and the description below. Other features and advantages will become apparent from the description, the drawings, and the claims.

Brief Description of the Drawings

[0010] FIG. 1 is a diagrammatic view of an automated clinical documentation computer system and an automated clinical documentation process coupled to a distributed computing network;

[0011] FIG. 2 is a diagrammatic view of a modular ACD system incorporating the automated clinical documentation computer system of FIG. 1;

[0012] FIG. 3 is a diagrammatic view of a mixed-media ACD device included within the modular ACD system of FIG. 2;

[0013] FIG. 4 is a flow chart of one implementation of the automated clinical documentation process of FIG. 1;

[0014] FIG. 5 is a user interface of the automated clinical documentation process of FIG. 1;

[0015] FIG. 6 is a user interface of the automated clinical documentation process of FIG. 1;

[0016] FIG. 7 is a user interface of the automated clinical documentation process of FIG. 1;

[0017] FIG. 8 is a user interface of the automated clinical documentation process of FIG. 1;

[0018] FIG. 9 is a flow chart of one implementation of the automated clinical documentation process of FIG. 1;

[0019] FIG. 10 is a flow chart of one implementation of the automated clinical documentation process of FIG. 1;

[0020] FIG. 11 is a diagrammatic view of example peripheral devices that may be used with the automated clinical documentation process of FIG. 1;

[0021] FIG. 12 is a flow chart of one implementation of the automated clinical documentation process of FIG. 1;

[0022] FIG. 13 is a user interface of the automated clinical documentation process of FIG. 1;

[0023] FIG. 14 is a user interface of the automated clinical documentation process of FIG. 1; and

[0024] FIG. 15 is a flow chart of another implementation of the automated clinical documentation process of FIG. 1.

[0025] Like reference symbols in the various drawings indicate like elements.

Detailed Description of the Preferred Embodiments

System Overview:

[0026] Referring to FIG. 1, there is shown automated clinical documentation process 10. As will be discussed below in greater detail, automated clinical documentation process 10 may be configured to automate the collection and processing of clinical encounter information to generate / store / distribute medical records.

[0027] Automated clinical documentation process 10 may be implemented as a server-side process, a client-side process, or a hybrid server-side / client-side process. For example, automated clinical documentation process 10 may be implemented as a purely server-side process via automated clinical documentation process 10s. Alternatively, automated clinical documentation process 10 may be implemented as a purely client-side process via one or more of automated clinical documentation process 10c1, automated clinical documentation process 10c2, automated clinical documentation process 10c3, and automated clinical documentation process 10c4. Alternatively still, automated clinical documentation process 10 may be implemented as a hybrid server-side / client-side process via automated clinical documentation process 10s in combination with one or more of automated clinical documentation process 10c1, automated clinical documentation process 10c2, automated clinical documentation process 10c3, and automated clinical documentation process 10c4.

[0028] Accordingly, automated clinical documentation process 10 as used in this disclosure may include any combination of automated clinical documentation process 10s, automated clinical documentation process 10c1, automated clinical documentation process 10c2, automated clinical documentation process 10c3, and automated clinical documentation process 10c4.

[0029] Automated clinical documentation process 10s may be a server application and may reside on and may be executed by automated clinical documentation (ACD)

computer system 12, which may be connected to network 14 (e.g., the Internet or a local area network). ACD computer system 12 may include various components, examples of which may include but are not limited to: a personal computer, a server computer, a series of server computers, a mini computer, a mainframe computer, one or more Network Attached Storage (NAS) systems, one or more Storage Area Network (SAN) systems, one or more Platform as a Service (PaaS) systems, one or more Infrastructure as a Service (IaaS) systems, one or more Software as a Service (SaaS) systems, a cloud-based computational system, and a cloud-based storage platform.

[0030] As is known in the art, a SAN may include one or more of a personal computer, a server computer, a series of server computers, a mini computer, a mainframe computer, a RAID device and a NAS system. The various components of ACD computer system 12 may execute one or more operating systems, examples of which may include but are not limited to: Microsoft Windows Server ™; Redhat Linux ™, Unix, or a custom operating system, for example.

[0031] The instruction sets and subroutines of automated clinical documentation process 10s, which may be stored on storage device 16 coupled to ACD computer system 12, may be executed by one or more processors (not shown) and one or more memory architectures (not shown) included within ACD computer system 12. Examples of storage device 16 may include but are not limited to: a hard disk drive; a RAID device; a random access memory (RAM); a read-only memory (ROM); and all forms of flash memory storage devices.

[0032] Network 14 may be connected to one or more secondary networks (e.g., network 18), examples of which may include but are not limited to: a local area network; a wide area network; or an intranet, for example.

[0033] Various IO requests (e.g. IO request 20) may be sent from automated clinical documentation process 10s, automated clinical documentation process 10c1, automated clinical documentation process 10c2, automated clinical documentation process 10c3 and/or automated clinical documentation process 10c4 to ACD computer system 12.

Examples of IO request 20 may include but are not limited to data write requests (i.e. a request that content be written to ACD computer system 12) and data read requests (i.e. a request that content be read from ACD computer system 12).

[0034] The instruction sets and subroutines of automated clinical documentation process 10c1, automated clinical documentation process 10c2, automated clinical documentation process 10c3 and/or automated clinical documentation process 10c4, which may be stored on storage devices 20, 22, 24, 26 (respectively) coupled to ACD client electronic devices 28, 30, 32, 34 (respectively), may be executed by one or more processors (not shown) and one or more memory architectures (not shown) incorporated into ACD client electronic devices 28, 30, 32, 34 (respectively). Storage devices 20, 22, 24, 26 may include but are not limited to: hard disk drives; optical drives; RAID devices; random access memories (RAM); read-only memories (ROM), and all forms of flash memory storage devices. Examples of ACD client electronic devices 28, 30, 32, 34 may include, but are not limited to, personal computing device 28 (e.g., a smart phone, a personal digital assistant, a laptop computer, a notebook computer, and a desktop computer), audio input device 30 (e.g., a handheld microphone, a lapel microphone, an embedded microphone (such as those embedded within eyeglasses, smart phones, tablet computers and/or watches) and an audio recording device), display device 32 (e.g., a tablet computer, a computer monitor, and a smart television), machine vision input device 34 (e.g., an RGB imaging system, an infrared imaging system, an ultraviolet imaging system, a laser imaging system, a SONAR imaging system, a RADAR imaging system, and a thermal imaging system), a hybrid device (e.g., a single device that includes the functionality of one or more of the above-references devices; not shown), an audio rendering device (e.g., a speaker system, a headphone system, or an earbud system; not shown), various medical devices (e.g., medical imaging equipment, heart monitoring machines, body weight scales, body temperature thermometers, and blood pressure machines; not shown), and a dedicated network device (not shown).

[0035] Users 36, 38, 40, 42 may access ACD computer system 12 directly through network 14 or through secondary network 18. Further, ACD computer system 12 may be connected to network 14 through secondary network 18, as illustrated with link line 44.

[0036] The various ACD client electronic devices (e.g., ACD client electronic devices 28, 30, 32, 34) may be directly or indirectly coupled to network 14 (or network 18). For example, personal computing device 28 is shown directly coupled to network 14 via a hardwired network connection. Further, machine vision input device 34 is shown directly coupled to network 18 via a hardwired network connection. Audio input device 30 is shown wirelessly coupled to network 14 via wireless communication channel 46 established between audio input device 30 and wireless access point (i.e., WAP) 48, which is shown directly coupled to network 14. WAP 48 may be, for example, an IEEE 802.11a, 802.11b, 802.11g, 802.11n, Wi-Fi, and/or Bluetooth device that is capable of establishing wireless communication channel 46 between audio input device 30 and WAP 48. Display device 32 is shown wirelessly coupled to network 14 via wireless communication channel 50 established between display device 32 and WAP 52, which is shown directly coupled to network 14.

[0037] The various ACD client electronic devices (e.g., ACD client electronic devices 28, 30, 32, 34) may each execute an operating system, examples of which may include but are not limited to Microsoft Windows ™, Apple Macintosh ™, Redhat Linux ™, or a custom operating system, wherein the combination of the various ACD client electronic devices (e.g., ACD client electronic devices 28, 30, 32, 34) and ACD computer system 12 may form modular ACD system 54.

[0038] Referring also to FIG. 2, there is shown a simplified example embodiment of modular ACD system 54 that is configured to automate clinical documentation. Modular ACD system 54 may include: machine vision system 100 configured to obtain machine vision encounter information 102 concerning a patient encounter; audio recording system 104 configured to obtain audio encounter information 106 concerning

the patient encounter; and a computer system (e.g., ACD computer system 12) configured to receive machine vision encounter information 102 and audio encounter information 106 from machine vision system 100 and audio recording system 104 (respectively). Modular ACD system 54 may also include: display rendering system 108 configured to render visual information 110; and audio rendering system 112 configured to render audio information 114, wherein ACD computer system 12 may be configured to provide visual information 110 and audio information 114 to display rendering system 108 and audio rendering system 112 (respectively).

[0039] Example of machine vision system 100 may include but are not limited to: one or more ACD client electronic devices (e.g., ACD client electronic device 34, examples of which may include but are not limited to an RGB imaging system, an infrared imaging system, a ultraviolet imaging system, a laser imaging system, a SONAR imaging system, a RADAR imaging system, and a thermal imaging system). Examples of audio recording system 104 may include but are not limited to: one or more ACD client electronic devices (e.g., ACD client electronic device 30, examples of which may include but are not limited to a handheld microphone, a lapel microphone, an embedded microphone (such as those embedded within eyeglasses, smart phones, tablet computers and/or watches) and an audio recording device). Examples of display rendering system 108 may include but are not limited to: one or more ACD client electronic devices (e.g., ACD client electronic device 32, examples of which may include but are not limited to a tablet computer, a computer monitor, and a smart television). Examples of audio rendering system 112 may include but are not limited to: one or more ACD client electronic devices (e.g., audio rendering device 116, examples of which may include but are not limited to a speaker system, a headphone system, and an earbud system).

[0040] As will be discussed below in greater detail, ACD computer system 12 may be configured to access one or more datasources 118 (e.g., plurality of individual datasources 120, 122, 124, 126, 128), examples of which may include but are not

limited to one or more of a user profile datasource, a voice print datasource, a voice characteristics datasource (e.g., for adapting the automated speech recognition models), a face print datasource, a humanoid shape datasource, an utterance identifier datasource, a wearable token identifier datasource, an interaction identifier datasource, a medical conditions symptoms datasource, a prescriptions compatibility datasource, a medical insurance coverage datasource, and a home healthcare datasource. While in this particular example, five different examples of datasources 118, are shown, this is for illustrative purposes only and is not intended to be a limitation of this disclosure, as other configurations are possible and are considered to be within the scope of this disclosure.

[0041] As will be discussed below in greater detail, modular ACD system 54 may be configured to monitor a monitored space (e.g., monitored space 130) in a clinical environment, wherein examples of this clinical environment may include but are not limited to: a doctor's office, a medical facility, a medical practice, a medical lab, an urgent care facility, a medical clinic, an emergency room, an operating room, a hospital, a long term care facility, a rehabilitation facility, a nursing home, and a hospice facility. Accordingly, an example of the above-referenced patient encounter may include but is not limited to a patient visiting one or more of the above-described clinical environments (e.g., a doctor's office, a medical facility, a medical practice, a medical lab, an urgent care facility, a medical clinic, an emergency room, an operating room, a hospital, a long term care facility, a rehabilitation facility, a nursing home, and a hospice facility).

[0042] Machine vision system 100 may include a plurality of discrete machine vision systems when the above-described clinical environment is larger or a higher level of resolution is desired. As discussed above, examples of machine vision system 100 may include but are not limited to: one or more ACD client electronic devices (e.g., ACD client electronic device 34, examples of which may include but are not limited to an RGB imaging system, an infrared imaging system, an ultraviolet imaging system, a

laser imaging system, a SONAR imaging system, a RADAR imaging system, and a thermal imaging system). Accordingly, machine vision system 100 may include one or more of each of an RGB imaging system, an infrared imaging systems, an ultraviolet imaging systems, a laser imaging system, a SONAR imaging system, a RADAR imaging system, and a thermal imaging system.

[0043] Audio recording system 104 may include a plurality of discrete audio recording systems when the above-described clinical environment is larger or a higher level of resolution is desired. As discussed above, examples of audio recording system 104 may include but are not limited to: one or more ACD client electronic devices (e.g., ACD client electronic device 30, examples of which may include but are not limited to a handheld microphone, a lapel microphone, an embedded microphone (such as those embedded within eyeglasses, smart phones, tablet computers and/or watches) and an audio recording device). Accordingly, audio recording system 104 may include one or more of each of a handheld microphone, a lapel microphone, an embedded microphone (such as those embedded within eyeglasses, smart phones, tablet computers and/or watches) and an audio recording device.

[0044] Display rendering system 108 may include a plurality of discrete display rendering systems when the above-described clinical environment is larger or a higher level of resolution is desired. As discussed above, examples of display rendering system 108 may include but are not limited to: one or more ACD client electronic devices (e.g., ACD client electronic device 32, examples of which may include but are not limited to a tablet computer, a computer monitor, and a smart television). Accordingly, display rendering system 108 may include one or more of each of a tablet computer, a computer monitor, and a smart television.

[0045] Audio rendering system 112 may include a plurality of discrete audio rendering systems when the above-described clinical environment is larger or a higher level of resolution is desired. As discussed above, examples of audio rendering system 112 may include but are not limited to: one or more ACD client electronic devices (e.g.,

audio rendering device 116, examples of which may include but are not limited to a speaker system, a headphone system, or an earbud system). Accordingly, audio rendering system 112 may include one or more of each of a speaker system, a headphone system, or an earbud system.

[0046] ACD computer system 12 may include a plurality of discrete computer systems. As discussed above, ACD computer system 12 may include various components, examples of which may include but are not limited to: a personal computer, a server computer, a series of server computers, a mini computer, a mainframe computer, one or more Network Attached Storage (NAS) systems, one or more Storage Area Network (SAN) systems, one or more Platform as a Service (PaaS) systems, one or more Infrastructure as a Service (IaaS) systems, one or more Software as a Service (SaaS) systems, a cloud-based computational system, and a cloud-based storage platform. Accordingly, ACD computer system 12 may include one or more of each of a personal computer, a server computer, a series of server computers, a mini computer, a mainframe computer, one or more Network Attached Storage (NAS) systems, one or more Storage Area Network (SAN) systems, one or more Platform as a Service (PaaS) systems, one or more Infrastructure as a Service (IaaS) systems, one or more Software as a Service (SaaS) systems, a cloud-based computational system, and a cloud-based storage platform.

[0047] Referring also to FIG. 3, audio recording system 104 may include directional microphone array 200 having a plurality of discrete microphone assemblies. For example, audio recording system 104 may include a plurality of discrete audio acquisition devices (e.g., audio acquisition devices 202, 204, 206, 208, 210, 212, 214, 216, 218) that may form microphone array 200. As will be discussed below in greater detail, modular ACD system 54 may be configured to form one or more audio recording beams (e.g., audio recording beams 220, 222, 224) via the discrete audio acquisition devices (e.g., audio acquisition devices 202, 204, 206, 208, 210, 212, 214, 216, 218) included within audio recording system 104.

[0048] For example, modular ACD system 54 may be further configured to steer the one or more audio recording beams (e.g., audio recording beams 220, 222, 224) toward one or more encounter participants (e.g., encounter participants 226, 228, 230) of the above-described patient encounter. Examples of the encounter participants (e.g., encounter participants 226, 228, 230) may include but are not limited to: medical professionals (e.g., doctors, nurses, physician's assistants, lab technicians, physical therapists, scribes (e.g., a transcriptionist) and/or staff members involved in the patient encounter), patients (e.g., people that are visiting the above-described clinical environments for the patient encounter), and third parties (e.g., friends of the patient, relatives of the patient and/or acquaintances of the patient that are involved in the patient encounter).

[0049] Accordingly, modular ACD system 54 and/or audio recording system 104 may be configured to utilize one or more of the discrete audio acquisition devices (e.g., audio acquisition devices 202, 204, 206, 208, 210, 212, 214, 216, 218) to form an audio recording beam. For example, modular ACD system 54 and/or audio recording system 104 may be configured to utilize audio acquisition device 210 to form audio recording beam 220, thus enabling the capturing of audio (e.g., speech) produced by encounter participant 226 (as audio acquisition device 210 is pointed to (i.e., directed toward) encounter participant 226). Additionally, modular ACD system 54 and/or audio recording system 104 may be configured to utilize audio acquisition devices 204, 206 to form audio recording beam 222, thus enabling the capturing of audio (e.g., speech) produced by encounter participant 228 (as audio acquisition devices 204, 206 are pointed to (i.e., directed toward) encounter participant 228). Additionally, modular ACD system 54 and/or audio recording system 104 may be configured to utilize audio acquisition devices 212, 214 to form audio recording beam 224, thus enabling the capturing of audio (e.g., speech) produced by encounter participant 230 (as audio acquisition devices 212, 214 are pointed to (i.e., directed toward) encounter participant 230). Further, modular ACD system 54 and/or audio recording system 104 may be

configured to utilize null-steering precoding to cancel interference between speakers and/or noise.

[0050] As is known in the art, null-steering precoding is a method of spatial signal processing by which a multiple antenna transmitter may null multiuser interference signals in wireless communications, wherein null-steering precoding may mitigate the impact off background noise and unknown user interference.

[0051] In particular, null-steering precoding may be a method of beamforming for narrowband signals that may compensate for delays of receiving signals from a specific source at different elements of an antenna array. In general and to improve performance of the antenna array, in incoming signals may be summed and averaged, wherein certain signals may be weighted and compensation may be made for signal delays.

[0052] Machine vision system 100 and audio recording system 104 may be stand-alone devices (as shown in FIG. 2). Additionally / alternatively, machine vision system 100 and audio recording system 104 may be combined into one package to form mixed-media ACD device 232. For example, mixed-media ACD device 232 may be configured to be mounted to a structure (e.g., a wall, a ceiling, a beam, a column) within the above-described clinical environments (e.g., a doctor's office, a medical facility, a medical practice, a medical lab, an urgent care facility, a medical clinic, an emergency room, an operating room, a hospital, a long term care facility, a rehabilitation facility, a nursing home, and a hospice facility), thus allowing for easy installation of the same. Further, modular ACD system 54 may be configured to include a plurality of mixed-media ACD devices (e.g., mixed-media ACD device 232) when the above-described clinical environment is larger or a higher level of resolution is desired.

[0053] Modular ACD system 54 may be further configured to steer the one or more audio recording beams (e.g., audio recording beams 220, 222, 224) toward one or more encounter participants (e.g., encounter participants 226, 228, 230) of the patient encounter based, at least in part, upon machine vision encounter information 102. As discussed above, mixed-media ACD device 232 (and machine vision system 100 / audio

recording system 104 included therein) may be configured to monitor one or more encounter participants (e.g., encounter participants 226, 228, 230) of a patient encounter.

[0054] Specifically, machine vision system 100 (either as a stand-alone system or as a component of mixed-media ACD device 232) may be configured to detect humanoid shapes within the above-described clinical environments (e.g., a doctor's office, a medical facility, a medical practice, a medical lab, an urgent care facility, a medical clinic, an emergency room, an operating room, a hospital, a long term care facility, a rehabilitation facility, a nursing home, and a hospice facility). And when these humanoid shapes are detected by machine vision system 100, modular ACD system 54 and/or audio recording system 104 may be configured to utilize one or more of the discrete audio acquisition devices (e.g., audio acquisition devices 202, 204, 206, 208, 210, 212, 214, 216, 218) to form an audio recording beam (e.g., audio recording beams 220, 222, 224) that is directed toward each of the detected humanoid shapes (e.g., encounter participants 226, 228, 230).

[0055] As discussed above, ACD computer system 12 may be configured to receive machine vision encounter information 102 and audio encounter information 106 from machine vision system 100 and audio recording system 104 (respectively); and may be configured to provide visual information 110 and audio information 114 to display rendering system 108 and audio rendering system 112 (respectively). Depending upon the manner in which modular ACD system 54 (and/or mixed-media ACD device 232) is configured, ACD computer system 12 may be included within mixed-media ACD device 232 or external to mixed-media ACD device 232.

[0056] As discussed above, ACD computer system 12 may execute all or a portion of automated clinical documentation process 10, wherein the instruction sets and subroutines of automated clinical documentation process 10 (which may be stored on one or more of e.g., storage devices 16, 20, 22, 24, 26) may be executed by ACD computer system 12 and/or one or more of ACD client electronic devices 28, 30, 32,

34.

