



(86) **Date de dépôt PCT/PCT Filing Date:** 2013/11/14
(87) **Date publication PCT/PCT Publication Date:** 2014/05/22
(85) **Entrée phase nationale/National Entry:** 2015/05/05
(86) **N° demande PCT/PCT Application No.:** US 2013/070002
(87) **N° publication PCT/PCT Publication No.:** 2014/078480
(30) **Priorité/Priority:** 2012/11/16 (US61/727,217)

(51) **Cl.Int./Int.Cl. G06F 3/01** (2006.01),
G06F 3/16 (2006.01)

(71) **Demandeur/Applicant:**
AETHER THINGS, INC., US

(72) **Inventeurs/Inventors:**
LAMB, DUNCAN, GB;
JACOBSEN, KENNETH, US;
EVANS, JOHN, US;
MOLTONI, THOMAS, US;
MANCINO, FELICE, US;
ROSENBERG, ARON, US;
LONG, JOHN, US

(74) **Agent:** HICKS INTELLECTUAL PROPERTY LAW

(54) **Titre : STRUCTURE UNIFIEE POUR LA CONFIGURATION, L'INTERACTION ET LA COMMANDE D'UN DISPOSITIF, ET
PROCEDES, DISPOSITIFS ET SYSTEMES ASSOCIES**

(54) **Title: UNIFIED FRAMEWORK FOR DEVICE CONFIGURATION, INTERACTION AND CONTROL, AND ASSOCIATED METHODS,
DEVICES AND SYSTEMS**

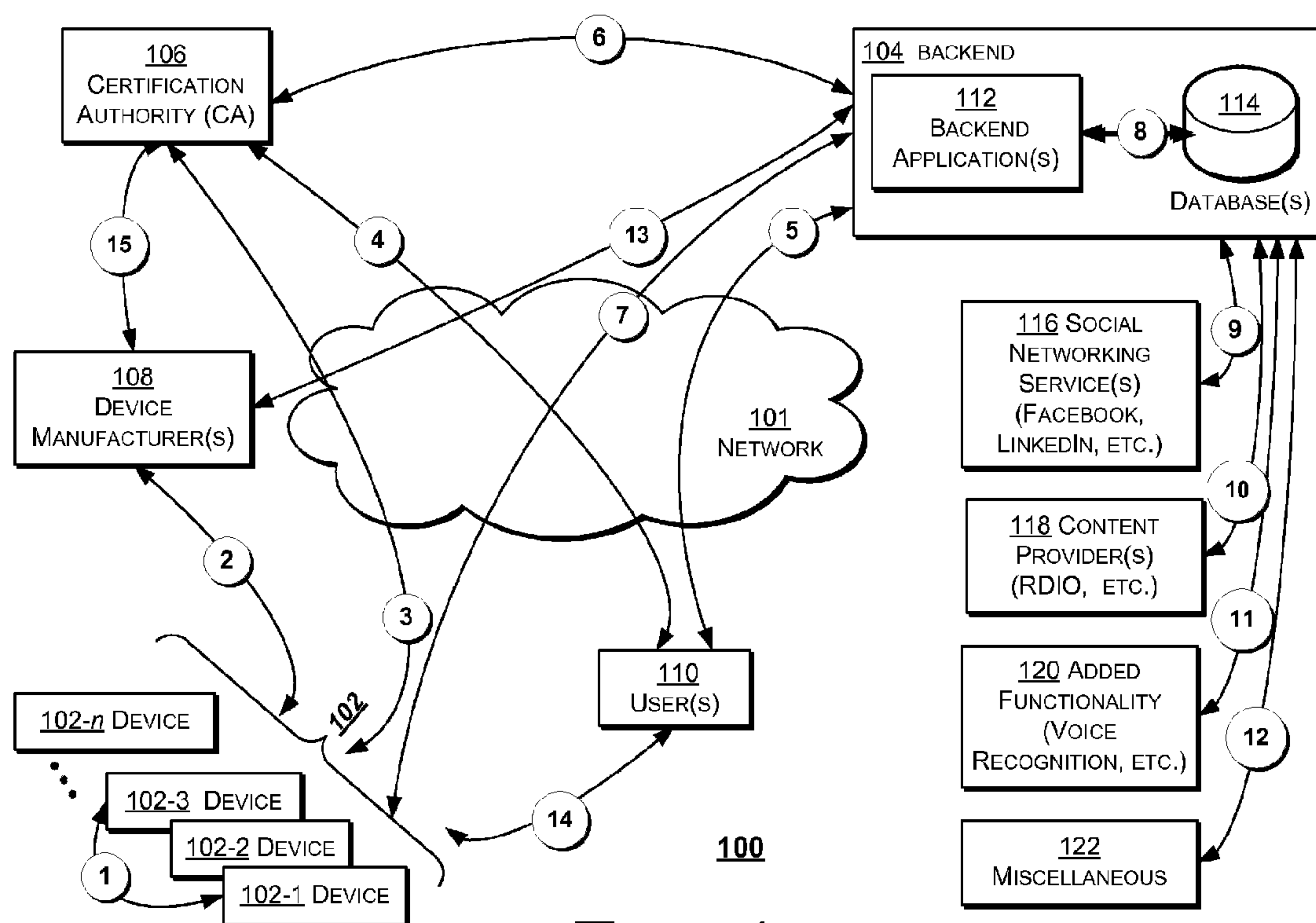
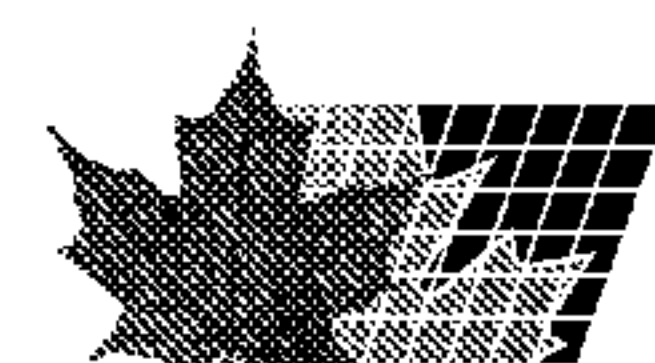


FIG. 1

(57) **Abrégé/Abstract:**

A method of operating a device includes buffering sensor information from an environment around the device as buffered sensor information, detecting a gaze of a person in the environment, based on the gaze detected, initiating recognition of some sensor information including the buffered sensor information; the recognition determining at least one instruction in the sensor information including the buffered sensor information; and operating the device based on the at least one instruction.



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

(19) World Intellectual Property
Organization
International Bureau(10) International Publication Number
WO 2014/078480 A1(43) International Publication Date
22 May 2014 (22.05.2014)(51) International Patent Classification:
G06F 3/01 (2006.01) *G06F 3/16* (2006.01)(21) International Application Number:
PCT/US2013/070002(22) International Filing Date:
14 November 2013 (14.11.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/727,217 16 November 2012 (16.11.2012) US

(71) Applicant: AETHER THINGS, INC. [US/US]; 149 New Montgomery, 5th Floor, San Francisco, CA 94105 (US).

(72) Inventors: LAMB, Duncan; Saariniemenkatu 1B40, FI-00530 Helsinki (FI). JACOBSEN, Kenneth; 3209 Longfellow Drive, Belmont, California 94002 (US). EVANS, John; 73 Cumberland Street, San Francisco, California 94110 (US). MOLTONI, Thomas; 363 Diamond Street, Apartment A, San Francisco, California 94114 (US). MANCINO, Felice; 23 Hancock Street, San Francisco, California 94114 (US). ROSENBERG, Aron; 3209 Andreasen Drive, Lafayette, California 94549 (US). LONG, John; 16211 Highland Drive, San Jose, California 95127 (US).

(74) Agent: SIRITZKY, Brian; Siritzky Law, PLLC, 8300 Greensboro Drive, Suite 800, McLean, Virginia 22102 (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, IR, IS, JP, KE, KG, 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, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

Declarations under Rule 4.17:

— of inventorship (Rule 4.17(iv))

[Continued on next page]

(54) Title: UNIFIED FRAMEWORK FOR DEVICE CONFIGURATION, INTERACTION AND CONTROL, AND ASSOCIATED METHODS, DEVICES AND SYSTEMS

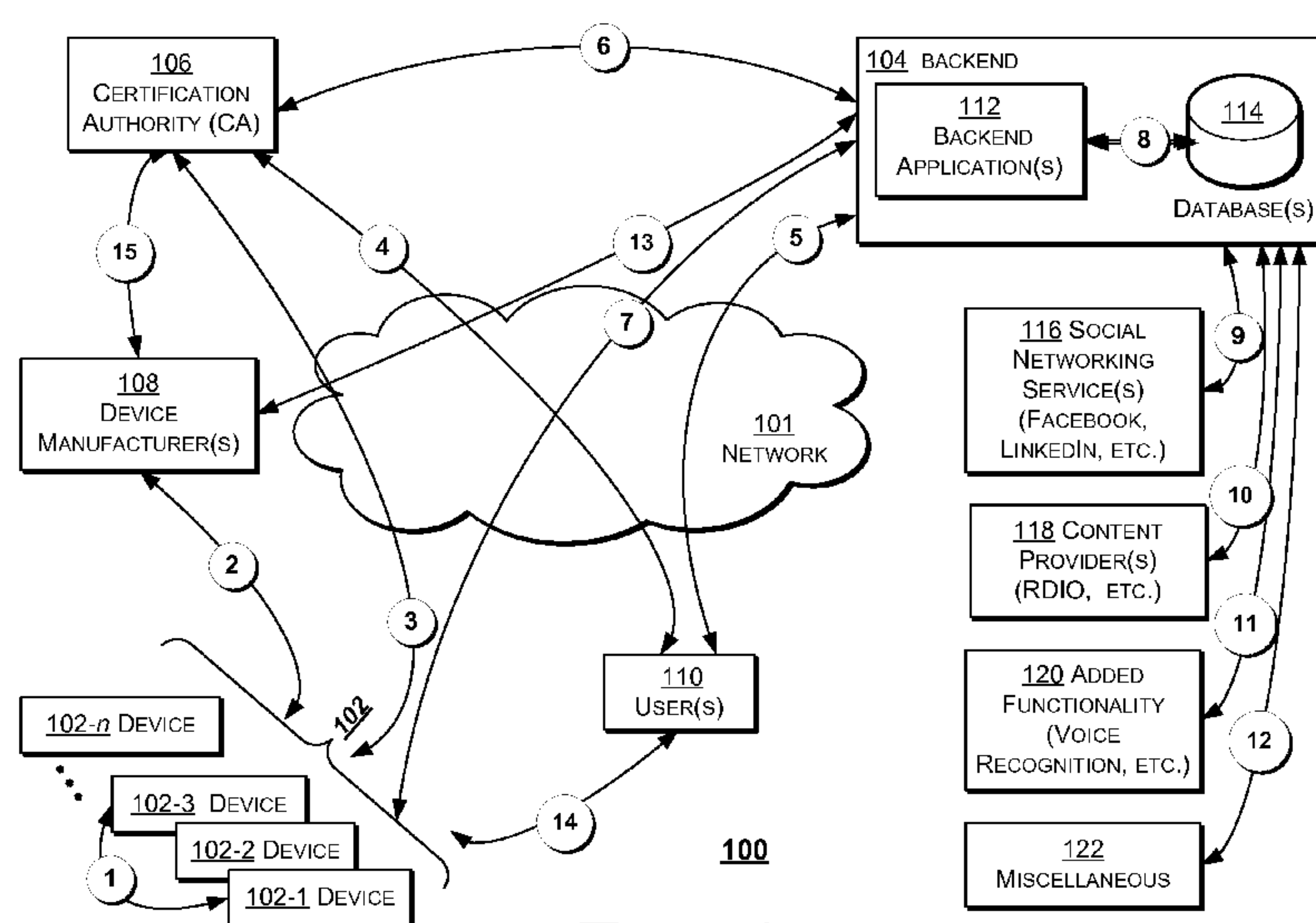


FIG. 1

(57) Abstract: A method of operating a device includes buffering sensor information from an environment around the device as buffered sensor information, detecting a gaze of a person in the environment, based on the gaze detected, initiating recognition of some sensor information including the buffered sensor information; the recognition determining at least one instruction in the sensor information including the buffered sensor information; and operating the device based on the at least one instruction.

WO 2014/078480 A1

Published:

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

UNIFIED FRAMEWORK FOR DEVICE CONFIGURATION, INTERACTION AND CONTROL, AND ASSOCIATED METHODS, DEVICES AND SYSTEMS

BACKGROUND OF THE INVENTION

COPYRIGHT STATEMENT

[0001] This patent document contains material subject to copyright protection. The copyright owner has no objection to the reproduction of this patent document or any related materials in the files of the United States Patent and Trademark Office, but otherwise reserves all copyrights whatsoever.

RELATED APPLICATIONS

[0002] This application claims priority to U.S. Provisional Patent Application No. 61/727,217, filed November 16, 2012, titled “Unified Framework For Device Configuration, Interaction And Control, And Associated Methods, Devices And Systems,” the entire contents of both of which are fully incorporated herein by reference for all purposes.

FIELD OF THE INVENTION

[0003] This invention relates to a unified framework for device configuration, interaction and control, and to related methods, devices, and systems.

BACKGROUND AND OVERVIEW

[0004] Computer-based devices, in particular consumer computer-based devices are ubiquitous. These devices are typically designed and manufactured independent of each other, and, to the extent that they can interact or make use of each other, they tend to try to use ad hoc or standardized techniques to do so. As a consequence, consumers are often forced to perform complicated setup procedures

for devices they acquire or when they try to use different devices together, even when the same vendor makes the devices. Even for single, stand-alone devices, setup procedures are typically complex. Some companies have tried to simplify the use of their own devices, but not the devices of others.

[0005] It is desirable and an object of this invention to provide a system / framework within which devices may be easily provisioned, configured, and assimilated.

[0006] It is further desirable an object of this invention to provide a system / framework within which devices of different types and different users can interact.

[0007] It is further desirable an object of this invention to provide a core system within devices to support their interaction with each other and with a common framework.

BRIEF DESCRIPTION OF THE DRAWINGS

[0008] Other objects, features, and characteristics of the present invention as well as the methods of operation and functions of the related elements of structure, and the combination of parts and economies of manufacture, will become more apparent upon consideration of the following description and the appended claims with reference to the accompanying drawings, all of which form a part of this specification, and wherein:

[0009] **FIG. 1** depicts a framework for an exemplary system according to embodiments hereof;

[0010] **FIG. 2** depicts aspects of configuration of a user's device within the framework of **FIG. 1** according to embodiments hereof;

[0011] **FIGS. 3A – 3C** depict details of the database(s) of the framework of **FIG. 1** according to embodiments hereof;

[0012] **FIGS. 4A, 4A-1, 4A-2, 4B, and 4C** depict aspects of a typical device for use within the framework of **FIG. 1** according to embodiments hereof;

[0013] **FIG. 4D – 4H** depict exemplary organization of corpora within the system according to embodiments hereof;

[0014] **FIGS. 4I – 4N** are flowcharts depicting exemplary operation of a device within the framework of **FIG. 1** according to embodiments hereof;

[0015] **FIGS. 5A – 5E** depict aspects of typical computer systems upon which embodiments of the present disclosure may be implemented and carried out;

[0016] **FIGS. 6A – 6I** show exemplary aspects of device provisioning and configuration within the framework of **FIG. 1** according to embodiments hereof;

[0017] **FIGS. 7A – 7J** show aspects of the interaction of devices within the framework of **FIG. 1** according to embodiments hereof;

[0018] **FIGS. 8A – 8D** show aspects of an exemplary specific device for sound rendering according to embodiments hereof; and

[0019] **FIGS. 9A – 9C** shown aspects of cooperation between sound-rendering devices according to embodiments hereof.

DETAILED DESCRIPTION OF THE PRESENTLY PREFERRED EXEMPLARY EMBODIMENTS

[0020] The term “***M***” is used herein as a name for the system, devices, processes, interactions, etc. being described. It should be appreciated that this name is only being used to aid in the description, and is not intended to limit the scope of the system in any way.

[0021] **FIG. 1** depicts a framework / system for an exemplary ***M*** system **100** in which multiple ***M***-enabled devices (**102-1, 102-2 ... 102-*n*** — collectively **102**) are in use. An ***M***-enabled device (or ***M*** device) **102** may be any device and any kind of device and may be a stand-alone device or integrated into or combined with any kind of device or system. For example, an ***M***-enabled device **102** may be, without limitation, a device that captures and/or creates content (including digital or analog content), a device that produces and/or renders content (again

including digital and/or analog content). An *M*-enabled device may be (or be incorporated in), for example, a camera, a speaker, a computer, a phone, a set-top box, a television, an appliance, etc. Combinations of these devices are also contemplated, *e.g.*, a set-top box that includes a camera and a speaker and a television monitor and a computer may be one or more *M*-enabled devices. It should be appreciated that the examples given here for devices **102** are merely exemplary and are not intended to be in any way limiting. *M*-enabled devices (or *M* devices) **102** will be described in greater detail below.

[0022] A system **100** preferably includes or has access to one or more Certification Authorities (CAs) **106** which may be part of a larger collection or hierarchy of certification authorities in a public key infrastructure (PKI) scheme (not shown). Each CA **106** is an entity that issues digital certificates for or on behalf of the system **100**.

[0023] *M*-enabled devices **102** are preferably manufactured (at least in part) by one or more authorized device manufacturers **108**. It should be appreciated that devices **102** in **FIG. 1** may be different types of devices made by different manufacturers.

[0024] *M*-enabled devices **102** may interact with each other and with a backend system **104**. It should be appreciated that different types of devices may interact with each other (*e.g.*, an *M* device embodied in a mobile phone may interact with another *M* device embodied in a speaker).

[0025] *M*-enabled devices **102** are each associated with a user **110**. A particular user **110** may have multiple devices **102** associated therewith. In some implementations each device **102** may be associated with only one user **110**. It should be appreciated that the term “user” is an internal entity within the system **100** and is used to define a binding between devices within the system **100**. A user **110** may be considered to be an entity that has a certain relationship within the system **100**. That entity may correspond to a person or a group of people or any other kind of entity (*e.g.*, a company, a school, a club, etc.). In a preferred

implementation, a user **110** is an entity that has registered with the system. A particular person (or entity) may correspond to more than one user **110**. For example, a person may choose to have two users **110** within the system **100**.

[0026] The backend **104** comprises one or more backend applications **112**, at least some of which interact with one or more databases **114** storing and maintaining information about devices **102** and users **110**. The backend **104** may also interact with various other entities, including social networking services **116** (such as Facebook, LinkedIn, and the like), and content providers **118** (such as RDIO, Pandora, Spotify, and the like). It should be appreciated that while the social networking services **116** are shown as separate from the content providers **118**, there may be some overlap between these entities. The backend **104** may interact with entities **120** that can provide added functionality to users **110** (or to devices **102**). The added functionality may include, *e.g.*, voice or instruction recognition functionality, face recognition, and the like. The backend **104** may also interact with other miscellaneous external components **122**.

[0027] The added functionality may provide or enhance or improve on functionality provided on some or all devices **102** or to provide functionality in addition to that provided on devices. The added functionality **120** may be used, *e.g.*, to provide functionality that is beyond the hardware capability of the device **120**. Thus, for example, and as explained in greater detail below, the added functionality **120** may provide voice recognition beyond that which is provided on a particular device **102** (or beyond that which is possible using the hardware of the particular device **102**). Effectively, the added functionality **120** may be used, at least in part, to extend the functionality of any particular *M*-enabled device **102**.

[0028] It should be appreciated that some or all of the functionality provided by the entities **120** may be integrated into the backend **104** (or into backend applications **112**).

[0029] It should also be appreciated that no ownership or management or control (or lack thereof) is implied by the separation of components in the

drawings. Thus, for example, there is no requirement that the CA(s) **104** be owned or operated by (or not be owned by) the same entity that operates the backend **104**. Similarly, there is no requirement that the social networking services **116** and/or the content providers **118** and/or the entities providing added functionality **120** be separately owned and operated (or that they be commonly owned and operated). Some or all of the system **100** may be integrated into a social networking service **116** and/or a content provider **118**.

SYSTEM AND COMPONENT INTERACTIONS

[0030] Various interactions may take place in the system **100** between the various components. These interactions are shown by numbered arcs or lines (denoted **1** to **15**) in **FIG. 1**. In the system **100** depicted in **FIG. 1**:

- Arc **#1** refers to interactions between two devices **102**.
- Arc **#2** refers to interactions between device manufacturer(s) **108** and devices **102**.
- Arc **#3** refers to interactions between CA(s) **106** and devices **102**.
- Arc **#4** refers to interactions between CA(s) **106** and users **110**.
- Arc **#5** refers to interactions between users **110** and the backend **104**.
- Arc **#6** refers to interactions between the CA(s) **106** and the backend **104**.
- Arc **#7** refers to interactions between devices **102** and the backend **104**.
- Arc **#8** refers to interactions between the backend and the database(s) **114**.
- Arc **#9** refers to interactions between the backend and the social networking services **116**.
- Arc **#10** refers to interactions between the backend and the content provider(s) **118**.

- Arc #11 refers to the interactions between the backend **104** and entities **120** providing added functionality.
- Arc #12 refers to interactions between the backend **104** and other miscellaneous external components **122**.
- Arc #13 refers to the interactions between device manufacturer(s) **108** and the Backend **104**.
- Arc #14 refers to interaction between devices **102** and users **110**.
- Arc #15 refers to interactions between CA(s) **106** and device manufacturer(s) **108**.

[0031] The various interactions described here may take place using any known method(s) and protocol(s), and may be wired, wireless or any combination thereof. Some interactions may take place, at least in part, via a network **101** (e.g., a packet-based network such as the Internet). The network **101** may be a public network or a private network or some combination thereof. The network may include one or more cellular and/or satellite components. It should be appreciated that the communications media or protocols through which the various components interact do not limit the system.

[0032] Interactions between the backend **104** and various other components or entities (e.g., the CAs **106**, social networking services **116**, content provider(s) **118**, device manufacturer(s) **108**, entities **120** providing added functionality, and other miscellaneous external components **122**) may make use of application program interfaces (APIs) or other interfaces of those components or entities. In some embodiments, the backend **104** may be integrated with some or all of the other components. It should be appreciated that the manner in which the backend communicates with any other components is not intended to be in any way limiting of the system **100**, and different modes and manners of communication are contemplated herein. Furthermore, the degree of integration (or lack thereof) between the backend **104** and any other components is not intended to be in any

way limiting of the system **100**, and different modes, manners, and degrees of integration are contemplated herein.

[0033] Although, as noted above, the various interactions and communications may take place using any known method(s) and protocol(s), inter-device interactions (Arc #1 in **FIG. 1**) preferably use the quickest and cheapest communications techniques – typically local (*e.g.*, Bluetooth, and the like, or Wi-Fi on a local network) when available, and preferably avoid communication via other networks.

[0034] Those of skill in the art will realize and understand, upon reading this description, that different and/or other interactions are possible in the system, and are not precluded by omission from the drawing. Thus, it should be appreciated that various entities may have other interactions (not shown in the drawing), and furthermore, that not all implementations of the system **100** need support all interactions shown in the drawing.

User IDs

[0035] Each user **110** must have at least one unique User Identity (ID) within the system **100**. A user's User ID may be based on or derived from that user's social network ID (*e.g.*, from their Facebook ID). With reference to **FIG. 2**, a user's User ID may be asserted in the form of a user certificate **175** issued by a CA **106**. The CA(s) **106** may include one or more device CAs **124** (shown in **FIG. 6B**) and one or more user CAs **126** (shown in **FIG. 2**), although it should be appreciated that the device and user CAs **124**, **126**, may be the same entity. If the CA(s) **106** includes user CAs **126**, then a user's user certificate **175** is preferably provided by the backend during a registration process or the like.

[0036] The user certificate **175** may be stored in a user's device **174** (*e.g.*, a smart phone, a smart device enabled by *M* technology, or the like) in a location **176** for such certificates and user identification information.

Users' Friends

[0037] Within the system **100** a user **110** may have one or more friends in the system **100**. A friend of a user within the system **100** is another user **110** with which the first user has established a relationship via the system **100**. The relationship may be established in various ways, *e.g.*, by sharing or interacting with each other's devices **102**. A user may establish permissions with that user's friends. The permissions may depend, *e.g.*, on the type of device(s) **102** associated with the user and may differ from friend to friend.

[0038] A user's "friend" relationship with another user within the system **100** may be of limited scope, *e.g.*, determined by the permissions. The scope of a friendship may be used to limit various aspects of the friendship, *e.g.*, duration of the friend relationship within the system, rights of the friend within the system, etc.

[0039] A user's friends within the system **100** may be distinguished from the user's social network friends (*e.g.*, the user's Facebook friends), although there may be an overlap. *E.g.*, if the user's User ID is based on or derived from that user's social network ID (*e.g.*, from their Facebook ID), the user's social network friends may overlap with the user's friends in the system **100**.

[0040] It should be appreciated that the system is not limited in any way by the manner in which users can establish a "friend" relationship with other users. It should also be appreciated that the system notion of a friend need not bear any relation to actual friendships or other relationships outside of the system **100**, and that friend relationship within the system is used to make associations and bindings within the system **100**.

Device IDs

[0041] Each device **102** has a Device Identity (ID) that must be unique within the system **100**. Creation and storage of device IDs is discussed in greater detail below.

The Databases

[0042] With reference to **FIG. 3A**, the backend **104** includes backend application(s) **112**, including one or more applications that interact with one or more database(s) **114**. Database(s) **114** includes devices database(s) **128** and user database(s) **130**, storing and maintaining information about devices **102** and users **110**, respectively.

[0043] The databases may use any kind of database technology, and no particular database implementation is described or required here, and it should be appreciated that the system **100** is not limited in any way by the database implementation. In some cases, some or all of the database implementation may use third party databases. It should also be appreciated that the user and device databases **128**, **130** may be integrated or made up of multiple different, possibly distributed databases.

Device Database(s) **128**

[0044] As shown in **FIGS. 3A** and **3B**, the database(s) **114** may maintain (in device database(s) **128**), information about each device **102** within the system **100**, including information about the device's owner (a user **110**), the device's capabilities, the device's history, and other information such as, *e.g.*, the devices last known location, the device's manufacturer, etc. Preferably the device ID is used as a key (or index) into the device database **128**. When a device **102** is manufactured it has no owner (unless specifically provisioned at manufacture). A device's capabilities may include its type (*e.g.*, speaker, etc.). It should be appreciated that the terms "own" and "owner" are not used here to imply or require any legal notion of ownership of a device by a user, and they refer to an association or binding between devices **102** and users **110** within the system **100**.

[0045] The devices database **128** may include device corpora for each device. As will be explained in greater detail below, a device's corpora may

correspond to corpora stored on the device, and may be used by the device or by users of the device.

The User Database 130

[0046] Preferably the User ID acts as a primary database key to the user database **130**. The user database **130** preferably associates certain information with each user **110**, including one or more of the following (with reference to **FIG. 3C**):

- A user profile describing information about that user. The user profile may be linked to or obtain information from social network data of the user (*e.g.*, from the user's Facebook profile). The user profile may also contain information about content providers that the user is associated with (*i.e.*, has accounts with), such as, *e.g.*, RDIO, etc.
- The user's devices **102** and, optionally, information about those devices (*e.g.*, the device's capabilities). Those of skill in the art will realize and understand, upon reading this description, that some information in the user database **130** may be obtained from the devices database **128**, and vice versa. For example, if the user database **130** stores device identifiers for the devices of each user, then those device identifiers can be used to obtain device-specific information for the corresponding devices from the devices database **128**.
- The user's "friends" within the system **100** and possibly permissions associated with those friends. In preferred implementations, a user's "friends" within the system are other users **110**, so that the information about each user's friends may be stored within the user database **130** as a list of zero or more user IDs. Users may be able to set customized permissions for each friend. In some embodiments, users may be able to define or use classes or groups or types of friends and assign permissions based on classification or group or type. In this manner a user may be able

to make use of a template or the like to quickly set or change permissions for one or more users. It should be appreciated that the system is not limited by the manner in which users select their friends or set permissions for their friends. Users may also remove other users as friends.

- The user's history, preferably stored or searchable by one or more of: device **102**, friend (*i.e.*, using one or more devices or friends as keys for a search of the history), and time. The device history may be a sequence of time-stamped events relating to each of the user's devices. For example, if a particular device **102** is a speaker, that device's history may be a time-stamped list of what was played over that speaker. The user's history relating to friends may include times and locations at which a device of the user interacted with (or was used by) a device of the user's friend (and/or vice versa). The examples given here for the kinds of historic details that may be searched are not intended to limit the scope of this description in any way, and those of skill in the art will realize and understand, upon reading this description, that different and/or other information and/or factors may be used to search a user's history.
- The user's IDs and certificates (including those issued by user CA **126**).
- User corpora relating to various device interface mechanisms (described in greater detail below). The user corpora preferably include local corpora and extended corpora. As will be explained in greater detail below, a user's local corpora may correspond to that user's corpora on one or more devices, and may be used by devices when under control of the user. A user's extended corpora may extend beyond the capacity of some or all devices **102**, and correspond to that user's corpora that may be used by the added functionality **120**.
- Configuration information, preferably including configuration details about the user's devices and information that may be used to configure other (*e.g.*, new) devices. The configuration information may contain

information about the user's wireless network settings in various locations, including passwords and other information that may be needed for a device to connect to those networks.

[0047] Those of skill in the art will realize and understand, upon reading this description, that different and/or other information, associations and relationships may be stored in device and user databases **128**, **130**.

DEVICES

[0048] Each device **102** has certain device-specific functionality / components associated therewith. For example, if the device **102** is a speaker (or camera or phone or the like), then the device-specific functionality / components will include functionality / components used to operate the device as a speaker (or camera or phone, etc.). With reference now to **FIG. 4A**, a device **102** includes mechanisms **132** supporting the device's device-specific functionality / components.

[0049] As used herein, the term "mechanism" means hardware, alone or in combination with software, and includes firmware.

[0050] A device **102** also includes various *M* system mechanisms / data **134** used to support and implement functionality within the system **100**. The system mechanisms / data **134** may interact (at **136**) with the device's device-specific functionality **132**, although the system mechanisms / data **134** are preferably distinct from the device-specific functionality **132**.

[0051] A device **102** preferably includes one or more sensors **138** that may be used by the system mechanisms **134** to implement aspects of the system functionality. As used herein, a "sensor" means any mechanism or device that can detect and/or measure a physical property or stimulus (*e.g.*, heat, light, sound, pressure, magnetism, motion, touch, capacitance). A sensor preferably provides an indication (*e.g.*, as an electrical signal) of the property or stimulus it detects/measures. A sensor may also be able to record, indicate, or otherwise

respond to the physical property or stimulus it detects/measures. A sensor may be implemented in hardware, software, or combinations thereof. A sensor may be provided as a stand-alone device or chip or may be integrated into other devices (or sensors) or chips. The specific sensors **138** may be implemented, at least in part, using specialized chips or circuitry or the like. Those of skill in the art will realize and understand, upon reading this description, that the system is not limited in any way by the manner in which sensors are implemented or integrated. It should also be understood that a particular sensor may detect and/or measure more than one kind of physical property or stimulus.

[0052] With reference to **FIG. 4A-1**, the sensors **138** may include, for example, one or more cameras, microphones, motion sensors (external motion and/or device motion), accelerometers, a compass, location positioning systems (LPSs), and the like. As used herein, LPS refers generally to any location positioning system that can be used to determine location of the device, and includes the United States' Global Positioning System (GPS) and the Russian GLObal NAVigation Satellite System (GLONASS), the European Union Galileo positioning system, the Chinese Compass navigation system, and the Indian Regional Navigational Satellite System. LPS also includes Wi-Fi and/or cellular phone system(s) that can provide position data, as well as assisted and augmented positioning systems, *e.g.*, using Wi-Fi and/or a cellular phone system. Although only one camera may be used for the face and gesture processing, more than one camera allows for three-dimensional (3D) processing, if needed.

[0053] The sensors **138** may interact with the system mechanisms **134** via standard interfaces that are provided for each sensor (at **140**). It should be appreciated that not every device **102** need have every kind of sensor **138**, and the different kinds or devices (or different implementations of the same kind of device) may have different sensors or kinds of sensors.

[0054] A device **102** preferably includes one or more communications mechanisms **142** that may be used by the system mechanisms **134** (at **144**) to

implement aspects of the system functionality. With reference to **FIG. 4A-2**, the communications mechanisms **142** may include, *e.g.*, one or more of: mechanisms for local communication (*e.g.*, Bluetooth, including Bluetooth Low Energy (BLE), ZigBee, etc.), mechanisms for Wi-Fi communication (*e.g.*, 802.11, etc.), mechanisms for cellular communication (*e.g.*, modems or other devices using a cellular telephone network, etc.); and mechanisms for wired communication (*e.g.*, Ethernet or the like). The communications mechanisms **142** may be implemented in protocol-specific chips or the like. It should be appreciated that not every device **102** need have every kind of communications mechanism **142**, and the different kinds or devices (or different implementations of the same kind of device) may have different sensors or kinds of communications mechanisms. However, at a minimum, each device **102** should be able to communicate, at least some of the time, in some manner with the backend **104** (arc #7 in **FIG. 1**) (whether by cellular / telephone network, by Wi-Fi, by wire, or in some other manner). Also, preferably each device **102** has at least one communications mechanism that allows it to communicate, at least some of the time, with other devices **102** in the system **100**. Each device also includes the required/appropriate connectors and/or antenna(e) for its various communication mechanism(s). These connectors and/or antenna(e) may be incorporated into the various communication mechanisms, especially when the communication mechanisms are provided in specialized chips or circuits.

[0055] The system mechanisms **134** (including any required sensors **138**) may be integrated into a device **102** as a separate board or chipset or they may share components with the device's mechanisms used to implement the device-specific functionality (as shown by the dotted-line oval shape connecting them). For example, if the device-specific functionality requires a microphone, that microphone may be shared with (or used by) the system mechanisms **134** as a sensor **138**. Similarly, at least some of the communications mechanisms may be shared between system mechanisms **134** and the device-specific functionality **132**

(as shown by the dotted-line oval shape connecting them). It should be appreciated, however, that the system mechanisms **134**, sensors **138**, and communications mechanisms **142** need to be able to operate and be controlled independent of the device-specific mechanisms, and that this need may override or prevent sharing of components in some implementations.

[0056] A device **102** may include a computer system **146** (which is described in greater detail below). The computer system **146** may interact (at **148**) with the system mechanisms **134** and may implement aspects of those mechanisms. Although shown as separate components in the drawing in **FIG. 4A**, it should be appreciated that some or all of the computer system **146** may be shared with and be part of the system mechanisms / data **134** (as shown by the dotted oval shape connecting them). Similarly, at least some of the computer system **146** may overlap with the device-specific functionality **132** of the device **102** (as shown by the dotted oval shape connecting them).

[0057] A device's Device Identity (ID) and other identity and certificate information may be stored (at **150**) in the system mechanisms / data **134**.

[0058] Each device **102** preferably includes bootstrapping / provisioning mechanisms **152** and various operational mechanisms **154** (both described in greater detail below). The various operational mechanisms **154** have corresponding operational storage **155** on the device. The operational storage **155** may store data used by the various operational mechanisms, and may be used for persistent and/or temporary storage. Those of skill in the art will realize and understand, upon reading this description, that some or all of the operational storage **155** may be integrated with other storage on the device, and operational storage **155** is shown in the drawing as a separate component to aid with this description.

[0059] Each device **102** includes at least one power system **157** that can power the device, including the system mechanism(s) **134**, the computer system **146**, sensors **138**, communications **142** and the device specific functionality **132**.

The power system **157** may include separate systems for some or all of the components, and may include battery power supplies alone or in conjunction with external power supplies. Preferably the system mechanisms have a separate power supply (*e.g.*, a battery) from the device specific functionality. When an external power supply is used (*e.g.*, A/C power via an adaptor), all components of the system should use the external power source even if they have separate internal power systems for use when not connected to an external source. However, it should be appreciated that the system is not limited by the manner in which power is supplied to the components.

Operational Mechanisms

[0060] With reference to **FIGS. 4A – 4C**, a device's operational mechanisms **154** may include some or all of the following mechanisms:

- Mechanisms **156** supporting device-to-device interaction and communication (this corresponds to the interaction shown by arc #1 in **FIG. 1**). The device-to-device interaction mechanisms **156** have and may use corresponding device-to-device storage **169** (**FIG. 4C**).
- Mechanisms **158** for command and/or control of the device **102**. Command and/or control mechanisms **158** have and may use corresponding command and/or control storage **159** (**FIG. 4C**).
- Mechanisms **160** supporting device-to-backend interaction and communication (this corresponds to the interaction shown by arc #7 in **FIG. 1**). Device-to-backend mechanisms **160** have and may use corresponding device-to-backend storage **161** (**FIG. 4C**).
- Interface mechanisms **162** used to operate the device within the system **100** and to support interaction (preferably including human interaction) with the device's other mechanisms and functionality. The interface mechanisms **162** have and may use corresponding interface mechanisms' storage **163** (**FIG. 4C**).

- Other operational mechanisms **164** used to operate the device within the system **100** that may have and use corresponding storage **165** (FIG. 4C).

[0061] The interface mechanisms **162** may include one or more of:

- Gesture mechanisms **166** that may be used by the operational mechanisms **154** to implement operational and/or functional features that make use of users' gestures (*e.g.*, gesture commands and the like). Gesture mechanisms **166** may have and use gesture storage **167**. Gesture mechanisms **166** preferably include one or more gesture detection mechanisms **168** and one or more gesture recognition mechanisms **170**, and the gesture storage **167** preferably includes associated gesture corpora **172** for use by the various gesture mechanisms. A gesture corpus (plural "corpora") refers to a collection of gestures or gesture samples usable by gesture mechanism(s) to detect and/or recognize gestures. In preferred systems, the gesture detection and/or recognition mechanisms **168**, **170** may be trained and adapted to detect and/or recognize gestures of one or more people (who may be users), and the associated gesture corpora may be modified based on this training. The gestures mechanism(s) **166** may use one or more sensors **138**, including, *e.g.*, camera sensor(s).
- Voice/speech mechanisms **174** that may be used by the operation mechanisms **154** to implement operational and/or functional features that make use of human (*e.g.*, users') voices (*e.g.*, for voice commands and the like). Voice/speech mechanisms **174** may have and use voice/speech storage **175**. Voice/speech mechanisms **174** preferably include one or more voice/speech detection mechanisms **176** and/or voice/speech recognition mechanisms **178**, and the voice/speech storage **175** preferably

includes associated corpora **180** for use by the various voice/speech mechanisms. A voice/speech corpus refers to a collection of word or phrase samples usable by voice/speech mechanism(s) to detect and/or recognize voice/speech. In preferred systems, the voice/speech recognition mechanisms may be trained and adapted to recognize voice/speech of one or more people (*e.g.*, users), and the associated speech corpora may be modified based on this training. The voice/speech mechanisms **174** may use one or more sensors **138**, including, *e.g.*, microphone sensor(s).

- Face/gaze mechanism(s) **182** that may be used by the operation mechanisms **154** to implement operational and/or functional features that make use of peoples' faces and/or gazes. Face/gaze mechanism(s) **182** may use face/gaze storage **183**. The face/gaze mechanism(s) **182** may include face/gaze detection mechanism(s) **184** and/or face/gaze recognition mechanism(s) **186**, and/or face movement detection mechanism(s) **187**, and the face/gaze storage **183** preferably includes associated face/gaze corpora **188** for use by the face/gaze recognition/detection mechanism(s) **182**. A face/gaze corpus refers to a collection of face and/or gaze samples usable by face/gaze mechanism(s) to detect and/or recognize faces and/or gazes. In preferred systems, the face/gaze recognition/detection mechanisms **182** may be trained and adapted to recognize one or more faces, and the associated face/gaze corpora **182** may be modified based on this training. The face movement detection mechanism(s) **187** may detect movement of parts of a face, *e.g.*, movement of the mouth, eyes, etc., and may be used, *e.g.*, to try to confirm that a person is

speaking. Face/gaze mechanism(s) **182** may use sensors **138**, including, *e.g.*, camera sensor(s).

- Other interface mechanisms **190** that may be used by the operation mechanisms **154** to implement operational and/or functional features that make use of other types of user interactions (*e.g.*, touch, typing, device movement, etc.). Other interface mechanisms **190** may use other interface mechanisms storage **191**. The other interface mechanism(s) **190** may use sensors **138**.

[0062] While corpora for speech, gesture, and face / gaze recognition / detection are mentioned above, those of skill in the art will realize and understand, upon reading this description, that other mechanisms, especially the interface mechanisms **162**, may have associated corpora and may also learn and adapt based on interactions the device has within (or outside of the system **100**), including interactions with humans, interactions with other devices **102**, interactions with the backend **104**, and interactions with users **110**. In general, a corpus for a particular interface mechanism refers to a collection of samples usable by that particular gaze mechanism to function and perform. Thus, *e.g.*, as shown in **FIG. 4C**, the other interface mechanisms **190** may include associated corpora **192**.

[0063] Unless stated otherwise, as used herein the word “corpora” refers to a single “corpus” and/or plural corpora. Thus, *e.g.*, when a device is described as having corpora for some feature, that should be read as the device having at least one corpus for that feature, or the device having a single corpus for that feature or multiple corpora for that feature.

[0064] In some preferred embodiments, a user can train their device(s) to recognize certain phrases and/or gesture patterns. One or more of these signature phrase/gesture patterns may be used (and required), *e.g.*, to trigger some or all of the commands for the device(s). It should be appreciated that a device may learn (and thus be trained) without any specific user intervention or request, and,

preferably, each device learns to recognize user interactions without being specifically requested to do so.

[0065] A device **102** also preferably includes heartbeat (HB) mechanism(s) **194** (described in greater detail below). The HB mechanism(s) may interact with other operational mechanisms **154**, including the device-to-device mechanism(s) **156** and device-to-backend mechanism(s) **158**.

[0066] Those of skill in the art will realize and understand, upon reading this description, that a particular device may use only some of the interface mechanisms **162**. It should also be appreciated, as will be described in greater detail below; various of the interface mechanisms **162** may be used in combination.

[0067] Preferably each operational mechanism **154** on a device **102** is able to operate on the device without any external interaction (that is, without any interaction with any other devices **102** or the backend **104**). However, it should be appreciated that the various operational mechanisms **154** may operate or use (or be enhanced by) added functionality **120** provided, *e.g.*, via the backend **104** or one or more other devices **102**. Thus, for example, various voice/speech mechanisms **174** may support a limited use of voice commands and the like when used without any external interaction. These voice commands may be limited, *e.g.*, by the capabilities of the various mechanisms and by the capacity (such as memory and computational capacity) of the device. The system **100** may provide extended voice commands and interactions to a device when enhanced by added functionality **120**. The limited local voice/speech interactions would be reflected in the voice/speech corpora **180** stored in the voice/speech storage **175** on the device. In the case of voice interactions, *e.g.*, a device **102** may have a limited corpus of words (*e.g.*, stored as voice/speech corpora **180** in voice/speech storage **175**) that it can recognize and that can be used to control aspects of the device. Using the added functionality **120** provides the device access to a potentially much larger corpus as well as to the ability to parse more complex instructions / queries.

For example, if the device **102** is a speaker, the corpus on a device **102** along with the voice/speech mechanisms **174** may support commands / instructions such as “Play louder”, “Softer”, etc., whereas an external corpus, provided by the added functionality **120** may support more complex instructions such as, *e.g.*, “Play the song I was listening to last night.” This latter request is more complex grammatically (therefore possibly requiring more speech recognition capability than the device can provide) and may also require access to external data (*e.g.*, in the database(s) **114**).

[0068] The interface mechanisms **162** may be provided with or include learning mechanisms that enable them to learn about the users of the device and about how those users interact with the device. For example, the voice/speech mechanisms **174** may learn to recognize the voices of particular people (preferably including a person corresponding to the user **110** associated with the device).

[0069] A device is preferably initially configured with generic corpora for its various interface mechanisms **162**. As a device’s interface mechanisms **162** learn, they may update the various corpora to reflect what has been learned. For example, as a device learns to recognize the voice of a particular person, it may update the voice/speech corpora for that person. Preferably a device retains a copy of the initial generic corpora and/or the original corpora may be saved in some location distinct from the device (*e.g.*, in the “cloud”).

[0070] A device may have different corpora for different users or people. However, it should be appreciated that as each device is associated with a user, preferably each interface mechanism on a device has at least one corpus associated with that user. Other people who use the device may have their own corpora associated therewith.

Monitoring for Interactions

[0071] As shown in **FIG. 4I**, a device **102** monitors for possible interactions (at **S402**). The device may monitor continuously (at **S402**) or it may monitor at

specific times and/or under specific conditions. This monitoring preferably uses one or more of the devices sensors **138** (*e.g.*, the devices camera(s), microphone(s), etc.). In order to avoid missing some interaction, the device preferably buffers or stores potential interactions (at **S404**). These buffered interactions may be stored in any known and appropriate manner for subsequent use, if needed, by the various interface mechanisms **162**. For example, sound detected by the devices microphone(s) may be buffered in a manner suitable for subsequent use, if needed, by the voice/speech mechanism(s) **174**. Similarly, external movement detected by the devices camera(s) may be buffered in a manner suitable for subsequent use, if needed, by the device's gesture mechanism(s) **166**, and images detected by the device may be stored in a manner suitable for use, if needed, by the devices face/gaze mechanism(s) **182** and/or other mechanism(s) **190**. It should be appreciated and understood that not all sensor input may be buffered for every device or type of device. However, it should also be appreciated that buffering sensor information may allow a device to provide more accurate interactions via its various interface mechanism(s) **162**, since the device may be able to reconstruct an interaction even if the device only realizes that there is an interaction occurring some time after the interaction has begun.

