

# (12)Indian Patent Application

---

(21) Application Number: 202241015587

(22) Filing Date: 21/03/2022      (43) Publication Date: 22/09/2023

(71) Applicant(s): L&T TECHNOLOGY SERVICES LIMITED

(72) Inventor(s): Motwani, Ashish  
Gawde, Neha Vidhyadhar

(51) International Classifications: G10L 15/22      H04W 56/00      H01M 16/00      H04N 17/04      H04M 1/24

(54) Title: A METHOD AND SYSTEM FOR TESTING SPEECH DETECTION

(57) Abstract: A method and a system for testing speech detection is disclosed. The method may include generating a speech-based test command using a text-based test command, and inputting the speech-based test command to a device-under-test. The speech-based test command may be processed by the device-under-test to generate an output command. The method may further include receiving the output command from the device-under-test corresponding to the speech-based test command, and comparing the text-based test command inputted to the device-under-test and the output command generated by the device-under-test and determine an accuracy of the output command generated by the device-under-test corresponding to the speech-based test command based on the comparison. The method may further include generating a report corresponding to the accuracy of the output command generated by the device-under-test corresponding to the speech-based test command based on the comparison.

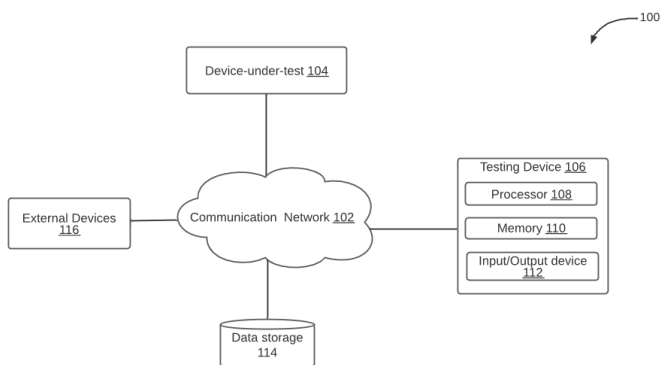


FIG. 1 DUANU DPACAN (INDA No. 2252)

# **FORM 2**

THE PATENTS ACT 1970  
(39 OF 1970)  
&  
The Patent Rules, 2003  
**Complete Specification**  
(See Section 10 and Rule 13)

## **1. TITLE OF THE INVENTION**

**A METHOD AND SYSTEM FOR TESTING SPEECH DETECTION**

## **2. APPLICANT(S)**

(a) NAME : **L&T TECHNOLOGY SERVICES LIMITED**

(b) NATIONALITY : **INDIAN**

(c) ADDRESS : **DLF IT SEZ Park, 2nd Floor – Block 3**

**1/124, Mount Poonamallee Road,**

**Ramapuram, Chennai – 600 089,**

**INDIA.**

## **3. PREAMBLE TO THE DESCRIPTION**

**COMPLETE**

The following specification particularly describes the invention and the manner in which it is to be performed.

## **DESCRIPTION**

### **Technical Field**

**[001]** This disclosure relates generally to speech detection, and more particularly to a method and system of testing of speech detection performed by a device-under-test.

### **Background**

**[002]** Voice assistant technology has seen rapid adoption in electronic products ranging from smart televisions and infotainment systems to Bluetooth speakers. Voice assistance allows the users to control these electronic products through speech, thereby allowing hands-free operation of these electronic products. The basic form of voice control through a predefined set of commands such as “increase temperature”, “decrease temperature”, “play next song” etc. has become an indispensable requirement for the users. Voice assistance tends to be quicker, more reliable, and less invasive as compared to touch-based or mechanical button-based interfaces, as they do not need the users to take their eyes and hands off their primary job. Constant efforts are made to make the customer’s experience as smooth and as appealing as possible by the use of speech assistance interactions. Further, the speech assistance technology is frequently updated throughout the development cycle as well based on the usage.

**[003]** However, in some cases, the speech assistance is not able to correctly identify the voice commands provided by the user. For example, the speech assistance either is unable to interpret the command or interprets it inaccurately. Detecting the inaccuracies in speech detection plays an important role in updating the speech detection systems to overcome the inaccuracies. However, efficiently detecting the inaccuracies is a challenge, especially when the commands are in multiple different languages.

**[004]** Accordingly, there is a need for an automated solution for effectively testing speech detection performed by devices, especially with respect to different dialects and with speech modulations associated different speakers.

## SUMMARY

**[005]** In an embodiment, a method of testing speech detection is disclosed. The method may include generating a speech-based test command using a text-based test command. The method may further include inputting the speech-based test command to a device-under-test. The speech-based test command may be processed by the device-under-test to generate an output command. The method may further include receiving the output command from the device-under-test corresponding to the speech-based test command. The output command may be text-based. The method may further include comparing the text-based test command inputted to the device-under-test and the output command generated by the device-under-test, and determining an accuracy of the output command generated by the device-under-test corresponding to the speech-based test command based on the comparison. The method may further include generating a report corresponding to the accuracy of the output command generated by the device-under-test corresponding to the speech-based test command based on the comparison.

**[006]** In another embodiment, a system for testing speech detection is disclosed. The system includes a device-under-test and a testing device. The testing device includes a processor and a memory communicatively coupled to the processor. The memory stores processor-executable instructions, which, on execution, cause the processor to generate a speech-based test command using a text-based test command. The processor-executable instructions may further cause the processor to input the speech-based test command to a device-under-test. The speech-based test command is to be processed by the device-under-test to generate an output command. The processor-executable instructions may further cause the processor to receive the output command from the device-under-test corresponding to the speech-based test command. The processor-executable instructions may further cause the processor to compare the text-based test command inputted to the device-under-test and the output command generated by the device-under-test, and further determine accuracy of the output

command generated by the device-under-test corresponding to the speech-based test command based on the comparison. The processor-executable instructions may further cause the processor to generate a report corresponding to the accuracy of the output command generated by the device-under-test corresponding to the speech-based test command based on the comparison.

[007] It is to be understood that both the foregoing general description and the following detailed description are exemplary and explanatory only and are not restrictive of the invention, as claimed.

### **BRIEF DESCRIPTION OF THE DRAWINGS**

[008] The accompanying drawings, which are incorporated in and constitute a part of this disclosure, illustrate exemplary embodiments and, together with the description, serve to explain the disclosed principles.

[009] **FIG. 1** illustrates a block diagram of an exemplary system for testing speech detection, in accordance with some embodiments of the present disclosure.

[010] **FIG. 2** illustrates a functional block diagram of a testing device, in accordance with some embodiments of the present disclosure.

[011] **FIG. 3** illustrates a process flow diagram of a process of generating a speech-based test command using a text-based test command, in accordance with an embodiment of the present disclosure.

[012] **FIG. 4A** illustrates a process flow diagram of a process of training and testing of a ML (deep learning) model, in accordance with an embodiment of the present disclosure.

[013] **FIG. 4B** illustrates a process flow diagram of an overall process of generating a speech-based test command using the trained ML model, in accordance with an embodiment of the present disclosure.

[014] **FIG. 5** illustrates a snapshot of a Graphical User Interface (GUI) associated with the generation of report, in accordance with an embodiment of the present disclosure.

[015] FIG. 6 is a flowchart of a method of testing speech detection, in accordance with some embodiments of the present disclosure.

[016] FIG. 7 is a flowchart of a method of generating a report, in accordance with some embodiments of the present disclosure.