[0057] As discussed above, automated clinical documentation (ACD) process 10 may be configured to automate the collection and processing of clinical encounter information to generate / store / distribute medical records. Accordingly and referring also to FIGS. 4-8, ACD process 10 may be configured to obtain 300 encounter information (e.g., machine vision encounter information 102 and/or audio encounter information 106) of a patient encounter (e.g., a visit to a doctor's office) of at least a first encounter participant, wherein the encounter information may include audio encounter information obtained from at least a first encounter participant (e.g., encounter participant 228, 226, 230, and/or 242). ACD process 10 may further be configured to process 302 the encounter information (e.g., machine vision encounter information 102 and/or audio encounter information 106) obtained from at least the first encounter participant, e.g., to generate an encounter transcript (e.g., encounter transcript 234) and/or generate 304 a user interface displaying a plurality of layers associated with the audio encounter information obtained from at least the first encounter participant. In some implementations, ACD process 10 may process at least a portion of the encounter transcript (e.g., encounter transcript 234) to populate at least a portion of a medical record (e.g., medical record 236) associated with the patient encounter (e.g., the visit to the doctor's office). Encounter transcript 234 and/or medical record 236 may be reviewed by a medical professional involved with the patient encounter (e.g., a visit to a doctor's office) to determine the accuracy of the same and/or make corrections to the same.

[0058] As noted above, ACD process 10 may process 302 the audio encounter information obtained from at least the first encounter participant. In some implementations, processing 302 the first audio encounter information may include defining 306 linkages between each of the plurality of layers associated with the audio encounter information. For example, the first layer of the plurality of layers may be an audio signal associated with the audio encounter information (e.g., complete audio of

the encounter, encompassing and clearly delineating each participant), wherein the second layer of the plurality of layers may be a transcript associated with the audio encounter information (e.g., a diarized audio transcript (*verbatim*) for each participant in the encounter), and wherein the third layer of the plurality of layers may be a medical report associated with the audio encounter information (e.g., a draft medical report in the appropriate clinical output format). In some implementations, additional layers may include, e.g., the above-noted machine vision-based recording of the encounter, including various signal formats and features, and discrete, standardized, actionable data resulting from the encounter, including, but not limited to medication plans (for example, in RxNorm) or lab orders (for example, LOINC) or diagnoses (for example, ICD10, CPT etc). In the example, the signals captured from the encounter information may be processed 304 into at least the above-noted three separate, yet closely linked and interdependent layers.

[0059] In some implementations, ACD process 10 may include an ASR portion that may process 302 the audio encounter information producing an approximate (e.g., diarized) verbatim transcript along with alignment information indicating the audio interval corresponding to each transcript word. In some implementations, a deep learning (e.g., sequence to sequence) model associated with ACD process 10 may convert the transcript to a medical report. It will be appreciated that various attribution techniques may be employed by ACD process 10 that may effectively softly assign responsibility for a given output (e.g., medical report) word to input (e.g., conversation transcript) words (e.g. attention weights, integrated gradient, etc.) according to the model. As a result, this may provide a soft mapping from the transcript word positions to report word positions. In some implementations, the input word position assigned maximal attribution for a given output word may be interpreted as being aligned (linked) to that output (e.g., when a hard mapping is required). Based on the ASR time alignment, a word in the draft medical report, aligned to a word in the ASR conversation transcript, may now be associated with an audio time interval of the associated audio

signal of the audio encounter information.

[0060] In some implementations, ACD process 10 may also may link (i.e., align) the ASR conversation transcript words with the draft medical report words. For transcript words that may have maximal attribution value for some set of medical report words, ACD process 10 may link them with the first word in that set. For the remaining transcript words, ACD process 10 may link them to the same word that the nearest preceding (or if none, nearest subsequent) conversation transcript word is linked to.

[0061] In some implementations, a visual recording (e.g., video stream of the patient encounter), if available, may also be a layer and may be time indexed and thus a given point in the recording may be associated with the same time in the audio recording and thus a conversation transcript word and draft report word. In some implementations, if discrete, standardized, actionable data is produced as a second (parallel) output sequence of the sequence to sequence model, then a similar model output attribution technique may be used to align tokens in this actionable data with the ASR conversation transcript words, and thus the audio intervals.

[0062] In some implementations, ACD process 10 may generate 304 a user interface displaying a plurality of layers associated with the audio encounter information obtained from at least the first encounter participant. For example, and referring at least to the example implementation of Fig. 5, an example user interface (e.g., user interface 500) is shown. It will be appreciated that UI 500 is shown merely for example purposes only, and that more or less features and arrangements of features may be used without departing from the scope of the present disclosure.

[0063] As can be seen from Fig. 5, UI 500 includes the first layer of the plurality of layers (e.g., first layer 502 which may be a visualization and/or direct play-back of an audio signal associated with the audio encounter information (e.g., complete audio of the encounter, encompassing and clearly delineating each participant), second layer 504 of the plurality of layers which may be a transcript (e.g., encounter transcript 234) associated with the audio encounter information (e.g., a diarized audio transcript

(*verbatim*) for each participant in the encounter), and third layer 506 of the plurality of layers which may be a medical report (e.g., medical report/record 236) associated with the audio encounter information (e.g., a draft medical report in the appropriate clinical output format). It will be appreciated that that each layer (audio, transcript, draft report, etc.) may be rendered in multiple different ways, and in the most appropriate way for a given use / preference of the end-user. As such, the specific rendering of layers should be taken as example only and not to limit the scope of the disclosure.

[0064] As noted above, encounter transcript 234 (via second layer 504) and/or medical record/record 236 (via third layer 506) may be reviewed by a medical professional involved with the patient encounter (e.g., a visit to a doctor's office) to determine the accuracy of the same and/or make corrections to the same. For example, a scribe involved with (or assigned to) the patient encounter (e.g., a visit to a doctor's office) may review encounter transcript 234 and/or medical record 236 to confirm that the same was accurate and/or make corrections to the same. In the event that corrections are made to encounter transcript 234 and/or medical record 236, automated clinical documentation process 10 may utilize these corrections for training / tuning purposes (e.g., to adjust the various profiles associated the participants of the patient encounter) to enhance the future accuracy / efficiency / performance of automated clinical documentation process 10.

[0065] Alternatively / additionally, a doctor involved with the patient encounter (e.g., a visit to a doctor's office) may review encounter transcript 234 and/or medical record 236 to confirm that the same was accurate and/or make corrections to the same. In the event that corrections are made to encounter transcript 234 and/or medical record 236, automated clinical documentation process 10 may utilize these corrections for training / tuning purposes (e.g., to adjust the various profiles associated the participants of the patient encounter) to enhance the future accuracy / efficiency / performance of automated clinical documentation process 10.

[0066] As such, and referring at least to the example implementation of Fig. 6,

assume for example purposes only that a scribe is reviewing the encounter information using UI 500. In some implementations, ACD process 10 may receive 308 a selection of a first portion of the audio encounter information at a first layer of the plurality of layers on the user interface, and ACD process 10 may display 310 an annotation of at least one of a second layer of the plurality of layers and a third layer of the plurality of layers corresponding to the first portion of the audio encounter information of the first layer of the plurality of layers selected on the user interface. For instance, the scribe may use, e.g., cursor 508 or other known input techniques, to select a portion of first layer 502 (i.e., the audio information of the audio encounter information), which may be received 308 by ACD process 10. As a result, in some implementations, ACD process 10 may display 310 some kind of annotation of either the second layer (e.g., the portion of the transcript associated with the selected portion of the first layer) and/or the third layer (e.g., the portion of the medical report associated with the selected portion of the first layer). In the example, ACD process 10 has displayed an annotation (e.g., bold font) of the second layer (e.g., the portion of the transcript associated with the selected portion of the first layer) and the third layer (e.g., the portion of the medical report associated with the selected portion of the first layer). It will be appreciated that any type of annotation may be used (e.g., bold, italics, underline, highlights, shading, transparency, etc.) without departing from the scope of the present disclosure. As such, the use of bolding as the annotation should be taken as example only and not to otherwise limit the scope of the disclosure.

[0067] Similarly, in some implantations, ACD process 10 may receive 312 a selection of the first portion of the audio encounter information at one of the second layer of the plurality of layers and the third layer of the plurality of layers on the user interface, and ACD process 10 may provide 314 audio of the first layer corresponding to the first portion of the audio encounter information of one of the second layer of the plurality of layers and the third layer of the plurality of layers selected on the user interface. For example, and referring to the example implementation of Fig. 7, the

scribe may use, e.g., cursor 508 or other known input techniques, to select a portion of second layer 504 (e.g., the transcript), and/or the third layer (e.g., the medical report) which may be received 312 by ACD process 10. As a result, in some implementations, ACD process 10 may provide 314 the audio of first layer 502 (e.g., the audio information of the audio encounter associated with the selected portion of second layer 504).

[0068] As such, ACD process 10 may leverage the above-noted defined linkages, such that the encounter recording as captured above may be provided to a human scribe (or physician) in such a fashion, so that they may navigate through each of these layers independently and be always in sync. For instance, if the scribe navigates through first layer 502 (audio) by listening, the part of the transcript that corresponds to that part of the audio may be bolded, highlighted, etc. accordingly, as well as the resulting medical report (and if relevant actionable data) from that information. Additionally / alternatively, if the scribe selects a part of the report, the audio / transcript sections that affected that part of the report may be bolded, highlighted, emphasized, etc. This annotation also may correspond to the audio and report "cursor" (e.g., audio cursor 510) moving, e.g., audio may next play from that position and typing / keyboard-navigation may commence from that position in the report. As a result, since each of the plurality of layers may be synchronized, a scribe (or physician, etc.) may later playback the whole patient encounter, navigate (e.g., "clicking" a mouse cursor of other pointing device at any position) in any portion of any of the layers, make edits, and would know which edit / correction belongs to the other sections (layers). More generally, "clicking" (i.e. resetting the cursor) in one viewed layer may annotate and update the cursor in the linked points / excerpts in the other layers.

[0069] In some implementations, at least a portion of the audio encounter information determined to lack relevance to the medical report may be one of sped up and skipped 316 by ACD process 10. For example, in some implementations, it may be possible for the scribe to "playback" the whole encounter and see the draft medical

report "emerge" (or simply see each associated report words highlighted, in sync with the cursor) so that they may ascertain which signals from the encounter caused / triggered which parts of the report, and react (e.g., correct / edit, etc.) accordingly. However, ACD process 10 may sped up or skip 316 the audio not only where no speech is detected by the ASR portion of ACD process 10, but also where the model (e.g., sequence to sequence / abstractive summarization, etc.) does not attribute any significant responsibility for the draft report to that excerpt of the conversation transcript's aligned audio (e.g., accumulated attribution across all outputs (e.g., medical report text) for an input (e.g., ASR word / excerpt) being below some threshold). For instance, assume for example purposes only that Mrs. Smith (i.e., the patient in this example) asks the Doctor about his golf game, which the Doctor spends about 30 seconds discussing. In the example, ACD process 10 may determine that the discussion of the Doctor's golf game is not relevant for the draft report to that excerpt of the conversation transcript's aligned audio, and may either skip or speed up (fast forward) through that portion of the audio (and the associated transcript).

[0070] Similarly, at least a portion of the audio encounter information determined to lack relevance to the medical report may be annotated 318 by ACD process 10. For instance, and referring to the example implementation of Fig. 8, UI 500 is shown. In the example, ACD process 10 may annotate 318 the portion of the audio encounter information (e.g., transcript and/or medical report) where the model (e.g., sequence to sequence / abstractive summarization, etc.) model does not attribute any significant responsibility for the draft report to that excerpt of the conversation transcript's aligned audio. For instance, assume for example purposes only that Mrs. Smith (i.e., the patient in this example) asks the Doctor "By the way, how has your golf game been lately?", to which the Doctor replies "Oh it's been terrible. People are starting to call me "captain Hook!" In the example, ACD process 10 may determine that the discussion of the Doctor's golf game is not relevant for the draft report to that excerpt of the conversation transcript's aligned audio, and may annotate 318 (e.g., bold) that portion of the

transcript. This may be especially advantageous for direct editing by a physician, who typically would not listen to the audio if reviewing / finalizing the report directly subsequent to the patient encounter. Thus, ACD process 10 may annotate 318 the subset of the transcript for which the model does not (significantly) attribute any portion of the medical report, allowing the physician to quickly scan for clinically relevant information potentially inadvertently missed by the model (e.g., not captured in some form in the report).

[0071] As discussed above, automated clinical documentation (ACD) process 10 may be configured to automate the collection and processing of clinical encounter information to generate / store / distribute medical records. Accordingly, as discussed above and referring also at least to FIG. 9, ACD process 10 may be configured to obtain 300 encounter information (e.g., machine vision encounter information 102 and/or audio encounter information 106) of a patient encounter (e.g., a visit to a doctor's office) of at least a first encounter participant, wherein the encounter information may include audio encounter information obtained from at least a first encounter participant (e.g., encounter participant 228, 226, 230, and/or 242). ACD process 10 may further be configured to process 302 the encounter information (e.g., machine vision encounter information 102 and/or audio encounter information 106) obtained from at least the first encounter participant, e.g., to generate an encounter transcript (e.g., encounter transcript 234) and/or generate 305 a user interface displaying a plurality of layers associated with the audio encounter information obtained from at least the first encounter participant, wherein at least one of the plurality of layers is one of exposed to the user interface and not exposed to the user interface based upon, at least in part, a confidence level. In some implementations, ACD process 10 may process at least a portion of the encounter transcript (e.g., encounter transcript 234) to populate at least a portion of a medical record/report (e.g., medical record 236) associated with the patient encounter (e.g., the visit to the doctor's office). Encounter transcript 234 and/or medical record 236 may be reviewed by a medical professional involved with the

patient encounter (e.g., a visit to a doctor's office) to determine the accuracy of the same and/or make corrections to the same.

[0072] In some implementations, processing 302 the first audio encounter information may include defining 306 linkages between each of the plurality of layers associated with the audio encounter information. For example, the first layer of the plurality of layers may be a visualization and/or direct play-back of an audio signal associated with the audio encounter information (e.g., complete audio of the encounter, encompassing and clearly delineating each participant), wherein the second layer of the plurality of layers may be a transcript associated with the audio encounter information (e.g., a diarized audio transcript (*verbatim*) for each participant in the encounter), and wherein the third layer of the plurality of layers may be a medical report associated with the audio encounter information (e.g., a draft medical report in the appropriate clinical output format). In some implementations, additional layers may include, e.g., the above-noted machine vision-based recording of the encounter, including various signal formats and features, and discrete, standardized, actionable data resulting from the encounter, including, but not limited to medication plans (for example, in RxNorm) or lab orders (for example, LOINC) or diagnoses (for example, ICD10, CPT etc). In the example, the signals captured from the encounter information may be processed 304 into at least the above-noted three separate, yet closely linked and interdependent layers.

[0073] In some implementations, ACD process 10 may include an ASR portion that may process 302 the audio encounter information producing an approximate (e.g., diarized) verbatim transcript along with alignment information indicating the audio interval corresponding to each transcript word. In some implementations, a deep learning (e.g., sequence to sequence) model associated with ACD process 10 may convert the transcript to a medical report. It will be appreciated that various attribution techniques may be employed by ACD process 10 that may effectively softly assign responsibility for a given output (e.g., medical report) word to input (e.g., conversation

transcript) words (e.g. attention weights, integrated gradient, etc.) according to the model. As a result, this may provide a soft mapping from the transcript word positions to report word positions. In some implementations, the input word position assigned maximal attribution for a given output word may be interpreted as being aligned (linked) to that output (e.g., when a hard mapping is required). Based on the ASR time alignment, a word in the draft medical report, aligned to a word in the ASR conversation transcript, may now be associated with an audio time interval of the associated audio signal of the audio encounter information.

[0074] In some implementations, ACD process 10 may also may link (i.e., align) the ASR conversation transcript words with the draft medical report words. For transcript words that may have maximal attribution value for some set of medical report words, ACD process 10 may link them with the first word in that set. For the remaining transcript words, ACD process 10 may link them to the same word that the nearest preceding (or if none, nearest subsequent) conversation transcript word is linked to. As a result, since each of the plurality of layers may be synchronized, a scribe (or physician, etc.) may later playback the whole patient encounter, navigate in any of the layers, make edits, and would know which edit / correction belongs to the other sections (layers).

[0075] In some implementations, a visual recording (e.g., video stream of the patient encounter), if available, may also be a layer and may be time indexed and thus a given point in the recording may be associated with the same time in the audio recording and thus a conversation transcript word and draft report word. In some implementations, if discrete, standardized, actionable data is produced as a second (parallel) output sequence of the sequence to sequence model, then a similar model output attribution technique may be used to align tokens in this actionable data with the ASR conversation transcript words, and thus the audio intervals.

[0076] In some implementations, ACD process 10 may generate 305 a user interface displaying a plurality of layers associated with the audio encounter information obtained from at least the first encounter participant, wherein at least one

of the plurality of layers is one of exposed to the user interface and not exposed to the user interface based upon, at least in part, a confidence level. For example, and referring again at least to the example implementation of Fig. 5, an example user interface (e.g., user interface 500) is shown. It will be appreciated that UI 500 is shown merely for example purposes only, and that more or less features and arrangements of features may be used without departing from the scope of the present disclosure.

[0077] As can be seen from Fig. 5, UI 500 includes the first layer of the plurality of layers (e.g., first layer 502 which may be a visualization and/or direct play-back of an audio signal associated with the audio encounter information (e.g., complete audio of the encounter, encompassing and clearly delineating each participant), second layer 504 of the plurality of layers which may be a transcript (e.g., encounter transcript 234) associated with the audio encounter information (e.g., a diarized audio transcript (*verbatim*) for each participant in the encounter), and third layer 506 of the plurality of layers which may be a medical report (e.g., medical report/record 236) associated with the audio encounter information (e.g., a draft medical report in the appropriate clinical output format). It will be appreciated that that each layer (audio, transcript, draft report, etc.) may be rendered in multiple different ways, and in the most appropriate way for a given use / preference of the end-user. As such, the specific rendering of layers should be taken as example only and not to limit the scope of the disclosure.

[0078] As noted above, encounter transcript 234 (via second layer 504) and/or medical record/record 236 (via third layer 506) may be reviewed by a medical professional involved with the patient encounter (e.g., a visit to a doctor's office) to determine the accuracy of the same and/or make corrections to the same. For example, a scribe involved with (or assigned to) the patient encounter (e.g., a visit to a doctor's office) may review encounter transcript 234 and/or medical record 236 to confirm that the same was accurate and/or make corrections to the same. In the event that corrections are made to encounter transcript 234 and/or medical record 236, automated clinical documentation process 10 may utilize these corrections for training / tuning purposes

(e.g., to adjust the various profiles associated the participants of the patient encounter) to enhance the future accuracy / efficiency / performance of automated clinical documentation process 10.

[0079] Alternatively / additionally, a doctor involved with the patient encounter (e.g., a visit to a doctor's office) may review encounter transcript 234 and/or medical record 236 to confirm that the same was accurate and/or make corrections to the same. In the event that corrections are made to encounter transcript 234 and/or medical record 236, automated clinical documentation process 10 may utilize these corrections for training / tuning purposes (e.g., to adjust the various profiles associated the participants of the patient encounter) to enhance the future accuracy / efficiency / performance of automated clinical documentation process 10.

[0080] As noted above, at least one of the plurality of layers is one of exposed to the user interface and not exposed to the user interface based upon, at least in part, a confidence level. For example, ACD process 10 may employ a confidence model that may only expose (e.g., display) at least one layer (e.g., the medical report layer 506 or a section of the medical report layer section) to the user interface if ACD process 10 (via the confidence model) determines there is sufficient confidence, such as a threshold confidence level, that the draft report (or draft report section) will be faster than being typed from scratch. For instance, exposing (e.g., displaying) the medical report/section to UI 500 (e.g., for editing) may enable the ability to change some incorrectly labeled terms/sentences, which may be faster than manually typing out the entire section from the beginning; however, if so many terms/sentences are incorrect (and therefore need to be changed), it may be faster to simply type out the entire section from the beginning, rather than making substantial changes to the draft report section. As such, in the example where ACD process 10 determines an estimated time that it will take to make corrections compared to an estimated time that it will take to type out the entire section from the beginning, and determines it may be faster to simply type out the entire section from the beginning (rather than make corrections), ACD process 10 may not display

the report/section by not exposing the report/section to UI 500 in the first place. In some implementations, the use of the confidence level for exposure may be only for outputs expected to be part of the actual case record, e.g., the medical report layer and the optionally concurrently generated discrete / actionable structured data. In some implementations, the decision to expose or not expose may be per section of the medical report (or structured data), e.g., Exam section (report) or Lab Orders (structured), etc.

[0081] In some implementations, ACD process 10 may leverage various features of the confidence model, including the statistics (e.g., min, max, standard deviation, etc.) of the difference in probability between the first and second most likely word hypothesized at each position in the report/section via, e.g., a model (e.g., sequence to sequence model), the log likelihood of the draft report/section per the model (e.g., in addition to a length normalized version of this), confidence values from the ASR portion of ACD process 10 (e.g., particularly for segments to which the report/section content is attributed) and information regarding the typical (e.g., mean / standard deviation) number of edits required for a relevant physician and typical number of edits made by a relevant scribe. For draft reports/sections that ACD process 10 may actually expose to be edited (rather than typed from scratch), ACD process 10 may collect the number of edits made and time required to make them, which may be added to training and improvement of the confidence model over time.