[0072] Those of skill in the art will realize and understand, upon reading this description, that the amount of information buffered (at **S404**) depends on the kind of information (*e.g.*, voice, images, movement history, etc.). Similarly, those of skill in the art will realize and understand, upon reading this description, that different information may be buffered for different periods of time. For example, sounds detected by the devices microphone(s) may be buffered for 30 seconds, whereas images detected by the camera(s) may be buffered for 15 seconds. Buffering may use any technique, *e.g.*, circular or wrap-around buffering, and the system is not limited by the kind of buffering used or the implementation of the buffering. Different buffering techniques may be used for different kinds of

information. The device **102** should have sufficient memory to store the required amount of buffered information.

[0073] Having buffered possible interaction(s) (at **S404**), the device tries to determine if a possible (buffered) interaction is an actual interaction (at **S406**). If the device determines that a possible interaction is (or may be) an actual interaction (at **S406**), the device may continue (at **S408**) to process that actual interaction. The processing of the actual interaction (at **S408**) may use the corpora selection process described above with reference to **FIG. 4I**.

[0074] The determination (at **S406**) of whether there is (or may be) an actual interaction going on with the device may make use of one or more triggers. For example, if a person starts to give a device voice commands, the device may not know that the sound it is detecting corresponds to those voice commands. However, if the person is also looking at the device while talking (as detectable using gaze detection), then the device may rely on that detected gaze, in combination with the detected speech, as a trigger to process an actual interaction.

[0075] A device **102** may operate in lots of different kinds of environments, and environmental factors and associated noise may affect interaction detection and processing. As used here, the term “noise” with respect to any kind of information or signal refers to information and/or signals that that may degrade processing of the information or signal (*e.g.*, that may degrade the corresponding detection and/or recognition of information in the signal). For example, the background sound of an air conditioner or fan may interfere with or degrade voice/speech recognition; or a constantly flashing light may interfere with or degrade the face or gesture mechanisms. It is therefore useful for a device to try to filter or remove noise prior to processing. In this regard, the other mechanism(s) **164** may include one or more noise removal filtering/cleaning mechanisms to remove noise from the inputs detected by the various sensors. One such mechanism is sound-noise cancellation that removes ambient sounds (*e.g.*, from air conditioners and fans) from sound detected (and buffered) by the devices

microphone(s). Different and or other noise removal filters may be used, and these filter mechanisms may adapt and learn from the environment in which the device is situated.

[0076] In cases where filtering/cleaning mechanisms are used, preferably all buffered information is filtered/cleaned.

[0077] The flowchart in **FIG. 4J** describes an exemplary process a device **102** may use to determine whether an interaction is taking place (at **S406**, **FIG. 4I**). Recall that various of the sensors **138** are monitored (at **S402**, **FIG. 4I**) and the output of the sensors is buffered (at **S404**, **FIG. 4I**) in case an actual interaction is detected and the sensor output is needed. Output from the various sensors **138** is provided (directly or via buffers) to the various interface mechanisms **162**, each of which attempts to determine whether or not an interaction is taking place. For example, the gesture detection mechanism(s) **168** try to determine (at **S410**) whether any movement that they are detecting via the camera sensor(s) corresponds to a gesture. The face/gaze detection mechanism(s) **184** try to determine (at **S412**) whether any images that they are detecting via the camera sensor(s) corresponds to at least one face. The voice/speech detection mechanism(s) **176** try to determine (at **S414**) whether any sound that they are detecting via the microphone sensor(s) corresponds to a speech. The other mechanism(s) **190** try to determine (at **S416**) whether they are detecting via the sensor(s) **138** (e.g., touch, movement, proximity of other user device) corresponds to an interaction with the device. The user detection mechanism(s) try to detect (at **S418**) whether a user is interacting with the device.

[0078] If a face (or more than one face) is detected (at **S412**), the gaze detection mechanism(s) may be used (at **S420**) to determine if a detected face is looking at the device, and face movement detection mechanism(s) **187** may be used to determine (at **S422**) if a detected face is moving in a manner that corresponds to a gesture and/or speech.

[0079] Face movement detection mechanism(s) **187** may also be used alone or in conjunction with face / gaze detection mechanism(s) **184** to determine whether any detected face is moving in such a way as to correspond to speech and/or gestures.

[0080] The various interaction detection mechanisms may operate concurrently (as shown in the flow chart depicted in **FIG. 4J**), although some of the mechanisms (*e.g.*, gaze detection and face movement detection) may depend on determinations of other mechanism(s).

[0081] In some implementations the various detection mechanisms may produce a Boolean value (true or false) reflecting their detection decision. In those cases the final determination as to whether an interaction has been detected is determined by the logical OR of those values. *E.g.*, if there are N detection mechanisms, and the i -th detection mechanism produces a Boolean value b_i , then the final determination as to whether an interaction has been detected may be computed using:

$$\text{Interaction Detected} = OR\ b_i, i=1 \dots N \quad (1)$$

[0082] In this implementation, any *true* value (*i.e.*, any detection by any detection mechanism) will result in a determination that an interaction is taking place. With reference to the flow chart in **FIG. 4J**, any “Yes” value by any mechanism will result in a positive determination that an interaction is taking place.

[0083] It should be appreciated that the label “Yes” on the lines in the flowchart of **FIG. 4J** represent a decision or determination by each of the respective mechanisms that they have most likely detected some feature (*e.g.*, gesture, face, speech, etc.). Similarly, the label “No” on lines in that flowchart represent a decision or determination by each of the respective mechanisms that they have most likely not detected some feature (*e.g.*, gesture, face, speech, etc.). It should therefore be appreciated that the various labels on these lines should not be construed to mean that a feature is or is not occurring, only that it was or was

not detected with some degree of certainty. Thus, those of skill in the art will realize and understand, upon reading this description, that the “Yes” labels on the flow lines in **FIG. 4J** may be read as “Maybe” or “Probably” or “More likely than not,” and that the “No” labels on the flow lines in that figure may be read as “Probably not.” Accordingly, the value *Interaction Detected* (determined using equation 1 above), when “True”, should be construed to mean “Most likely true” or “More likely true than false”, and, when “False,” should be construed to mean “Most likely false.”

[0084] The device may use the probabilistic nature of the interaction detection. Thus, in some implementations each detection mechanism may produce a value or number reflecting a degree of certainty of the detection result (*e.g.*, a real number from 0.0 to 1.0, where 0.0 means no interaction was detected and 1.0 means an interaction was certainly detected, or an integer from 0 to 100, with 0 meaning no interaction and 100 meaning definite interaction). In those cases the final determination as to whether an interaction has been detected is determined, at least in part, using a score determined as a function of the values produced by the various detection mechanisms. For example, if there are N detection mechanisms, and the i -th detection mechanism produces a real value r_i in the range 0.0 to 1.0, and the i -th detection mechanism’s score has a weight of w_i , then

$$\text{Interaction Score} = \mathbf{F}(r_i, w_i, i=1 \dots N) \quad (2)$$

For some function $\mathbf{F}$. It should be appreciated that the weights w_i in formula 2 are such that the *Interaction Score* ≤ 1.0 . The value *Interaction Detected* may be determined, *e.g.*, by comparing the *Interaction Score* (*e.g.*, as determined in equation 2) to a static or dynamic threshold. For example,

$$\text{Interaction Detected} = \text{Interaction Score} > \text{Threshold } (T) \quad (3)$$

[0085] In some implementations the function $\mathbf{F}$ (formula 2) may produce a weighted sum or average of the values produced by the various detection mechanisms, where different weights may be given to different detection mechanisms, depending, *e.g.*, on their known or perceived or historical accuracy.

For example, if there are N detection mechanisms, and the i -th detection mechanism produces a real value r_i in the range 0.0 to 1.0 , and the i -th detection mechanism's score has a weight of w_i , then the final determination as to whether an interaction has been detected may be computed using:

$$\text{Interaction Score} = \sum_{i=1..N} r_i w_i \quad (2')$$

As noted above (with reference to formula 2), the weights w_i are such that the $\text{Interaction Score} \leq 1.0$

[0086] Although a weighted sum is used in this example, those of skill in the art will realize and understand, upon reading this description, that different and/or other functions may be used to determine whether an interaction is taking place. For example, as shown in formula 2'' below, a weighted average score may be used to determine the interaction score:

$$\text{Interaction Score} = \frac{\sum_{i=1..N} r_i w_i}{N} \quad (2'')$$

[0087] The flow chart depicted in **FIG. 4K** shows the detect interaction step (**S406'**) in an implementation in which each detection mechanism produces a score (*e.g.*, a real value R_i in the range 0.0 to 1.0), and the final determination as to whether an interaction is detected is determined, at least in part, as a function of those scores (*e.g.*, per equation 2 or 2' above). In this example, the gesture detection mechanism(s) **168** try to determine (at **S410'**) whether any movement that they are detecting via the camera sensor(s) corresponds to a gesture, and produce a score ($\mathbf{R}_1 \in [0 .. 1]$) indicative of this determination. The face/gaze detection mechanism(s) **184** try to determine (at **S412'**) whether any images that they are detecting via the camera sensor(s) corresponds to at least one face, and produce a score ($\mathbf{R}_2 \in [0 .. 1]$) indicative of this determination. The voice/speech detection mechanism(s) **176** try to determine (at **S414'**) whether any sound that they are detecting via the microphone sensor(s) corresponds to a speech, and produce a score ($\mathbf{R}_3 \in [0 .. 1]$) indicative of this determination. The other interface mechanism(s) **190** try to determine (at **S416**) whether they are detecting via the

sensor(s) **138** (*e.g.*, touch, movement, proximity of other user device) corresponds to an interaction with the device, and produce a score ($\mathbf{R}_4 \in [0 .. 1]$) indicative of these determinations. The user detection mechanism(s) try to detect (at **S418'**) whether a user is interacting with the device and produces a score ($\mathbf{R}_7 \in [0 .. 1]$) indicative of its determinations.

[0088] As in the previous example (**FIG. 4J**), if a face is detected (at **S412'**), the gaze detection mechanism(s) may be used (at **S420'**) to determine if a detected face is looking at the device. In this case the gaze detection mechanism(s), if used, produce a score ($\mathbf{R}_5 \in [0 .. 1]$) indicative of their determinations. Similarly, face movement detection mechanism(s) **187** may be used to determine (at **S422'**) if a detected face is moving in a manner that corresponds to a gesture and/or speech and to produce a score ($\mathbf{R}_6 \in [0 .. 1]$) indicative of their determinations. The decision as to whether or not to initiate or use the gaze detection and/or face movement detection mechanisms may be based, *e.g.*, on the score ($\mathbf{R}_2$) produced by the face detection mechanism(s) (at **S412'**), where predetermined threshold values (denoted $\mathbf{T}_G$ and $\mathbf{T}_M$ in **FIG. 4K**) may be used to initiate the gaze and / or face movement detections. The values of these thresholds ($\mathbf{T}_G$ and $\mathbf{T}_M$) may be the same. The threshold values may be preset and fixed or they may be dynamic, based, *e.g.*, on information the system learns about its detection success.

[0089] As noted above, each detection mechanism may have a corresponding detection weight associated therewith. For example, as summarized in the following table:

Mechanism	Detection Weight
Gesture Detection	w_1
Face detection	w_2
Voice / Speech detection	w_3
Other detection	w_4
Gaze detection	w_5
Face movement detection	w_6
User detection	w_7

[0090] The system preferably produces (at **S407**) a running value corresponding to the most recent scores (in this example implementation, $R_1 \dots R_7$) produced by each of the various detection mechanisms. Each detection mechanism may have a corresponding detection weight associated therewith, and the running value may be produced using a weighted function of the various scores (*e.g.*, using formulas **2**, **2'** or **2''** above). The decision as to whether or not an interaction has been detected may be based, at least in part, on a comparison between this score (computed at **S407**) and another threshold value (denoted $T_{\text{Interaction}}$ in the drawing in **FIG. 4K**). The threshold value $T_{\text{Interaction}}$ may be preset and fixed or dynamic, based, *e.g.*, on information the system learns about its detection success.

[0091] The various interaction detection mechanisms / processes may proceed in parallel with and independent of each other. However, it should be appreciated that the various mechanisms / processes may, in some cases, have access to and use the scores produced by the other mechanisms. These other mechanisms' scores may be used, *e.g.*, to trigger other detection (as in the exemplary cases in **FIG. 4K** where the score (R_2) produced by the face detection at **S412'** is used as a trigger for gaze and face movement detection).

[0092] The weights for the various detection mechanisms may vary dynamically based on information the device has learned from prior detections. Each weight is a value in the range 0 to *max_weight* for some value of *max_weight*. For the purposes of this description, assume that *max_weight* is 1.0, although any value can be used, as long as the score threshold value ($T_{\text{Interaction}}$ in the example above) is set accordingly. Initially all mechanisms may be given the same weight (*e.g.*, *max_weight*), and then the weights assigned to certain mechanisms may be adjusted based on the accuracy or usefulness of those mechanisms. For example, in a darkened room, the gesture and face detection mechanisms may be given reduced weight, or in a noisy room (where the device

cannot adequately filter out the noise), the speech detection mechanism(s) may be given reduced weight. Once the light in the room changes or the noise in the room is reduced, the weights of the corresponding mechanisms may be adjusted upwards. In addition, as the device learns from its prior decisions, weights can be adjusted.

[0093] It should be appreciated that it is generally preferable to have false positive detection of possible interactions than to miss actual interactions. If the device does not provide adequate (*e.g.*, near real time) response to interactions then users will stop using the interactions and possibly the device. Accordingly, preferably the weights should be set conservatively high and the corresponding threshold values should be conservatively low in order to detect possible interactions with the device.

[0094] Each of the detection mechanisms consumes power, and the constant use of these mechanisms by the device may cause it to use too much power. The need for accurate interaction detection must therefore be balanced with the need to conserve power on the device. When a device is connected to an external power source, however, no limit need be placed on the use of detection mechanisms. Triggers for the various mechanisms may be used to conserve power. For example, the speech detection mechanisms may be triggered by a simpler sound detection mechanism (that consumes less power), and the face and gesture detection mechanisms may be triggered by a simpler mechanism (that consumes less power) that detects changes in the images captured by the camera sensor(s).

[0095] In some implementations, devices may be put into a mode (*e.g.*, a sleep mode or an ignore mode) during which they perform minimal interaction detection. This mode may be entered, *e.g.*, based on a time of day, location, user instruction, or some other factor. For example, a device may be set to perform minimal interaction detection between midnight and 7 AM on weekdays. In some cases, where a device is used in a crowded or noisy location (*e.g.*, a speaker at a party or a phone at a busy airport), the device may be set to ignore certain

interactions or to require triggers or confirmations for interactions. In some cases, certain interactions may be disabled based on what the device is currently doing, *e.g.*, speech input may be disabled while a music playing device is actively playing music, but reactivated between pieces of music. This kind of setting may be done by adjusting the weights associated with some of the detection mechanisms (*e.g.*, set the weight for gesture detection very low at a dance party). Setting a low weight for a particular detection mechanism will not disable that mechanism and so it will still operate and consume power. Furthermore, even with a low weight value, a mechanism may detect a gesture that the device may construe as a possible interaction. In preferred implementations certain interaction detection mechanisms can be turned off or disabled so that they do not operate or consume any power. In some cases this may be achieved by setting the corresponding weight for a detection mechanism to zero (0.0).

[0096] In preferred implementations, each detection mechanism resets its score to zero (or *false*) after some period of time (*e.g.*, 5 – 10 seconds) or after a detected interaction has actually been carried out.

[0097] The ongoing process of checking for possible interactions (*e.g.*, **S406** in **FIG. 4I**) is typically performed without knowledge of which person might initiate the interaction. Accordingly, to the extent this process requires (or could benefit from) the use of corpora, the device preferably uses the corpora of the user associated with the device to perform the recognition processes. As will be explained below, once a possible interaction is detected, the device may use (or try to use) different corpora to recognize and process the actual interaction.

[0098] Upon detection of possible interaction(s) (**S406**, **FIG. 4I**), the device preferably maintains an indication of what kind of interaction(s) may be going on. This may be achieved by having the process of determining possible interactions (**S406**, **FIGS. 4J, 4K**) produce and provide information for use by subsequent processing. In the case of the Boolean implementation of **FIG. 4J**, the process may set a bit vector corresponding the values b_i , $i=1 \dots N$ used to determine the

value of *Interaction Detected* in formula 1 above. In the case of the real values used in the implementation in **FIG. 4K**, the process may set an array or vector of values corresponding to the score produced by each detection mechanism for use by subsequent processing. Other approaches may be used for the device to maintain information about possible interactions, and the system is not limited by the manner in which this information is maintained or communicated.

[0099] While trying to determine whether an interaction is taking place (at **S406, FIG. 4I**), the device may not know (or need to know) which corpora to use for the various interface mechanisms **162**. In preferred implementations, if interaction detection requires a corpus, the device **102** uses corpora associated with the user **110** with which the device is associated. In some cases, a device may use other corpora (*e.g.*, generic corpora or corpora associated with other authorized users) at this stage of processing.

[00100] Having detected a possible interaction (**S406, FIG. 4I**), the device proceeds to process the interaction. The processing of the actual interaction preferably uses the interface mechanisms **162** along with other operational mechanisms **154** to process the interaction. In order to process an interaction, the device may need to use corpora for the various interface mechanisms **162**. Thus, as shown in the flowchart in **FIG. 4L**, the device first determines (at **S409**) which corpora to use for the various interface mechanisms.

[00101] The corpora (both on the device and at the backend) may be organized, *e.g.*, as shown in **FIGS. 4D – 4H**, so that when the device recognizes a particular person by their face, gestures, voice, or in some other way, the other corpora associated with that person may be determined and used by the corresponding mechanisms. For example, if a device first recognizes a person's face, the device may, if needed, access and use the other corpora associated with that person (**FIG. 4D**) (or it may make an association based on a user's smart device (*e.g.*, phone or tablet). Similarly, if a device first recognizes a person's gestures (**FIG. 4E**), voice/speech (**FIG. 4F**), or some other aspect of the person

(**FIG. 4G**), the device may, if needed, access and use the other corpora associated with that person. Additionally, if the device can determine the user it is interfacing with, then the device may, if needed, access the corpora for that user (**FIG. 4H**). It should be remembered that within the system **100** a “user” **110** is a notion used to form some kind of association or binding of one or more devices within the system. A user **110** typically corresponds to a person, however not all people who interact with a device or who attempt to interact with a device are necessarily users within the system **100**.

[00102] An exemplary flow process used to select corpora (**S409** in **FIG. 4L**) is shown in **FIG. 4M**. For this exemplary implementation it is assumed that the device includes at least some generic corpora for the various interface mechanisms. These generic corpora may be included with the device at time of manufacture or during subsequent provisioning or configuration. To reach this process (selecting corpora), the device has detected some kind of interaction (at **S406**, **FIG. 4I**). The device may have detected one or more of sound, external movement, touch, movement of the device, etc. This detection may use some or all of the devices sensors **138**, so that, *e.g.*, the devices microphone may detect sound, the devices camera(s) may detect external movement, the devices accelerometer(s) may detect movement of the device, the devices touch sensors may detect that the device is being touched by a person or another device, the device may detect interaction from a user, etc. Recall that at this point the device may have an indication of what possible type of interaction was detected (*e.g.*, using the bit vector (for the implementation of **FIG. 4J**) or a vector of score values (for the implementation of **FIG. 4K**)). The device may use this information to try to determine what the interaction is and which corpora to use.

[00103] Depending on what potential interaction(s) the device detected (at **S406**, **FIG. 4J**; **S406'**, **FIG. 4K**), the device may then determine (at **S426**, **FIG. 4M**) whether it can recognize a particular user or person. If the device does recognize a user or person then (at **S428**, **FIG. 4M**) the device selects the

corresponding corpora for that user or person. On the other hand, if the device does not recognize the potential interaction as corresponding to any user or person known to the device (at **S426, FIG. 4M**), then the device selects the device's generic corpora (at **S430, FIG. 4M**).

[00104] Depending on the type of potential interaction detected by the device (at **S424**), the device will use the appropriate interface mechanism(s) (at **S426**) in order to try to recognize a person or user known to the device. A person or user is considered to be known to a device if the device has at least one corpus stored for that person or user. A person may be known to a device based on previous interactions with the device, because the person is the user associated with the device, or because the person is associated with (*e.g.*, a friend of) the user associated with the device and has been given permission to access the device in certain ways. In preferred implementations, the user associated with the device should always be known to the device.

[00105] Thus, for example, as shown in **FIG. 4I**, having detected that some interaction is taking place (at **S424**), the device may use gesture detection and recognition mechanisms **168, 170** (at **S432**) to try to determine whether the gestures detected correspond to those of a known user/person. If the device does recognize the gestures (at **S432**) as those of a known user/person, then the device selects the corresponding corpora (at **S434**). The device may use the mapping shown in **FIG. 4E** to determine corpora based on recognized gestures.

[00106] The device may also (or instead) use the face/gaze detection and recognition mechanisms **184, 186**, in conjunction with sensors **138** (*e.g.*, sensors that are camera(s)), (at **S436**) to try to determine whether the interaction is that of a user/person known whose face is known to the device or to the system. If the device determines that a nearby or visible face is known to the device (*i.e.*, corresponds to a user / person known to the device), then the device may select corresponding corpora based on that face (at **S438**). The device may use the mapping shown in **FIG. 4D** to determine corpora based on a recognized face.

[00107] The device may also (or instead) use detected sound (using sensors **138** that are microphone(s)) and the voice/speech detection and recognition mechanisms **176, 178** to try (at **S440**) to determine whether the detected sound is speech, and, if so, whether it corresponds to speech of a person/user known to the device. If the detected sound is determined (at **S440**) to be speech of a person/user known to the device, then the device may select appropriate corpora based on that speech (at **S442**). The device may use the mapping shown in **FIG. 4F** to determine corpora based on recognized speech.

[00108] The device may also (or instead) recognize a person / user using some other interface mechanism(s) **190** (at **S444**), in which case the device may select the appropriate corresponding corpora for that person / user based (at **S446**), *e.g.*, using the mapping shown in **FIG. 4G**.

[00109] The device may also (or instead) recognize that the interaction is with a known user (at **S448**), in which case the device may select the appropriate corresponding corpora for that user based (at **S450**), *e.g.*, using the mapping shown in **FIG. 4H**.

[00110] Although **FIG. 4M** shows various recognition attempts (at **S426**), it should be appreciated that not all of these steps are performed (or even be available) in every device or type of device. Furthermore, it should be appreciated that in some devices or types of devices, the steps may be performed in parallel, in series, or in some combination thereof. In some cases, multiple tests may be used to recognize and/or confirm recognition (in **S426**) before corpora is selected. Thus, for example, a particular device or type of device may first use face recognition and then, only if that fails, use some other technique. As another example, a particular device or type of device may simultaneously try to recognize faces (at **S412, FIG. 4J; S412', FIG. 4K**) and users (at **S418, FIG. 4J; S418', FIG. 4K**), and then, optionally, follow up with other recognition approaches.

[00111] Those of skill in the art will realize and understand, upon reading this description, that the system is not limited by the manner or order (including

concurrent, series, or combinations thereof) in which the various user/person recognition steps are carried out, nor is the system in any way limited by the mechanisms available or used to try to recognize a user/person.

[00112] If more than one recognition mechanism is used (in **S426**), then the device needs to be able to deal with conflicts. A conflict may occur, *e.g.*, when one mechanism identifies one user/person and another recognition mechanism identifies a different user/person. This may occur when the device is unable to determine enough information about the person/user interacting with the device, and may be because the device does not have enough information about that person and/or because there is more than one person potentially interacting with the device. For example, a new device may not yet have enough information about its potential users to make accurate recognition decisions, so that the different mechanisms may make different recognition decisions. A device may also be confused because there is more than one person in the vicinity, so that its sensors are picking up details from different people. For example, if there are multiple people in the device's vicinity, its microphone(s) may be picking up the voice of one person while its camera(s) are picking up the face of another person.

[00113] It is also possible that a particular sensor will recognize more than one person. For example, the camera(s) in the device may recognize more than one face or they may find multiple faces.

[00114] A device **102** preferably has at least one conflict resolution strategy. A conflict resolution strategy may be adaptive, with the device learning based on prior interactions and recognition decisions. In some cases, the device may use a weighted function of the recognition decisions, for example, giving most weight to the user recognition (at **S448**), less weight to face recognition (at **S436**), still less weight to voice/speech recognition (at **S440**), and so on. Those of skill in the art will realize and understand, upon reading this description, that different and/or other functions may be used, and that different and/or other conflict resolution

strategies may be provided. For example, a conflict resolution strategy may be dynamic, changing over time (*e.g.*, based on learning).

[00115] In addition to conflict resolution strategies, a device **102** may also include various optimizations to improve recognition of whether a person is trying to interact with the device, and, if so, which person. One exemplary optimization is to use gaze detection (using the face/gaze mechanism(s) **182**) to determine whether someone is actually looking at the device. Gaze detection may be used, *e.g.*, to select a corpus, as a trigger for other recognition and interactions, and/or as part of a conflict resolution strategy. For example, if the device detects sound (using its microphone(s)), the device may not know whether that sound corresponds to commands or queries for the device. If the device can also detect that someone is looking at the device then that gaze, alone or along with the detected sound, can be used to trigger speech recognition. It should be appreciated that gaze detection does not require or depend of face recognition.

[00116] As an additional optimization, in some implementations, the device may use face movement detection mechanism(s) **187** to determine whether a face that it has detected (using the face/gaze detection mechanism(s) **184**) is moving in a way that might be used to confirm (or reject) that detection. For example, if the system finds multiple faces (using face/gaze detection mechanism(s) **184**), and also detects speech (using voice/speech mechanism(s) **174**), then any faces that show movement corresponding to speech (*e.g.*, where the mouth is moving) are preferred candidates for selection. It should be appreciated that the face movement detection mechanism(s) **187** need not interpret speech, they need only detect mouth or jaw movement or some other kind of movement that may correspond to speech.

[00117] In some implementations face movement detection mechanism(s) **187** may be used to detect movement of other parts of a face, *e.g.*, eyebrows, eyes, jaw, mouth, etc., for use by the gesture mechanism(s) **166**. Face movements may

themselves be gestures recognized by a device, or they may be used to confirm other detected gestures.

[00118] Similarly, gestures (as determined by gesture mechanism(s) **166**) may be used alone or to confirm other detected information (*e.g.*, faces, voices, movements, etc.)

[00119] Those of skill in the art will realize and understand, upon reading this description, that, in general, each interface mechanism **162** may be used alone or in combination with other interface mechanisms. When used together, the various interface mechanisms may be used to confirm each other and/or as triggers for each other.

[00120] It should be appreciated that not all devices or kinds of devices need have all of the interface mechanisms, and that these organizations may not all be used on all devices or kinds of devices. In addition, it should be appreciated that some devices may use different and/or other techniques to determine which corpora to use.

[00121] Although the interaction detection (*e.g.*, **S406** in **FIG. 4J**, **S406'** in **FIG. 4K**) and corpus/corpora selection (**FIG. 4L**) are shown as separate processes, those of skill in the art will realize and understand, upon reading this description, that some of the determinations made the interaction detection may be used in subsequent processing, *e.g.*, in determining whether a person / user is recognized (in **S426**, **FIG. 4M**). Thus, in some cases, in the process of detecting a possible interaction (**S406** in **FIG. 4J**, **S406'** in **FIG. 4K**), the device may have determined sufficient information about a person / user interacting with the device to skip (or simplify) the process of user / person recognition (**S426**).

[00122] With reference again to **FIG. 4L**, having determined (at **S409**) which corpora to use for the actual interaction, the device may proceed (at **S411**) to determine the actual interaction, using the corpora determined (at **S409**) and the various interface mechanisms **162**. As will be apparent to those of skill in the art, upon reading this description, the actual interaction may be determined (in full or

in part) by the processes of person recognition (**S426**) and corpus selection (**S428**). If these two processes do not result in a determination of the actual interaction, then the device proceeds to determine the actual interaction (at **S411**).

[00123] An alternate implementation / approach to processing the actual implementation assumes that the corpora to be used are those of the device's user **110** unless some other authorized user is specifically recognized or identified. In this exemplary implementation, the scores produced by the interaction detection mechanisms are used to control which recognition mechanisms are used. With reference to the flowchart in **FIG. 4N**, the device determines (at **S409''**) which corpora to use for the actual interaction. If the detect user process (*e.g.*, **S418'** in **FIG. 4K**) produced a score greater than some threshold value (denoted T_{user} in **FIG. 4N**), then the device tries to recognize the user (at **S418''**) and, if successful, then selects corpora for the recognized user (at **S450''**). If the device fails to recognize a user (preferably an authorized user), or if score (R_7) produced by the detect user process/mechanism does not exceed the threshold value (T_{user}), then the device selects corpora associated with the device's user (which may be generic corpora). With corpora selected (at **S409''**), the device proceeds to determine the actual interaction (at **S411''**). In the exemplary implementation of **FIG. 4N**, the device may use scores produced by various interface detection mechanisms to determine whether or not to invoke corresponding recognition mechanisms. For example, as shown in **FIG. 4N**, the device may invoke the gesture recognition mechanism(s) **170** (at **S410''**) if the score (R_1) produced (*e.g.*, at **S410'** in **FIG. 4K**) by the gesture detection mechanism(s) **168** exceeds a threshold value (denoted $T_{Gesture}$ in **FIG. 4N**). Similarly, the device may invoke the face recognition mechanism(s) **184** (at **S412''**) if the score (R_2) produced (*e.g.*, at **S412'** in **FIG. 4K**) by the face detection mechanism(s) **168** exceeds a threshold value (denoted T_{Face} in **FIG. 4N**). And similarly, the device may invoke the voice/speech recognition mechanism(s) **178** (at **S414''**) if the score (R_3) produced (*e.g.*, at **S414'** in **FIG. 4K**) by the voice/speech detection mechanism(s) **176**

exceeds a threshold value (denoted T_{Speech} in **FIG. 4N**); and the device may invoke the other interface mechanism(s) **190** (at **S416''**) if the score (R_4) produced (*e.g.*, at **S416'** in **FIG. 4K**) by the detect other interface mechanism(s) **190** exceeds a threshold value (denoted T_{Other} in **FIG. 4N**).

[00124] It should be understood and appreciated that the various mechanisms used to determine an interaction may proceed concurrently or in some predefined order (*e.g.*, based on the score or weighted score produced by the detection mechanisms). Those of skill in the art will realize and understand, upon reading this description, that the system is not limited in any way by the order in which various recognition mechanisms are invoked.

[00125] It should also be understood and appreciated that some of the recognition mechanisms may work together with others or may be integrated with others.

[00126] The threshold values described with reference to the exemplary implementations (*e.g.*, in **FIG. 4N**) may be static or dynamic, and the invocation of any particular recognition mechanism(s) may be based on different and/or other factors. In some implementations, some or all of the threshold values may be modified based on information the device learns from prior interactions. Those of skill in the art will realize and understand, upon reading this description, that the system is not limited in any way by the values or triggers used to invoke various recognition mechanisms.

[00127] As is apparent from the diagram in **FIG. 4N**, some of the mechanisms that were used to detect possible interactions (*e.g.*, gaze detection and face movement detection) may not be needed or used by the interaction recognition mechanisms. Thus, *e.g.*, once a gaze is detected, that information may be used to trigger interaction recognition, but it may no longer be needed by the actual recognition. It should be appreciated that some of the information used during the detection process (in addition to the scores produced by the various mechanisms) may be provided the recognition mechanisms. So, *e.g.*, the face

recognition mechanism (invoked at **S412''** in **FIG. 4N**) may already have information from the gaze detection mechanism in order to know which face to try to recognize. Alternatively, the face and/or gesture recognition mechanisms may, themselves, invoke the gaze detection and/or face movement detection mechanisms.

Process Actual Interaction

[00128] Having determined the actual interaction (at **S411**), the device proceeds (at **S413**) to carry out the instructions or commands or queries associated with the actual interaction. Determining and/or carrying out the actual interaction (at **S411**, **S413**) may be done locally, on the device, or it may require (or benefit from) information from the backend **104** (*e.g.*, from the databases **114**), interaction via the backend with some other entity (*e.g.*, social networking service(s) **116**, or content provider(s) **118**, etc.), as well as processing or assistance from the added functionality entities **120**. Carrying out the actual interaction may also involve interacting with another device **102**. Interactions between the device and the backend correspond to arc #7 in **FIG. 1**; interactions with the backend database(s) **114** correspond to arc #8; interactions with the social networking service(s) **116**, or content provider(s) **118** correspond to arcs #9 and #10, respectively, and interactions with added functionality entities **120** correspond to arc #11. A device's interactions with other devices correspond to arc #1 in **FIG. 1**.

[00129] Carrying out the actual interaction (at **S413**) may require the device to use previously buffered information (**S404**, **FIG. 4I**). For example, if a device detects and buffers sound that may be a speech interaction, the device may not start the speech recognition until (or unless) the device also detects a person looking (gazing) at the device (*e.g.*, **S420**, **FIG. 4J**, **S420'**, **FIG. 4K**). The gaze thus acts as a trigger for subsequent speech recognition (*e.g.*, **S414''**, **FIG. 4N**). This approach allows a device to capture speech or other information (*e.g.*, gestures) that a person starts to give while not yet looking at the device, and

continues to provide while looking at the device. It should be appreciated that other interactions (*e.g.*, gestures, facial movements, tapping the device, etc.) may be used as triggers for subsequent recognition of interactions, including interactions that may have already begun. Since the system has buffered potential interactions, it is able to use the buffered information once one or more triggering events occur.

[00130] Those of skill in the art will realize and understand, upon reading this description, that any kind of information that can be detected and/or measured by the sensors **138** may be buffered and then used after one or more triggering events are detected.

Learning

[00131] As noted earlier, the various mechanisms used for detection and recognition of interactions may include learning features so that they may become more accurate over time and use. In some cases, as a mechanism learns about a particular user, that mechanism may update corresponding corpora for that user. For example, as a speech recognition mechanism learns the speech of a particular user, it may update the speech corpora associated with that particular user.

[00132] In addition to each individual mechanism learning from its interactions with users, the device itself may learn how a particular user interacts with the device and may optimize its handling based on that learning. For example, if a particular user always uses the same gesture (*e.g.*, pointing to the device) in combination with speech commands, then the device can learn that pattern. In that example, the pointing gesture may be given more weight as a trigger for speech recognition. As another example, if a particular user always uses a particular hand gesture along with face movements (*e.g.*, eyebrow raising) and certain words, then those features, in combination, can be given higher weight by the device.

[00133] While various operational mechanisms are shown in the diagram in **FIG. 4B** as separate mechanisms, these separations are only provided by way of example and to aid in this description and are not intended to limit the scope of this description in any way. Those of ordinary skill in the art will appreciate and understand, upon reading this description, that the separations in **FIG. 4B** are not required, and that some or all of the mechanisms may be combined into different and/or other functional units, including into a single functional operational mechanism. For example, the gesture detection and recognition mechanisms **168**, **170** may be part of a single gesture mechanism. Similarly, the voice/speech detection and recognition mechanisms **176**, **178** may be part of a single voice/speech mechanism. And similarly, the face/gaze detection and recognition mechanisms **184**, **186** may be part of a single face/gaze detection and recognition mechanism. Similarly, the logical depiction of the operational mechanisms' storage **155** in **FIG. 4C** is given to aid in this description and is not intended to limit the scope of this description in any way. Those of ordinary skill in the art will appreciate and understand, upon reading this description, that different and/or other storage organizations are possible and are contemplated herein.

DEVICE STATES, PROVISIONING AND CONFIGURING DEVICES

[00134] As used herein, with reference to a device **102** within the framework **100**, the term "provisioning" refers to the process of installing (or updating) the various system mechanisms used by the device. Provisioning may include installing and/or updating firmware or software on a device. As used herein, with reference to a device **102** within the framework **100**, the term "configuring" refers to the process of establishing or setting operational options and/or parameters for the various mechanisms used by the device. For example, configuring a device may include setting passwords for the device, establishing network parameters for the device, and so on.

[00135] With reference now to **FIG. 6A**, a device **102** may be considered to be in various provisioning and configuration states. A device **102** is in a *pre-provisioned* state when it has not yet been provisioned with system mechanisms. A device **102** is in a *generic provisioned* state when it has been provisioned with system mechanisms but is not yet associated with a user **110**. A device **102** is in an *unconfigured state* when it has not yet been configured. A device **102** is in a *generic configured* state when it has been configured for use on the system **100**, but not for a particular user **110**. A device **102** is said to be in a *user-provisioned* state when it has been provisioned for a particular user **110**. A device **102** is said to be in a *user-configured* state when it has been configured for a particular user **110**. Various possible state transitions (denoted **T1**, **T2** ... **T6**) for a device **102** are shown in **FIG. 6A**. These states and transitions are merely descriptive and exemplary, and are used here only to aid in this description. It should be appreciated that this system is not limited by these various states, what they are named or the state transitions described here. It should also be appreciated that these states and transitions may be independent of the device specific functionality.

[00136] There are two aspects to provisioning and configuring a device **102** within a framework **100**. The first aspect is essentially independent of any user and conforms the device **102** to then-current and/or localized versions of all system mechanisms. This aspect corresponds to the state transition **T1** for a device from *pre-provisioned* to *generic provisioned* and, possibly, transition **T2** from *unconfigured* to *generic configured*.

[00137] The second aspect provisioning and configuring a device **102** within a framework **100** conforms the device to the settings/requirements of a particular user **110** currently associated with the device. For a previously generically provisioned or generically configured device, this aspect corresponds to the state transition **T3** from *generic provisioned* to *user provisioned* and, possibly, transition **T4** from *generic configured* to *user configured*. A device's association

with a particular user may change, possibly only temporarily or under limited conditions (*e.g.*, location, duration, use, etc.). Thus, this second aspect of provisioning and configuring may also correspond to the state transition **T5** where a user-provisioned device is provisioned for a different user, and/or to the state transition **T6** where a user-configured device is configured for a different user.

[00138] The process of provisioning a device thus may include installing the latest versions of all mechanisms (*e.g.*, software, firmware, etc.) on the device. When a device **102** is manufactured, the manufacturer **108** typically installs a version of the system mechanisms **134** on the device (along with versions of other mechanisms such as those for the device-specific functionality). However, those versions are often out of date even at the time of manufacture. Accordingly, in a provisioning process, the bootstrap / provisioning mechanism **152** may update some or all mechanisms (*e.g.*, software, firmware, etc.), either at first power on, or even before the device is shipped or sold. In some cases, a device **102** may be shipped in a low power mode with a timer set to power up after a fixed period (*e.g.*, 36 hours) and to then to run the bootstrap / provisioning mechanism **152** in order to try to update all mechanisms. The device **102**, when powered on, will search for known wireless (Wi-Fi) networks or other ways to connect to the Internet. Preferably the device is pre-configured (as part of its generic configuration) with information about known and trusted networks, and the device uses those networks to connect, via the network **101**) to a known and trusted source location from which updates can be obtained. Once a network connection is found and established, the device **102** can begin updating itself from the trusted source. In this manner, when a device reaches a user (*e.g.*, after it is sold to the user); the device should have the most current version (or a recent version) of all mechanisms and be fully (or substantially fully) provisioned for that user.

[00139] For those mechanisms that may require corpora (*e.g.*, speech recognition, gesture recognition, etc.), initially generic corpora are installed on the device.

[00140] Preferably each device **102** is configured (*e.g.*, at time of manufacture) with sufficient information to allow the device to establish secure contact with the backend **104**. To this end, an initial generic configuration for the device may include names and passwords for various wireless network and/or generic credentials to support secure wireless (*e.g.*, cellular, WiFi, Bluetooth, or BLE) communication between the device and the backend **104**. The term secure is used here to refer to communications channels that can be trusted and that are preferably not spoofed. The degree of security is a function of the type of device, and different types of devices may require different degrees of security for device-to-device and device-to-backend communications.

[00141] The processing by a manufacturer and until the device is associated with any user corresponds to the state transitions **T1** (and possibly **T2**) in **FIG. 6A**.

[00142] In order to operate in such a way as to make use of the system **100**, each device **102** is preferably associated with a user **110**. A device may be associated with no user (*e.g.*, when it is first manufactured), but a device may not be associated with more than one owner (and preferably not with more than one user at a time). As explained herein, a device **102** may be used by more than one person or user **110**, but, within the system **100**, the device is only associated with a single user.

[00143] In some cases, a device may be associated with a user as part of a provisioning step of the manufacturing and/or provisioning processes (*e.g.*, if a user orders or purchases the device and provides the user's identification (User Identity) prior to manufacture of the device **102**).

[00144] The first provisioning process (described above) may take place without the device being associated with a user. A second level of provisioning and configuration preferably takes place once the device **102** is associated with a user **110**. In particular, once a device **102** is associated with a user, that device **102** can obtain configuration information (*e.g.*, wireless network information such

as network IDs and passwords) from that user. When a device is configured for a particular user, the device may obtain information from the user database **130** about that user. For example, the device **102** may obtain user profile information, user local corpora, and/or configuration information about that user. This information may be stored by the device for use, *e.g.*, by the operational mechanisms **155**. For example, the device **102** may obtain user local corpora from the user database **130** and store those corpora in the corresponding appropriate interface mechanisms' storage **163** on the device. In this manner, if a user has already used a particular device or kind of device, a newly acquired device may not have to be re-trained to detect and recognize various interactions with that user.

[00145] Some information from the user database **130** (*e.g.*, a user's list of devices – a list of device IDs of devices owned by that user, and the user's list of friends – a list of user IDs of friends of the user) may be encoded in a one-way encoding (*e.g.*, using a cryptographic hash such as MD5) on the device. In this manner the device IDs and user IDs are not exposed, but (as will be explained later) information in the lists may be used (*e.g.*, to evaluate possible relationships between two devices).

[00146] The device and user databases are preferably updated (preferably in near real time) to reflect the provisioning and configuration states of each device, so that preferably the device database provides a current view of the state and configuration of each device in the system. It should be appreciated that devices that are not connected or that cannot connect to a network **101** or use some other method (*e.g.*, a cellular system), may defer providing information to the backend. However, preferably each device updates the backend with its state information while connected and when reconnecting to a network **101**. Similarly, when a device **102** connects to the backend **104** in some manner, that device preferably obtains updates from the backend, including any updates relating to the device's user.

[00147] In particular, each device **102** should regularly update the corpora in the device database **128** to reflect the current corpora on the device. These updates may take place on a regular scheduled basis (*e.g.*, once a day) or whenever the device can connect to the databases (via the backend **104**) and determine that the device database needs updating.

[00148] In addition, the corpora information in the user database **130** should reflect the current state of the corpora on each user's device(s). Thus, the user database **130** should be updated (*e.g.*, in the same manner as the device database) to reflect the latest corpora for that user. Recall that the user database **130** may include user local corpora and user extended corpora. The user local corpora correspond to the corpora on the user's devices. The user extended corpora correspond to corpora used by the backend or other external systems (*e.g.*, added functionality **120**) to process user interactions. For example, the user local corpora may include limited speech recognition corpora that a device **102** can use, whereas the user extended corpora may include extended speech recognition corpora for that user that can be used by, *e.g.*, the added functionality **120**.

[00149] Corpora for each user may be organized or stored on the user database **130** based on the kind or capabilities of the user's devices. This allows the system to support multiple kinds of devices. When a user updates the user database **130** with corpora from a particular device, the database preferably stores those corpora based on the kind and abilities of the device.

[00150] Preferably the device database **128** and the user database **130** maintain prior versions of corpora.

[00151] In some cases a user may have multiple devices of the same type (*e.g.*, multiple speakers). It is preferable for the corpora on each of the user's devices (especially devices of the same type) to have the most current version of that user's corpora for that type of device. Accordingly, each device should routinely contact the backend to determine whether it has the latest version of the

corpora. If not, the device may obtain the latest version of the corpora from the user database **130**.

[00152] The provisioning and configuration of the device, once first associated with a user, corresponds to the state transitions **T3** (and possibly **T4**) in **FIG. 6A**.

DEVICE MANUFACTURE & PROVISIONING

[00153] **FIGS. 6B – 6I** show examples of manufacture of devices **102**. Device manufacturer **108** is preferably authorized by the system **100** to make system-enabled devices **102**.

[00154] In the example in **FIG. 6B**, each device has a unique serial number provided by the manufacturer, and each device has a Device ID that is a function of the device's unique serial number. In the example in **FIG. 6C**, as will be described below, each device has a unique serial number provided by the *M* system. It should be appreciated that the unique serial number used by the *M* system may differ from other serial numbers used by the device manufacturer.

[00155] With reference to the diagrams in **FIGS. 6B** and **6C**, in one exemplary embodiment a device manufacturer **108** provides (at **S601**) the device serial numbers (either individually or in batch) to a Device Certificate Generator **172** (which could be part of the backend **104**). The Device Certificate Generator **172** uses device information provided by the manufacturer (*e.g.*, the serial number) to create (at **S602**) a unique Device ID for that device (*i.e.*, for the device associated with or to be associated with the serial number). That unique Device ID is sent to a device CA **124** (at **S603**) which puts it into a certificate signed by the Device CA **124**. The Device CA **124** then sends the signed certificate back to the manufacturer (at **S604, S605**) to be put into the device. The signed certificate with the unique Device ID is stored in the device's certificates **150**.