### **DETAILED DESCRIPTION**

[017] Exemplary embodiments are described with reference to the accompanying drawings. Wherever convenient, the same reference numbers are used throughout the drawings to refer to the same or like parts. While examples and features of disclosed principles are described herein, modifications, adaptations, and other implementations are possible without departing from the spirit and scope of the disclosed embodiments. It is intended that the following detailed description be considered as exemplary only, with the true scope and spirit being indicated by the following claims. Additional illustrative embodiments are listed below.

[018] One or more techniques for testing speech detection performed by a device-under-test are disclosed. The scope of the present disclosure extends to a wide range of devices including car infotainment systems, smart devices like smart televisions, smartphones, etc. The techniques disclosed provide one stop solution for speech testing providing comprehensive coverage across languages. Further, the present techniques are supported by various operating systems (OS) like Linux, Android, etc. Furthermore, the present techniques cover voice samples with different languages, accents, pitch, amplitude, gender, environment, etc.

[019] The present techniques help in the reduction of the overall time and cost associated with testing and validating the device-under-test as no natural language speaker is required to test and validate the speech utterances. Further, the present techniques offer reduction in analysis efforts since the results are evaluated based on set of rules for testing gender-specific commands, environment-specific commands, etc. Furthermore, the present techniques provide for generating speech-based

commands in multiple languages with possible configuration changes. Moreover, the present techniques support commonly used NLP languages (along with modulation, pitch, and gender variation) as well as non-NLP languages.

**[020]** The present techniques further aid in the generation and execution of speech commands in one or more supported languages. Further, by training the system to understand multiple dialects and different speaker categories, the present techniques cut down on analysis time while also lowering the number of false failure cases. Furthermore, the present techniques provide for report generation based on the specifics on the failures of speech detection with device-under-test that have been observed during validating the speech utterances. For example, a user may select specific speech modulation types (based on gender, age group, pitch, amplitude of the speaker. As such, the present techniques may help identifying the root cause with respect to the failures. Therefore, overall, the present techniques provide for ease of implementation, maintainability, and scalability.

**[021]** Referring now to **FIG. 1**, a block diagram of an exemplary system 100 for testing speech detection is illustrated, in accordance with an embodiment of the present disclosure. The system 100 may include a testing device 106 and a device-under-test 104. The testing device 106 may be a computing device having data processing capability. In particular, the testing device 106 may have capability for testing the speech detection performed by the device-under-test 104. Examples of the testing device 106 may include, but are not limited to a desktop, a laptop, a notebook, a netbook, a tablet, a smartphone, a mobile phone, an application server, a web server, or the like.

**[022]** The system 100 may further include a data storage 114. For example, the data storage 114 may store various types of data required by the testing device 106 for testing speech detection. The testing device 106 may be communicatively coupled to the data storage 114 and the device-under-test 104 via a communication network 102. The device-under test 104 may be one of an infotainment system, a smart device like a smart television, or a smartphone. The communication network 102 may include

a communication medium through which the testing device 106, the device-under-test 104, the data storage 114, and the external device 116 may communicate with each other. Examples of the communication network 102 may include, but are not limited to, the Internet, a cloud network, a Wireless Fidelity (Wi-Fi) network, a Personal Area Network (PAN), a Local Area Network (LAN), or a Metropolitan Area Network (MAN). Various devices in the system 100 may be configured to connect to the communication network 102, in accordance with various wired and wireless communication protocols. Examples of such wired and wireless communication protocols may include, but are not limited to, a Transmission Control P Protocol and Internet Protocol (TCP/IP), User Datagram Protocol (UDP), Hypertext Transfer Protocol (HTTP), File Transfer Protocol (FTP), Zig Bee, EDGE, IEEE 802.11, light fidelity(Li-Fi), 802.16, IEEE 802.11s, IEEE 802.11g, multi-hop communication, wireless access point (AP), device to device communication, cellular communication protocols, and Bluetooth (BT) communication protocols.

**[023]** The testing device 106 may be configured to perform one or more functionalities that may include generating a speech-based test command using a text-based test command, and inputting the speech-based test command to the device-under-test 104. The speech-based test command is to be processed by the device-under-test 104 to generate an output command. The output command may be, for example, a set of executable instructions for carrying out operations associated with the command. For example, when a voice command “increase volume” is provided to the device-under-test 104 (e.g. an infotainment system of a car), the device-under-test 104 may interpret this voice command to generate an processor-executable (output command) instruction to carry out the operation of increasing the volume.

**[024]** The one or more functionalities may further include receiving the output command from the device-under-test 104 corresponding to the speech-based test command. The one or more functionalities may further include comparing the text-based test command inputted to the device-under-test 104 and the output command generated by the device-under-test 104. The one or more functionalities may further

include determining an accuracy of the output command generated by the device-under-test 104 corresponding to the speech-based test command.

**[025]** Additionally, the one or more functionalities may further include generating a report corresponding to the accuracy of the output command generated by the device-under-test 104 corresponding to the speech-based test command based on the comparison.

**[026]** In order to perform the above-discussed functionalities, the testing device 106 may include a processor 108 and a memory 110. The processor 108 may include suitable logic, circuitry, interfaces, and/or code that may be configured to assess the quality of the multimedia content. The processor 108 may be implemented based on temporal and spatial number of processor technologies, which may be known to one ordinarily skilled in the art. Examples of implementations of the processor 108 may be a Graphics Processing Unit (GPU), a Reduced Instruction Set Computing (RISC) processor, an Application-Specific Integrated Circuit (ASIC) processor, a Complex Instruction Set Computing (CISC) processor, a microcontroller, Artificial Intelligence (AI) accelerator chips, a co-processor, a central processing unit (CPU), and/or a combination thereof. The memory 110 may be a non-volatile memory or a volatile memory. Examples of non-volatile memory may include, but are not limited to a flash memory, a Read Only Memory (ROM), a Programmable ROM (PROM), Erasable PROM (EPROM), and Electrically EPROM (EEPROM) memory. Examples of volatile memory may include but are not limited to Dynamic Random Access Memory (DRAM), and Static Random-Access memory (SRAM). The memory 110 may also store various asset data (e.g. asset identification ID) that may be captured, processed, and/or required by the system 100.

**[027]** The testing device 106 may further include Input/Output (I/O) devices 112. Examples may include, but are not limited to a display, keypad, microphone, audio speakers, vibrating motor, LED lights, etc. The I/O devices 112 may receive input from a user and also display an output of the computation performed by the processor 108. For example, the user input may include selection of at least one speech modulation

type set from the plurality of speech modulation type sets for the selection of the speech-based test command. Further, the display of the I/O device 112 may include a display screen which is capable of displaying the report representing the test results of the accuracy of the output command generated by the device-under-test 104 corresponding to the speech-based test command.

**[028]** Additionally, the testing device 106 may be communicatively coupled to an external device 116 for sending and receiving various data. Examples of the external device 116 may include, but is not limited to, a remote server, digital devices, and a computer system. The testing device 106 may connect to the external device 116 over the communication network 102. The testing device 106 may connect to external device 116 via a wired connection, for example via Universal Serial Bus (USB).

**[029]** Referring now to **FIG. 2**, a functional block diagram of a testing device 106 is illustrated, in accordance with an embodiment of the present disclosure. As mentioned above, the testing device 106 may be configured to test speech detection performed by the device-under-test 104. The testing device 106 may include a script generating module 202, command generating module 204, a command inputting module 206, an output command receiving module 208, a comparing module 210, an accuracy determining module 212, a report generating module 214, and a report analyzing module 216.