[0082] For example, in some implementations, ACD process 10 may train 408 a model that maps the transcript to the medical report from pairs of prior transcripts and medical reports. For instance, the core model (e.g., sequence to sequence model) of ACD process 10 that may map the conversation transcript to the medical reports may be trained 408 from pairs of (e.g., ASR approximate) transcripts and medical reports. For example, as additional reports are generated / edited / typed by scribes / physicians, these additional reports may be automatically (or manually) added to the training pool for the sequence to sequence model, which may be automatically intermittently retrained. In some implementations, the model may be metadata-dependent (e.g.,

physician, institution, scribe, etc.), either by (e.g., regularized) adaptation to appropriate subsets of the training pool or by using this metadata as inputs or combination thereof. That is, the training pool may be specific to the user, type of user, or institution of the user making the changes to the medical report. In some implementations, the number and types of edits made by the scribe (or other user) may also be used by the confidence model to better predict the number of edits (and thus editing efficiency) for subsequent drafts.

[0083] In some implementations, e.g., towards the end of the patient encounter, as the confidence of the information regarding the medical report section rises, more items may be documented (e.g., added) in this section. For instance, assume for example purposes only that one of the encounter participants (e.g., the physician) mentions "examination." Using the techniques described above, ACD process 10 may recognize this fact, and may display some associated points about the physical examinations under the "Examination" section (e.g., in the medical report layer 506 shown at least in Fig. 5), which may be first grayed. In the example, as the physician continues to verbalize further associated points that make the confidence level of what was entered into the Examination section higher, the greyed points may become solid, and further adjustments / edits may be made by ACD process 10 shortly thereafter (e.g., in the following seconds). Similarly, related to the "Assessment/Plan" section (e.g., in the medical report layer 506) after the physician verbalizes his assessment, ACD process 10 may start documenting that section, first grayed, but then the grayed words may become more solid (e.g., darker or normal solid color) with more points being added as ACD process 10 knows more and the confidence level rises. In some implementations, showing the transition from grey to solid wording may help train scribes (or other users) to use ACD process 10 and recognize how it works.

[0084] In some implementations, ACD process 10 may update 410 an output of the user interface based upon, at least in part, one or more modifications made at least one layer of the plurality of layers. For instance, rather than the output being static during

the editing process, ACD process 10 may update 410 the output based on the modifications made by the editor so far (e.g., based on the decoder being autoregressive in nature). In particular, the decoder (e.g., sequence to sequence decoder) output may depend on its preceding output. As such, if the scribe (or other user) makes a correction in one part of the draft report, ACD process 10 may update a best guess at the subsequent content. In some implementations, this may be distracting to the user, and so to make it less distracting, it may be optionally limited to a toggle-able mode (e.g., online / synchronous vs. global review mode) and/or limited to only modifying the draft output for subsequent sections of the report. In some implementations, there may be at least four ways that corrections to case-record-persistent system outputs (e.g., medical reports, structured data) may be utilized. For example, pairs of ASR transcripts and corrected reports may be used for offline (sequence to sequence, transcript->report) model training. If in the typing acceleration mode as opposed to draft-report-correction mode, as the user types, the prediction for next sentence may be updated, which may not be model training / adaptation, but rather a reflection of the (auto-regressive) model predicting next output based on report content so far. If in the draft-report-correction mode, as the user makes corrections, prediction of subsequent content (sentences or perhaps less distracting, sections) and thus the draft report content (in this report) may be updated, again utilizing the same auto-regressive nature of the sequence to sequence model (and not some result of model training). Tuples of ASR transcripts (along with ASR confidence information), draft reports and noted user edits (e.g., corrections) may be used for an offline confidence model (e.g., required number of edits) training.

[0085] In some implementations, ACD process 10 may update 412 a prediction for a next sentence to be typed based upon, at least in part, current content of at least one layer of the plurality of layers, where in some implementations, the at least one layer of the plurality of layers may be the medical report. For example, for reports that the above-noted confidence model (via ACD process 10) determines are best typed from scratch rather than editing a draft, ACD process 10 may leverage the summarization

model to accelerate composition by having it predict the next sentence to type and updating that prediction based on, and to be consistent with, what the scribe (or other user) has typed so far. For instance, ACD process 10 (e.g., via UI 500) may allow the user to select the model's prediction of the current/next sentence being presented to the user in the report by selecting the tab key (or other shortcut) to complete the sentence (e.g., auto-completion based typing accelerant). In this way, ACD process 10 (e.g., via the model) may only predict the draft report a sentence at a time based on the scribes typing / feedback. In some implementations, the prediction may be based not just upon what the scribe is currently writing, but may additionally (or alternatively) be based upon the context of what the scribe has previous written in previous sentences of the report (or based upon a training pool specific to the scribe). In some implementations, the sentence prediction may be based on the same model that produces the draft medical report. In some implementations, when based on the confidence model and it is decided to suppress the draft medical report (section), the ACD process 10 based auto-completion typing accelerant may allow the salvage of some efficiency improvement.

[0086] As discussed above, automated clinical documentation (ACD) process 10 may be configured to automate the collection and processing of clinical encounter information to generate / store / distribute medical records. Accordingly and referring also at least to FIGS. 10-11, ACD process 10 may be configured to obtain 300 encounter information (e.g., machine vision encounter information 102 and/or audio encounter information 106) of a patient encounter (e.g., a visit to a doctor's office) of at least a first encounter participant, wherein the encounter information may include audio encounter information obtained from at least a first encounter participant (e.g., encounter participant 228, 226, 230, and/or 242). ACD process 10 may further be configured to process 302 the encounter information (e.g., machine vision encounter information 102 and/or audio encounter information 106) obtained from at least the first encounter participant, e.g., to generate an encounter transcript (e.g., encounter transcript 234) and/or generate 304 a user interface displaying a plurality of layers

associated with the audio encounter information obtained from at least the first encounter participant. In some implementations, ACD process 10 may receive 307 a user input from a peripheral device to navigate through each of the plurality of layers associated with the audio encounter information displayed on the user interface. In some implementations, ACD process 10 may process at least a portion of the encounter transcript (e.g., encounter transcript 234) to populate at least a portion of a medical record (e.g., medical record 236) associated with the patient encounter (e.g., the visit to the doctor's office). Encounter transcript 234 and/or medical record 236 may be reviewed by a medical professional involved with the patient encounter (e.g., a visit to a doctor's office) to determine the accuracy of the same and/or make corrections to the same.

[0087] As noted above, ACD process 10 may process 302 the audio encounter information obtained from at least the first encounter participant. In some implementations, processing 302 the first audio encounter information may include defining 306 linkages between each of the plurality of layers associated with the audio encounter information. For example, the first layer of the plurality of layers may be a visualization and/or direct play-back of an audio signal associated with the audio encounter information (e.g., complete audio of the encounter, encompassing and clearly delineating each participant), wherein the second layer of the plurality of layers may be a transcript associated with the audio encounter information (e.g., a diarized audio transcript (*verbatim*) for each participant in the encounter), and wherein the third layer of the plurality of layers may be a medical report associated with the audio encounter information (e.g., a draft medical report in the appropriate clinical output format). In some implementations, additional layers may include, e.g., the above-noted machine vision-based recording of the encounter, including various signal formats and features, and discrete, standardized, actionable data resulting from the encounter, including, but not limited to medication plans (for example, in RxNorm) or lab orders (for example, LOINC) or diagnoses (for example, ICD10, CPT etc). In the example, the signals

captured from the encounter information may be processed 304 into at least the above-noted three separate, yet closely linked and interdependent layers.

[0088] In some implementations, ACD process 10 may include an ASR portion that may process 302 the audio encounter information producing an approximate (e.g., diarized) verbatim transcript along with alignment information indicating the audio interval corresponding to each transcript word. In some implementations, a deep learning (e.g., sequence to sequence) model associated with ACD process 10 may convert the transcript to a medical report. It will be appreciated that various attribution techniques may be employed by ACD process 10 that may effectively softly assign responsibility for a given output (e.g., medical report) word to input (e.g., conversation transcript) words (e.g. attention weights, integrated gradient, etc.) according to the model. As a result, this may provide a soft mapping from the transcript word positions to report word positions. In some implementations, the input word position assigned maximal attribution for a given output word may be interpreted as being aligned (linked) to that output (e.g., when a hard mapping is required). Based on the ASR time alignment, a word in the draft medical report, aligned to a word in the ASR conversation transcript, may now be associated with an audio time interval of the associated audio signal of the audio encounter information.

[0089] In some implementations, ACD process 10 may also may link (i.e., align) the ASR conversation transcript words with the draft medical report words. For transcript words that may have maximal attribution value for some set of medical report words, ACD process 10 may link them with the first word in that set. For the remaining transcript words, ACD process 10 may link them to the same word that the nearest preceding (or if none, nearest subsequent) conversation transcript word is linked to.

[0090] In some implementations, a visual recording (e.g., video stream of the patient encounter), if available, may also be a layer and may be time indexed and thus a given point in the recording may be associated with the same time in the audio recording and thus a conversation transcript word and draft report word. In some

implementations, if discrete, standardized, actionable data is produced as a second (parallel) output sequence of the sequence to sequence model, then a similar model output attribution technique may be used to align tokens in this actionable data with the ASR conversation transcript words, and thus the audio intervals.

[0091] In some implementations, ACD process 10 may generate 304 a user interface displaying a plurality of layers associated with the audio encounter information obtained from at least the first encounter participant. For example, and referring again at least to the example implementation of Fig. 5, an example user interface (e.g., user interface 500) is shown. It will be appreciated that UI 500 is shown merely for example purposes only, and that more or less features and arrangements of features may be used without departing from the scope of the present disclosure.

[0092] As can be seen from Fig. 5, UI 500 includes the first layer of the plurality of layers (e.g., first layer 502 which may be a visualization and/or direct play-back of an audio signal associated with the audio encounter information (e.g., complete audio of the encounter, encompassing and clearly delineating each participant), second layer 504 of the plurality of layers which may be a transcript (e.g., encounter transcript 234) associated with the audio encounter information (e.g., a diarized audio transcript (*verbatim*) for each participant in the encounter), and third layer 506 of the plurality of layers which may be a medical report (e.g., medical report/record 236) associated with the audio encounter information (e.g., a draft medical report in the appropriate clinical output format). It will be appreciated that that each layer (audio, transcript, draft report, etc.) may be rendered in multiple different ways, and in the most appropriate way for a given use / preference of the end-user. As such, the specific rendering of layers should be taken as example only and not to limit the scope of the disclosure.

[0093] As noted above, encounter transcript 234 (via second layer 504) and/or medical record/record 236 (via third layer 506) may be reviewed by a medical professional involved with the patient encounter (e.g., a visit to a doctor's office) to determine the accuracy of the same and/or make corrections to the same. For example,

a scribe involved with (or assigned to) the patient encounter (e.g., a visit to a doctor's office) may review encounter transcript 234 and/or medical record 236 to confirm that the same was accurate and/or make corrections to the same. In the event that corrections are made to encounter transcript 234 and/or medical record 236, automated clinical documentation process 10 may utilize these corrections for training / tuning purposes (e.g., to adjust the various profiles associated the participants of the patient encounter) to enhance the future accuracy / efficiency / performance of automated clinical documentation process 10.

[0094] Alternatively / additionally, a doctor involved with the patient encounter (e.g., a visit to a doctor's office) may review encounter transcript 234 and/or medical record 236 to confirm that the same was accurate and/or make corrections to the same. In the event that corrections are made to encounter transcript 234 and/or medical record 236, automated clinical documentation process 10 may utilize these corrections for training / tuning purposes (e.g., to adjust the various profiles associated the participants of the patient encounter) to enhance the future accuracy / efficiency / performance of automated clinical documentation process 10.

[0095] Therefore, and referring at least to the example implementation of Fig. 11, to help review encounter transcript 234 and/or medical record 236 to confirm that the same was accurate and/or make corrections to the same, in some implementations, ACD process 10 may receive 307 a user input from a peripheral device to navigate through each of the plurality of layers associated with the audio encounter information displayed on the user interface. In some implementations, the peripheral device may include, e.g., a keyboard (e.g., keyboard 1100), a pointing device (e.g., mouse 1102), a foot pedal (e.g., foot pedal 1104), a dial (e.g., dial 1106), or combination thereof. It will be appreciated that other peripheral devices may be used without departing from the scope of the present disclosure. One or more of the peripheral devices may be physically or wirelessly connected to the computing device displaying UI 500.

[0096] As such, and referring again at least to the example implementation of Fig.

6, assume for example purposes only that a scribe is reviewing the encounter information using UI 500. In some implementations, receiving 307 the user input from the peripheral device may include receiving 309, via the user input from the peripheral device, a selection of a first portion of the audio encounter information at a first layer of the plurality of layers on the user interface, and ACD process 10 may display 310 an annotation of at least one of a second layer of the plurality of layers and a third layer of the plurality of layers corresponding to the first portion of the audio encounter information of the first layer of the plurality of layers selected on the user interface. For instance, the scribe may use, e.g., any of the example peripheral devices, to select a portion of first layer 502 (i.e., the audio information of the audio encounter information), which may be received 309 by ACD process 10. As a result, in some implementations, ACD process 10 may display 310 some kind of annotation of either the second layer (e.g., the portion of the transcript associated with the selected portion of the first layer) and/or the third layer (e.g., the portion of the medical report associated with the selected portion of the first layer). In the example, ACD process 10 has displayed an annotation (e.g., bold font) of the second layer (e.g., the portion of the transcript associated with the selected portion of the first layer) and the third layer (e.g., the portion of the medical report associated with the selected portion of the first layer). It will be appreciated that any type of annotation may be used (e.g., bold, italics, underline, highlights, shading, transparency, etc.) without departing from the scope of the present disclosure. As such, the use of bolding as the annotation should be taken as example only and not to otherwise limit the scope of the disclosure.

[0097] Similarly, in some implantations, receiving 307 the user input from the peripheral device may include receiving 313, via the user input from the peripheral device, a selection of the first portion of the audio encounter information at one of the second layer of the plurality of layers and the third layer of the plurality of layers on the user interface, and ACD process 10 may provide 314 audio of the first layer corresponding to the first portion of the audio encounter information of one of the

second layer of the plurality of layers and the third layer of the plurality of layers selected on the user interface. For example, and referring again at least to the example implementation of Fig. 7, the scribe may use, e.g., any of the example peripheral devices, to select a portion of second layer 504 (e.g., the transcript), and/or the third layer (e.g., the medical report) which may be received 312 by ACD process 10. As a result, in some implementations, ACD process 10 may provide 314 the audio of first layer 502 (e.g., the audio information of the audio encounter associated with the selected portion of second layer 504).

[0098] As noted above, the physician (or scribe) may need to verify the correctness of the medical report created by ACD process 10. This may involve, e.g., checking for correctness of each sentence/word in the medical report generated by looking for evidence in the conversation transcript (attribution), checking the correctness of the conversational transcript generated by an ASR) by listening to the audio segment. The use of dedicated hardware (e.g., a peripheral device) may be utilized by ACD process 10 to improve editing efficiency for medical transcriptionists / physicians, etc. to help navigate and browse (e.g., sentence by sentence) through the draft medical report, the associated conversation transcript excerpt, and/or associated audio file, annotated (e.g., highlighted, bolded, etc.) with its corresponding audio cued up for easy playback.

[0099] For example, in some implementations, the user input from the peripheral device may include a keyboard shortcut when the peripheral device is keyboard 1100. For example, default and/or user defined keyboard shortcuts may be used to interact with UI 500. For instance, a combination of keys (e.g., control + shift + S) may execute a particular command for UI 500. For instance, such a shortcut may cause UI 500 to, e.g., switch between sentences in an output of the medical report, switch between sections in the output of the medical report, switch between the medical report and the transcript, provide/playback audio of the audio signal, cease/stop playback of the audio of the audio signal, speeding up playback of the audio of the audio signal, or slow down playback of the audio of the audio signal. It will be appreciated that various other

keyboard shortcuts, as well as single key inputs (e.g., arrows, spacebar, etc.) may execute the above-noted commands for UI 500 without departing from the scope of the disclosure.

[00100] In some implementations, the user input from the peripheral device may include a pointing device action when the peripheral device is pointing device 1102. For example, default and/or user defined mouse actions may be used to interact with UI 500. For instance, movement of the mouse cursor over a particular portion of one of the above-noted layers, clicking or double clicking on a particular portion of one of the above-noted layers, hovering over a particular portion of one of the above-noted layers, using a scroll wheel of the mouse, etc. may execute a particular command for UI 500. For instance, any one of the example uses of the mouse may cause UI 500 to, e.g., switch between sentences in an output of the medical report, switch between sections in the output of the medical report, switch between the medical report and the transcript, provide/playback audio of the audio signal, cease/stop playback of the audio of the audio signal, speeding up playback of the audio of the audio signal, or slow down playback of the audio of the audio signal. It will be appreciated that various other keyboard shortcuts, as well as single key inputs (e.g., arrows, spacebar, etc.) may execute the above-noted commands for UI 500 without departing from the scope of the disclosure.

[00101] In some implementations, the user input from the peripheral device may include raising and lowering of the foot pedal when the peripheral device is foot pedal 1104. For example, default and/or user defined foot pedal actions may be used to interact with UI 500. For instance, raising and lowering of the foot pedal (similar to how the foot pedal in a vehicle may function) may execute a particular command for UI 500. For instance, lowering the pedal (e.g., using the user's foot to press down on pedal 1104) may cause UI 500 to, e.g., provide/playback audio of the audio signal, and may speed up playback of the audio of the audio signal depending on the amount the pedal is lowered. Conversely, raising the pedal (e.g., lifting the user's foot off the pedal

1104) may cause UI 500 to, e.g., cease/stop playback of the audio of the audio signal, and may or slow down playback of the audio of the audio signal depending on the amount the pedal is raised.

[00102] In some implementations, the user input from the peripheral device may include at least one of a rotating action, an up action, a down action, a left action, a right action, and a pressing action of the dial when the peripheral device is dial 1106. For example, default and/or user defined dial actions may be used to interact with UI 500. For instance, UI 500 (via ACD process 10) may switch the focus between sentences in the output medical report by rotating the dial, e.g., one click clockwise may move the focus to the next sentence and counter clockwise may move the focus to the previous sentence. As another example, UI 500 (via ACD process 10) may switch the focus between different sections of the medical report using the up/down arrows on dial 1106. Similarly, as discussed above at least with regard to Figs. 6 and 7, evidence in conversational transcripts for the selected sentence in the report/transcript may be annotated (e.g., highlighted, bolded, etc.). Evidence may include a word or phrase or sentence and may be obtained from attention weights (e.g., byproduct of a sequence to sequence based summarization system) or other dedicated models for attribution. As another example, UI 500 (via ACD process 10) may switch focus from the medical report to the conversational transcript using, e.g., left/right arrows on dial 1106. As another example, UI 500 (via ACD process 10) may browse may switch the focus between sentences in the transcript by rotating the dial, e.g., one click clockwise may move the focus to the next sentence and counter clockwise may move the focus to the previous sentence. As yet another example, UI 500 (via ACD process 10) may, when the annotated sentence in the transcript is selected by pressing dial 1106, play back the corresponding audio chunk (e.g., identified by time stamps from the ASR decoding).

[00103] It will be appreciated that any of the example peripheral devices may be used in any combination with each other's functions without departing from the scope of the disclosure. For example, using pointing device 1102 to hover over a particular

portion of the transcript, medical report section, or audio signal may be the starting location where, the functions of e.g., dial 1106 may be used. As such, the use of any particular peripheral device and its particular function should be taken as example only and not to otherwise limit the scope of the disclosure.

[00104] As such, ACD process 10 may leverage the above-noted defined linkages and peripheral devices, such that the encounter recording as captured above may be provided to a human scribe (or physician) in such a fashion, so that they may navigate through each of these layers independently and be always in sync. For instance, if the scribe navigates through first layer 502 (audio) by listening, the part of the transcript that corresponds to that part of the audio may be bolded, highlighted, etc. accordingly, as well as the resulting medical report (and if relevant actionable data) from that information. Additionally / alternatively, if the scribe selects a part of the report, the audio / transcript sections that affected that part of the report may be bolded, highlighted, emphasized, etc. This annotation also may correspond to the audio and report "cursor" (e.g., audio cursor 510) moving, e.g., audio may next play from that position and typing / keyboard-navigation may commence from that position in the report. As a result, since each of the plurality of layers may be synchronized, a scribe (or physician, etc.) may later playback the whole patient encounter, navigate (e.g., "clicking" a mouse cursor of other pointing device at any position) in any portion of any of the layers, make edits, and would know which edit / correction belongs to the other sections (layers). More generally, "clicking" (i.e. resetting the cursor) in one viewed layer may annotate and update the cursor in the linked points / excerpts in the other layers.