[00156] Those of skill in the art will realize and understand, upon reading this description, that different and/or other information (other than or in addition to

that provided by the manufacturer **108** at **S601**) may be used by the device certificate generator (at **S602**) to generate device certificates.

[00157] Preferably the information provided by the manufacturer (at **S602**) also includes information about the device's capabilities and/or components.

[00158] Once the device certificate has been generated and signed, information about that certificate (including possibly a copy of the certificate) is sent to the backend **112** (at **S606**). The device ID associated with the certificate may be used by the devices database **128** as a key or index into the database. Since the backend **104** provided information used to generate the device IDs, information about the device (*e.g.*, its capabilities) may already be known by the backend, and so these can be associated with the device ID in the devices database **128**.

[00159] Preferably information in the certificate is encrypted so that it may only be read with an appropriate decryption key by the device.

[00160] It should be appreciated that once a device certificate is generated by the backend **104** (*e.g.*, via the device certificate generator **172**), that certificate can only be associated with a single device (with the unique device ID that was associated with the certificate). If, for some reason, the device becomes inactive or lost (or is never actually manufactured), that device ID and certificate should not be reused.

[00161] The interactions between the manufacturer **108** and the backend **104** (via the device certificate generator **172**) (at **S601**, **S605** and **S606**) correspond to the interaction #13 in **FIG. 1**. The interactions between the device CA **124** and the backend **104** (via the device certificate generator **172**) correspond to the interaction #6 in **FIG. 1**.

[00162] In some cases, the device certificate generator **172** may provide certificates to the manufacturer in bulk, based on a list of serial numbers from the manufacturer.

[00163] In an alternate and presently preferred exemplary embodiment the system generates (or obtains) blocks of serial numbers and associated certificates, and provides those serial numbers and certificates in blocks to the manufacturers. Thus, with reference to the diagrams in **FIGS. 6D** and **6E**, the backend **104** generates serial numbers and provides them (at **S651**) to a device certificate generator **172'**. As in the previous example, the device certificate generator **172'** may be part of and/or co-located with the backend **104**.

[00164] The Device Certificate Generator **172'** uses information provided by the backend (*e.g.*, the serial number) to create a unique Device ID for the device associated with or to be associated with the serial number. That unique Device ID is sent to a device CA **124** (at **S653**) which puts it into a certificate signed by the Device CA **124**. The Device CA **124** then sends the signed certificate back to the Device Certificate Generator **172'** (at **S654**) which sends it back to the backend (at **S655**). The signed certificate with the unique Device ID is to be stored in a device's certificates **150**.

[00165] The information in the certificates is preferably encrypted so that it may only be read by authorized devices. The Device Certificate Generator **172'** may encrypt the serial number and unique Device ID before providing the certificate to the Device CA **124**.

[00166] Those of ordinary skill in the art will realize and appreciate, upon reading this description, that since the backend **104** is providing the serial numbers, those numbers may be used, at least in part, to form the device IDs.

[00167] The backend **104** provides the device manufacturer **108** with serial numbers and corresponding certificates (at **S656**), preferably a block of such numbers and certificates. The manufacturer may use some or all of those certificates in devices, and provides the backend **108** (at **S657**) with a list of the serial numbers/certificates it uses.

[00168] It should be appreciated that since the backend has a list of all serial numbers and copies of all certificates, it can track numbers used and can thereby

verify information provided by manufacturers. For example, if a manufacturer fails to report the use of a particular serial number in a particular device, the backend will detect use of that serial number when that particular device connects to the system. Similarly, if a manufacturer uses the same serial number/certificate combination in multiple devices, the backend will detect the duplication.

[00169] The flowchart in **FIG. 6E**, depicts aspects of the manufacturing embodiment described in **FIG. 6D**. The dashed vertical line in **FIG. 6E** is provided to show which aspects are carried out by the system and which are carried out by the manufacturer. The system generates serial numbers and certificates (at **S661**) and provides these to the backend (at **S662**), preferably in blocks of numbers/certificates. The manufacturer gets the block of numbers/certificates (at **S663**) and uses some of them in devices (at **S664**). The manufacturer reports which serial numbers/certificates it has used (at **S665**), and the backend gets the lists of used serial numbers/certificates (at **S666**).

[00170] In some implementations the manufacturer may report additional information (at **S665**) such as device type, capability, etc.

Provisioning, Configuring, and Associating with a User

[00171] Exemplary initial provisioning and configuring of a device **102** (corresponding to state changes **T1** and **T2**, respectively, in **FIG. 6A**) are shown in the flow diagram in **FIG. 6F**. At this stage it is assumed that the device has a device ID associated therewith (as described above). The manufacturer provides the required mechanisms and other components (*e.g.*, system mechanisms **134**, sensors **138**, communications **142**) as needed for the device **102** (at **S608**). These mechanisms and other components may be provided in the form of a single board or kit with connections for the device specific components, or they may be fully or partially integrated with the device specific components. When provided as a kit or board, not all components may be active or activated. For example, certain components may be deactivated based on the location of the device, the kind of

device, or because users will be charged additional amounts for their subsequent activation.

[00172] With the mechanisms installed, the device is provided with an initial configuration (at **S610**). The initial configuration may include generic corpora, as needed, for various mechanisms. The initial configuration may also include information supporting the device's connection to the backend (*e.g.*, information about known or trusted networks, etc.).

[00173] Preferably the device maintains a list or manifest of all mechanisms and their current state of configuration, including version information and corpora details. Once a device is initially provisioned and configured (*e.g.*, during manufacture), its current manifest is provided to the backend for storage in the database entry associated with the device (at **S612**).

[00174] As noted above, a provisioned/configured device may update its mechanisms and configuration prior to being associated with a user (at **S614**). For example, a device may use a known/trusted Wi-Fi connection to update its firmware during shipping. Any such updates should be reflected in the device's manifest, both on the device and in the device database (at **S616**). The device database may maintain a history of updates to the device. Some mechanisms (*e.g.*, the voice/speech recognition, gesture recognition, etc.) used by a device may be provided by third parties. In those cases the mechanism may have firmware and/or corpora included therewith, and updates to those components may have to be obtained from the third parties.

User Registration

[00175] Recall that each user **110** must have at least one unique User Identity (ID) within the system **100**. Preferably each user **110** obtains their User ID by registering with the system **100**. User registration may take place via an offline process, via a web interface to the backend, or via a device **102**. A user may register prior to having any device(s) **102**. As part of a user's registration (as

explained above), each user has a user ID that must be unique within the system **100**. Once a user registers, an entry for that user is made in the user database **130**, preferably keyed or indexed primarily on that user's user ID. The user's database entry (corresponding to their User ID) is populated during the registration process to include information provided by the user (*e.g.*, via a form or questionnaire) and/or information that the system can obtain or deduce based on information the user provides (directly or indirectly). For example, if the user uses a social network ID or registers via social network, then information about the user from that social network may be included in the database.

[00176] Although the user database **130** is preferably indexed using the User ID, those of ordinary skill in the art will appreciate and understand, upon reading this description, that different and/or other keys may be used to access data in the user database.

[00177] As noted, multiple devices **102** may be associated with each user **110**. While the devices preferably have no hierarchical status, it is useful to describe the process of configuring a first device that a user obtains.

[00178] As shown in **FIG. 2**, each user preferably has at least one user device **174** (*e.g.*, a smart phone or the like) that has that user's user ID and associated user certificate **175** stored thereon. The user device **174** may also have one or more system mechanisms **178** (*e.g.*, in the form of an application (app) running on the user device **174**). The system mechanism / user app **178** provides the user with a way to configure the user device **174** within the system **100** as well as to configure other aspects of the user device. In particular, the user device **174** may provide the user with a way to set Wi-Fi and other network information (such as identification data, service set identifiers (SSIDs) and passwords for local Wi-Fi networks). This information may be stored as configuration information **180** on the user device **174** (**FIG. 2**) and may also be stored as configuration information associated with the user in the user database **130** (**FIG. 3A**). Those of skill in the art will realize and understand, upon reading this description, that this

configuration information is sensitive and should be maintained in secrecy (*e.g.*, via encryption).

[00179] It should be appreciated that the user device **174** may be an instance of a device **102**.

[00180] Exemplary user registration is described with reference to **FIG. 6G**. First, (at **S620**) a user obtains a user ID from the system (*e.g.*, as described above). The system then creates a user database entry for that user (preferably keyed or indexed on the User ID) (at **S622**). The system then populates that database fields for that user (at **S624**).

Associating a device with a user

[00181] Each device **102** needs to be associated with a user **110** in order for the device to operate fully within the system **100**. It should be appreciated that a device that is not associated with any user may still be able to provide some or all of its device-specific functionality.

[00182] When a user acquires a device **102** (*e.g.*, a new device), that device needs to be associated with that user. Exemplary association of a device with a user is described with reference to **FIG. 6H**. First (at **S626**), the device is associated with the user in the device and user databases **128**, **130**. In the device database **128**, the Owner of the device is set to the user's User ID. In the user database **130**, the device's unique device ID is added to the devices associated with the user's User ID. Information about the device may be added to other fields in the user database **130**, and the history may be updated to reflect the user's association with this device.

[00183] In some embodiments, a device may become associated with a user by having the user touch (or tap) the device with another device of the user's. In preferred implementations, when a particular device is not yet owned by any user, the first time that particular device is tapped by another device, the particular device becomes associated with the user of the other device. The particular device

may obtain the user's configuration information (**S628** in **FIG. 6H**) from the backend database(s) and/or from the user's device that tapped it. When a particular device is already associated with a user, then a subsequent touch (or tap) from another device may be used to provide temporary permissions to the particular device, *e.g.*, to allow the devices to be combined in some way or to allow the particular device to inherit some configuration information (preferably temporarily) from the device that touched it. Once devices have been paired (*e.g.*, by touch), they may then share information via a Bluetooth, BLE, or WiFi signal or the like. It should be appreciated that sharing information may be in multiple forms, for example, metadata may be shared via Bluetooth while content (*e.g.*, music or video content) may be shared over WiFi. Furthermore, having been previously paired, two devices may detect each other's presence (*e.g.*, by a Bluetooth signal within their range) to continue or re-establish collaboration.

[00184] Preferably each device maintains some information about the user(s) associated with that device (*e.g.*, system mechanism(s) / data **134**, **FIG. 4A**). The information on the device may be updated (at **S628**) to reflect information about the user. This information may include at least some of the information associated with the user's User ID in the user database **130**. For example, the information stored on the device **102** may include the user's user ID, information from the user's profile, information about other devices associated with the user, information about the user's friends (*e.g.*, by their respective user IDs), user certificates, user corpora, and user configuration information. It should be appreciated that at least some of the user's information on the device should preferably be stored and maintained in a secure manner, *e.g.*, in encrypted form.

[00185] Once the device has been updated (at **S628**) to include the user's information, the device database **128** may need to be updated (at **S630**) to reflect changes made to the device. For example, if the user's local corpora were stored on the device (in place of whatever corpora were already there, *e.g.*, the generic

corpora), then the device database **128** should be updated to reflect this information (in Device Corpora and in Device History).

[00186] The exemplary process described here to associate a device with a particular user **110** in the system **100** corresponds to the state changes **T3**, from *generic provisioned* to *user provisioned*, and **T4**, from *generic configured* to *user configured* in **FIG. 6A**.

[00187] If a user acquires a device that has been previously used with the system (and is therefore associated with another user), the device may first need to be restored to a state in which it has no user information associated therewith. The device may be restored by any technique of restoring it to its factory settings (or pre-user settings). This kind of reset corresponds to the state changes **T3'**, from *user provisioned* back to *generic provisioned*, and **T4'**, from *user configured* back to *generic configured* in **FIG. 6A**.

Associating Configuration information with a device

[00188] As noted, a device may obtain user configuration information from the user database **130** (or from another device) when the device first becomes associated with the user (**S628** in **FIG. 6H**). However, as the user information may change (*e.g.*, the user gets a new friend within the system or the user has updated or new wireless network information, or the user has new cellular communication information, etc.), that information should propagate from the database to the device (and vice versa). Thus, as shown in **FIG. 6H**, the providing of the user's configuration information to the device (at **S628**), and updating the user and device databases **128**, **130** (at **S630**), is repeated as needed (when possible). For example, a device may check the databases when it can (*e.g.*, when it is able to connect to the backend **104**) to determine if it has the latest version of the user's information. In some cases, the backend may try to push notifications to the device to alert it about updates to the user information. Since a device may change (or cause a change) to a user's information (*e.g.*, a device may have

updated corpora or network configuration information), these changes also need to propagate back to the device and user databases. This process is reflected in **FIG. 6I** which shows repeatedly (as needed or when possible) providing information from the device (at **S632**) to the backend **104**, and then updating the user and device databases accordingly (at **S634**). It should be appreciated that any update to the user database **130** for a particular user may require corresponding updates to be sent to that user's devices and corresponding updates to the device database **128**. For example, if a user has multiple devices, and the corpora and/or configuration information are changed on one of those devices, the corpora and/or configuration information should propagate to that user's other devices. In this manner, a user need not have to independently or separately train or configure all of their devices, and the each device can benefit from the training and configurations applied to that user's other devices.

[00189] It should be appreciated and understood that a user's devices may get out of synch with each other and / or with the information in the user and device databases **128**, **130**. This may happen, *e.g.*, when devices are unable to connect to the backend for some period of time. The system preferably applies a conflict resolution technique in order to synchronize devices and the databases. An exemplary conflict resolution approach may use time stamps to select the most current versions of configuration and corpora information. Another exemplary conflict resolution approach may always assume that the versions of configuration and corpora information in the user database are correct.

[00190] Preferably any conflict resolution technique can be performed without user intervention, although the user may be provided with an interface to the backend and/or to that user's devices (*e.g.*, via an app on a user device such as phone **174** or via a web interface) to allow the user to select specific configurations and/or corpora. In some implementations a user may be able to force (*e.g.*, push) updates to their devices.

Note on corpora

[00191] Corpora are used, *e.g.*, by the various interface mechanisms **162** in the devices **102**. For example, voice / speech recognition mechanism(s) **178** may use local speech corpora (stored on the device **102**). However, as is understood by those of skill in the art, voice / speech recognition may be affected by a number of factors, even for the same voice / speech recognition mechanism(s) **178**. For example, different quality or kinds of input sensor (*e.g.*, microphones) may result in different corpora, even for the same voice / speech recognition mechanism(s) **178**. To this end, if needed, corpora in the user database **128** may be organized based on hardware specifics of the users' devices. In these cases, when certain corpora of a particular user change (based, *e.g.*, on a learning process on a specific device associated with that particular user), those corpora are updated in the database, but are only propagated to comparable devices of that user (*i.e.*, to devices of that user having a comparable hardware and sensor configuration for the recognition mechanism(s) associated with those corpora).

[00192] For example, suppose a particular user **110** has multiple devices, some of which have a first hardware configuration for their system mechanism(s) **134** and/or sensors **138**, and others have a second hardware configuration for their system mechanism(s) **134** and/or sensors **138**. The devices with the first hardware configuration use a first set of corpora for their corresponding operational mechanisms, and the devices with the second hardware configuration use a second set of corpora (distinct from the first set of corpora) for their corresponding operational mechanisms. In this example, when a first device with the first hardware configuration updates its corpora (*e.g.*, a speech recognition corpus or a gesture recognition corpus), that update should be sent to the user database **130** and to the device database **128**, but it should only propagate to other devices of that particular user having the first hardware configuration.

[00193] Those of skill in the art will realize and understand, upon reading this description, that devices of different types (*i.e.*, having different underlying

device-specific functionality) may have the same hardware configurations for their system mechanisms. Similarly, devices of the same type (*e.g.*, speakers), may have different hardware configurations for their system mechanisms

[00194] The inventors realized that it would be useful to have a device learn various setup and configuration information with minimal user intervention and action. Thus, they realized, it is advantageous and preferable to have devices learn from each other. Accordingly, in some aspects, a device **102-A** may obtain configuration information from another device **102-B**. In some cases, a device may obtain information from another device by having the two devices touch each other. These interactions correspond to the device-to-device interactions #1 depicted in **FIG. 1**, and may be implemented, at least in part, by device-to-device mechanisms **156**. In other cases, a device may obtain information from another device by being instructed to obtain it by a user command. A user's device may also obtain configuration and other information from the user database.

Device Heartbeat and Interactions

[00195] Recall that each device **102** preferably includes heartbeat (HB) mechanism(s) **194** (**FIG. 4B**). The Heartbeat mechanism(s) **194** on a device **102** have two primary functions: (1) to generate heartbeat messages (heartbeats), and (2) to monitor for heartbeats from other devices.

[00196] Thus, heartbeat mechanism(s) **194** on a particular device **102** may be used to provide various signals to the system (*e.g.*, the backend **104**) and/or to other devices **102** about the state or existence or presence of the particular device **102**. A device's heartbeat (HB) mechanism(s) **194** may use the device's communications mechanisms **142** to broadcast the device's heartbeat (and associated information) via one or more of the device's mechanisms for local communication (*e.g.*, Bluetooth, including BLE, ZigBee, etc.), the device's mechanisms for Wi-Fi communication (*e.g.*, 802.11, etc.), the device's mechanisms for cellular communication (*e.g.*, modems or other devices using a

cellular telephone network, etc.); and the device's mechanisms for wired communication (*e.g.*, Ethernet or the like). Each heartbeat message may contain information that allows other components of the system **100** (*e.g.*, the backend **104**, other devices **102**) to recognize (and possibly confirm) that it is a heartbeat message, and information identifying the device so that other components of the system **100** can recognize (and possibly confirm the device identifying information).

[00197] Different heartbeat messages (with different formats and having different information, and at different frequencies) may be broadcast via the different communication mechanisms. For example, a heartbeat message intended for the backend **104** and sent via the network **101** or a cellular network may be sent out daily or when some historical information is to be provided. On the other hand, a heartbeat message intended for other devices and broadcast via the device's local communications mechanisms (*e.g.*, Bluetooth, BLE, or the like) or sent on a local network to which the device is connected may go out every minute (or at some other regular and short time interval).

[00198] A heartbeat signal should include some information about the device, preferably at least the device's Device ID. For example, as shown in **FIG. 7A**, a heartbeat signal **700** from a device includes an encoding of the corresponding device ID and, optionally, an encoding of the user ID for the owner of the device. The heartbeat signal may include additional information (shown by the dotted line in the drawing in **FIG. 7A**). The signal sent to the backend may include additional information such as, *e.g.*, the device's location, history, etc. A local heartbeat signal may include only the device ID. Information in a heartbeat signal is preferably protected (*e.g.*, via encryption). The device ID and user ID may also be encoded with a one-way encoding (*e.g.*, a cryptographic hash such as MD5) to prevent their exposure.

[00199] Each device **102** should also routinely (preferably continuously) monitor for heartbeats from other devices **102** that may be nearby or on the same

network (*e.g.*, using the device's mechanisms for local communication as well as the device's mechanisms for Wi-Fi and wired communication). Another device may be said to be nearby a particular device if the particular device can pick up the other devices heartbeat via the particular device's local communications mechanism(s). Other notions of nearness may be used and are contemplated herein.

[00200] **FIG. 7B** shows an example device A (**102-A**) broadcasting a heartbeat signal via the device's heartbeat (HB) mechanism(s) **194-A** using the device's communications mechanisms **142-A** (*e.g.*, a local communications mechanism). As shown in the drawing, a second device (device B – **102-B**), detects device-A's heartbeat signal via device-B's communications mechanism **142-B**. Although not shown in the drawing, it should be appreciated that device B is also broadcasting its heartbeat signal and that device A may detect device B's heartbeat signal. Furthermore, and also not shown in the drawing, each of the devices may also be sending other heartbeat signals via different communications mechanisms.

[00201] **FIG. 7C** shows exemplary processing by each device **102** (*e.g.*, by the heartbeat mechanism **194** of the device) to monitor for heartbeats from other devices (at **S702**). If a heartbeat from another device is detected (at **S704**), then that heartbeat is processed (at **S706**), otherwise the system continues to monitor for heartbeats (at **S702**). Once the detected heartbeat is processed (or once the device begins to process a detected heartbeat), the device continues to monitor for heartbeats from other devices.

Devices Cooperate/Collaborate/Join

[00202] Some devices **102** may operate alone and in combination with one or more other devices. It should be appreciated that devices do not have to be homogeneous or even of the same kind to operate together. For example, devices that are speakers may operate together (as will be described below). As another

example, a device that is a speaker may operate together with a device that is a video display.

[00203] Devices **102** that operate together in some way are said to cooperate or collaborate. For the remainder of this document the terms “cooperate”, “cooperation”, and “cooperating” refer to “cooperate and/or collaborate”, “cooperation and/or collaboration,” and “cooperating and/or collaborating,” respectively.

[00204] As used herein, devices **102** are said to be joined if they are combined and cooperate for the purpose of at least some aspects of their operation. Devices may be joined in various ways. In some cases, a device may join one or more other devices merely by being put in proximity to the other devices. In some cases, devices may be joined by specific instructions via a user interface. In some cases, devices may be joined by having one of them touch the other.

[00205] It should be appreciated that devices may cooperate without changing their ownership. That is, a device of one user may cooperate or be joined with the device of another user without either of the devices changing ownership (*i.e.*, without either device becoming associated with a different user in the system **100**).

Processing a heartbeat

[00206] A device's processing of a heartbeat detected from another device (**S704**) may depend on a number of factors, including, *e.g.*, at least some of the following:

- Whether the first device knows the second device (*e.g.*, from a prior interaction).
- Whether the first and second devices are commonly owned (by the same user).
- Whether the devices are owned by friends.

- Whether the devices can cooperate in some way (this may depend, at least in part, on the kind of device or on each device's specific functionality, the devices' proximity and/or what each device is already doing). For example, a smartphone device and a speaker device may cooperate to play music from the smartphone device on the speaker device if they are in the same room; or two speaker devices may cooperate to both play the same music that one of them is already playing. Two headphone devices may, on the other hand, not cooperate if they are both already playing (*i.e.*, rendering) sound.

[00207] The factors and examples given here are provided only by way of example, and different factors may be used to determine how to process detected heartbeats.

[00208] **FIG. 7D** shows exemplary processing (at **S706**) by a device (referred to here for the purposes of this discussion as device A) of the heartbeat of another device (referred to here as device B). First, device A (on which this process is running) determines the device ID of device B from the received heartbeat message (*e.g.*, heartbeat signal **700-A** in **FIG. 7B**). The device ID may be encoded in the signal in such a way that it can be extracted by other devices. In some cases a device's heartbeat may contain a cryptographic hash (*e.g.*, an MD5 hash) of the device's device ID. In these cases, other devices can use the hash of the device ID as the basis of their decisions and the device ID itself does not get exposed.

[00209] Having determined the other device's device ID (at **S708**) (or a hash thereof), device A then determines if it and device B are owned by the same user (*i.e.*, if they are co-owned) (at **S710**). Recall that each device stores and maintains information from the user database **130** for the user of that device. This information preferably includes a list of the user's devices, *e.g.*, by device ID. Device A can determine if the device ID in the heartbeat message matches a device ID in the list of that user's devices. If the device ID is hashed or one-way

encoded in some other manner, then the device A may store the list of co-owned devices using the same encoding.

[00210] If the devices are determined to be co-owned (at **S710**), then device A evaluates possible cooperation with device B (at **S712**). Possible cooperation may depend on a number of factors, as noted above. In addition to cooperation with respect to their underlying functionality (*e.g.*, as speakers, etc.), when co-owned devices find each other (*e.g.*, via a heartbeat), the devices may share and/or update configuration information, as needed.

[00211] Whether or not the devices actually cooperate (as determined in **S712**), device A preferably updates its history to reflect its encounter with device B (at **S714**). Device A may then advise the backend of that encounter (at **S716**). Note that preferably the processing of a detected heartbeat by a device takes place without the device having to contact the backend. Therefore advising the backend of the encounter need not take place until another routine connection is made with the backend. Note too that if either device updates its configuration as a result of the encounter then that update should eventually propagate back to the backend.

[00212] If the devices are not co-owned (as determined in **S710**), then device A tries to determine (at **S718**) if device B is owned by a friend of the owner of device A. Recall that the heartbeat message may contain an encoding of the user ID of the user of device B, and that each device stores information from the user database **130**, including a list of friends. The user ID in the heartbeat message can be compared to the user IDs in the list of friends to determine if there is a match. It should be appreciated that if the user ID in the heartbeat message is one-way encoded then the user IDs in the friends list should be similarly encoded.

[00213] If it is determined (in **S718**) that device B is owned by a friend of the owner of device A, then some cooperation between the devices may still occur. Accordingly, device A then evaluates possible cooperation between the devices (as cooperation between friends' devices) (at **S720**). This kind of cooperation may include the same kinds of cooperation as between co-owned devices (in **S712**),

however it may depend on permissions associated with the friend. Friends' devices may even share some configuration and corpora information, however this is preferably done on a limited and temporary basis.

[00214] As with cooperation between co-owned devices (at **S712**), whether or not friends' devices actually cooperate (as determined in **S720**), device A preferably updates its history to reflect its encounter with device B (at **S714**), and again, device A may then advise the backend of that encounter (at **S716**).

[00215] If it is determined (at **S718**) that the devices are not owned by friends, then device A determines (at **S722**) whether it has encountered device B before. Device A may then record information about device B (at **S724**) and then proceeds to update its history (at **S714**) and, eventually, to advise the backend of the encounter (at **S716**).

[00216] As will be appreciated, cooperation between devices (co-owned or devices of friends) may require additional communication between the devices. For example, co-owned devices may need to communicate with each other in order to synchronize their configuration information. Each device **102** therefore preferably has at least one mechanism or process running that is listening for communications from other devices via the various communications mechanisms. Two devices that encounter each other (*e.g.*, via one or both of their heartbeats) may then interact further, as needed. Communication between two devices (*e.g.*, via a local wireless mechanism or a local wired network, etc.) preferably uses secure communication over whatever channel is used. In some cases two devices may first encounter each other via a heartbeat on one communications mechanism (*e.g.*, Bluetooth or BLE) and then have subsequent communication using a different communication mechanism (*e.g.*, Wi-Fi).

[00217] Exemplary processing for potential cooperation between co-owned devices (at **S712**) is shown in **FIG. 7E**. In this example, the device A (the one that detected the heartbeat of device B) first tries (at **S726**) to contact and establish a connection with device B. Once connected, the devices may (at **S728**) update /

synchronize their configuration information (if needed). (The update / synchronization (at **S728**) corresponds to state transition **T6** in **FIG. 6A**.) the device also determines information about itself and the other device in order to determine if any cooperation is possible and desired. The device may determine the device-specific functionality of the other device (at **S730**), what the other device is doing (at **S732**), and what the device itself is doing (at **S734**). This information may be used to determine (at **S736**) possible cooperation(s) between the devices. Protocols may be established for various devices or types of devices to support their cooperation. For example, in some implementations one device is touched to another to establish or indicate a desired cooperation between them. Therefore, device A may also determine (at **S738**) if there have been any indications of desired cooperation (*e.g.*, if one device has touched the other or if a person has instructed the devices to cooperate with certain devices if and when found). Based on the information determined (at **S730**, **S732**, **S734**, **S736**, **S738**), the device may select and initiate a possible cooperation (at **S740**). If multiple cooperation(s) are possible (as determined at **S736**), then the selection of one of them may be favored by an indication of desired cooperation (as determined at **S738**).

[00218] Note that since device A detected device B's heartbeat, it is possible that device B also detected device A's heartbeat. In the event that two devices are evaluating cooperation with each other, a convention may need to be established as to which device makes certain decisions. One exemplary convention is that the device that first initiates a contact with the other device takes the lead, if/when needed, in decision making. Another possible approach is to have the device with the highest device ID take the lead, if/when needed, in decision making. Those of skill in the art will realize and understand, upon reading this description, that different and/or other conflict resolution approaches may be used, if/when needed.

[00219] It should be appreciated that just because device A initiates a cooperation with device B (at **S740**) does not mean that device B will go along

with the cooperation. In some cases, devices A and B may negotiate and agree cooperation before it is initiated, in which cases device B goes along as agreed.

[00220] Those of skill in the art will realize and understand, upon reading this description, that just because certain cooperation is possible between co-owned devices does not mean that it will be done. Furthermore, it should be appreciated that different and/or other factors may be used to determine and possibly initiate cooperation between co-owned devices.

[00221] Exemplary processing for potential cooperation between friends' devices (at S720 in FIG. 7D) is shown in FIG. 7F. This process is similar to that for co-owned devices (described above with reference to FIG. 7E), but (i) requires that the devices have permission to cooperate, and (ii) preferably only updates configuration information needed to support any chosen cooperation. As shown in FIG. 7F, if permitted (based on the friend permissions associated with the owner of device B), device A establishes a connection with device B (at S742). If permitted, device A determines the specific functionality of the device B (at S744) and what device B is doing (at S746). Device A determines what it is currently doing (at S748). Based at least in part on some of the information it has determined, and on the friend permissions associated with the owner of device B, device A then determines (at S750) possible cooperation with device B. Device A also determines (at S752) if there have been any other indications of desired cooperation between the devices. Based at least in part on these determinations, device A selects (at S754) a possible and permitted cooperation with device B. If multiple cooperation(s) are possible (as determined at S736), then the selection of one of them may be favored by an indication of desired cooperation (as determined at S738). Devices A and B update (at S756) their configuration information (as and if needed) to support the selected/permitted cc cooperation. Then the permitted cooperation is initiated (at S758).

[00222] It should be appreciated that, as with co-owned devices, just because device A initiates a cooperation with device B (at S758) does not mean that device

B will go along with the cooperation. In some cases, devices A and B may negotiate and agree upon cooperation before it is initiated, in which cases device B goes along as agreed.

[00223] Those of skill in the art will realize and understand, upon reading this description, that just because certain cooperation is possible between friend's devices does not mean that it will be done. Furthermore, it should be appreciated that different and/or other factors may be used to determine and possibly initiate cooperation between friend's devices.

[00224] Although device cooperation was described above based on heartbeat detection (with other factors possibly being used to confirm or select possible device interactions), those of skill in the art will realize and understand, upon reading this description, that (as noted above), other factors may be used to initiate device cooperation. For example, as noted earlier, in some cases a device may join one or more other devices merely by being put in proximity to the other devices; in some cases, devices may be joined by specific instructions via a user interface; and in some cases, devices may be joined by having one of them touch the other. These other factors (*e.g.*, touch, proximity, specific command, etc.) may, in some cases, override other factors including what either of the devices is doing at the time.

[00225] Thus, for example, as shown in **FIG. 7G**, each device **102** may routinely monitor for some type of contact from other devices (at **S760**). The type of contact detected may depend on one or more factors such as, *e.g.*, the type of device (*i.e.*, on its underlying device-specific functionality). Some devices may attempt to initiate contact with other devices based on physical touch, user instructions, etc. Thus, the detection of a contact attempt from another device (at **S762**) may involve interpretation of voice and/or gesture instructions (using voice/speech mechanism(s) **174** and/or gesture mechanism(s) **166**), from sensor input (from one or more sensors **138**), etc. Having detected a possible contact

attempt (at **S762**), the device proceeds (at **S764**) to process the contact with the other device.

[00226] Processing of a possible contact attempt (at **S764**) may be similar to the heartbeat processing described above (with reference to **FIGS. 7D – 7F**). In particular, when a first device detects a contact attempt from a second device, the first device will still need to determine the device ID of the second device (and vice versa), determine if the devices are co-owned or owned by friends, and process the contact attempt accordingly. However, in the case of a contact attempt, a device may assume that the other device (the device that initiated the contact) is trying to establish some form of cooperation between the devices.

[00227] Thus, in the case of device initiated contact, the processing to evaluate cooperation between co-owned devices (at **S712** in **FIGS. 7D, 7E**) and evaluate cooperation between friends' devices (at **S720** in **FIGS. 7D, 7F**) may be modified as described here with reference to **FIGS. 7H – 7J**. Note that the primary difference between the processing is that the desired cooperation is given precedence in selecting a possible cooperation. Thus, in evaluating device-initiated cooperation between co-owned devices (at **S712'** in **FIG. 7I**), the device-initiated desired cooperation is selected and initiated (at **S740'** in **FIG. 7I**) if it is determined to be a possible cooperation; and in evaluating device-initiated cooperation between friends' devices (at **S720'** in **FIG. 7J**), the desired device-initiated cooperation is selected and initiated (at **S754', S758'**) if the device-initiated cooperation is possible and permitted.

[00228] Devices that are cooperating also need to be able to end their cooperation. Cooperation may be terminated in various ways, including, without limitation, by one or more of the devices being powered off, by explicit user instruction, by a change in permissions (as determined from updated configuration information received by a device), by the devices becoming separated in such a way that they can no longer cooperate (*e.g.*, one device is removed to a different room in a house). Those of skill in the art will realize and understand, upon

reading this description, that different and/or other ways of terminating device cooperation may be used.

[00229] Interaction between devices corresponds to arc #1 in **FIG. 1**.

[00230] The processing described above for device-to-device interaction is given only by way of example, and is not intended to limit the scope of the system in any way. Those of skill in the art will realize and understand, upon reading this description, that different and/or other interactions may be performed and are within the scope of this system.

COMPUTING

[00231] One of ordinary skill in the art will readily appreciate and understand, upon reading this description, that the various processes described herein may be implemented by, *e.g.*, appropriately programmed general purpose computers, special purpose computers and computing devices of any size or complexity. One or more such computers or computing devices may be referred to as a computer system.

[00232] The services, mechanisms, operations and acts shown and described above are implemented, at least in part, by software running on one or more computers of system **100**. For example, functionality associated with the backend **104** may be implemented by software running on one or more computers.

[00233] Programs that implement such methods (as well as other types of data) may be stored and transmitted using a variety of media (*e.g.*, computer readable media) in a number of manners. Hard-wired circuitry or custom hardware may be used in place of, or in combination with, some or all of the software instructions that can implement the processes of various embodiments. Thus, various combinations of hardware and software may be used instead of software only.

[00234] One of ordinary skill in the art will readily appreciate and understand, upon reading this description, that the various processes described

herein may be implemented by, *e.g.*, appropriately programmed general purpose computers, special purpose computers and computing devices. One or more such computers or computing devices may be referred to as a computer system.

[00235] **FIG. 5A** is a schematic diagram of a computer system **500** upon which embodiments of the present disclosure may be implemented and carried out.

[00236] According to the present example, the computer system **500** may include a bus **502** (*i.e.*, interconnect), one or more processors **504**, one or more communications ports **514**, a main memory **506**, read-only memory **508**, removable storage media **510**, and a mass storage **512**.

[00237] As used herein, a “processor” means one or more microprocessors, central processing units (CPUs), computing devices, microcontrollers, digital signal processors, or like devices or any combination thereof, regardless of their architecture. An apparatus that performs a process can include, *e.g.*, a processor and those devices such as input devices and output devices that are appropriate to perform the process.

[00238] Processor(s) **504** can be custom processors or any known processor, such as, but not limited to, an Intel® Itanium® or Itanium 2® processor(s), AMD® Opteron® or Athlon MP® processor(s), or Motorola® lines of processors, ARM-based processors, and the like. Communications port(s) **514** can be any of an RS-**232** port for use with a modem based dial-up connection, a 10/100 Ethernet port, a Gigabit port using copper or fiber, or a USB port, and the like. Communications port(s) **514** may be chosen depending on a network such as a Local Area Network (LAN), a Wide Area Network (WAN), or any network to which the computer system **500** connects. The computer system **500** may be in communication with peripheral devices (*e.g.*, display screen **516**, input device(s) **518**) via Input / Output (I/O) port **520**.

[00239] While referred to herein as peripheral devices, it should be appreciated that such devices may be integrated into the form of a device comprising the computer system **500**. For example, a computer system that is

used in a cellular phone has the display screen and input device as part of the phone. It should also be appreciated that the peripheral devices, if provided, may be combined (*e.g.*, in the case of a touch screen or the like).

[00240] Those of skill in the art will realize and understand, upon reading this description, that not every computer system **500** needs to include all of the components. For example, not every computer system **500** requires removable storage media **510** or mass storage **512**. Similarly, not every computer system **500** will have a display screen **516**.

[00241] Main memory **506** can be Random Access Memory (RAM), or any other dynamic storage device(s) commonly known in the art. Read-only memory **508** can be any static storage device(s) such as Programmable Read-Only Memory (PROM) chips for storing static information such as instructions for processor **504**. Mass storage **512** can be used to store information and instructions. For example, hard disks such as the Adaptec® family of Small Computer Serial Interface (SCSI) drives, an optical disc, an array of disks such as Redundant Array of Independent Disks (RAID), such as the Adaptec® family of RAID drives, or any other mass storage devices may be used.

[00242] Bus **502** communicatively couples processor(s) **504** with the other memory, storage and communications blocks. Bus **502** can be a PCI / PCI-X, SCSI, a Universal Serial Bus (USB) based system bus (or other) depending on the storage devices used, and the like. Removable storage media **510** can be any kind of external hard-drives, floppy drives, IOMEGA® Zip Drives, Compact Disc – Read Only Memory (CD-ROM), Compact Disc – Re-Writable (CD-RW), Digital Video Disk – Read Only Memory (DVD-ROM), SDRAM, etc.

[00243] Embodiments herein may be provided as one or more computer program products, which may include a machine-readable medium having stored thereon instructions, which may be used to program a computer (or other electronic devices) to perform a process. As used herein, the term “machine-readable medium” refers to any medium, a plurality of the same, or a combination

of different media, which participate in providing data (*e.g.*, instructions, data structures) which may be read by a computer, a processor or a like device. Such a medium may take many forms, including but not limited to, non-volatile media, volatile media, and transmission media. Non-volatile media include, for example, optical or magnetic disks and other persistent memory. Volatile media include dynamic random access memory, which typically constitutes the main memory of the computer. Transmission media include coaxial cables, copper wire and fiber optics, including the wires that comprise a system bus coupled to the processor. Transmission media may include or convey acoustic waves; light waves and electromagnetic emissions, such as those generated during radio frequency (RF) and infrared (IR) data communications.

[00244] The machine-readable medium may include, but is not limited to, floppy diskettes, optical discs, CD-ROMs, magneto-optical disks, ROMs, RAMs, erasable programmable read-only memories (EPROMs), electrically erasable programmable read-only memories (EEPROMs), SDRAMs, magnetic or optical cards, flash memory, or other type of media/machine-readable medium suitable for storing electronic instructions. Moreover, embodiments herein may also be downloaded as a computer program product, wherein the program may be transferred from a remote computer to a requesting computer by way of data signals embodied in a carrier wave or other propagation medium via a communication link (*e.g.*, modem or network connection).

[00245] Various forms of computer readable media may be involved in carrying data (*e.g.*, sequences of instructions) to a processor. For example, data may be (i) delivered from RAM to a processor; (ii) carried over a wireless transmission medium; (iii) formatted and/or transmitted according to numerous formats, standards or protocols; and/or (iv) encrypted in any of a variety of ways well known in the art.

[00246] A computer-readable medium can store (in any appropriate format) those program elements which are appropriate to perform the methods.

[00247] As shown, main memory **506** is encoded with application(s) **522-1** that supports the functionality as discussed herein (the application **522-1** may be an application that provides some or all of the functionality of the services described herein, *e.g.*, backend processing). Application(s) **522-1** (and/or other resources as described herein) can be embodied as software code such as data and/or logic instructions (*e.g.*, code stored in the memory or on another computer readable medium such as a disk) that supports processing functionality according to different embodiments described herein.

[00248] For example, as shown in **FIGS. 5B-5C**, when a computer system **500** is used to implement functionality of the backend **104**, then application(s) **522-1** may include backend applications **524-1**, and when a computer system **500** is used to implement functionality of a device, then applications **522-1** may include device applications **526-1**.

[00249] During operation of one embodiment, processor(s) **504** accesses main memory **506** via the use of bus **502** in order to launch, run, execute, interpret or otherwise perform the logic instructions of the application(s) **522-1**. Execution of application(s) **522-1** produces processing functionality of the service related to the application(s). In other words, the process(es) **522-2** represent one or more portions of the application(s) **522-1** performing within or upon the processor(s) **504** in the computer system **500**.

[00250] For example, as shown in **FIGS. 5D-5E**, when a computer system **500** is used to implement functionality of the backend **104**, then process(es) **522-2** may include backend process(es) **524-2**; and when a computer system **500** is used to implement functionality of a device, then process(es) **522-2** may include device process(es) **526-2**.

[00251] It should be noted that, in addition to the process(es) **522-2** that carries(carries) out operations as discussed herein, other embodiments herein include the application **522-1** itself (*i.e.*, the un-executed or non-performing logic instructions and/or data). The application **522-1** may be stored on a computer

readable medium (*e.g.*, a repository) such as a disk or in an optical medium.

According to other embodiments, the application **522-1** can also be stored in a memory type system such as in firmware, read only memory (ROM), or, as in this example, as executable code within the main memory **506** (*e.g.*, within Random Access Memory or RAM). For example, application **522-1** may also be stored in removable storage media **510**, read-only memory **508**, and/or mass storage device **512**.

[00252] Those skilled in the art will understand that the computer system **500** can include other processes and/or software and hardware components, such as an operating system that controls allocation and use of hardware resources. For example, operating system (OS) programs including a kernel may be processes on the computer system.

[00253] As discussed herein, embodiments of the present invention include various steps or operations. A variety of these steps may be performed by hardware components or may be embodied in machine-executable instructions, which may be used to cause a general-purpose or special-purpose processor programmed with the instructions to perform the operations. Alternatively, the steps may be performed by a combination of hardware, software, and/or firmware. The term “module” refers to a self-contained functional component, which can include hardware, software, firmware or any combination thereof.

[00254] One of ordinary skill in the art will readily appreciate and understand, upon reading this description, that embodiments of an apparatus may include a computer/computing device operable to perform some (but not necessarily all) of the described process.

[00255] Embodiments of a computer-readable medium storing a program or data structure include a computer-readable medium storing a program that, when executed, can cause a processor to perform some (but not necessarily all) of the described process.

[00256] Where a process is described herein, those of skill in the art will appreciate that the process may operate without any user intervention. In another embodiment, the process includes some human intervention (*e.g.*, a step is performed by or with the assistance of a human).

[00257] With reference again to **FIGS. 4A – 4B**, recall that a device **102** includes a computer system **146**. In some cases, computer system **146**, alone or in combination with system mechanism(s) / data **134**, may correspond to a computer system **500** as described above (with reference to **FIGS. 5A-5C**), although, as should be appreciated, the computer system **146** may not include all of the components shown in **FIG. 5A**, and the computer system **146** may include additional components (*e.g.*, related to special processing required by the device **102**). For example, a computer system **146** may include multiple processors, multiple memories, etc. It should also be appreciated that a computer system **146** may be formed of multiple computer systems **500**. In addition, the computer system **146** may implement some or all of the device-specific functionality **132**.

CONTROLLING DEVICES

[00258] System-enabled devices **102** may be controlled by one or more of voice control, gesture control, contact control (*e.g.*, using buttons and the like). In addition, certain kinds of system-enabled devices **102**, when in the presence of other like devices, may be fully or partially controlled by one or more of the other devices or by instructions given to one or more of the other devices. For example, when the devices **102** are speakers, multiple devices may be combined to operate together. In such cases, certain commands (*e.g.*, raise volume) may be given by a user to one of the devices (speakers) but should be followed by all of the cooperating devices.

Voice Control

[00259] A device's voice mechanism(s) **166** may be used to support voice control of the device. Voice mechanism(s) **166** preferably include voice recognition mechanisms for basic commands appropriate to the kind of device. For example, for a device that is primarily a speaker, the voice commands may include commands to power the device off (or on from a low power mode), play louder, softer, etc. The voice mechanism(s) **166** may be implemented using special hardware or circuitry and DSPs (Digital Signal Processors).

[00260] Each device preferably maintains a corpus of recognized words from users. In some cases the device may maintain multiple corpora of words, one for each of a number of users. Since a device may be controlled by more than one person (and, depending on permissions set in the device, the person controlling a device may not be a known user of the system), the device needs to be able to associate certain commands with appropriate users. In this manner the device can determine which corpus of words to use for the voice / command recognition.

[00261] The device **102** may use face recognition mechanism(s) **168** in combination with one or more cameras (sensors **138**) to associate a voice with a particular user in order to select an appropriate corpus.

[00262] In some cases the device may not be able to process a voice command/request. This may be due to any number of factors including complexity of the command/request, environmental causes (*e.g.*, noise), the speaker's accent, etc. In such cases the device may, if possible (*e.g.*, if connected to the network), and if permissible, send the voice command/request to the backend for processing (*e.g.*, by voice recognition provided added functionality **120**). The voice may be sent in a raw form or in some preprocessed form. The result of such processing may be a command/request for the backend (*e.g.*, a database query) or a command for the device itself. It should be appreciated that device commands processed remotely via the backend may not be quick enough to control certain aspects of a device (*e.g.*, for a speaker, "Play louder"), and that

backend processing is more useful for more complex commands, especially those involving database queries.