**[030]** It should be noted that all such aforementioned modules 202-216 may be represented as a single module or a combination of different modules. Further, as will be appreciated by those skilled in the art, each of the modules 202-216 may reside, in whole or in parts, on one device or multiple devices in communication with each other. In some embodiments, each of the modules 202-216 may be implemented as dedicated hardware circuit comprising custom application-specific integrated circuit (ASIC) or gate arrays, off-the-shelf semiconductors such as logic chips, transistors, or other discrete components. Each of the modules 202-216 may also be implemented in a programmable hardware device such as a field programmable gate array (FPGA), programmable array logic, programmable logic device, and so forth. Alternatively,

each of the modules 202-216 may be implemented in software for execution by various types of processors. An identified module of executable code may, for instance, include one or more physical or logical blocks of computer instructions, which may, for instance, be organized as an object, procedure, function, or other construct. Nevertheless, the executables of an identified module or component need not be physically located together, but may include disparate instructions stored in different locations which, when joined logically together, include the module and achieve the stated purpose of the module. Indeed, a module of executable code could be a single instruction, or many instructions, and may even be distributed over several different code segments, among different applications, and across several memory devices.

**[031]** The script generating module 202 may be configured to generate one or more text-based test commands. In particular, the script generating module 202 may generate a plurality of text variations corresponding to a text-based test command. It should be noted that the plurality of text variations may be generated by varying one or more parameters associated with each of plurality of text variations. By way of an example, the one or more parameters may include a script, a grammar usage, and a speech style. By way of an example, machine learning (ML)-based techniques may be used for generating the plurality of text variations. By way of another example, the script generating module 202 may receive one or more grammar files or text-based commands in one or more languages, and apply external functions based on command wise preconditions and postcondition execution code, for generating the plurality of text variations. The script generating module 202 may read/parse input data and then perform command-wise execution based on the external functions, and create a script corresponding to the execution flow with verification points. In some embodiments, a Graphical User Interface (GUI) (which is explained in detail in conjunction with FIG. 5) may be used by a user to select a language (e.g. “English UK” as shown in FIG. 5) via a tab provided on the GUI, for generating the text-based test command.

**[032]** The command generating module 204 may be configured to generate a speech-based test command using the text-based test command. It should be noted

that a plurality of speech-based test commands may be generated corresponding to each of the plurality of text variations using a trained ML (deep learning) model. Each speech-based test command of the plurality of speech-based test commands may be associated with a unique speech modulation type set. It should be noted that the plurality of speech modulation type sets may be associated with one of a gender type, an accent type, an age type, a pitch type, an amplitude type, an environment type, etc. The process of generation the speech-based test command using the ML model is further explained in detail, in conjunction with FIGs. 3 and 4A-4B.

**[033]** Referring now to **FIG. 3**, a process flow diagram of a process 300 of generating a speech-based test command using a text-based test command is illustrated, in accordance with an embodiment of the present disclosure. By way of an example, the process 300 may be performed by the command generating module 204 of the testing device 106, using a trained ML model 310.

**[034]** A voice sample 302 may be received and processed by a speech encoder 304. The speech encoder 304 may create an embedding vector (a fixed dimensional vector representation) corresponding to the inputted voice sample 302. Further, a text command 306 may be received and processed by a text encoder 308. The text encoder 308 may create an embedding vector corresponding to the text command 306. It should be noted that the text command 306 may include a command in text format for which one or more speech-based commands are to be generated. Further, the voice sample 302 may be provided to create the speech-based commands corresponding to the speech modulation type associated with the voice sample 302.

**[035]** The embedding vector corresponding to the inputted voice sample 302 and the embedding vector corresponding to the text command 306 may be fed to the trained ML model 310. The trained ML model 310 may combine the embedding vector corresponding to the inputted voice sample 302 and the embedding vector corresponding to the text command 306 to generate a combined output. The decoder 312 may decode the combined output into a spectrogram. The vocoder 314 may transform the spectrogram into an audio waveform, i.e. the speech-based command.

[036] Referring now to **FIG. 4A**, a process flow diagram of a process 400 of training and testing of a ML (deep learning) model is illustrated, in accordance with an embodiment of the present disclosure. Further, as already discussed above, the ML model may be configured to generate a plurality of speech-based test commands corresponding to each of the plurality of text variations. To this end, the ML model may be trained using a dataset 402.

[037] The dataset 402 may include a training voice and text dataset 402A. It should be noted that a customized configuration of the ML model 404 may be first trained using the training voice and text dataset 402A, before carrying out the testing. Based on the training, a trained ML model 406 is created. As such, the training voice and text dataset 402A may include a plurality of text-based test commands and their corresponding speech-based test commands. Further, the training voice and text dataset 402A may include multiple speech-based test commands with varying speech modulation types, corresponding to each text-based test command.

[038] The dataset 402 may further include a testing voice and text dataset 402B. Once the ML model is trained, the ML model may be used for generating the speech-based test commands based on the testing voice and text dataset 402B. The speech-based test commands generated by the ML model may then be verified manually. As such, the testing voice and text dataset 402B may also act as training dataset, that plays a part in the training of the ML model.

[039] Referring now to **FIG. 4B**, a process flow diagram of an overall process 400B of generating a speech-based test command using the trained ML model 406 is illustrated, in accordance with an embodiment of the present disclosure. An input text and the corresponding speaker embedding 408 may be received for generating the output audio 410 (speech-based test command). The input text and the corresponding speaker embedding 408 may be created for the testing voice and text dataset 402B using any embedding/encoding application. Thereafter, the input text and the corresponding speaker embedding 408 may be fed to the trained ML model 406

(obtained via process 400A). The trained ML model 406 may then output the speech-based test command (output audio) 410.

**[040]** Returning to FIG. 2, the command inputting module 206 may be configured to input the speech-based test command to the device-under-test 104. In other words, the speech-based test command may be inputted by the testing device 106 to the device-under-test 104. As will be understood, this speech-based test command inputted by the testing device 106 may be an alternative to speech-based test command otherwise provided by a human user to the device-under-test 104. Further, in order to test the speech detection capability of the device-under-test 104, the device-under-test 104 may be tested with multiple speech-based test commands varying in terms of different speech modulation types. For example, the speech modulation types may vary based on a gender type (e.g. a male voice or a female voice), an accent type (e.g. regional accent like American accent, or Indian accent, etc.), an age type (e.g. young speaker, middle-aged speaker, elderly speaker, etc.), a pitch type (i.e. degree of highness or lowness of a tone), an amplitude type (i.e. loudness), and an environment type (e.g. a noisy environment, a echoing environment, etc.).

**[041]** The output command receiving module 208 may be configured to receive an output command from the device-under-test 104 corresponding to the speech-based test command. It should be noted that the output command may be text-based. As explained above, the output command may be, for example, a set of executable instructions for carrying out operations associated with the command. For example, when a voice command “increase volume” is provided to the device-under-test 104 (e.g. an infotainment system of a car), the device-under-test 104 may interpret this voice command to generate an processor-executable (output command) instruction to carry out the operation of increasing the volume.

**[042]** The comparing module 210 may be configured to compare the text-based test command inputted to the device-under-test 104 and the output command corresponding to the speech-based test command generated by the device-under-test 104. As mentioned above, in some embodiments, the output command may be in text

format. As such, the comparing module 210 may perform a text-based comparison between the text-based test command (inputted to the device-under-test 104) and the output command corresponding to the speech-based test command generated by the device-under-test 104.

**[043]** The accuracy determining module 212 may determine accuracy of the output command generated by the device-under-test 104 (corresponding to the speech-based test command) based on the comparison. The accuracy determining module 212 may therefore determine how accurately the device-under-test 104 interprets the speech-based test command.