[00105] As discussed above, automated clinical documentation (ACD) process 10 may be configured to automate the collection and processing of clinical encounter information to generate / store / distribute medical records. Accordingly, as discussed above and referring also at least to FIGS. 12-14, ACD process 10 may be configured to obtain 300 encounter information (e.g., machine vision encounter information 102

and/or audio encounter information 106) of a patient encounter (e.g., a visit to a doctor's office) of at least a first encounter participant, wherein the encounter information may include audio encounter information obtained from at least a first encounter participant (e.g., encounter participant 228, 226, 230, and/or 242). ACD process 10 may further be configured to process 302 the encounter information (e.g., machine vision encounter information 102 and/or audio encounter information 106) obtained from at least the first encounter participant, e.g., to generate an encounter transcript (e.g., encounter transcript 234) and/or generate 505 a user interface displaying a plurality of layers associated with the audio encounter information obtained from at least the first encounter participant, wherein at least a portion of one of the plurality of layers is added to a section of the one of the plurality of layers based upon, at least in part, a confidence level. In some implementations, ACD process 10 may process at least a portion of the encounter transcript (e.g., encounter transcript 234) to populate at least a portion of a medical record/report (e.g., medical record 236) associated with the patient encounter (e.g., the visit to the doctor's office). Encounter transcript 234 and/or medical record 236 may be reviewed by a medical professional involved with the patient encounter (e.g., a visit to a doctor's office) to determine the accuracy of the same and/or make corrections to the same.

[00106] In some implementations, processing 302 the first audio encounter information may include defining 306 linkages between each of the plurality of layers associated with the audio encounter information. For example, the first layer of the plurality of layers may be a visualization and/or direct play-back of an audio signal associated with the audio encounter information (e.g., complete audio of the encounter, encompassing and clearly delineating each participant), wherein the second layer of the plurality of layers may be a transcript associated with the audio encounter information (e.g., a diarized audio transcript (*verbatim*) for each participant in the encounter), and wherein the third layer of the plurality of layers may be a medical report associated with the audio encounter information (e.g., a draft medical report in the appropriate clinical

output format). In some implementations, additional layers may include, e.g., the above-noted machine vision-based recording of the encounter, including various signal formats and features, and discrete, standardized, actionable data resulting from the encounter, including, but not limited to medication plans (for example, in RxNorm) or lab orders (for example, LOINC) or diagnoses (for example, ICD10, CPT etc). In the example, the signals captured from the encounter information may be processed 304 into at least the above-noted three separate, yet closely linked and interdependent layers.

[00107] In some implementations, ACD process 10 may include an ASR portion that may process 302 the audio encounter information producing an approximate (e.g., diarized) verbatim transcript along with alignment information indicating the audio interval corresponding to each transcript word. In some implementations, a deep learning (e.g., sequence to sequence) model associated with ACD process 10 may convert the transcript to a medical report. It will be appreciated that various attribution techniques may be employed by ACD process 10 that may effectively softly assign responsibility for a given output (e.g., medical report) word to input (e.g., conversation transcript) words (e.g. attention weights, integrated gradient, etc.) according to the model. As a result, this may provide a soft mapping from the transcript word positions to report word positions. In some implementations, the input word position assigned maximal attribution for a given output word may be interpreted as being aligned (linked) to that output (e.g., when a hard mapping is required). Based on the ASR time alignment, a word in the draft medical report, aligned to a word in the ASR conversation transcript, may now be associated with an audio time interval of the associated audio signal of the audio encounter information.

[00108] In some implementations, ACD process 10 may also may link (i.e., align) the ASR conversation transcript words with the draft medical report words. For transcript words that may have maximal attribution value for some set of medical report words, ACD process 10 may link them with the first word in that set. For the remaining

transcript words, ACD process 10 may link them to the same word that the nearest preceding (or if none, nearest subsequent) conversation transcript word is linked to. As a result, since each of the plurality of layers may be synchronized, a scribe (or physician, etc.) may later playback the whole patient encounter, navigate in any of the layers, make edits, and would know which edit / correction belongs to the other sections (layers).

[00109] In some implementations, a visual recording (e.g., video stream of the patient encounter), if available, may also be a layer and may be time indexed and thus a given point in the recording may be associated with the same time in the audio recording and thus a conversation transcript word and draft report word. In some implementations, if discrete, standardized, actionable data is produced as a second (parallel) output sequence of the sequence to sequence model, then a similar model output attribution technique may be used to align tokens in this actionable data with the ASR conversation transcript words, and thus the audio intervals.

[00110] As noted above, in some implementations, ACD process 10 may generate a user interface displaying a plurality of layers associated with the audio encounter information obtained from at least the first encounter participant. For example, and referring again at least to the example implementation of Fig. 5, an example user interface (e.g., user interface 500) is shown. It will be appreciated that UI 500 is shown merely for example purposes only, and that more or less features and arrangements of features may be used without departing from the scope of the present disclosure.

[00111] As can be seen from Fig. 5, UI 500 includes the first layer of the plurality of layers (e.g., first layer 502 which may be a visualization and/or direct play-back of an audio signal associated with the audio encounter information (e.g., complete audio of the encounter, encompassing and clearly delineating each participant), second layer 504 of the plurality of layers which may be a transcript (e.g., encounter transcript 234) associated with the audio encounter information (e.g., a diarized audio transcript (*verbatim*) for each participant in the encounter), and third layer 506 of the plurality of

layers which may be a medical report (e.g., medical report/record 236) associated with the audio encounter information (e.g., a draft medical report in the appropriate clinical output format). It will be appreciated that that each layer (audio, transcript, draft report, etc.) may be rendered in multiple different ways, and in the most appropriate way for a given use / preference of the end-user. As such, the specific rendering of layers should be taken as example only and not to limit the scope of the disclosure.

[00112] As noted above, encounter transcript 234 (via second layer 504) and/or medical record/record 236 (via third layer 506) may be reviewed by a medical professional involved with the patient encounter (e.g., a visit to a doctor's office) to determine the accuracy of the same and/or make corrections to the same. For example, a scribe involved with (or assigned to) the patient encounter (e.g., a visit to a doctor's office) may review encounter transcript 234 and/or medical record 236 to confirm that the same was accurate and/or make corrections to the same. In the event that corrections are made to encounter transcript 234 and/or medical record 236, automated clinical documentation process 10 may utilize these corrections for training / tuning purposes (e.g., to adjust the various profiles associated the participants of the patient encounter) to enhance the future accuracy / efficiency / performance of automated clinical documentation process 10.

[00113] Alternatively / additionally, a doctor involved with the patient encounter (e.g., a visit to a doctor's office) may review encounter transcript 234 and/or medical record 236 to confirm that the same was accurate and/or make corrections to the same. In the event that corrections are made to encounter transcript 234 and/or medical record 236, automated clinical documentation process 10 may utilize these corrections for training / tuning purposes (e.g., to adjust the various profiles associated the participants of the patient encounter) to enhance the future accuracy / efficiency / performance of automated clinical documentation process 10.

[00114] As noted above, at least one of the plurality of layers may be one of exposed to the user interface and not exposed to the user interface based upon, at least

in part, a confidence level. For example, ACD process 10 may employ a confidence model that may only expose (e.g., display) at least one layer (e.g., the medical report layer 506 or a section of the medical report layer section) to the user interface if ACD process 10 (via the confidence model) determines there is sufficient confidence, such as a threshold confidence level, that the draft report (or draft report section) will be faster than being typed from scratch. For instance, exposing (e.g., displaying) the medical report/section to UI 500 (e.g., for editing) may enable the ability to change some incorrectly labeled terms/sentences, which may be faster than manually typing out the entire section from the beginning; however, if so many terms/sentences are incorrect (and therefore need to be changed), it may be faster to simply type out the entire section from the beginning, rather than making substantial changes to the draft report section. As such, in the example where ACD process 10 determines an estimated time that it will take to make corrections compared to an estimated time that it will take to type out the entire section from the beginning, and determines it may be faster to simply type out the entire section from the beginning (rather than make corrections), ACD process 10 may not display the report/section by not exposing the report/section to UI 500 in the first place.

[00115] In some implementations, ACD process 10 may leverage various features of the confidence model, including the statistics (e.g., min, max, standard deviation, etc.) of the difference in probability between the first and second most likely word hypothesized at each position in the report/section via, e.g., a model (e.g., sequence to sequence model), the log likelihood of the draft report/section per the model (e.g., in addition to a length normalized version of this), confidence values from the ASR portion of ACD process 10 (e.g., particularly for segments to which the report/section content is attributed) and information regarding the typical (e.g., mean / standard deviation) number of edits required for a relevant physician and typical number of edits made by a relevant scribe. For draft reports/sections that ACD process 10 may actually expose to be edited (rather than typed from scratch), ACD process 10

may collect the number of edits made and time required to make them, which may be added to training and improvement of the confidence model over time.

[00116] For example, in some implementations, ACD process 10 may train 408 a model that maps the transcript to the medical report from pairs of prior transcripts and medical reports. For instance, the core model (e.g., sequence to sequence model) of ACD process 10 that may map the conversation transcript to the medical reports may be trained 408 from pairs of (e.g., ASR approximate) transcripts and medical reports. For example, as additional reports are generated / edited / typed by scribes / physicians, these additional reports may be automatically (or manually) added to the training pool for the sequence to sequence model, which may be automatically intermittently retrained. In some implementations, the model may be metadata-dependent (e.g., physician, institution, scribe, etc.), either by (e.g., regularized) adaptation to appropriate subsets of the training pool or by using this metadata as inputs or combination thereof. That is, the training pool may be specific to the user, type of user, or institution of the user making the changes to the medical report. In some implementations, the number and types of edits made by the scribe (or other user) may also be used by the confidence model to better predict the number of edits (and thus editing efficiency) for subsequent drafts.

[00117] As noted above, in some implementations, ACD process 10 generate 505 UI 500 displaying the plurality of layers associated with the audio encounter information obtained from at least the first encounter participant, and in some implementations, ACD process 10 may annotate 509 at least the portion of the one of the plurality of layers added to the section of the one of the plurality of layers based upon, at least in part, the confidence level. For example, in some implementations, e.g., during a patient encounter, toward the end of the patient encounter, or during scribe training during playback, as the confidence of the information regarding the medical report section rises, more items may be documented (e.g., added) in this section. For instance, assume for example purposes only that one of the encounter participants (e.g.,

the physician) mentions "assessment" or "plan" or alternatively ACD process 10 may use NLU to determine from the conversational transcript 234 that the context of what the physician is saying (e.g., "and come back for a follow up in two weeks") may apply to the "Assessment/Plan" section (e.g., in medical report layer 506) of medical report 236. Using the techniques described above, ACD process 10 may recognize this, and may add (display) some associated points about the recommendation under the "Assessment/Plan" section (e.g., in the medical report layer 506 shown at least in the example implementation of UI 500 in Fig. 13), which may be first added and grayed based upon a first confidence level of what was said. That is, "recommending she come back for a follow up in 2 weeks" (shown at portion 1300a) may be first be added and grayed based upon the first confidence level being below a threshold confidence level. In the example, and referring at least to the example implementation of UI 500 shown in Fig. 14, as the physician continues to verbalize further associated points of what was respectively entered into the Assessment/Plan section, which therefore makes the confidence level (e.g., an updated confidence level above the threshold confidence level) of what was entered become higher, some of the grayed points (e.g., "recommending she come back" shown at portion 1300b) associated with the higher confidence level may be annotated (e.g., become solid or other type of annotation), or further adjustments / edits may be made by ACD process 10 shortly thereafter (e.g., in the following seconds) to correct the grayed portion based on the confidence level now taking into account the additional context of the further verbalization, which may then become solid. As such, after the physician verbalizes his assessment, ACD process 10 may start documenting that section, first grayed, but then the grayed words may become more solid (e.g., darker or normal solid color) with more points being added and becoming more solid as ACD process 10 knows more and the confidence level of what was said rises. In some implementations, showing the transition from gray to solid (or other annotation) wording may help train scribes (or other users) to use ACD process 10 and recognize how it works; however, it will be appreciated that the annotation of

wording based upon confidence level may occur at any stage of the patient encounter and/or post patient encounter during scribe/physician playback/editing.

[00118] In some implementations, ACD process 10 may update 410 an output of the user interface based upon, at least in part, one or more modifications made at least one layer of the plurality of layers. For instance, rather than the output being static during the editing process, ACD process 10 may update 410 the output based on the modifications made by the editor so far (e.g., based on the decoder being autoregressive in nature). In particular, the decoder (e.g., sequence to sequence decoder) output may depend on its preceding output. As such, if the scribe (or other user) makes a correction in one part of the draft report, ACD process 10 may update a best guess at the subsequent content. In some implementations, this may be distracting to the user, and so to make it less distracting, it may be optionally limited to a toggle-able mode (e.g., online / synchronous vs. global review mode) and/or limited to only modifying the draft output for subsequent sections of the report.

[00119] In some implementations, ACD process 10 may update 412 a prediction for a next sentence to be typed based upon, at least in part, current content of at least one layer of the plurality of layers, where in some implementations, the at least one layer of the plurality of layers may be the medical report. For example, for reports that the above-noted confidence model (via ACD process 10) determines are best typed from scratch rather than editing a draft, ACD process 10 may leverage the summarization model to accelerate composition by having it predict the next sentence to type and updating that prediction based on, and to be consistent with, what the scribe (or other user) has typed so far. For instance, ACD process 10 (e.g., via UI 500) may allow the user to select the model's prediction of the current/next sentence being presented to the user in the report by selecting the tab key (or other shortcut) to complete the sentence. In this way, ACD process 10 (e.g., via the model) may only predict the draft report a sentence at a time based on the scribes typing / feedback. In some implementations, the prediction may be based not just upon what the scribe is currently

writing, but may additionally (or alternatively) be based upon the context of what the scribe has previous written in previous sentences of the report (or based upon a training pool specific to the scribe).As discussed above,

[00120] As discussed above, automated clinical documentation (ACD) process 10 may be configured to automate the collection and processing of clinical encounter information to generate / store / distribute medical records. Additionally, as discussed above and referring also at least to the example implementation of FIG. 15, ACD process 10 may be configured to obtain 300 encounter information (e.g., machine vision encounter information 102 and/or audio encounter information 106) of a patient encounter (e.g., a visit to a doctor's office) of at least a first encounter participant, wherein the encounter information may include audio encounter information obtained from at least a first encounter participant (e.g., encounter participant 228, 226, 230, and/or 242). ACD process 10 may further be configured to process 302 the encounter information (e.g., machine vision encounter information 102 and/or audio encounter information 106) obtained from at least the first encounter participant, e.g., generate 602 an alert to obtain additional encounter information of the patient encounter, to generate an encounter transcript (e.g., encounter transcript 234) and/or generate 304 a user interface displaying a plurality of layers associated with the audio encounter information obtained from at least the first encounter participant. In some implementations, ACD process 10 may process at least a portion of the encounter transcript (e.g., encounter transcript 234) to populate at least a portion of a medical record (e.g., medical record 236) associated with the patient encounter (e.g., the visit to the doctor's office). Encounter transcript 234 and/or medical record 236 may be reviewed by a medical professional involved with the patient encounter (e.g., a visit to a doctor's office) to determine the accuracy of the same and/or make corrections to the same.

[00121] As noted above, ACD process 10 may process 302 the audio encounter information obtained from at least the first encounter participant. In some

implementations, processing 302 the first audio encounter information may include defining 306 linkages between each of the plurality of layers associated with the audio encounter information. For example, the first layer of the plurality of layers may be a visualization and/or direct play-back of an audio signal associated with the audio encounter information (e.g., complete audio of the encounter, encompassing and clearly delineating each participant), wherein the second layer of the plurality of layers may be a transcript associated with the audio encounter information (e.g., a diarized audio transcript (*verbatim*) for each participant in the encounter), and wherein the third layer of the plurality of layers may be a medical report associated with the audio encounter information (e.g., a draft medical report in the appropriate clinical output format). In some implementations, additional layers may include, e.g., the above-noted machine vision-based recording of the encounter, including various signal formats and features, and discrete, standardized, actionable data resulting from the encounter, including, but not limited to medication plans (for example, in RxNorm) or lab orders (for example, LOINC) or diagnoses (for example, ICD10, CPT etc). In the example, the signals captured from the encounter information may be processed 304 into at least the above-noted three separate, yet closely linked and interdependent layers.

[00122] In some implementations, ACD process 10 may include an ASR portion that may process 302 the audio encounter information producing an approximate (e.g., diarized) verbatim transcript along with alignment information indicating the audio interval corresponding to each transcript word. In some implementations, a deep learning (e.g., sequence to sequence) model associated with ACD process 10 may convert the transcript to a medical report. It will be appreciated that various attribution techniques may be employed by ACD process 10 that may effectively softly assign responsibility for a given output (e.g., medical report) word to input (e.g., conversation transcript) words (e.g. attention weights, integrated gradient, etc.) according to the model. As a result, this may provide a soft mapping from the transcript word positions to report word positions. In some implementations, the input word position assigned

maximal attribution for a given output word may be interpreted as being aligned (linked) to that output (e.g., when a hard mapping is required). Based on the ASR time alignment, a word in the draft medical report, aligned to a word in the ASR conversation transcript, may now be associated with an audio time interval of the associated audio signal of the audio encounter information.

[00123] In some implementations, ACD process 10 may also may link (i.e., align) the ASR conversation transcript words with the draft medical report words. For transcript words that may have maximal attribution value for some set of medical report words, ACD process 10 may link them with the first word in that set. For the remaining transcript words, ACD process 10 may link them to the same word that the nearest preceding (or if none, nearest subsequent) conversation transcript word is linked to.

[00124] In some implementations, a visual recording (e.g., video stream of the patient encounter), if available, may also be a layer and may be time indexed and thus a given point in the recording may be associated with the same time in the audio recording and thus a conversation transcript word and draft report word. In some implementations, if discrete, standardized, actionable data is produced as a second (parallel) output sequence of the sequence to sequence model, then a similar model output attribution technique may be used to align tokens in this actionable data with the ASR conversation transcript words, and thus the audio intervals.

[00125] In some implementations, ACD process 10 may generate 304 a user interface displaying a plurality of layers associated with the audio encounter information obtained from at least the first encounter participant. For example, and referring again at least to the example implementation of Fig. 5, an example user interface (e.g., user interface 500) is shown. It will be appreciated that UI 500 is shown merely for example purposes only, and that more or less features and arrangements of features may be used without departing from the scope of the present disclosure.

[00126] As can be seen from Fig. 5, UI 500 includes the first layer of the plurality of layers (e.g., first layer 502 which may be a visualization and/or direct play-back of

an audio signal associated with the audio encounter information (e.g., complete audio of the encounter, encompassing and clearly delineating each participant), second layer 504 of the plurality of layers which may be a transcript (e.g., encounter transcript 234) associated with the audio encounter information (e.g., a diarized audio transcript (*verbatim*) for each participant in the encounter), and third layer 506 of the plurality of layers which may be a medical report (e.g., medical report/record 236) associated with the audio encounter information (e.g., a draft medical report in the appropriate clinical output format). It will be appreciated that that each layer (audio, transcript, draft report, etc.) may be rendered in multiple different ways, and in the most appropriate way for a given use / preference of the end-user. As such, the specific rendering of layers should be taken as example only and not to limit the scope of the disclosure.

[00127] As noted above, encounter transcript 234 (via second layer 504) and/or medical record/record 236 (via third layer 506) may be reviewed by a medical professional involved with the patient encounter (e.g., a visit to a doctor's office) to determine the accuracy of the same and/or make corrections to the same. For example, a scribe involved with (or assigned to) the patient encounter (e.g., a visit to a doctor's office) may review encounter transcript 234 and/or medical record 236 to confirm that the same was accurate and/or make corrections to the same. In the event that corrections are made to encounter transcript 234 and/or medical record 236, automated clinical documentation process 10 may utilize these corrections for training / tuning purposes (e.g., to adjust the various profiles associated the participants of the patient encounter) to enhance the future accuracy / efficiency / performance of automated clinical documentation process 10.

[00128] Alternatively / additionally, a doctor involved with the patient encounter (e.g., a visit to a doctor's office) may review encounter transcript 234 and/or medical record 236 to confirm that the same was accurate and/or make corrections to the same. In the event that corrections are made to encounter transcript 234 and/or medical record 236, automated clinical documentation process 10 may utilize these corrections for

training / tuning purposes (e.g., to adjust the various profiles associated the participants of the patient encounter) to enhance the future accuracy / efficiency / performance of automated clinical documentation process 10.

[00129] As such, and referring again at least to the example implementation of Fig. 6, assume for example purposes only that a scribe is reviewing the encounter information using UI 500. In some implementations, ACD process 10 may receive a selection of a first portion of the audio encounter information at a first layer of the plurality of layers on the user interface, and ACD process 10 may display an annotation of at least one of a second layer of the plurality of layers and a third layer of the plurality of layers corresponding to the first portion of the audio encounter information of the first layer of the plurality of layers selected on the user interface. For instance, the scribe may use, e.g., cursor 508 or other known input techniques, to select a portion of first layer 502 (i.e., the audio information of the audio encounter information), which may be received by ACD process 10. As a result, in some implementations, ACD process 10 may display some kind of annotation of either the second layer (e.g., the portion of the transcript associated with the selected portion of the first layer) and/or the third layer (e.g., the portion of the medical report associated with the selected portion of the first layer). In the example, ACD process 10 has displayed an annotation (e.g., bold font) of the second layer (e.g., the portion of the transcript associated with the selected portion of the first layer) and the third layer (e.g., the portion of the medical report associated with the selected portion of the first layer). It will be appreciated that any type of annotation may be used (e.g., bold, italics, underline, highlights, shading, transparency, etc.) without departing from the scope of the present disclosure. As such, the use of bolding as the annotation should be taken as example only and not to otherwise limit the scope of the disclosure.