Gesture Control

[00263] A device's gesture mechanism(s) **164** may be used, alone or in combination with the voice mechanism(s) **166**, to support gesture control of the device. Gesture mechanism(s) **164** preferably include gesture recognition mechanisms for basic commands appropriate to the kind of device. The gesture mechanism(s) **164** may use one or more of the sensors **138**, including, *e.g.*, one or more cameras. Special purpose gesture detection / recognition hardware and circuitry may be used.

Face and Gaze Detection

[00264] In some cases there may be multiple people talking simultaneously near a device. Some of what they are saying may not be intended as commands for the device. Accordingly, in some cases, a device may use a combination of gaze detection (determined by face/gaze mechanism(s) **168**) to determine whether a voice command is intended for the device. Face/gaze mechanism(s) **168** may use one or more sensors (*e.g.*, one or more cameras) to determine whether or not a person talking is actually looking at the device **102**. Since a person may begin talking (to a device) before they completely face the device, preferably each device constantly buffers a period of sound so that once a gaze is detected, the device can begin voice recognition of the buffered stream.

[00265] In some cases mouth movement detection can be used in combination with gaze detection to confirm that a person looking at a device is the one talking to the device.

[00266] Those of skill in the art will realize and understand, upon reading this description, that voice recognition, gesture detection and recognition, and face and/or gaze detection may be used in various combinations to control a device.

REPORTING BACK TO THE BACKEND

[00267] In operation, each device reports information to the backend **102** (corresponding to arc #7 in **FIG. 1**). The information preferably includes the unique device ID of the reporting device and, if the device is associated with a user, the unique user ID associated with the owner of the reporting device. Some or all of the information reported by each device may be stored in the devices database **128** and/or the user database **130**, *e.g.*, as device history and/or user history, respectively. Since each device has a unique device ID, and since each user has a unique user ID, information from a device may be stored in the database(s) keyed on the device and user IDs.

[00268] In some cases, a device **102** may include information about its location at the time of its reporting. In those cases, the location information may be stored in the device database (both as current device location and as device history). Similarly, the location information may be stored in the user database as per device history. In this manner, queries to the database may include queries about device location.

[00269] When a user registers with the system, the user may provide location identification information associated with their current location at the time of registration. A user may also store multiple locations in the system, each with different identification provided by the user. For example, a user may store GPS location information for their home, their work, their friends' homes, etc. In this manner, the system can support database queries based on named locations (*e.g.*, "Where is my device?" to which the system's response may be "At Joe's home."). Preferably a user need not specifically request storage of location information, as location (*e.g.*, GPS) data are preferably stored automatically as part of history data or context metadata.

[00270] A device **102** may also report information that is specific to the kind of device (*e.g.*, the device's specific functionality). For example, a device that is primarily a speaker may report to the backend information about what it plays and

where. In some cases, the information may include information about device settings and what other devices were involved (*e.g.*, joined). In this manner, the database(s) will support queries of the kind “What was I playing at Joe’s house last night at around 10 o’clock?”, to which the system may provide a list of songs.

[00271] A device **102** may also report information about proximate devices or users.

EXAMPLE DEVICE – SOUND RENDERING DEVICE

[00272] **FIGS. 8A – 8D** depict aspects of the architecture of an exemplary device **800** (an embodiment of device **102**) in which the specific functionality of the device is sound rendering. Device **800** may be used, *e.g.*, as a speaker. As shown in **FIGS. 8A—8B**, sound rendering device **800** includes components **832** supporting device-specific functionality. These components **832** include one or more speaker drivers **860**, one or more signal processors **862**, one or more processors **864**, memory/storage **866**, and controls **868**.

[00273] As shown in **FIGS. 8A, 8C**, the device **800** may include communications mechanisms, including Bluetooth mechanisms (BLE mechanisms), Ethernet mechanisms, ZigBee mechanisms, cellular mechanisms, and Wi-Fi mechanisms. In one implementation, the communications mechanisms of a device **800** include Bluetooth mechanisms and do not include Ethernet, ZigBee, or cellular mechanisms.

[00274] As shown in **FIGS. 8A and 8D**, sound rendering device **800** may also include sensors **838**, including one or more cameras **870**, one or more microphones **872**, device motion sensor(s), location / position sensor(s), external motion sensor(s), touch/contact sensor(s), light sensor(s), temperature sensor(s), and other sensors. In one implementation, the sensors of a device **800** do not include cameras or temperature sensors.

[00275] In one exemplary implementation, the following may be used for some of these components:

Component	Part Description
<i>Components for device-specific functionality</i>	
speaker drivers 860	Subwoofer (<i>e.g.</i> , TangBand W3-2000) and two tweeters (<i>e.g.</i> , Vifa Ox20SC00)
signal processors 862	<i>e.g.</i> , NXP Class-D amplifier and DSP
processors 864	Single or multicore ARM-based SOC (<i>e.g.</i> , FreeScale i.MX6)
memory/storage 866	Flash NAND (<i>e.g.</i> , 4 GBytes from Micron) and DDR (<i>e.g.</i> , 1 GByte DDR3 from Micron)
controls 868	Capacitive touch buttons, strips and surfaces, haptic and digital encoders, voice and gesture
<i>Sensor components</i>	
cameras 870	5 MPixel sensor, <i>e.g.</i> , OmniVision 5640
microphones 872	Two for active noise cancellation (<i>e.g.</i> , Panasonic WM-61A)
device motion	Accelerometer and compass (<i>e.g.</i> , FreeScale MMA8453)
location / position	Through GPS associated with WiFi/BT combo SIP, or through proximity to another GPS device
external motion	Accelerometer and compass (<i>e.g.</i> , FreeScale MMA8453)
touch/contact	Capacitive touch sensors (<i>e.g.</i> , Azoteq IQS253DNR)
light	Sensing via camera sensor, producing via multiple LEDs
<i>Communications components</i>	
Bluetooth	WiFi/BT combo SIP <i>e.g.</i> , Atheros AR6233
Wi-Fi	WiFi/BT combo SIP <i>e.g.</i> , Atheros AR6233
<i>Computer System Components</i>	
	<i>e.g.</i> , FreeScale i.MX6 ARM-based MCU
<i>Power Components</i>	
	<i>e.g.</i> , FreeScale power management chip, coulomb counter and battery power management chip

[00276] In another exemplary implementation, the following may be used for some of these components:

Component	Part Description
<i>Components for device-specific functionality</i>	

Component	Part Description
speaker drivers 860	Subwoofer (<i>e.g.</i> , GGEC W0200A) and two tweeters (<i>e.g.</i> , GGEC T20N4A)
signal processors 862	<i>e.g.</i> , TI Class-D amplifier and DSP
processors 864	Single or multicore ARM-based SOC (<i>e.g.</i> , FreeScale i.MX6)
memory/storage 866	Flash NAND (<i>e.g.</i> , 4 GBytes from Micron) and DDR (<i>e.g.</i> , 1 GByte DDR3 from Micron)
controls 868	Capacitive touch buttons, strips and surfaces, haptic and digital encoders, voice and gesture. Accelerometers, and device motion)
<i>Sensor components</i>	
cameras 870	5 MPixel sensor, <i>e.g.</i> , OmniVision 5640
microphones 872	Three for active noise cancellation (<i>e.g.</i> , Knowles SPK0833LM4H-B)
device motion	Accelerometer and compass (<i>e.g.</i> , FreeScaleMMA8451)
location / position	Through GPS associated with WiFi/BT combo SIP, or through proximity to another GPS device
external motion	Accelerometer and compass (<i>e.g.</i> , FreeScale MMA8451) (Gyro and compass)
touch/contact	Capacitive touch sensors (<i>e.g.</i> , Cypress CY8C5267LTI-LP089)
light	Sensing via camera sensor or ambient light sensor (<i>e.g.</i> , Capella CM32180), producing via multiple LEDs
<i>Communications components</i>	
Bluetooth	WiFi/BT combo SIP <i>e.g.</i> , Marvell 88W8797
Wi-Fi	WiFi/BT combo SIP <i>e.g.</i> , Marvell 88W8797
<i>Computer System Components</i>	
	<i>e.g.</i> , FreeScale i.MX6 ARM-based MCU (MCU is Cypress CY8C5267LTI-LP089)
<i>Power Components</i>	
	<i>e.g.</i> , FreeScale power management chip, coulomb counter and battery power management chip

[00277] It should be appreciated that the above list is given only by way of example, and is not intended to limit the scope of the device in any way.

[00278] Any known mechanism may be used for the various interface mechanisms **162**. For example, the face movement detection may use the CANDIDE system for model-based coding of human faces. CANDIDE uses a

facial model with a small number of polygons (approximately 100) that allows for fast reconstruction with moderate computing power.

[00279] The sound-rendering device **800** may operate as a device **102** as described above.

[00280] Those of skill in the art will realize and understand, upon reading this description, that different and/or other specific components may be used within the sound rendering device **800**, and such other components are contemplated herein and are within the scope of the system. It should be appreciated that the various components may be implemented and packaged in multiple ways, and that the device is not limited by the manner in which the components are implemented or packaged. It should further be appreciated that the device is not limited by the form that the packaging or device takes (*i.e.*, by the device's form factor).

[00281] The following table contains the syntax of a list of exemplary commands (*i.e.*, phrases) that a device can interpret locally using voice recognition mechanisms. As used herein, a phrase means one or more words. In the corpus table below, phrases in ***bold italic*** font are in the local corpus; phrases in square parentheses (“[“, “]”) are optional. The vertical bar (“|”) between phrases means “or” (*i.e.*, one of the phrases). Thus, *e.g.*, “A” | “B” means the phrase “A” or the phrase “B”. A phrase denoted “#n” means a number. A phrase followed by a star (“*”) means that the phrase may be repeated. A word phrase followed by “(s)” means that the singular or plural of the word may be used. Thus, *e.g.*, item number 2 in the table,

[[***a***] ***little*** | ***much**** | #n] ***softer*** | ***lower*** | ***higher*** | ***quiet***[***er***]

could correspond to “a little lower”, “much higher”, “softer”, “five lower”, “little softer”, “much much softer”, “quiet”, “quieter”, “a little quieter”, etc.

	Voice Instructions (local corpus syntax)
1	<i>Play</i>
2	[[<i>a</i>] <i>little</i> <i>much</i> * <i>#n</i>] <i>softer</i> <i>lower</i> <i>higher</i> <i>quiet</i> [<i>er</i>]
3	[[<i>re</i>] <i>play</i>] [[<i>the</i>] [<i>next</i> <i>previous</i>] [<i>any</i>] <i>a</i> [<i>a</i>] <i>random</i>] [<i>#n</i>] [<i>song</i> (<i>s</i>) <i>tune</i> (<i>s</i>)]
4	<i>stop</i>
5	[<i>skip</i>] [[[<i>the</i>] [<i>next</i> <i>previous</i>] [<i>#n</i>] <i>a</i>] [<i>song</i> (<i>s</i>) <i>tune</i> (<i>s</i>)]
6	<i>mute</i>
7	[<i>more</i> <i>less</i>] <i>bass</i> <i>treble</i>
8	[<i>adjust</i>] [<i>tone</i> <i>volume</i> <i>treble</i> <i>bass</i>] [<i>up</i> <i>high</i> (<i>er</i>) <i>low</i> (<i>er</i>) <i>down</i>]

[00282] As an example, item no. 3 in the table could mean any of: “Play the next ten tunes”, “Play any random song”, “play next tune”, “replay the previous song,” “Play random tunes”, “Replay”, “Play”, etc. As another example, item no. 5 could mean any of: “Skip”, “next three songs”, “a tune”, “skip seven”, “previous tune”, etc. As a further example, item no. 8 could mean any of “tone”, “adjust treble up”, “bass lower”, etc.

[00283] In the following:
 <*musical entity*> can be a specific song, artist, or album,
 <*artist entity*> is the name of an artist (*e.g.* Pink Floyd)
 <*album entity*> is a specific collection of songs in order (*e.g.*, “Dark side of the moon”)

[00284] The voice commands in some embodiments may include:
“play” (<musical entity> | <musical entity> “by” <artist entity> | <album entity>
by <artist entity> | “something” | “something new” | “something different” | “my favorites”)

[00285] The voice/speech recognition mechanism(s) 178 may thus recognize certain spoken phrases and will then have to determine their corresponding semantics (*i.e.*, meaning) and provide corresponding instructions to the other operational mechanisms (*e.g.*, to the command / control mechanism(s) 158) to actually control the device.

[00286] It should be appreciated that this exemplary corpus provides the syntax of recognized phrases, and that not all phrases will have meaning (or reasonable meaning) for the device. For example, no. 3 above would support recognition of the phrase “replay the next any three song”, and no. 5 above would support recognition of the phrase “skip the previous a tune”. While both of these phrases are syntactically correct (according to the syntax in the corpus), they may not correspond to any meaningful command and may be ignored by the device.

[00287] The corpus syntax given above for voice/speech recognition mechanism(s) **178** is only provided as an example, and those of skill in the art will realize and understand, upon reading this description, that different and/or other voice instructions may be understood by the sound rendering device **800**, and are contemplated herein.

[00288] Sound rendering devices **800** may cooperate with each other to render the same sounds (*e.g.*, to play the same music from the same source and at the same time – preferably synchronized). When two or more sound rendering devices **800** are cooperating to render sound from the same source, they need not all render exactly the same sound. For example, multiple sound rendering devices **800** may cooperate to render sound from the same source as a surround sound system. As another example, multiple sound rendering devices **800** may cooperate to render sound from the same source such that some of them render some sound (*e.g.*, from some musical instruments) while others render other sound (*e.g.*, from other musical instruments).

[00289] It should be appreciated that a sound-rendering device **800** may also be a source of the signal used to produce the sound. For example, a smartphone (such as an iPhone or the like) may have a speaker (albeit a small one) and produce a signal that can be used (both by the phone itself and by other devices **800**) to produce sound.

[00290] Examples of cooperation between sound-rendering devices **800** are given below with reference to **FIGS. 9A – 9C**.

Example cooperation – Stereo

[00291] With reference to the drawing in **FIG. 9A**, two sound-rendering devices **800-A** and **800-B** may cooperate to provide a stereo effect. The DSPs in the devices cooperate to produce, *e.g.*, a Haas effect. It should be appreciated that the devices may determine their own relative positions (*e.g.*, using echo location or some other mechanism), and they may use this relative position information to optimize the cooperative effect.

Example cooperation – separate instruments

[00292] With reference to the drawing in **FIG. 9B**, multiple sound-rendering devices **800-A – 800-D** may cooperate such that each one of them plays only some of the instruments in the source signal. The source signal may provide separate streams for each instrument or each DSP may be programmed to filter out certain instruments. As devices join this group, each device is allocated one or more instruments to render. For example, assume that initially device A was playing alone and rendering all sounds in the source signal. When device B joins device A, then device A may render, *e.g.*, Bass and violin, and device B may render Cello and vocals. When device C joins devices A and B, then device C can be given responsibility for violin, leaving device A with only bass. Then when device D joins, it can take responsibility for vocals from device B (as shown in the drawing in **FIG. 9B**). If other devices join the group they can combine with one or more of the already-present devices or they can take on some other responsibility. If a device leaves the group then the part of the signal for which it was responsible should be re-assigned to another device still in the group.

[00293] Although this example shows selective responsibility for different instruments in an audio signal, those of skill in the art will realize and understand, upon reading this description, that devices **800** may be given responsibility for different and/or other aspects of an audio stream. It should be appreciated that a device **800** may render (or not render) any part or parts of an audio stream that its

DSP can filter (in or out). Furthermore, a device **800** may enhance or modify any part or parts of an audio stream that its DSP can filter.

Example cooperation – arbitrary arrangement

[00294] With reference to the drawing in **FIG. 9C**, multiple sound-rendering devices **800-A – 800-E**, placed or located in an arbitrary (haphazard) arrangement may cooperate. The devices may determine their own relative positions (*e.g.*, using echo location or some other approach), and they may use this relative position information to optimize the cooperative effect. The devices may also cooperate to produce an optimal or beneficial cooperative effect for a listener (if the position of the listener is known or can be determined).

[00295] If the devices assume that the listener is the person giving them commands (voice, gesture, etc.), then devices can use their respective cameras to locate (and follow) the listener, adjusting the sound accordingly. A single camera in a single device may be able to determine the direction in which the listener is located. Some techniques allow single cameras to determine approximate distance. Multiple cameras (in a single device or in multiple devices) can more accurately locate a listener (*e.g.*, by face location and/or movement tracking). In addition to (or instead of) using one or more cameras to locate a person, location detection may be achieved using voice input and echo detection. Thus, *e.g.*, in a device that does not have a camera voice input, echo detection may be used alone to determine location. In a device that has a camera, voice input, echo detection may be used alone or in combination with the camera(s) to determine location.

[00296] Various options for cooperation are possible in this example. For example, different instruments can be placed along different stereo lines (**L1**, **L2**, **L3** in the drawing); some of the devices **800-A** to **800-E** may be used to add effects such as echo, reverberation, or the like; some of the devices may be used to cancel out room noise or echoes, etc.

[00297] For example, as shown in **FIG. 9C**, bass may be rendered by devices A and C (along the stereo line L1), violin and cello may be rendered by devices A and E (along the stereo line L2), and vocals may be rendered by devices D and E (along the stereo line L3). Effects (*e.g.*, cancellation of room noise or echoes) may be performed by device B.

[00298] If the audio signal contains multiple channels or other encodings, the devices may cooperate to render these channels.

[00299] The nature of sound rendering devices **800**, especially portable devices, lends to their cooperation. In some implementations, a user may grant a friend guest privileges to share their devices **800**. In some implementations a user may grant temporary (“party mode”) privileges to all any other device to share their sound rendering devices **800**.

Conflict resolution

[00300] As described above (with reference to heartbeat processing), it may be necessary for one device to take control when multiple devices try to cooperate. In the case of sound rendering devices **800**, a preferred convention is that the device that first initiates a contact with the other device (*e.g.*, by touching it, etc.) takes the lead, if/when needed, in decision making. Those of ordinary skill in the art will realize and appreciate, upon reading this description, that different and/or other techniques of determining which device takes control may be used and are contemplated herein.

Genre Classification

[00301] Sound may be classified into genre (*e.g.*, vocal, instrumental, jazz, classical, spoken voice, etc.) and these genre may be provided with the sound source signal and may be used to automatically set or adjust the DSPs in a sound rendering device **800**. In some cases, the preset genre information may be combined with or overridden by user preferences (which may be provided via

some user interface or learned by the device based on user interactions with the device). For example, if a user always adjusts the DSP settings for a certain genre of music, always overriding the preset DSP settings, then the device **800** may learn the user's desired settings and always use those instead of the system's preset settings for that genre.

[00302] Genre information may be set, *e.g.*, in an advance, offline process that analyzes the source sound. For example, the provider of a source library may pre-analyze all music in their library to classify the genre of each item of music. That classification may be stored, *e.g.*, a bit vector representing the genre, and may be provided with the source data. It should be appreciated, however, that the processing of genre information in a source signal is independent of the manner in which that genre information was obtained or set.

[00303] Cooperating devices **800** may use genre information in the source signal to determine and adjust how they cooperate. Thus, when rendering sound corresponding to multiple songs, cooperating devices **800** may modify the way in which they cooperate depending on the genre of each song.

[00304] Those of skill in the art will realize and understand, upon reading this description, that different and/or other kinds of cooperation may be used by multiple sound rendering devices **800**, and such cooperation(s) are contemplated herein and are within the scope of the system.

History and Learning

[00305] As noted above, the system **100** may obtain information from each device **102**. In the case of sound rendering devices **800**, the devices preferably inform the backend **184** what sound they are rendering (*e.g.*, what music, etc. they are playing), as well as when and where it is being rendered. To this end, each device **800** retains a history of its activities and provides that history to the backend **184** on a regular basis and/or when it is able to. The history may be provided as a time-stamped, ordered list of activities and device settings that may

be used to reproduce the device's activities. If a device is cooperating with another device, that information is also included in the history and both (all) cooperating devices provide their own history information to the backend.

[00306] The backend stores device history information in the device and user databases **128, 130**.

[00307] This kind of device history information supports subsequent queries (via the backend **184** and possibly the added functionality **120**) of the kind:

1. "Play what I was listening to on Monday at 4:30 PM."
2. "Play what I was listening to with [user] Joe on Sunday morning."
3. "Set my device to the settings I had on July 1, 2012."
4. "Play what [user and friend] Mary is listening to now."

[00308] Note that query #2 may require that Joe and the user making the query be friends and may require permission from Joe. Query #4 may require that Mary and the user making the query be friends and may require permission from Mary. Note too that Query #4 assumes that the system has been updated to know (in near real time) what Mary is listening to.

[00309] These queries are only provided as examples, and are not intended to limit the scope of the system in any way.

Noise cancellation

[00310] As noted above, a device may try to filter out environmental noise in order to process voice interactions more precisely. A sound-rendering device **800** poses additional problems, since the device itself may be a source of environmental noise. In any case, the device should not perceive sound rendered by the device as commands to the device itself. Accordingly, in preferred implementations, a sound-rendering device **800** filters out the sound it produces from sound obtained by its sound sensors (microphones).

[00311] As used herein the words "first", "second", and so on, when used as adjectives before a term, are merely used to distinguish similar terms, and their use does not imply or define any numerical limits or any ordering (temporal or otherwise). Thus, *e.g.*, the terms "first device" and "second device" are merely used to refer to and distinguish between different devices.

[00312] As used herein, including in the claims, the phrase "based on" means "based at least in part on," unless specifically stated otherwise. Thus, *e.g.*, the phrase "based on XYZ" means "based at least in part on XYZ."

[00313] Thus are described a unified framework for device configuration, interaction and control, along with systems, devices and mechanisms used therein. While the invention has been described in connection with what is presently considered to be the most practical and preferred embodiments, it is to be understood that the invention is not to be limited to the disclosed embodiment, but on the contrary, is intended to cover various modifications and equivalent arrangements included within the spirit and scope of the appended claims.

WHAT IS CLAIMED:**We claim:**

1. A device comprising:

(A) first mechanisms supporting device-specific functionality of the device; and

(B) second mechanisms supporting control of said first mechanisms, said second mechanisms including:

(B)(1) sensors configured to obtain information relating to physical and environmental properties of said device and/or an environment around said device;

(B)(2) control mechanisms; and

(B)(3) human-interface mechanisms configured to obtain sensor input from at least some of said sensors, to determine sensor information based on said sensor input, and to provide said sensor information to said control mechanisms,

(C) wherein said control mechanisms are configured to

(C)(i) determine control information based on said sensor information from said human-interface mechanisms, and

(C)(ii) provide at least some of said control information to said first mechanisms, and

(D) wherein said first mechanisms are configured and adapted to obtain said control information from said control mechanisms and to operate, at least in part, in accordance with said control information obtained from said control mechanisms, and

(E) wherein said sensors include:

(E)(1) one or more cameras configured to obtain image information from said environment around said device and to provide said image information to said human-interface mechanisms, and

(E)(2) one or more microphones configured to obtain sound information from said environment around said device and to provide said sound information to said human-interface mechanisms, and

(F) wherein said human-interface mechanisms comprise:

(F)(1) speech mechanisms configured to recognize speech in said sound information and to provide information about recognized speech in said sound information as speech information to said control mechanisms, wherein said control mechanisms determine said control information based on said speech information; and

(F)(2) face mechanisms configured to detect face information and/or gaze information in said image information;

wherein said speech mechanisms (F)(1) are configured to initiate speech recognition based on information detected by said face mechanisms.

2. The device of claim 1 wherein said speech mechanisms are configured to initiate speech recognition based on gaze information detected by said face mechanisms.

3. The device of claims 1 or 2 wherein said speech mechanisms are configured to buffer sound information from said environment around said device and to initiate speech recognition of buffered sound information based on gaze information detected by said face mechanisms.

4. The device of claim 1 or 2 wherein said speech mechanisms use at least one speech corpus and wherein said speech mechanisms are configured to select a speech corpus from said at least one speech corpus based on face information provided by said face mechanisms.

5. The device of claim 1 or 2 wherein said human-interface mechanisms further comprise:

(F)(3) gesture mechanisms configured to detect and recognize gestures in said image information; and to provide said information about recognized gestures in said image information as gesture information to said control mechanisms, and

wherein said control mechanisms are configured to determine said control information based on said gesture information.

6. A device comprising:

(A) device-specific mechanisms supporting device-specific functionality of the device; and

(B) second mechanisms supporting control of said device-specific mechanisms, said second mechanisms including:

(B)(1) control mechanisms;

(B)(2) sensors, including one or more cameras configured to obtain image information from an environment around said device, and one or more microphones configured to obtain sound information from said environment around said device; and

(B)(3) human interface mechanisms comprising:

(B)(3)(1) face mechanisms configured to determine gaze information in image information obtained from said one or more cameras, and

(B)(3)(2) speech mechanisms configured to recognize speech in sound information obtained from said one or more microphones and to provide information about recognized speech in said sound information as speech information to said control mechanisms,

wherein said speech mechanisms are configured to initiate speech recognition based on gaze information detected by said face mechanisms, and

wherein said human-interface mechanisms are configured to determine interface information based on said information about recognized speech obtained from said speech mechanisms, and to provide said interface information to said control mechanisms, and

wherein said control mechanisms are configured:

to determine control information based on said interface information from said human-interface mechanisms, and

to provide at least some of said control information to said device-specific mechanisms, and

wherein said device-specific mechanisms are configured to obtain said control information from said control mechanisms and to operate based, at least in part, in accordance with said control information obtained from said control mechanisms.

7. The device of any of the preceding claims wherein the device-specific functionality comprises sound rendering.

8. The device of claim 7 wherein the device is a speaker.

9. A method of operating a device, the method implemented, at least in part, by hardware, including at least one processor and a memory, the method comprising:

- (A) buffering sensor information from an environment around said device as buffered sensor information in a buffer of said memory;
- (B) detecting a gaze of a person in said environment;
- (C) based on said gaze detected in (B), initiating recognition of some sensor information including said buffered sensor information;
- (D) said recognition initiated in (C) determining at least one instruction in said sensor information including said buffered sensor information; and
- (E) operating said device based on said at least one instruction.

10. The method of claim 9 wherein the sensor information comprises image information and wherein the recognition initiated in (C) comprises gesture recognition, and wherein said at least one instruction is determined based on at least one gesture recognized by said gesture recognition.

11. The method of claims 8 or 9 wherein the sensor information comprises sound information and wherein the recognition initiated in (C) comprises speech recognition, and wherein said at least one instruction is determined based on at least one spoken phrase recognized by said speech recognition.

12. A method of operating a device, the method implemented, at least in part, by hardware, including at least one processor and a memory, the method comprising:

- (A) buffering sound from an environment around said device as buffered sound in said memory;
- (B) detecting a gaze of a person in said environment;
- (C) based on said gaze detected in (B), initiating speech recognition of some sound including said buffered sound;
- (D) said speech recognition determining at least one instruction in said sound including said buffered sound; and
- (E) operating said device based on said at least one instruction.

13. A method of operating a device, the method implemented, at least in part, by hardware, including at least one processor and a memory, the method comprising:

- (A) buffering image information from an environment around said device as buffered image information in said memory;
- (B) detecting a gaze of a person in said environment;
- (C) based on said gaze detected in (B), initiating gesture recognition of some image information including said buffered image information;
- (D) said gesture recognition determining at least one instruction in said image information including said buffered image information; and
- (E) operating said device based on said at least one instruction.

14. The method of any of claims 9 to 13 wherein the device is a sound rendering device.

15. The method of claim 14 wherein the device is a speaker.

16. A method, operable in a framework, the method comprising:

(A) providing a device, said device having a device-specific functionality;

(B) associating the device with a user of said framework, said user having user-specific configuration information associated therewith, said user-specific configuration information for said user comprising: at least one speech corpus supporting recognition of speech of said user; and

(C) automatically configuring said device with configuration information based on said user-specific configuration information associated with the user.

17. The method of claim 16 wherein the user-specific configuration information for said user further comprise one or more of:

network configuration information;

password information;

at least one gesture corpus for gesture recognition associated with the user;
and

face information for face recognition of the user.

18. The method of one of claims 16 and 17 further comprising:

(D) updating at least some of said configuration information on said device based on updated user-specific configuration information associated with the user.

19. The method of claim 18 wherein said device obtains said updated user-specific configuration information from another device.

20. The method of claim 19 further comprising:

(D) providing at least some user-specific configuration information to another device.

21. The method of claim 18 wherein the device updates some of the user-specific configuration information based on one or more human interactions with said device.

22. The method of claim 21 further comprising:

(D) providing updated user-specific configuration information to another location.

23. The method of claim 22 wherein the updated user-specific configuration information is stored at the other location.

24. A method, operable in a framework, the method comprising:

(A) associating a first device with a user of said framework, said user having user-specific configuration information associated therewith, the user-

specific configuration information including at least one corpus for recognition of interactions associated with the user; and

(B) automatically configuring said device with at least some of said user-specific configuration information associated with the user.

25. The method of claim 24 wherein the at least one corpus comprises one or more of:

at least one speech corpus for speech recognition associated with the user;
and

at least one gesture corpus for gesture recognition associated with the user.

26. The method of claim 24 or 25 wherein the user-specific configuration information further comprises one or more of:

network configuration information;

password information; and

face information for face recognition of the user.

27. The method of claim 26 wherein the first device updates some of the user-specific configuration information based on one or more human interactions.

28. The method of claim 27 wherein said first device provides updated user-specific configuration information to another location.

29. The method of claim 28 further comprising:

(C) at said other location, associating updated user-specific configuration information received from said first device with said user.

30. The method of claim 29 further comprising:

(D) associating a second device with said user; and

(E) automatically configuring said second device with said updated user-specific configuration information.

31. The method of any of claims 24 – 30 wherein the first device is a sound rendering device.

32. The method of claim 30 wherein the first device and the second device are sound rendering devices.

33. A framework supporting operation of multiple devices on behalf of a plurality of users, wherein each device of said multiple devices is configured to be primarily associated in the framework with one user of said plurality of users, said framework comprising:

a backend system comprising hardware and software, including at least one processor and memory, said backend system including: a database system, and backend applications, said backend applications running on said hardware and configured:

(a) to interface with said database system and with said multiple devices, and

(b) to maintain, in said database system, device information about each device of said multiple devices, said device information about each particular device including information about any user associated with said particular device, and

(c) to maintain, in said database system, user information about each user of said plurality of users, said user information for each particular user including information supporting at least one human-interface control mechanism in devices associated with that particular user; and

(d) to provide to at least one device associated with a specific user of said plurality of users, at least some of said specific user's user information from said database system.

34. The framework of claim 33 wherein the user information for each particular user further includes configuration information associated with the particular user, and wherein said specific user's user information provided by said backend system in (d) includes configuration information associated with the specific user.

35. The framework of claim 33 wherein said specific user's user information provided by said backend system includes said information supporting said at least one human-interface control mechanism in devices associated with the specific user.

36. The framework of any one of claims 33 to 35 wherein the at least one human-interface control mechanism includes a speech recognition mechanism, and wherein, for a specific user, the information supporting said at

least one human-interface control mechanism comprises at least one speech corpus associated with the specific user and usable by a speech recognition mechanism to support recognition of speech by said specific user.

37. The framework of any one of claims 33 to 35 wherein the at least one human-interface control mechanism includes a gesture recognition mechanism, and wherein, for a specific user, the information supporting said at least one human-interface control mechanism comprises at least one gesture corpus associated with the specific user and usable by a gesture recognition mechanism to support recognition of gestures by said specific user.

38. The framework of claim 33 wherein the backend applications are further configured:

- (e) to obtain updated user information from devices; and
- (f) to associate, in said database system, said updated user information with corresponding users.

39. The framework of claim 38 wherein the updated user information from a particular device comprises updated information supporting a human-interface control mechanism on the particular device.

40. The framework of claim 39 wherein the backend applications associate, in (f), the updated information supporting the human-interface control mechanism on the particular device with a user associated with the particular device.

41. The framework of claim 39 wherein the human-interface control mechanism on the particular device comprises a speech recognition mechanism, and wherein the updated user information comprises an updated speech corpus for the speech recognition mechanism.

42. The framework of claim 41 wherein the backend applications associate, in (f), the updated speech corpus for the speech recognition mechanism on the particular device with a user associated with the particular device.

43. A method, operable in a framework supporting operation of multiple devices on behalf of a plurality of users, wherein each device of said multiple devices is configured to be primarily associated in the framework with one user of said plurality of users, said framework comprising: a backend system comprising hardware and software, including at least one processor and memory, said backend system including: a database system, and backend applications running on said hardware and configured to interface with said database system and with said multiple devices,

the method comprising:

(A) maintaining, in said database system, device information about each device of said multiple devices, said device information about each particular device including information about any user associated with said particular device;

(B) maintaining, in said database system, user information about each user of said plurality of users, said user information for each particular user including information supporting at least one human-interface control mechanism in devices associated with that particular user; and

(C) providing, to at least one device associated with a specific user of said plurality of users, at least some of said specific user's user information from said database system.

44. The method of claim 43 wherein the user information for each particular user further includes configuration information associated with the particular user, and wherein said specific user's user information provided in (C) includes configuration information associated with the specific user.

45. The method of claim 43 wherein said specific user's user information provided in (C) includes said information supporting said at least one human-interface control mechanism in devices associated with the specific user.

46. The method of any one of claims 43 to 45 wherein the at least one human-interface control mechanism includes a speech recognition mechanism, and wherein, for a specific user, the information supporting said at least one human-interface control mechanism comprises at least one speech corpus associated with the specific user and usable by a speech recognition mechanism to support recognition of speech by said specific user.

47. The method of any one of claims 33 to 35 wherein the at least one human-interface control mechanism includes a gesture recognition mechanism, and wherein, for a specific user, the information supporting said at least one human-interface control mechanism comprises at least one gesture corpus associated with the specific user and usable by a gesture recognition mechanism to support recognition of gestures by said specific user.

48. The method of claim 43 further comprising, by said the backend applications:

- (D) obtaining updated user information from a particular device; and
- (E) associating, in said database system, said updated user information with a corresponding user of said particular device.

49. The method of claim 48 wherein the updated user information from the particular device comprises updated information supporting a human-interface control mechanism on the particular device.

50. The method of claim 49 wherein the human-interface control mechanism on the particular device comprises a speech recognition mechanism, and wherein the updated user information comprises an updated speech corpus for the speech recognition mechanism.

51. The method of claim 43 further comprising:

- (D) associating, in said database system, a new device with a user.

