

(51) International Patent Classification:
Not classified

(21) International Application Number:
PCT/IL2024/050578

(22) International Filing Date:
11 June 2024 (11.06.2024)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/507,480 12 June 2023 (12.06.2023) US

(71) Applicant: **TECHNOLOGIES OF VOICE INTER-FACE LTD** [IL/IL]; 35 Jabotinsky st, 5251108 Ramat Gan (IL).

(72) Inventors: **GIBSON, Matthew**; 84 Slayleigh Lane, Sheffield Yorkshire S10 3RH (GB). **JURCICEK, Filip**; Luční 322, 252 17 Tachlovice (CZ). **KARAULOV, Ievgen**; Julio César 1318, apt. 206, Montevideo, 1130 (UY).

(74) Agent: **GOLDRAICH, Marganit et al.**; GOLD-PATENTS & FINANCIAL SERVICES LTD, 15 Yohanan Hasandlar St., P.O.B. 25267, 31251 Haifa (IL).

(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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,

HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, 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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

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

Published:
— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

(54) Title: SPEAKER EMBEDDINGS FOR PARKINSON'S DISEASE ON/OFF CLASSIFICATION

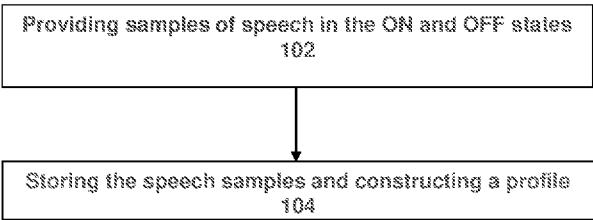


FIG. 1

(57) Abstract: A method is provided for classifying speaker embeddings for a user having Parkinson's disease. The method comprises, after enrolling data with speech samples in an ON state and in an OFF state, constructing a profile of the user in an auxiliary model, monitoring a state of the user and receiving a test signal of the user, extracting embeddings from a test signal, and eventually, classifying the recognition signal as an ON state or an OFF state.

WO 2024/257097 A2



SPEAKER EMBEDDINGS FOR PARKINSON'S DISEASE ON/OFF CLASSIFICATION

TECHNICAL FIELD

[0001] The present disclosure relates to speaker embeddings classification in general, and to systems that automatically classifies the state of a person with Parkinson's disease (PwP) as ON or OFF state, given a sample of a speech, in particular.

BACKGROUND

[0002] A person with Parkinson's disease may experience alterations to their speech as a symptom of the current condition. Such alterations occur in approximately 75-90% of the PwP population. These alterations are a particular sub-category of dysarthric speech characterized by reduced loudness and pitch variability, breathy or hoarse voice, imprecise articulation, and more general features such as abnormalities of speech rate and pause ratio.

[0003] Levodopa (L-DOPA) is a medication used to treat the symptoms of Parkinson's disease. When medicated, the displayed symptoms - including speech related symptoms - of a PwP can be successfully suppressed. This state is described as the ON state. When the medication wears off, the symptoms return; this is known as the OFF state.

SUMMARY

[0004] The present subject matter is directed towards a method for speech-based classification of the state of a PwP.

[0005] According to an embodiment of the present disclosure, it is provided a method for classifying speaker embeddings for a user having Parkinson's disease, the method comprising:

- receiving, by a computerized system, enrolment data wherein the data comprises speech samples of the user in an ON state and in an OFF state;
- storing the speech samples of the user;
- constructing from the enrolment data a profile of the user in an auxiliary model;
- monitoring a state of the user;
- receiving, by a computerized system, a test signal of the user;

extracting embeddings from a test signal;
computing distances between the test signal embeddings and the ON state and the OFF auxiliary models;
computing the difference between the two distances;
classifying the recognition signal as an ON state or an OFF state.

[0006] In accordance with another embodiment, the ON state is defined to be when the user is sufficiently medicated and wherein the OFF state is defined to be when the user is not medicated.

[0007] In accordance with another embodiment, the speech samples are about 25 phrases in each one of the ON state and the OFF state.