[00130] Similarly, in some implantations, ACD process 10 may receive a selection of the first portion of the audio encounter information at one of the second layer of the plurality of layers and the third layer of the plurality of layers on the user

interface, and ACD process 10 may provide audio of the first layer corresponding to the first portion of the audio encounter information of one of the second layer of the plurality of layers and the third layer of the plurality of layers selected on the user interface. For example, and referring to the example implementation of Fig. 7, the scribe may use, e.g., cursor 508 or other known input techniques, to select a portion of second layer 504 (e.g., the transcript), and/or the third layer (e.g., the medical report) which may be received by ACD process 10. As a result, in some implementations, ACD process 10 may provide the audio of first layer 502 (e.g., the audio information of the audio encounter associated with the selected portion of second layer 504).

[00131] As such, ACD process 10 may leverage the above-noted defined linkages, such that the encounter recording as captured above may be provided to a human scribe (or physician) in such a fashion, so that they may navigate through each of these layers independently and be always in sync. For instance, if the scribe navigates through first layer 502 (audio) by listening, the part of the transcript that corresponds to that part of the audio may be bolded, highlighted, etc. accordingly, as well as the resulting medical report (and if relevant actionable data) from that information. Additionally / alternatively, if the scribe selects a part of the report, the audio / transcript sections that affected that part of the report may be bolded, highlighted, emphasized, etc. This annotation also may correspond to the audio and report "cursor" (e.g., audio cursor 510) moving, e.g., audio may next play from that position and typing / keyboard-navigation may commence from that position in the report. As a result, since each of the plurality of layers may be synchronized, a scribe (or physician, etc.) may later playback the whole patient encounter, navigate (e.g., "clicking" a mouse cursor of other pointing device at any position) in any portion of any of the layers, make edits, and would know which edit / correction belongs to the other sections (layers). More generally, "clicking" (i.e. resetting the cursor) in one viewed layer may annotate and update the cursor in the linked points / excerpts in the other layers.

[00132] As noted above, ACD process 10 may generate 602 an alert (e.g., an audio alert or a visual alert such as a pop-up message, text message, email, etc.) to obtain additional encounter information of the patient encounter. For example, ACD process 10 may be aware of the relevant sections of the above-noted medical report sections when processing the obtained encounter information, as well as explicit data points (such as physical exam findings, lab orders, etc.). In the example, as will be discussed below, ACD process 10 may determine that, e.g., additional information may be needed from the patient, since some of the current encounter information obtained so far may be vague, missing, could not be documented/verified (e.g., neurological section, extremities, etc.), or information which would make the patient encounter more compliant to clinical / administrative guidelines. In some implementations, such as the above example, ACD process 10 may generate 602 an alert (e.g., for the physician) to obtain additional encounter information of the patient encounter to address the encounter information deficiencies.

[00133] In some implementations, the alert may be generated 602 and provided to the physician (or other healthcare professional) while the patient is still in the examination room to make it easier to then obtain the additional encounter information; however, the alert may be generated at any time. In some implementations, a virtual assistant (e.g., virtual assistant 238 from Fig. 2) may generate and provide the alert to inform any medical professional (e.g., doctors, nurses, physician's assistants, lab technicians, physical therapists, scribes/transcriptionist, etc.) and/or staff members involved in the patient encounter) to gather the additional encounter information. In some implementations, the alert may be provided during a pre-visit portion of the patient encounter (e.g., patient check in) and/or a post-visit portion of the patient encounter (e.g., check out).

[00134] In some implementations, processing 302 the first audio encounter information may include comparing 606 at least one layer of the plurality of layers with at least one data source, which may include at least one of physical exam finding

information, lab order information, medical condition checklist information, and compliance information. For example, as discussed above with regard to at least Fig. 2, ACD process 10 may be configured to access one or more data sources 118 (e.g., plurality of individual data sources 120, 122, 124, 126, 128), examples of which may include but are not limited to one or more a medical conditions symptoms data source, a prescriptions compatibility data source, a medical insurance coverage data source, a data source storing past/present physical exam findings of the patient, past/present lab order information for the patient, a medical compliance data source, combination thereof, etc.

[00135] Continuing with the above example, assume for example purposes only that the information documented in the medical report section of UI 500 denotes that the physician has conducted a standard physical examination for the patient. In the example, assume that the standard physical examination has nine medical points that should be checked for a standard physical examination (e.g., eyes, throat, reflexes, blood pressure, etc.). Further assume in the example that only eight of the nine medical points that should be checked for a standard physical examination have been completed by the physician. In the example, ACD process 10 may use the patient encounter information (e.g., from the medical report) to determine that a standard physical examination has been conducted with only eight medical points being documented, obtain one of the above-noted data sources that indicates that there are nine medical points that should be checked for a standard physical examination, compare 606 the information in the medical report (e.g., the eight documented medical issues) with one of the obtained data sources that indicates that there are nine medical points that should be checked for a standard physical examination, determine that one of the medical points that should be checked for a standard physical examination has not been documented in the medical report (e.g., blood pressure), and generate an alert for the physician to obtain the missing ninth medical point (e.g., blood pressure). In some implementation, the encounter information in the medical report section that was used

(at least in part) to determine that an alert should be generated may be annotated (as discussed above) to help the physician confirm that additional information is indeed needed.

[00136] As another example, assume that for example purposes only that the information documented in the medical report section of UI 500 denotes that the physician has conducted a standard physical examination for the patient. In the example, assume that the standard physical examination includes a mandatory blood test. Further assume in the example that the physician has not ordered a blood test. In the example, ACD process 10 may use the patient encounter information (e.g., from the medical report) to determine that a standard physical examination has been conducted without a blood test being documented, obtain one of the above-noted data sources that indicates that a blood test should be included for a standard physical examination, compare 606 the information in the medical report (e.g., showing a lack of blood being drawn or blood test being ordered) with one of the obtained data sources that indicates that a blood test should be included for a standard physical examination, determine that the patient's blood should be drawn and a blood test should be ordered for a standard physical examination that has not been documented in the medical report (e.g., blood pressure), and generate an alert for the physician to obtain the missing information (e.g., blood draw/test).

[00137] It will be appreciated that while the above example involves an indication of missing information from a medical examination, other examples of missing (or vague) information may be used without departing from the scope of the present disclosure. For instance, ACD process may similarly generate an alert if the information in the medical report indicates additional information may need to be obtained from the patient encounter to be more compliant with clinical / administrative guidelines. As another example, ACD process may similarly generate an alert if the information in the medical report indicates a particular prescription has not been filled/refilled where one of the above-noted data sources indicates that the particular

prescription should be filled/refilled. As another example, ACD process may similarly generate an alert if the information in the medical report indicates symptoms of a particular illness that could be verified by examining some part of the patient that has been examined. As such, the specific disclosure of generating an alert for a missed medical examination point should be taken as example only and not to otherwise limit the scope of the present disclosure.

General:

[00138] As will be appreciated by one skilled in the art, the present disclosure may be embodied as a method, a system, or a computer program product. Accordingly, the present disclosure may take the form of an entirely hardware embodiment, an entirely software embodiment (including firmware, resident software, micro-code, etc.) or an embodiment combining software and hardware aspects that may all generally be referred to herein as a "circuit," "module" or "system." Furthermore, the present disclosure may take the form of a computer program product on a computer-usable storage medium having computer-usable program code embodied in the medium.

[00139] Any suitable computer usable or computer readable medium may be utilized. The computer-usable or computer-readable medium may be, for example but not limited to, an electronic, magnetic, optical, electromagnetic, infrared, or semiconductor system, apparatus, device, or propagation medium. More specific examples (a non-exhaustive list) of the computer-readable medium may include the following: an electrical connection having one or more wires, a portable computer diskette, a hard disk, a random access memory (RAM), a read-only memory (ROM), an erasable programmable read-only memory (EPROM or Flash memory), an optical fiber, a portable compact disc read-only memory (CD-ROM), an optical storage device, a transmission media such as those supporting the Internet or an intranet, or a magnetic storage device. The computer-usable or computer-readable medium may also be paper or another suitable medium upon which the program is printed, as the program can be electronically captured, via, for instance, optical scanning of the paper or other medium,

then compiled, interpreted, or otherwise processed in a suitable manner, if necessary, and then stored in a computer memory. In the context of this document, a computer-usable or computer-readable medium may be any medium that can contain, store, communicate, propagate, or transport the program for use by or in connection with the instruction execution system, apparatus, or device. The computer-usable medium may include a propagated data signal with the computer-usable program code embodied therewith, either in baseband or as part of a carrier wave. The computer usable program code may be transmitted using any appropriate medium, including but not limited to the Internet, wireline, optical fiber cable, RF, etc.

[00140] Computer program code for carrying out operations of the present disclosure may be written in an object oriented programming language such as Java, Smalltalk, C++ or the like. However, the computer program code for carrying out operations of the present disclosure may also be written in conventional procedural programming languages, such as the "C" programming language or similar programming languages. The program code may execute entirely on the user's computer, partly on the user's computer, as a stand-alone software package, partly on the user's computer and partly on a remote computer or entirely on the remote computer or server. In the latter scenario, the remote computer may be connected to the user's computer through a local area network / a wide area network / the Internet (e.g., network 14).

[00141] The present disclosure is described with reference to flowchart illustrations and/or block diagrams of methods, apparatus (systems) and computer program products according to embodiments of the disclosure. It will be understood that each block of the flowchart illustrations and/or block diagrams, and combinations of blocks in the flowchart illustrations and/or block diagrams, may be implemented by computer program instructions. These computer program instructions may be provided to a processor of a general purpose computer / special purpose computer / other programmable data processing apparatus, such that the instructions, which execute via

the processor of the computer or other programmable data processing apparatus, create means for implementing the functions/acts specified in the flowchart and/or block diagram block or blocks.

[00142] These computer program instructions may also be stored in a computer-readable memory that may direct a computer or other programmable data processing apparatus to function in a particular manner, such that the instructions stored in the computer-readable memory produce an article of manufacture including instruction means which implement the function/act specified in the flowchart and/or block diagram block or blocks.

[00143] The computer program instructions may also be loaded onto a computer or other programmable data processing apparatus to cause a series of operational steps to be performed on the computer or other programmable apparatus to produce a computer implemented process such that the instructions which execute on the computer or other programmable apparatus provide steps for implementing the functions/acts specified in the flowchart and/or block diagram block or blocks.

[00144] The flowcharts and block diagrams in the figures may illustrate the architecture, functionality, and operation of possible implementations of systems, methods and computer program products according to various embodiments of the present disclosure. In this regard, each block in the flowchart or block diagrams may represent a module, segment, or portion of code, which comprises one or more executable instructions for implementing the specified logical function(s). It should also be noted that, in some alternative implementations, the functions noted in the block may occur out of the order noted in the figures. For example, two blocks shown in succession may, in fact, be executed substantially concurrently, or the blocks may sometimes be executed in the reverse order, not at all, or in any combination with any other flowcharts depending upon the functionality involved. It will also be noted that each block of the block diagrams and/or flowchart illustrations, and combinations of blocks in the block diagrams and/or flowchart illustrations, may be implemented by

special purpose hardware-based systems that perform the specified functions or acts, or combinations of special purpose hardware and computer instructions.

[00145] The terminology used herein is for the purpose of describing particular embodiments only and is not intended to be limiting of the disclosure. As used herein, the singular forms "a", "an" and "the" are intended to include the plural forms as well, unless the context clearly indicates otherwise. It will be further understood that the terms "comprises" and/or "comprising," when used in this specification, specify the presence of stated features, integers, steps, operations, elements, and/or components, but do not preclude the presence or addition of one or more other features, integers, steps, operations, elements, components, and/or groups thereof.

[00146] The corresponding structures, materials, acts, and equivalents of all means or step plus function elements in the claims below are intended to include any structure, material, or act for performing the function in combination with other claimed elements as specifically claimed. The description of the present disclosure has been presented for purposes of illustration and description, but is not intended to be exhaustive or limited to the disclosure in the form disclosed. Many modifications and variations will be apparent to those of ordinary skill in the art without departing from the scope and spirit of the disclosure. The embodiment was chosen and described in order to best explain the principles of the disclosure and the practical application, and to enable others of ordinary skill in the art to understand the disclosure for various embodiments with various modifications as are suited to the particular use contemplated.

[00147] A number of implementations have been described. Having thus described the disclosure of the present application in detail and by reference to embodiments thereof, it will be apparent that modifications and variations are possible without departing from the scope of the disclosure defined in the appended claims.

What Is Claimed Is:

1. A computer-implemented method comprising:

obtaining, by a computing device, encounter information of a patient encounter, wherein the encounter information includes audio encounter information obtained from at least a first encounter participant;

processing the audio encounter information obtained from at least the first encounter participant; and

generating an alert to obtain additional encounter information of the patient encounter.

2. The computer-implemented method of claim 1 wherein processing the first audio encounter information includes defining linkages between each of a plurality of layers associated with the audio encounter information.

3. The computer-implemented method of claim 1 further comprising generating a user interface displaying the plurality of layers associated with the audio encounter information obtained from at least the first encounter participant.

4. The computer-implemented method of claim 3 wherein the first layer of the plurality of layers is an audio signal associated with the audio encounter information, wherein the second layer of the plurality of layers is a transcript associated with the audio encounter information, and wherein the third layer of the plurality of layers is a medical report associated with the audio encounter information.

5. The computer-implemented method of claim 4 wherein processing the first audio encounter information includes comparing at least one layer of the plurality of layers with at least one data source.

6. The computer-implemented method of claim 5 wherein the at least one data source includes at least one of physical exam finding information, lab order information, medical condition checklist information, and compliance information.

7. The computer-implemented method of claim 1 wherein the alert is at least one of an audio alert and a visual alert.

8. A computer program product residing on a computer readable medium having a plurality of instructions stored thereon which, when executed by a processor, cause the processor to perform operations comprising:

obtaining encounter information of a patient encounter, wherein the encounter information includes audio encounter information obtained from at least a first encounter participant;

processing the audio encounter information obtained from at least the first encounter participant; and

generating an alert to obtain additional encounter information of the patient encounter.

9. The computer program product of claim 8 wherein processing the first audio encounter information includes defining linkages between each of a plurality of layers associated with the audio encounter information.

10. The computer program product of claim 8 wherein the instructions further comprise generating a user interface displaying the plurality of layers associated with the audio encounter information obtained from at least the first encounter participant.

11. The computer program product of claim 10 wherein the first layer of the plurality of layers is an audio signal associated with the audio encounter information, wherein the second layer of the

plurality of layers is a transcript associated with the audio encounter information, and wherein the third layer of the plurality of layers is a medical report associated with the audio encounter information.

12. The computer program product of claim 11 wherein processing the first audio encounter information includes comparing at least one layer of the plurality of layers with at least one data source.

13. The computer program product of claim 12 wherein the at least one data source includes at least one of physical exam finding information, lab order information, medical condition checklist information, and compliance information.

14. The computer program product of claim 8 wherein the alert is at least one of an audio alert and a visual alert.

15. A computing system including a processor and memory configured to perform operations comprising:

obtaining encounter information of a patient encounter, wherein the encounter information includes audio encounter information obtained from at least a first encounter participant;

processing the audio encounter information obtained from at least the first encounter participant; and

generating an alert to obtain additional encounter information of the patient encounter.

16. The computing system of claim 15 wherein processing the first audio encounter information includes defining linkages between each of a plurality of layers associated with the audio encounter information.

17. The computing system of claim 15 wherein the instructions further comprise generating a user interface displaying the plurality of layers associated with the audio encounter information obtained from at least the first encounter participant.

18. The computing system of claim 17 wherein the first layer of the plurality of layers is an audio signal associated with the audio encounter information, wherein the second layer of the plurality of layers is a transcript associated with the audio encounter information, and wherein the third layer of the plurality of layers is a medical report associated with the audio encounter information.

19. The computing system of claim 18 wherein processing the first audio encounter information includes comparing at least one layer of the plurality of layers with at least one data source.

20. The computing system of claim 19 wherein the at least one data source includes at least one of physical exam finding information, lab order information, medical condition checklist information, and compliance information.

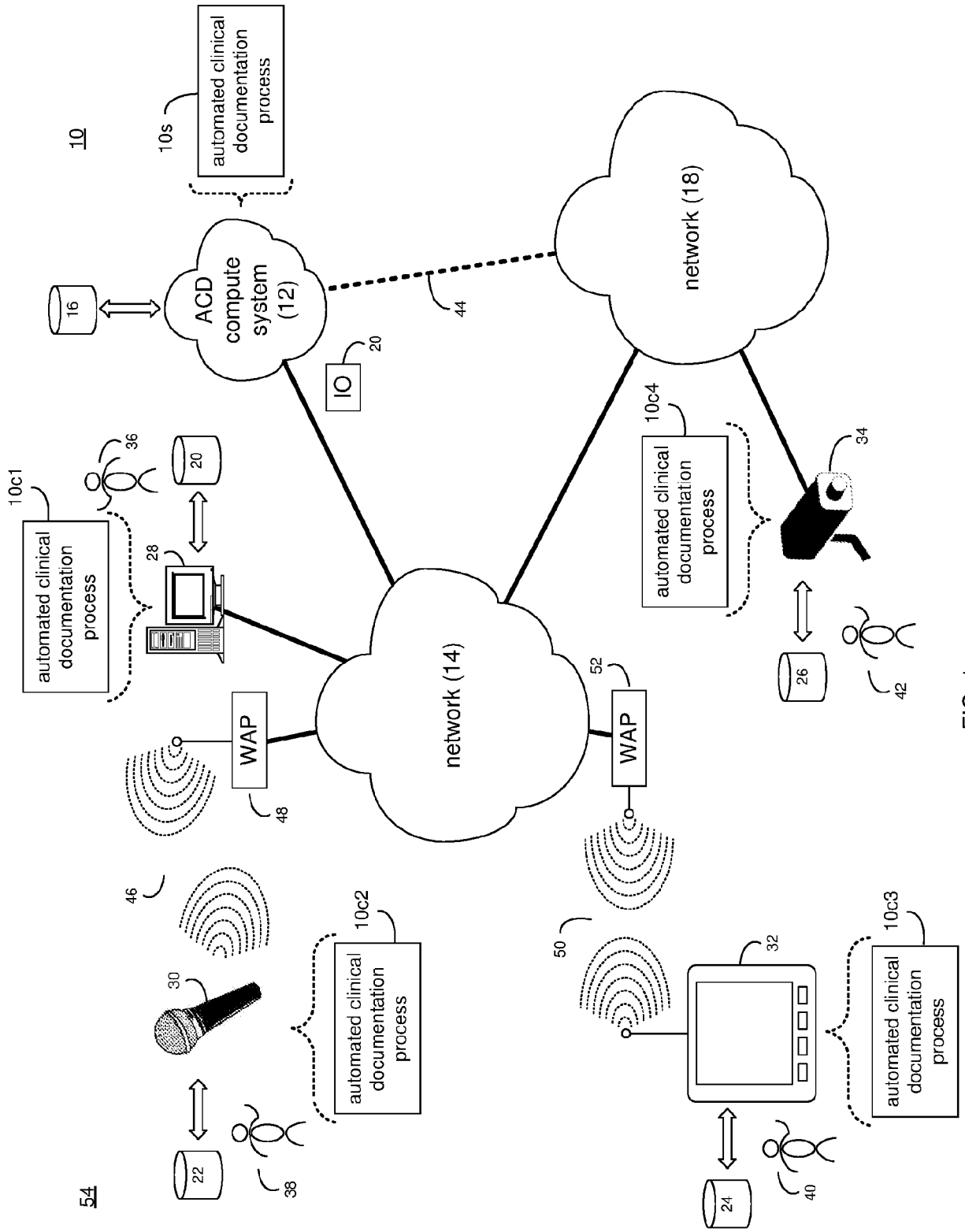
10
10s
automated clinical documentation process
ACD compute system (12)
16
IO
20
44
network (18)
10c1
automated clinical documentation process
36
20
28
WAP
48
network (14)
52
WAP
10c4
automated clinical documentation process
34
26
42
46
30
10c2
automated clinical documentation process
22
38
54
50
32
10c3
automated clinical documentation process
24
40

