



US 20180315414A1

(19) **United States**

(12) **Patent Application Publication**
Hardee et al.

(10) **Pub. No.: US 2018/0315414 A1**

(43) **Pub. Date: Nov. 1, 2018**

(54) **ADAPTIVE DIGITAL ASSISTANT AND SPOKEN GENOME**

G10L 15/22 (2006.01)

G10L 13/047 (2006.01)

G06F 17/30 (2006.01)

(71) Applicant: **INTERNATIONAL BUSINESS MACHINES CORPORATION, ARMONK, NY (US)**

(52) **U.S. Cl.**

CPC *G10L 13/0335* (2013.01); *G10L 25/63* (2013.01); *G10L 15/22* (2013.01); *G10L 2015/223* (2013.01); *G06F 17/3002* (2013.01); *G06F 17/30029* (2013.01); *G06F 17/30026* (2013.01); *G10L 13/047* (2013.01)

(72) Inventors: **Christopher J. Hardee**, Raleigh, NC (US); **Steve Joroff**, Rivervale, NJ (US); **Pamela A. Nesbitt**, Ridgefield, CT (US); **Scott E. Schneider**, Rolesville, NC (US)

(57) **ABSTRACT**

Embodiments of the invention include a context sensitive adaptive digital assistant for personalized interaction. Embodiments of the invention also include a spoken genome for characterization and analysis of human voice. Aspects of the invention include selecting a starter vocabulary, receiving voice communications from a user, and modifying the starter vocabulary to generate a personalized lexicon. Aspects of the invention also include analyzing and categorizing human voice according to a plurality of characteristics, and creating a spoken genome database.