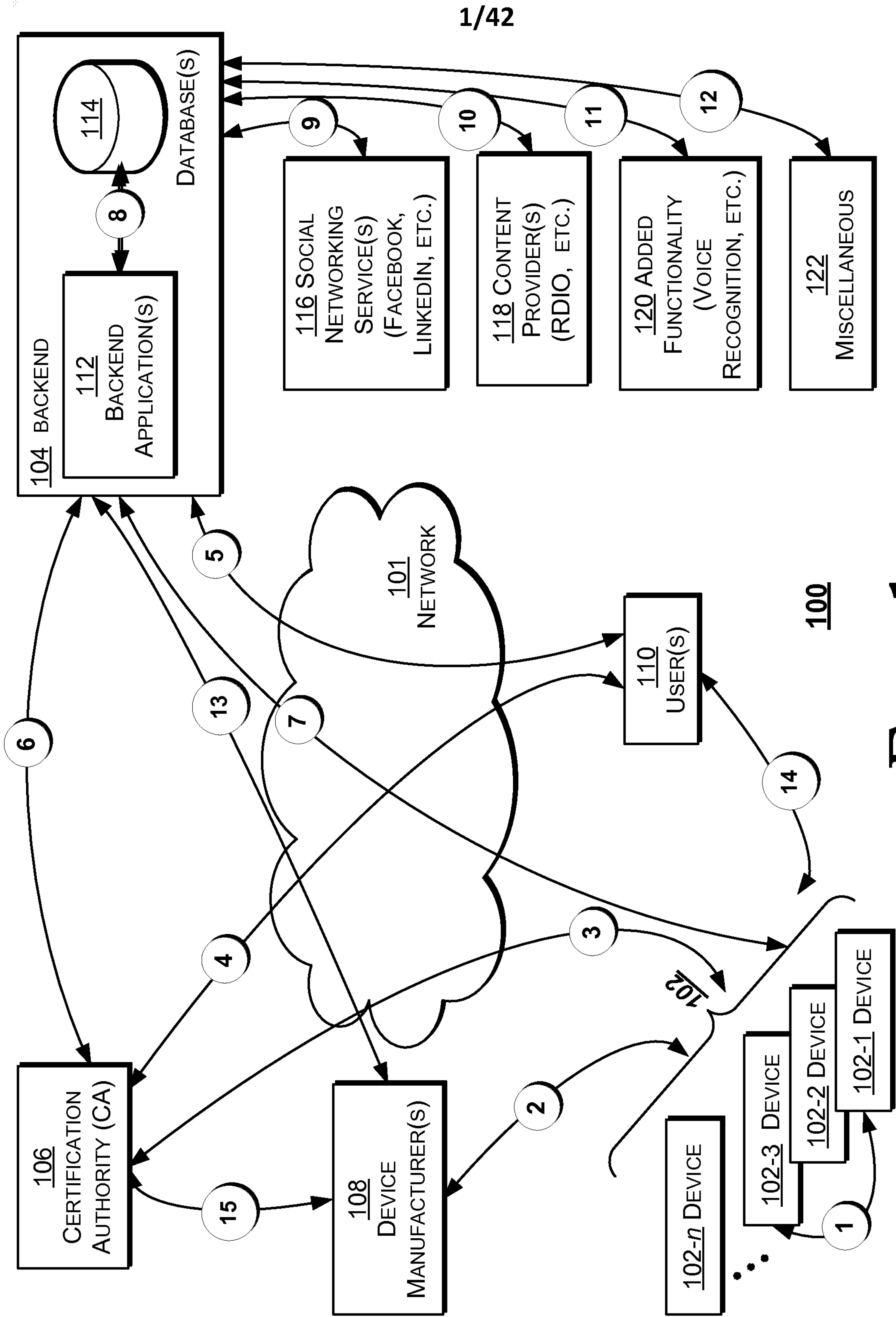


FIG. 1

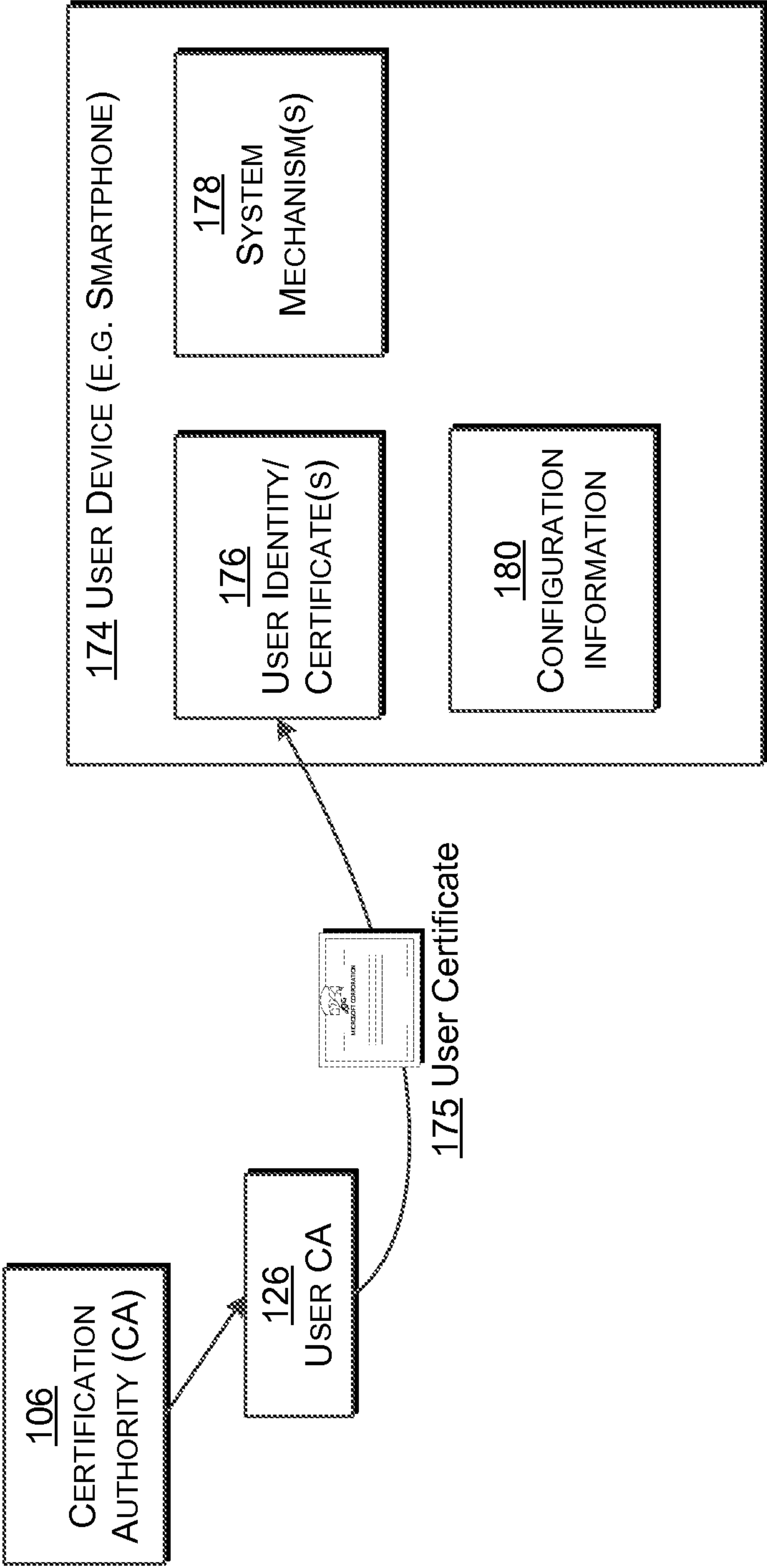


FIG. 2

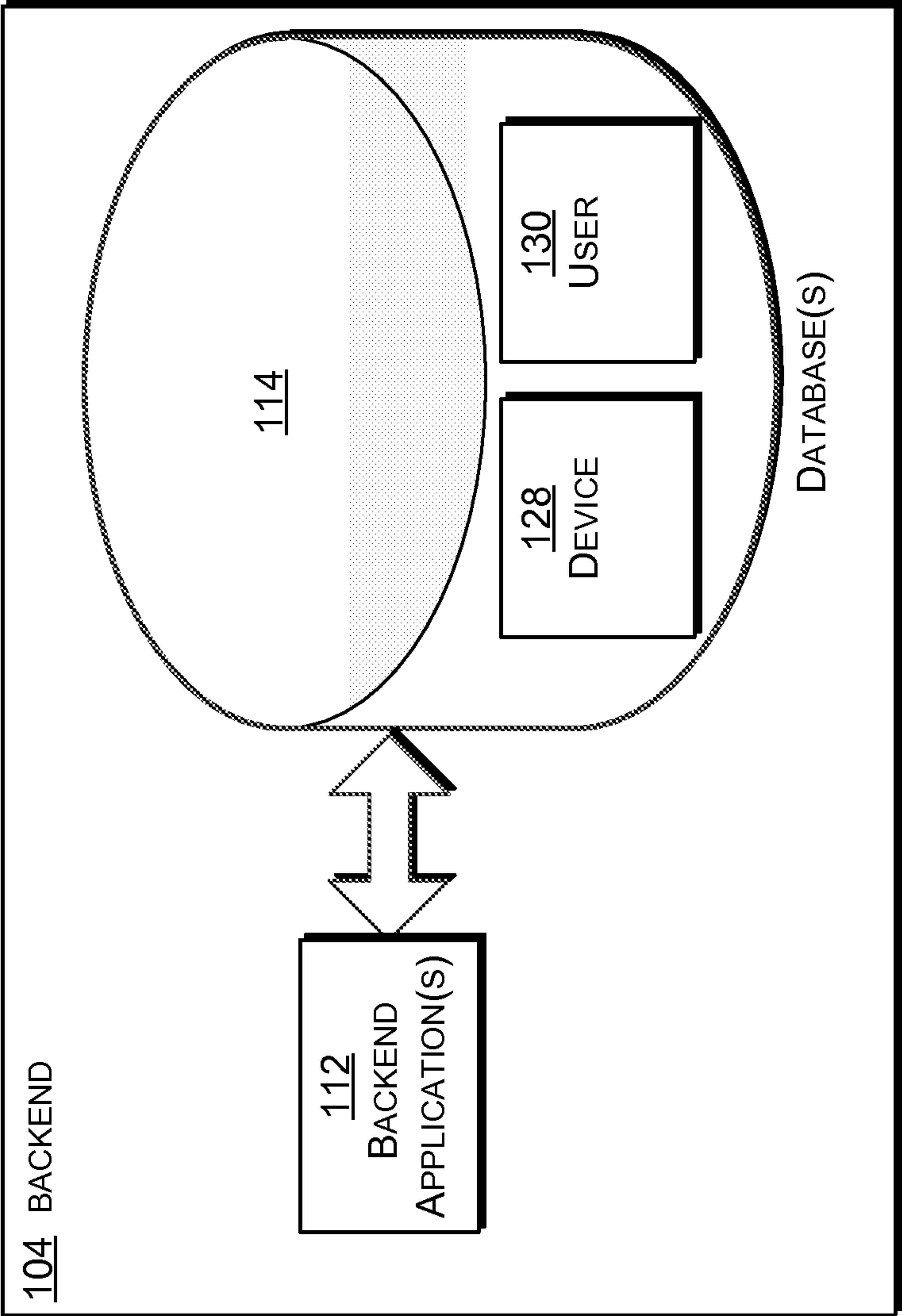


FIG. 3A

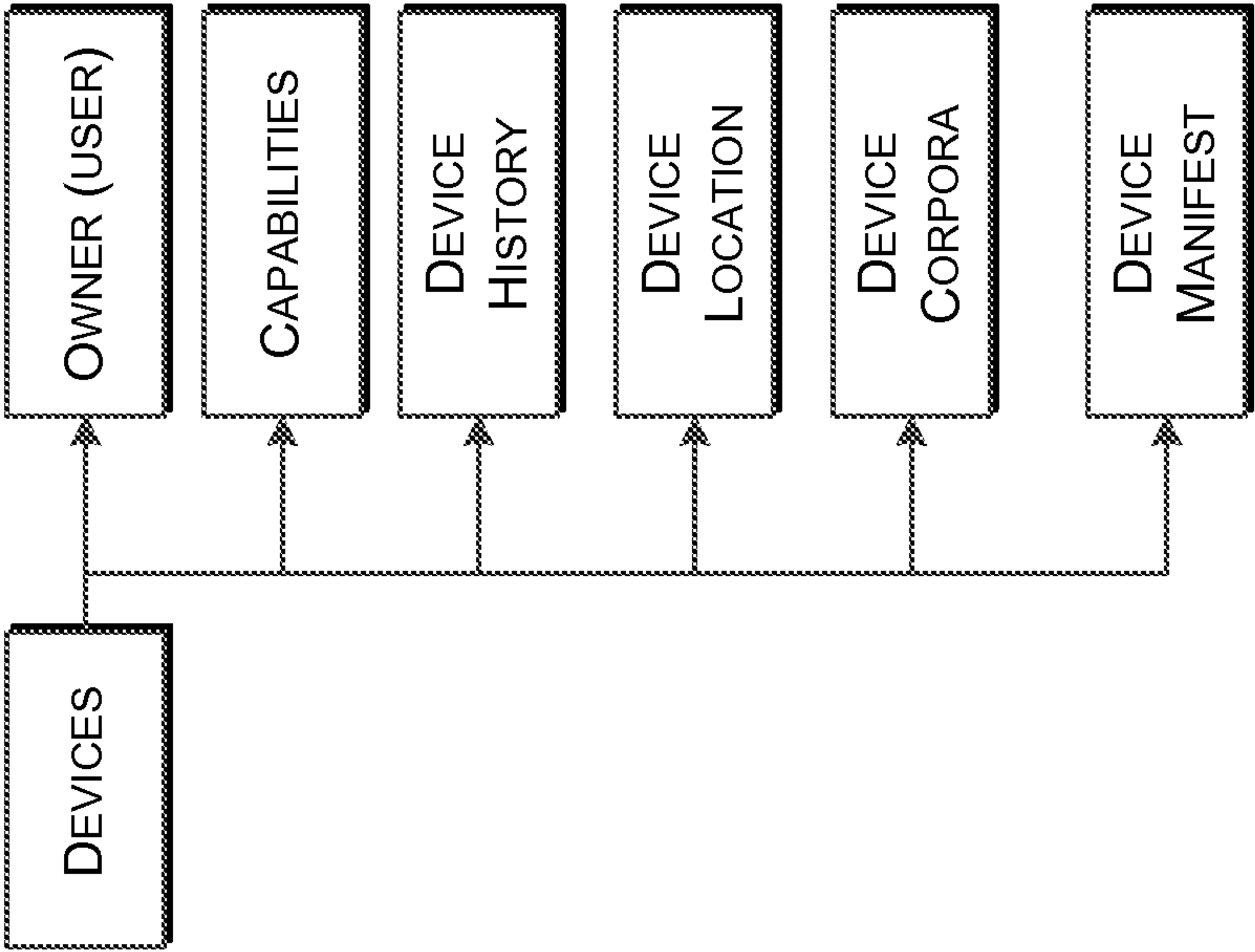


FIG. 3B

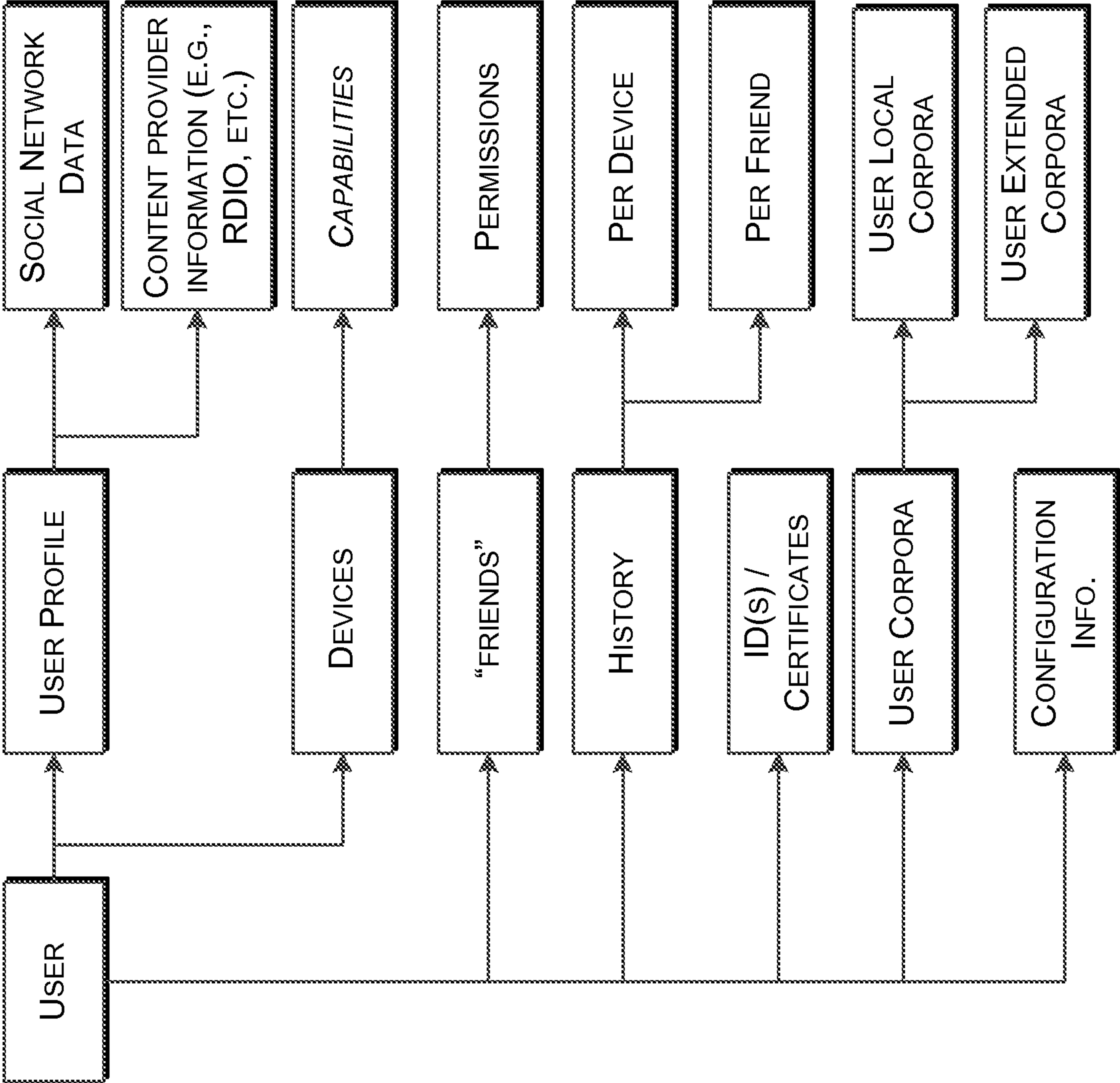


FIG. 3C

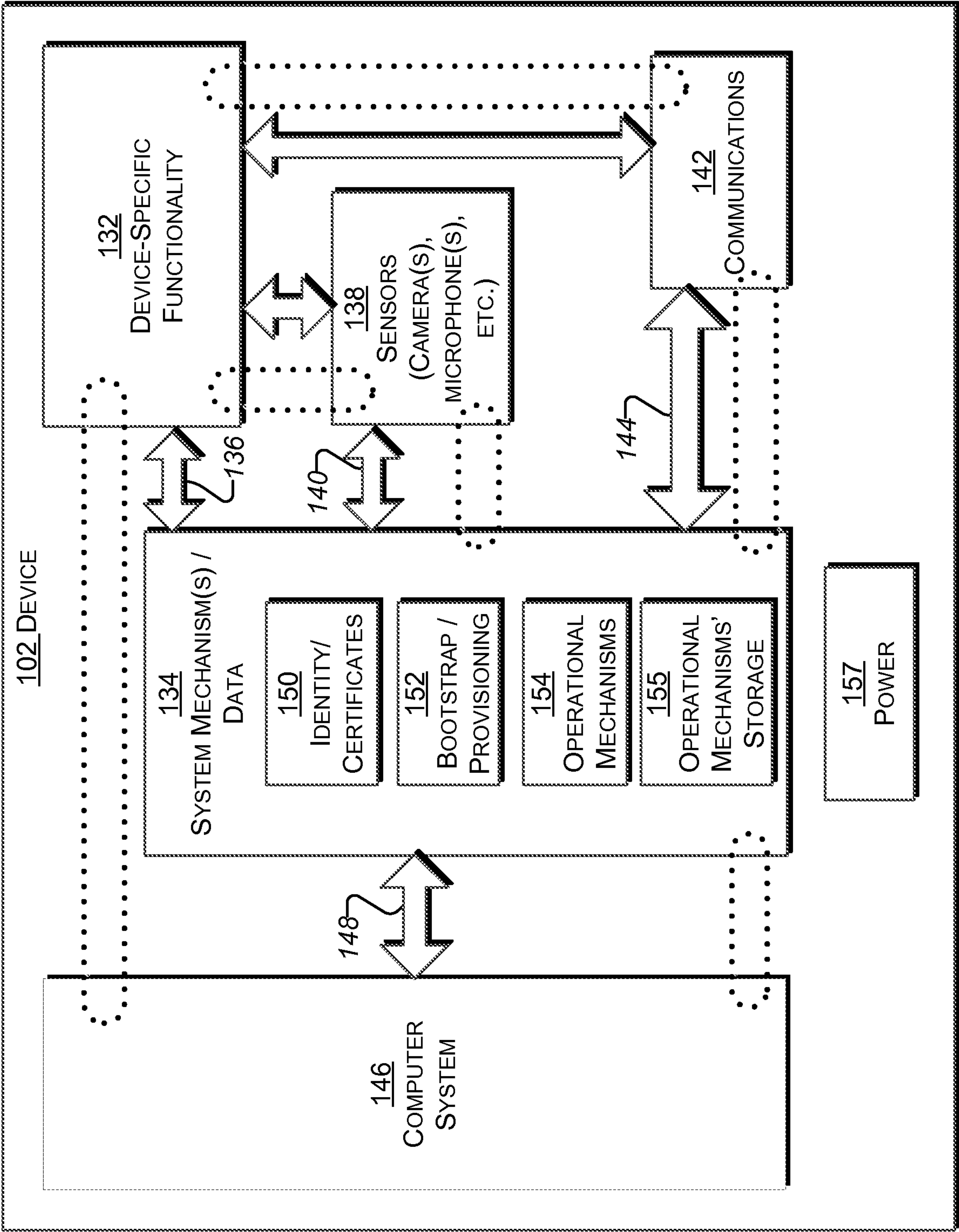


FIG. 4A

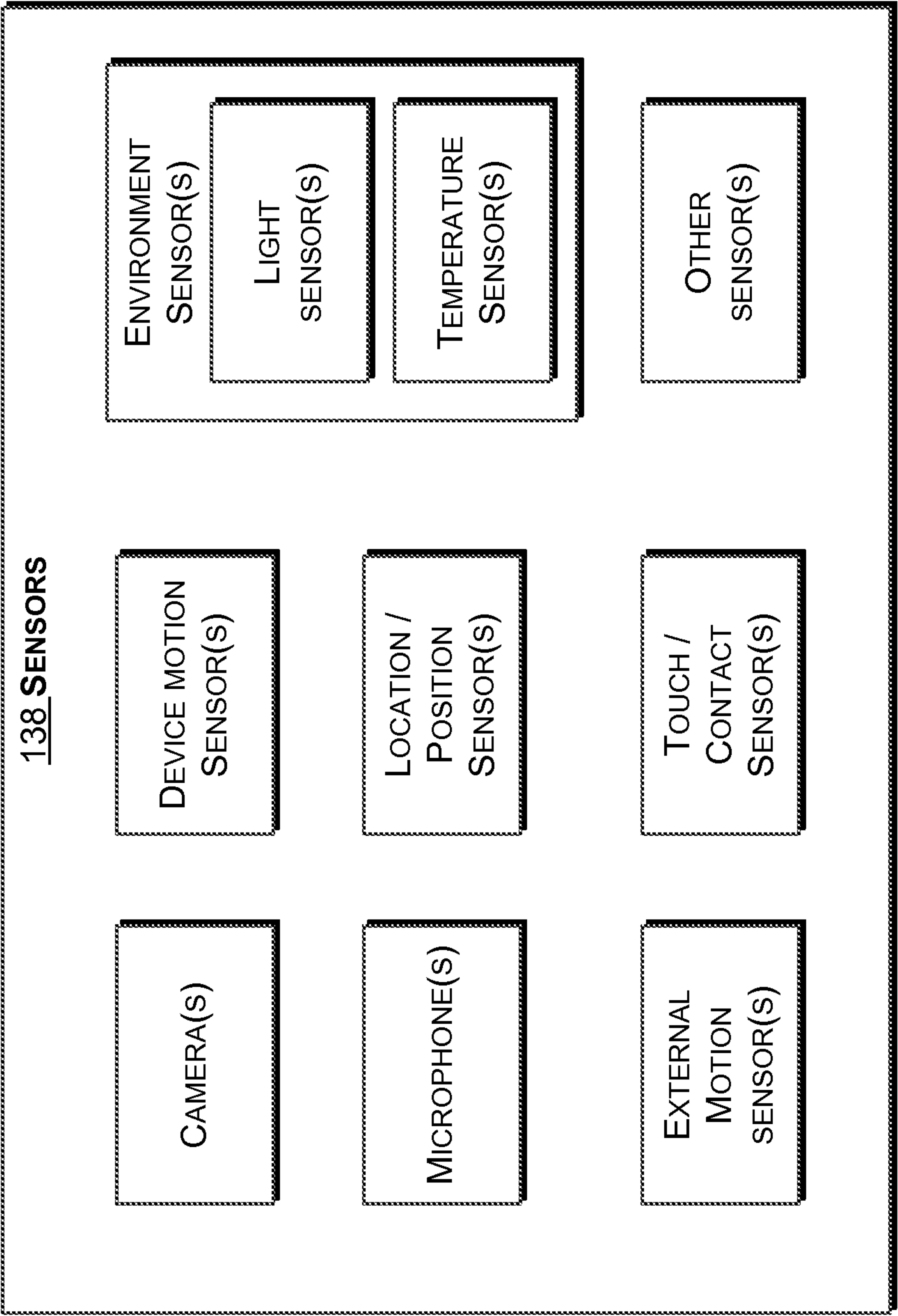


FIG. 4A-1

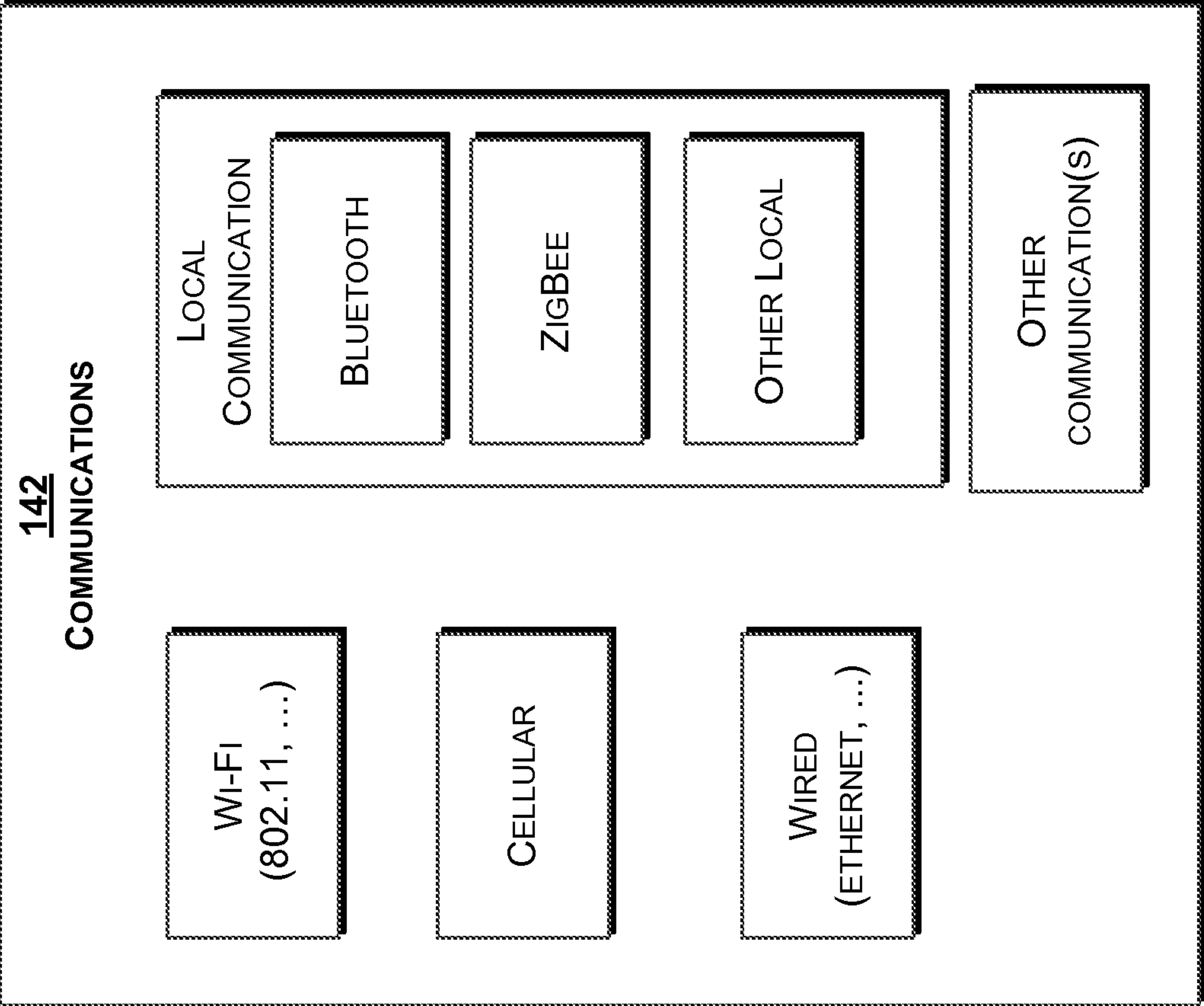


FIG. 4A-2

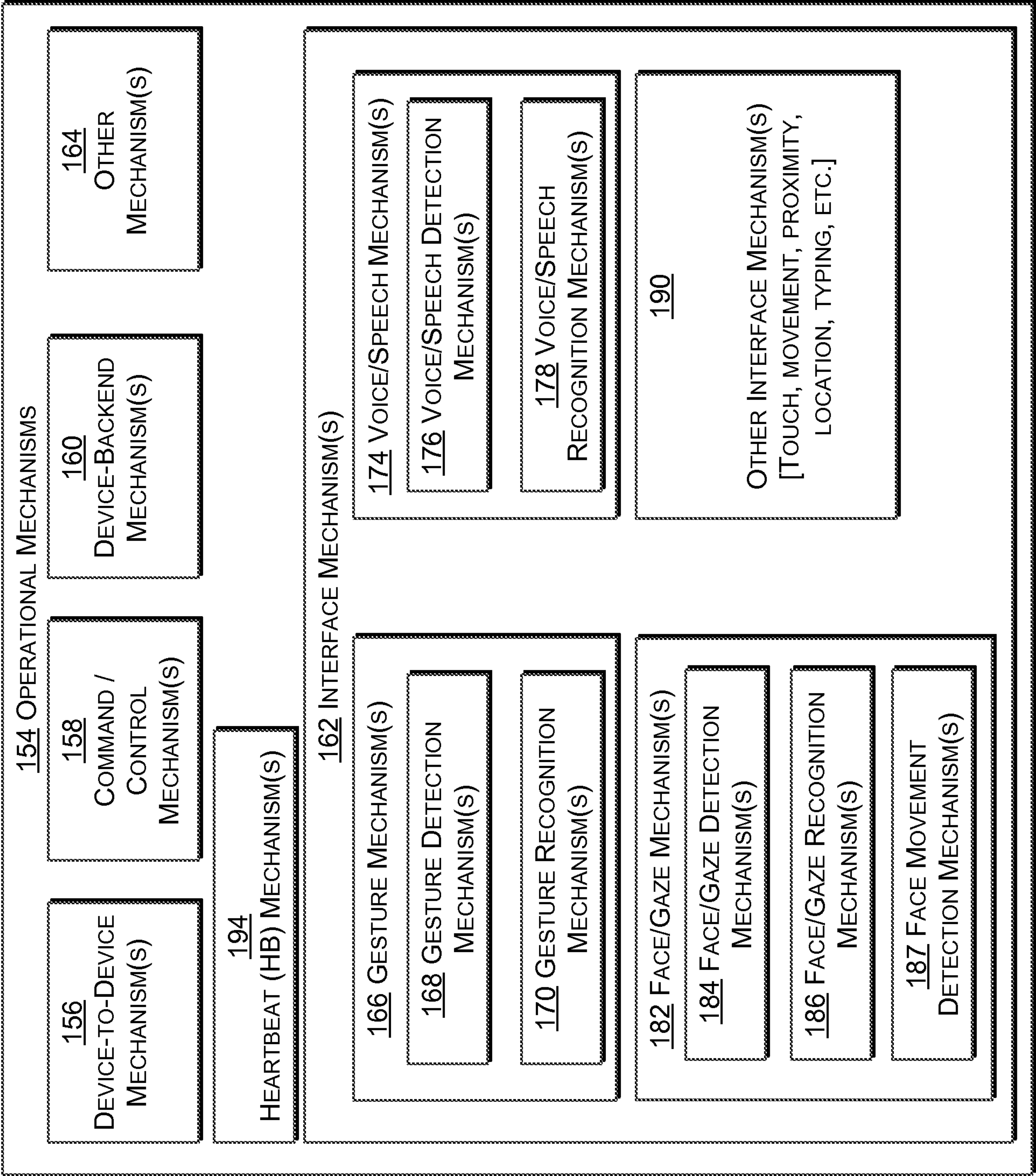


FIG. 4B

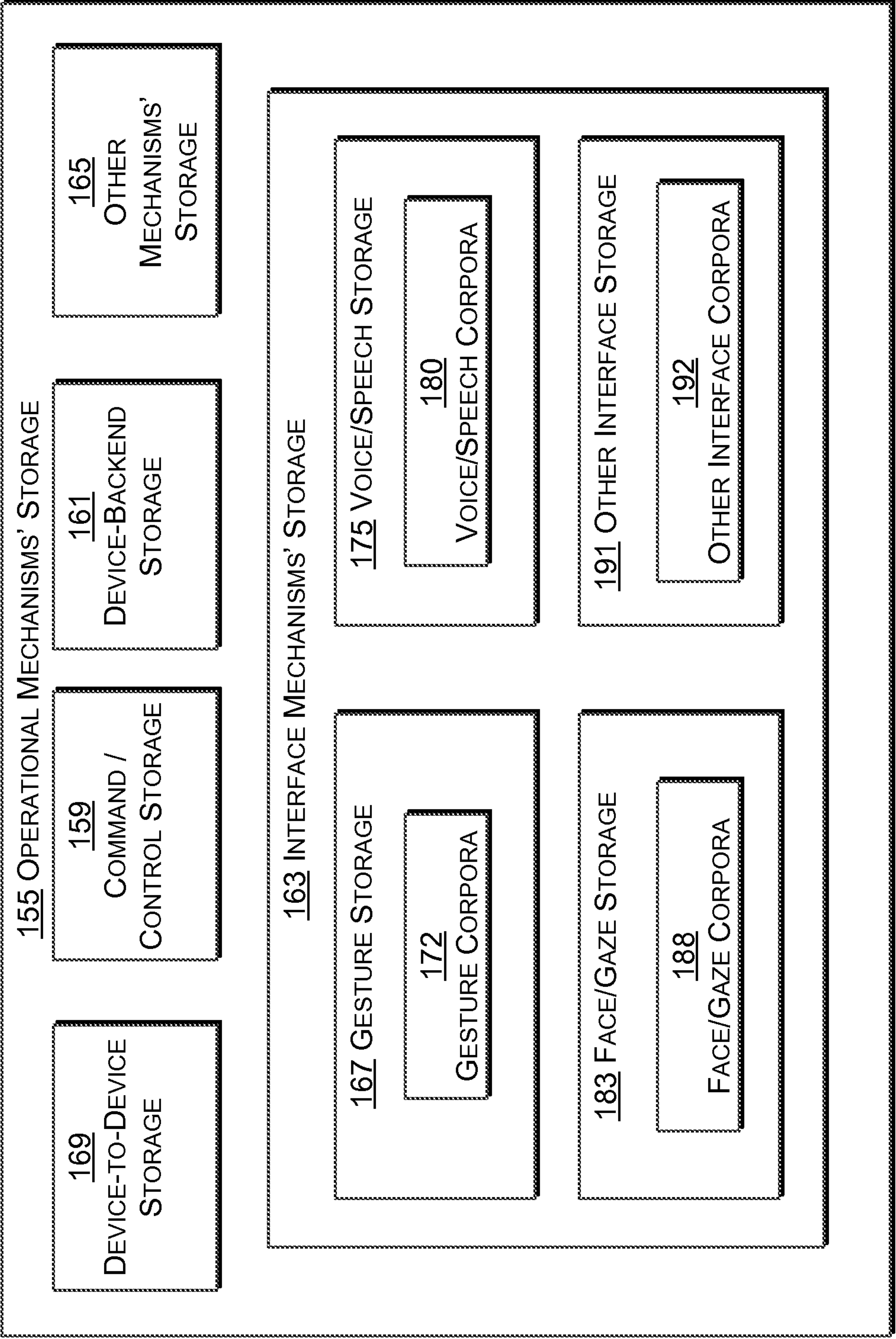


FIG. 4C

