

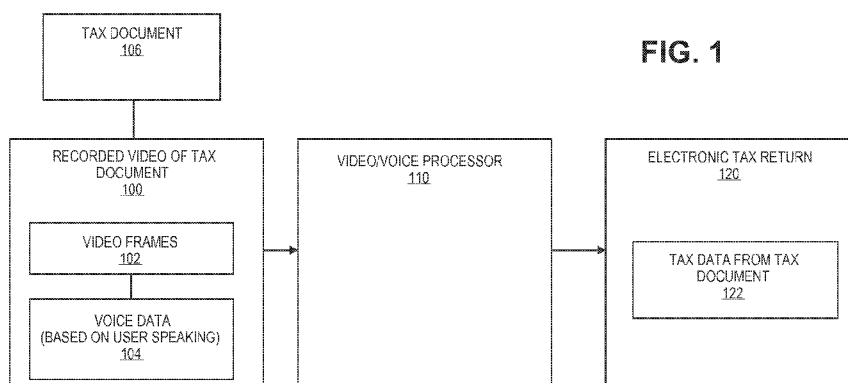


- (51) **International Patent Classification:**
G06F 17/00 (2006.01) G11B 20/10 (2006.01)
G10L 15/25 (2013.01)
- (21) **International Application Number:**
PCT/US2013/039936
- (22) **International Filing Date:**
7 May 2013 (07.05.2013)
- (25) **Filing Language:** English
- (26) **Publication Language:** English
- (30) **Priority Data:**
13/874,382 30 April 2013 (30.04.2013) US
- (71) **Applicant: INTUIT INC.** [US/US]; 2700 Coast Avenue, Mountain View, CA 94043 (US).
- (72) **Inventors; and**
- (71) **Applicants :** **MORI, Kenichi** [US/US]; 2032 Ladera Court, Carlsbad, CA 92009 (US). **MARR, Justin** [US/US]; 1130 32nd Street, San Diego, CA 92102 (US). **HARRIS, Catherine** [US/US]; 15696 Beltaire Lanie, San Diego, CA 92127 (US).
- (74) **Agent: LUECK, Gary;** 2040 Main Street, Ste 710, Irvine, CA 92614 (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, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, 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, 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, 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, ML, MR, NE, SN, TD, TG).

Published:
— with international search report (Art. 21(3))

(54) **Title:** VIDEO-VOICE PREPARATION OF ELECTRONIC TAX RETURN



(57) **Abstract:** Methods, systems and computer program products for processing video of tax documents and associated verbal input provided by a user and populating at least a portion of an electronic tax return with processing results.

WO 2014/178888 A1

VIDEO-VOICE PREPARATION OF ELECTRONIC TAX RETURN

SUMMARY

[0001] Embodiments are related to preparation of an electronic tax return based
5 on video / audio data of tax documents.

[0002] Certain embodiments are directed to processing a video taken of one or
more tax documents and populating at least a portion of an electronic tax return with
results of video processing.

[0003] Certain embodiments are directed to acquiring or receiving and processing
10 video or photograph of a tax document and audio data in the form voice or speech
by a user of a tax preparation application for the purpose of populating one or more
fields of an electronic tax return or interview screen generated by the tax preparation
application will be populated. Voice or speech data may be part of the video or
recorded and received separately of the video or photograph of a tax document. At
15 least a portion of an electronic tax return is populated with results of video/voice
processing, which may involve analysis of document structure within the video or
photograph and character recognition. For example, a video may be analyzed to
determine the type or name of the tax document, and voice data may be analyzed to
determine or confirm the type or name of the tax document by providing additional
20 contextual information about the tax documents such as a description of the tax
document, the source of the tax document, or other verbal explanation regarding the
document's relevance to the electronic tax return. A character recognition process
may be applied to video frames, images or photographs to determine tax data
therein, and that recognized tax data may be used to populate one or more fields of
25 an electronic tax return for that tax document.

[0004] Certain embodiments are directed to use of multiple video or image
capture devices for preparation of an electronic tax return including video and/or
voice processing for determining document types and tax data for electronic tax
returns, and for detecting user expressions or gestures during electronic dialogue
30 with a computing apparatus or tax preparation application. The electronic dialogue
may occur during recording or imaging of the documents and/or during preparation
of an electronic tax return or during which video and/or voice processing is
performed. For example, a first or front-facing camera of a computing apparatus
may be used to acquire a video or photograph of a tax document, and a second or

rear-facing camera facing the user may be used to detect a facial expression or gesture of the user during the process or electronic dialogue generated according to embodiments. With embodiments, the tax preparation application generates and presents responses related to the detected expression or gesture. The first and second cameras and associated processing may be operated or performed simultaneously or by the same processor, or independently of each other or at different times. For example, the rear-facing camera may be used to detect user frustration or apprehension such as by an inquisitive look or frown, and embodiments may determine appropriate responses to alleviate user concerns such as providing affirmation to the user that the process is proceeding as it should, or by providing positive news about a refund the user will receive.

[0005] Certain embodiments are directed to monitoring biometric data during tax document processing and/or preparation of an electronic tax return. One embodiment involves detection of pulse rate or change thereof, which may indicate user reaction or emotion during various stages of preparing an electronic tax return including recording of video of documents and/or voice descriptions thereof, electronic dialogue and preparation of the electronic tax return. For example, a sudden increase in pulse rate may indicate user frustration or apprehension, and embodiments may determine appropriate responses to alleviate user concerns or provide positive news about the user's refund, which may be detected or confirmed by a detected reduction in pulse rate or return to a prior pulse rate.

[0006] Certain embodiments are directed to receiving user-provided voice descriptions of tax documents such as a description of the type of document, data within the document, source of the document, or other description of or explanation why the document is relevant to the electronic tax return. User-provided voice descriptions can be utilized in conjunction with video or photograph processing to identify or confirm identification of a tax document and can be used as part of the electronic dialogue about tax documents being processed or to be processed.

[0007] Certain embodiments are directed to processing of a single video of one tax document or processing of a single video of multiple tax documents (e.g., a continuous video taken of different tax documents, and the user moving the video acquisition or computing device to scan different tax documents, or placing different tax documents in front of the video acquisition or computing device). Thus, embodiments may be utilized to partially prepare an electronic tax return based on

video and/or voice processing, with the remaining sections being manually completed by the user as necessary, or an electronic tax return can be completed based exclusively on processing of video or video and voice data without requiring the user to manually enter data, thus greatly simplifying and expediting preparation
5 of electronic tax returns.

[0008] Certain embodiments are directed to utilizing multiple and opposite facing video or image capture elements such as front and rear facing cameras of a mobile communication device such as a smartphone or tablet computing device. With
10 embodiments, the user can acquire video or photographs of one or more tax documents with one video or image capture element while the opposite facing video or image capture element is used to detect facial expression or gesture of the user. The user may also speak to provide voice data that is recorded and processed according to embodiments. Thus, different video or image capture elements can be used for different purposes, one of which does not involve imaging or video of a tax
15 document. Such embodiments may be executed, for example, during electronic dialogue generated according to embodiments such that system generated or selected responses related to facial expressions or gestures indicative of reaction or emotion can be presented to the user. For example, the rear-facing camera may be used to detect that user frustration or apprehension as indicated by a facial
20 expression of an inquisitive look or frown, or change from a smile to a frown, and appropriate responses can be generated or selected and presented by the tax preparation application to the user to alleviate the detected concerns of the user.

[0009] Certain embodiments are directed to generating and presenting an electronic dialogue that is presented to the user while the user is taking a video of
25 tax documents and/or verbally describing tax documents. The electronic dialogue may be in the form of follow up questions or instructions given a determined tax document type and/or voice input spoken by the user. For example, if a tax document involves data indicating the user is married, embodiments may determine, with reference to a tree structure, table or other data structure, follow up questions to
30 be audibly presented to the user. With the example of the user being married, follow-up questions may include, for example, questions about children of the user and deductions related to home ownership. Thus, the user may take a video of tax documents, while verbally describing those documents, and respond verbally or via the computing apparatus or mobile communication device to questions or

instructions of the generated electronic dialogue presented by the tax preparation application. The questions or instructions selected for the electronic dialogue may also be pre-determined and based on the user's response or input, facial expression or reaction and/or biometric data.

5 [0010] Certain other embodiments are directed to allowing a user to provide a voice description of tax documents such as a description of the type of document, data within the document, source of the document, or other description of why the document is relevant to the electronic tax return. These voice descriptions can be utilized in conjunction with video and photograph processing to identify or confirm
10 identification of a tax document.

[0011] Certain other embodiments are directed to various combinations of embodiments discussed above. For example, certain embodiments may involve one or more or all of tax document type determination, tax document character recognition and population of fields of an electronic tax return. As a further example,
15 certain embodiments may involve one or more or all of processing a video of a tax document, processing voice data of a user, which may be part of a video file or clip or processed independently of a video file or clip, detection of facial expressions or gestures, and detection of biometric data and electronic dialogue.

[0012] A further embodiment is directed to a computer-implemented method for
20 populating an electronic tax return and comprises receiving an image of a tax document and receiving voice data of a user of a tax preparation application executed to prepare the electronic tax return. The method further comprises analyzing the image and the voice data to determine a document type and tax data contained within the image, identifying a field of the electronic tax return to be
25 populated with determined tax data of the determined document type, and populating the field with the determined tax data. For example, one embodiment is directed to a computer-implemented method for populating an electronic tax return and comprises a computing apparatus acquiring and/or receiving a video of a tax document. The video includes a plurality of frames and voice data of a user preparing the electronic
30 tax return, or the voice data may be separate from the video frames. The method further comprises analyzing at least one frame and the voice data to determine the type of tax document captured in the recorded video and tax data contained therein. The method further comprises identifying a field of the electronic tax return to be

populated with determined tax data of the determined document type and populating the field with the determined tax data.

[0013] Yet other embodiments are directed to systems configured to perform video/audio processing of tax documents and populating an electronic tax return.

5 For example, a computing apparatus or processor thereof, e.g., in the form of a computer, mobile communication device, or tablet computing device may be configured to execute embodiments, which may be a module or component of a tax preparation application. Systems may include only the computing apparatus, or the computing apparatus and one or more databases or tables that are used, for example, for the purpose of comparing a detected structure of a tax document in a
10 photograph or video frame and pre-determined tax document templates, or to compare voice data segments and pre-determined voice segments or words for the purpose of determining the meaning or context of a verbal description of the tax document provided by the user. System embodiments may also include a camera or
15 other video or image capture element if such elements are not part of or integrated within the computing apparatus.

[0014] Further embodiments are directed to articles of manufacture or computer program products comprising a non-transitory, computer readable storage medium having instructions which, when executed by a computer or computing apparatus,
20 cause one or more processors of the computer or computing apparatus to execute a process for processing video and/or audio data of a tax document and populating an electronic tax return according to embodiments.

[0015] In a single or multiple embodiments, the computing apparatus receives a previously recorded video. In other embodiments, the computing apparatus is
25 involved in recording or acquiring the video. For example, a video may be previously recorded and stored on and received from a memory card (such as a memory card for a camera or camcorder). The video can also be acquired and processed by the same component such as a mobile communication or tablet device having one or more cameras for video acquisition. The video of one or more tax documents and
30 voice data may be simultaneously recorded, received and stored by the computing apparatus. The video and voice data may also be acquired or recorded separately and subsequently synchronized or associated with each other as needed based on, for example, timestamp and elapsed time data.

[0016] In a single or multiple embodiments, individual video frames or static images of a video are analyzed to determine the type of tax document that was captured in a frame or image. The document type can be determined by determining a structure of the tax document that was captured (e.g., structure of boxes, lines, symbols, dimensions, spacing, or other structural attributes) and comparing that determined structure to respective structures of pre-determined tax document templates. A template having a structure matching or that most closely matches the determined structure can be selected, and the tax document type is determined based on the type of tax document represented by the selected template. If the document type cannot be determined, according to one embodiment, the user-provided voice data is analyzed to determine additional document related or contextual information about the imaged document. Voice data analysis may involve converting voice into text, parsing the text data into segments such as individual words, comparing segments and pre-determined segments (such as selected keywords indicative of particular document types), and determining or confirming the document type based at least in part upon the comparison. If the document type still cannot be determined, frames for that document can be tagged, and the user can be requested to provide input to identify the document type. The determined structure of this document can then be stored as a template for future comparisons.

[0017] In a single or multiple embodiments, a video frame is selected (e.g., the first frame, or the middle frame of a plurality of frames, or a frame having the best optical attribute such as contrast), and that frame or image is analyzed to determine the document type and the tax data, e.g., using template comparisons and a character recognition process or program. Thus, document type and tax data determinations involve different recognition processes or video or frame analyses.

[0018] In a single or multiple embodiments, voice or speech data provided or spoken by the user can be associated or synchronized with video data or frames thereof such that the determined meaning or context of voice data is matched to respective videos or frames thereof. Such association or synchronization may be based on one or more of video and voice data start times or timestamps and elapsed times. Synchronization may also be based on pre-determined voice segments. For example, the user may be taking a video of a first tax document and provides a first verbal description, then switches to taking a video of a second tax document and provides a second verbal description. Synchronization can be based on a pre-

determined term spoken by the user such as “next” or “next document” to signal the beginning of a video and/or voice description of the next or new tax document, or by the user clicking a user interface element such as “done” to signify that the user 215 is finished with a tax document 106 and that any following verbal description is for the next tax document to be recorded.

[0019] In a single or multiple embodiments, processing is performed by a computing apparatus of a mobile communication device such as a smartphone or tablet computing device or other computing apparatus, which executes or accesses a tax preparation application to acquire the video of one or more tax documents 10 using a camera or other video or image capture element. In one embodiment, the video is acquired and received by the computing apparatus as a result of the user positioning the tax document and the mobile communication device such that the tax document is in view of the camera. The video can be acquired while the user speaks, and this voice data is detected by a microphone of the mobile 15 communication device. Other camera or video or image acquisition devices may also be utilized such as a web camera or a camcorder.

[0020] In a single or multiple embodiments, the computing apparatus engages in an interactive electronic dialogue with the user. Dialogue communications may be in response to at least one of the determined document type and determined tax data. 20 For example, when a document type determination is made, the computing apparatus may determine appropriate responses in the form related questions, instructions or explanations. These questions, instructions or explanations may be identified by the computing apparatus referring to a tree structure, database or table identifying the responses to be presented in response to different detected document 25 types or verbal input by the user. The responses can be presented audibly and/or via a display or user interface. For these purposes, the computing apparatus may utilize text-to-speech processing to present the questions, instructions or explanations to the user, and voice recognition to interpret the user’s answers.

[0021] In a single or multiple embodiments, the computing apparatus determines 30 which response to present to the user based at least in part upon a detected facial expression or gesture of the user. The facial expression or gesture can be determined by analysis of a video or frame thereof. This detection may occur during, for example, the electronic dialogue with the user. Corresponding responses to the detected facial expression or gesture can then be determined, e.g., by reference to a

tree structure, database or table, and presented to the user. In certain embodiments, a first video or image capture element of the computing apparatus is used to take a video of a tax document, and facial expression or gesture is performed using a different second video or image capture element of the computing apparatus, e.g., using front and rear facing cameras. In one embodiment, tax document recording and facial or gesture recognition are performed simultaneously using different, opposite facing cameras.

[0022] In a single or multiple embodiments, the computing apparatus determines which response to present to the user based at least in part upon a detected biometric data. In one embodiment, the biometric data is the user's pulse. Other biometric data may include the user's temperature, skin color (e.g., whether the user is pale or red) or change thereof, and whether the user is perspiring. These types of biometric data are indicative of a certain feelings or reactions to embodiment processing, electronic dialogue or certain questions or explanations presented to the user.

[0023] In one embodiment, the biometric data such as the pulse or temperature of the user of a mobile communication device is detected while the user is holding the mobile communication device. In other embodiments, the user may have or be wearing a biometric sensor, or such sensors may be integrated within the user's clothing, and communications between the worn sensor and a mobile communication device may be by wireless communications such as BLUETOOTH communications. Embodiments may involve detection of one or multiple types of biometric data, and in the case in which multiple types of biometric data are acquired, analysis and processing may be based on, for example, a weighted average or other statistical methodology or priority to a certain type of biometric data if a detected reading surpasses a pre-determined threshold.

DESCRIPTION OF THE DRAWINGS

[0024] The foregoing and other aspects of embodiments are described in further detail with reference to the accompanying drawings, wherein:

[0025] Fig. 1 is a block diagram generally illustrating video/voice processing for preparation of an electronic tax return according to embodiments;

[0026] Fig. 2 is a block diagram of system components utilized in embodiments for processing video and voice data to prepare an electronic tax return;

[0027] Fig. 3 illustrates an embodiment of a method for populating an electronic tax return based at least in part upon results of analyzing video frames and voice data;

5 [0028] Fig. 4 is a system flow diagram showing receiving or acquisition of video and voice data, processing or analysis of video and voice data; mapping to an electronic tax return, and population of an electronic tax return according to embodiments;

[0029] Fig. 5 illustrates an example of a tax document that can be processed according to embodiments, and how a tax document may be structured;

10 [0030] Figs. 6A-C illustrate how a mobile communication device can be used to record video and user descriptions of multiple tax documents according to embodiments;

[0031] Fig. 7 illustrates an embodiment of a method for determining a type or name of a tax document captured in a video;

15 [0032] Fig. 8 illustrates steps that may be executed for synchronizing or associating video frames and voice data according to embodiments;

[0033] Figs. 9A-B generally illustrate how video frames and voice data can be synchronized or associated with each other during a continuous recording or recording of different video clips according to embodiments;

20 [0034] Fig. 10 illustrates how fields of a determined tax document are mapped to fields of an electronic tax return or interview screens generated by a tax preparation application according to embodiments;

[0035] Fig. 11 illustrates an embodiment of a method for determining a type or name of a tax document based at least in part upon comparisons of determined tax document structure and tax document templates according to embodiments;

[0036] Fig. 12 further illustrates comparison of determined tax document structure and tax document templates according to embodiments;

[0037] Fig. 13 illustrates a method for processing voice data according to embodiments;

30 [0038] Fig. 14 further illustrates one manner in which voice data can be processed according to embodiments;

[0039] Fig. 15 illustrates a table or database including data of voice data text segments and information about how the text segments relate to tax documents;

[0040] Fig. 16 illustrates steps for populating, formatting and filing electronic tax returns prepared according to embodiments;

[0041] Fig. 17 illustrates an embodiment directed to generating an interactive electronic dialogue and communicating with a user;

5 [0042] Fig. 18 generally illustrates a tree structure of tax-related topics that may be utilized to determine questions, instructions or explanations to present to the user as part of an interactive electronic dialogue;

[0043] Fig. 19 illustrates an embodiment of a method for utilizing multiple cameras for recording a document and performing facial expression or gesture
10 recognition and determining responses to be communicated to the user based at least in part upon the detected facial expression or gesture;

[0044] Fig. 20 illustrates how a user may manipulate a mobile communication device with opposite facing cameras to record a video of a tax document with one camera while the user's facial expression or gesture recognition is detected with
15 another camera;

[0045] Fig. 21 illustrates a table or database including data of facial expressions or gestures and corresponding responses to be communicated to the user;

[0046] Fig. 22 illustrates an embodiment of a method for detecting biometric data of the user and determining responses to be communicated to the user based at
20 least in part upon the detected biometric data;

[0047] Fig. 23 illustrates a table or database including data biometric attributes and data thereof and corresponding responses to be communicated to the user; and

[0048] Fig. 24 is a block diagram of components of a computing apparatus or system in which various embodiments may be implemented or that may be utilized to
25 execute various embodiments.