[0008] In accordance with another embodiment, the profile is an average of the speaker embeddings.

[0009] In accordance with another embodiment, the profile comprises two centroid vectors.

[0010] In accordance with another embodiment, one of the two centroid vectors is the ON centroid defined as an average of the speaker embeddings derived from the enrollment data in the ON state, and another one of the two centroid vectors is the OFF centroid defined as an average of the speaker embeddings derived from the enrollment data in the OFF state.

[0011] In accordance with another embodiment, the auxiliary model comprises a trainable part of 384 parameters.

[0012] In accordance with another embodiment, the method further comprising setting a threshold level defining the difference between the ON state and the OFF state.

[0013] In accordance with another embodiment, the monitoring of the state of the user is conducted in a passive manner.

[0014] In accordance with another embodiment, the threshold can be tuned.

[0015] In accordance with another embodiment, the distances are a first distance between the speaker embeddings and the ON centroid and a second distance between the speaker embeddings and the OFF centroid.

[0016] In accordance with another embodiment, if a difference in the distances between the first distance and the second distance exceeds the threshold, an ON state is defined, otherwise, an OFF state is defined.

BRIEF DESCRIPTION OF THE DRAWINGS

[0017] The subject matter is herein described, by way of example only, with reference to the accompanying drawings. With specific reference now to the drawings in detail, it is stressed that the particulars shown are by way of example and for purposes of illustrative discussion of the preferred embodiments of the present subject matter only, and are presented in the cause of providing what is believed to be the most useful and readily understood description of the principles and conceptual aspects of the subject matter. In this regard, no attempt is made to show structural details of the subject matter in more detail than is necessary for a fundamental understanding of the subject matter, the description taken with the drawings making apparent to those skilled in the art how the several forms of the subject matter may be embodied in practice.

In the drawings:

[0018] Fig. 1 shows a flow chart of enrollment method, in accordance with some exemplary embodiments of the disclosed subject matter;

[0019] Fig. 2 shows a flow chart of recognition method, in accordance with some exemplary embodiments of the disclosed subject matter;

DETAILED DESCRIPTION

[0020] The disclosed subject matter is described below with reference to flowchart illustrations and/or block diagrams of methods, apparatus (systems) and computer program products according to embodiments of the subject matter. It will be understood that each block of the flowchart illustrations and/or block diagrams, and combinations of blocks in the flowchart illustrations and/or block diagrams, can be implemented by computer program instructions. These computer program instructions may be provided to a processor of a general-purpose computer, special purpose computer, or other programmable data processing apparatus to produce a machine, such that the instructions, which execute via the processor of the computer or other programmable data processing apparatus, create means for implementing the functions/acts specified in the flowchart and/or block diagram block or blocks.

[0021] These computer program instructions may also be stored in a computer-readable medium that can direct a computer or other programmable data processing apparatus to function in a

particular manner, such that the instructions stored in the computer-readable medium produce an article of manufacture including instruction means which implement the function/act specified in the flowchart and/or block diagram block or blocks.

[0022] The computer program instructions may also be loaded onto a computer or other programmable data processing apparatus to cause a series of operational steps to be performed on the computer or other programmable apparatus to produce a computer implemented process such that the instructions which execute on the computer or other programmable apparatus provide processes for implementing the functions/acts specified in the flowchart and/or block diagram block or blocks.

[0023] Before explaining at least one embodiment of the subject matter in detail, it is to be understood that the subject matter is not limited in its application to the details of construction and the arrangement of the components set forth in the following description or illustrated in the drawings. The subject matter is capable of other embodiments or of being practiced or carried out in various ways. Also, it is to be understood that the phraseology and terminology employed herein is for the purpose of description and should not be regarded as limiting. The drawings are generally not to scale. For clarity, non-essential elements were omitted from some of the drawings.

[0024] There are several speaker embeddings models, such as ECAPA-TDNN, x-vectors, d-vectors and c-vectors speaker embedding models. The current subject matter uses such speaker identification models, preferably ECAPA-TDNN speaker model for the purpose of Parkinson's disease ON/OFF state detection.