**[044]** The report generating module 214 may generate a test report corresponding to the accuracy of the output command generated by the device-under-test 104 corresponding to the speech-based test command based on the comparison. For example, the test report may be displayed using the display screen of the input/output device 112. In particular, in some embodiments, the report generating module 214 may receive a selection of at least one speech modulation type set from a plurality of speech modulation type sets via a user-interface. Further, it should be noted that the user-interface may also be used for starting speech execution, i.e. by selecting the type of speech modulation type set on which the device-under-test is desired to be tested. The user-interface is further explained in detail conjunction with FIG. 5.

**[045]** Referring now to **FIG. 5**, a snapshot of a Graphical User Interface (GUI) 500 for receiving a selection of at least one speech modulation type set is illustrated, in accordance with an embodiment of the present disclosure. The GUI 500 may be implemented via a webpage in an example embodiment. For example, the GUI 500 may enable a user to configure their preferred speech modulation type sets for generating a script and a customized report.

**[046]** The GUI 500 may provide a plurality of options related with speech modulation type sets. By way of an example, the GUI 500 may include a gender type speech modulation type set 502A, a generation (i.e. age) type speech modulation type set 502B, a modulation type (i.e. a pitch and amplitude type) speech modulation type

set 502C, and an environment type speech modulation type set 502D. By way of an example, the gender type speech modulation type set 502A may include options for selecting a male voice type or a female voice or both. The generation type speech modulation type set 502B may include options for selecting an old (age) type, a child type, a young type, or all types (i.e. all age groups). Further, the modulation type speech modulation type set 502C may include options for selecting a soft (sound modulation) type, a loud (sound modulation) type, or both. The environment type speech modulation type set 502D may include options for selecting a noisy type, a silent type, or both. Additionally, the GUI 500 may include analysis categories what may allow the user to select a standard speech modulation type set 502E or a random speech modulation type set 502F. As will be understood, the standard speech modulation type set 502E may include certain one or more preselected speech modulation type sets. Further, the random speech modulation type set 502F may include one or more randomly selected speech modulation type sets.

**[047]** A user may provide their selection of at least one speech modulation type set from the plurality of speech modulation type sets 502A-502F via the GUI 500. To this end, for example, the GUI 500 include a radio button provided alongside each of the plurality of speech modulation type sets 502A-502F. Therefore, the selection from the user may be received by way of the user clicking/touching a radio button associated with each of plurality of speech modulation type sets 502A-502F that the user desires to select. Further, as mentioned earlier, the GUI 500 may allow the user to select a language, e.g. “English UK”, as shown in FIG. 5, via a tab 504. The text-based test commands will be generated based on the selected language. Once the text-based test command is generated, the GUI 500 may be used for speech execution, i.e. to start generating speech-based test commands.

**[048]** Returning to FIG. 2, the report generating module 214 may generate the report corresponding to the selection provided by the user. Once the selection is received, the report generating module 214 may coordinate with the command inputting module 206 to provide one or more speech-based test commands associated

with the selected at least one speech modulation type set to the device-under-test 104. Further, the report generating module 214 may coordinate with the output command receiving module 208 to obtain one or more output commands from the device-under-test 104 corresponding to the one or more speech-based test commands. Furthermore, the report generating module 214 may coordinate with the comparing module 210 to obtain comparison result of the comparison of the one or more output commands corresponding to the one or more speech-based test commands with the text-based test command. Further, the report generating module 214 may coordinate with the accuracy determining module 212 to obtain the accuracy of the one or more output commands corresponding to the one or more speech-based test commands with respect to the text-based test command. The report generating module 214 may further generate the report based on the accuracy of the one or more output commands corresponding to the one or more speech-based test commands with respect to the text-based test command.

**[049]** The report analyzing module 216 may analyze the report by comparing the output command generated by the device-under-test 104 with at least one of historical data and current data. The report may be analyzed to identify root cause of inaccuracies in the speech detection performed by the device-under-test 104.

**[050]** Referring now to **FIG. 6**, a flowchart of a method 600 for testing speech detection performed by the device-under-test 104 is illustrated, in accordance with an embodiment of the present disclosure. By way of an example, the method 600 may be performed by the testing device 106. The device-under-test 104 may be an infotainment system, or a smart device like a smart television, or a smartphone.

**[051]** At step 602, a speech-based test command may be generated using a text-based test command. In some embodiments, in order to generate the speech-based test command, steps 602A-602C may be performed. At step 602A, a plurality of text variations corresponding to the text-based test command may be generated. At step 602B, a plurality of speech-based test commands corresponding to each of the plurality of text variations may be generated, using a trained ML model. Each speech-based test command of the plurality of speech-based test commands may be associated with a

unique speech modulation type set of a plurality of speech modulation type sets. It should be noted that the speech modulation type set of the plurality of speech modulation type sets is associated with one of a gender type, an accent type, an age type, a pitch type, an amplitude type, and an environment type. Further, the plurality of text variations may be generated by varying one or more parameters associated with each of plurality of text variations. The one or more parameters may include a script, a grammar usage, and a speech style.

**[052]** In some embodiments, in order to generate the plurality of speech-based test commands corresponding to each of the plurality of text variations, a voice sample may be inputted to the trained ML model. In response to the inputting, the plurality of speech-based test commands corresponding to each of the plurality of text variations may be obtained, based on the voice sample. At step 602C, the speech-based test command from the plurality of speech-based test commands may be selected for testing the device-under-test.

**[053]** At step 604, the speech-based test command may be inputted to the device-under-test 104. The speech-based test command is to be processed by the device-under-test 104 to generate an output command. At step 606, the output command may be received from the device-under-test 104 corresponding to the speech-based test command. The output command may be text-based. At step 608, the text-based test command inputted to the device-under-test 104 may be compared with the output command generated by the device-under-test 104. At step 610, an accuracy of the output command generated by the device-under-test 104 may be determined corresponding to the speech-based test command, based on the comparison. Additionally, in some embodiments, at step 612, a report may be generated corresponding to the accuracy of the output command generated by the device-under-test 104 corresponding to the speech-based test command, based on the comparison. The step 612 of report generation is further explained in detail in conjunction with FIG. 7.

**[054]** Referring now to **FIG. 7**, a flowchart of a method 700 of generating the report corresponding to the accuracy of the output command generated by the device-under-test 104 is illustrated, in accordance with an embodiment of the present disclosure. By way of an example, the method 700 may be performed by the report generating module 214 of the testing device 106.

**[055]** At step 702, a selection of at least one speech modulation type set from the plurality of speech modulation type sets 502A-502F may be received, via a user-interface, for example, the GUI 500. At step 704, the one or more speech-based test commands associated with the at least one speech modulation type set may be inputted to the device-under-test 104. At step 706, one or more output commands from the device-under-test 104 may be received corresponding to the one or more speech-based test commands. At step 708, the one or more output commands corresponding to the one or more speech-based test commands may be compared with the text-based test command. At step 710, the accuracy of the one or more output commands corresponding to the one or more speech-based test commands may be determined with respect to the text-based test command. At step 712, the report may be generated based on the accuracy of the one or more output commands corresponding to the one or more speech-based test commands with respect to the text-based test command.

**[056]** Returning to **FIG. 6**, at step 614, the report may be analyzed by comparing the output command generated by the device-under-test 104 with at least one of historical data and current data. In particular, the comparison may include comparing results of the input command in different speech modulation types (i.e. languages, environment, gender, etc.) with results seen in the past execution runs. For example, the report may be analyzed to identify root cause of inaccuracies in the speech detection performed by the device-under-test 104. In other words, when an accuracy of the output command generated by the device-under-test 104 is low, the report may be analyzed for identifying the possible reason behind the low accuracy. For example, analysis of the report may help identify that the accuracy of speech detection performed by the device-under-test 104 is low for a certain speech modulation type as compared

to others speech modulation types. Accordingly, the analysis of the report may be used for taking a corrective action to improve the accuracy with respect to that speech modulation type. For example, the corrective action may include retraining the ML model with input dataset associated with that speech modulation type.