(21) Appl. No.: **15/497,536**

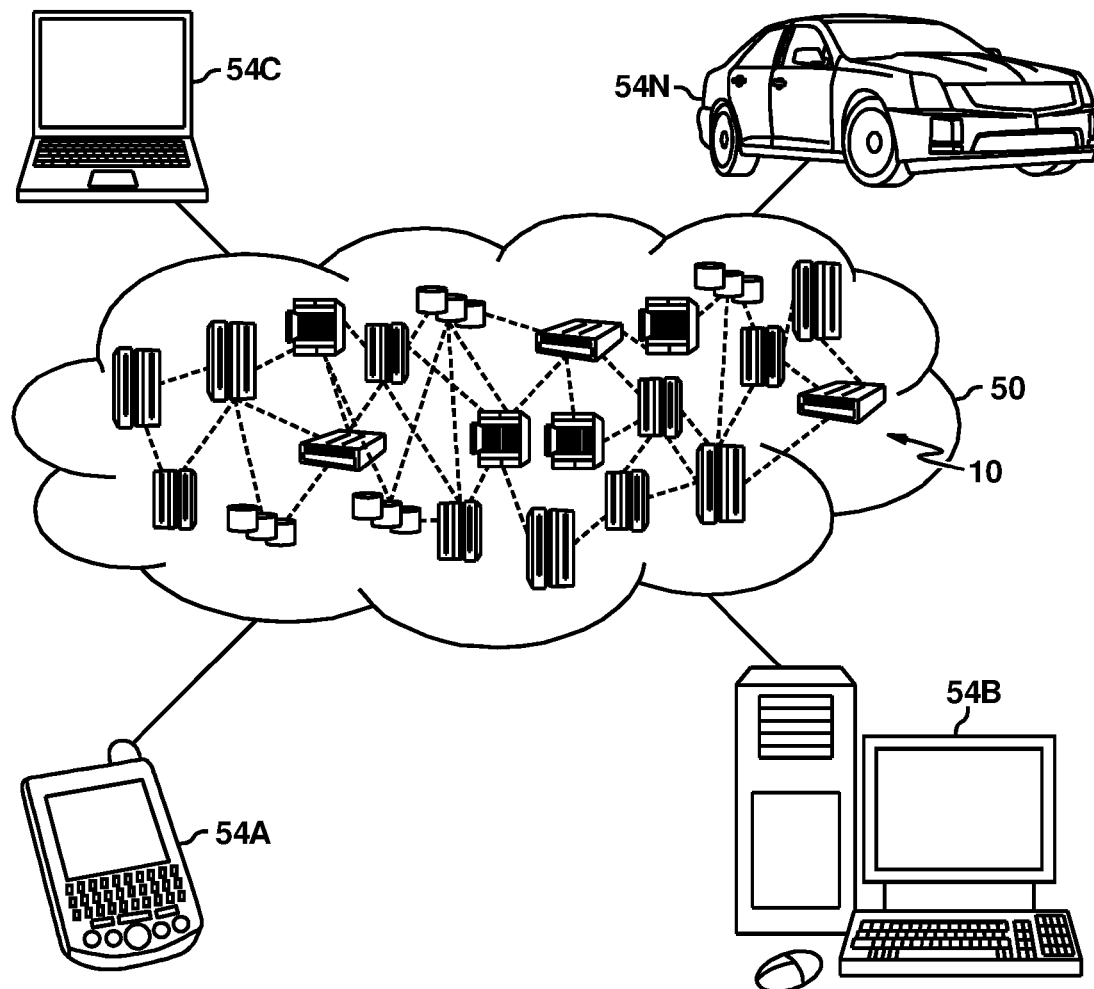
(22) Filed: **Apr. 26, 2017**

Publication Classification

(51) **Int. Cl.**

G10L 13/033 (2006.01)

G10L 25/63 (2006.01)