[0025] The current subject matter uses a deep neural network that maps a single input audio signal to a compressed vector of 192 floating point numbers. These vectors are referred to as speaker embeddings. Two stages of processing are involved; enrollment and recognition-time, as further described below.

[0026] Referring now to Fig. 1 showing a flow chart of enrollment method, according to exemplary embodiments of the present subject matter.

[0027] During enrollment stage, in step 102, a user provides to a computerized system data comprising examples of his/her speech in the ON state and in the OFF state. The ON state is when the user is sufficiently medicated and symptoms are suppressed; accordingly the speech is characterized in one specific manner and the OFF state is when the user is insufficiently medicated and symptoms are displayed; and accordingly the speech is characterized in a second specific manner. The computerized process does not need a specific format or storage method, and does not require additional hardware. The speech samples can be general, not part of a specific dictionary or group of words or expressions, and only a small number of samples in each state are necessary. In a preferred embodiment, only 25 phrases in each state are required. The variety of phrases the user can use are, for example, everyday used words such as ‘turn on the lights’, ‘Alexa’ and similar.

[0028] In step 104, the computerized process stores the user’s speech samples and a profile of the speaker is constructed from these enrolment signals. The speaker profile is an average of the speaker embeddings derived from enrolment signals. The speaker profile comprises two centroid vectors. The profile of the speaker is saved as an auxiliary model. The centroid vectors are the same size as the speaker embedding vector, having 192 dimensions. These centroid vectors are as follows: The ON centroid: the average of the set of speaker embeddings derived from the set of enrollment signals in the ON state and the OFF centroid: the average of the set of speaker embeddings derived from the set of enrollment signals in the OFF state.

[0029] The trainable part of the model is relatively small - two 192-dimensional vectors (one for the ON state, one for the OFF state), so only 384 parameters. Since it is so small, it can be robustly estimated from a very small amount of data.

[0030] The model is designed to extract characteristics related to the speaker from input speech (accent, gender, state-of-health, etc.). These speaker characteristics remain relatively constant and stable across utterances of different phrases and words.

[0031] Referring now to Fig. 2 showing a flow chart of recognition method, according to exemplary embodiments of the present subject matter.

[0032] In optional step 202, the computerized process conducts monitoring of the user's state. The monitoring step can be passive. The state of the user is assessed from the speech signal, which is an audio file containing the user's speech. The speaker characteristics (i.e. the speaker embeddings) are extracted from the speech signal by the ECAPA-TDNN model. The speaker characteristics are insensitive to what is being said (i.e. the word sequence); but are sensitive to how it is said (accent, gender, stat-of-health, etc.).

[0033] In step 204, a threshold- a level, point, or value above which something is true or will take place while below that value, it is false or will not take place - is determined to classify the user's state as an ON state or an OFF state. The threshold tuning can be done by the user or can be automatically tuned by the computerized process.

[0034] In step 206, the computerized process receives a test signal from the user.

[0035] In step 208, the computerized process extracts recognition signal embeddings and computes the test signal.

[0036] In step 210, the computerized process computes a distance between the test signal and the speaker profile ON and OFF states. The distance between the test embedding and the ON (dist_ON) and OFF (dist_OFF) centroids for the test user are computed using negative cosine similarity metric, as an example.

[0037] In step 212, the computerized process computes the difference between the two distances computed in step 210 ($\text{dist_OFF} - \text{dist_ON}$). If the distance difference exceeds the threshold set in step 204, the utterance is classified as the ON state, otherwise it is classified as OFF state.

[0038] The flowchart in the Figures illustrate the architecture, functionality, and operation of possible implementations of systems, methods and computer program products according to various embodiments of the present subject matter. In this regard, each block in the flowchart or block diagrams may represent a module, segment, or portion of program code, which comprises one or more executable instructions for implementing the specified logical function(s). It should also be noted that, in some alternative implementations, the functions noted in the block may occur out of the order noted in the figures. For example, two blocks shown in succession may, in

fact, be executed substantially concurrently, or the blocks may sometimes be executed in the reverse order, depending upon the functionality involved. It will also be noted that each block of the block diagrams and/or flowchart illustration, and combinations of blocks in the block diagrams and/or flowchart illustration, can be implemented by special purpose hardware based systems that perform the specified functions or acts, or combinations of special purpose hardware and computer instructions.