FIG. 1

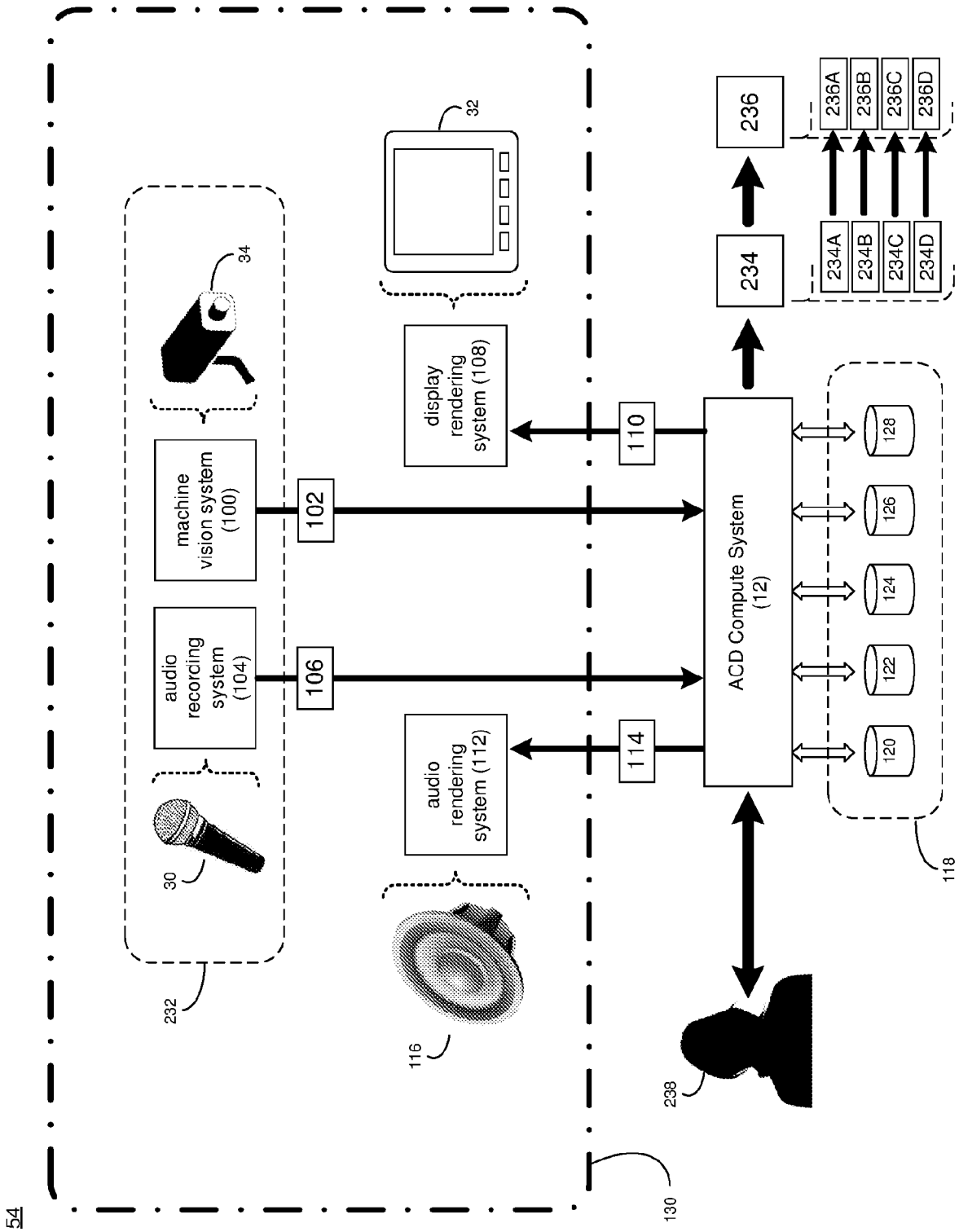
54
232
machine vision system (100)
audio recording system (104)
34
30
32
display rendering system (108)
audio rendering system (112)
116
102
106
110
114
ACD Compute System (12)
236
234
236A
236B
236C
236D
234A
234B
234C
234D
128
126
124
122
120
118
238
130

FIG. 2

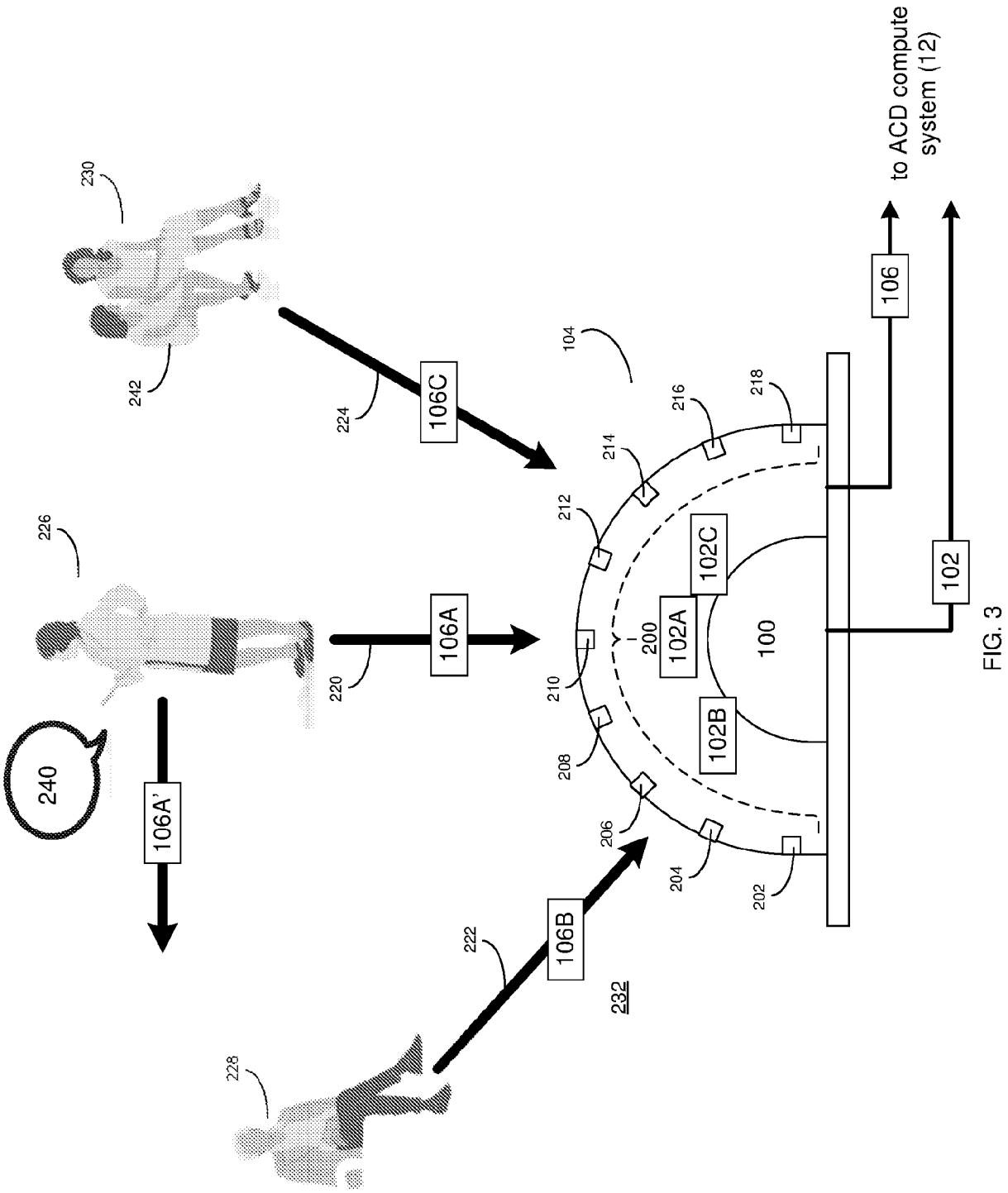
to ACD compute system (12)
106
102
104
100
102A
102B
102C
200
202
204
206
208
210
212
214
216
218
220
222
224
226
228
230
232
240
242
106A
106A'
106B
106C

FIG. 3

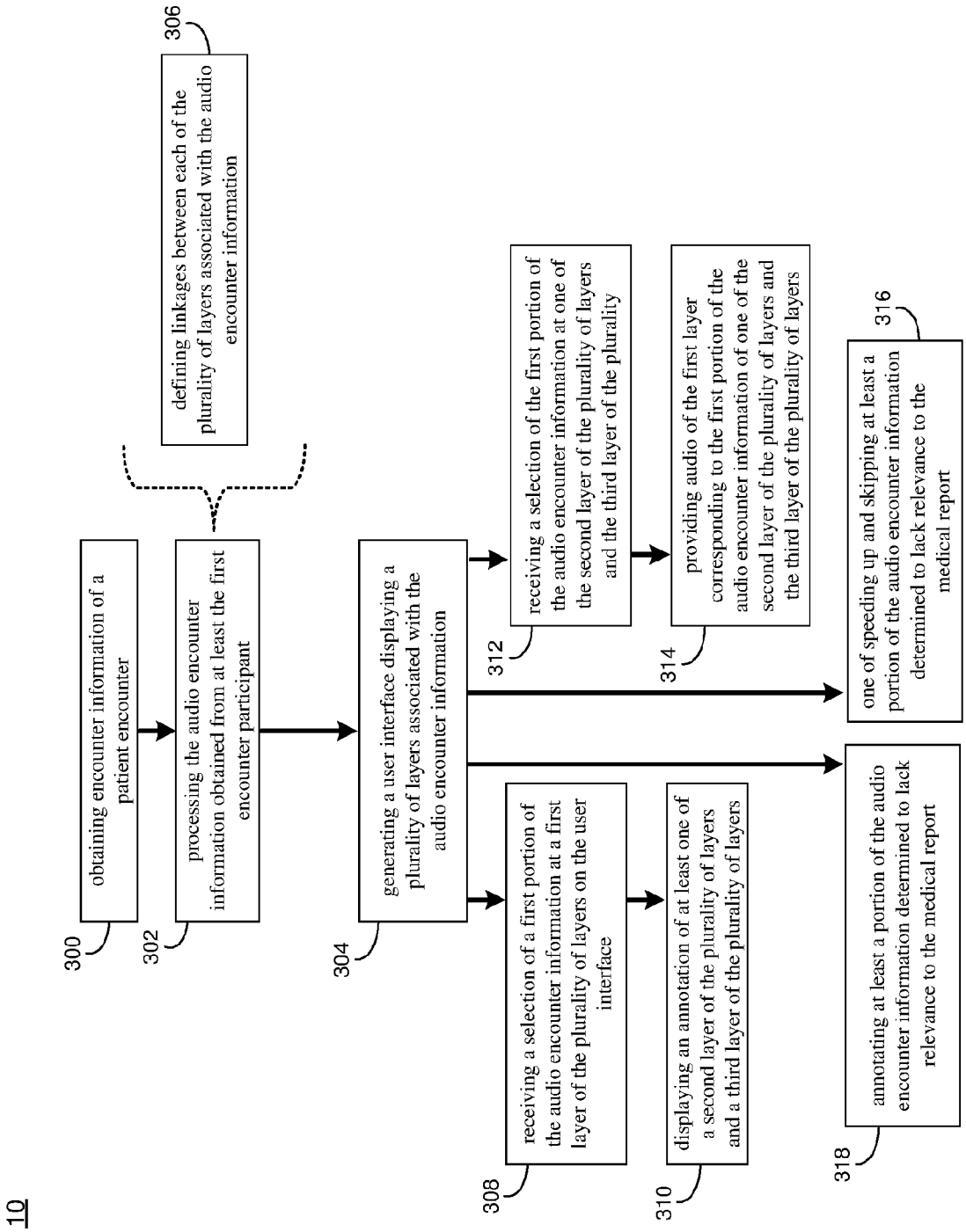
10
300
obtaining encounter information of a patient encounter
302
processing the audio encounter information obtained from at least the first encounter participant
306
defining linkages between each of the plurality of layers associated with the audio encounter information
304
generating a user interface displaying a plurality of layers associated with the audio encounter information
308
receiving a selection of a first portion of the audio encounter information at a first layer of the plurality of layers on the user interface
310
displaying an annotation of at least one of a second layer of the plurality of layers and a third layer of the plurality of layers
312
receiving a selection of the first portion of the audio encounter information at one of the second layer of the plurality of layers and the third layer of the plurality
314
providing audio of the first layer corresponding to the first portion of the audio encounter information of one of the second layer of the plurality of layers and the third layer of the plurality of layers
316
one of speeding up and skipping at least a portion of the audio encounter information determined to lack relevance to the medical report
318
annotating at least a portion of the audio encounter information determined to lack relevance to the medical report

FIG. 4

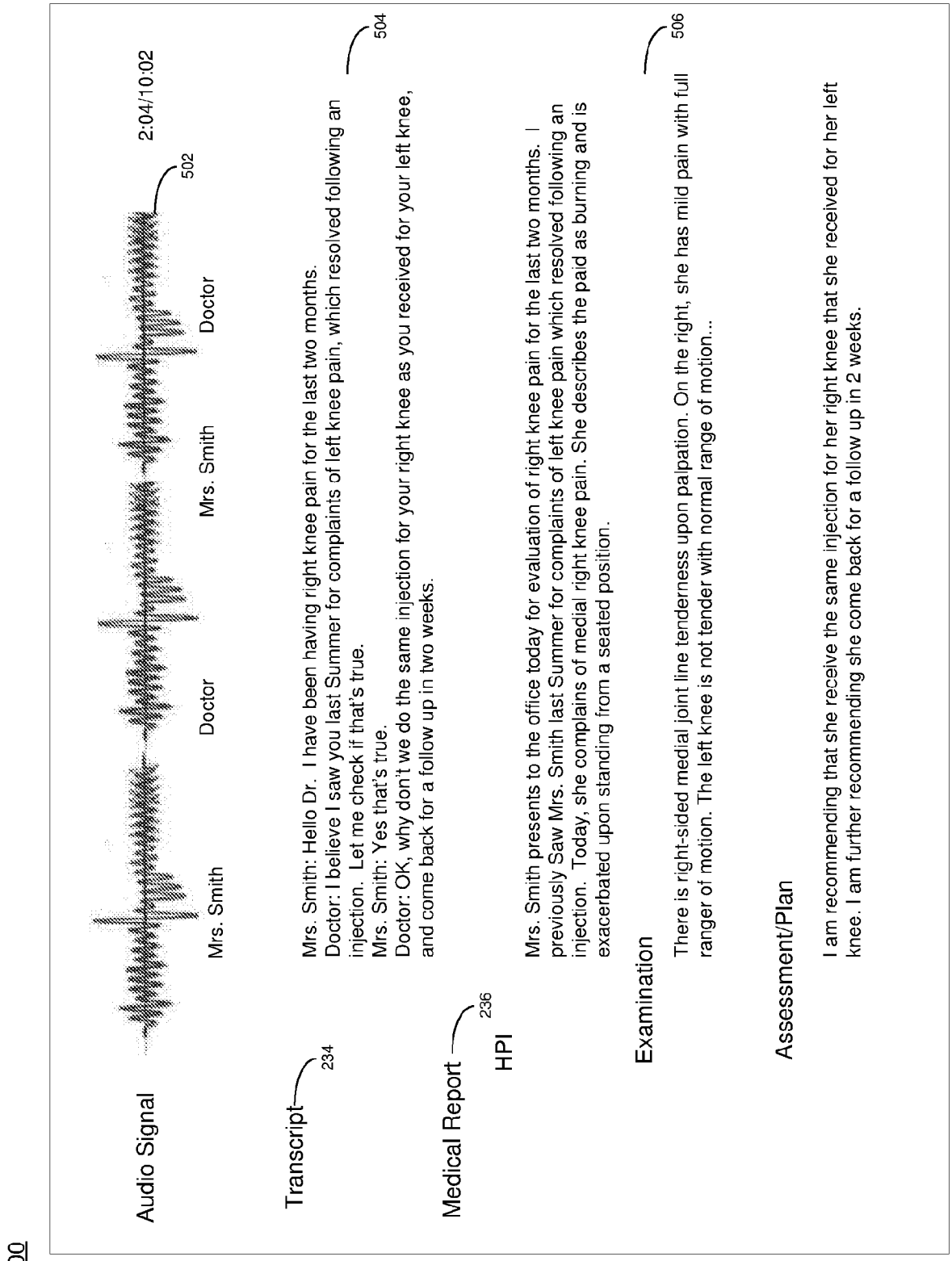
500
Audio Signal
2:04/10:02
502
Mrs. Smith
Doctor
Mrs. Smith
Doctor
Transcript
234
504
Mrs. Smith: Hello Dr. I have been having right knee pain for the last two months.
Doctor: I believe I saw you last Summer for complaints of left knee pain, which resolved following an injection. Let me check if that's true.
Mrs. Smith: Yes that's true.
Doctor: OK, why don't we do the same injection for your right knee as you received for your left knee, and come back for a follow up in two weeks.
Medical Report
236
HPI
Mrs. Smith presents to the office today for evaluation of right knee pain for the last two months. I previously Saw Mrs. Smith last Summer for complaints of left knee pain which resolved following an injection. Today, she complains of medial right knee pain. She describes the paid as burning and is exacerbated upon standing from a seated position.
Examination
506
There is right-sided medial joint line tenderness upon palpation. On the right, she has mild pain with full ranger of motion. The left knee is not tender with normal range of motion...
Assessment/Plan
I am recommending that she receive the same injection for her right knee that she received for her left knee. I am further recommending she come back for a follow up in 2 weeks.

FIG. 5

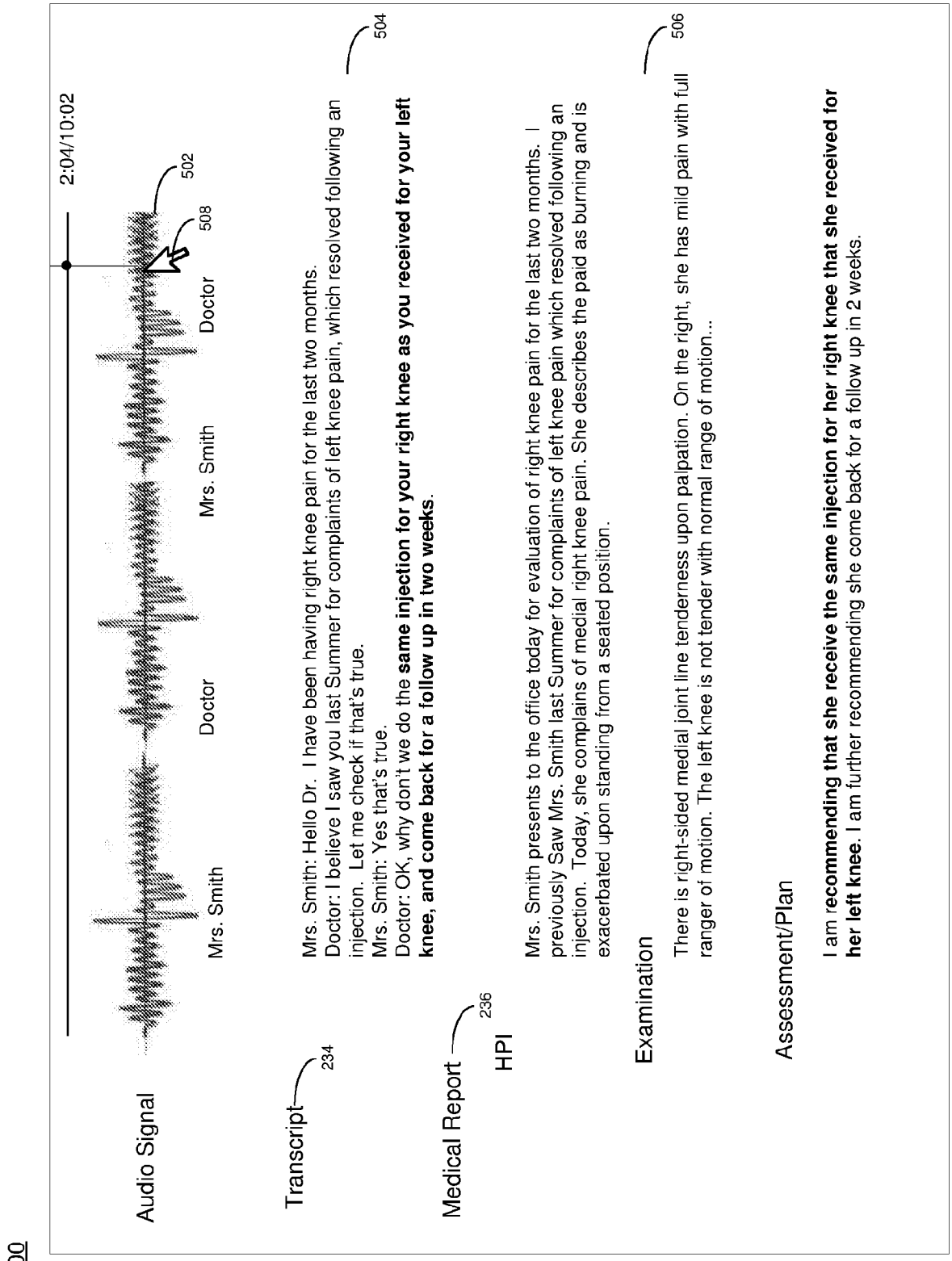
500
Audio Signal
2:04/10:02
502
508
Mrs. Smith
Doctor
Mrs. Smith
Doctor
Transcript
234
504
Mrs. Smith: Hello Dr. I have been having right knee pain for the last two months.
Doctor: I believe I saw you last Summer for complaints of left knee pain, which resolved following an injection. Let me check if that's true.
Mrs. Smith: Yes that's true.
Doctor: OK, why don't we do the **same injection for your right knee as you received for your left knee, and come back for a follow up in two weeks**.
Medical Report
236
HPI
Mrs. Smith presents to the office today for evaluation of right knee pain for the last two months. I previously Saw Mrs. Smith last Summer for complaints of left knee pain which resolved following an injection. Today, she complains of medial right knee pain. She describes the paid as burning and is exacerbated upon standing from a seated position.
Examination
506
There is right-sided medial joint line tenderness upon palpation. On the right, she has mild pain with full ranger of motion. The left knee is not tender with normal range of motion...
Assessment/Plan
I am **recommending that she receive the same injection for her right knee that she received for her left knee**. I am further recommending she come back for a follow up in 2 weeks.