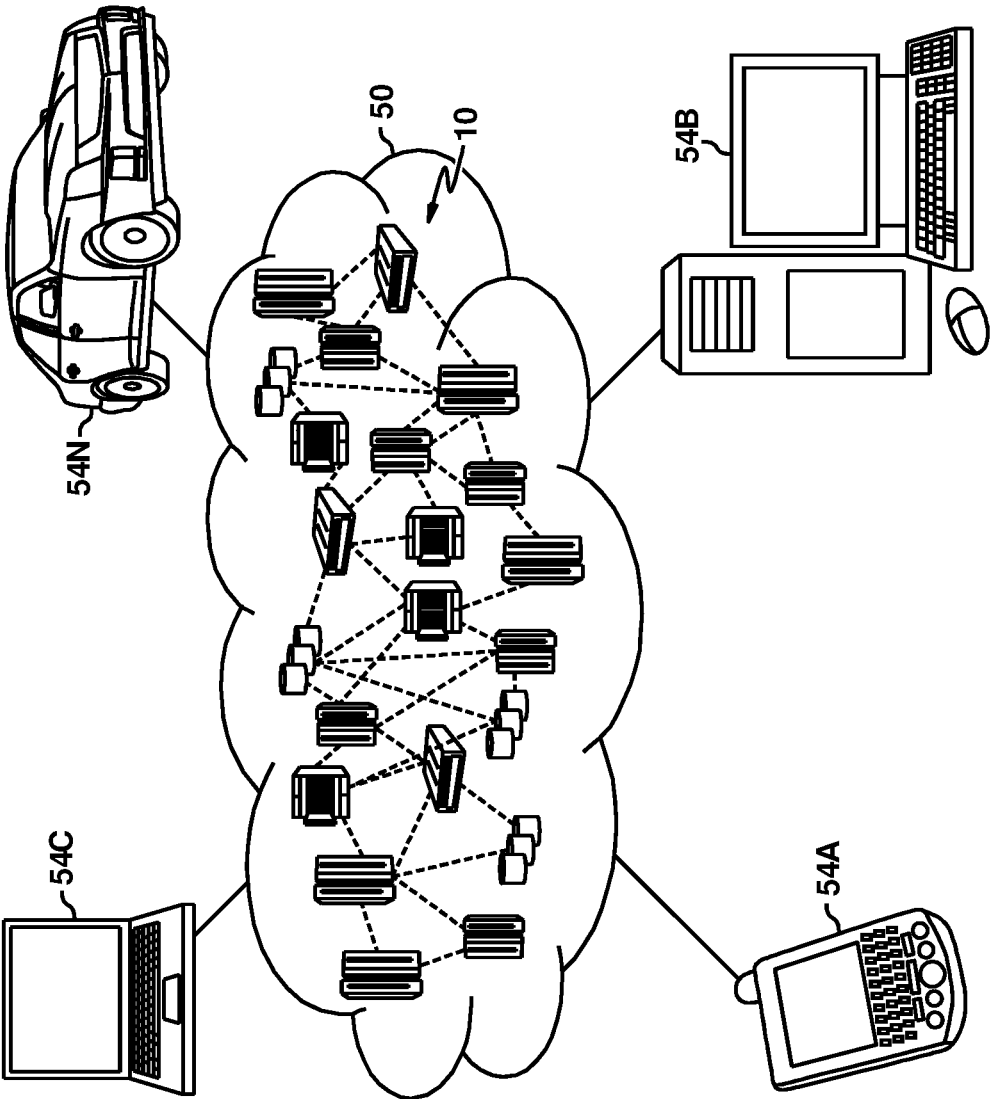


FIG. 1

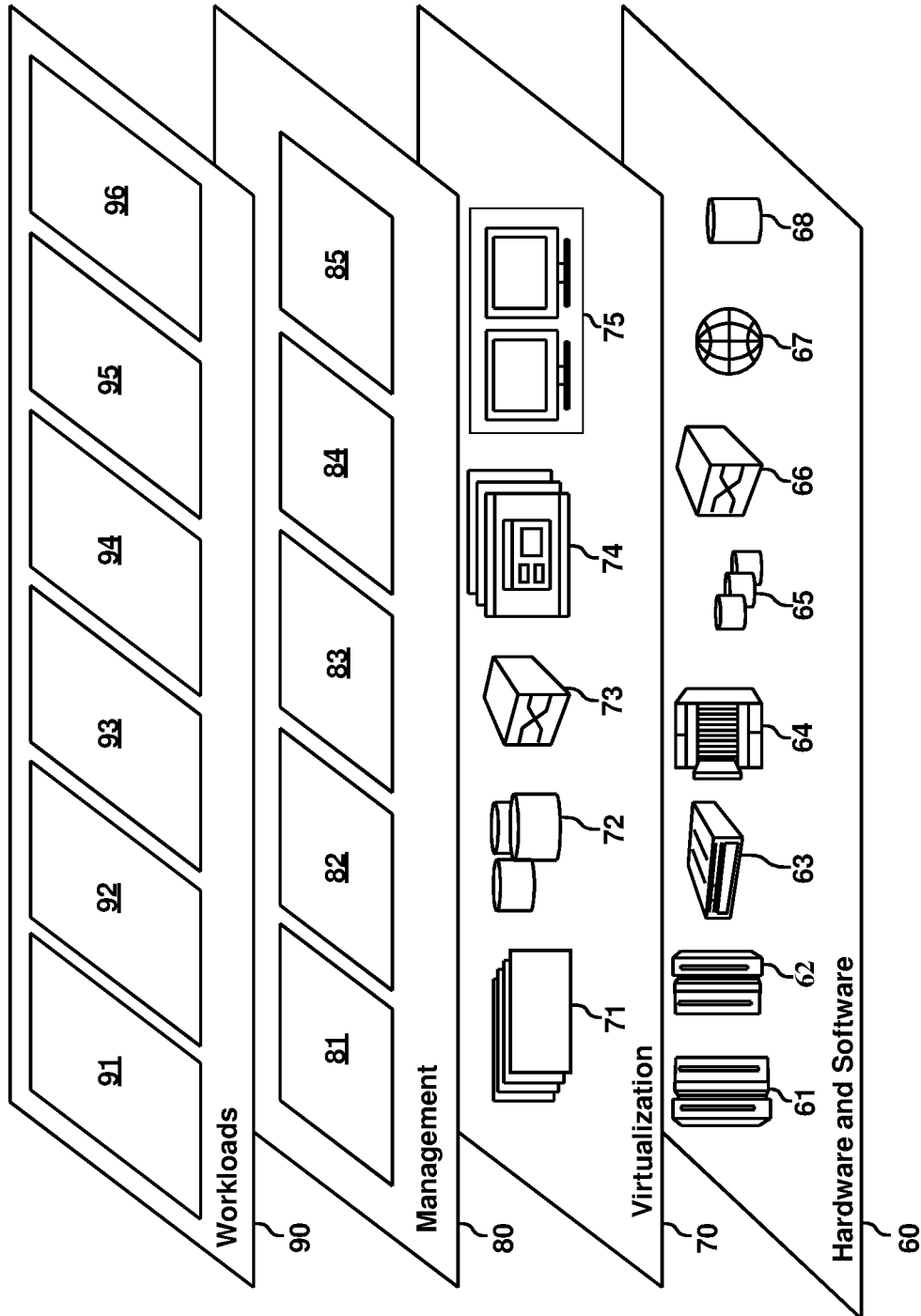