[0039] The terminology used herein is for the purpose of describing particular embodiments only and is not intended to be limiting of the subject matter. As used herein, the singular forms "a", "an" and "the" are intended to include the plural forms as well, unless the context clearly indicates otherwise. It will be further understood that the terms "comprises" and/or "comprising," when used in this specification, specify the presence of stated features, integers, steps, operations, elements, and/or components, but do not preclude the presence or addition of one or more other features, integers, steps, operations, elements, components, and/or groups thereof.

[0040] The embodiment was chosen and described in order to best explain the principles of the subject matter and the practical application, and to enable others of ordinary skill in the art to understand the subject matter for various embodiments with various modifications as are suited to the particular use contemplated.

[0041] The present subject matter may be a system, a method, and/or a computer program product. The computer program product may include a computer readable storage medium (or media) having computer readable program instructions thereon for causing a processor to carry out aspects of the present subject matter.

[0042] The computer readable storage medium can be a tangible device that can retain and store instructions for use by an instruction execution device. The computer readable storage medium may be, for example, but is not limited to, an electronic storage device, a magnetic storage device, an optical storage device, an electromagnetic storage device, a semiconductor storage device, or any suitable combination of the foregoing. A non-exhaustive list of more specific examples of the computer readable storage medium includes the following: a portable computer diskette, a hard disk, a random access memory (RAM), a read-only memory (ROM), an erasable programmable read-only.

[0043] memory (EPROM or Flash memory), a static random access memory (SRAM), a portable compact disc read-only memory (CD-ROM), a digital versatile disk (DVD), a memory stick, a floppy disk, a mechanically encoded device such as punch-cards or raised structures in a groove having instructions recorded thereon, and any suitable combination of the foregoing. A computer readable storage medium, as used herein, is not to be construed as being transitory signals per se, such as radio waves or other freely propagating electromagnetic waves, electromagnetic waves propagating through a waveguide or other transmission media (e.g., light pulses passing through a fiber-optic cable), or electrical signals transmitted through a wire.

[0044] Computer readable program instructions described herein can be downloaded to respective computing/processing devices from a computer readable storage medium or to an external computer or external storage device via a network, for example, the Internet, a local area network, a wide area network and/or a wireless network. The network may comprise copper transmission cables, optical transmission fibers, wireless transmission, routers, firewalls, switches, gateway computers and/or edge servers. A network adapter card or network interface in each computing/processing device receives computer readable program instructions from the network and forwards the computer readable program instructions for storage in a computer readable storage medium within the respective computing/processing device.

[0045] Computer readable program instructions for carrying out operations of the present subject matter may be assembler instructions, instruction-set-architecture (ISA) instructions, machine instructions, machine dependent instructions, microcode, firmware instructions, state-setting data, or either source code or object code written in any combination of one or more programming languages, including an object oriented programming language such as Java, Smalltalk, python, C++ or the like, and conventional procedural programming languages, such as the "C" programming language or similar programming languages. The computer readable program instructions may execute entirely on the user's computer, partly on the user's computer, as a stand-alone software package, partly on the user's computer and partly on a remote computer or entirely on the remote computer or server. In the latter scenario, the remote computer may be connected to the user's computer through any type of network, including a local area network (LAN) or a wide area network (WAN), or the connection may be made to an external computer

(for example, through the Internet using an Internet Service Provider). In some embodiments, electronic circuitry including, for example, programmable logic circuitry, field-programmable gate arrays (FPGA), or programmable logic arrays (PLA) may execute the computer readable program instructions by utilizing state information of the computer readable program instructions to personalize the electronic circuitry, in order to perform aspects of the present subject matter.

[0046] Aspects of the present subject matter are described herein with reference to flowchart illustrations and/or block diagrams of methods, apparatus (systems), and computer program products according to embodiments of the subject matter. It will be understood that each block of the flowchart illustrations and/or block diagrams, and combinations of blocks in the flowchart illustrations and/or block diagrams, can be implemented by computer readable program instructions.