FIG. 6

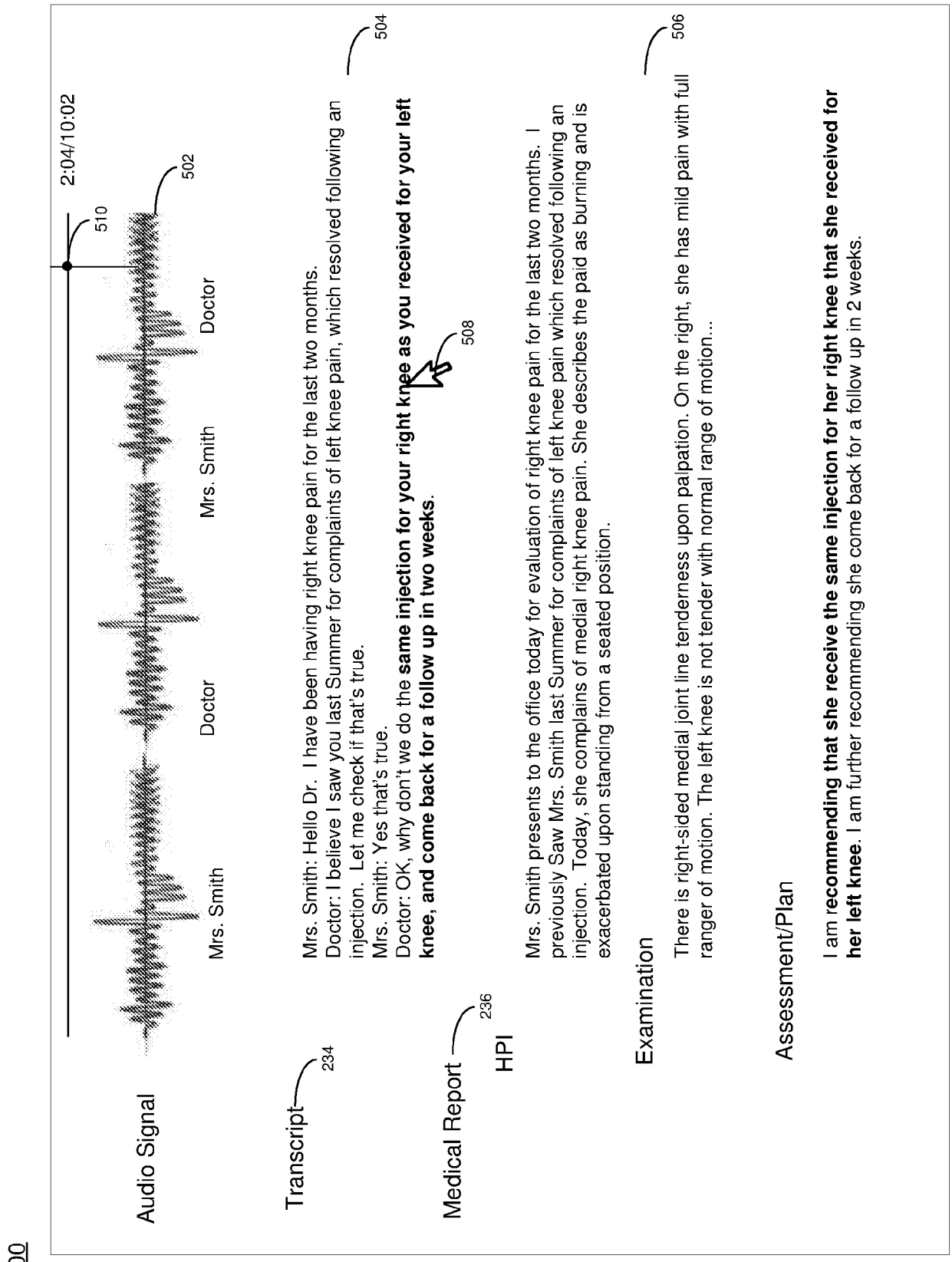
500
Audio Signal
2:04/10:02
510
502
Mrs. Smith
Doctor
Mrs. Smith
Doctor
Transcript
234
504
Mrs. Smith: Hello Dr. I have been having right knee pain for the last two months.
Doctor: I believe I saw you last Summer for complaints of left knee pain, which resolved following an injection. Let me check if that's true.
Mrs. Smith: Yes that's true.
Doctor: OK, why don't we do the **same injection for your right knee as you received for your left knee, and come back for a follow up in two weeks.**
508
Medical Report
236
HPI
Mrs. Smith presents to the office today for evaluation of right knee pain for the last two months. I previously Saw Mrs. Smith last Summer for complaints of left knee pain which resolved following an injection. Today, she complains of medial right knee pain. She describes the paid as burning and is exacerbated upon standing from a seated position.
Examination
506
There is right-sided medial joint line tenderness upon palpation. On the right, she has mild pain with full ranger of motion. The left knee is not tender with normal range of motion...
Assessment/Plan
I am **recommending that she receive the same injection for her right knee that she received for her left knee**. I am further recommending she come back for a follow up in 2 weeks.

FIG. 7

500

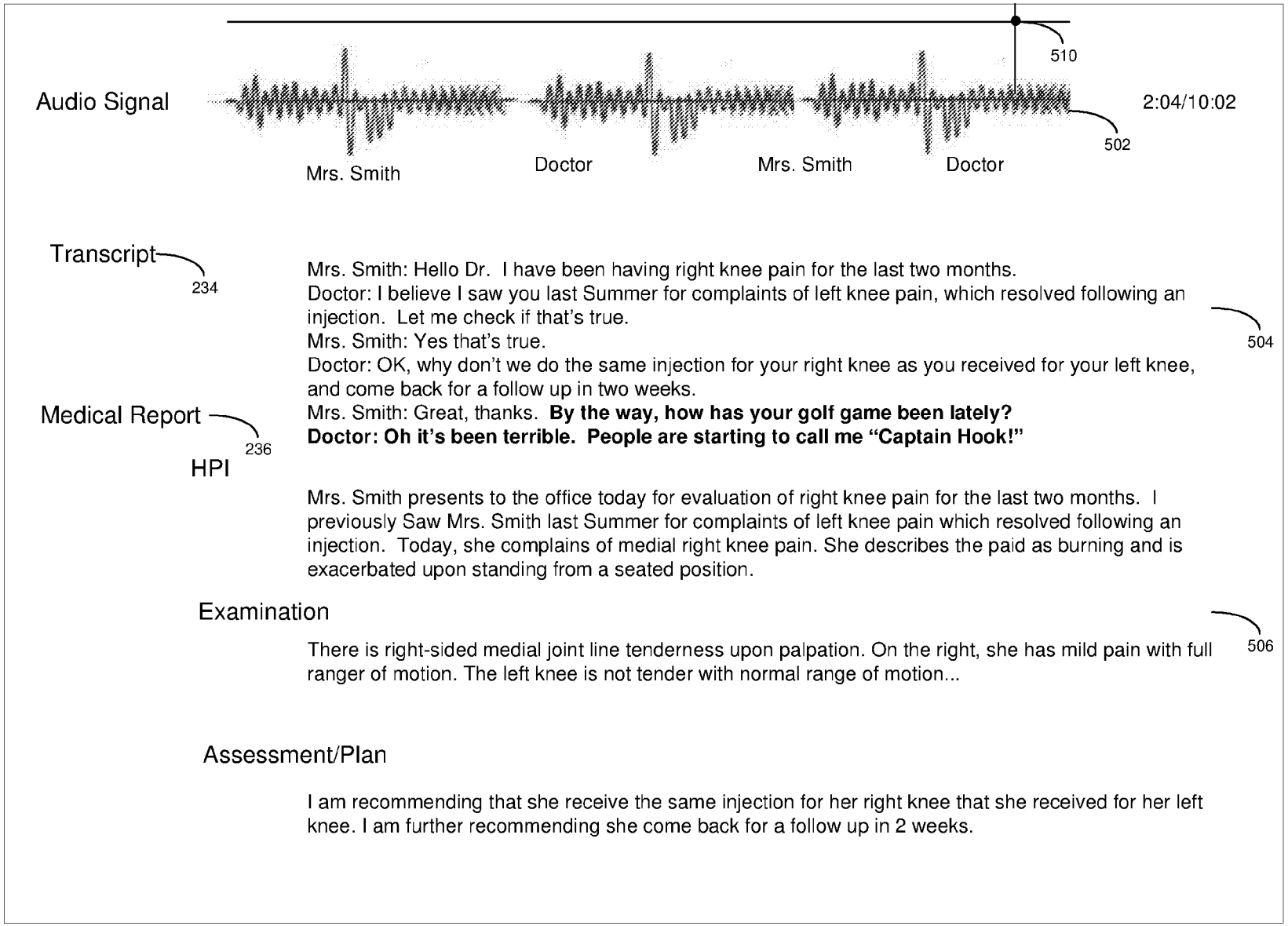
Audio Signal
510
2:04/10:02
502
Mrs. Smith
Doctor
Mrs. Smith
Doctor
Transcript
234
Mrs. Smith: Hello Dr. I have been having right knee pain for the last two months.
Doctor: I believe I saw you last Summer for complaints of left knee pain, which resolved following an injection. Let me check if that's true.
Mrs. Smith: Yes that's true.
Doctor: OK, why don't we do the same injection for your right knee as you received for your left knee, and come back for a follow up in two weeks.
Mrs. Smith: Great, thanks. By the way, how has your golf game been lately?
Doctor: Oh it's been terrible. People are starting to call me "Captain Hook!"
504
Medical Report
236
HPI
Mrs. Smith presents to the office today for evaluation of right knee pain for the last two months. I previously Saw Mrs. Smith last Summer for complaints of left knee pain which resolved following an injection. Today, she complains of medial right knee pain. She describes the paid as burning and is exacerbated upon standing from a seated position.
Examination
506
There is right-sided medial joint line tenderness upon palpation. On the right, she has mild pain with full ranger of motion. The left knee is not tender with normal range of motion...
Assessment/Plan
I am recommending that she receive the same injection for her right knee that she received for her left knee. I am further recommending she come back for a follow up in 2 weeks.

FIG. 8

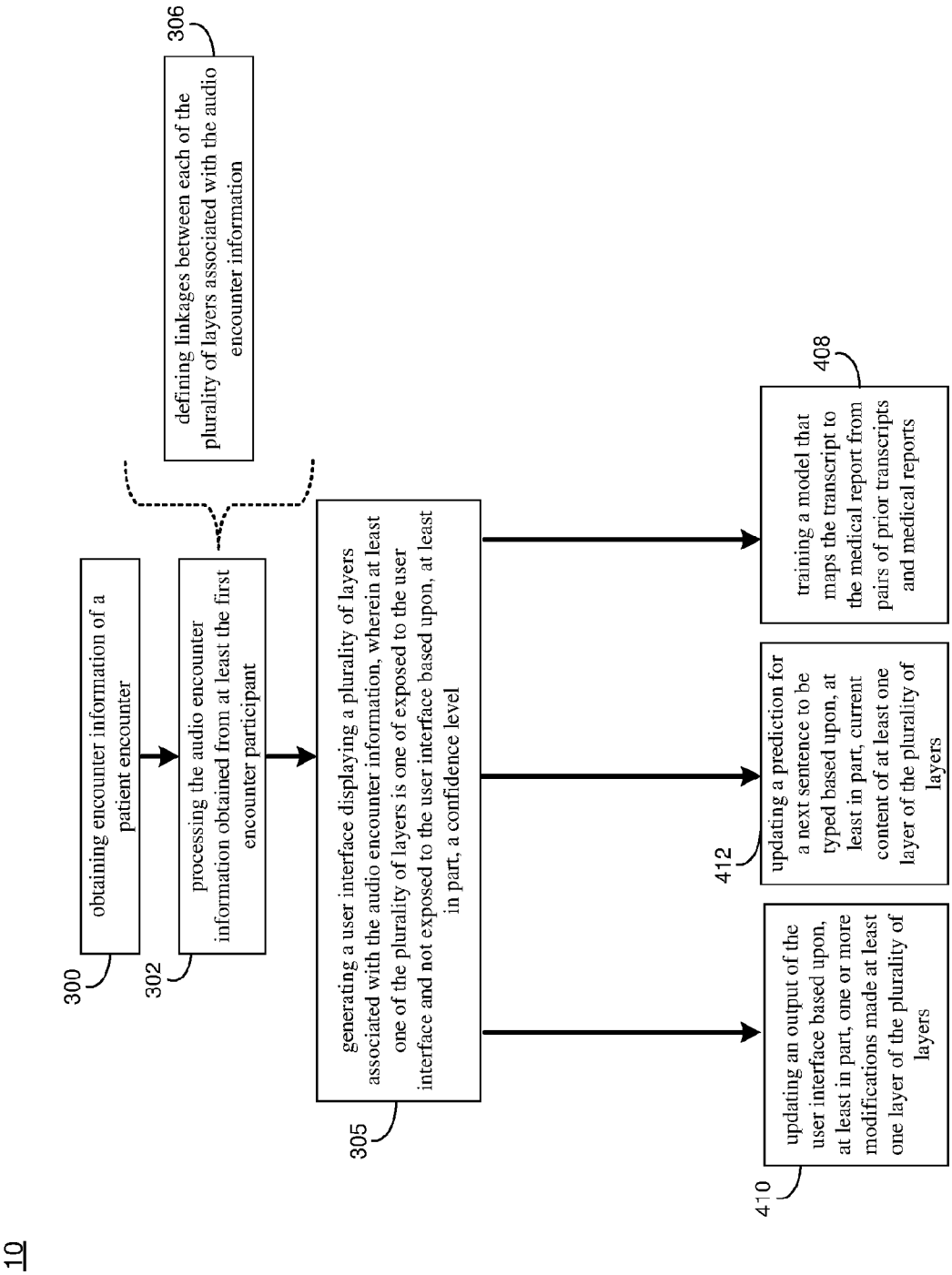
10
obtaining encounter information of a patient encounter
300
processing the audio encounter information obtained from at least the first encounter participant
302
defining linkages between each of the plurality of layers associated with the audio encounter information
306
generating a user interface displaying a plurality of layers associated with the audio encounter information, wherein at least one of the plurality of layers is one of exposed to the user interface and not exposed to the user interface based upon, at least in part, a confidence level
305
updating an output of the user interface based upon, at least in part, one or more modifications made at least one layer of the plurality of layers
410
updating a prediction for a next sentence to be typed based upon, at least in part, current content of at least one layer of the plurality of layers
412
training a model that maps the transcript to the medical report from pairs of prior transcripts and medical reports
408

FIG. 9

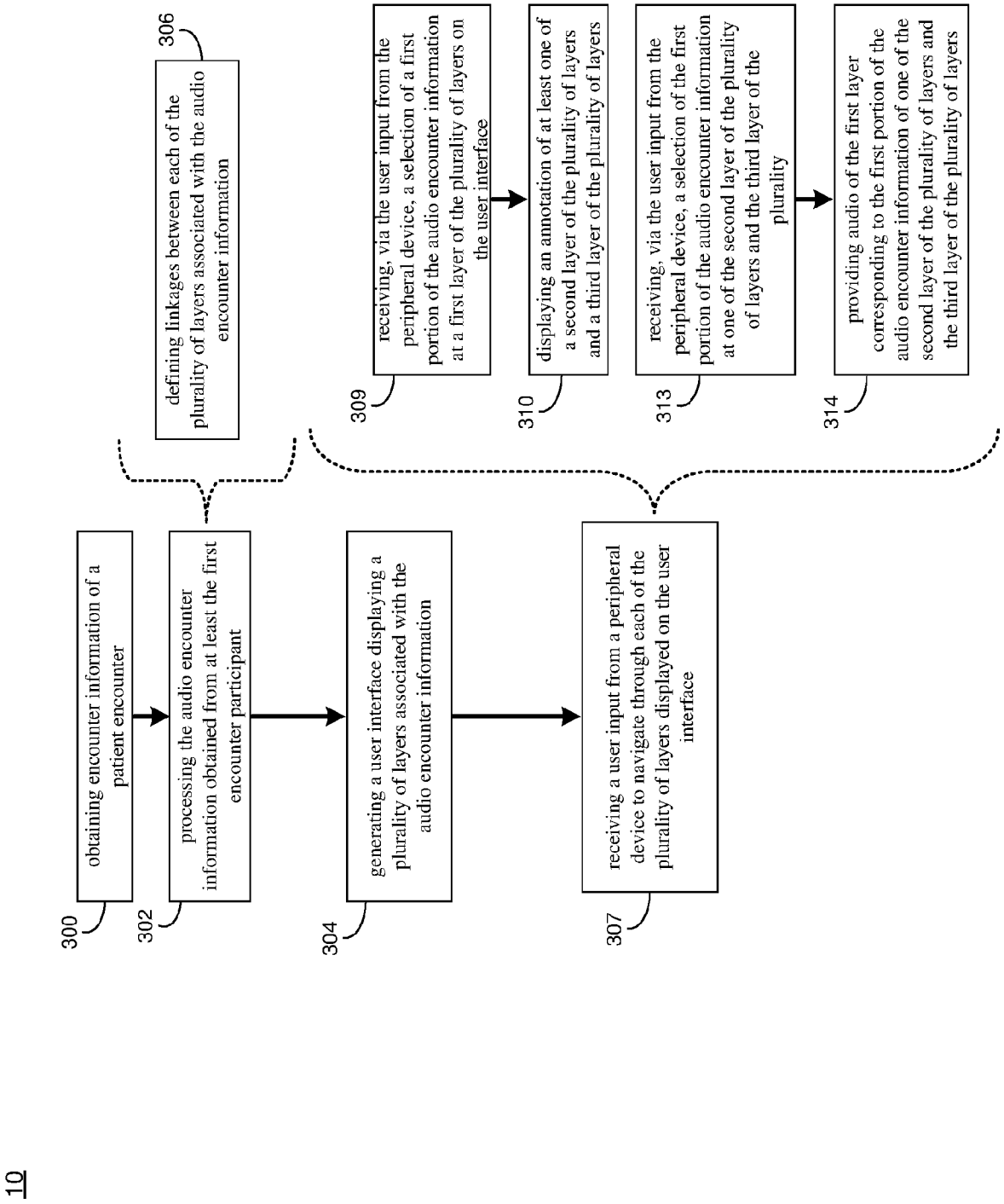
10
300
obtaining encounter information of a patient encounter
302
processing the audio encounter information obtained from at least the first encounter participant
306
defining linkages between each of the plurality of layers associated with the audio encounter information
304
generating a user interface displaying a plurality of layers associated with the audio encounter information
307
receiving a user input from a peripheral device to navigate through each of the plurality of layers displayed on the user interface
309
receiving, via the user input from the peripheral device, a selection of a first portion of the audio encounter information at a first layer of the plurality of layers on the user interface
310
displaying an annotation of at least one of a second layer of the plurality of layers and a third layer of the plurality of layers
313
receiving, via the user input from the peripheral device, a selection of the first portion of the audio encounter information at one of the second layer of the plurality of layers and the third layer of the plurality
314
providing audio of the first layer corresponding to the first portion of the audio encounter information of one of the second layer of the plurality of layers and the third layer of the plurality of layers