**[057]** It is intended that the disclosure and examples be considered as exemplary only, with a true scope and spirit of disclosed embodiments being indicated by the following claims.

## CLAIMS

We claim:

1. A method of testing speech detection, the method comprising:
  - generating, by a testing device, a speech-based test command using a text-based test command;
  - inputting, by the testing device, the speech-based test command to a device-under-test, wherein the speech-based test command is to be processed by the device-under-test to generate an output command;
  - receiving, by the testing device, the output command from the device-under-test corresponding to the speech-based test command, wherein the output command is text-based;
  - comparing, by the testing device, the text-based test command inputted to the device-under-test and the output command generated by the device-under-test; and
  - determining, by the testing device, an accuracy of the output command generated by the device-under-test corresponding to the speech-based test command, based on the comparison.
  
2. The method as claimed in claim 1, wherein generating the speech-based test command comprises:
  - generating a plurality of text variations corresponding to the text-based test command;
  - generating a plurality of speech-based test commands corresponding to each of the plurality of text variations, using a trained ML model;
  - wherein each speech-based test command of the plurality of speech-based test commands is associated with a unique speech modulation type set of a plurality of speech modulation type sets; and
  - selecting the speech-based test command from the plurality of speech-based test commands for testing the device-under-test.

3. The method as claimed in claim 2, wherein the speech modulation type set of the plurality of speech modulation type sets is associated with one of: a gender type, an accent type, an age type, a pitch type, an amplitude type, and an environment type.

4. The method as claimed in claim 2, wherein the plurality of text variations is generated by varying one or more parameters associated with each of plurality of text variations, wherein the one or more parameters comprise a script, a grammar usage, and a speech style.

5. The method as claimed in claim 2, wherein generating the plurality of speech-based test commands corresponding to each of the plurality of text variations comprises:

inputting a voice sample to the trained ML model; and

obtaining, from the trained ML model, the plurality of speech-based test commands corresponding to each of the plurality of text variations, based on the voice sample.

6. The method as claimed in claim 1, further comprising:

generating a report corresponding to the accuracy of the output command generated by the device-under-test corresponding to the speech-based test command, based on the comparison.

7. The method as claimed in claim 6, wherein generating the report further comprises:

receiving, via a user-interface, a selection of at least one speech modulation type set from the plurality of speech modulation type sets;

inputting, to the device-under-test, one or more speech-based test commands associated with the at least one speech modulation type set; and

receive one or more output commands from the device-under-test corresponding to the one or more speech-based test commands; and

comparing the one or more output commands corresponding to the one or more speech-based test commands with the text-based test command;

determining the accuracy of the one or more output commands corresponding to the one or more speech-based test commands with respect to the text-based test command; and

generating the report based on the accuracy of the one or more output commands corresponding to the one or more speech-based test commands with respect to the text-based test command.

8. The method as claimed in claim 1, wherein the device-under-test is one of an infotainment system and a smart device.

9. The method as claimed in claim 1 further comprising:

analyzing the report by comparing the output command generated by the device-under-test with at least one of historical data and current data.

10. A system for testing speech detection, the system comprising:

a device-under-test; and

a testing device comprising:

a processor; and

a memory communicatively coupled to the processor, wherein the memory stores a plurality of processor-executable instructions, which, upon execution, cause the processor to:

generate a speech-based test command using a text-based test command;

input the speech-based test command to the device-under-test, wherein the speech-based test command is to be processed by the device-under-test to generate an output command;

receive the output command from the device-under-test corresponding to the speech-based test command, wherein the output command is text-based;

compare the text-based test command inputted to the device-under-test and the output command generated by the device-under-test; and

determine an accuracy of the output command generated by the device-under-test corresponding to the speech-based test command, based on the comparison.

11. The system as claimed in claim 10, wherein generating the speech-based test command comprises:

generating a plurality of text variations corresponding to the text-based test command;

generating a plurality of speech-based test commands corresponding to each of the plurality of text variations, using a trained ML model;

wherein each speech-based test command of the plurality of speech-based test commands is associated with a unique speech modulation type set of a plurality of speech modulation type sets; and

selecting the speech-based test command from the plurality of speech-based test commands for testing the device-under-test.

12. The system as claimed in claim 11, wherein generating the plurality of speech-based test commands corresponding to each of the plurality of text variations comprises:

inputting a voice sample to the trained ML model; and

obtaining, from the trained ML model, the plurality of speech-based test commands corresponding to each of the plurality of text variations, based on the voice sample.

13. The system as claimed in claim 11, wherein the plurality of processor-executable instructions further cause the processor to:

generate a report corresponding to the accuracy of the output command generated by the device-under-test corresponding to the speech-based test command, based on the comparison, wherein generating the report further comprises:

receiving, via a user-interface, a selection of at least one speech modulation type set from the plurality of speech modulation type sets;

inputting, to the device-under-test, one or more speech-based test commands associated with the at least one speech modulation type set; and

receive one or more output commands from the device-under-test corresponding to the one or more speech-based test commands; and

comparing the one or more output commands corresponding to the one or more speech-based test commands with the text-based test command;

determining the accuracy of the one or more output commands corresponding to the one or more speech-based test commands with respect to the text-based test command; and

generating the report based on the accuracy of the one or more output commands corresponding to the one or more speech-based test commands with respect to the text-based test command.

14. The system as claimed in claim 10, wherein the plurality of processor-executable instructions further cause the processor to:

analyze the report by comparing the output command generated by the device-under-test with at least one of historical data and current data.