DETAILED DESCRIPTION OF ILLUSTRATED EMBODIMENTS

[0049] Embodiments are related preparing at least a portion of an electronic tax
30 return based at least in part upon processing a video taken of tax documents, or by processing video and voice data describing the tax document or how the tax document is relevant to the electronic tax return.

[0050] As generally illustrated in Fig. 1, according to one embodiment, a video 100 is taken of a tax document 106. The video 100 includes frames 102 and

associated audio data 104 in the form of speech or voice data 104 spoken by a user, e.g., spoken by the user while the user records the video 100 of the tax document 106. The video 100 is provided to a computing apparatus or processor 110, which analyses video frames 102 and voice data 104. Results of the video/voice analyses
5 may be applied to populate fields or sections of an electronic tax return 120 or interview screens of a tax preparation application utilized to prepare the electronic tax return 120 with determined tax data contained within the tax document 106.

[0051] Thus, embodiments allow users to record and provide data about tax documents 106 in a manner that is natural to the user, while also being more flexible
10 and convenient compared to traditional tax preparation programs that are based on non-verbal, display-based question and answer interaction. Not only do embodiments provide for a more natural and flexible tax return preparation experience, but embodiments are particularly suited for use with smaller devices such as mobile communication and computing devices such as smartphones and
15 tablets. Such mobile devices often have smaller display screens or display screens such that display-based question and answer interaction may be challenging or inconvenient. Embodiments are also able to solve this unmet need by leveraging video and audio capture components of the mobile communication or computing device and allowing the user to use the mobile device to take videos of tax
20 documents while describing how the tax documents are relevant to the electronic tax return. A user may begin preparation and even complete preparation of an electronic tax return utilizing video, or video and voice data acquired via a mobile communication device and processing according to embodiments, without typing tax data into fields of the electronic tax return or interview screen generated by the tax
25 preparation program. In the event that certain sections of the electronic tax return are not amenable to video or voice processing and tax return population, or if the tax document is not recognized, users can manually enter tax return data using the mobile or other device as desired. Further aspects of embodiments are described with reference to Figs. 2-24.

[0052] Fig. 2 illustrates an embodiment of a system 200 or components thereof
30 that are utilized or involved in execution of method embodiments. In the illustrated embodiment, the system 200 includes or involves a computer, computing apparatus, computing device or mobile communication or computing device 210 (generally, mobile communication device, computing device or computing apparatus 210) of a

user 215. The user 215 may be a taxpayer, preparer or accountant. For ease of explanation, reference is made to user 215. Mobile communication device is defined to include mobile communication and computing devices capable of communications such as smartphones, a laptop computer, a Personal Digital Assistant (PDA) and a tablet device. For ease of explanation, reference is made generally to mobile communication device 210 or computing apparatus 210.

[0053] The system 200 also comprises or involves one or more intermediate computers 220 (one intermediate computer 220 is illustrated for ease of explanation) of one or more hosts 215, a computer 230 of a tax authority 235, and a data store or database 240 (generally, data store 240) accessible by the mobile communication device 210 locally or via a network connection (not shown in Fig. 2).

[0054] The intermediate computer 220 or another intermediate computer can serve as an electronic filing server or clearing house, formats and electronically files the completed electronic tax return 120 with the tax authority computer 230. The tax authority 235 with which tax returns 120 are filed may be a federal tax authority such as the Internal Revenue Service (IRS), a state tax authority such as the State Franchise Board of California or other tax collecting entity to which taxes are paid. For ease of explanation, reference is made generally to a tax authority 235, and examples involve the IRS, but it will be understood that tax authority 235 refers to various tax collecting entities, including local, state and federal tax collecting entities, and embodiments may apply to generate estimates of when a taxpayer can expect to receive refunds from one or more or all of local, state and federal tax authorities. The intermediate computer 220 or processes acknowledgements and other data received from the tax authority 235 and routes related information back to the computing apparatus 210. One example of an intermediate computer 220 or electronic filing server that may be utilized for these purposes is an electronic filing server of Intuit Inc., Mountain View, California.

[0055] In the illustrated embodiment, a video/voice processor 214 is utilized to analyze video and/or voice data according to embodiments and may be part of or in communication with the tax preparation application 211, one example of which is TURBO TAX tax preparation application, available from Intuit Inc., Mountain View, California. TURBO TAX is a registered trademark of Intuit Inc. The tax preparation application 211 may be utilized by taxpayers for the purpose of preparing a tax or information return or other tax filing, and that may be operable to record, transmit,

transfer and organize data related to such tax filings, thus enabling taxpayers to file income tax returns or assisting them with managing their finances or tax-related matters. In embodiments involving a mobile communication device 210, the tax preparation application 211 may be in the form of a mobile application executing on the mobile communication device 210 or hosted by the intermediate computer 220 and accessible by the mobile communication device 210 executing a browser. It will be understood that the tax preparation application 211 and video/voice processor 214 (generally, processor 214) may execute locally on the mobile communication device 210 or accessed via a network connection.

[0056] While reference is made to a video/voice processor 214 ("processor" 214 generally), as shown in Fig. 2, the processor 214 may be used processing of different types of data according to embodiments. As described in further detail below, the processor 214 may be used for video frame processing, voice processing, video-voice synchronization, document structure recognition, character recognition, electronic dialogue, text to speech conversion, speech recognition, facial expression detection and biometric data detection, and for these purposes, may access respective programs and data hosted by the data store 240 as necessary. For ease of explanation, reference is made to a processor 214 or to a processor associated with a particular processing function such as voice processor that executes various aspects of embodiments.

[0057] As illustrated in Fig. 2, the mobile communication device is in communication through a network 250a with the intermediate computer 220, which may also be an electronic filing server operable to electronically file tax returns with the tax authority computer 230 and is in communication through a network 250b with the tax authority computer 230. Each of the networks 250a-b and other networks discussed herein (generally network 250) may be different, or two or more networks 250 may be the same depending on the system configuration and communication protocols employed. One or more or all of the networks 250 may be, for example, a cellular network, a wireless network, a Local Area Network (LAN) and/or a Wide Area Network (WAN) or a combination of two or more networks. Reference to a network 250 generally is not intended to refer to a specific network or communications protocol, and it should be understood that embodiments can be implemented using various networks and combinations thereof.

[0058] As shown in Fig. 2, the mobile communication device 210 also includes one or more cameras 216 or image or video capture elements (generally camera 216), operable to take videos and photographs of tax documents 106. The mobile communication device 210 also includes a microphone 217 ("Mic" in Fig. 2) that allows the user 215 to verbally describe information about the tax document 106 that is being recorded. The resulting video 100 may include both video frames 102 and voice data 104, or the voice data 104 may be acquired independently of the video 100 depending on how the computing device 210 is configured. Having generally described systems components, aspects thereof and their functionality and interoperability, various embodiments and aspects thereof are described in further detail with reference to Figs. 3-24.

[0059] Referring to Fig. 3, and with further reference to Fig. 4, in a method according to one embodiment, at 302, a video, video clip or file (generally, video 100) of one or more tax documents 106 is received by the processor 214. In the illustrated embodiment, the video 100 includes a plurality of frames 102 and voice data 104 of the user 215.

[0060] For example, referring to Fig. 5, the tax document 106 may be a Form W-2 tax document that includes tax data 122 such as the user's employer, wages and federal income tax withheld. While Form W-2 is provided as one example of a tax document 106 that may be processed according to embodiments, it will be understood that many other tax documents 106 may also be utilized.

[0061] Referring again to Fig. 4, according to one embodiment, the same mobile communication device 210 is used to acquire the video frames 102 and voice data 104 using the camera 216 and microphone 217 of the mobile communication device 210. For example, referring to Figs. 6A-C, the user 215 may have three tax documents 106a-c that are positioned on a desk or table 600. The user 215 positions the mobile communication device 210 such that the first tax document 106a is in view of the camera 216, initiates acquisition of a video 100, and speaks during the video to describe the first tax document 106a or how the first tax document 106a is relevant to the electronic tax return 120. After the user 215 has completed describing the first tax document 106a, the user 215 may move or direct the mobile communication device camera 216 to the next or second tax document 106b, continues the video or initiates a second video, and speaks during the video to describe the second tax document 106b or how the second tax document 106b is

relevant to the electronic tax return 120. Similarly, after the user 215 has completed describing the second tax document 106b, the user 215 moves or directs the mobile communication device camera 216 to the next or third tax document 106c, continues the video or initiates a third video, and speaks during the video to describe the third tax document 106c or how the third tax document 106c is relevant to the electronic tax return 120.

[0062] The user 215 may also have a stack of multiple tax documents 106, acquire a video and describe the first tax document 106a, then remove the first tax document 106a from the top of the stack to show the next or second tax document 106b, and so on for additional tax stacked tax documents 106. The user 215 may also hold tax documents in the air with one hand and hold the mobile communication device 210 in the other and record a video 100 while speaking. Thus, it will be understood that there are various ways the user 215 can be positioned or move, the mobile communication device 210 can be positioned or angled, and the tax documents 106 can be arranged or stacked in order to record videos 100 of tax documents 106 while user 215 describes the tax documents 106.

[0063] In another embodiment, the video 100 is received by the processor 214, e.g., read from a memory card of a camera or camcorder that was utilized to acquire and store the video 100, or in an electronic message from a computer having a web camera that was utilized to acquire the video 100 of tax documents 106. For ease of explanation, and not limitation, reference is made to acquiring, storing and processing a video 100 with the mobile communication device 210, but it will be understood that embodiments are not so limited.

[0064] Referring again to Fig. 3, and continuing with step 304, the processor 214 analyzes one or more frames 102 of the video 100 and the associated voice data 104 to determine the type of the tax document 106 and tax data 122 within the video frame 102 that is identified and used to populate fields of the electronic tax return 120.

[0065] Referring to Fig. 7, at 702, before video and voice analysis is performed and depending in the configuration of system components utilized, one or more frames 102 and voice data 104 may need to be synchronized or associated with each other if this has not already been done as a result of creation of the video 100. For this purpose, referring to Fig. 8, synchronization may involve one or more of steps 802-808 involving determining the starting times or when the video frames 102

and voice data 104 began, filtering frames 102 during periods of silence (no voice data spoken by user 215), determining the lengths or elapsed times of the video 100, frames 102 or voice data 104, or detecting a trigger or keyword indicating a transition to recording of a different tax document 106. For example, a trigger or keyword may be the user 215 speaking "next" or "next document" which is detected and identified for purposes of matching 810 or synchronization a video frames 102 of a corresponding tax document 106 and related voice data 104. The user may also indicate the end and/or beginning of recording of tax documents by tapping a user interface element displayed on a screen of the mobile communication device 210.

Thus, it will be understood that one or more video frames 102 can be matched at 810, to or associated or synchronized with voice data 104 of the user 215, and that one or more of steps 802-808 may be performed and are provided as examples of how synchronization may achieved if needed. Further, it will be understood that this matching or synchronization may be performed with respect to a continuous video taken of one or more tax documents 106 or with individual videos or clips for each tax document 106, as generally illustrated in Figs. 9A-B, wherein Fig. 9A shows how voice data 104a-b can be synchronized with respective groups of frames 102a-b during a continuous video recording, and Fig. 9B shows how voice data 104a-b can be synchronized with respective groups of frames 102a-b of multiple video clips.

[0066] Referring again to Fig. 3 (304) and Fig. 7 (704), the processor 214 analyzes video frames 102 and/or voice data 104 (video and voice processing described in further detail below with reference to Figs. 11-16) to confirm the determination of the type of tax document 106, modify the document type determination, or determine the document type if the type cannot be determined via video analysis. According to one embodiment, the type of the tax document 106 is determined based on analysis of one or more video frames 102 such that when video frames 102 are to be analyzed, one or more frames 102 can be selected for analysis. Frame 102 selection can be based on, for example, criteria such as the first frame, a group of frames at the beginning of a recording, the first frame in which it is detected to include an image of tax document 106, a frame within a middle portion of the video or video clip (e.g., if it is assumed that this is when the user 215 is most focused on the tax document 106 after initial set up or introduction). Frame 102 selection may also be based on one or more frames having the best or having certain optical quality such as frames 102 having the highest contrast or clarity as

determined by an optical analysis of frames 102 or application of an optical character recognition (OCR) process to the frames 102 and selection of one or more frames with the lowest error rates during recognition. Thus, while certain embodiments are described with reference to performing an OCR process after document type identification, it will be understood that OCR can be performed as part of the document type identification.

[0067] With continuing reference to Fig. 7, at 706, user input can be requested to confirm the document type determination, confirm a document type determination, or modify document type determination. At 708, and with further reference to Fig. 3 (306) and Fig. 10, having determined the document type, fields of that tax document 106 can be mapped 1002 to corresponding fields of the electronic tax return 120 or interview screens of the tax preparation application 211 utilized to prepare the electronic tax return 120 such that when the tax data 122 within the document 106 is determined, that determined tax data 122 can be populated within respective fields according to the mapping 1002. Further details regarding how document type determinations can be performed using video and/or voice processing are described with reference to Figs. 11-16.

[0068] Referring to Figs. 11 and 12, having selected one or more frames 102 of the video 100 of one or more tax documents 106 that was received or acquired, one or more frames 102 are processed to determine what type of tax document 106 has been imaged or recorded. For this purpose, at 1102, the processor 214 determines the configuration or structure 1201 of the tax document 106 that appears within the frame 102. The configuration, structure 1201 or other identifying data may involve boxes, lines, sections, symbols, and dimensions, location and arrangement thereof.

[0069] At 1104, the processor 214 accesses the data store 240 containing templates 1202 of pre-determined tax documents 106. For ease of explanation and illustration, four tax document templates 1202a-d are illustrated, but it will be understood that there may be hundreds and thousands of templates 1202. For example, referring again to Fig. 5, Form W-2 has a certain dimensions, field configuration and dimensions, and a template 1202 may include the structure of Form W-2 (boxes, lines) and any unique identifiers or symbols or alpha-numeric data (e.g., popular form names).

[0070] At 1106, and as generally illustrated in Fig. 12, the processor 214 compares the determined structure 1201 and the templates 1202a-d (generally,

template 1202), selects a template 1202 as a match or closest match, and at 1110, determines the type of the document 106 that was recorded based at least in part upon the document type of the selected template 1202.

[0071] According to one embodiment, processing of one or more frames 102 or static images of a video 100 may be used to determine the type of tax document 106 within the video or frame 102. According to another embodiment, the received voice data 104 is used to provide additional contextual information or meaning to the video analysis in order to determine what type of tax document 106 is captured in the video 100. Voice data 104 may include various types of information about the tax document 106 or about its relevance to the electronic tax return 120 such as describing the tax document 106 by type or name, the source of the tax document 106, or types of tax data 122 within the tax document 106. For example, the user 215 may take a video 100 of a first tax document 106a and explain "This is my W2 from my prior employer" and then when taking a video of a second tax document 106b, continue explaining "This is my W2 from my current employer – I found a new job last year." As another example, the user 215 may explain "I got married" or "We had our second child last year."

[0072] Referring to Fig. 13 and with further reference to Fig. 14, one method for processing voice data 104 for this purpose includes, at 1302, the processor 214 receiving the analog voice data 204 and applying speech recognition or voice to text conversion 1402. At 1304, the processor 214 parses 1404 the text into segments as necessary (referred to as segment or voice segment). A segment may be an individual word or groups or words having number of words less than a pre-determined number of words. At 1306, the processor 210 may filter 1406 or remove certain segments so that subsequent processing is applied to selected segments and not filtered segments. Filtering 1306 may be applied to certain words, words beginning with a certain letter or words having a certain number of letters, e.g., based on a determination that when a user 215 describes the most common tax documents 106, such descriptions do not involve these filtered segments, or do so a small percentage of the time. For example, the voice processor may filter "a," "the" or words determined as not providing useful information about the tax document 106, whereas terms such as "married," "children," mortgage," "interest," "deduction," "wages," "new," names of tax documents would not be filtered, and the processor 214 may also be configured to access the data store 240 containing a list of table of

selected segments (such as “married,” “children” etc.) that should not be filtered and should be subsequently processed such as names of tax documents (W2, 1099) or other information, terms, phrases or segments determined (e.g., empirically, or based on terms appearing within tax documents) or selected as being pertinent to
5 the tax document 106.

[0073] At 1308, and with further reference to Fig. 15, a comparator 1408 of the processor 214 accesses the data store 240 containing table or database 1500 including columns 1501-1503 for pre-determined voice segments and corresponding meanings or contextual information associated with those voice segments and/or
10 relevant or potential relevant tax documents 106 associated with those segments. The meanings or contextual information may, for example, be created by the administrator of the system or determined empirically. Thus, as shown in Fig. 15, the table or database 1500 includes column 1501 with various segments and their corresponding meaning or context or relevance to tax documents 106 or
15 identification of a tax document 106. At 1310, the comparator 1408 compares voice segments of the voice data 104 and the pre-determined voice segments in column 1501 of the table or database 1500, and at 1312, identifies any matches, and selects those rows or table entries and determines the corresponding meaning, context, interpretation or document identification 1408 with the data in column 1502 and/or a
20 relevant or potentially relevant tax document in column 1503 in the table 1500 at 1314.

[0074] Thus, the type of the tax document 106 that was recorded may be determined by analyzing video or frames 102 thereof or by analyzing both video or frames 102 thereof and voice or speech data 104 provided by the user 215 during
25 recording of the video 100 such that the voice or speech data 104 provides additional meaning, identification or confirmation of the video analysis.