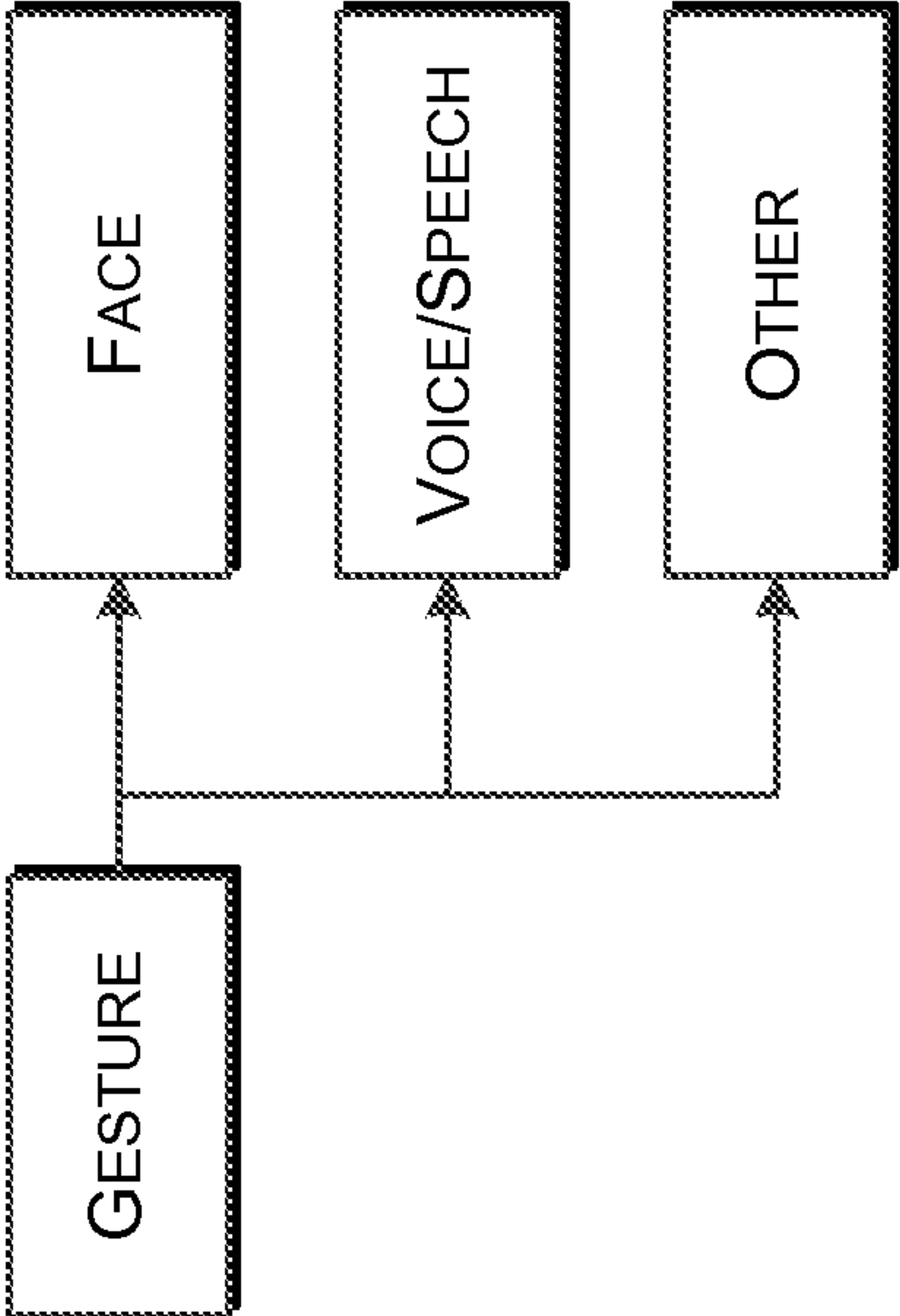


FIG. 4D

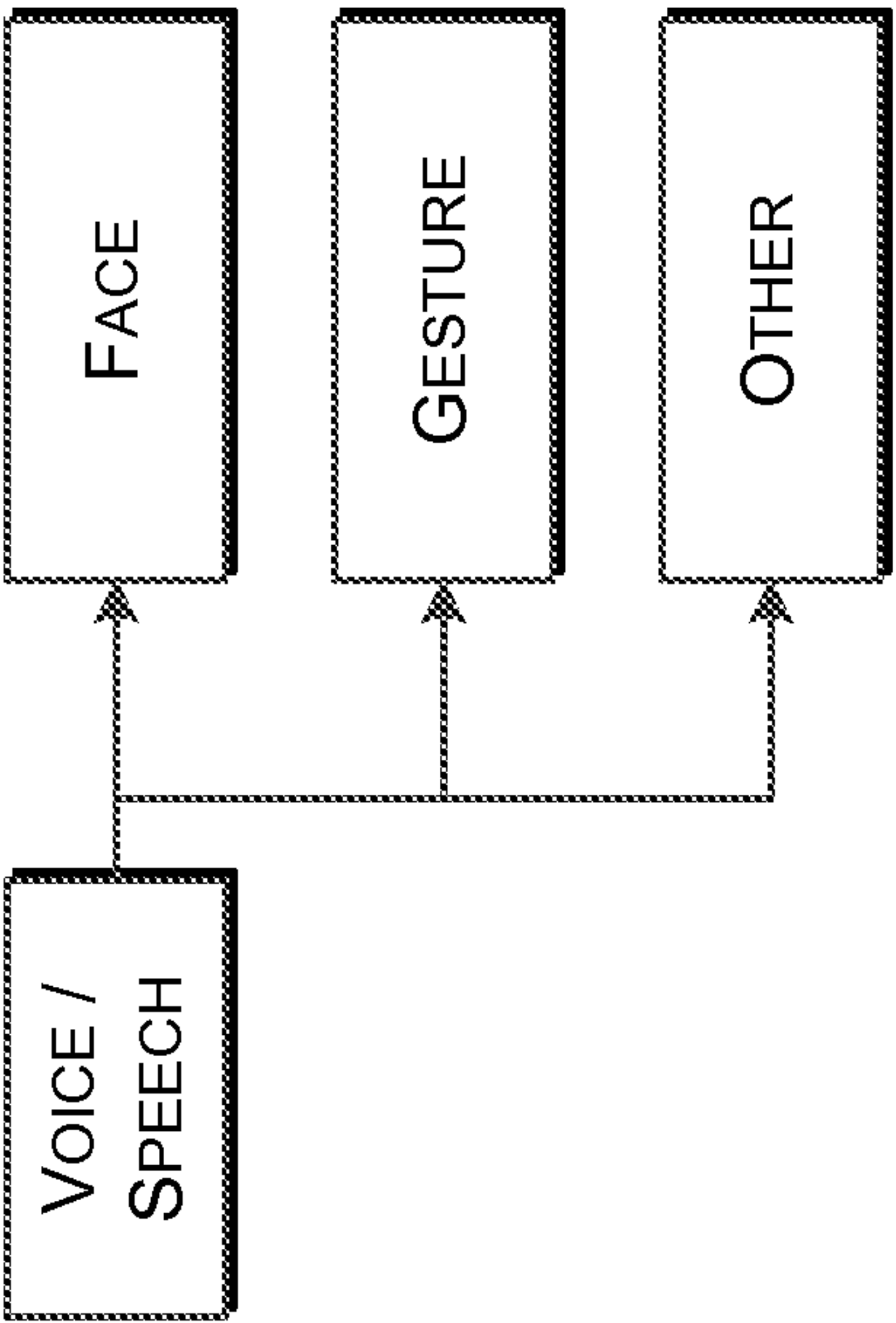


FIG. 4E

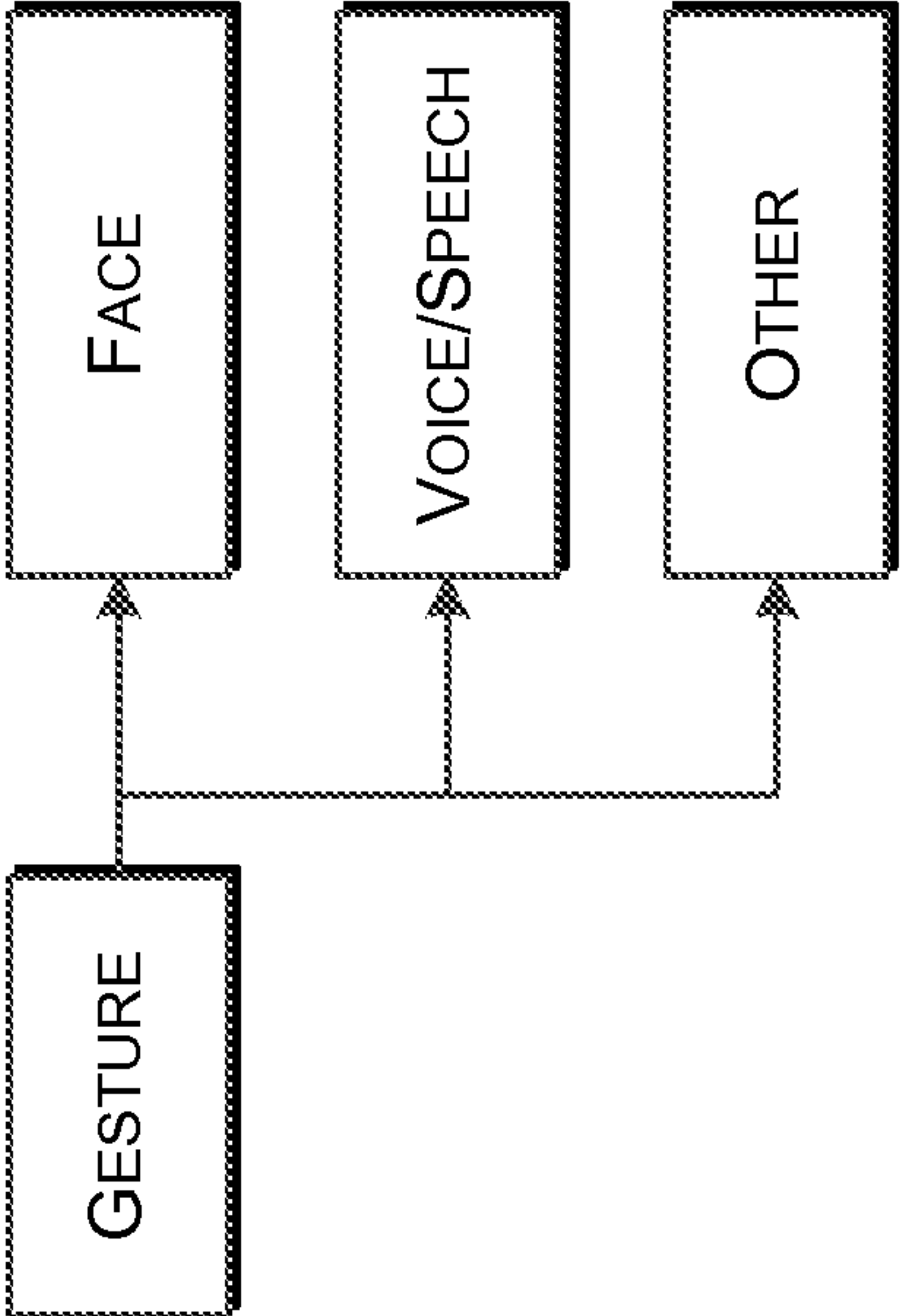


FIG. 4F

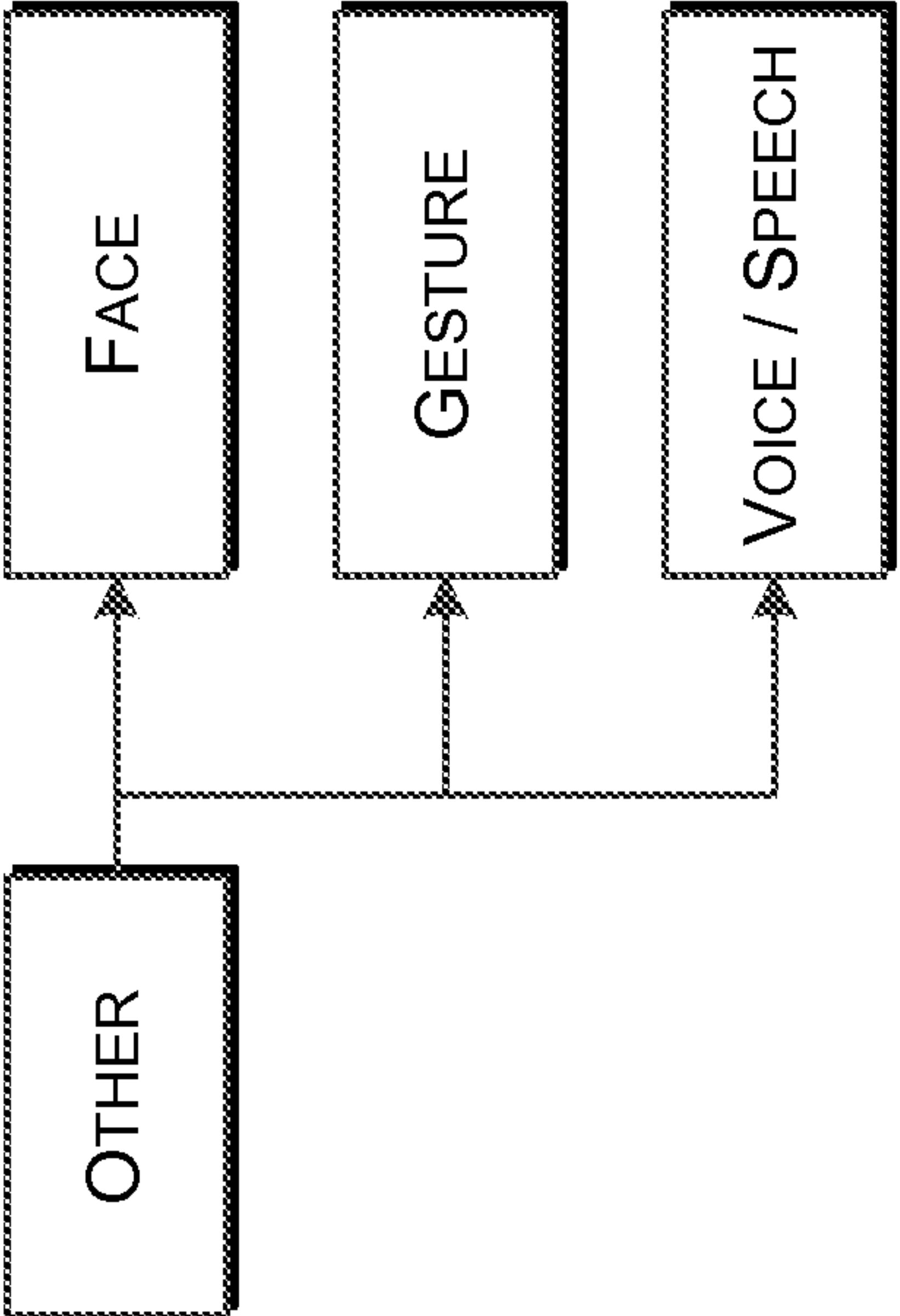


FIG. 4G

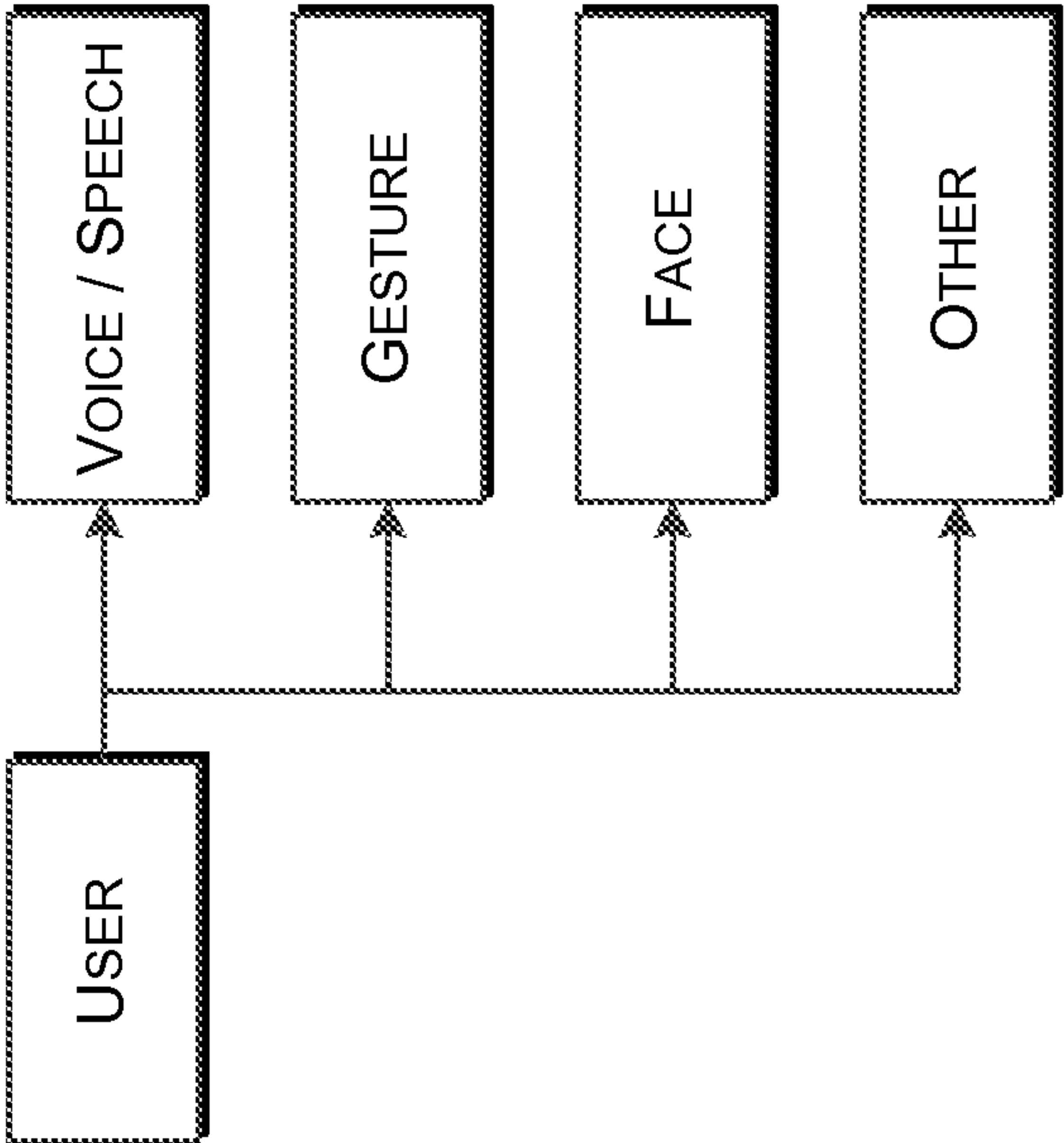


FIG. 4H

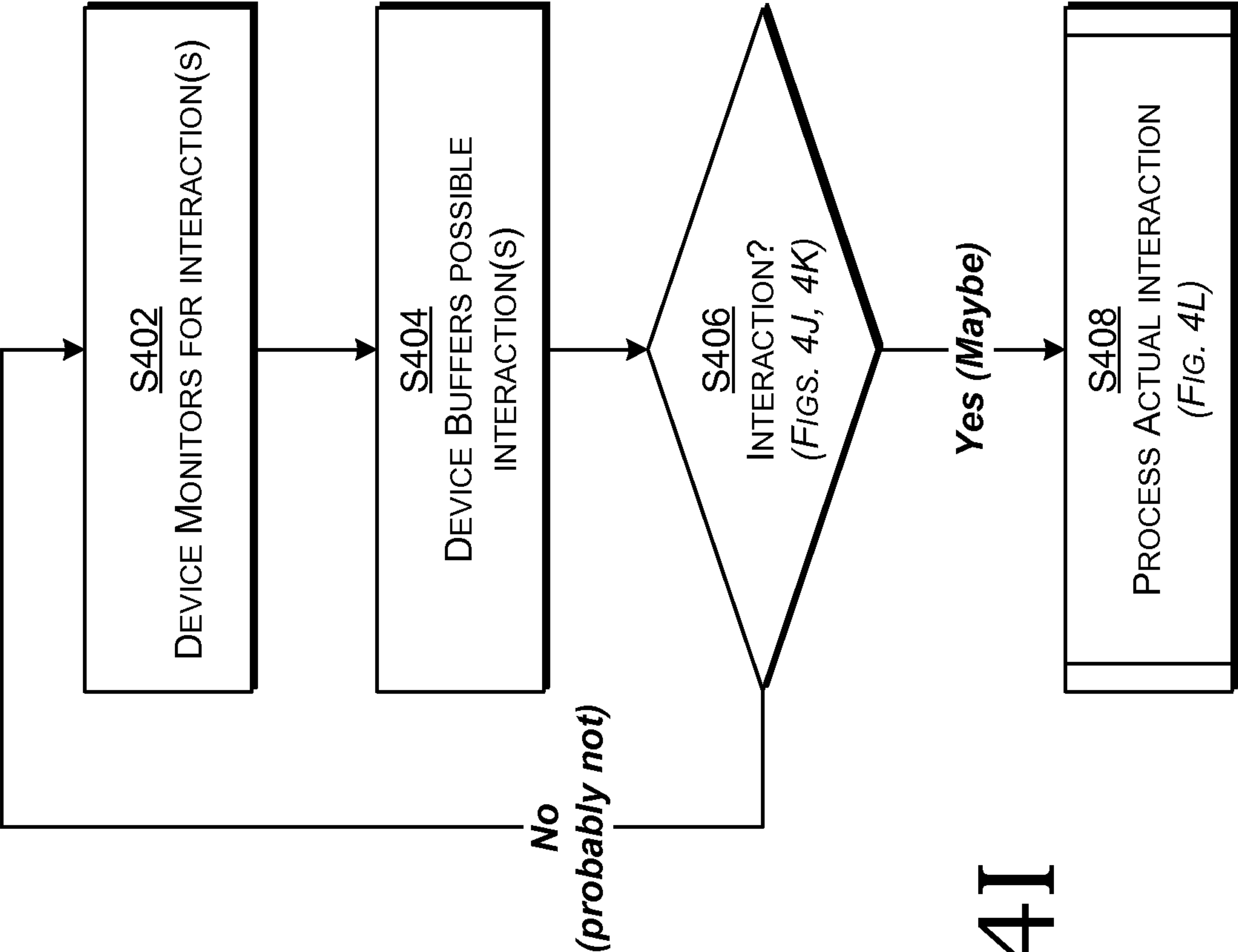
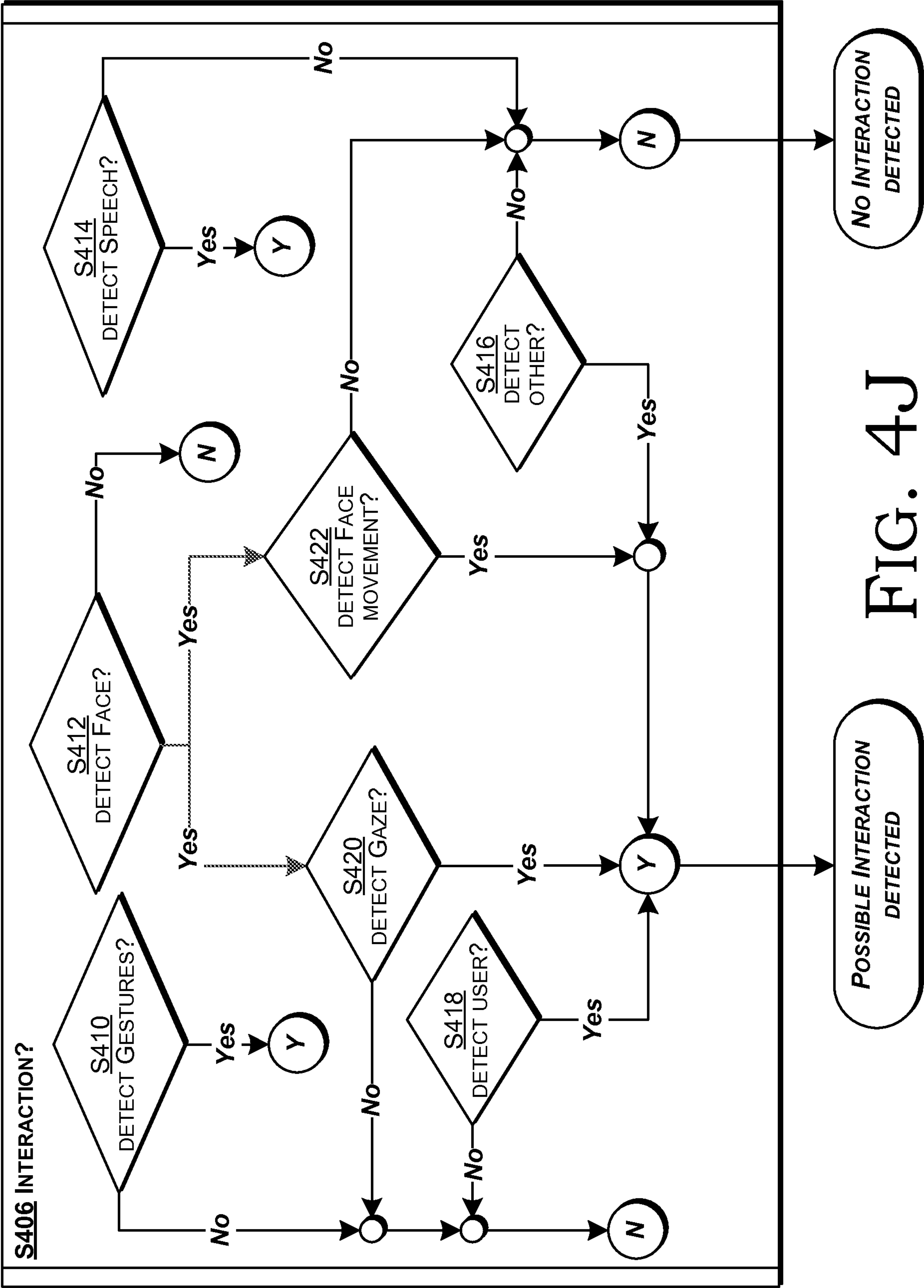


FIG. 4I



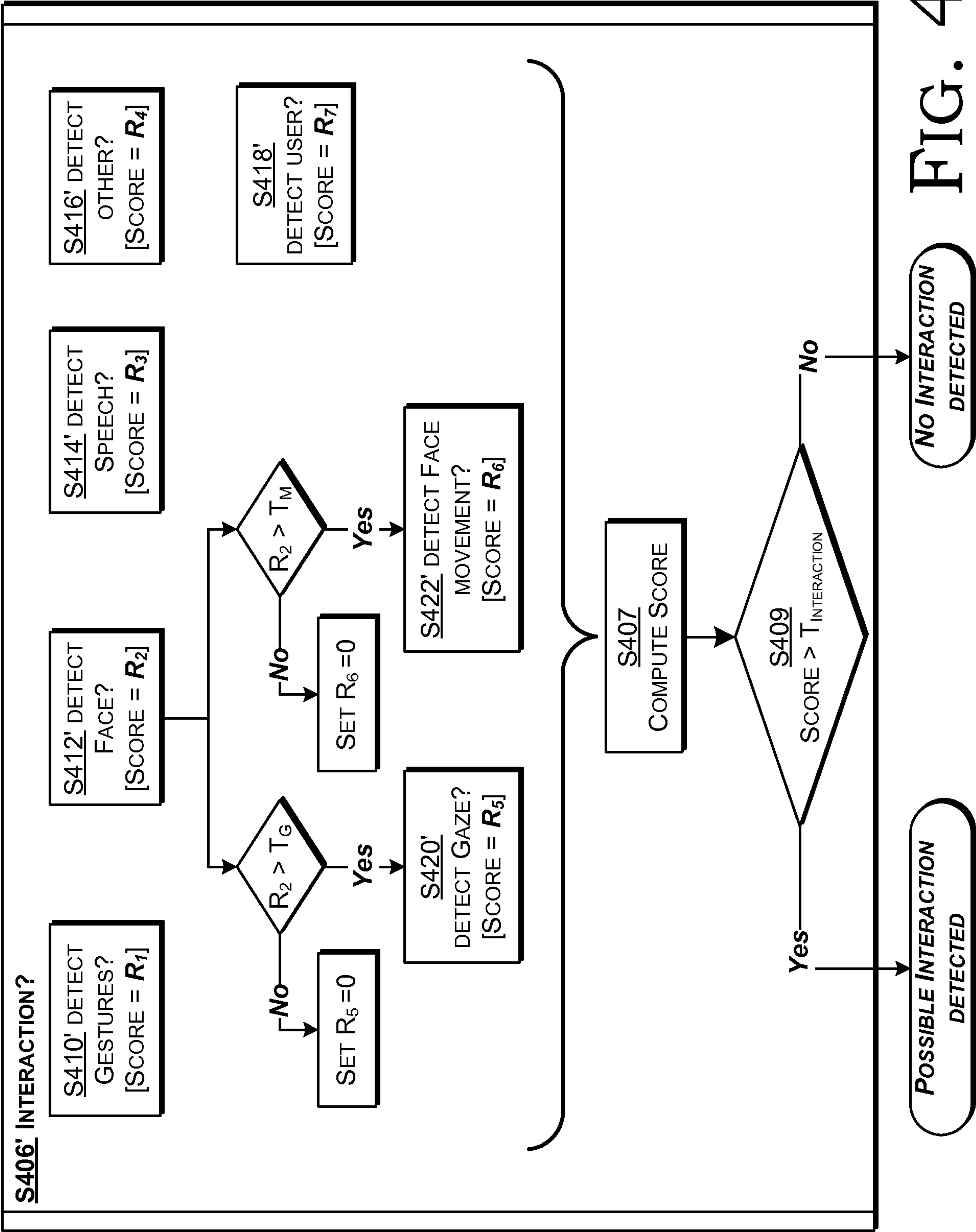


FIG. 4K

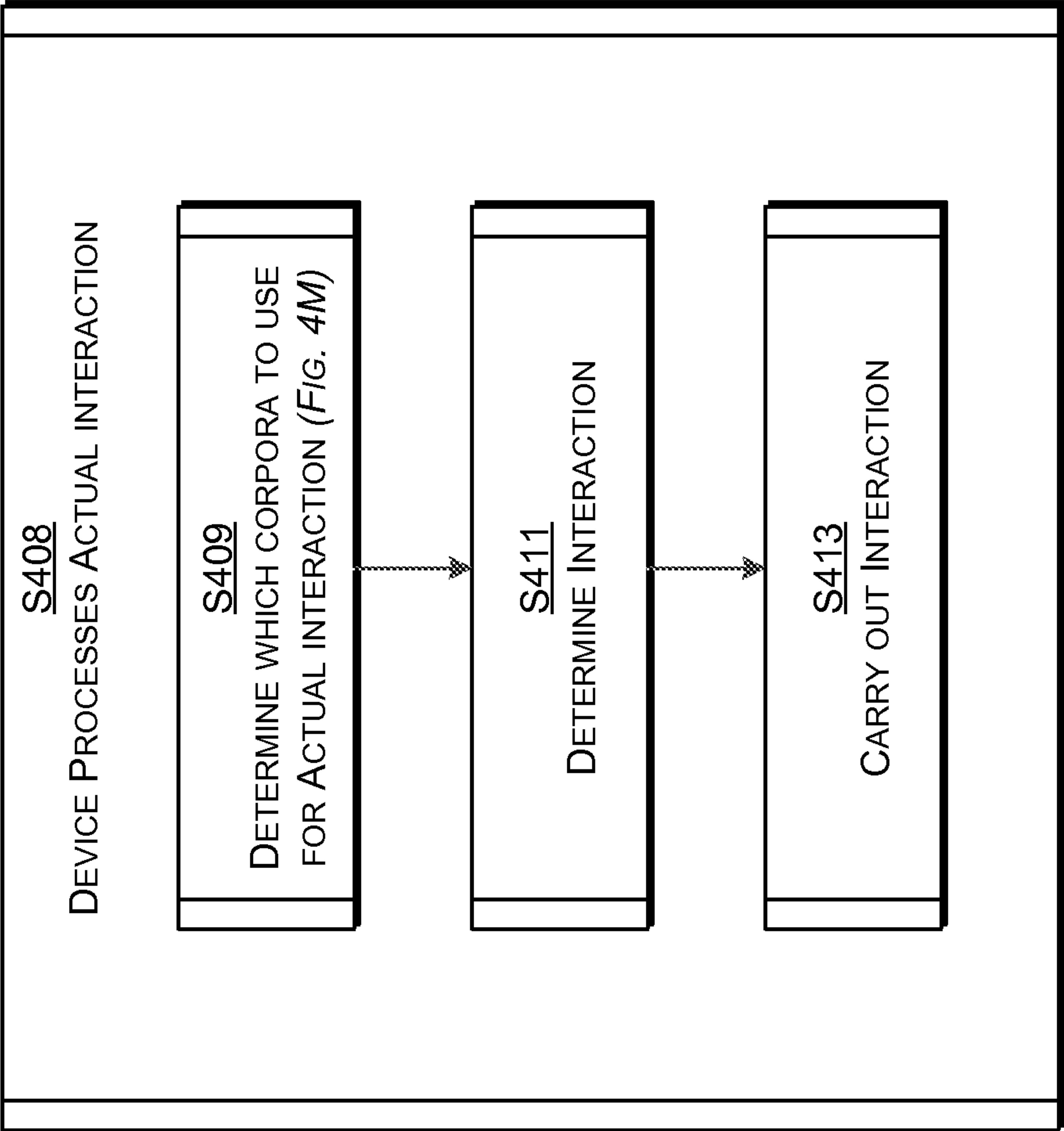


FIG. 4L

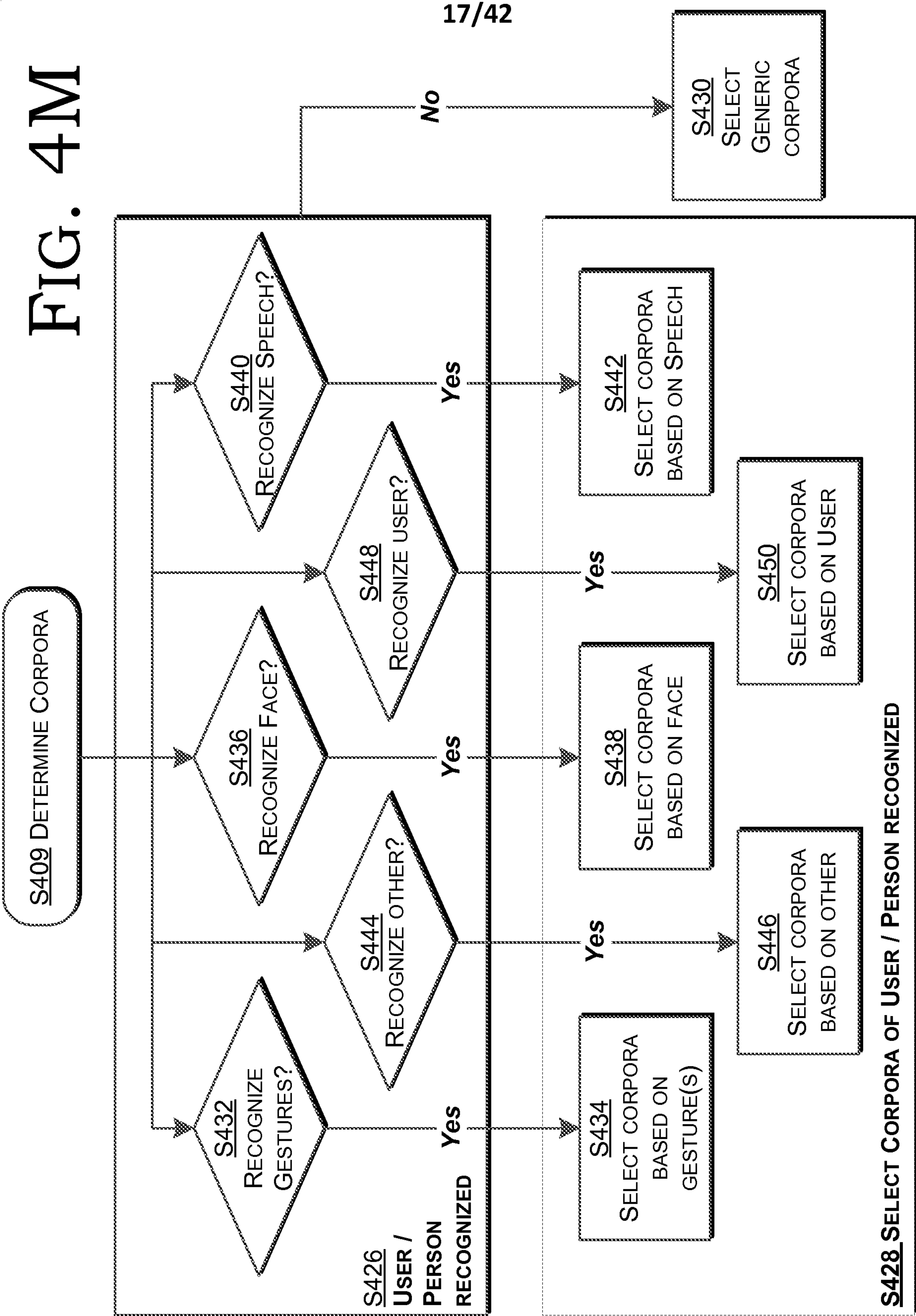
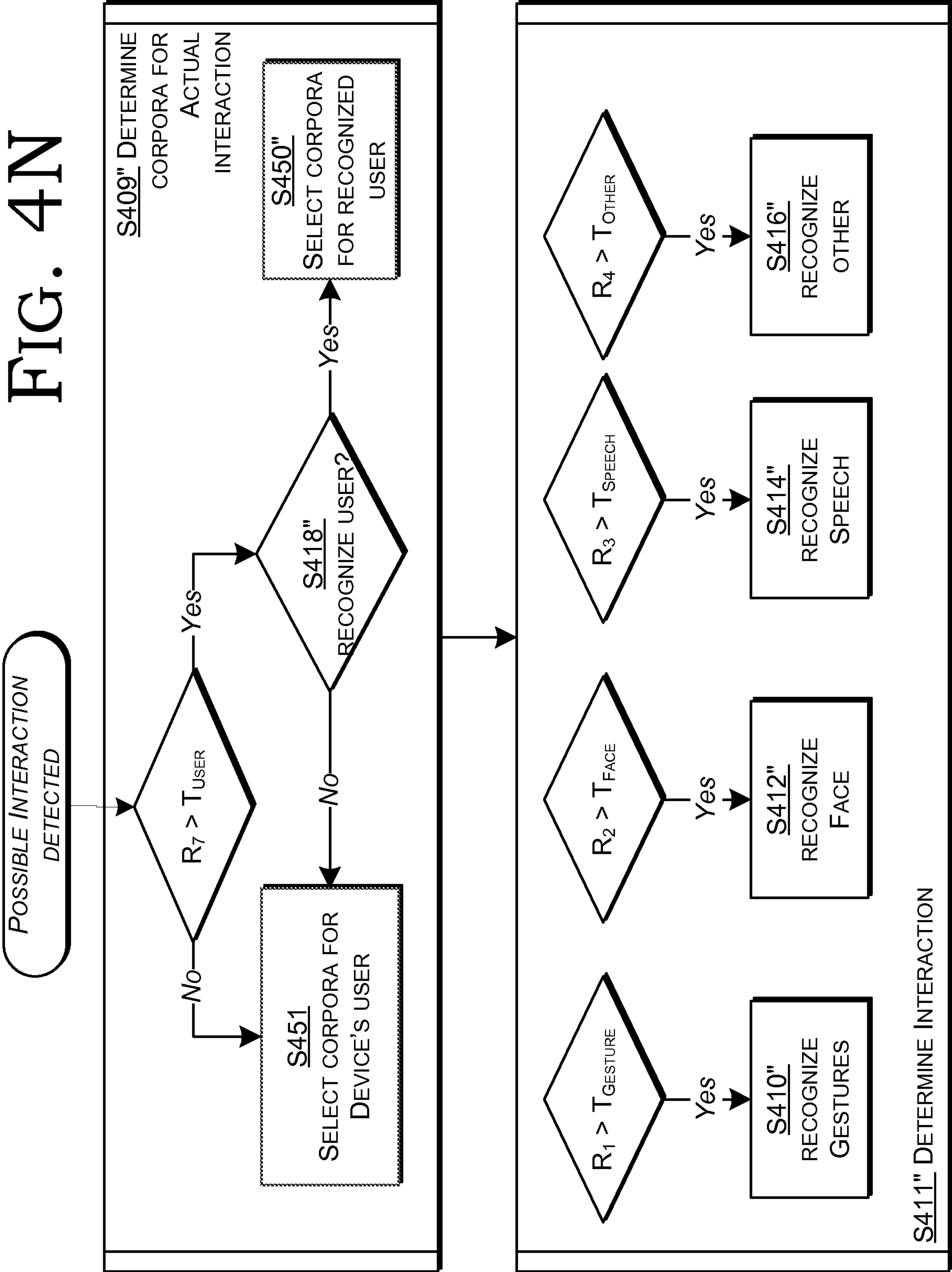