[0047] These computer readable program instructions may be provided to a processor of a general-purpose computer, special purpose computer, or other programmable data processing apparatus to produce a machine, such that the instructions, which execute via the processor of the computer or other programmable data processing apparatus, create means for implementing the functions/acts specified in the flowchart and/or block diagram block or blocks. These computer readable program instructions may also be stored in a computer readable storage medium that can direct a computer, a programmable data processing apparatus, and/or other devices to function in a particular manner, such that the computer readable storage medium having instructions stored therein comprises an article of manufacture including instructions which implement aspects of the function/act specified in the flowchart and/or block diagram block or blocks.

[0048] The computer readable program instructions may also be loaded onto a computer, other programmable data processing apparatus, or other device to cause a series of operational steps to be performed on the computer, other programmable apparatus or other device to produce a computer implemented process, such that the instructions which execute on the computer, other programmable apparatus, or other device implement the functions/acts specified in the flowchart and/or block diagram block or blocks.

CLAIMS

1. A method for classifying speaker embeddings for a user having Parkinson's disease, the method comprising:

- receiving, by a computerized system, enrolment data wherein the data comprises speech samples of the user in an ON state and in an OFF state;
- storing the speech samples of the user;
- constructing from the enrolment data a profile of the user in an auxiliary model;
- monitoring a state of the user;
- receiving, by a computerized system, a test signal of the user;
- extracting embeddings from a test signal;
- computing distances between the test signal embeddings and the ON state and the OFF auxiliary models;
- computing the difference between the two distances;
- classifying the recognition signal as an ON state or an OFF state.

2. The method for classifying speaker embeddings for a user having Parkinson's disease as claimed in Claim 1, wherein the ON state is defined to be when the user is sufficiently medicated and wherein the OFF state is defined to be when the user is not medicated.

3. The method for classifying speaker embeddings for a user having Parkinson's disease as claimed in Claim 1, wherein the speech samples are about 25 phrases in each one of the ON state and the OFF state.

4. The method for classifying speaker embeddings for a user having Parkinson's disease as claimed in Claim 1, wherein the profile is an average of the speaker embeddings.

5. The method for classifying speaker embeddings for a user having Parkinson's disease as claimed in Claim 1, wherein the profile comprises two centroid vectors.
6. The method for classifying speaker embeddings for a user having Parkinson's disease as claimed in Claim 5, wherein one of the two centroid vectors is the ON centroid defined as an average of the speaker embeddings derived from the enrollment data in the ON state, and another one of the two centroid vectors is the OFF centroid defined as an average of the speaker embeddings derived from the enrollment data in the OFF state.
7. The method for classifying speaker embeddings for a user having Parkinson's disease as claimed in Claim 1, wherein the auxiliary model comprises a trainable part of 384 parameters.
8. A method for classifying speaker embeddings for a user having Parkinson's disease as claimed in Claim 1, further comprising setting a threshold level defining the difference between the ON state and the OFF state.
9. A method for classifying speaker embeddings for a user having Parkinson's disease as claimed in Claim 1, wherein the monitoring of the state of the user is conducted in a passive manner.
10. The method for classifying speaker embeddings for a user having Parkinson's disease as claimed in Claim 9, wherein the threshold can be tuned.

11. The method for classifying speaker embeddings for a user having Parkinson's disease as claimed in Claim 9, wherein the distances are a first distance between the speaker embeddings and the ON centroid and a second distance between the speaker embeddings and the OFF centroid.