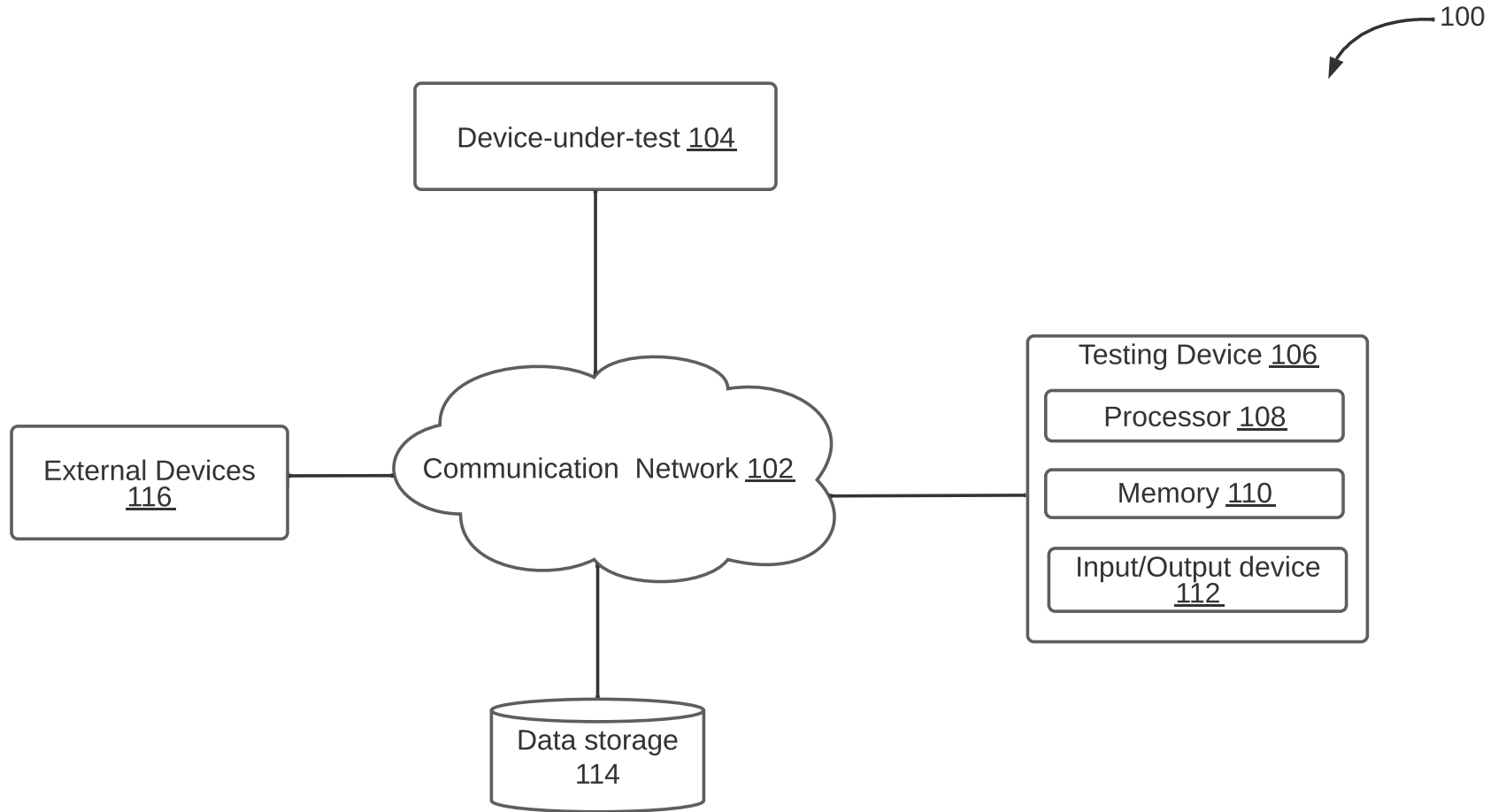
**Dated this 21<sup>st</sup> Day of March 2022**

**Bhanu Prasad (INPA No: 3253)  
Manager, IPR Dept.  
L&T Technology Services Ltd.**

## **A METHOD AND SYSTEM FOR TESTING SPEECH DETECTION**

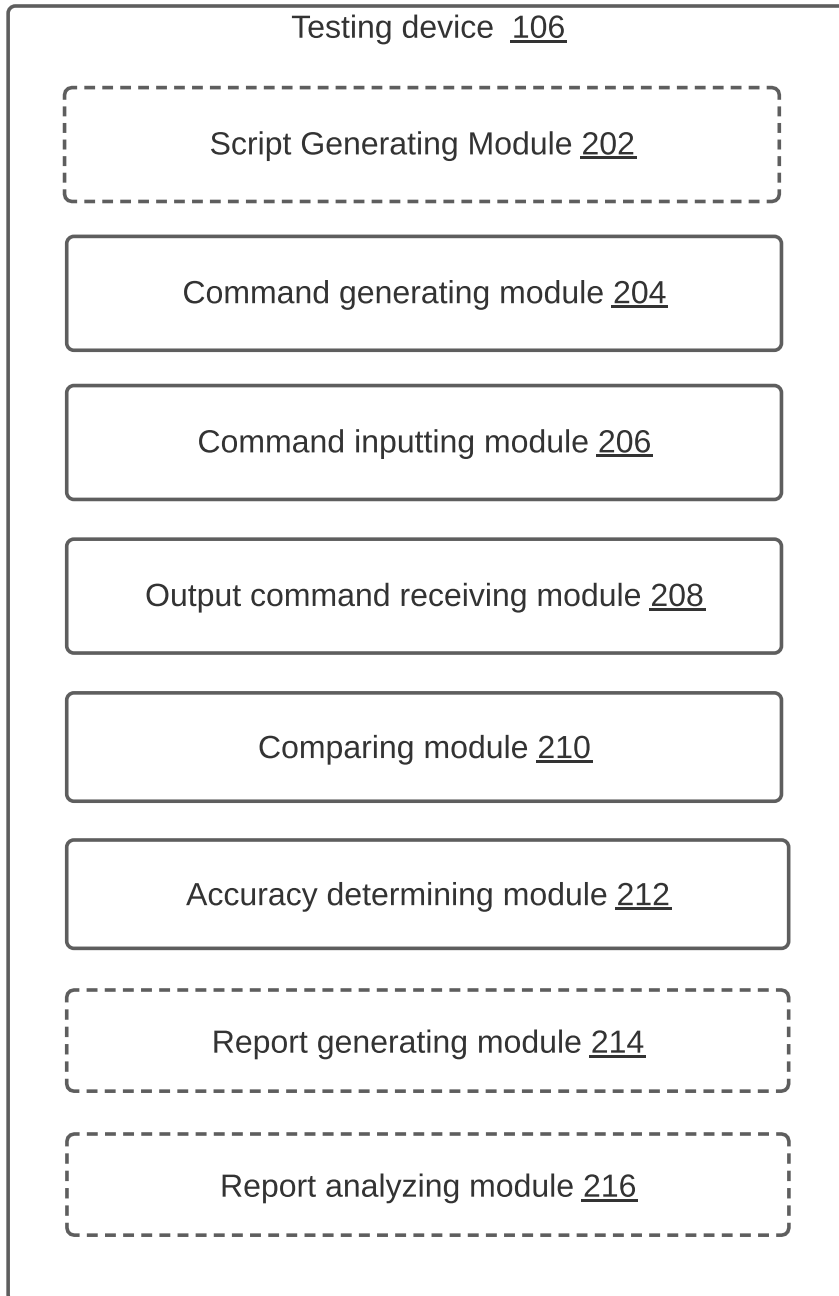
### **ABSTRACT**

A method and a system for testing speech detection is disclosed. The method may include generating a speech-based test command using a text-based test command, and inputting the speech-based test command to a device-under-test. The speech-based test command may be processed by the device-under-test to generate an output command. The method may further include receiving the output command from the device-under-test corresponding to the speech-based test command, and comparing the text-based test command inputted to the device-under-test and the output command generated by the device-under-test and determine an accuracy of the output command generated by the device-under-test corresponding to the speech-based test command based on the comparison. The method may further include generating a report corresponding to the accuracy of the output command generated by the device-under-test corresponding to the speech-based test command based on the comparison.