FIG. 4N



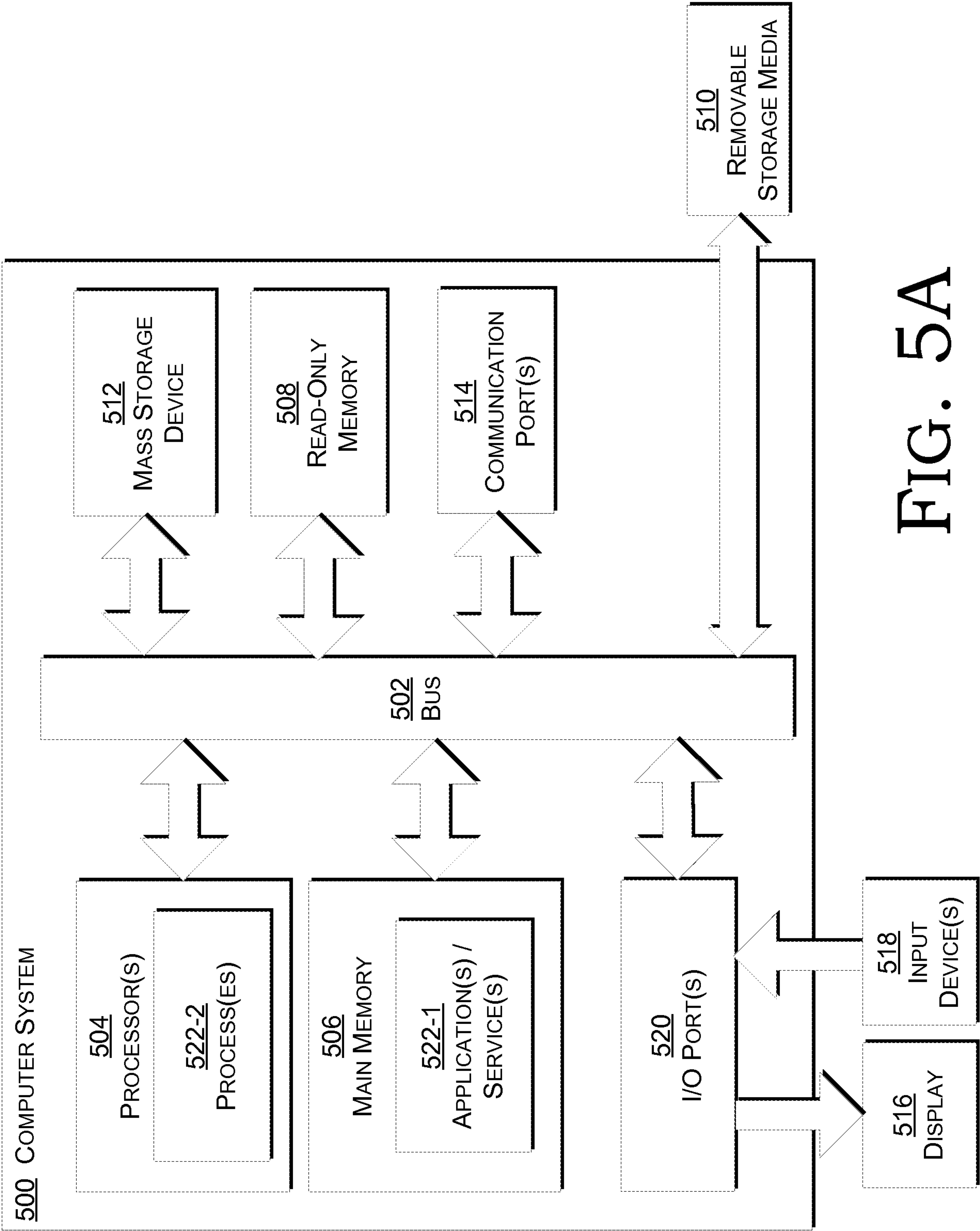


FIG. 5A

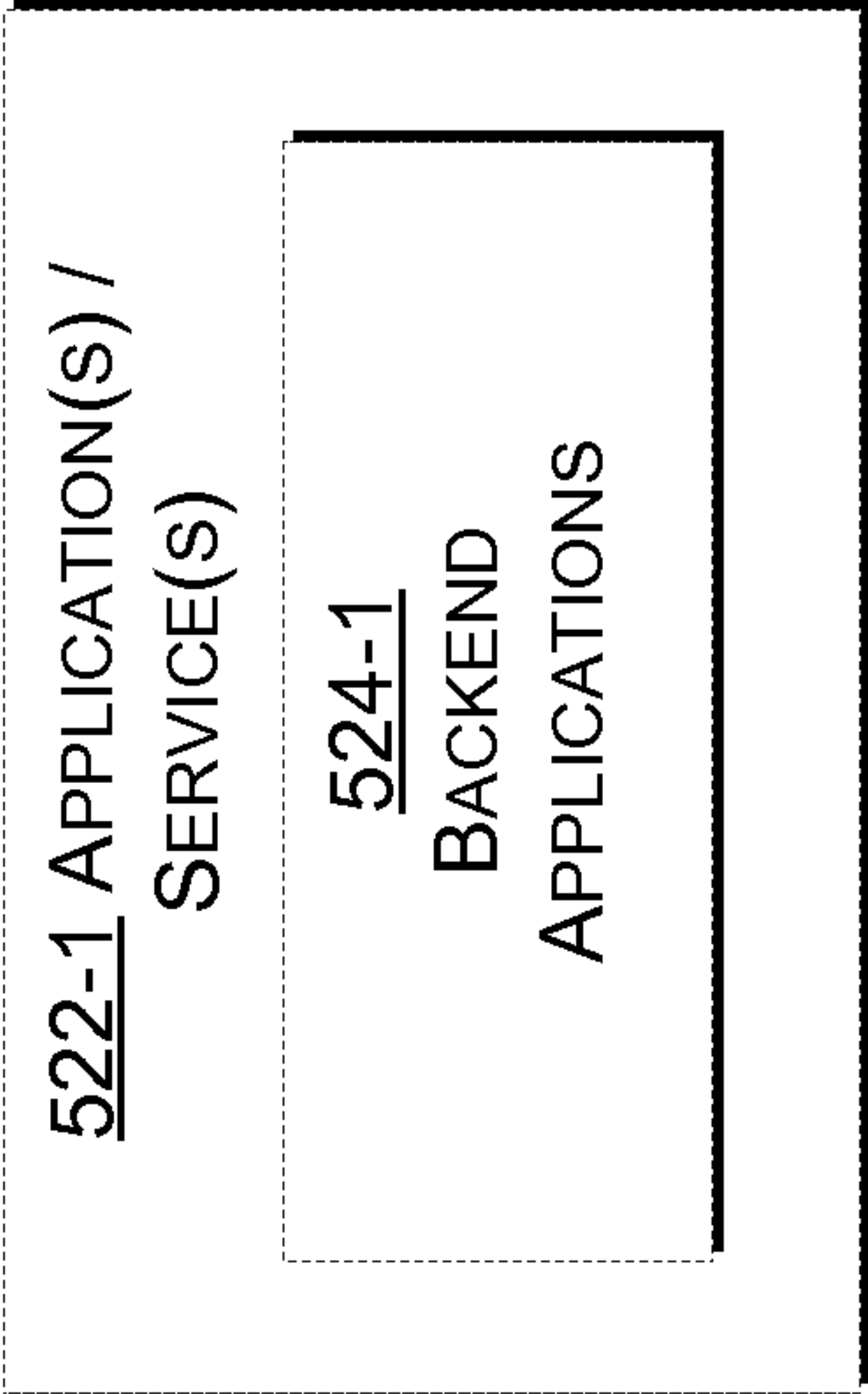


FIG. 5B

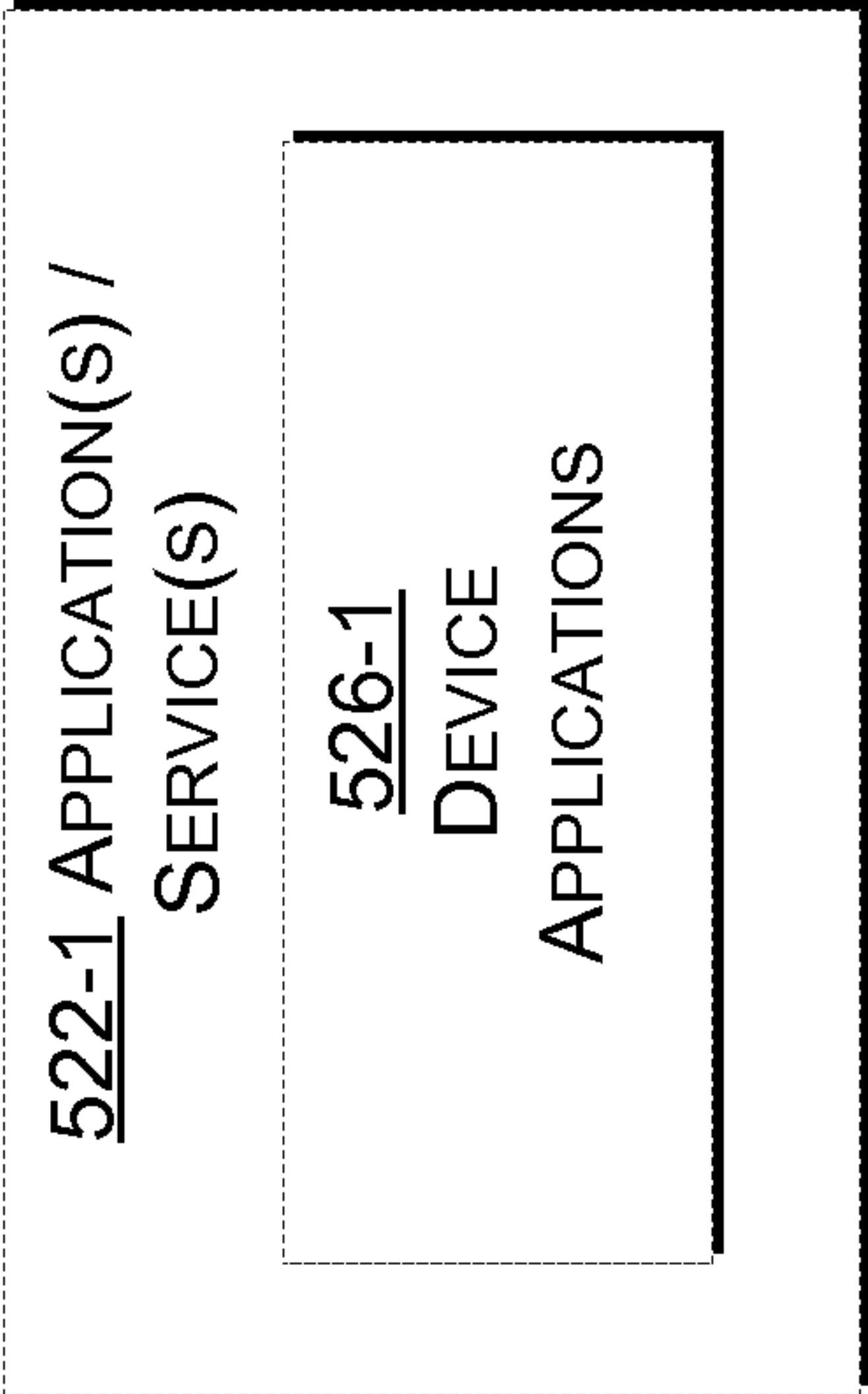


FIG. 5C

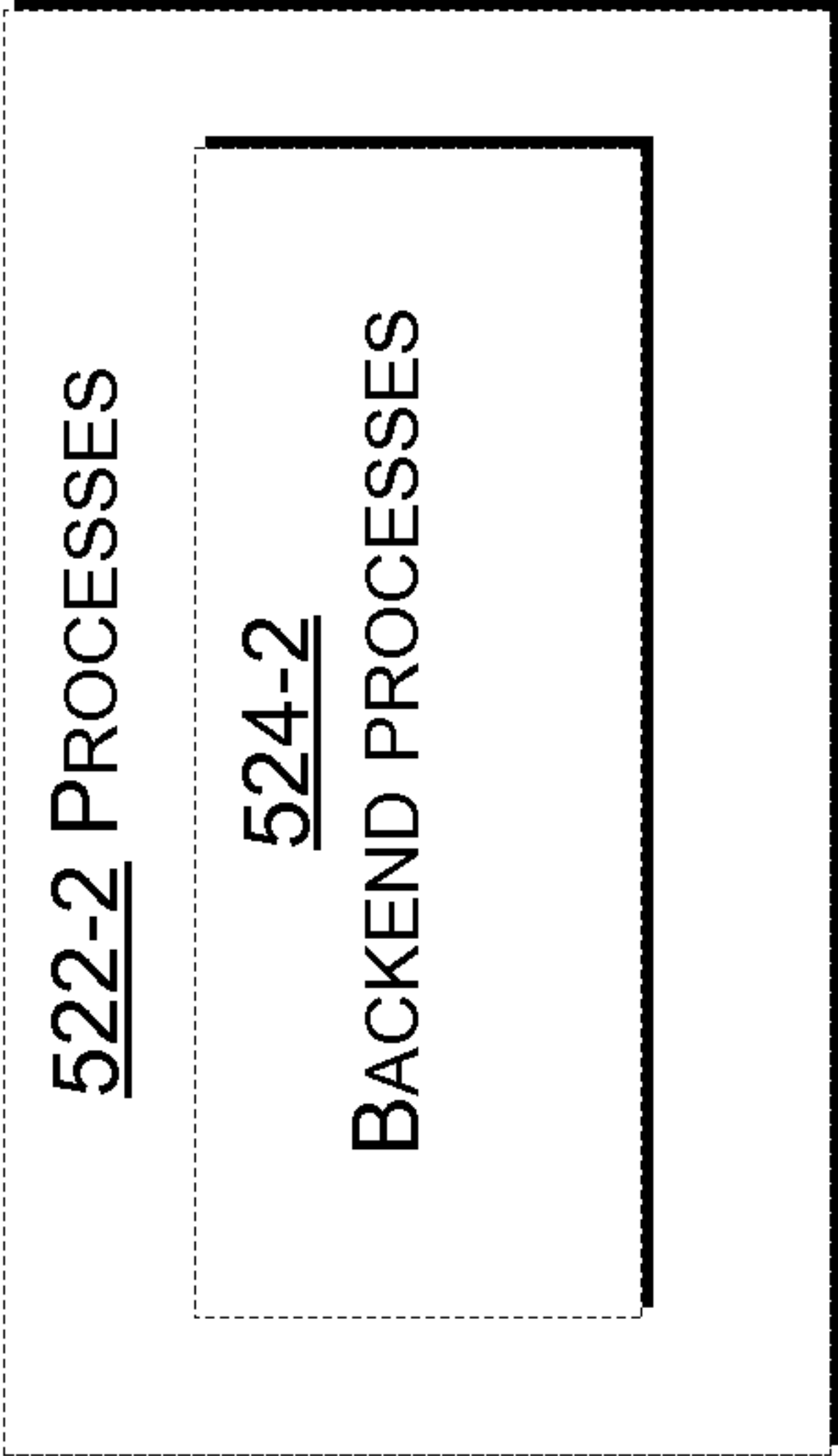


FIG. 5D

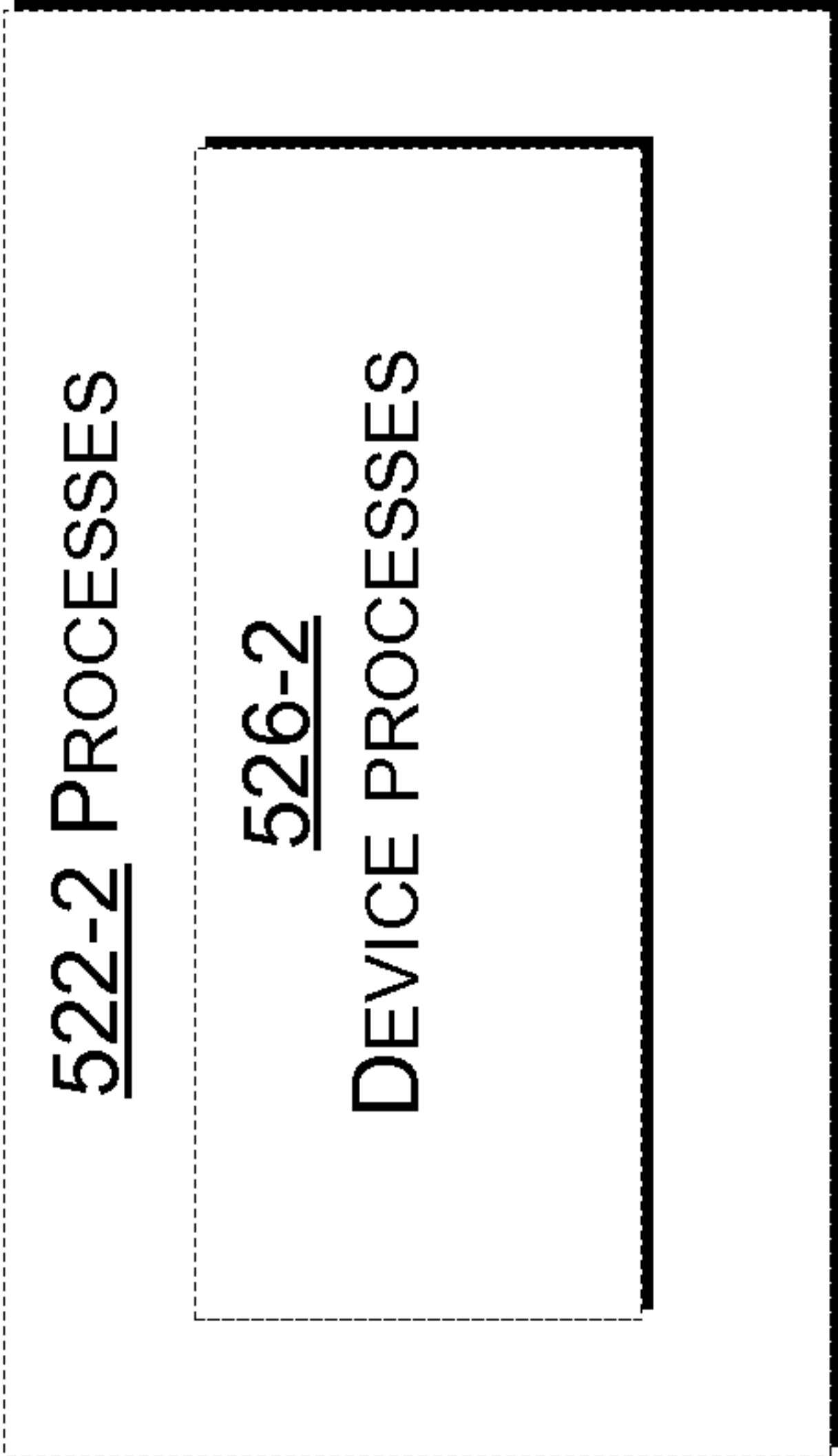


FIG. 5E

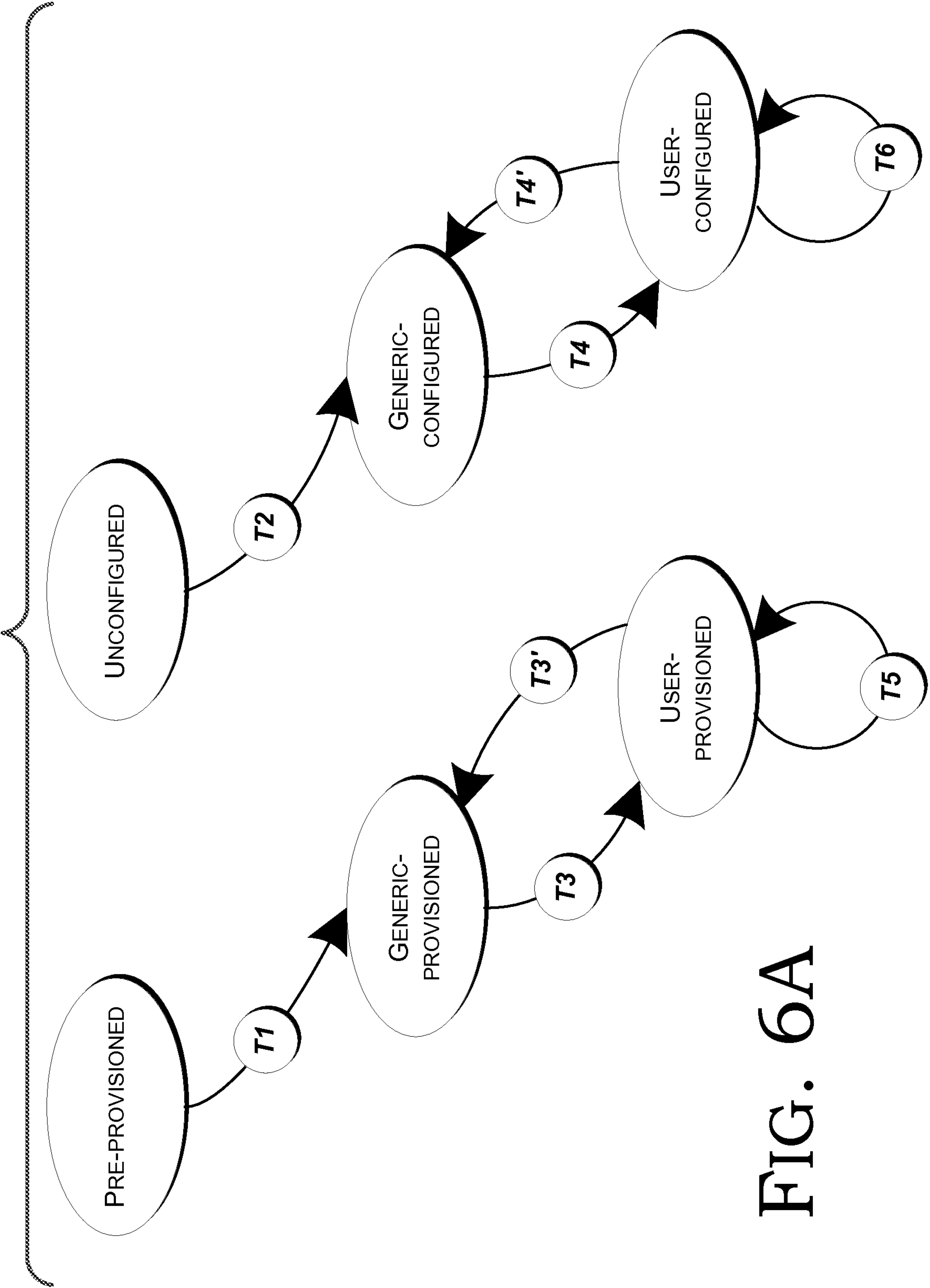


FIG. 6A

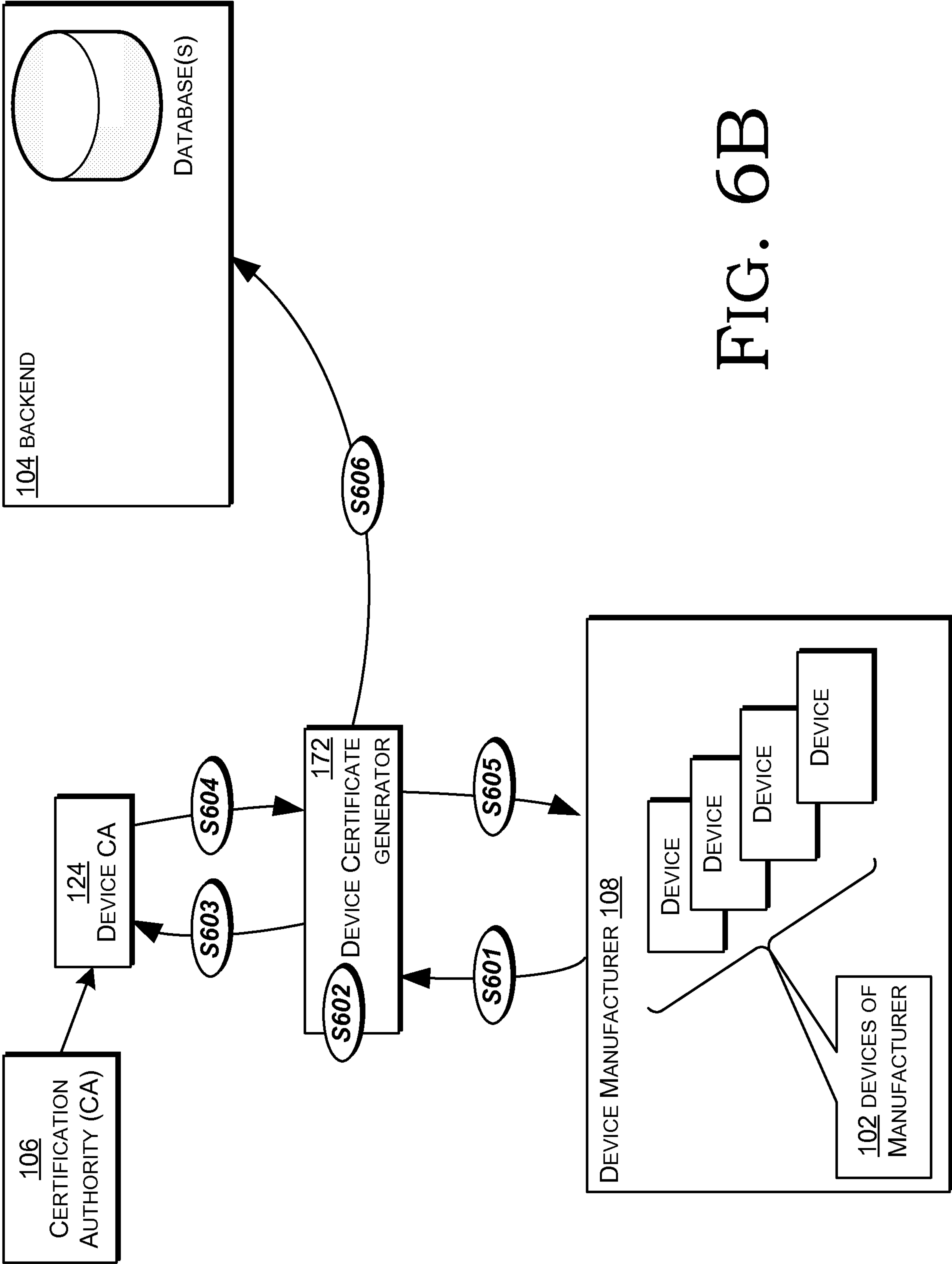


FIG. 6B

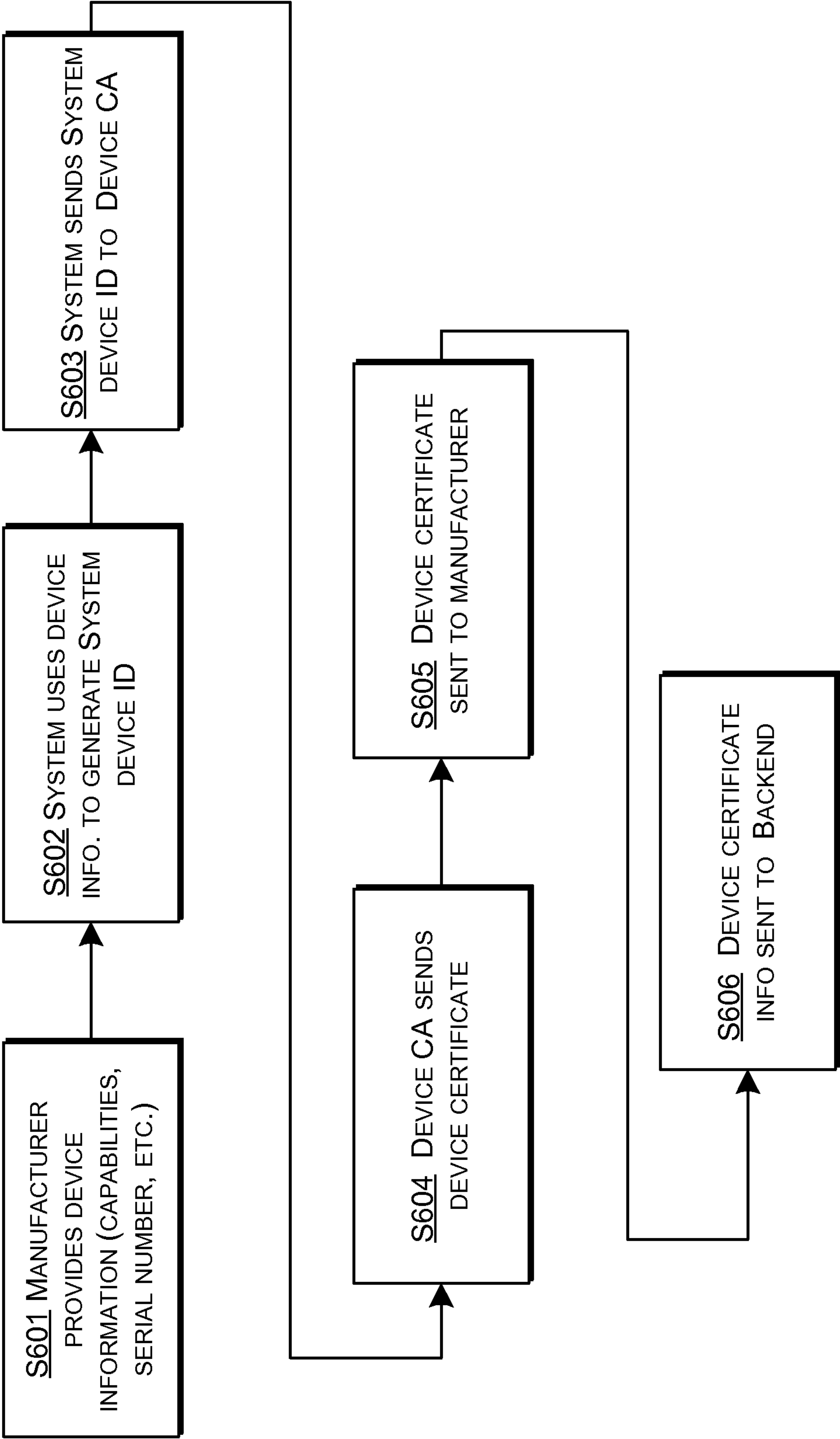


FIG. 6C

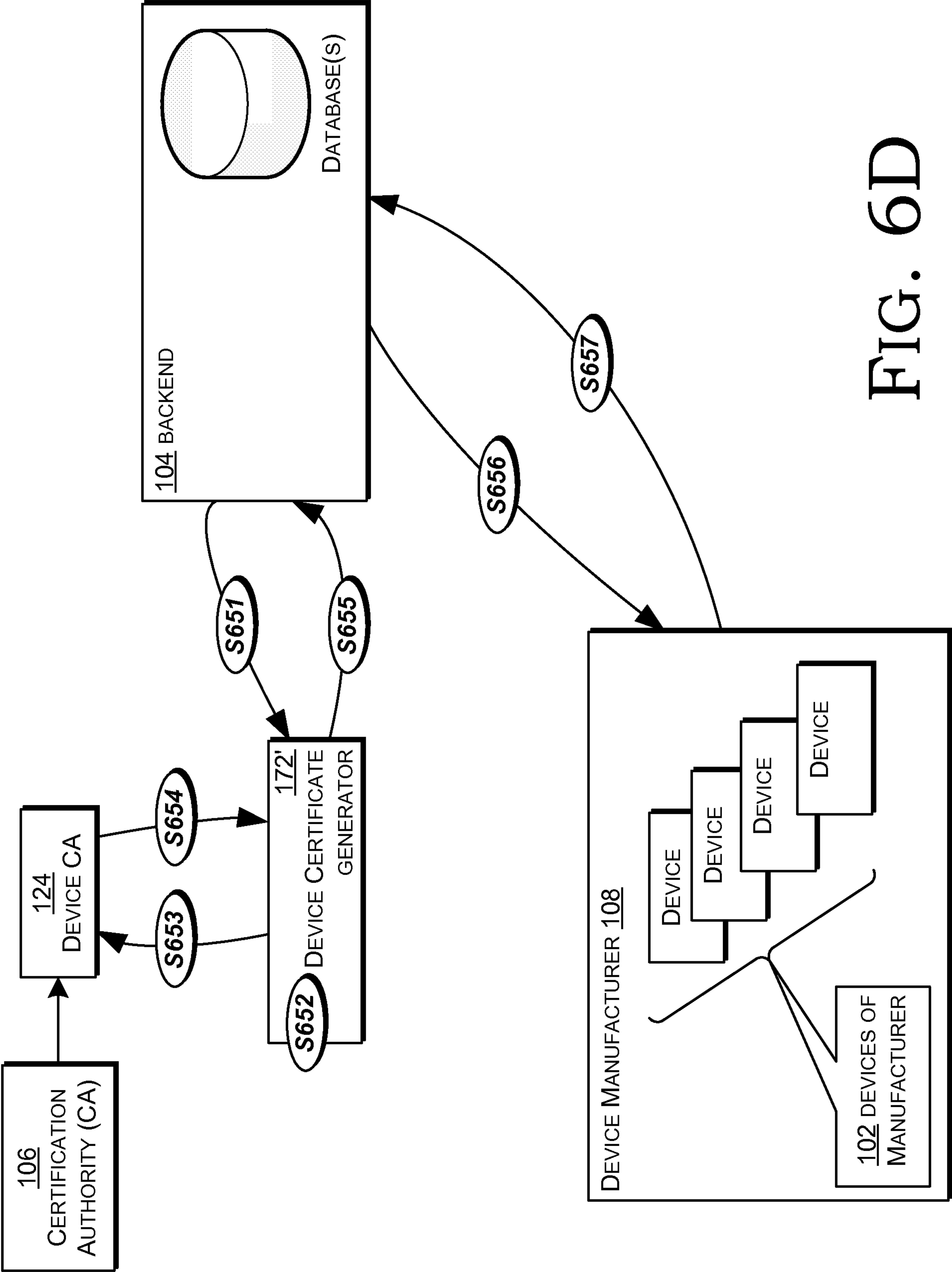


FIG. 6D

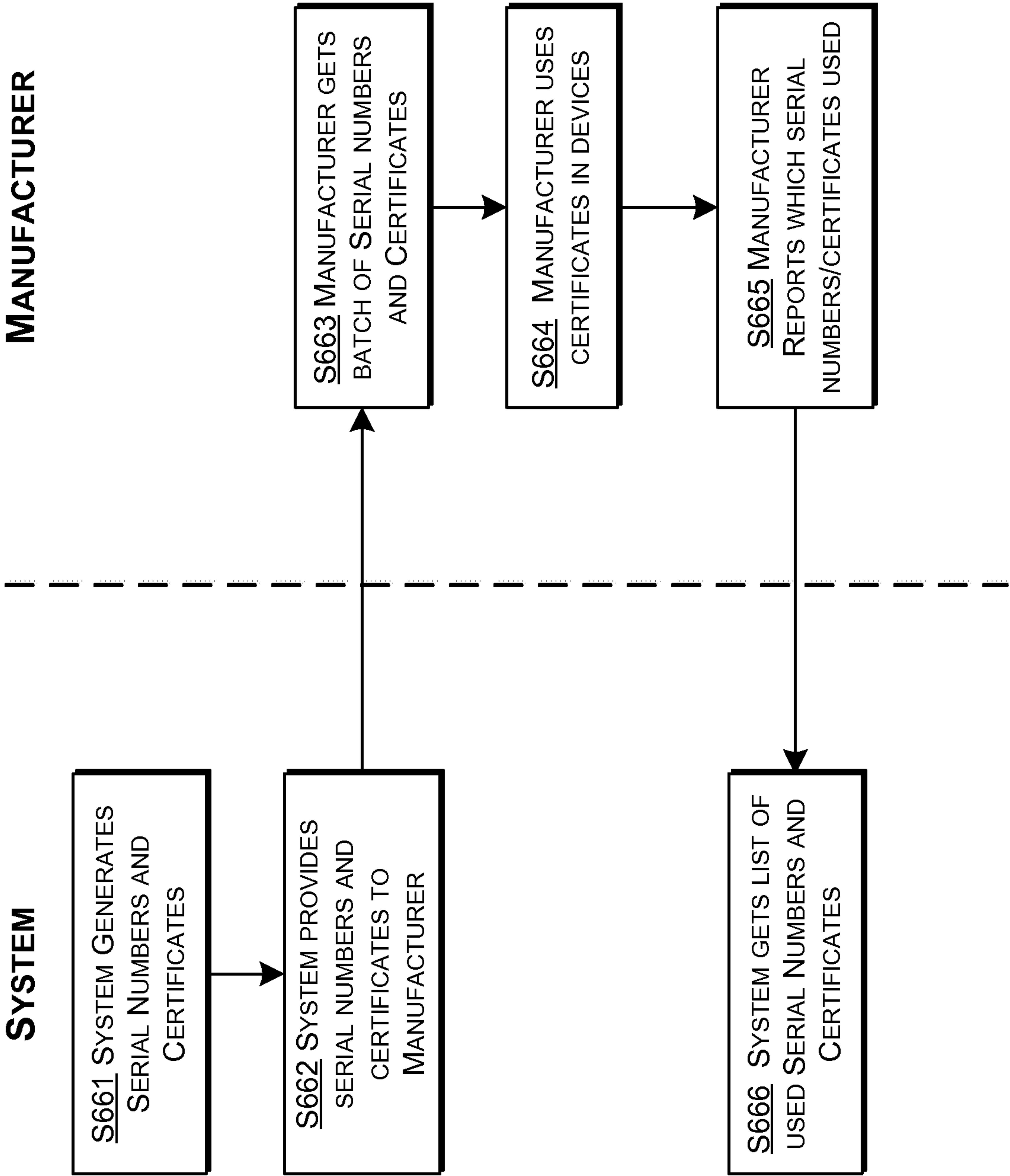


FIG. 6E

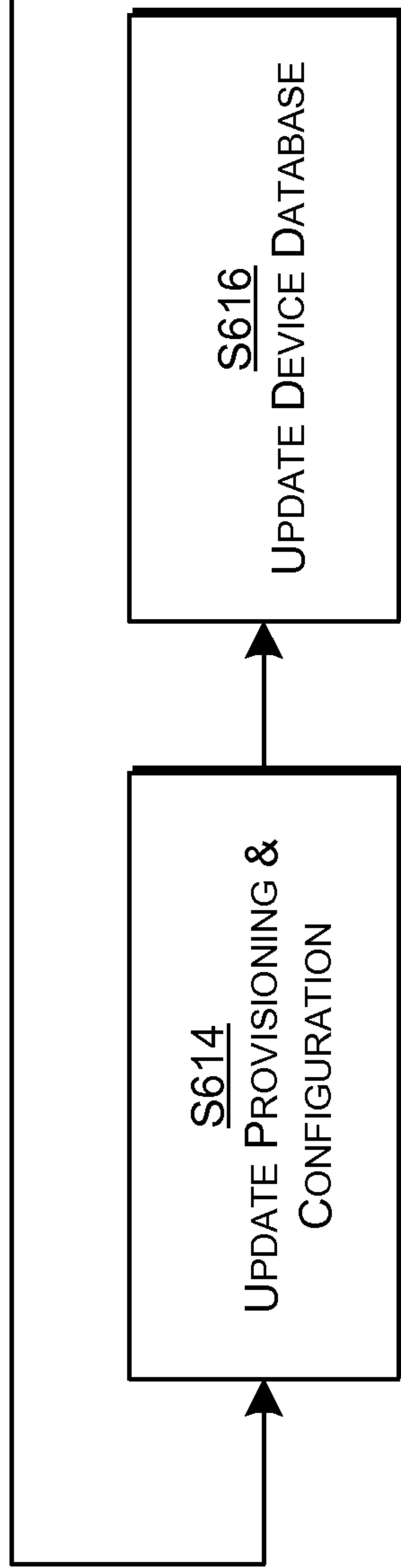
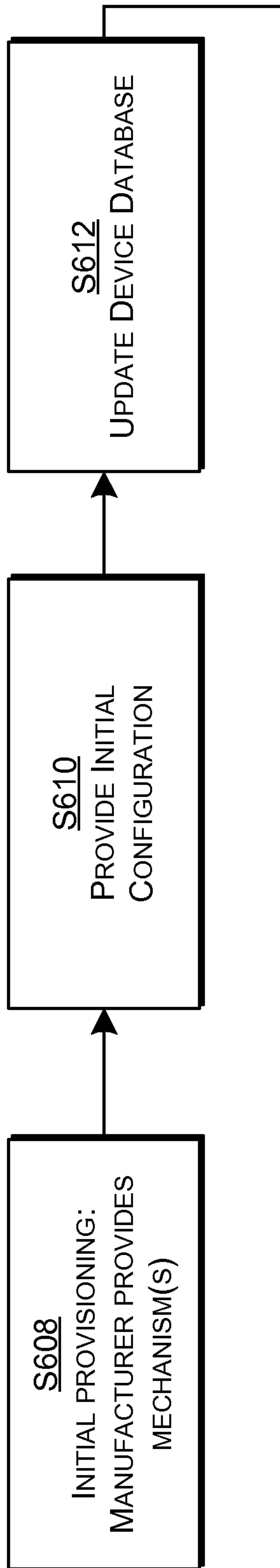


FIG. 6F.

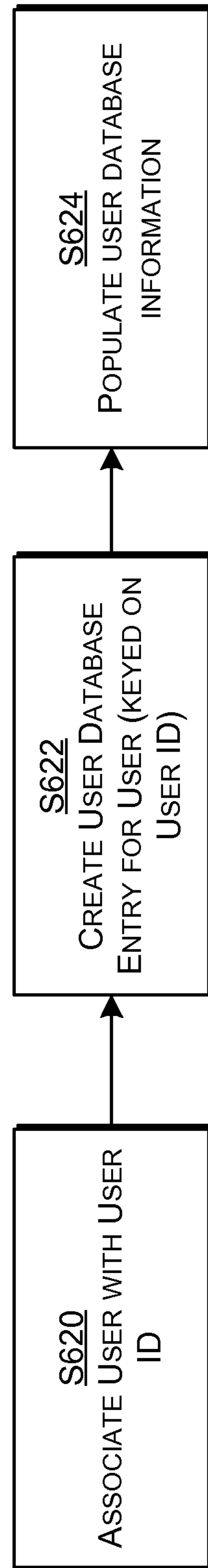


FIG. 6G.