12. The method for classifying speaker embeddings for a user having Parkinson's disease as claimed in Claim 9, wherein if a difference in the distances between the first distance and the second distance exceeds the threshold, an ON state is defined, otherwise, an OFF state is defined

1/1

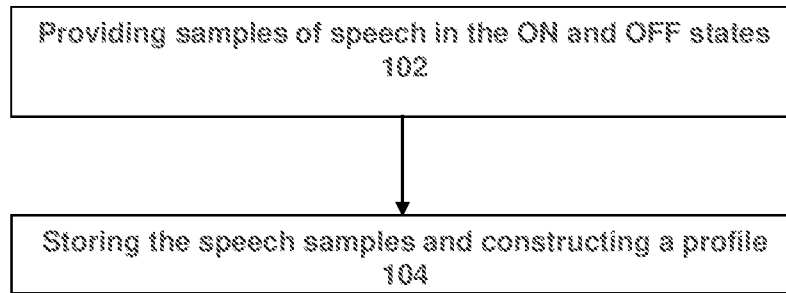


FIG. 1

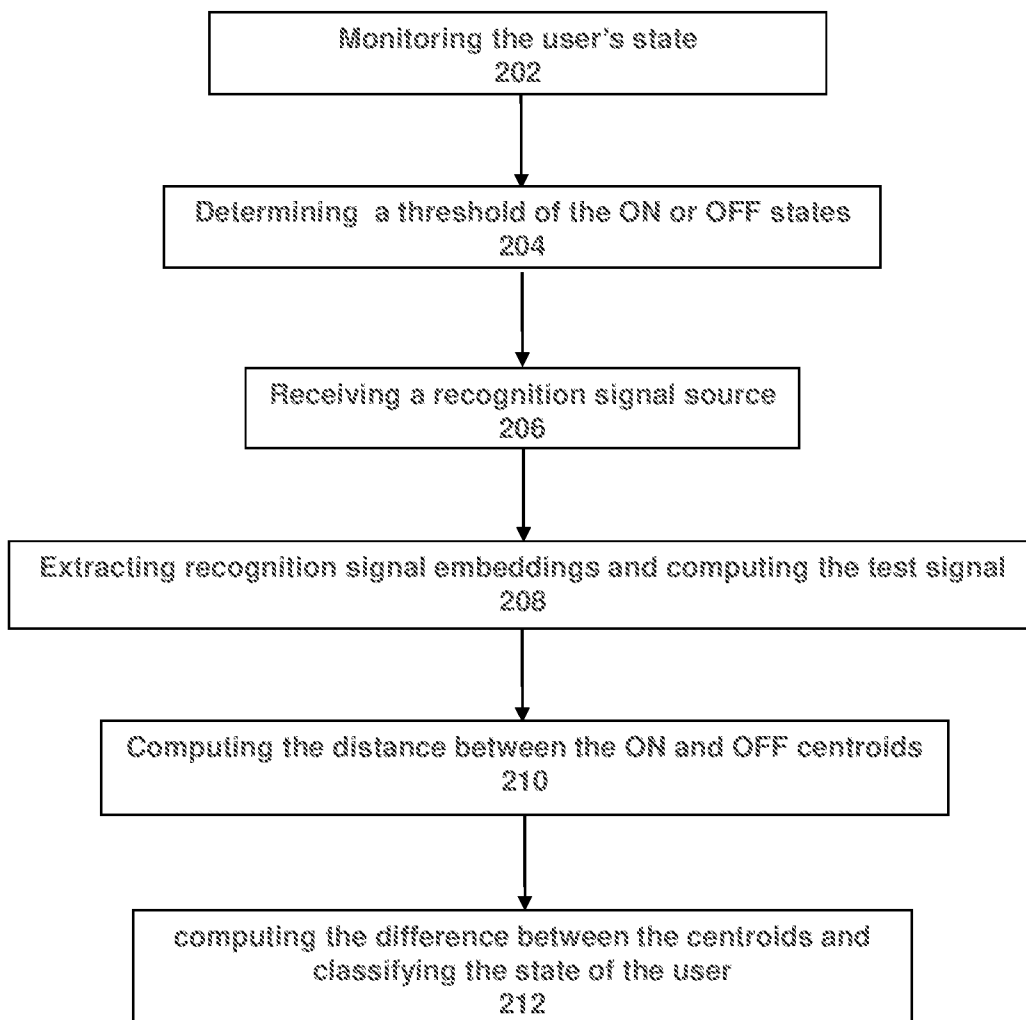


FIG. 2