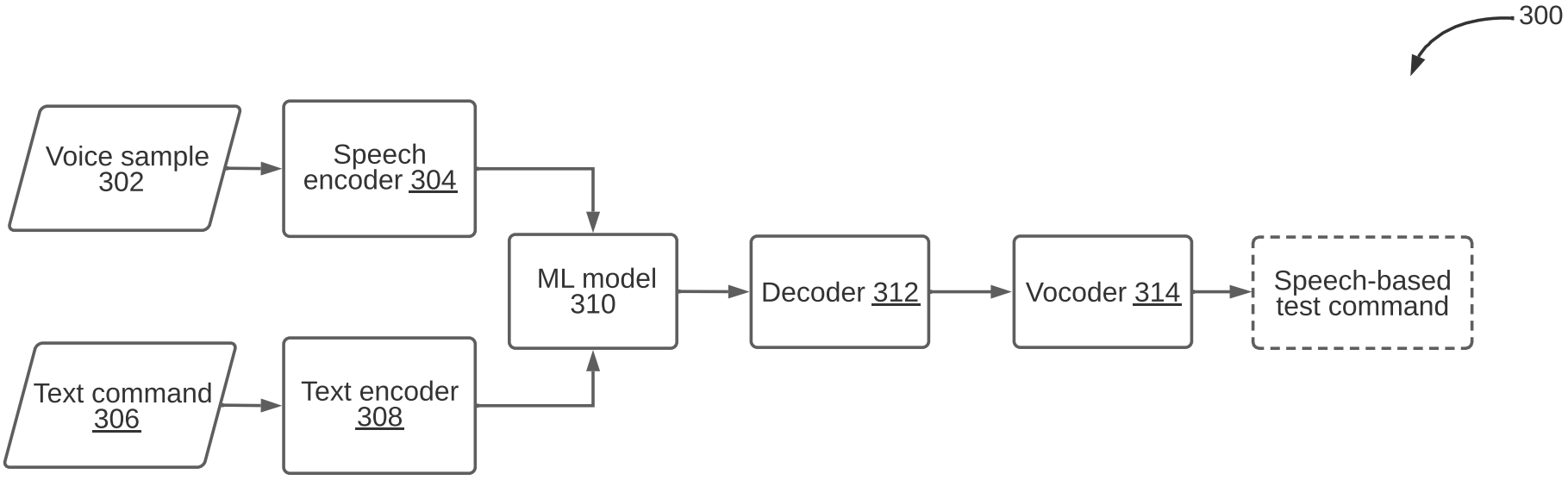
**FIG. 1**

**BHANU PRASAD (INPA No.: 3253)**  
Manager, IPR Dept.,  
L&T Technology Services Limited, DLF 3rd Block,  
2nd Floor,  
Manapakkam, TN, Chennai - 600 089.



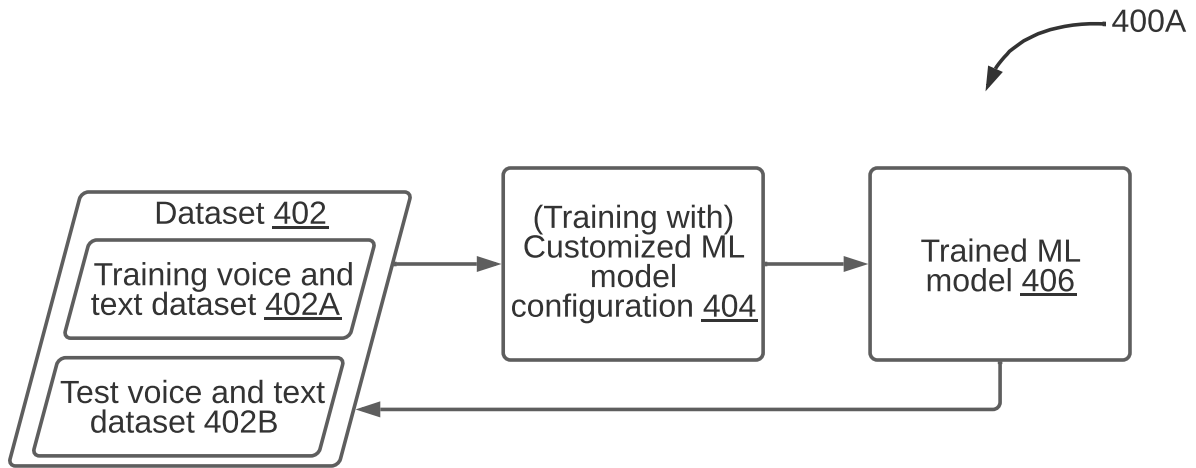
**FIG.2**

**BHANU PRASAD (INPA No.: 3253)  
Manager, IPR Dept.,  
L&T Technology Services Limited,  
DLF 3rd Block, 2nd Floor,  
Manapakkam, TN, Chennai - 600 089.**

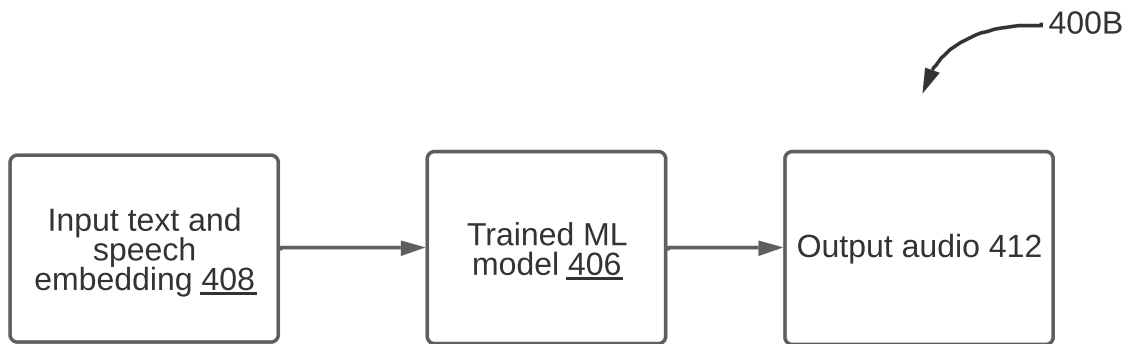


**FIG. 3**

**BHANU PRASAD (INPA No.: 3253)**  
Manager, IPR Dept.,  
L&T Technology Services Limited, DLF 3rd Block,  
2nd Floor,  
Manapakkam, TN, Chennai - 600 089.



**FIG. 4A**



**FIG. 4B**

**BHANU PRASAD (INPA No: 3253)**  
**Manager, IPR Dept.,**  
**L&T Technology Services Limited,**  
**DLF 3rd Block, 2nd Floor,**  
**Manapakkam, TN, Chennai - 600 089.**