[0075] Referring again to Figs. 3 and 10, and with further reference to Fig. 16, having determined the document type and having mapped 1002 sections of the tax document captured within the video frame 102 to fields of the electronic tax return
30 120 or interview screens of the tax preparation application 211, at 1310 / 1602, the processor 214 executes a further recognition process, e.g., Optical Character Recognition (OCR), which is applied to the video frame 102 to determine tax data 122 contained therein. Voice data 104 may also be analyzed for this purpose.

[0076] At 1604, the processor 214 determines corresponding alpha-numeric tax data 122 contained within tax document 106, and at 1606, populates the respective identified fields of electronic tax return 120 or interview screens generated by tax preparation application 211 with determined alpha-numeric tax data 122 according to the determined mapping 1002. At 1608, if the electronic tax return 120 is not completed exclusively with the video or video and voice processing, then those fields can be tagged by the processor 214, and the user 215 can then be presented with those fields or tax return documents and import or manually enter additional tax data into the electronic tax return 120 such that at 1610, the electronic tax return 120 is eventually completed, transmitted to the intermediate computer 220 at 1612, which formats and electronically files the electronic tax return 120 with the computer 230 of the tax authority 243 at 1614.

[0077] Referring to Figs. 17 and 18, a further embodiment, which may be utilized in conjunction with embodiments described above, involves the processor 214 generating questions or responses and engaging the user 215 in an interactive electronic dialogue while the user 215 is involved in one or more steps of embodiments such as recording tax documents 106 or providing verbal descriptions of the tax documents 106. In one embodiment, at 1702, during steps involving video and/or voice data processing involving document type determination, determining the meaning or contextual information of voice data 104, or determining tax data 122 within a tax document 106 during a recognition process, the processor 124 may process the results of these processing steps and at 1704, access the data store 240 containing a tree structure, table or database 1800. The tree structure, table or database 1800 cross-references document type / meaning / recognition results and system responses based on those types of tax documents 106 or tax data 122 that are to be presented to the user 215 as part of electronic dialogue with the user 215.

[0078] For example, the processor 124 may determine from the tax document 106 or recognition results that the user 215 is married. The processor 124 at 1706 then determines a corresponding electronic response to this married determination in the form of follow-up questions on same marriage topic or a different selected topic, explanation, instruction, reference to section of electronic tax return or tax preparation application. For example, as shown in Fig. 18, upon determining that the user 215 is married, the processor 214 may access a tree or other organization data

structure 1800 of follow up questions concerning tax-related issues that often arise with personal information such as marital status 1801, such as whether they have children 1802 (e.g., childcare expenses) or own a house 1803 (e.g., mortgage interest paid, property tax paid). While Fig. 18 illustrates a tree structure with follow up questions concerning two topics or categories, it will be understood that the tree structure may involve more topics and more detailed questions and further follow-ups for each category.

[0079] At 1708, having identified the electronic response(s) to be presented to the user 215, the processor 214 performs text to voice conversion such that the response can be verbally presented to the user 215 via a speaker of the mobile communication device or other computing apparatus utilized. This is particularly beneficial when users 215 have a mobile communication device since such devices often have smaller screens such that it may be difficult or less convenient to display the follow up questions via a user interface displayed on a screen of the mobile communication device, but it will be understood that the follow-up questions presented by the processor 214 can also be displayed to the user 215.

[0080] In the embodiment involving electronic dialogue in audible form, at 1710, the user 215 answers the computer generated question or provides an answer to a computer generated description, instruction or explanation. For example, the computer generated response or question may be "I noted you are married, do you have children?" in response to which the user 215 would audibly answer "Yes" or "No" for example, or provide a more detailed explanation "Yes, and we adopted a child last year." At 1712, the processor 214 converts the voice data or answer 104 into text and processes the text to determine the meaning of the user's answer, and determine subsequent follow up questions or responses as necessary, such as asking about additional tax documents that may be related to the user's answers. For example, if the user responded "Yes, and we adopted a child last year" the processor 214 may continue with additional follow-ups with audible questions concerning potential tax deductions or credits pertaining to adoption or other tax topics that may be related to having a child such as being married, deductions for mortgage interest, etc.

[0081] While certain embodiments involve determining what to ask or explain to a user given determined tax documents 106 or tax data 122, the interactive electronic dialogue may also be based on questions or statements made by the user 215.

Accordingly, it will be understood that Fig. 17 is provided as one example of how embodiments may be implemented, and that the electronic dialogue may be keyed to document type or tax data determinations and/or other voice data provided by the user 215.

5 [0082] Referring to Figs. 19-21, a further embodiment that may be utilized in conjunction with various embodiments described above involves the processor 214 detecting a facial expression, gesture, or other non-verbal communication of the user 202 (generally, facial expression 2002) during various steps described above such as during recording the video 100, the user 215 speaking or listening to follow-up
10 questions presented by the processor 214 during the interactive electronic dialogue, or the user 215 answering such follow-up questions, and the processor 214 initiating an action such determining and presenting an explanation or instruction for the user 215 in response to the detected facial expression 2002.

[0083] For this purpose, in the illustrated embodiment involving a mobile
15 communication device 210 such as a smartphone or a tablet computing device, different cameras or image/video acquisition elements (generally, cameras 216a, 216b) may be utilized – one for taking a video of tax documents 106, the other for detecting a facial expression 2002 of the user 215. Facial expressions 2002 may involve the user's eyes, lips/mouth, forehead, for example, and may be detected by
20 analyzing the facial structure or attributes within a video frame 102 and comparing that detected structure or attribute to pre-determined facial expression templates or other resource.

[0084] In the illustrated embodiment, the mobile communication device 210 includes a first or front facing camera 216a and a second or rear facing camera
25 216b. The video 100 or images acquired by these cameras 216 are provided to the processor 214 for analysis and facial expression 2002 detection, one embodiment of which is described below.

[0085] At 1902, the first camera 216a of the mobile communication device 210 or other computing apparatus (e.g., front facing camera facing away from user 215 and
30 directed to tax document 106) is used to acquire a video 100 of one or more tax documents 106 as described above. At 1904, during one or more stages of embodiments involving recording of video or voice/speech, document type determination, determining a meaning or contextual information of voice data, and determining tax data 122 within a tax document 106 resulting from recognition

process, at 1906, the second camera 216b (e.g., rear facing camera facing toward user's face and away from tax document 106) is used to take a video or photograph of the user's face. As discussed above, the video may include frames 102 comprising static images. At 1908, the processor 124 detects a facial expression 5 2002 or change thereof (e.g., detection of a frown, detection of a smile, detection of a change from a smile to a frown, a change from a frown to a smile, forehead wrinkles, or a smirk). With further reference to Fig. 21, at 1910, the processor 214 accesses the data store 240 containing table or database 2100. The table or database 2100 includes columns 2101-2103 of data for cross-referencing the facial 10 expressions or other non-verbal communications 2002 or changes thereof, and how the processor 214 should respond given the detected facial expression 2002, which is indicative of a meaning, feeling, reaction or emotion. For example, as shown in Fig. 21, embodiments may utilize a table or database 2100 with data or columns 2101-2103 for detected facial expressions or gestures, a description of a 15 corresponding emotion or reaction the user may be experiencing, and how the system should respond to the expression or gesture, or emotion or reaction. In the illustrated embodiment, for example if the user's facial expression 2002 is a frown, expressing frustration or disappointment, the table 2100 may be consulted by the processor 214 to determine that some type of positive reinforcement or message 20 should be communicated to the user 215.

[0086] At 1912, the processor 214 utilizes the table or database 240 to determine how the system should respond to the detected facial expression 2002, given the meaning, feeling or emotion associated with the expression or gesture, and at 1914, the determined question, instruction or explanation of the system response is 25 converted from text to voice, and the voice version of the question, instruction or explanation is then audibly presented to the user 215 as part of the interactive electronic dialogue.

[0087] For example, if the facial expression 2002 is one of fear or worry, such as a frown or facial expression showing tension in the user's face, the processor 214 30 may determine that a response reassuring the user 215 that the process is proceeding as it should can be selected and presented. As another example, if the user 215 changed jobs during the course of a year, and during the interactive electronic dialogue, the processor 214 may verbally explain "User, you will receive a \$1,500 this year, but we were not able to take advantage of your job search

expenses because the IRS requires that those expenses require you to itemize." Upon hearing this news, the user 215 may give a slight frown. Upon detecting the frown, and determining that this facial expression 2002, or change from a smile to a frown has a negative meaning, the processor 214 may respond with explaining
5 "However, the good news is that you were able to receive a larger refund by not itemizing – if you itemized to deduct job search expenses, your refund would be reduced from \$1,500 to \$1,300." In response to hearing this good news, despite not being able to deduct job search expenses, the user 215 nods satisfyingly and/or smiles, and upon detecting this, the processor 214 may determine that it can
10 proceed to electronic dialogue for the next topic. If the initial response does not address the negative facial expression 2002, further responses can be selected and communicated as necessary.

[0088] With continuing reference to Fig. 20, and with further reference to Figs. 22-23, a further embodiment that may be utilized in conjunction with various
15 embodiments described above involves detecting one or more types of biometric data 2012 of the user 215 such as pulse rate of the user 215. Biometric data 2012 may be detected during, 2202, recording a video, speaking, listening to follow-up questions presented by the system, or answering such follow-up questions, and initiating an action or response such as determining and presenting an explanation
20 or instruction for the user 215 in response to the detected facial expression 2002.

[0089] For this purpose, in one embodiment, a camera 216 and/or biometric data or sensor 2010 of the mobile communication device 210 may be utilized to determine or acquire biometric data 2012. While the sensor 2010 is shown in the context of
25 embodiments involving facial expression 2002 detection, embodiments involving facial expressions and biometric data may be utilized independently or separately of each other, or utilized in conjunction with each other. Accordingly, Fig. 20 is merely provided as an example to show the mobile communication device 210 including a
type of sensor 2010 for detecting biometric data 2012.

[0090] Continuing with reference to Fig. 22, at 2204, biometric data 2012 is
30 detected or received by the processor 214. For example, embodiments may leverage the camera 216, sensors 2010 and applications executing on the mobile communication device 210 for detecting a pulse rate of the user 215. In one application, the biometric sensor 2010 is in the form of or involves the user 215 placing a fingertip over a lens of the front facing camera 216 while a light is emitted

by the camera 216. Biometric data 2012 in the form of pulse rate is detected, e.g., using similar techniques utilized in pulse oximeters.

[0091] In other embodiments, the biometric sensor 2010 may be a thermal sensor for detecting biometric data 2012 in the form of temperature such that when the mobile communication device 210 is held by the user 215, or otherwise positioned
5 over a part of the user's body, the temperature of the user 215 is determined. Other biometric data may include the user's skin color (e.g., whether the user is pale or red) or change thereof, and whether the user is perspiring. These types of biometric data 2012 are indicative of a certain feelings or reactions to embodiment processing,
10 electronic dialogue or certain questions or explanations presented to the user 215.

[0092] In one embodiment, biometric data 215 such as the pulse or temperature of the user 215 of the mobile communication device 210 is detected while the user 215 is holding the mobile communication device 215, whereas in other embodiments, the user 215 may have or be wearing a separate biometric sensor (e.g., integrated into
15 clothing), and communications between the worn sensor and a mobile communication device 210 may be by wireless communications such as BLUETOOTH communications. Embodiments may involve detection of one or multiple types of biometric data, and in the case in which multiple types of biometric data are acquired, analysis and processing may be based on, for example, a
20 weighted average or other statistical methodology or priority to a certain type of biometric data if a detected reading surpasses a pre-determined threshold.

[0093] With further reference to Fig. 23, showing the example of biometric data 2012 in the form of pulse rate, the processor 214 may access the data store 240 containing a table or database 2300 with columns 2301, 2302 cross-referencing
25 pulse rate or other biometric data 2012 or changes thereof and corresponding system responses to be presented to the user 215. Upon detecting the pulse rate, the processor 214 searches the database or table 2300 for a match or matching range of pulse rates, to determine the response at 2108. The response 2108 is converted from text to voice and audibly presented to the user 215. For example, if it
30 is determined that the user's pulse rate is high, or increasing relative when the recording and description of tax documents 106 began, the system can identify and present a response related to addressing user 215 concerns or reassuring the user 215 that the video 100 and voice inputs 104 are being utilized to successfully prepare the electronic tax return 120.

[0094] Fig. 24 generally illustrates components of a computing device 2400 that may be utilized to execute embodiments and that includes a memory 2410, account processing program instructions 2412, a processor or controller 2420 to execute account processing program instructions 2412, a network or communications interface 2430, e.g., for communications with a network or interconnect 2440 between such components. The memory 2410 may be or include one or more of cache, RAM, ROM, SRAM, DRAM, RDRAM, EEPROM and other types of volatile or non-volatile memory capable of storing data. The processor unit 2420 may be or include multiple processors, a single threaded processor, a multi-threaded processor, a multi-core processor, or other type of processor capable of processing data. Depending on the particular system component (e.g., whether the component is a computer or a hand held mobile communications device), the interconnect 2440 may include a system bus, LDT, PCI, ISA, or other types of buses, and the communications or network interface may, for example, be an Ethernet interface, a Frame Relay interface, or other interface. The network interface 2430 may be configured to enable a system component to communicate with other system components across a network which may be a wireless or various other networks. It should be noted that one or more components of computing device 2400 may be located remotely and accessed via a network. Accordingly, the system configuration provided in Fig. 24 is provided to generally illustrate how embodiments may be configured and implemented.

[0095] Method embodiments may also be embodied in, or readable from, a computer-readable medium or carrier, e.g., one or more of the fixed and/or removable data storage data devices and/or data communications devices connected to a computer. Carriers may be, for example, magnetic storage medium, optical storage medium and magneto-optical storage medium. Examples of carriers include, but are not limited to, a floppy diskette, a memory stick or a flash drive, CD-R, CD-RW, CD-ROM, DVD-R, DVD-RW, or other carrier now known or later developed capable of storing data. The processor 2420 performs steps or executes program instructions 2412 within memory 2410 and/or embodied on the carrier to implement method embodiments.

[0096] Although particular embodiments have been shown and described, it should be understood that the above discussion is not intended to limit the scope of these embodiments. While embodiments and variations of the many aspects of the

invention have been disclosed and described herein, such disclosure is provided for purposes of explanation and illustration only. Thus, various changes and modifications may be made without departing from the scope of the claims.

[0097] For example, while certain embodiments are describe described with
5 reference to processing video (frames thereof) and voice data based on the user
speaking during recording of the video, embodiments may also involve the user
taking photographs of the tax documents, independently recording verbal
descriptions of the tax documents, synchronizing, matching or associating the
10 photographs with respective voice descriptions, and processing the images and
voice data. Thus, recording of tax documents, via video or photograph, may be done
simultaneously with a user speaking and providing verbal descriptions or
explanations, or the tax document video or image capture may be performed
independently of the user's verbal description.

[0098] While multiple embodiments and variations of the many aspects of the
15 invention have been disclosed and described herein, such disclosure is provided for
purposes of illustration only. Where methods and steps described above indicate
certain events occurring in certain order, those of ordinary skill in the art having the
benefit of this disclosure would recognize that the ordering of certain steps may be
modified and that such modifications are in accordance with the variations of the
20 invention. Additionally, certain of the steps may be performed concurrently in a
parallel process when possible, as well as performed sequentially.

[0099] Accordingly, embodiments are intended to exemplify alternatives,
modifications, and equivalents that may fall within the scope of the claims.

What is claimed is:

1. A computer-implemented method for populating an electronic tax return, the computer-implemented method comprising:

5 a computing apparatus receiving a video of a tax document, the video comprising a plurality of frames and voice data comprising the user speaking about the tax document;

the computing apparatus analyzing at least frame and the voice data to determine a document type and tax data contained within the video;

10 the computing apparatus identifying a field of the electronic tax return to be populated with determined tax data of the determined document type; and

the computing apparatus populating the field with the determined tax data.

2. The computer-implemented method of claim 1, receiving the video comprising the computing apparatus recording the video.

15

3. The computer-implemented method of claim 1, receiving the video comprising reading a previously recorded video from a memory card.

4. The computer-implemented method of claim 1, wherein the video and the voice data are simultaneously recorded, received and stored by the computing apparatus.

20

5. The computer-implemented method of claim 1, at least one frame comprising a static image of the tax document, the static image being analyzed to determine the document type.

25

6. The computer-implemented method of claim 5, the document type being determined based at least in part upon analysis of the static image by:

determining a structure of the tax document within the static image;

30 comparing the determined structure and a plurality of templates; and

selecting a template based at least in part upon the comparison, the document type being determined based at least in part upon the selected template.

7. The computer-implemented method of claim 5, wherein the computing apparatus is unable to determine the document type based on analysis of the static image, the method further comprising analyzing associated voice data to determine the document type.

5

8. The computer-implemented method of claim 7, analyzing associated voice data comprising:

parsing the voice data into voice segments;

comparing voice segments and pre-determined voice segments; and

10

determining the document type corresponding to the matched pre-determined segment based at least in part upon a voice segment matching a pre-determined voice segment.

9. The computer-implemented method of claim 5, wherein the computing apparatus determines the document type based on analysis of the static image and confirms the document type determination based at least in part upon analyzing the associated voice data.

15

10. The computer-implemented method of claim 8, wherein the computing apparatus is unable to determine the document type based on analysis of the static image and associated voice data, the method further comprising the computing apparatus requesting input by the user to identify the document type.

20

11. The computer-implemented method of claim 1, at least one frame comprising a static image of the tax document, wherein at least the static image is analyzed to determine the document type and the tax data.

25

12. The computer-implemented method of claim 11, the document type being determined utilizing a first recognition process, and the tax data being determined utilizing a second recognition process different than the first recognition process.

30

13. The computer-implemented method of claim 12, the first recognition process involving a comparison of a document structure determined from at least one frame, and the second recognition process involving character recognition of data within the at least one frame.

14. The computer-implemented method of claim 1, at least one frame comprising a static image of the tax document, wherein the static image and the voice data are both analyzed to determine the document type.

5

15. The computer-implemented method of claim 14, determining the document type based at least in part upon analysis of the static image and the voice data comprising:

determining a structure of the tax document within the static image;

10

comparing the determined structure and a plurality of tax document templates;

selecting a template based at least in part upon the comparison;

parsing the voice data into voice segments; and

comparing voice segments and pre-determined voice segments,

15

wherein the document type is determined based at least in part upon the selected template and a determination that the voice segment matches the pre-determined voice segment.

16. The computer-implemented method of claim 1, wherein the document type is determined based on a first type of video analysis, and the tax data is determined based on a second type of video analysis.