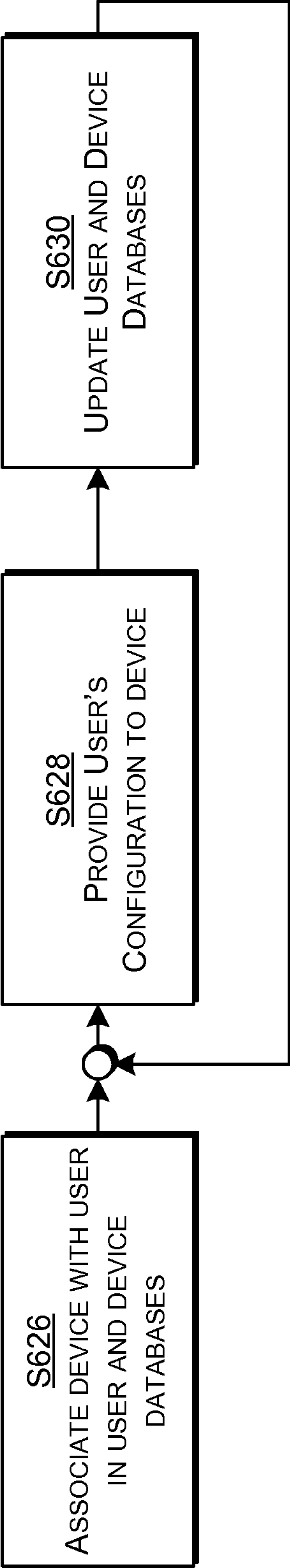


FIG. 6H

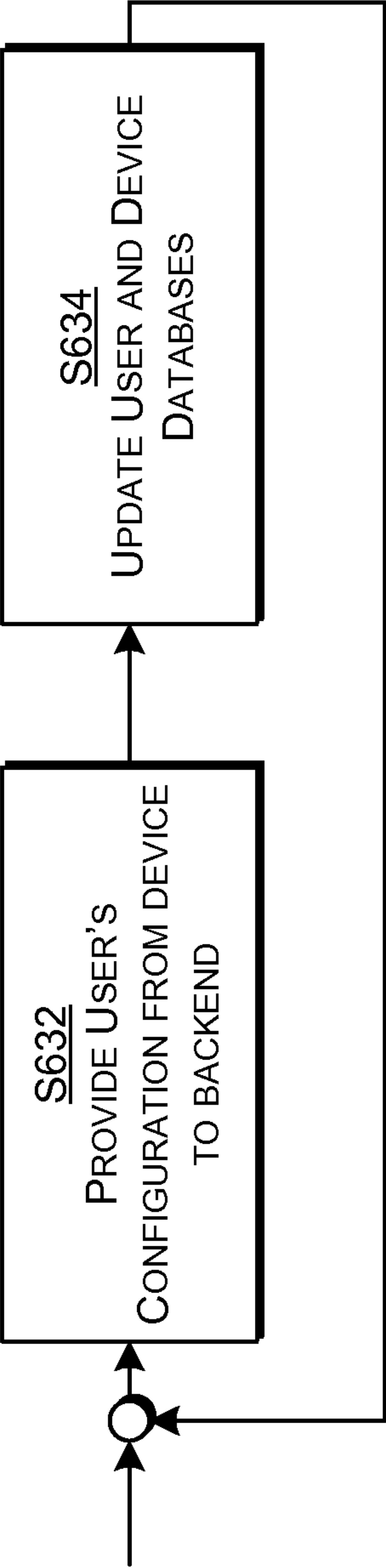


FIG. 6I

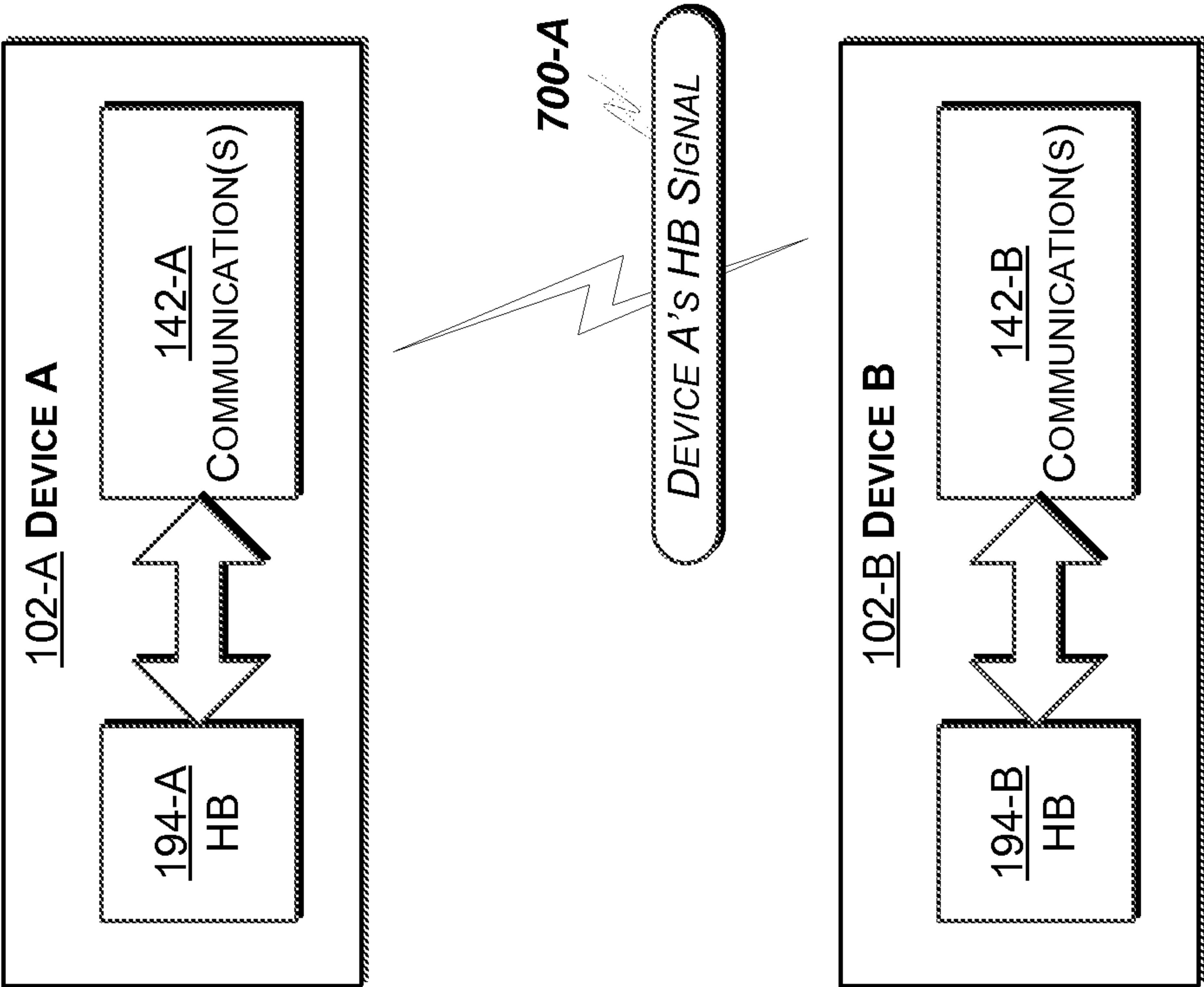


FIG. 7B

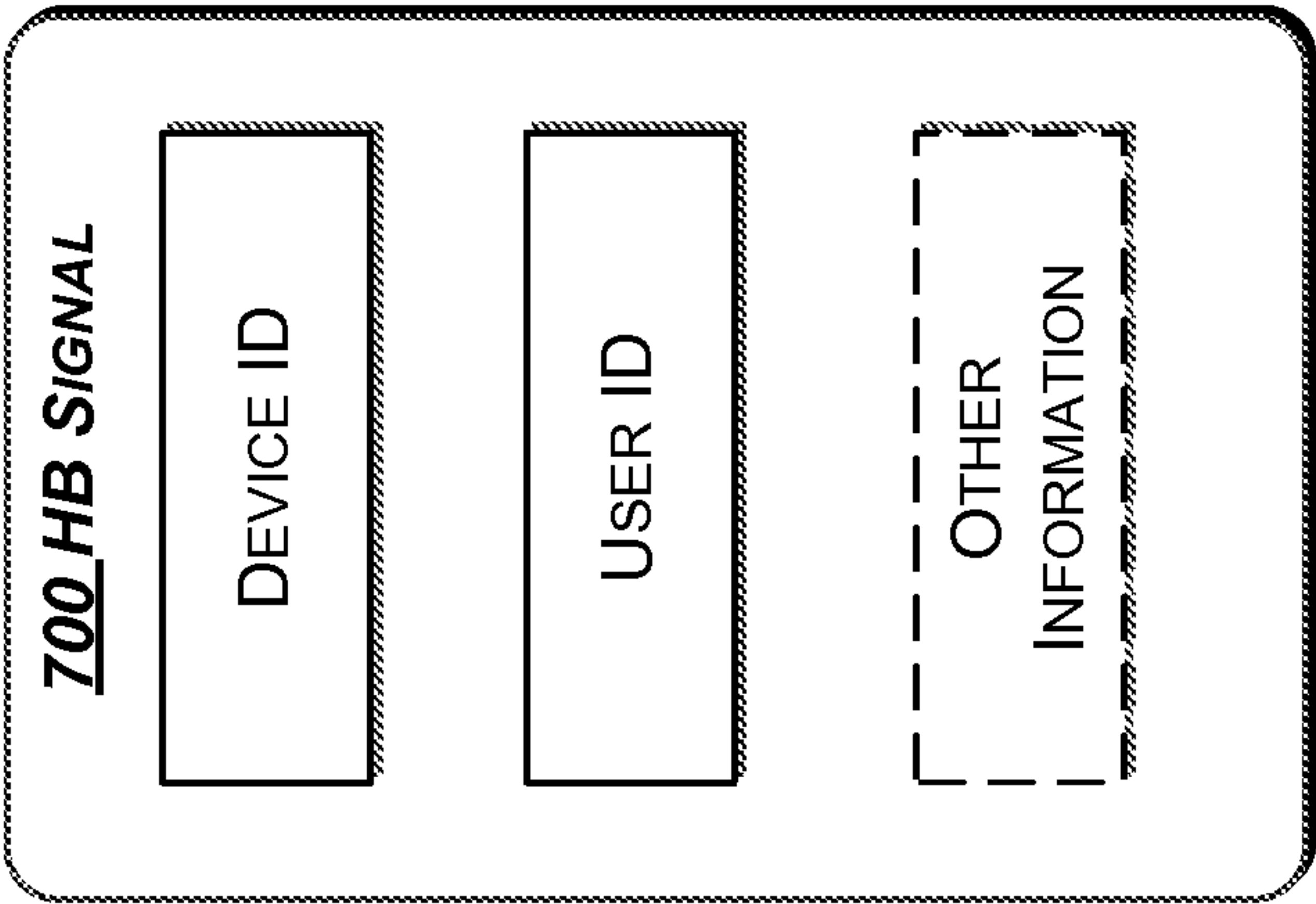


FIG. 7A

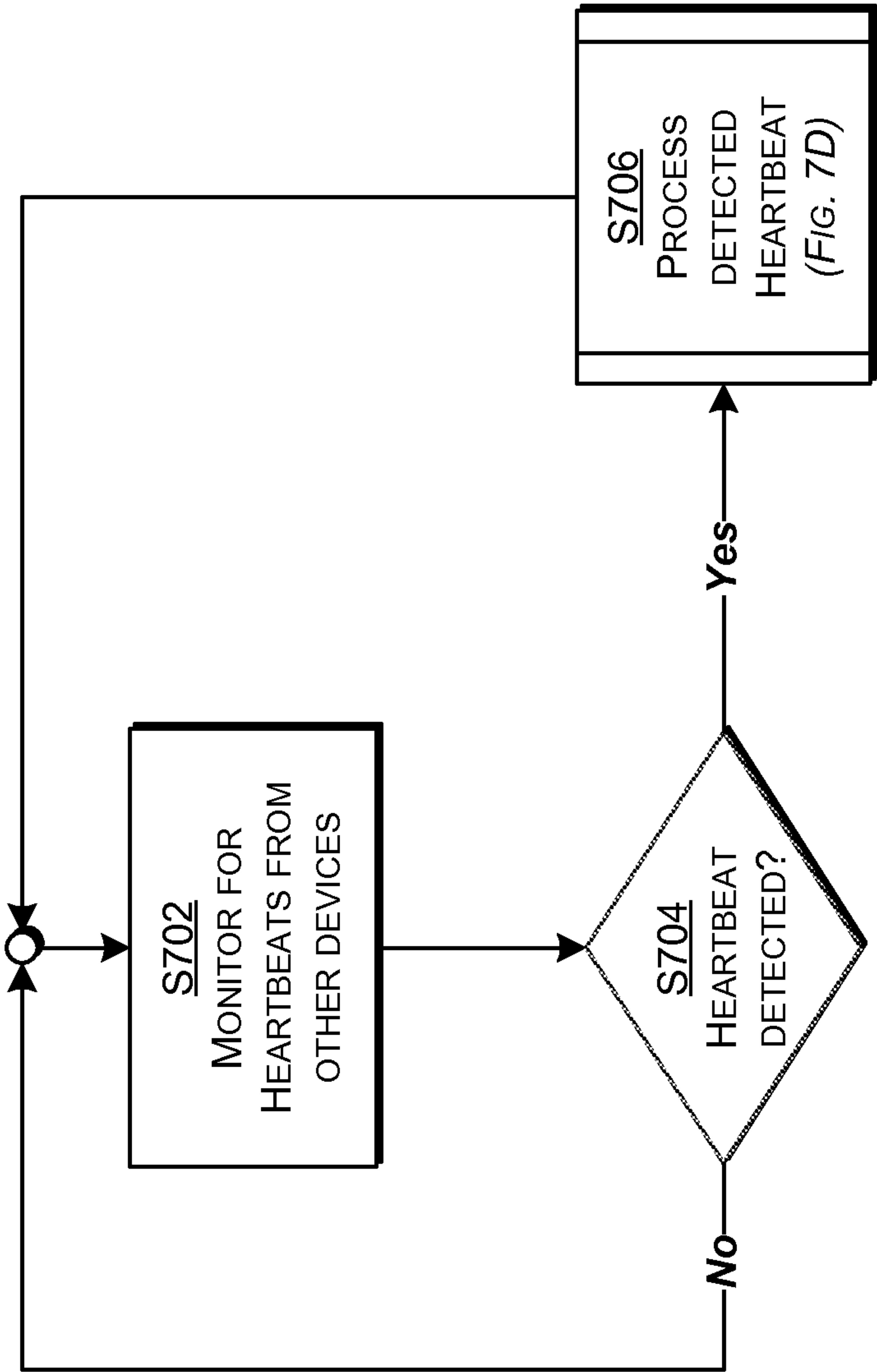


FIG. 7C

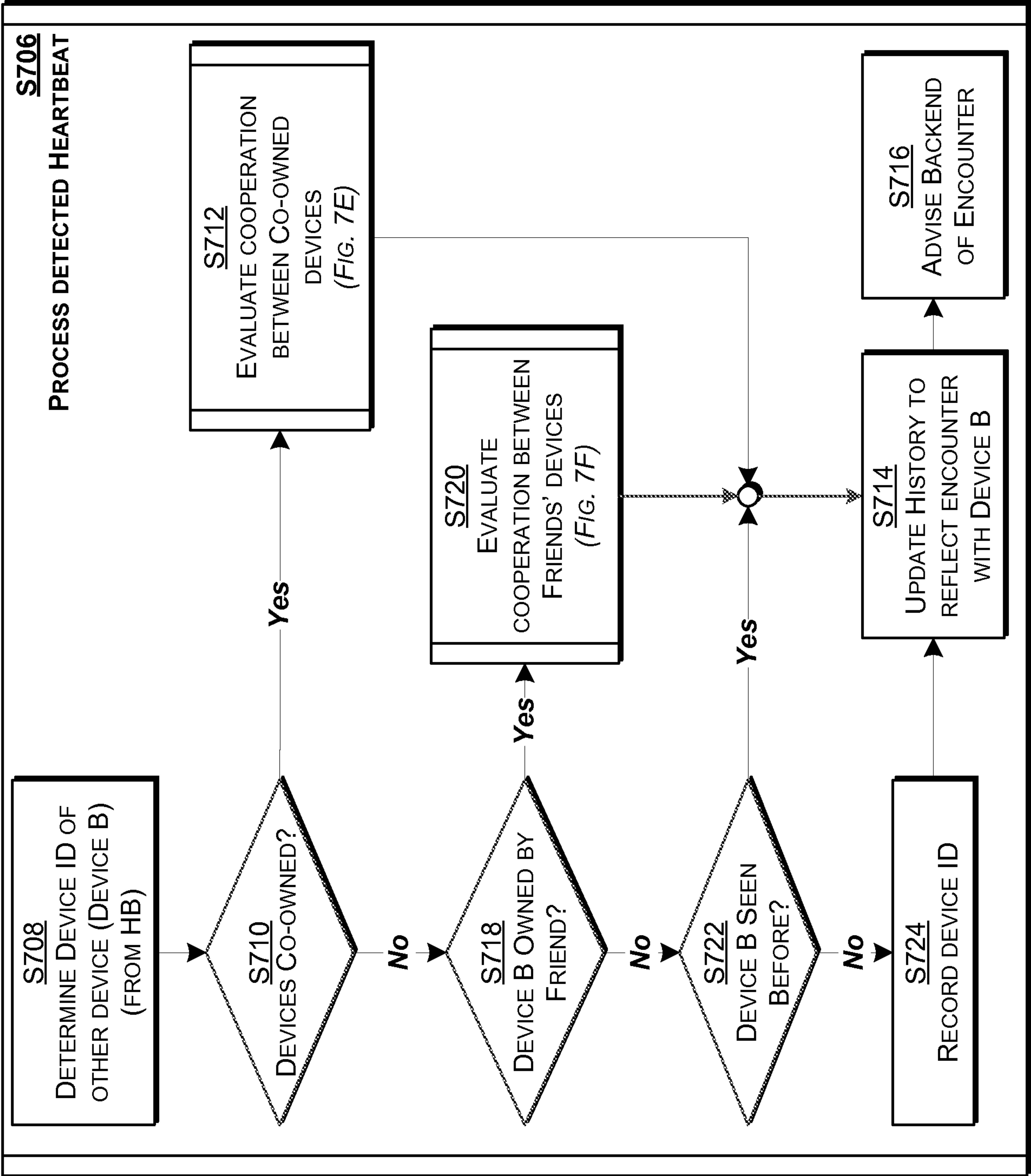


FIG. 7D

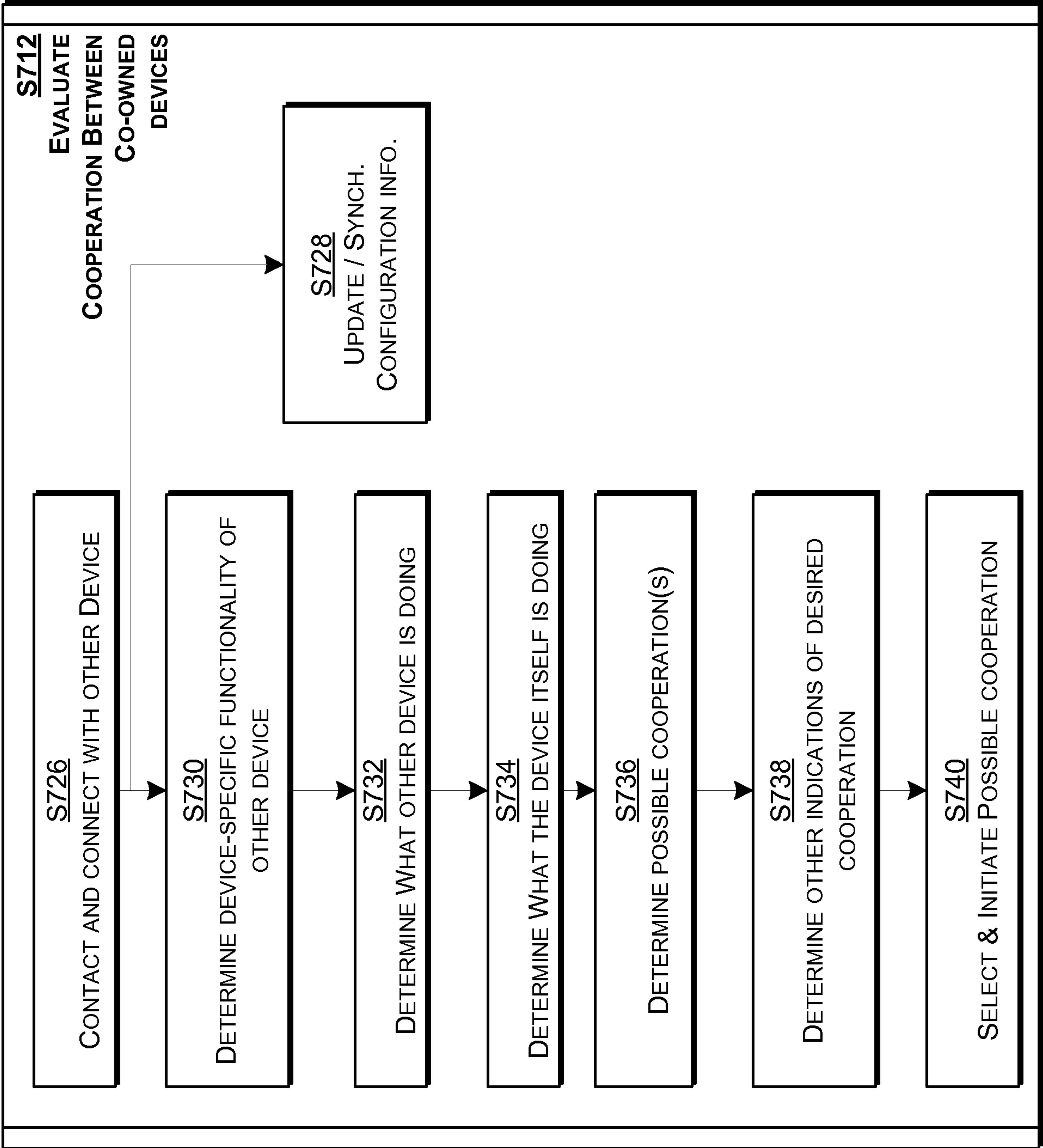


FIG. 7E

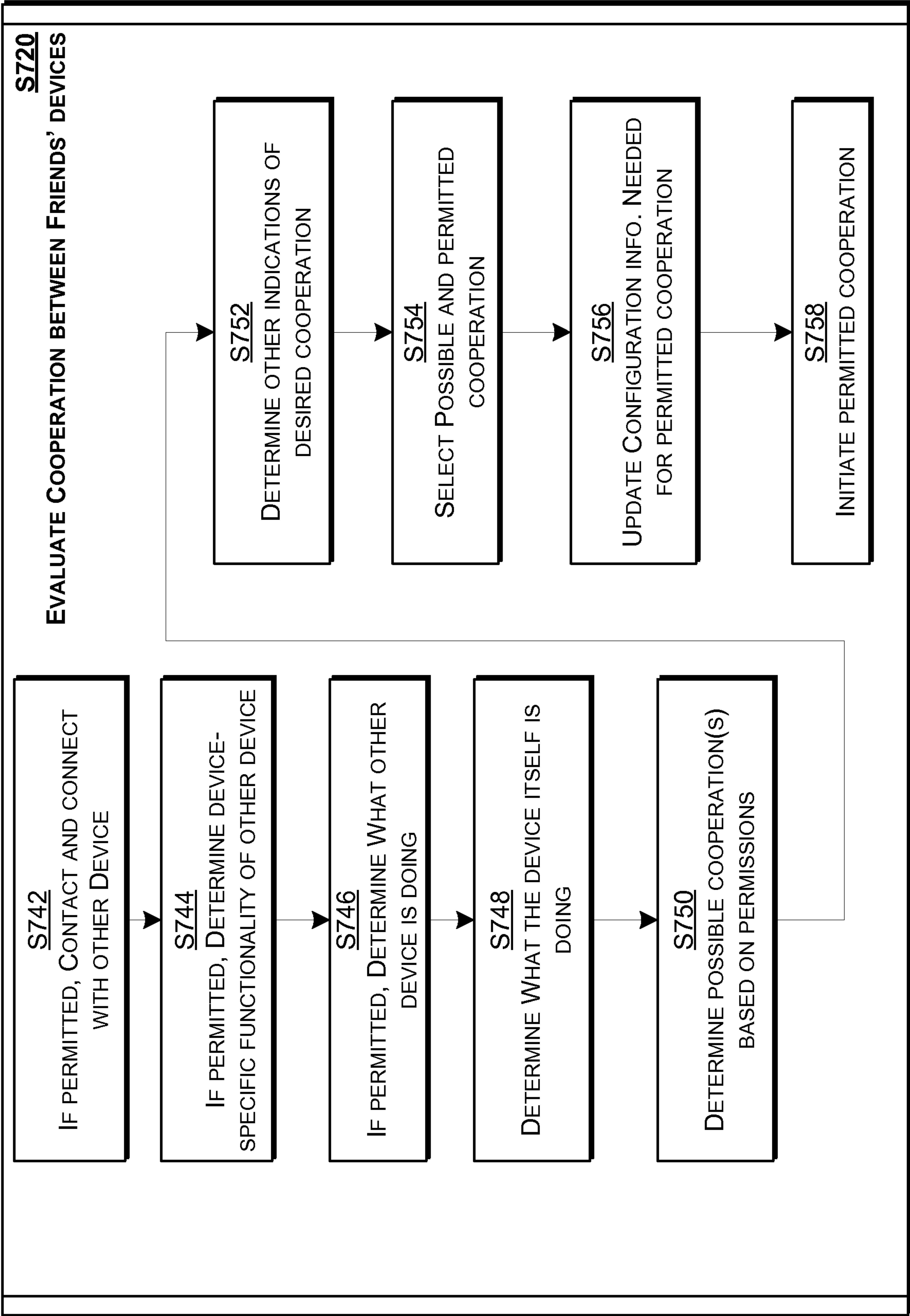


FIG. 7F

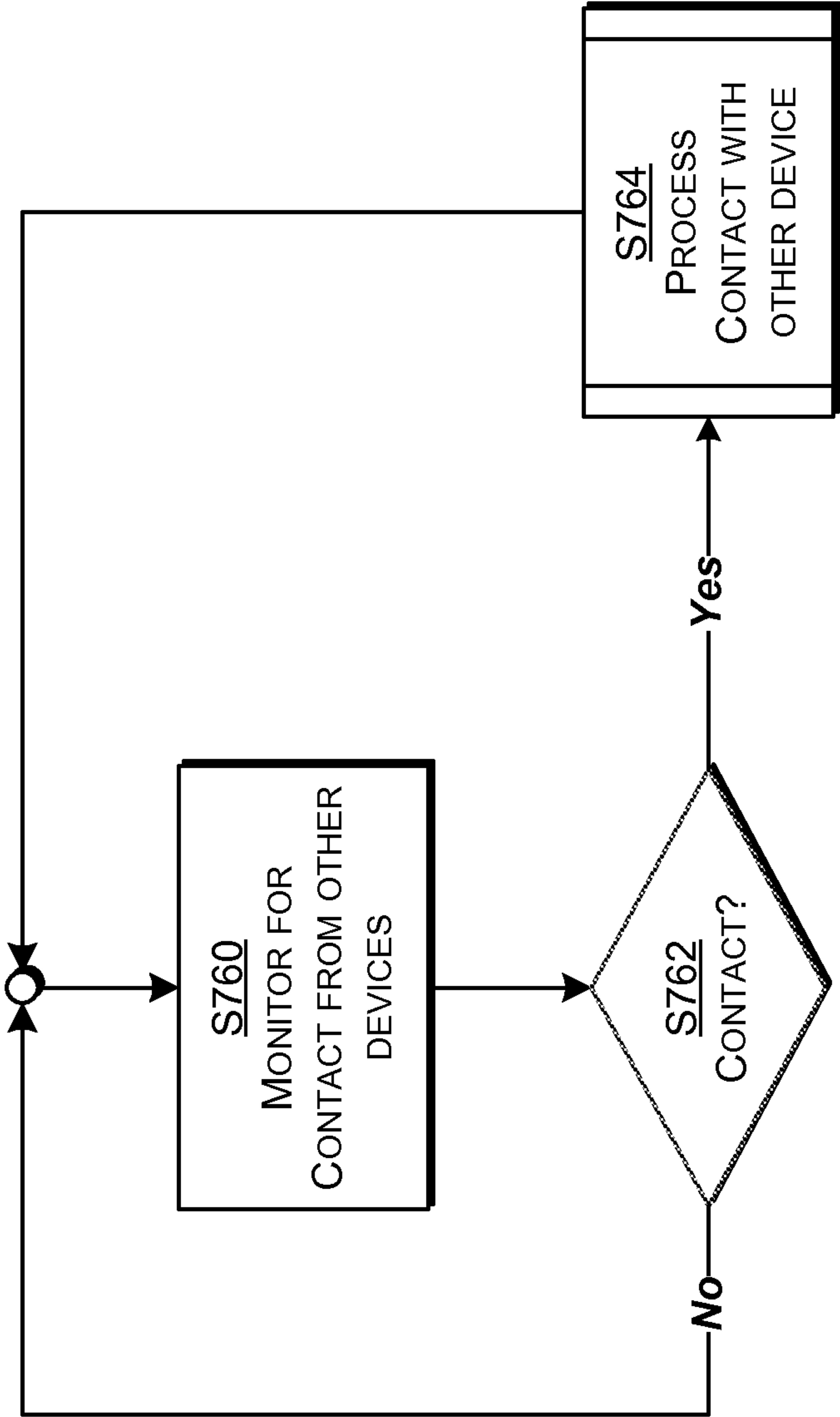


FIG. 7G

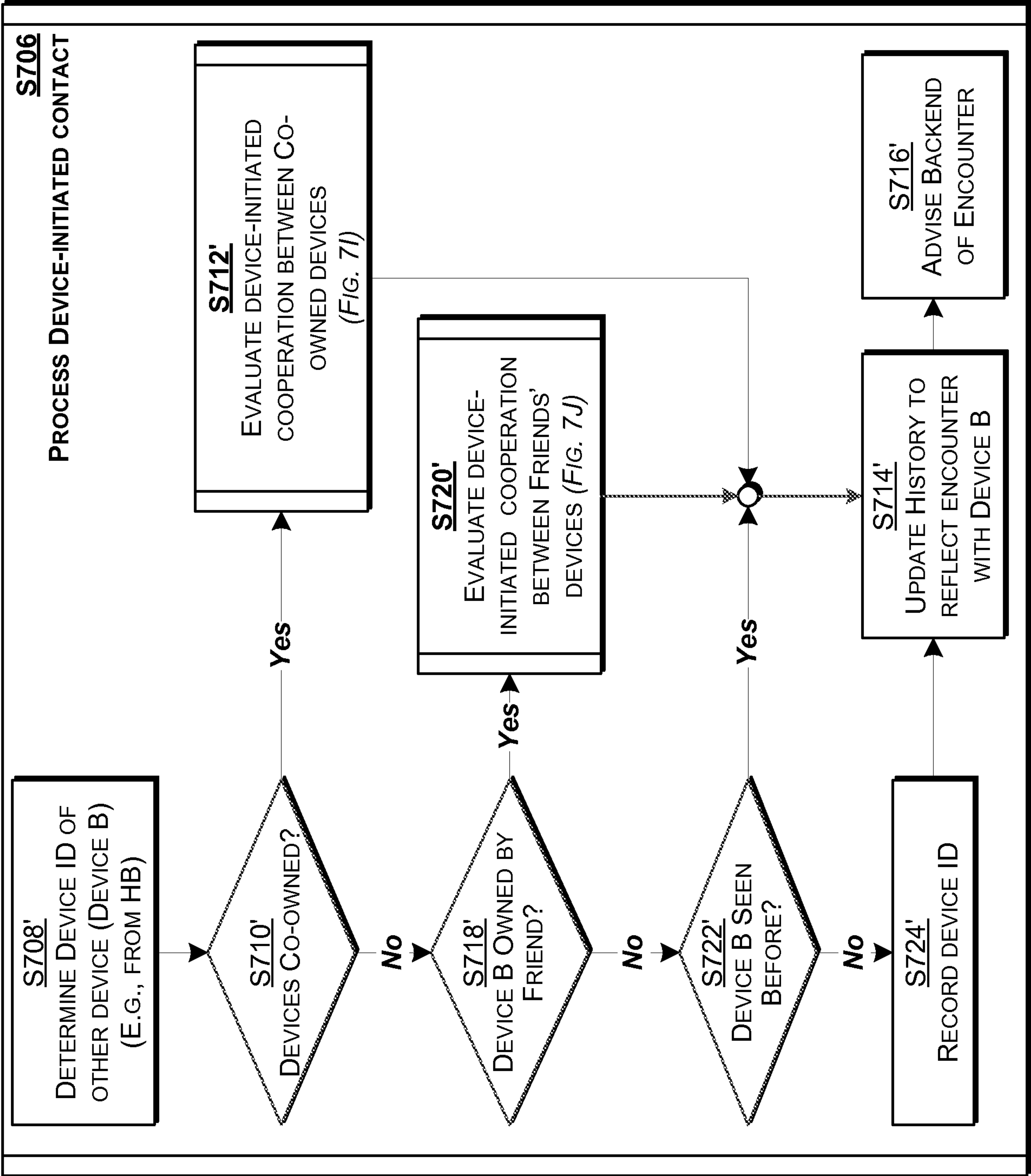
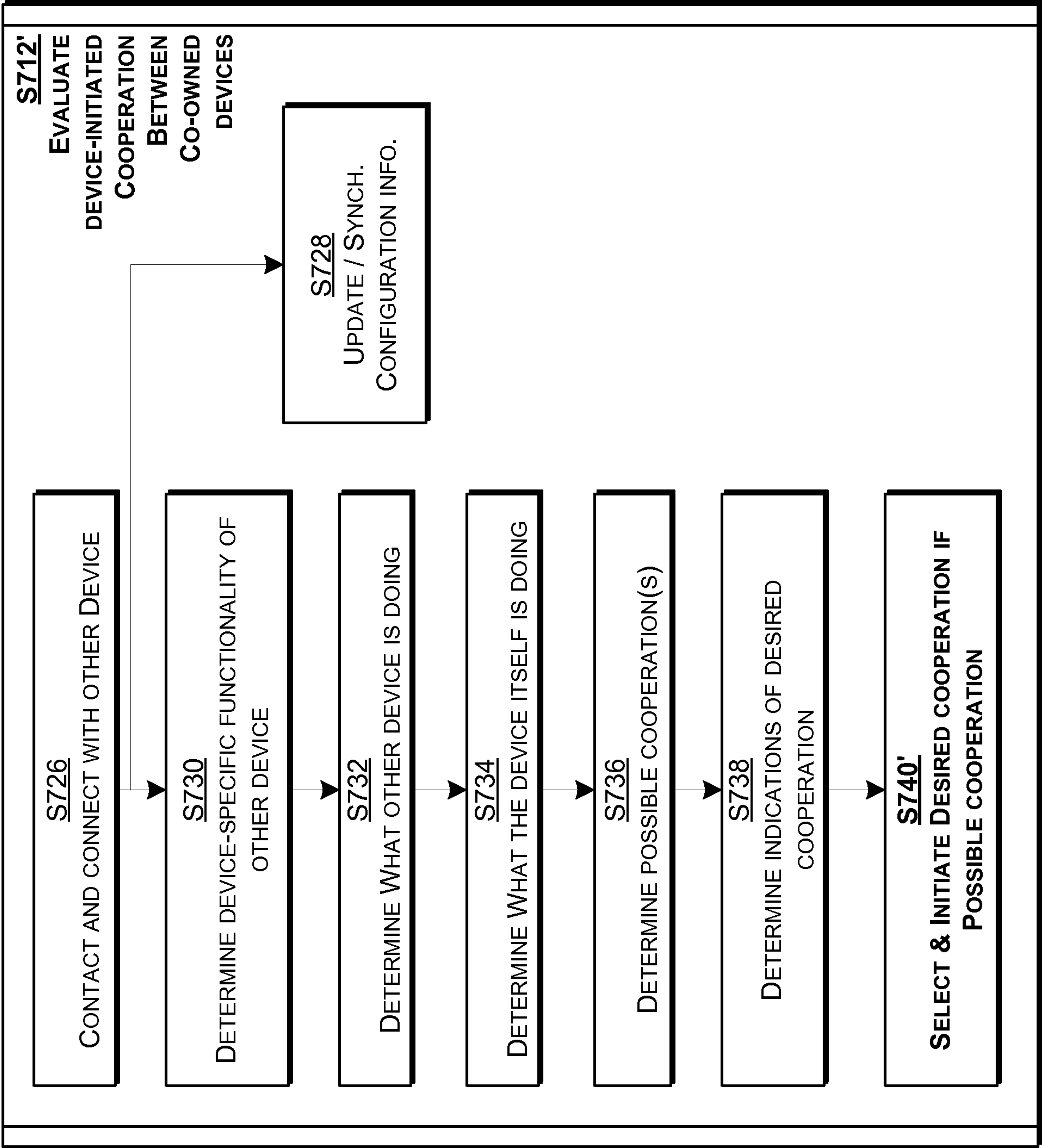


FIG. 7H

FIG. 7I



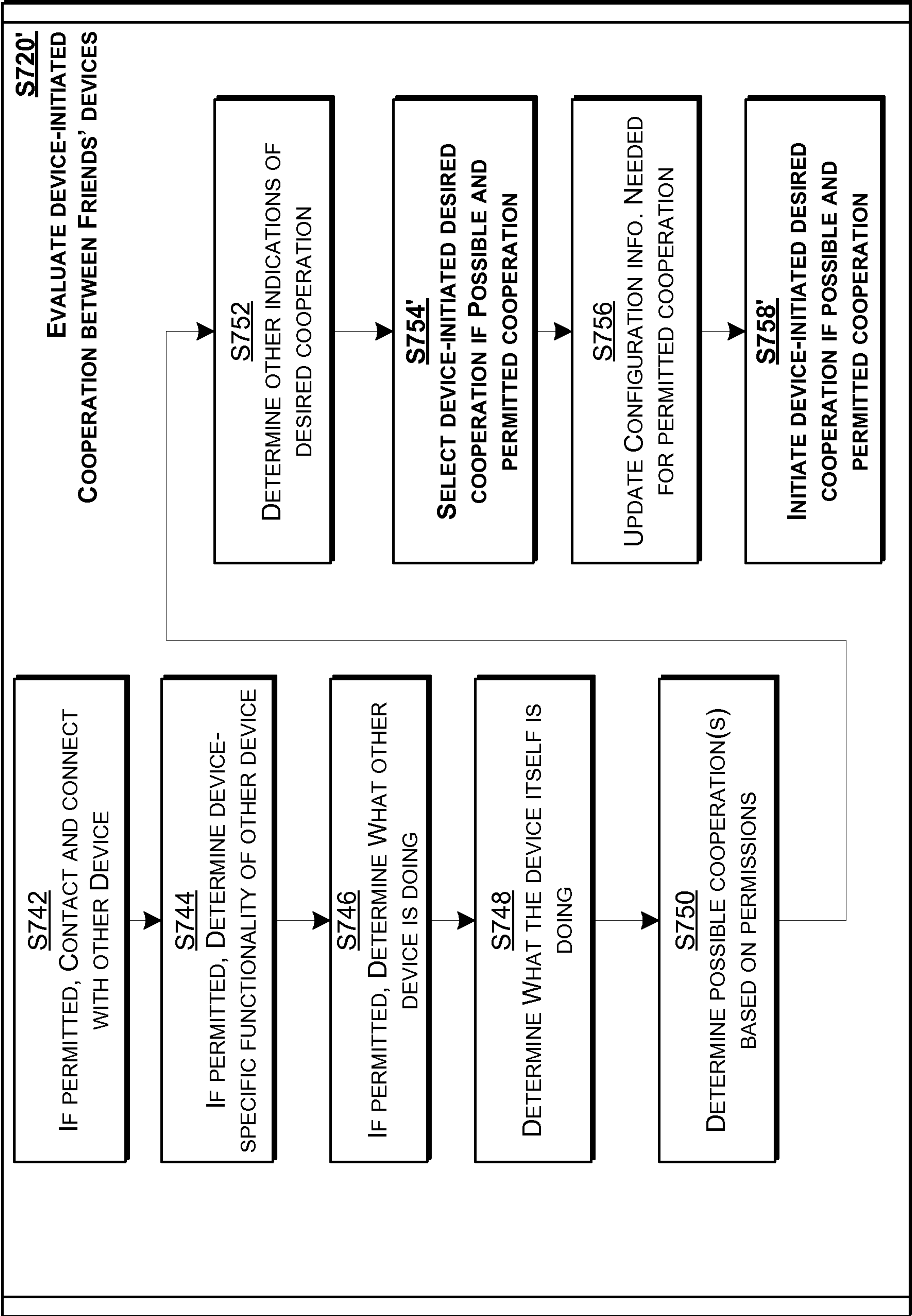


FIG. 7J

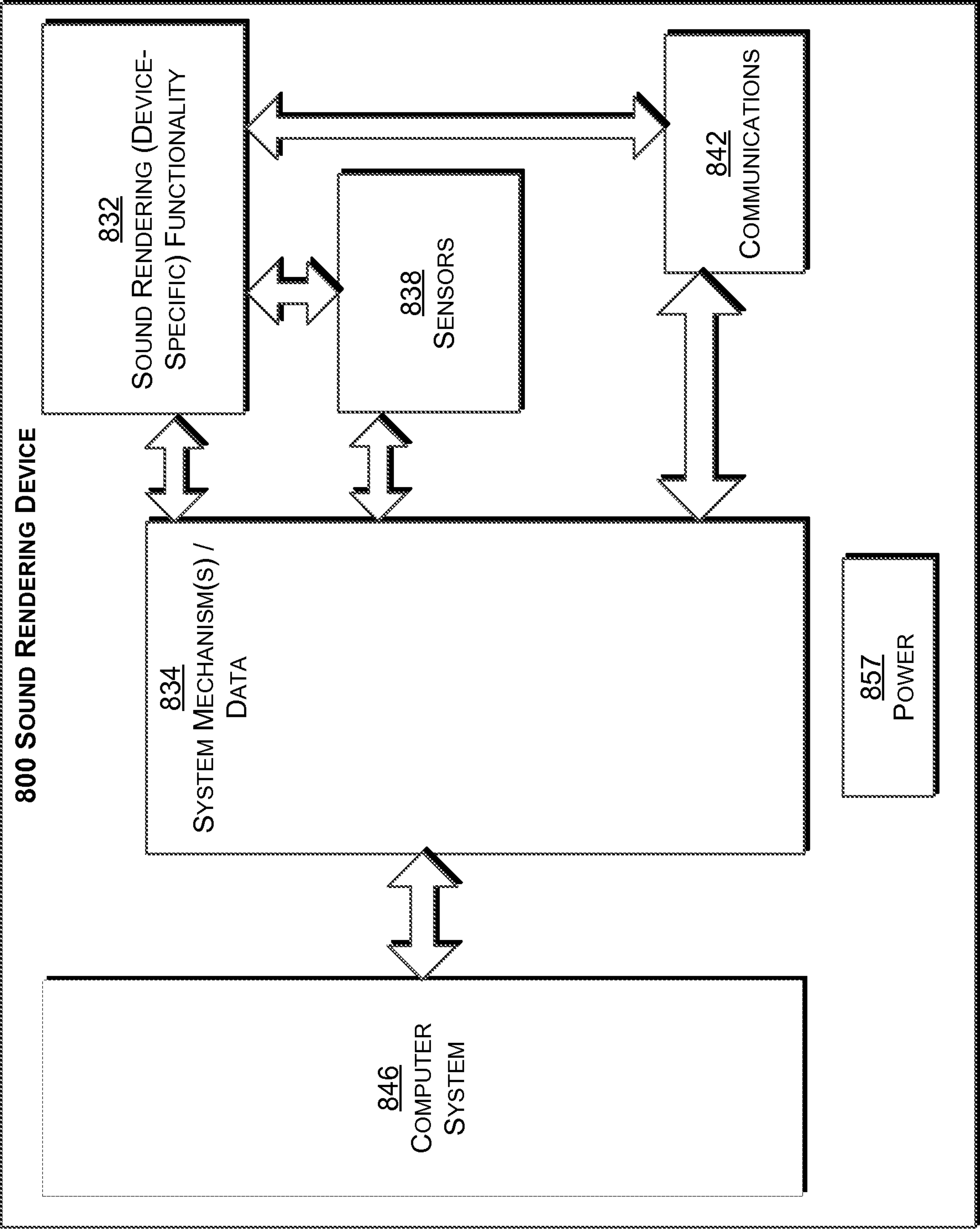
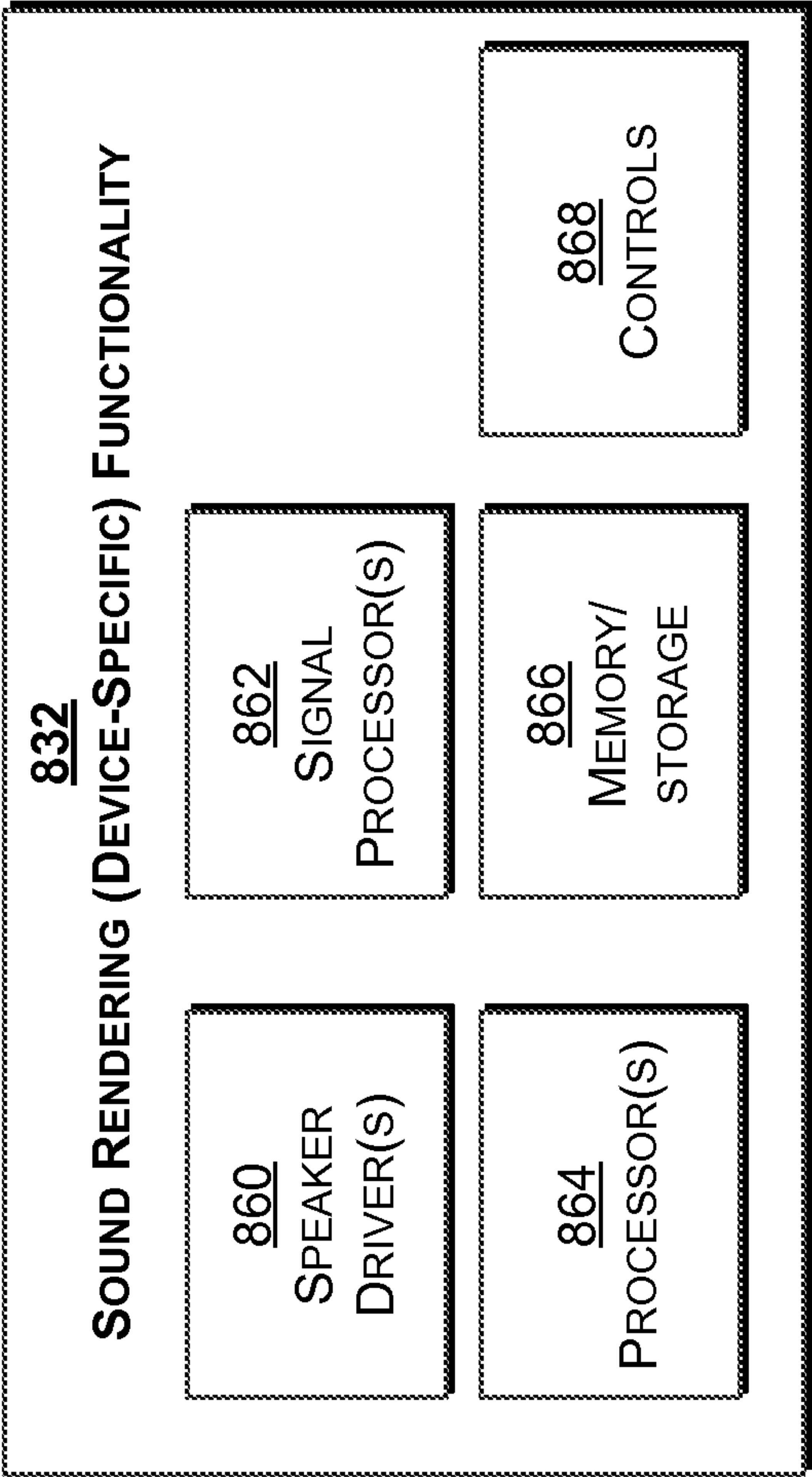


FIG. 8A

FIG. 8B



38/42

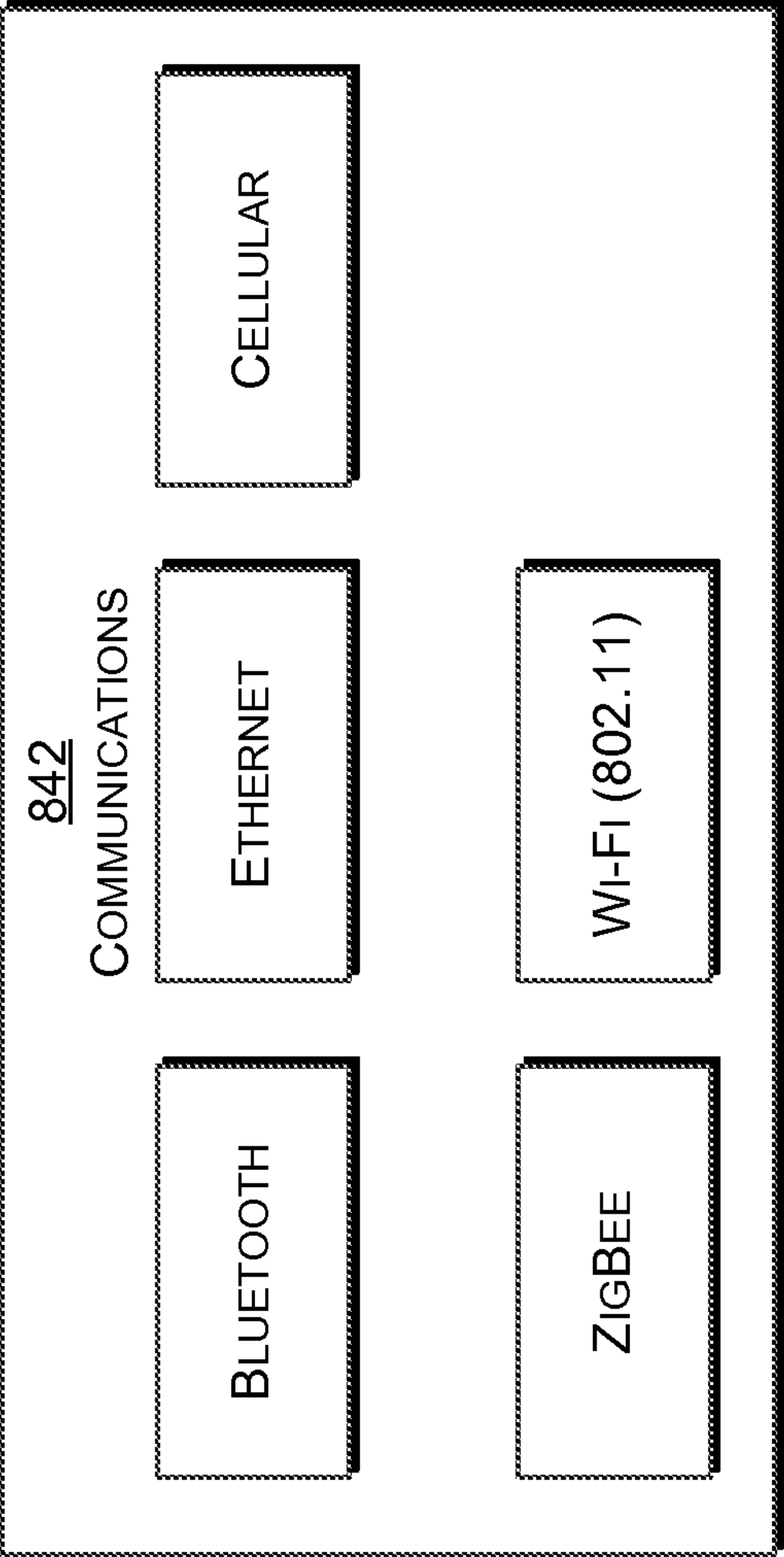


FIG. 8C

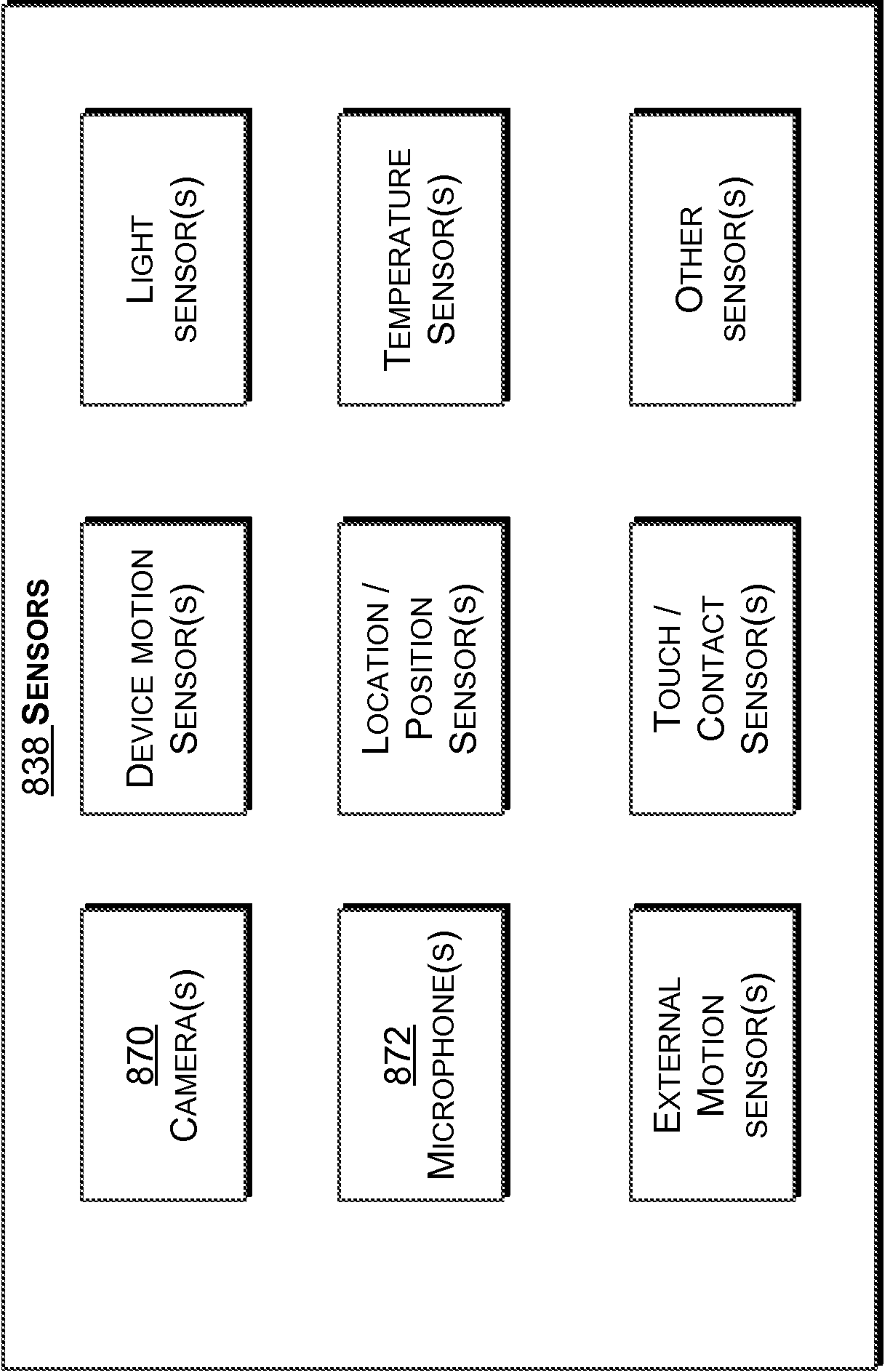


FIG. 8D

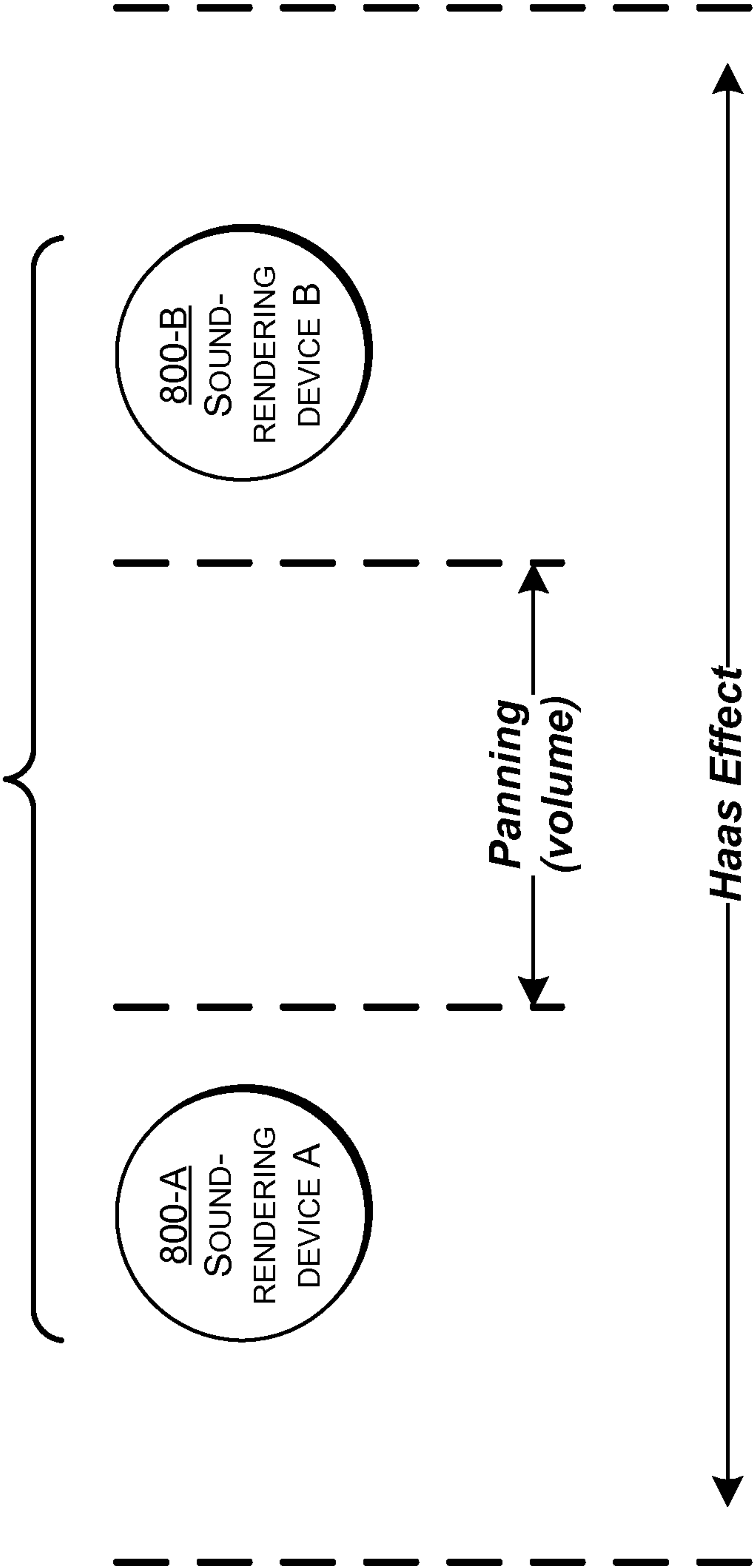


FIG. 9A

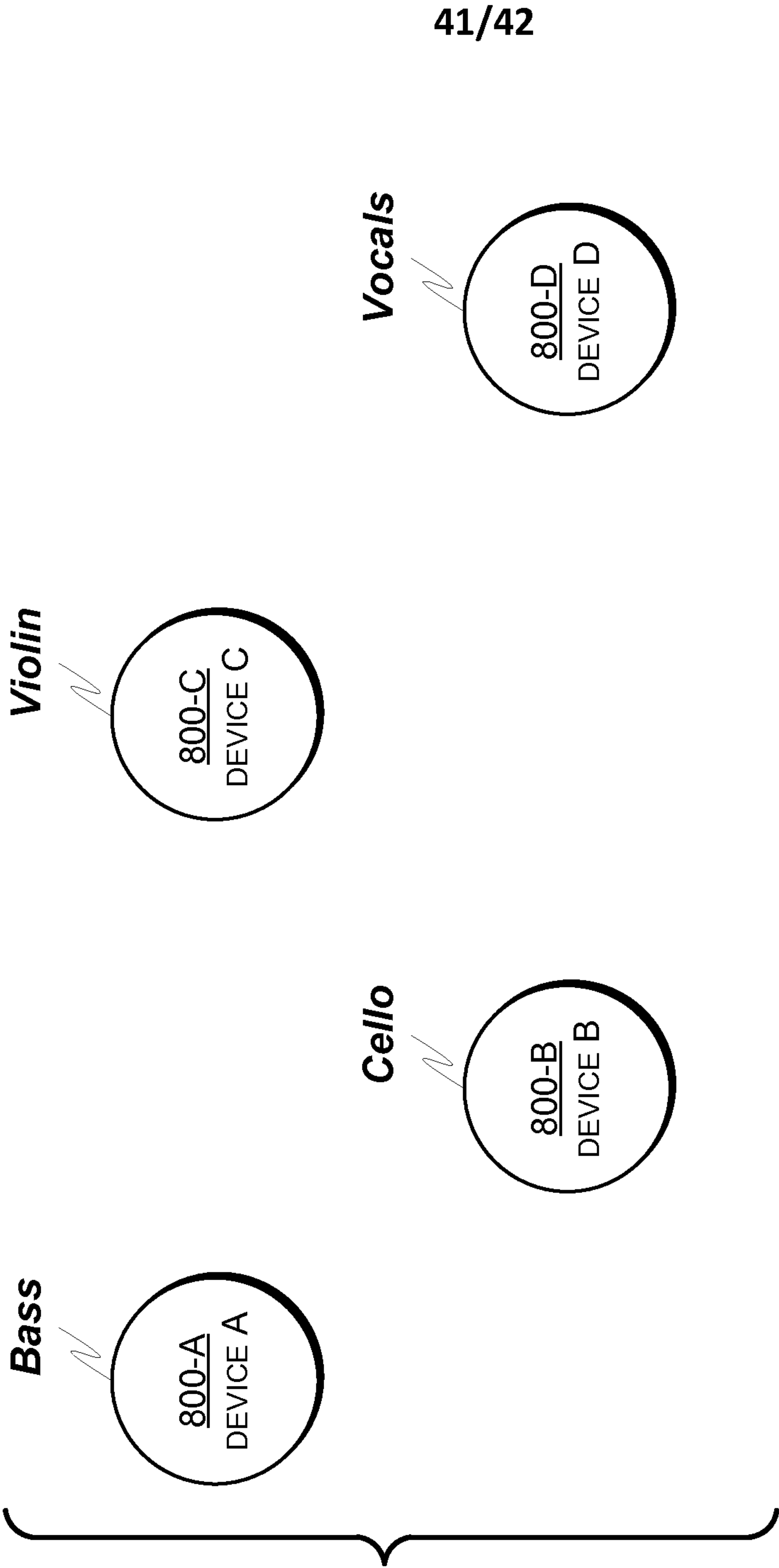


FIG. 9B

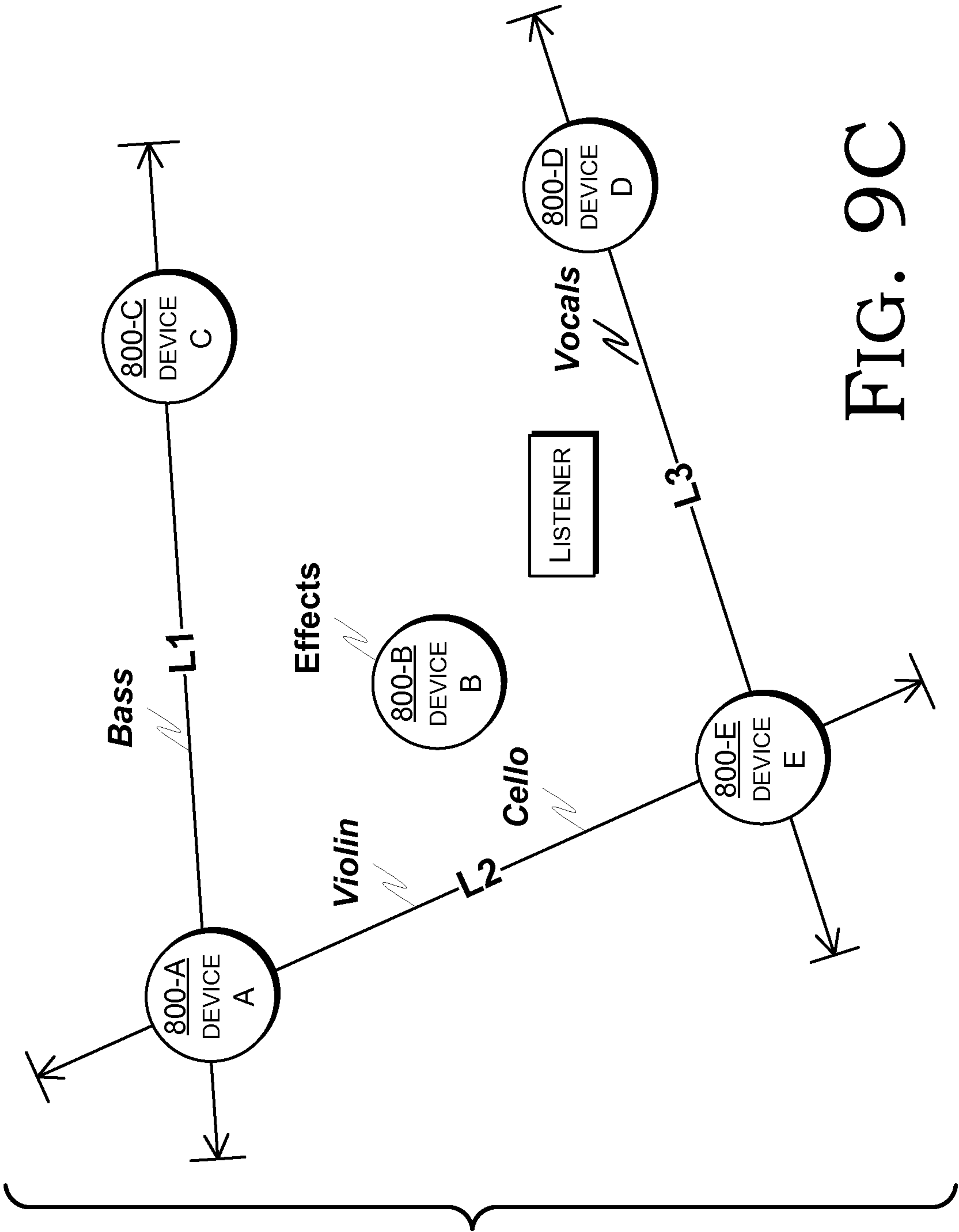


FIG. 9C

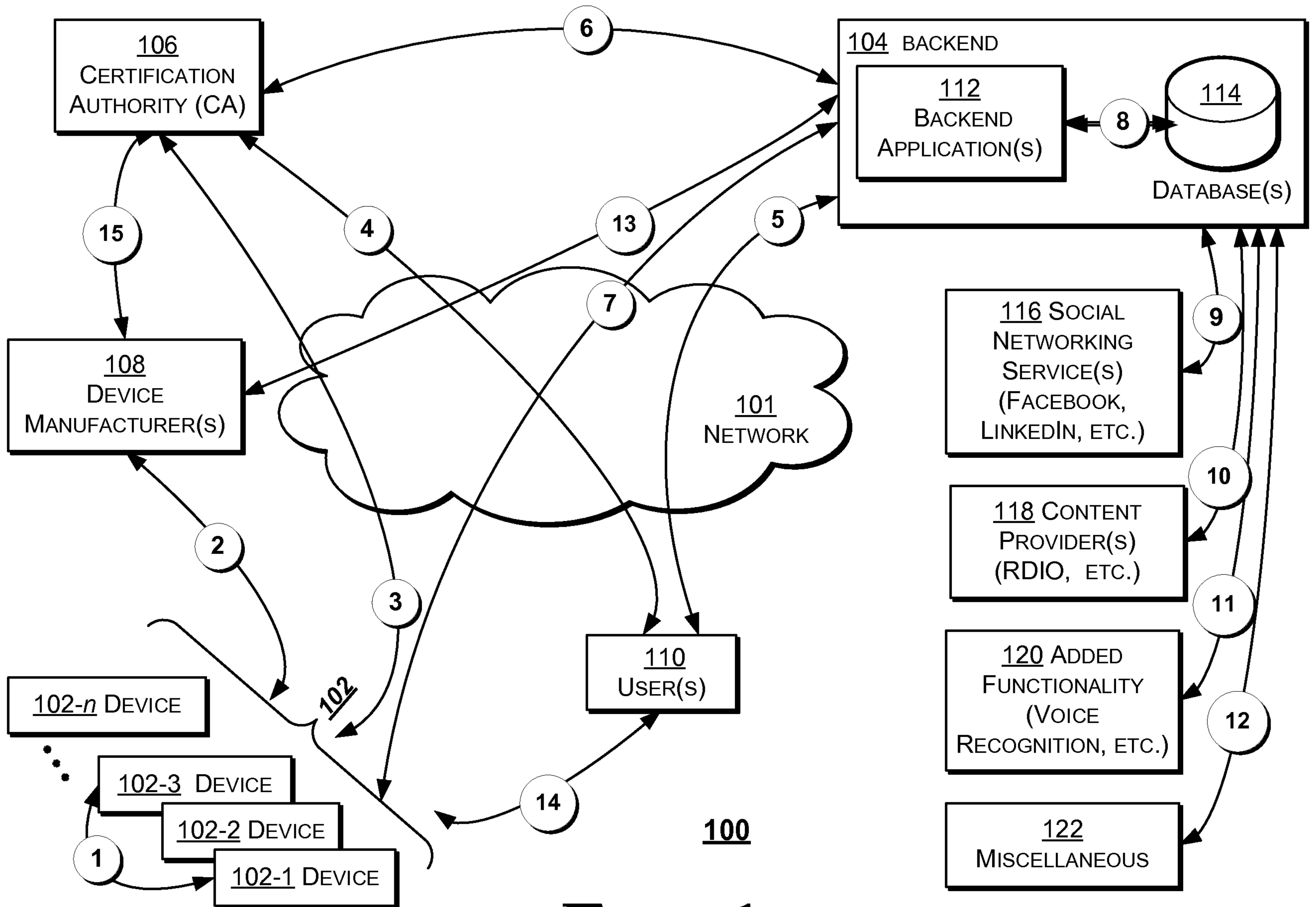


FIG. 1