500

---

**Configuration:**

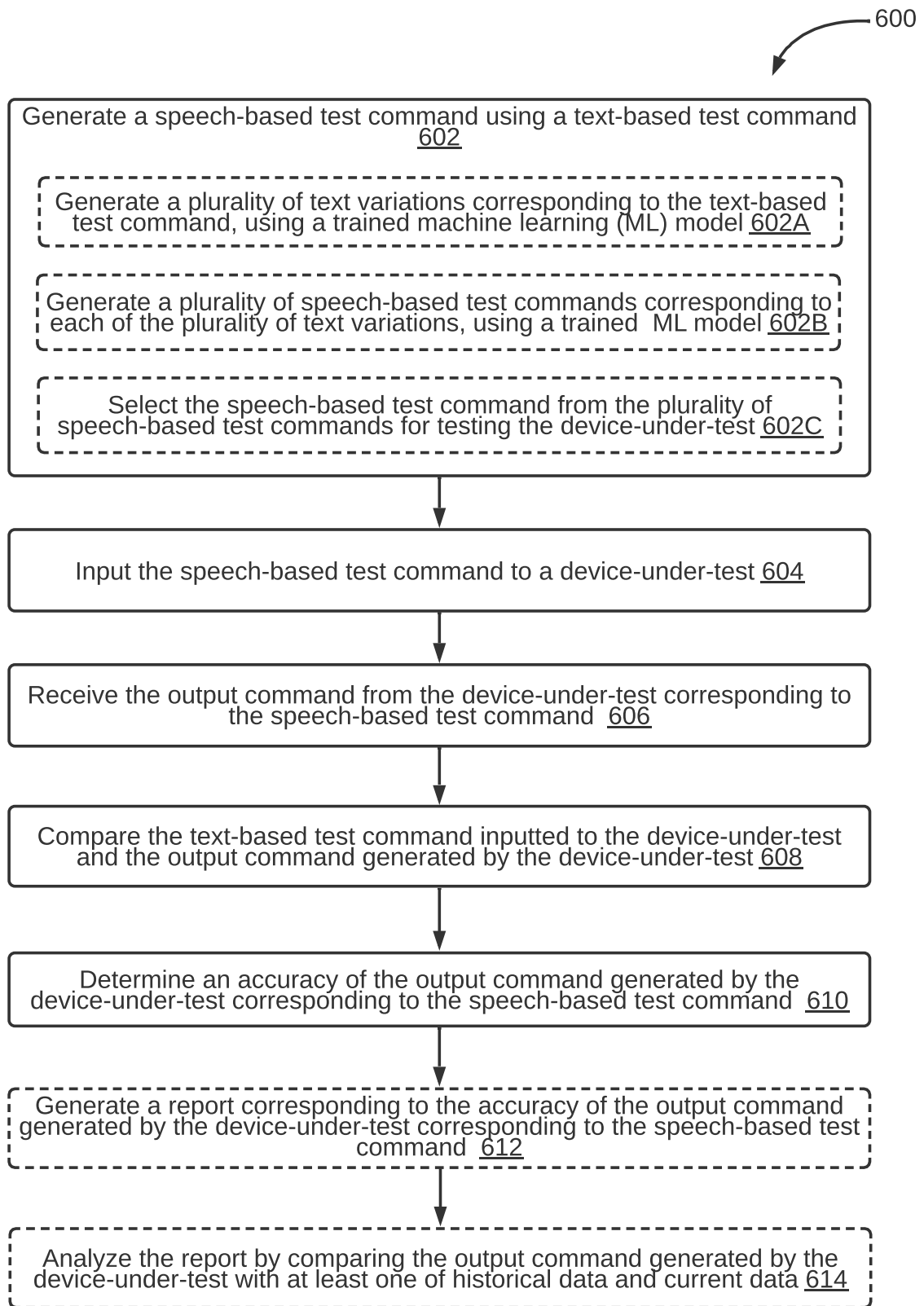
---

New Build:	<input type="text" value="21w50.5-1"/>
Previous Execution Build:	<input type="text" value="21w50.1-1"/>
Execution Category:	<input type="radio"/> Grammar_File <input type="radio"/> Master_Intent_File <input type="radio"/> LastBuildFailureTestcases
Execution Language:	<input type="text" value="English UK"/>
Derivate Type:	<input type="text" value="G05"/>
Rack Name:	<input type="text" value="Mumbai"/>
Recording Category:	<input checked="" type="radio"/> Male <input type="radio"/> Female <input type="radio"/> Both
	<input type="radio"/> Old <input type="radio"/> Child <input type="radio"/> Young <input checked="" type="radio"/> All
	<input checked="" type="radio"/> Soft <input type="radio"/> Loud <input type="radio"/> Both
	<input type="radio"/> Noisy <input checked="" type="radio"/> Silent <input type="radio"/> Both
Analysis Category:	<input checked="" type="radio"/> Standard <input type="radio"/> Random
<input type="button" value="Submit"/>	

502E      502F

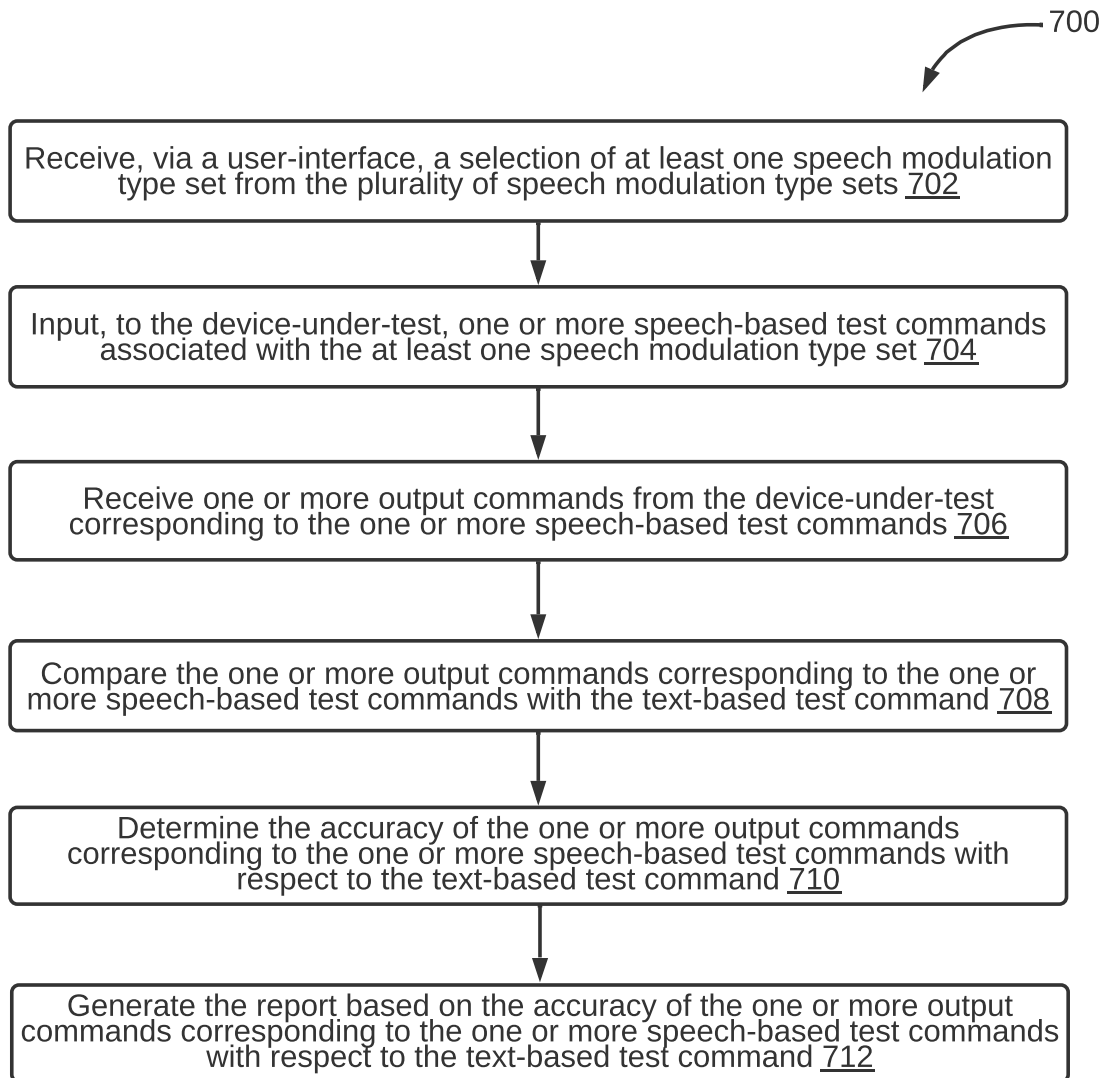
**FIG. 5**

**BHANU PRASAD (INPA No.: 3253)  
Manager, IPR Dept.,  
L&T Technology Services Limited, DLF  
3rd Block, 2nd Floor,  
Manapakkam, TN, Chennai - 600 089.**



**FIG. 6**

**BHANU PRASAD (INPA No.: 3253)**  
**Manager, IPR Dept.,**  
**L&T Technology Services Limited, DLF**  
**3rd Block, 2nd Floor,**  
**Manapakkam, TN, Chennai - 600 089.**



**FIG. 7**

**BHANU PRASAD (INPA No.: 3253)  
Manager, IPR Dept.,  
L&T Technology Services Limited,  
DLF 3rd Block, 2nd Floor,  
Manapakkam, TN, Chennai - 600 089.**