20

17. The computer-implemented method of claim 1, further comprising the computing apparatus synchronizing or associating the plurality of frames and the voice data.

25

18. The computer-implemented method of claim 17, the plurality of frames and the voice data being synchronized based at least in part upon respective elapsed times from respective start times of the plurality of frames and the voice data.

30

19. The computer-implemented method of claim 1, the video being acquired by a computing apparatus comprising a mobile communication device utilized by the user.

20. The computer-implemented method of claim 19, wherein the video is acquired and received by the computing apparatus of the mobile communication device as a

result of the user positioning the tax document and the mobile communication device such that the tax document is in view of a video capture element of the mobile communication device, the video of the tax document being acquired while the user provides voice data by speaking into a microphone of the mobile communication
5 device.

21. The computer-implemented method of claim 1, the computing apparatus receiving the video by reading the video from a memory card storing the video recorded by a camcorder.

10

22. The computer-implemented method of claim 1, the voice data comprising a user-spoken description of the tax document.

23. The computer-implemented method of claim 1, the voice data comprising a
15 user-spoken description of a source of the tax document.

24. The computer-implemented method of claim 1, further comprising the computing apparatus engaging in an electronic dialogue with the user in response to at least one of the determined document type and determined tax data.

20 25. The computer-implemented method of claim 24, engaging in the electronic dialogue comprising the computing apparatus:

determining a first response based at least in part upon at least one of the determined document type and determined tax data; and
presenting the first response to the user.

25

26. The computer-implemented method of claim 25, the electronic dialogue being executed utilizing text-to-speech and voice recognition.

27. The computer-implemented method of claim 25, the first response being a
30 question, an instruction, an explanation related to the electronic tax return, or identification of a section of the electronic tax return.

28. The computer-implemented method of claim 27, the first response comprising multiple questions, instructions, explanations or identification of multiple sections of the electronic tax return.

5 29. The computer-implemented method of claim 25, the first response being determined with reference to a tree structure or table identifying which response should be presented to the user based on at least one of respective determined document types and respective determined tax data.

10 30. The computer-implemented method of claim 25, further comprising the computing apparatus:

detecting a facial expression or gesture of the user during the electronic dialogue;

determining a second response to the facial expression or gesture; and

15 presenting the second response to the user.

31. The computer-implemented method of claim 30, the second response being determined with reference to a tree structure or table identifying which response should be presented to the user based at least in part upon the detected facial
20 expression or gesture.

32. The computer-implemented method of claim 30, wherein the video of the tax document is acquired with a first video or image capture element of the computing apparatus, and the facial expression or gesture is detected with a second video or
25 image capture element of the computing apparatus.

33. The computer-implemented method of claim 32, the computing apparatus comprising a mobile communication device, the video of the tax document is
30 acquired with the first video or image capture element comprising a front facing camera of the mobile communication device, and the facial expression or gesture is detected with the second video or image capture element comprising a rear facing camera of the mobile communication device, wherein the front facing camera faces

the tax document, and the rear facing camera faces the user while the user is holding the mobile communication device.

34. The computer-implemented method of claim 25, further comprising the
5 computing apparatus:

detecting biometric data of the user during the electronic dialogue;
determining a second response to the biometric data; and
presenting the second response to the user.

10 35. The computer-implemented method of claim 34, the biometric data comprising a pulse rate of the user determined by a detection component of the mobile communication device while the user is holding the mobile communication device.

36. The computer-implemented method of claim 1, further comprising the
15 computing apparatus

identifying fields of the electronic tax return that were not populated with determined tax data; and
presenting the identified fields to the user.

20 37. The method of claim 1, wherein the electronic tax return is completed based solely on tax data determined from the video.

38. The method of claim 1, further comprising the computing apparatus
electronically filing a completed electronic tax return with a computer of a tax
25 authority.

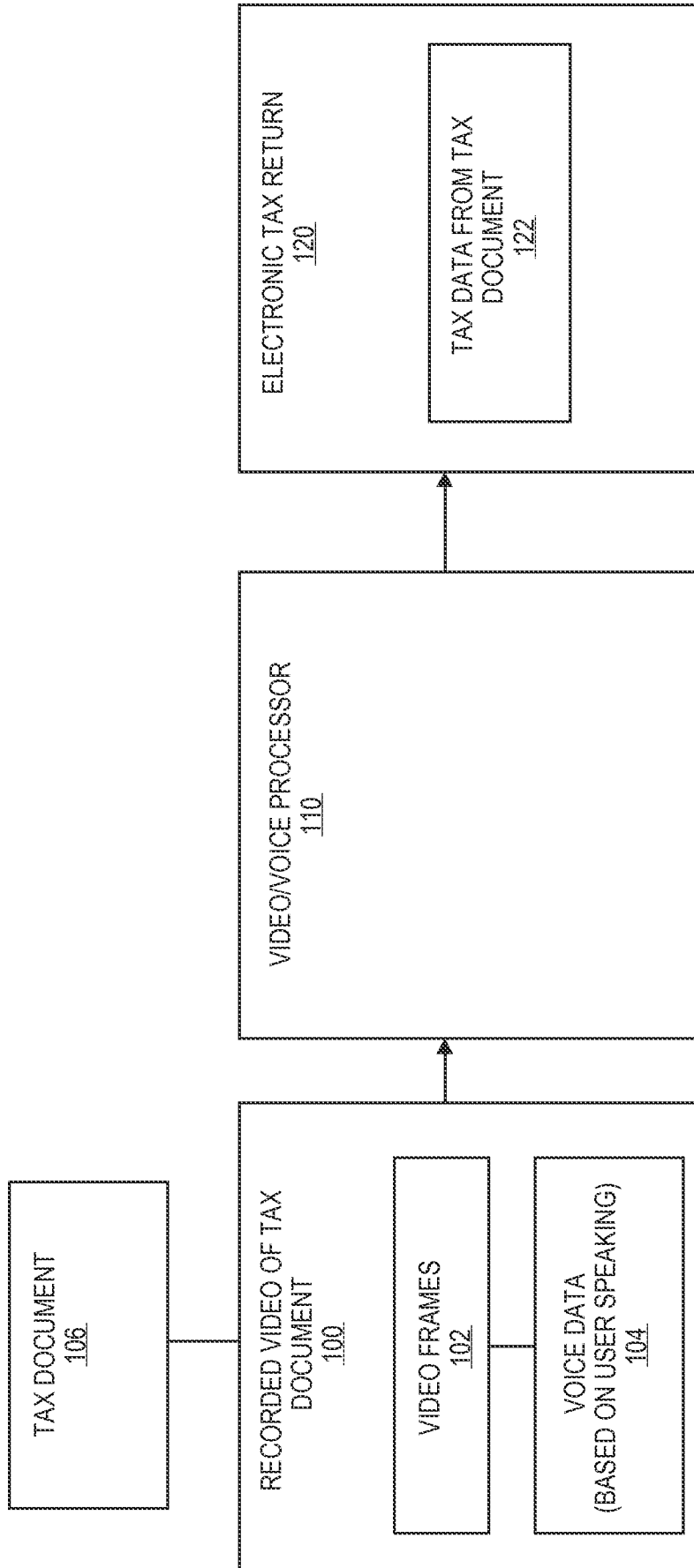


FIG. 1

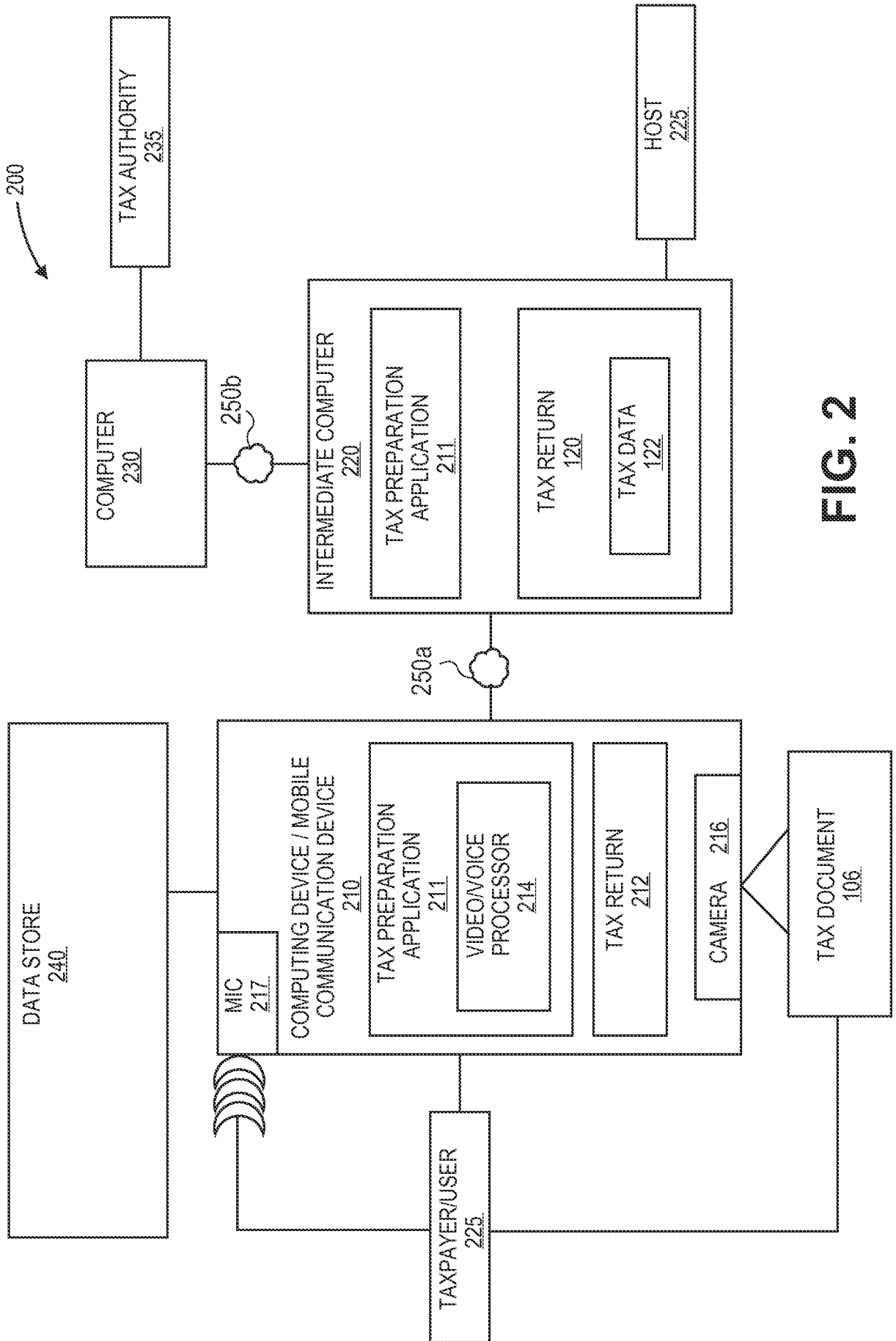
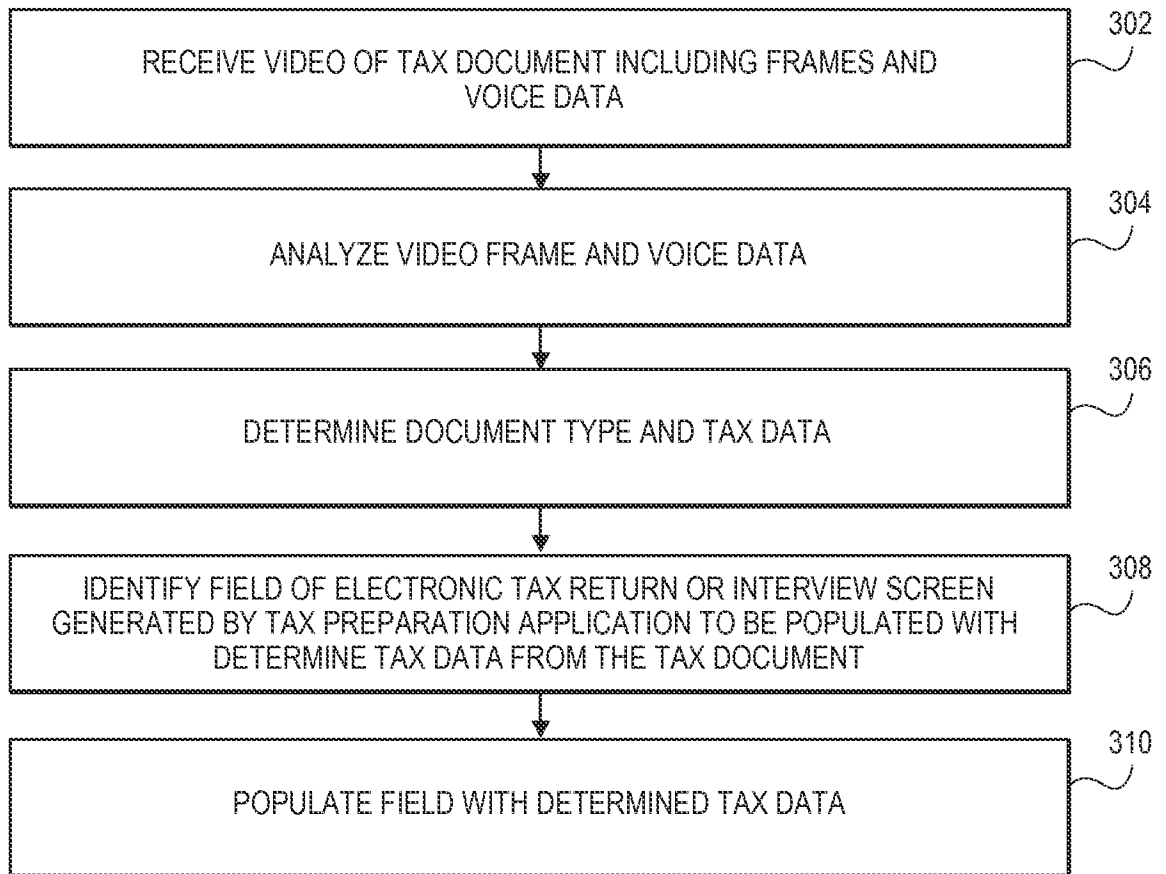


FIG. 2

3/27

**FIG. 3**

4/27

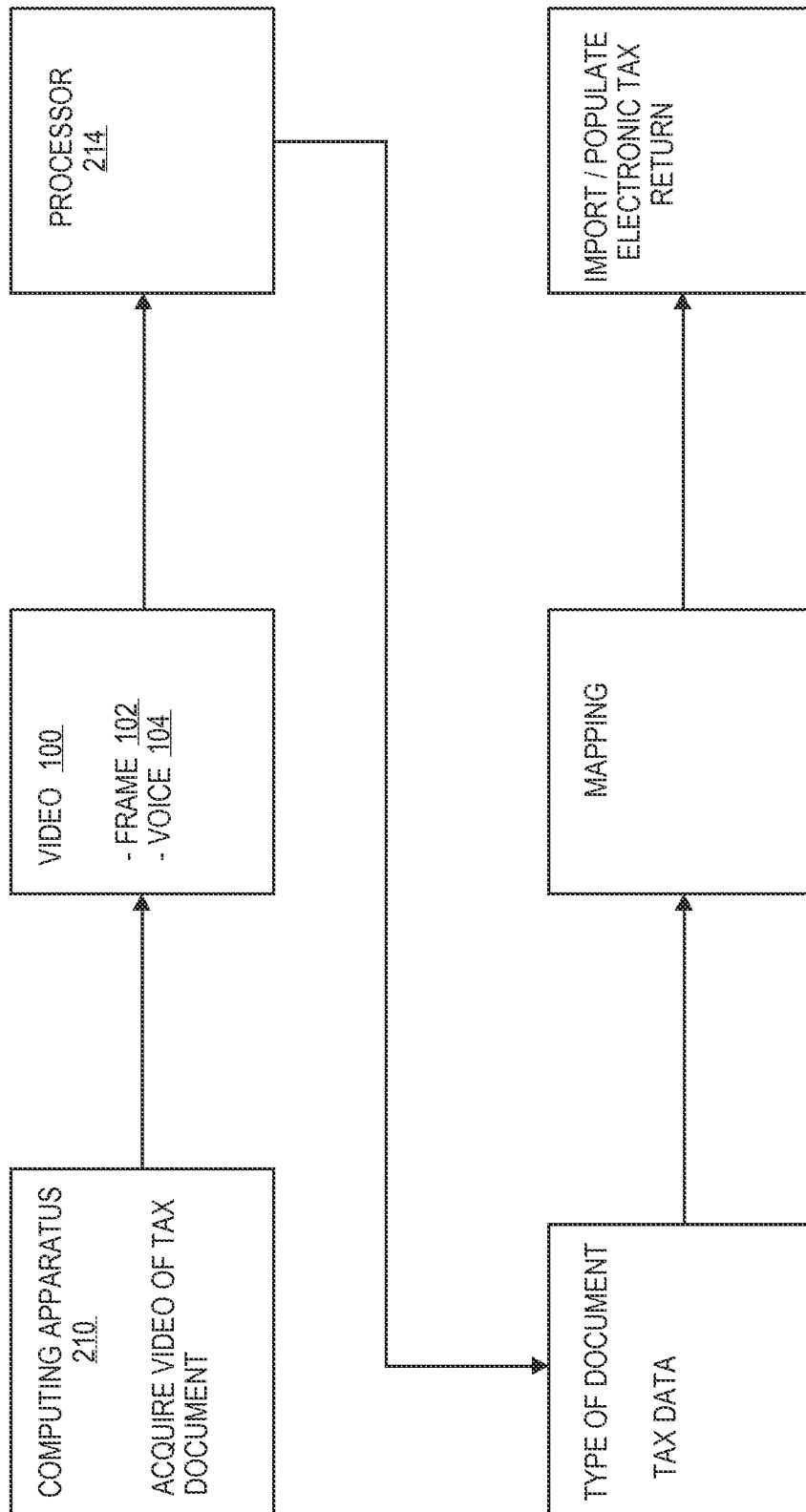


FIG. 4

5/27

106


a Employee's social security number 444-44-4444		OMB No. 1545-0008 		Visit the IRS website at www.irs.gov/efile	
b Employer identification number (EIN) 12-1234567		1 Wages, tips, other compensation 21689.20		2 Federal income tax withheld 2488.46	
c Employer's name, address, and ZIP code SOME COMPANY NAME 1111 SOME ADDRESS DRIVE SOMECITY, OK 00000		3 Social security wages 22360.00		4 Social security tax withheld 1386.32	
d Control number		5 Medicare wages and tips 22360.00		6 Medicare tax withheld 324.22	
e Employee's first name and initial CONNIE		7 Social security tips		8 Allocated tips	
Last name BARNES		9		10 Dependent care benefits	
Suff. P.O. BOX 11211 OKLAHOMA CITY, OK 73127		11 Nonqualified plans		12a See instructions for box 12	
f Employee's address and ZIP code		13a <input type="checkbox"/> Statutory employee 13b <input type="checkbox"/> Retirement plan 13c <input type="checkbox"/> First-time homebuyer's exclusion 13d <input type="checkbox"/> Sick pay 14 Other		12b 12c 12d	
15 State OK		16 State wages, tips, etc. 21689.20		17 State income tax 936.00	
18 State ID number 12-1234567		19 Local wages, tips, etc.		20 Local income tax	
W-2 Wage and Tax Statement		2009		Department of the Treasury—Internal Revenue Service	