FIG. 10

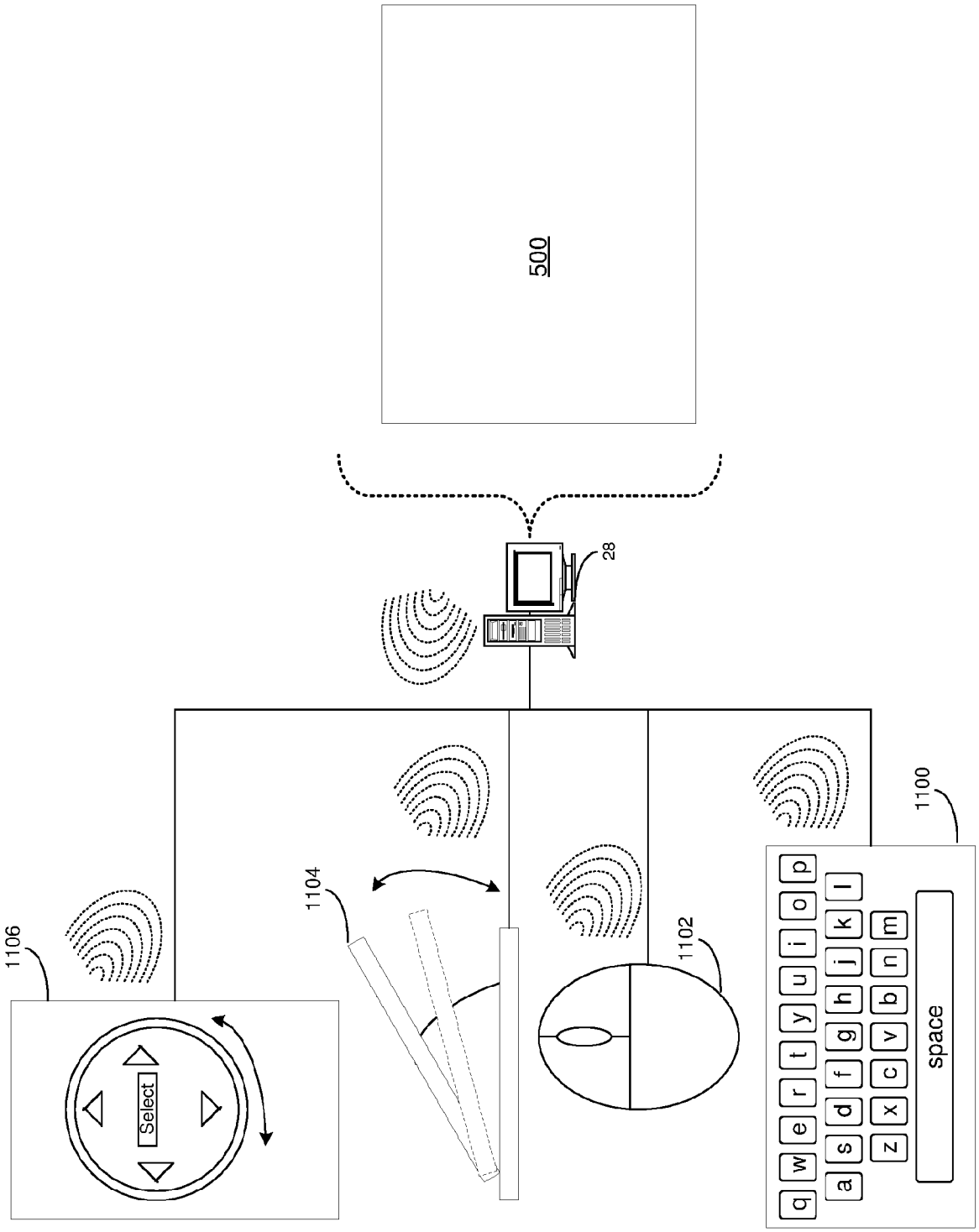

FIG. 11

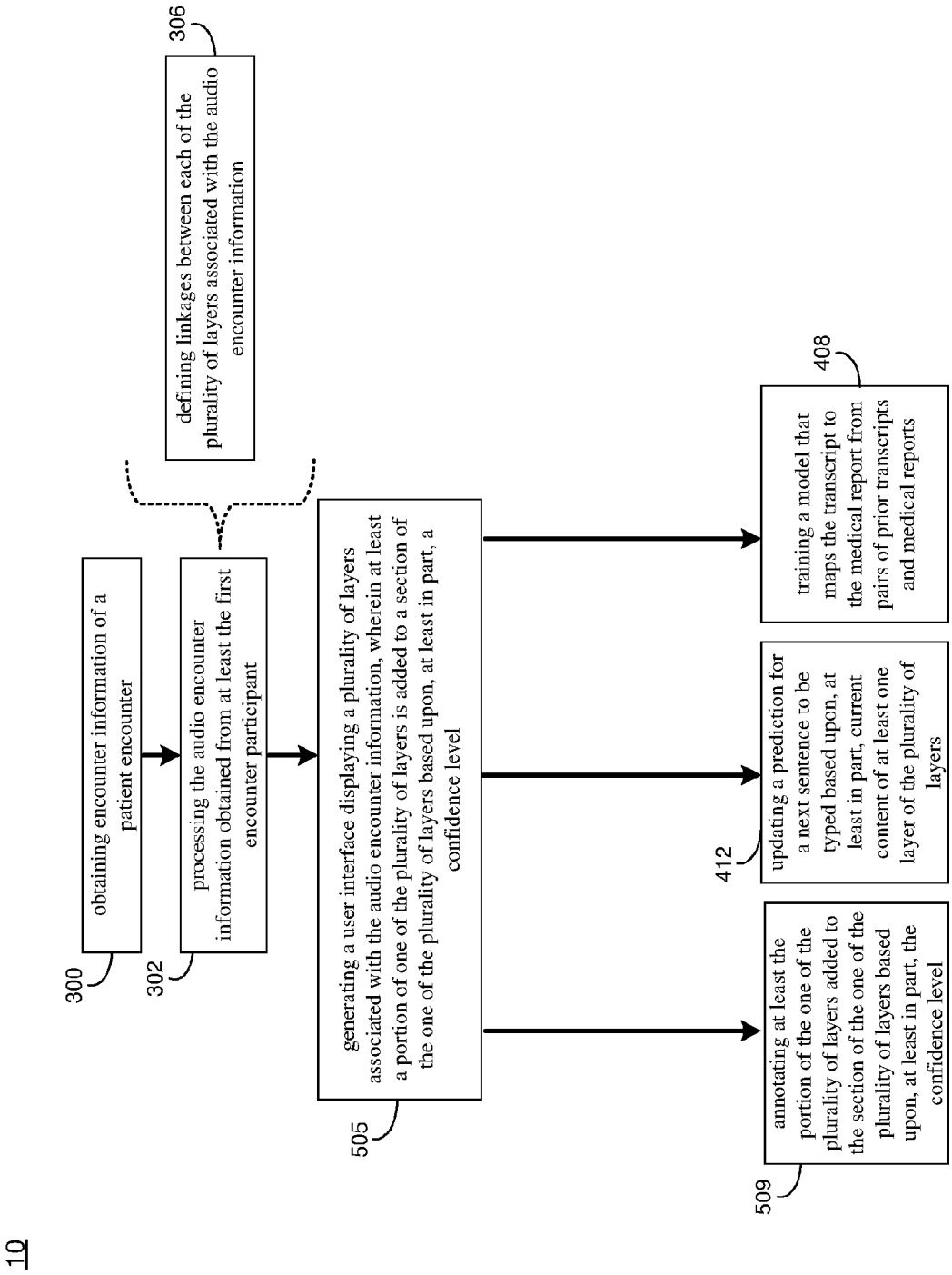
10
obtaining encounter information of a patient encounter
300
processing the audio encounter information obtained from at least the first encounter participant
302
defining linkages between each of the plurality of layers associated with the audio encounter information
306
generating a user interface displaying a plurality of layers associated with the audio encounter information, wherein at least a portion of one of the plurality of layers is added to a section of the one of the plurality of layers based upon, at least in part, a confidence level
505
annotating at least the portion of the one of the plurality of layers added to the section of the one of the plurality of layers based upon, at least in part, the confidence level
509
updating a prediction for a next sentence to be typed based upon, at least in part, current content of at least one layer of the plurality of layers
412
training a model that maps the transcript to the medical report from pairs of prior transcripts and medical reports
408

FIG. 12

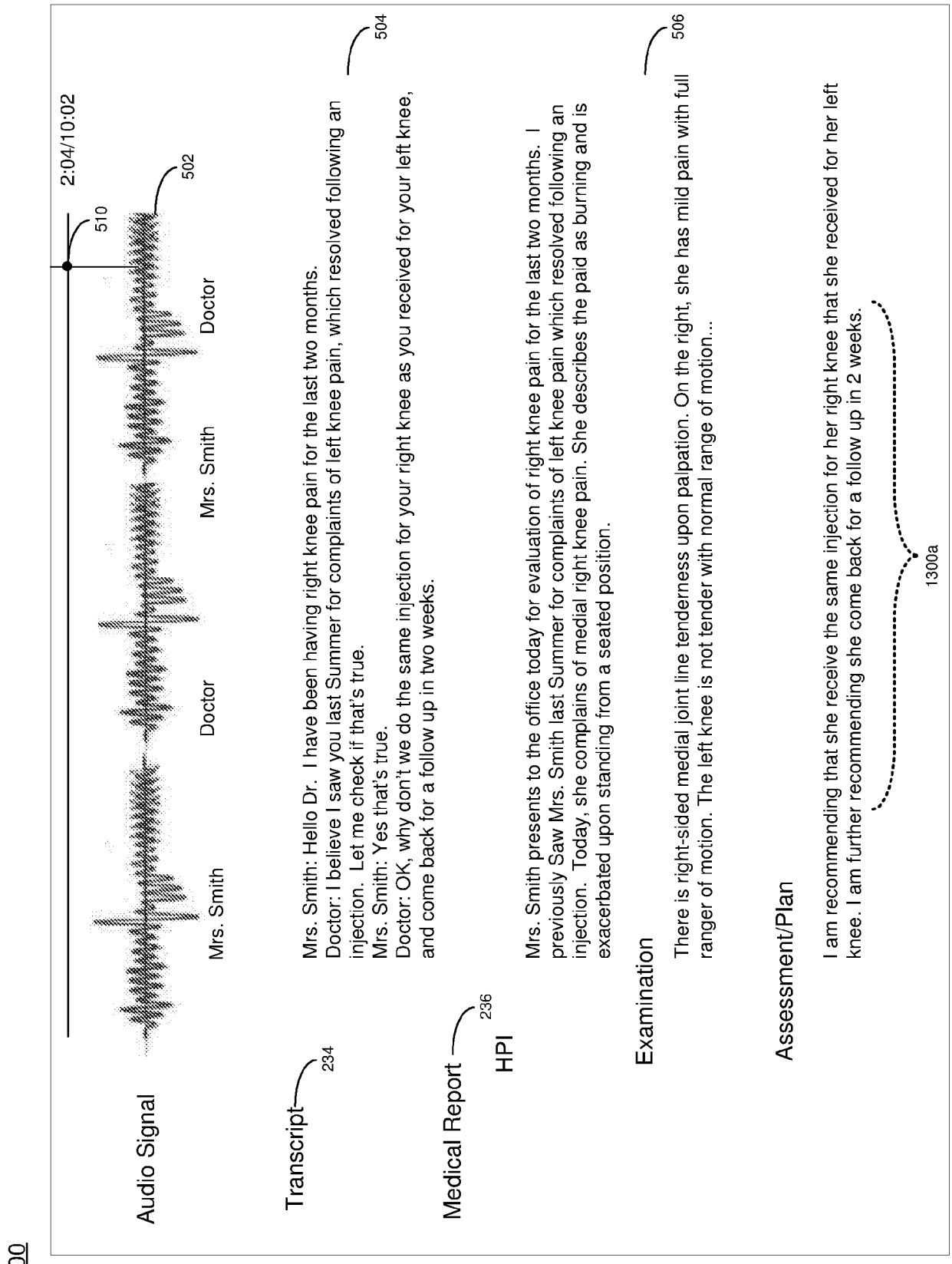
500
Audio Signal
2:04/10:02
510
502
Mrs. Smith
Doctor
Mrs. Smith
Doctor
Transcript
234
Mrs. Smith: Hello Dr. I have been having right knee pain for the last two months.
Doctor: I believe I saw you last Summer for complaints of left knee pain, which resolved following an injection. Let me check if that's true.
Mrs. Smith: Yes that's true.
Doctor: OK, why don't we do the same injection for your right knee as you received for your left knee, and come back for a follow up in two weeks.
504
Medical Report
236
HPI
Mrs. Smith presents to the office today for evaluation of right knee pain for the last two months. I previously Saw Mrs. Smith last Summer for complaints of left knee pain which resolved following an injection. Today, she complains of medial right knee pain. She describes the paid as burning and is exacerbated upon standing from a seated position.
Examination
506
There is right-sided medial joint line tenderness upon palpation. On the right, she has mild pain with full ranger of motion. The left knee is not tender with normal range of motion....
Assessment/Plan
I am recommending that she receive the same injection for her right knee that she received for her left knee. I am further recommending she come back for a follow up in 2 weeks.
1300a

FIG. 13

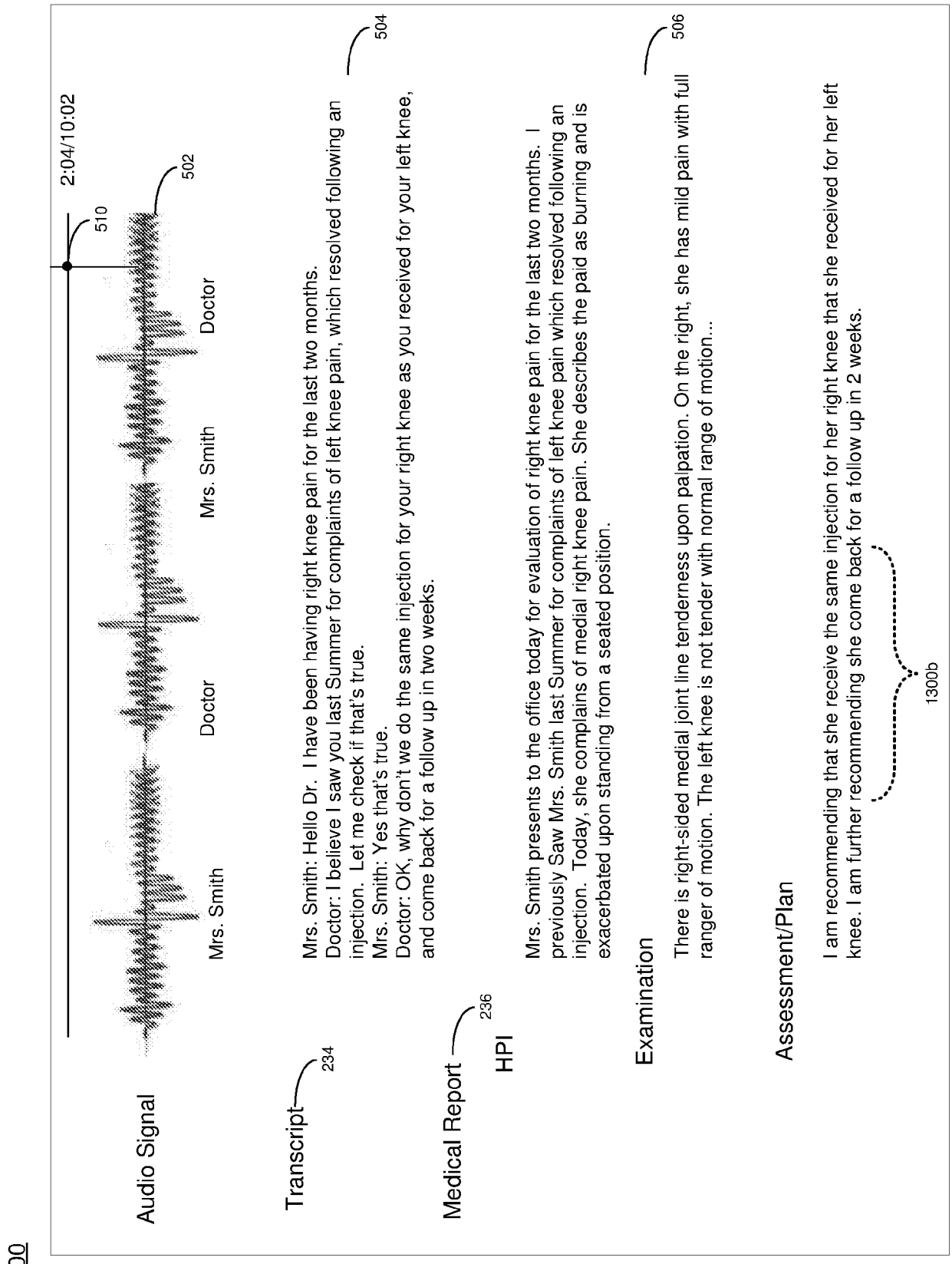
500
Audio Signal
2:04/10:02
510
502
Mrs. Smith
Doctor
Mrs. Smith
Doctor
Transcript
234
504
Mrs. Smith: Hello Dr. I have been having right knee pain for the last two months.
Doctor: I believe I saw you last Summer for complaints of left knee pain, which resolved following an injection. Let me check if that's true.
Mrs. Smith: Yes that's true.
Doctor: OK, why don't we do the same injection for your right knee as you received for your left knee, and come back for a follow up in two weeks.
Medical Report
236
HPI
Mrs. Smith presents to the office today for evaluation of right knee pain for the last two months. I previously Saw Mrs. Smith last Summer for complaints of left knee pain which resolved following an injection. Today, she complains of medial right knee pain. She describes the paid as burning and is exacerbated upon standing from a seated position.
Examination
506
There is right-sided medial joint line tenderness upon palpation. On the right, she has mild pain with full ranger of motion. The left knee is not tender with normal range of motion....
Assessment/Plan
I am recommending that she receive the same injection for her right knee that she received for her left knee. I am further recommending she come back for a follow up in 2 weeks.
1300b

FIG. 14

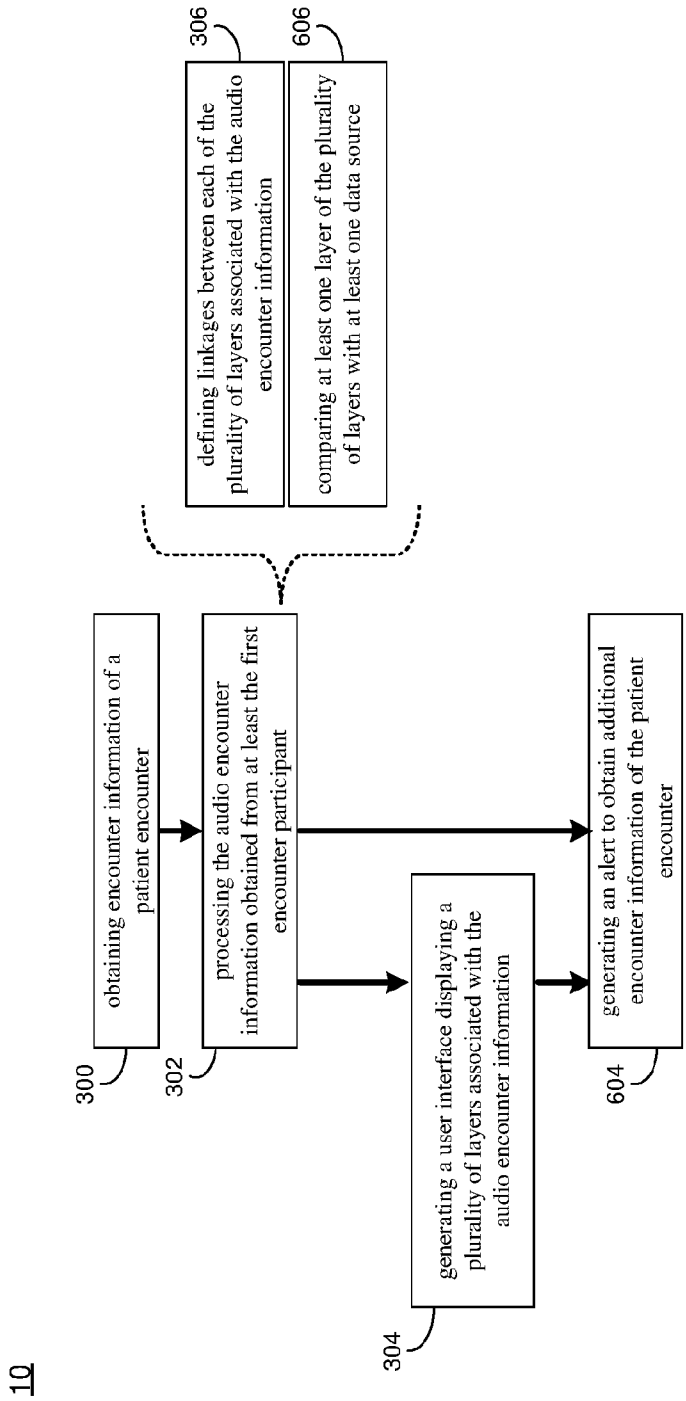
10
obtaining encounter information of a patient encounter
300
processing the audio encounter information obtained from at least the first encounter participant
302
defining linkages between each of the plurality of layers associated with the audio encounter information
306
comparing at least one layer of the plurality of layers with at least one data source
606
generating a user interface displaying a plurality of layers associated with the audio encounter information
304
generating an alert to obtain additional encounter information of the patient encounter
604

FIG. 15

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 19/20771

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - G06F 3/00 (2019.01)
CPC - G09B 7/04, G06F 3/0481, G06F 8/34, G06F 3/04845, G06F 9/4443, G06F 9/4443, G06F 3/0482, G06F 8/38, G06F 3/16, G06F 3/0481, G11B 27/34, G06F 17/30749, G06F 3/167

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
See Search History Document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
See Search History Document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
See Search History Document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2015/0379209 A1 (Kusuma et al.) 31 December 2015 (31.12.2015), entire document especially abstract; para [0068], [0188], [0204], [0206], [0235]	1-20
A	US 2016/0179770 A1 (MModal IP LLC) 23 June 2016 (23.06.2016), entire document	1-20
A	US 2011/0301982 A1 (GREEN, JR. et al.) 08 December 2011 (08.12.2011), entire document	1-20

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
07 May 2019 (07.05.2019)	30 MAY 2019

Name and mailing address of the ISA/US	Authorized officer:
Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774