FIG. 5

6/27

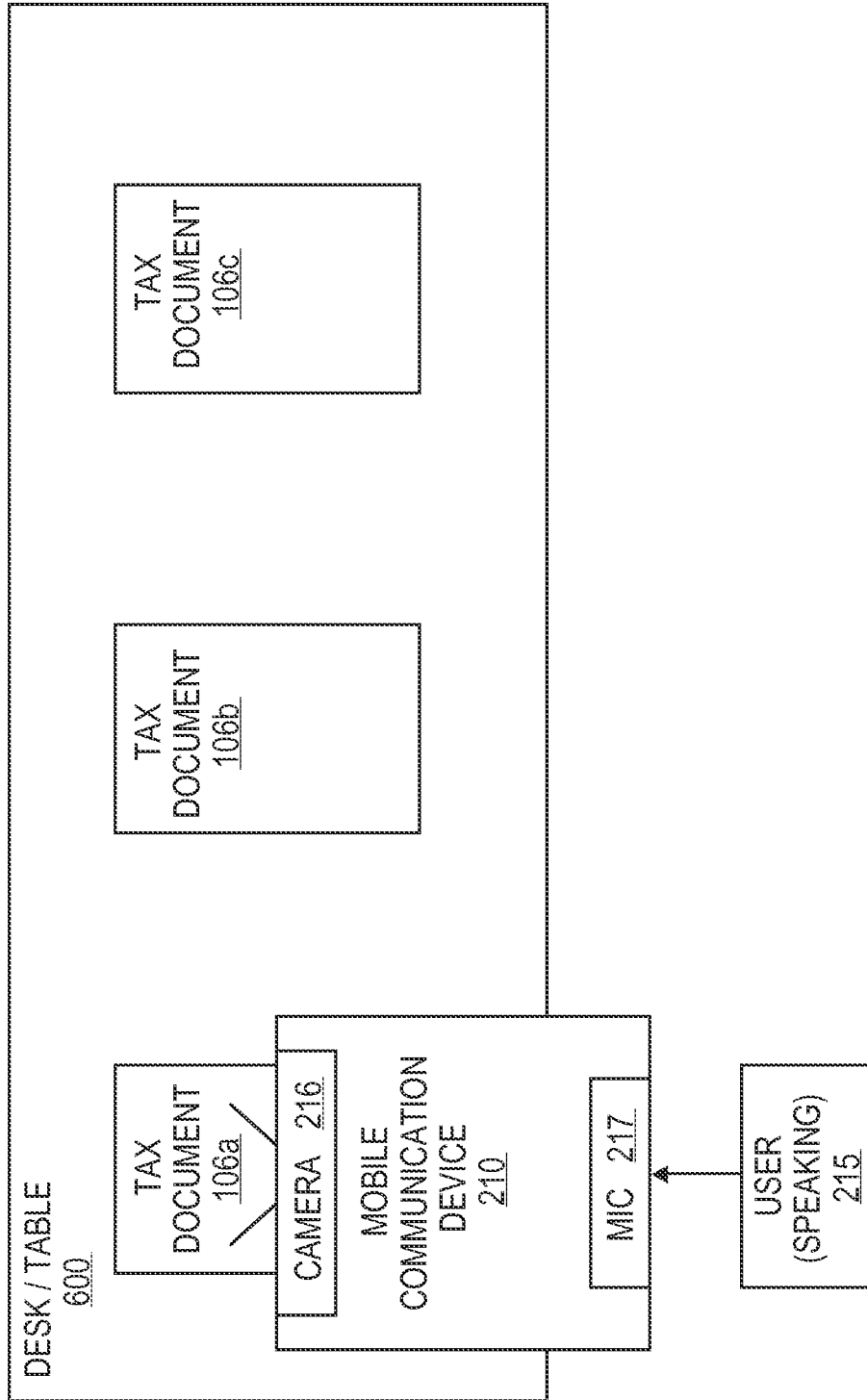


FIG. 6A

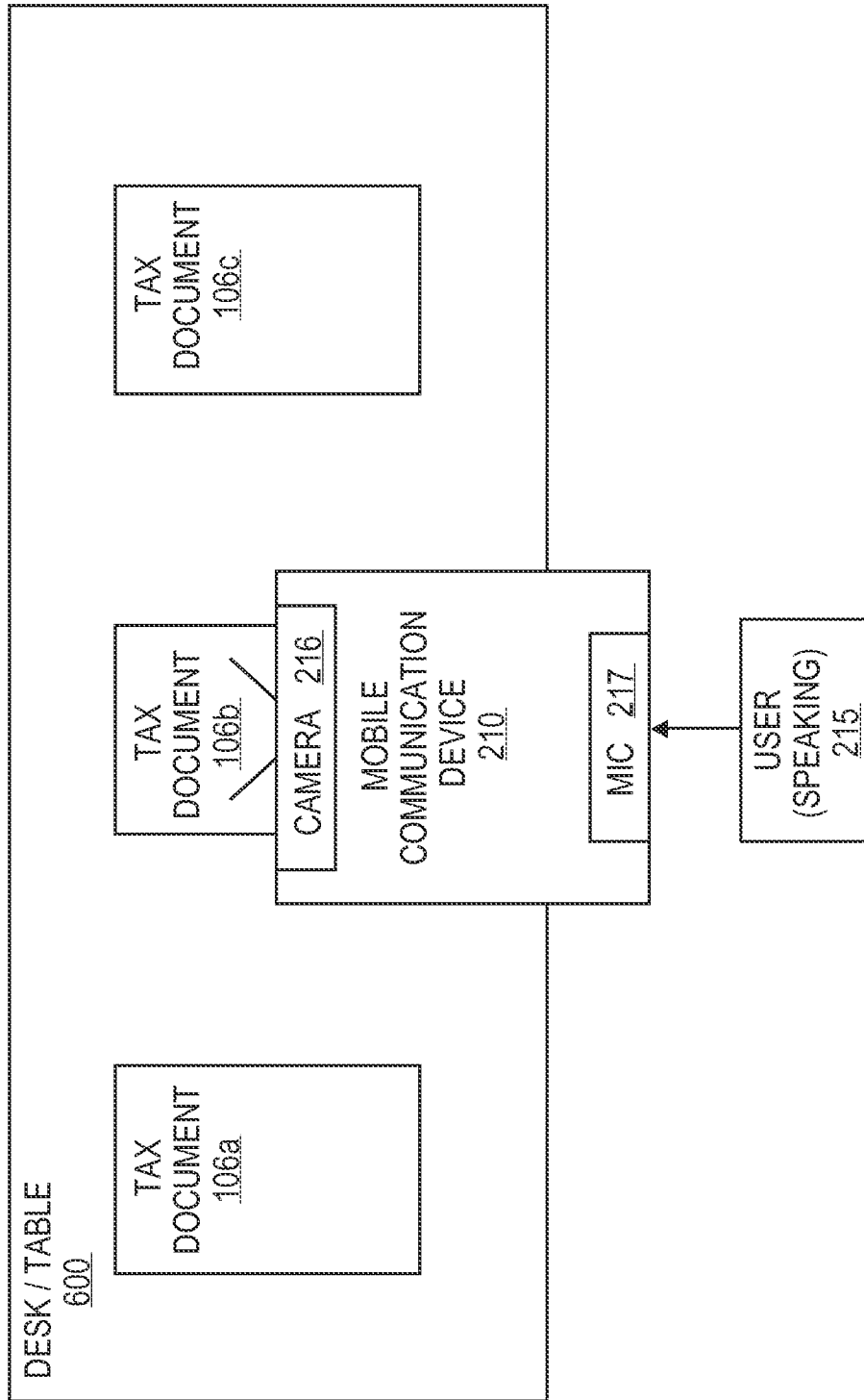


FIG. 6B

8/27

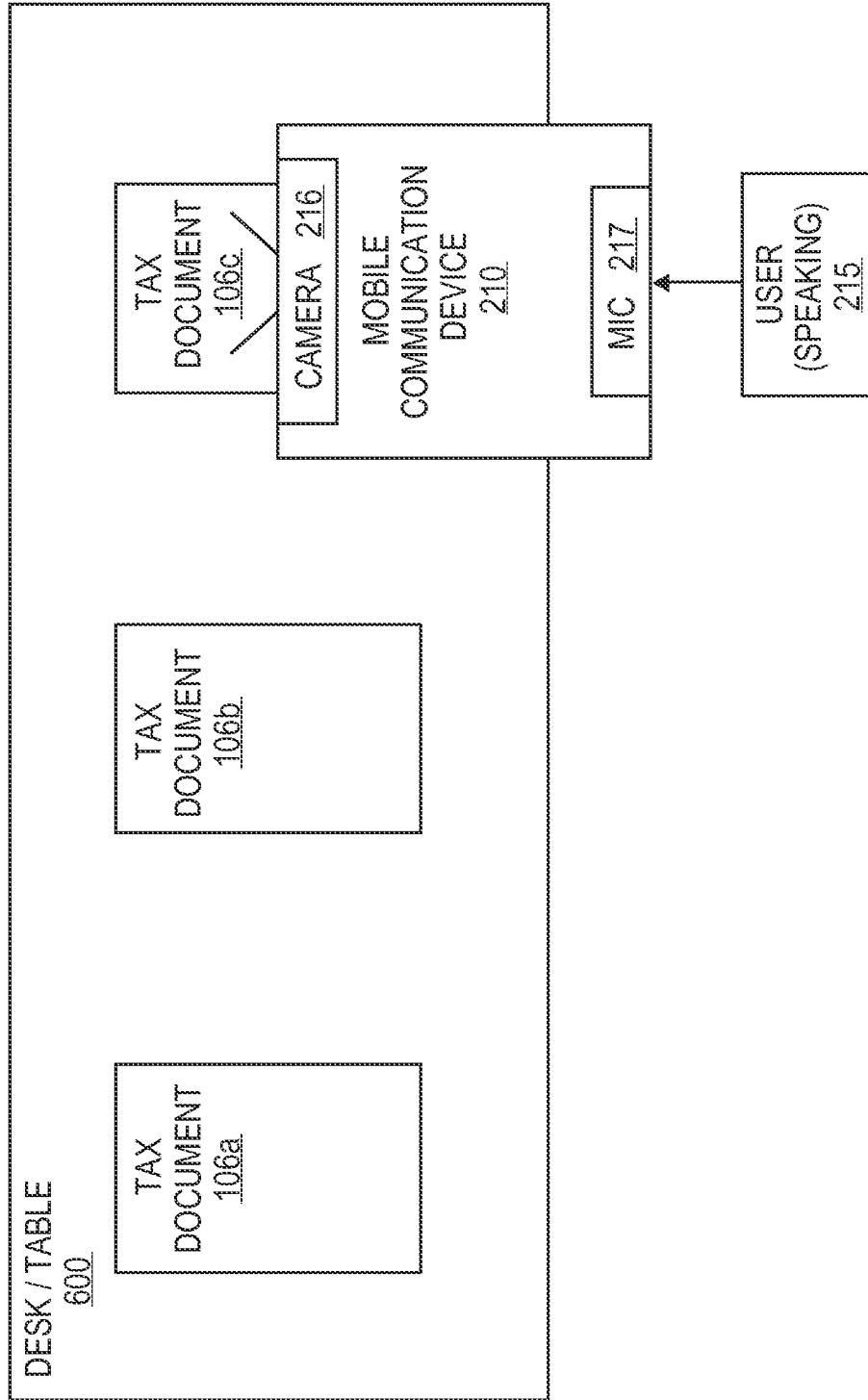


FIG. 6C

9/27

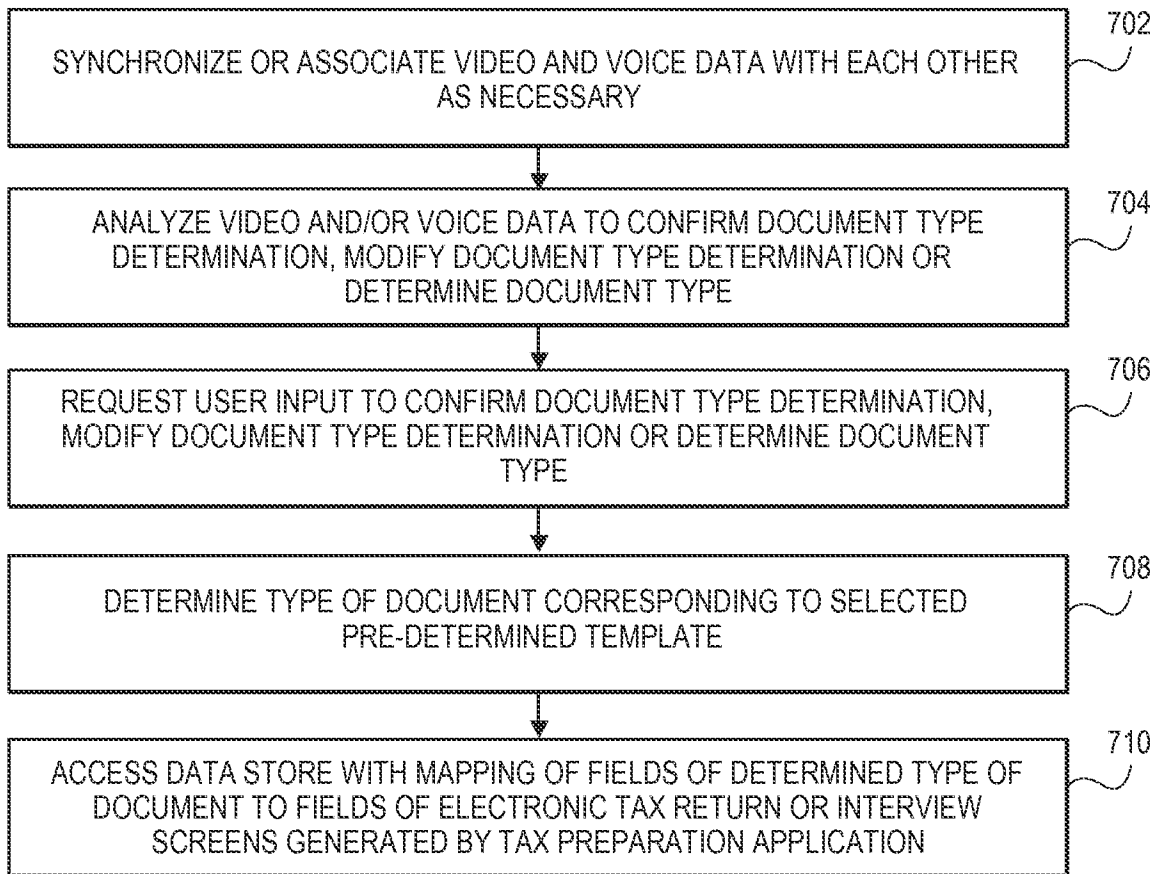
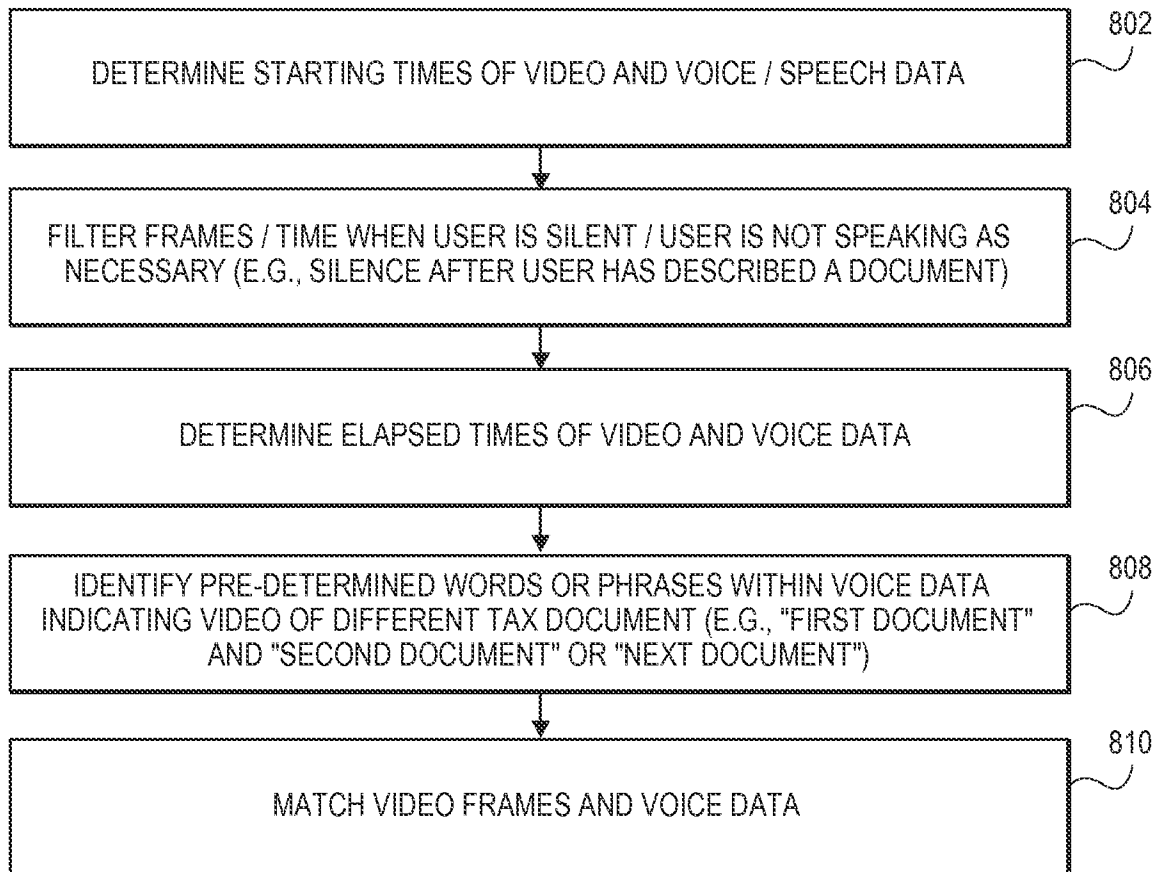


FIG. 7

10/27

**FIG. 8**

11/27

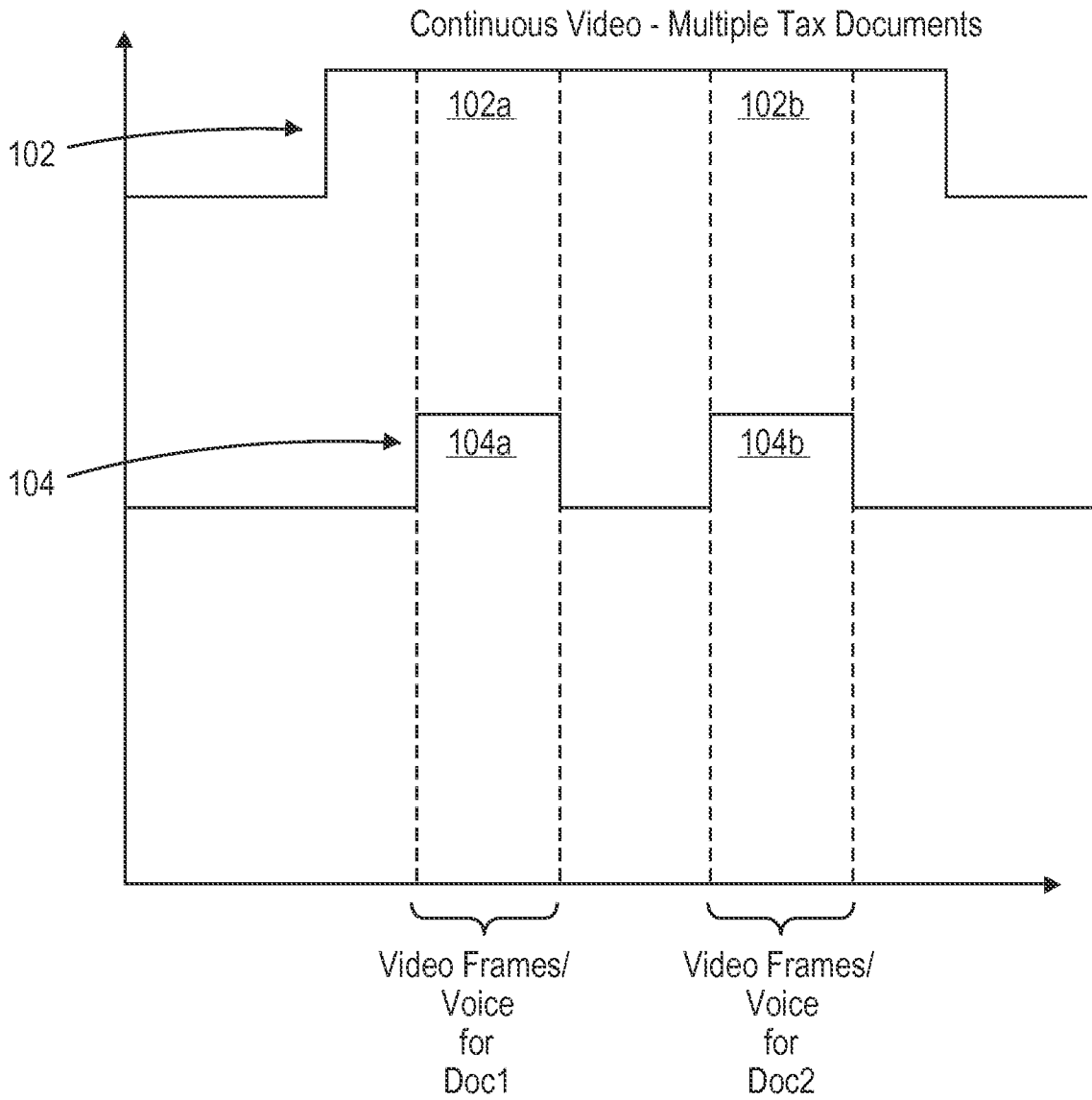


FIG. 9A

12/27

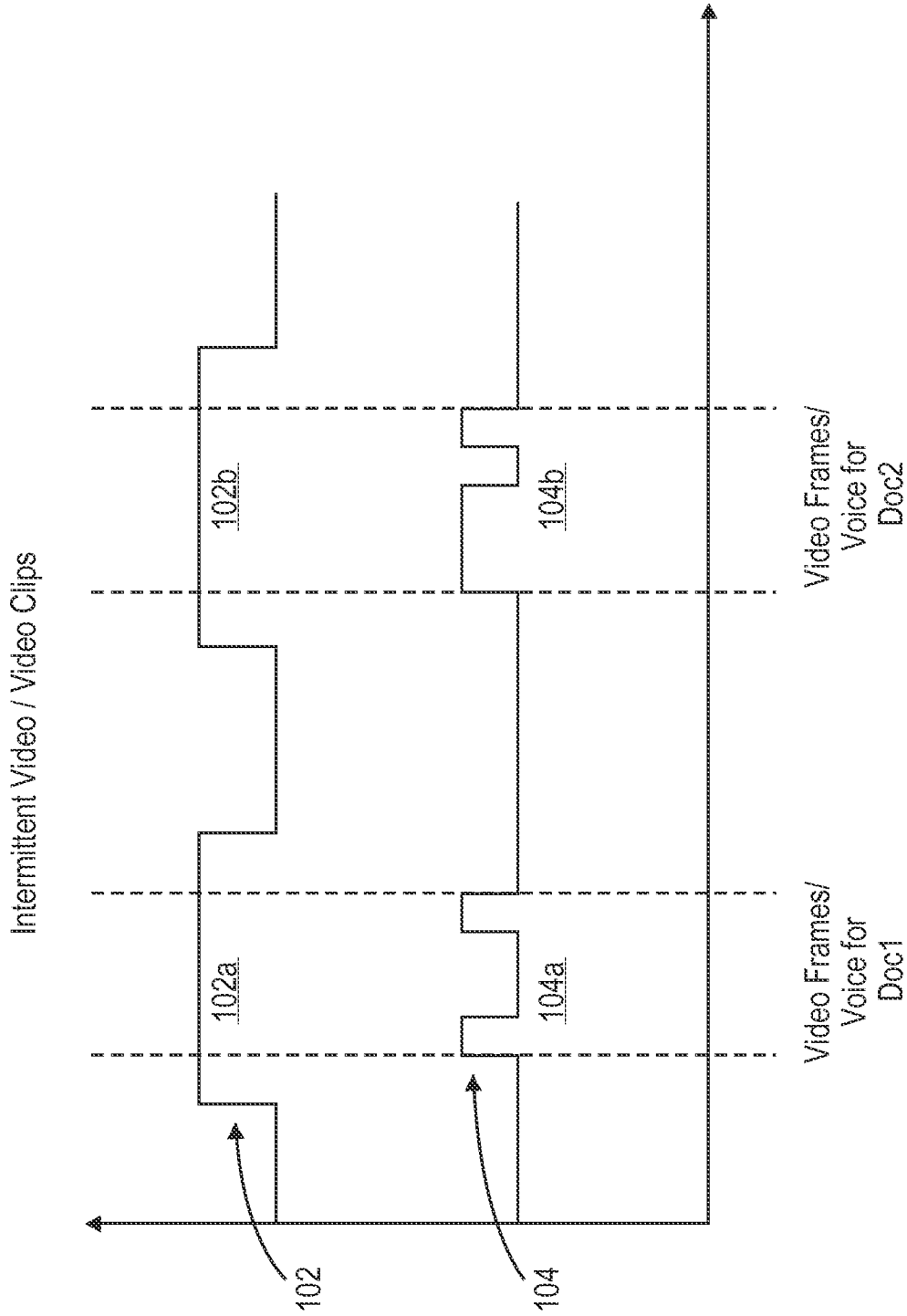


FIG. 9B

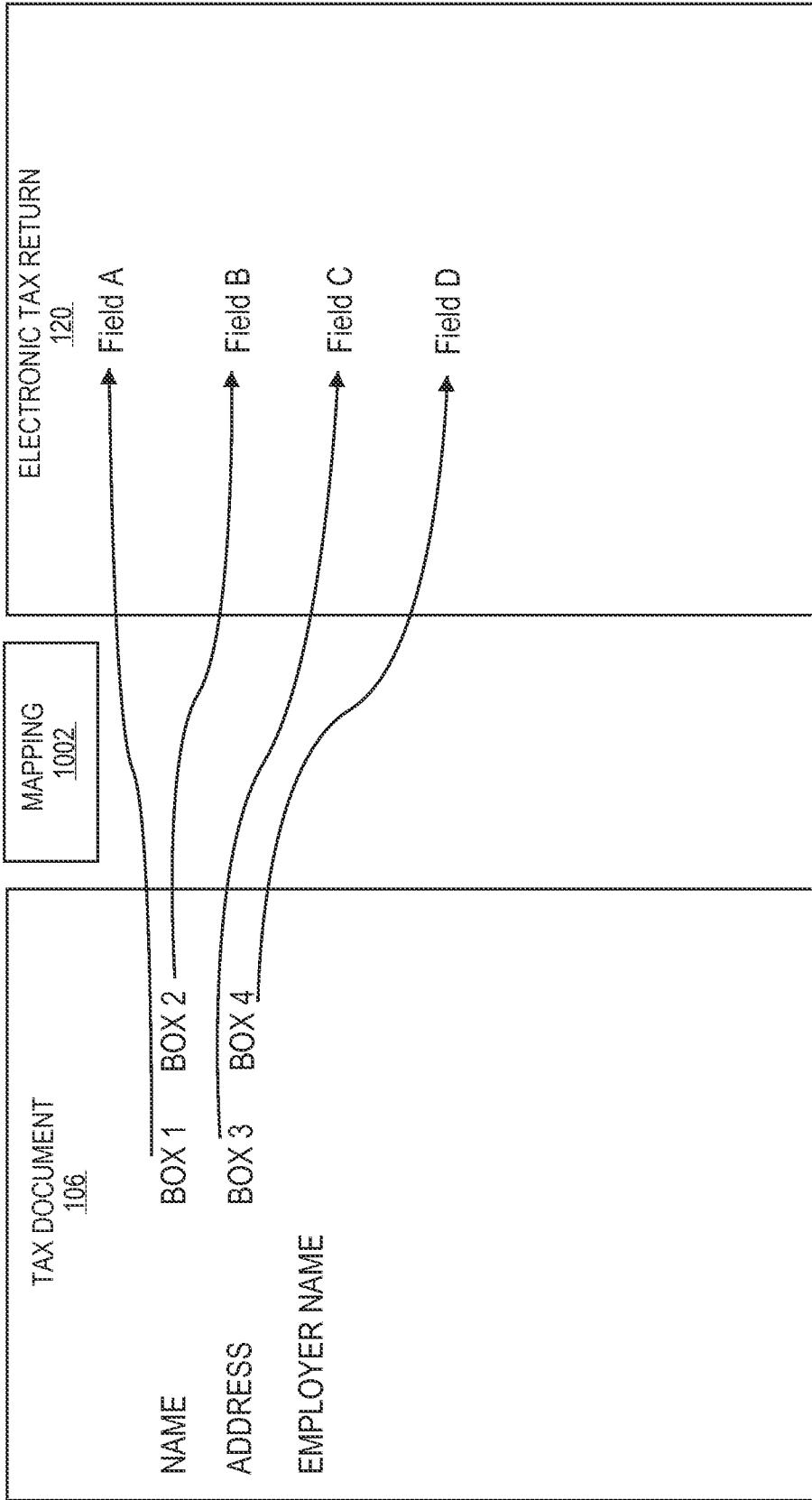
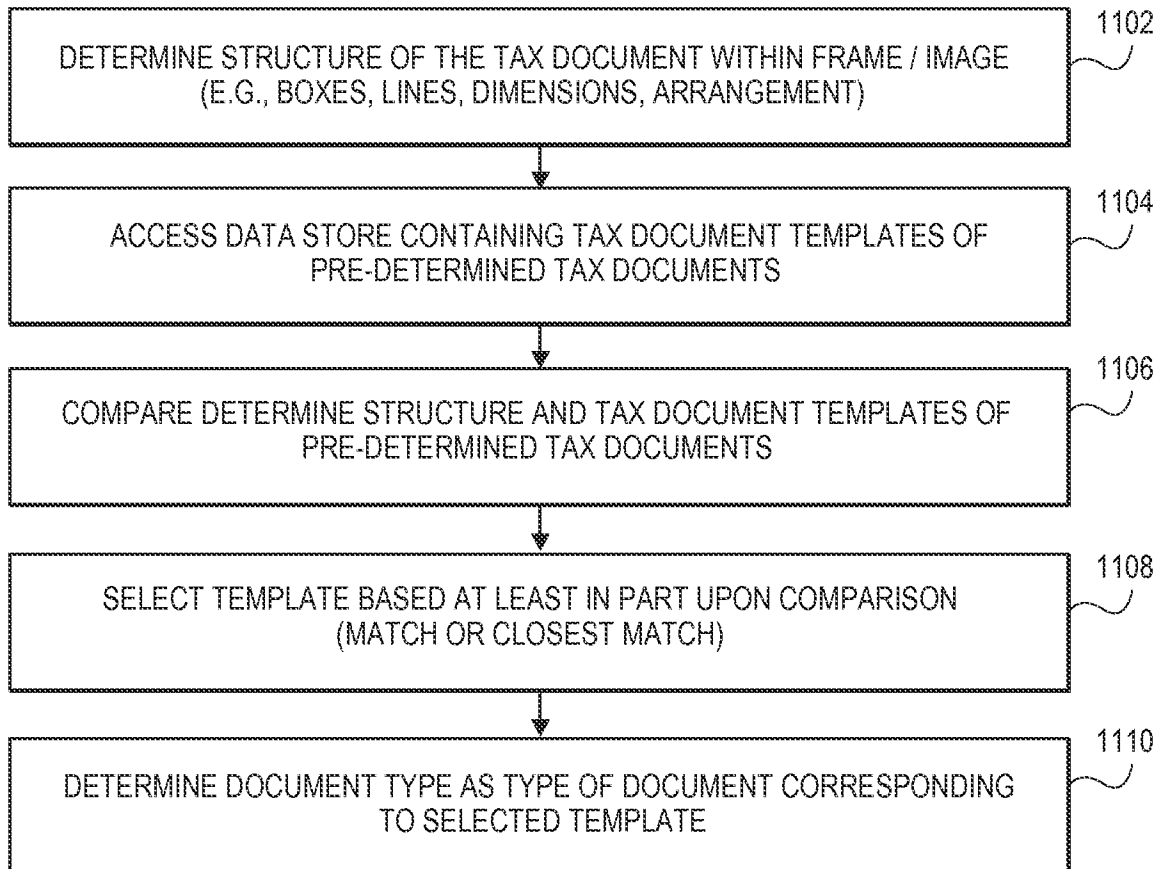


FIG. 10

14/27

**FIG. 11**

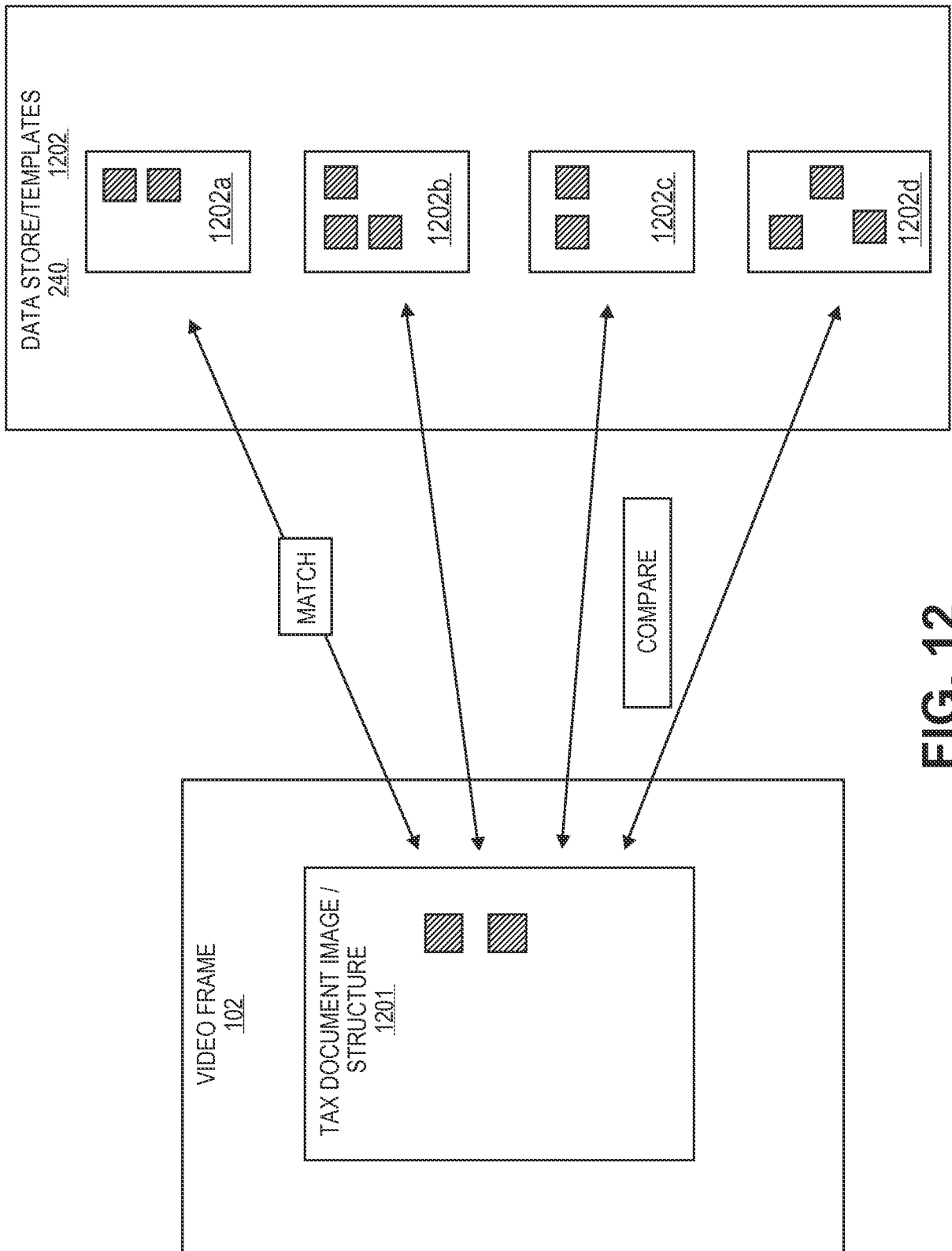
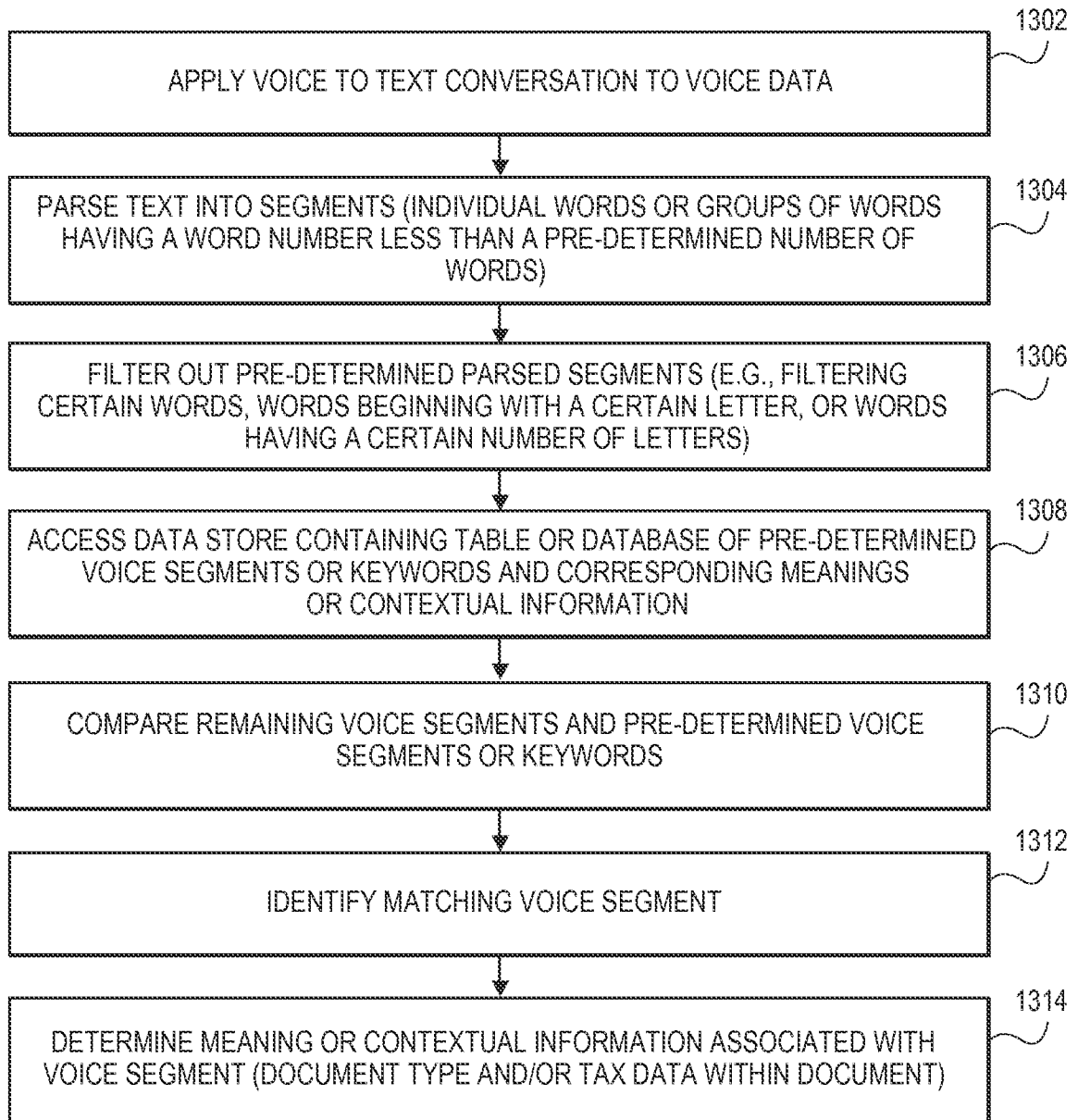


FIG. 12

16/27

**FIG. 13**

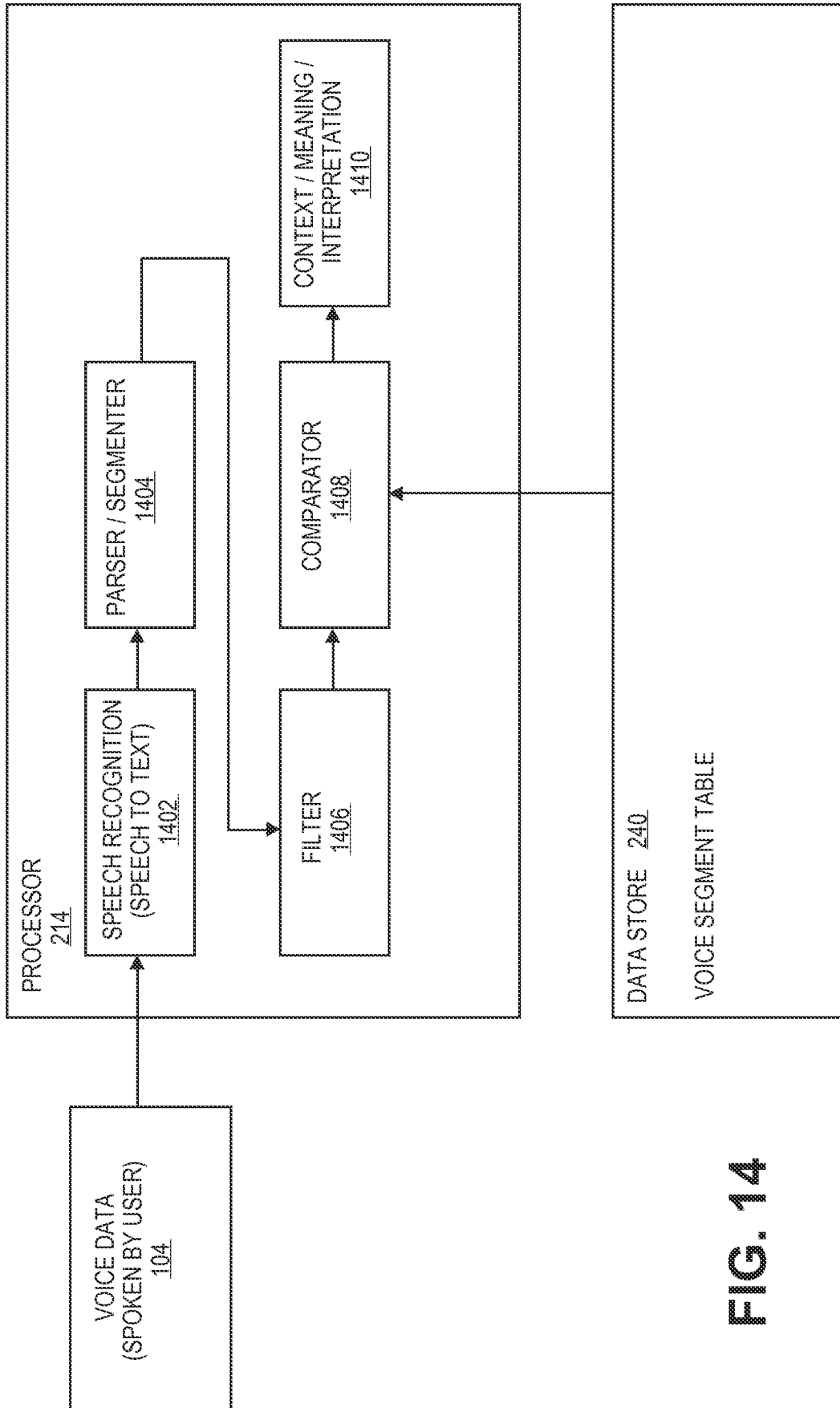


FIG. 14

Voice Segment Database / Table
1500



Segment <u>1501</u>	Meaning / Context / Interpretation <u>1502</u>	Tax Document <u>1503</u>
Segment 1	Meaning 1	Tax Document 1
Segment 2	Meaning 2	
Segment 3	Meaning 3	
Segment 4	Meaning 4	Tax Document 4
Segment n	Meaning 5	

FIG. 15

19/27

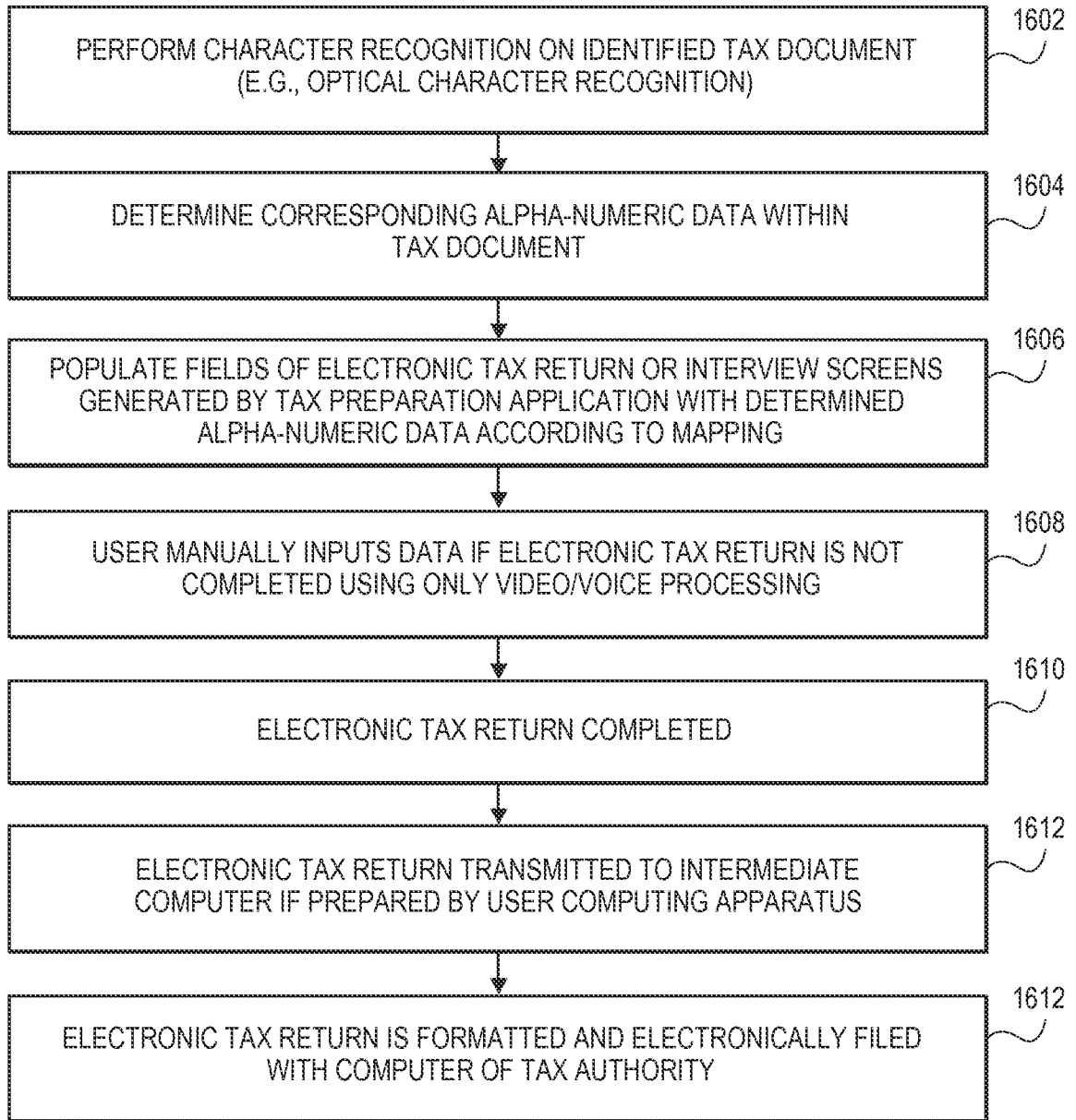


FIG. 16

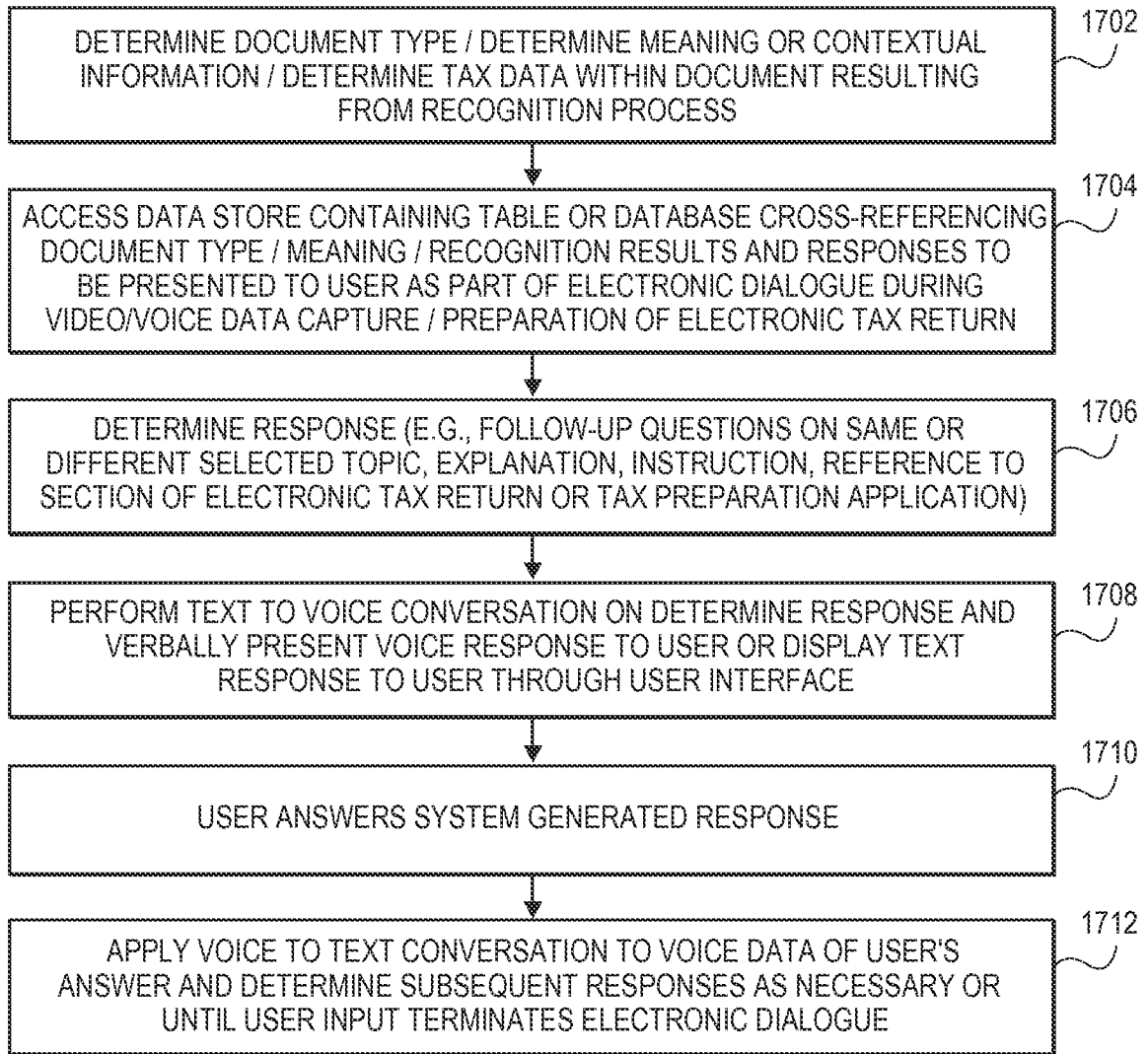


FIG. 17

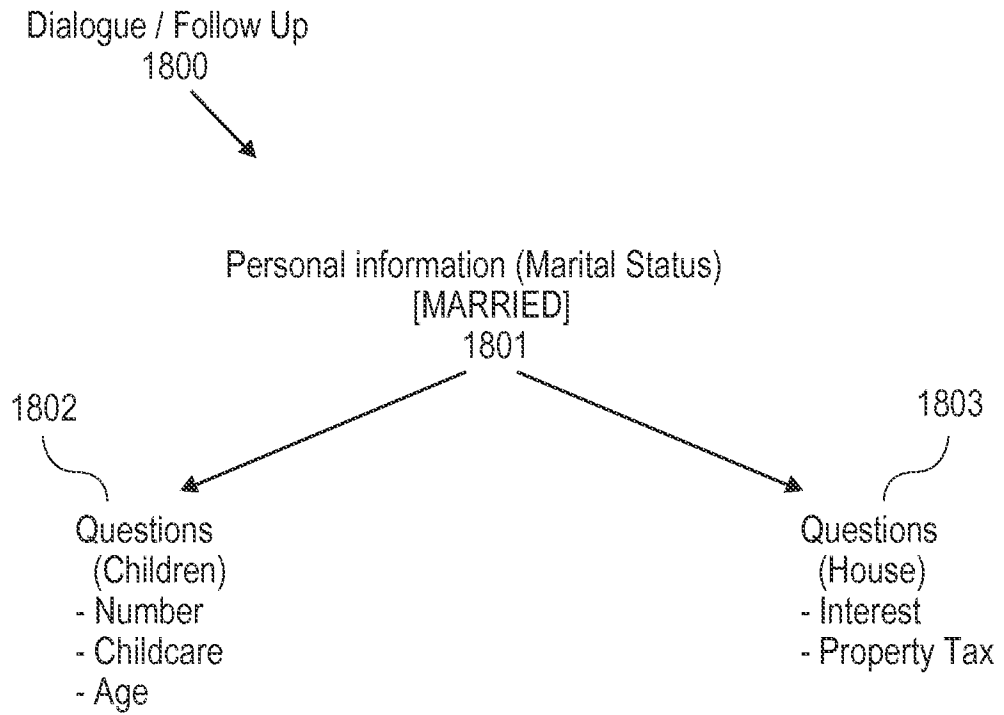


FIG. 18

22/27

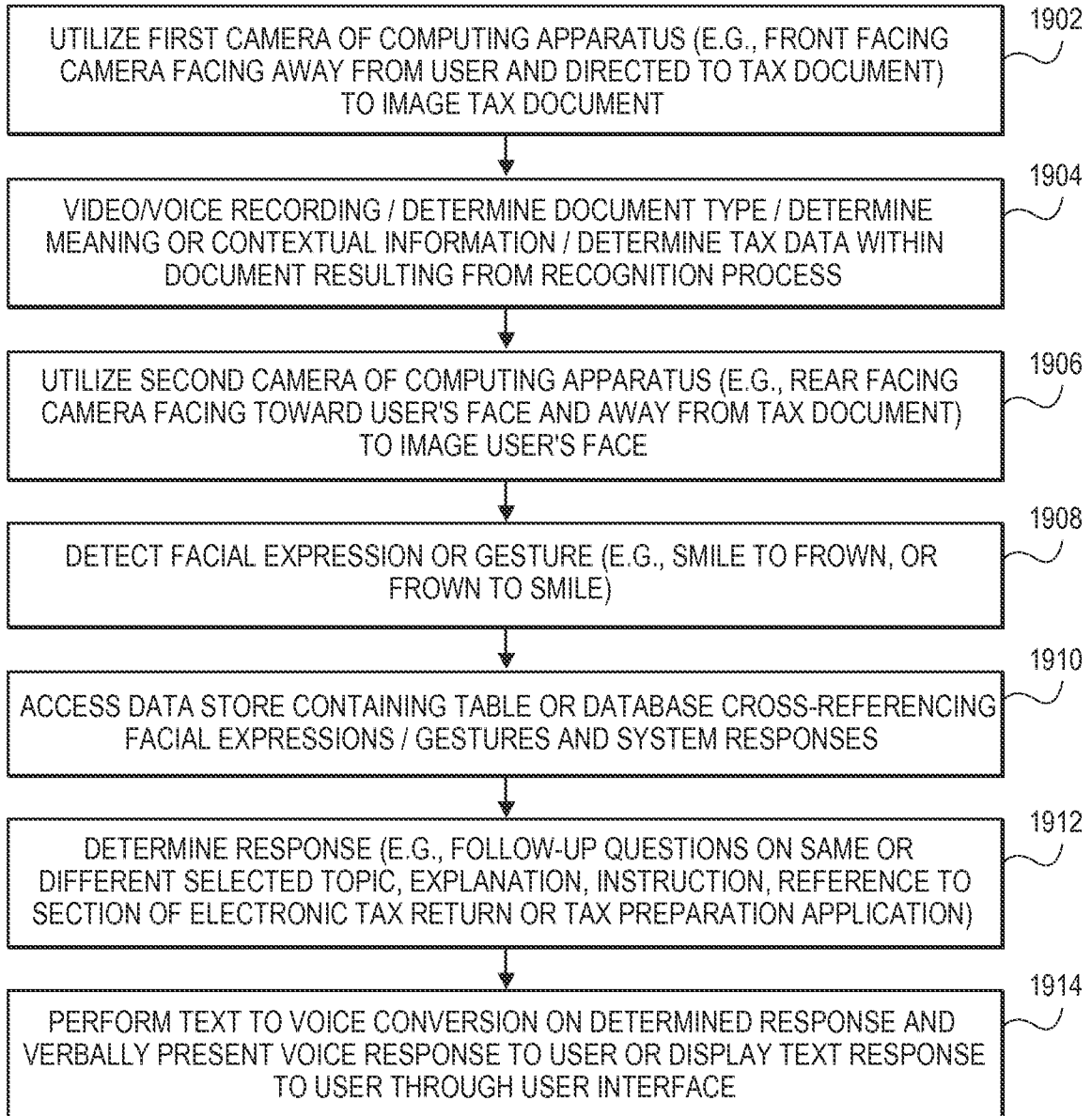


FIG. 19

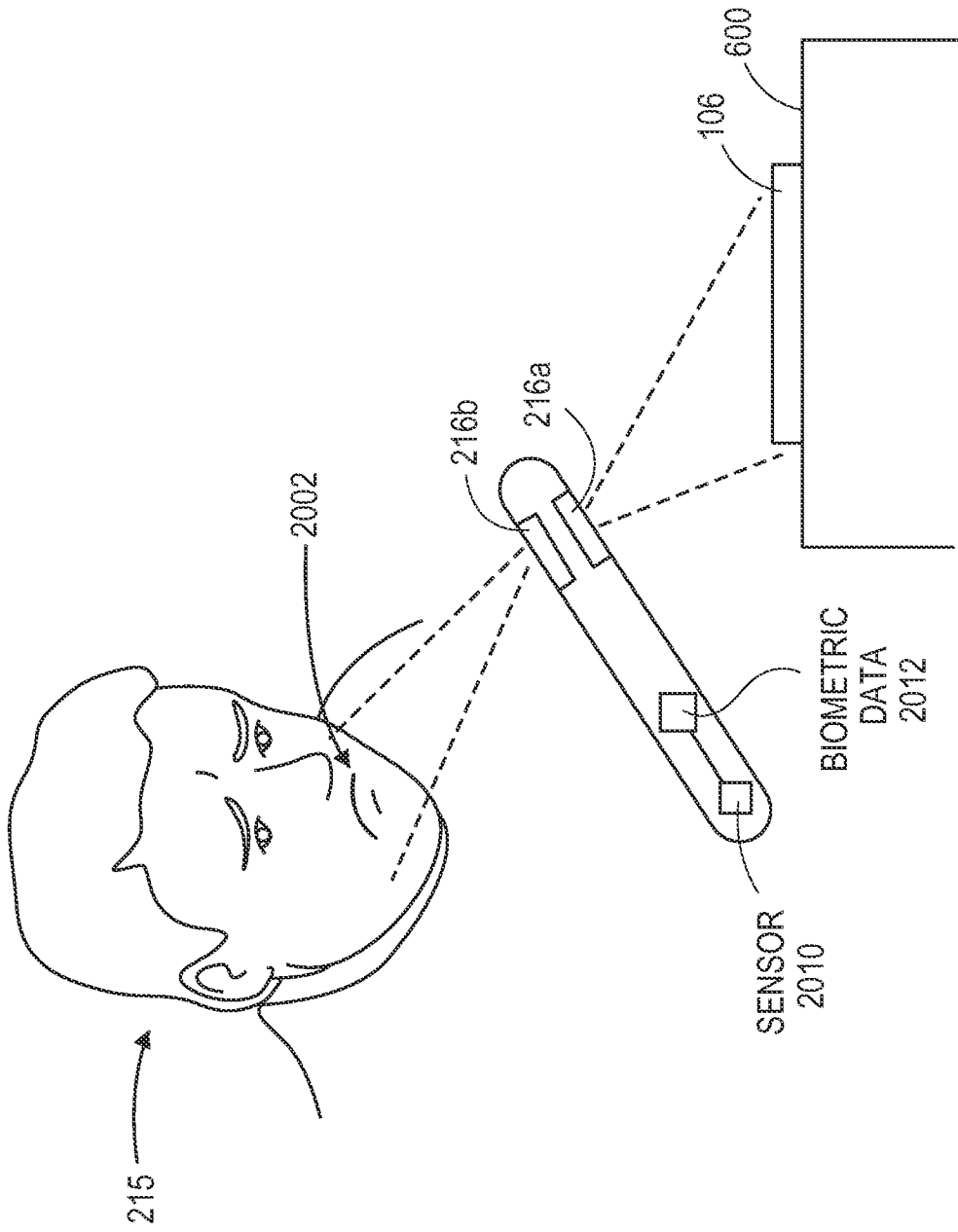


FIG. 20

Facial Expression and Gesture / Dialogue Response Table
2100



Facial Expression / Detected Facial Pattern 2101	Gesture 2102	System Response 2103
Frown Biting Lips Smile	Shake head side to side Cross Arms Hands on Head Pulling Hair Nod Head	Affirmation Response Positive Refund Notification Positive Progress - Almost Done Inquiry of whether user has question None Needed

FIG. 21

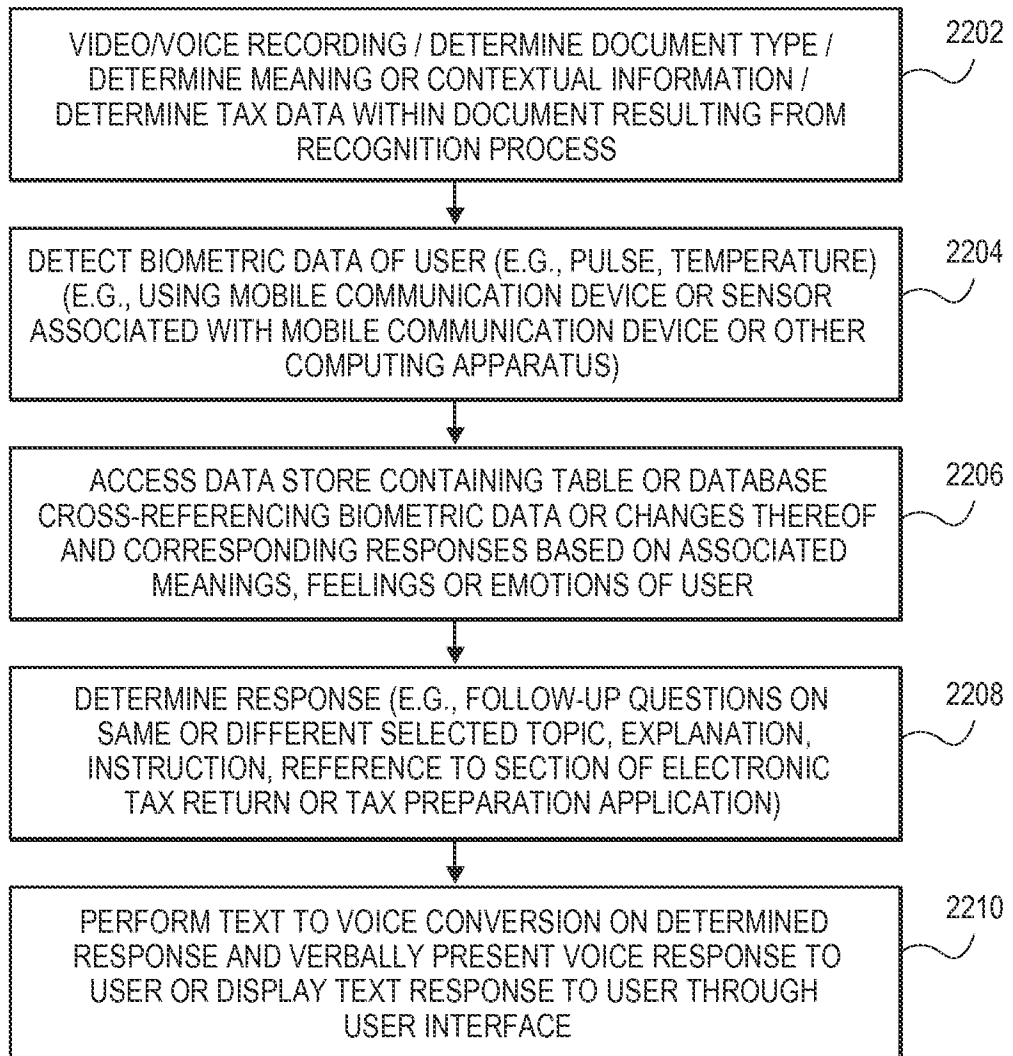


FIG. 22

Biometric Data / Dialogue Response Table
2300



Biometric Data / Range (e.g., Pulse Rate) 2301	System Response 2302
Pulse Rate Range 1 (Low / Normal Pulse Rate)	None Needed
Pulse Rate Range 2 (Moderate Pulse Rate)	Inquiry whether user has question
Pulse Rate Range 3 (High Pulse Rate)	Affirmation Response Positive Refund Notification Positive Progress - Almost Done Inquiry whether user has question

FIG. 23

27/27

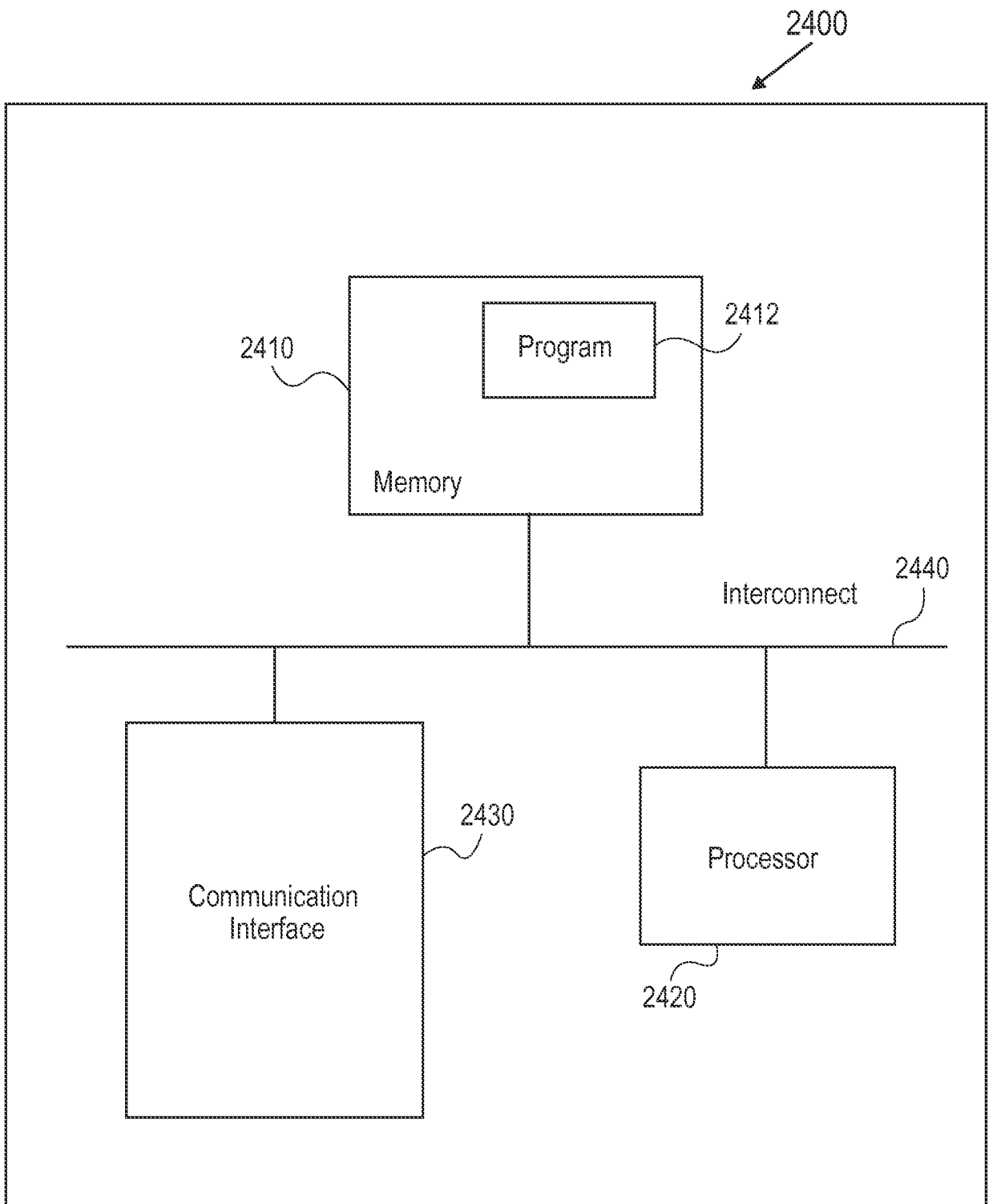


FIG. 24

A. CLASSIFICATION OF SUBJECT MATTER**G06F 17/00(2006.01)i, G10L 15/25(2013.01)i, G11B 20/10(2006.01)i**

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)

G06F 17/00; G06K 15/00; G10L 15/00; G06F 17/30; G06K 19/06; G06Q 40/00; G06K 9/72; G10L 15/25; G11B 20/10

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

Korean utility models and applications for utility models

Japanese utility models and applications for utility models

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

eKOMPASS(KIPO internal) & Keywords: electronic tax return, voice, speech, image, video, recognize, document, field, type, and similar terms

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2012-0211561 A1 (JONATHAN A. LIEBERMAN et al.) 23 August 2012 See paragraphs [0050] and [0068]; and figure 12C.	1-38
A	US 2009-0300068 A1 (DING-YUAN TANG) 03 December 2009 See paragraphs [0018] and [0022]; figure 1; and claim 1.	1-38
A	US 2010-0070394 A1 (DAVID A. WYLE) 18 March 2010 See paragraphs [0009] and [0032]-[0034]; figure 5; and claim 11.	1-38
A	US 2007-0022372 A1 (ZICHENG LIU et al.) 25 January 2007 See paragraphs [0004]-[0005] and [0030]-[0032]; and figures 1-2.	1-38
A	US 2002-0013701 A1 (THOMAS C. OLIVER et al.) 31 January 2002 See paragraphs [0005]-[0008] and [0030]-[0037]; and figure 1.	1-38

 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

26 February 2014 (26.02.2014)

Date of mailing of the international search report

26 February 2014 (26.02.2014)

Name and mailing address of the ISA/KR

International Application Division
Korean Intellectual Property Office
189 Cheongsu-ro, Seo-gu, Daejeon Metropolitan City, 302-701,
Republic of Korea

Facsimile No. +82-42-472-7140

Authorized officer

NHO, Ji Myong

Telephone No. +82-42-481-8528



INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2013/039936

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2012-0211561 A1	23/08/2012	US 8418920 B2 WO 2012-112168 A1	16/04/2013 23/08/2012
US 2009-0300068 A1	03/12/2009	None	
US 2010-0070394 A1	18/03/2010	US 2006-0155618 A1 US 7853494 B2 WO 2006-073833 A2 WO 2006-073833 A3	13/07/2006 14/12/2010 13/07/2006 20/09/2007
US 2007-0022372 A1	25/01/2007	US 7694214 B2	06/04/2010
US 2002-0013701 A1	31/01/2002	CN 1184614 C CN 1214612 C CN 1258061 A CN 1258162 A EP 1014258 A2 EP 1014258 A3 EP 1014338 A1 JP 2000-194533 A JP 2000-215021 A KR 10-2000-0057082 A US 6226422 B1 US 6289140 B1 US 6295391 B1 US 6580838 B2	12/01/2005 10/08/2005 28/06/2000 28/06/2000 28/06/2000 26/11/2003 28/06/2000 14/07/2000 04/08/2000 15/09/2000 01/05/2001 11/09/2001 25/09/2001 17/06/2003