FIG. 2

100

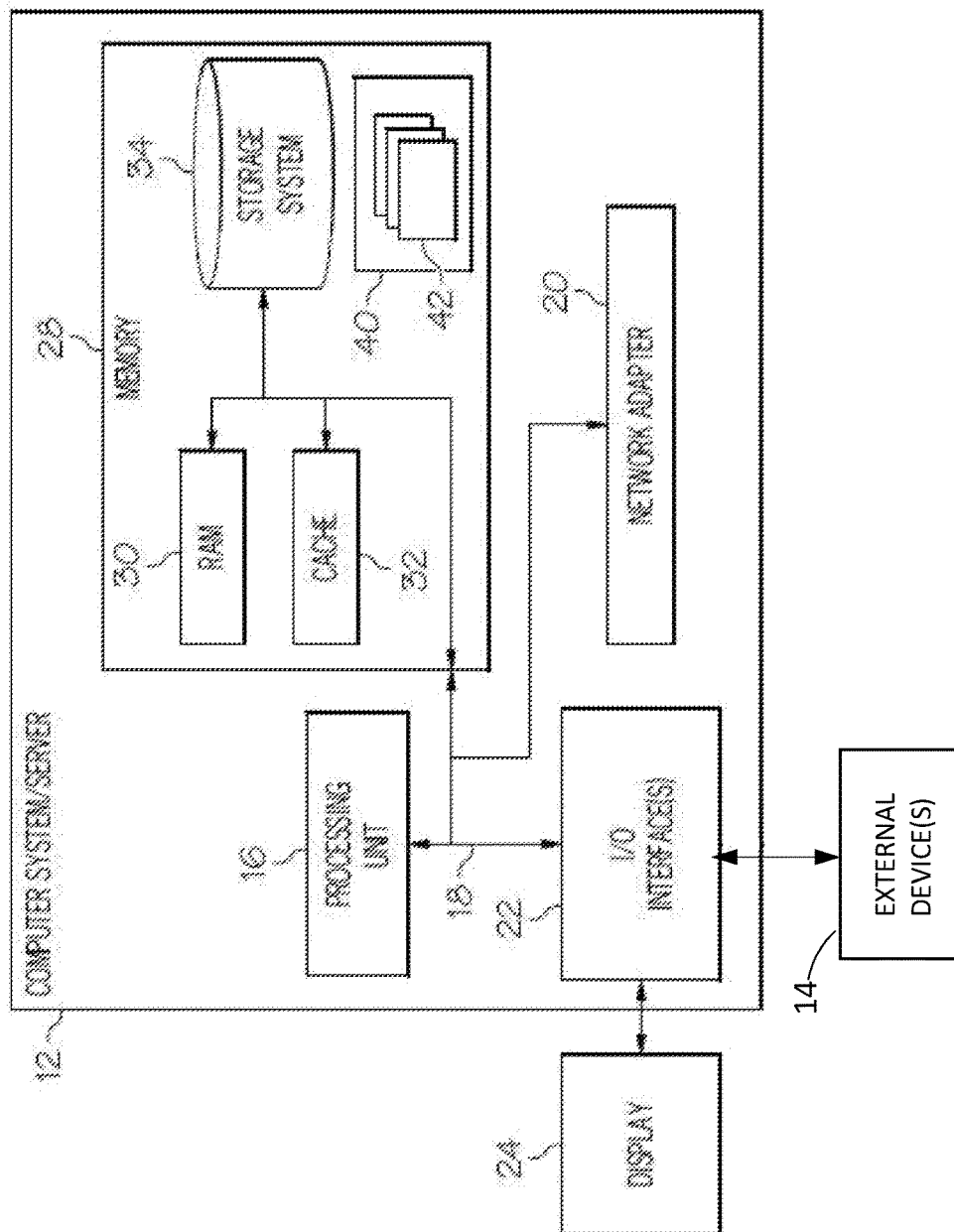


FIG. 3

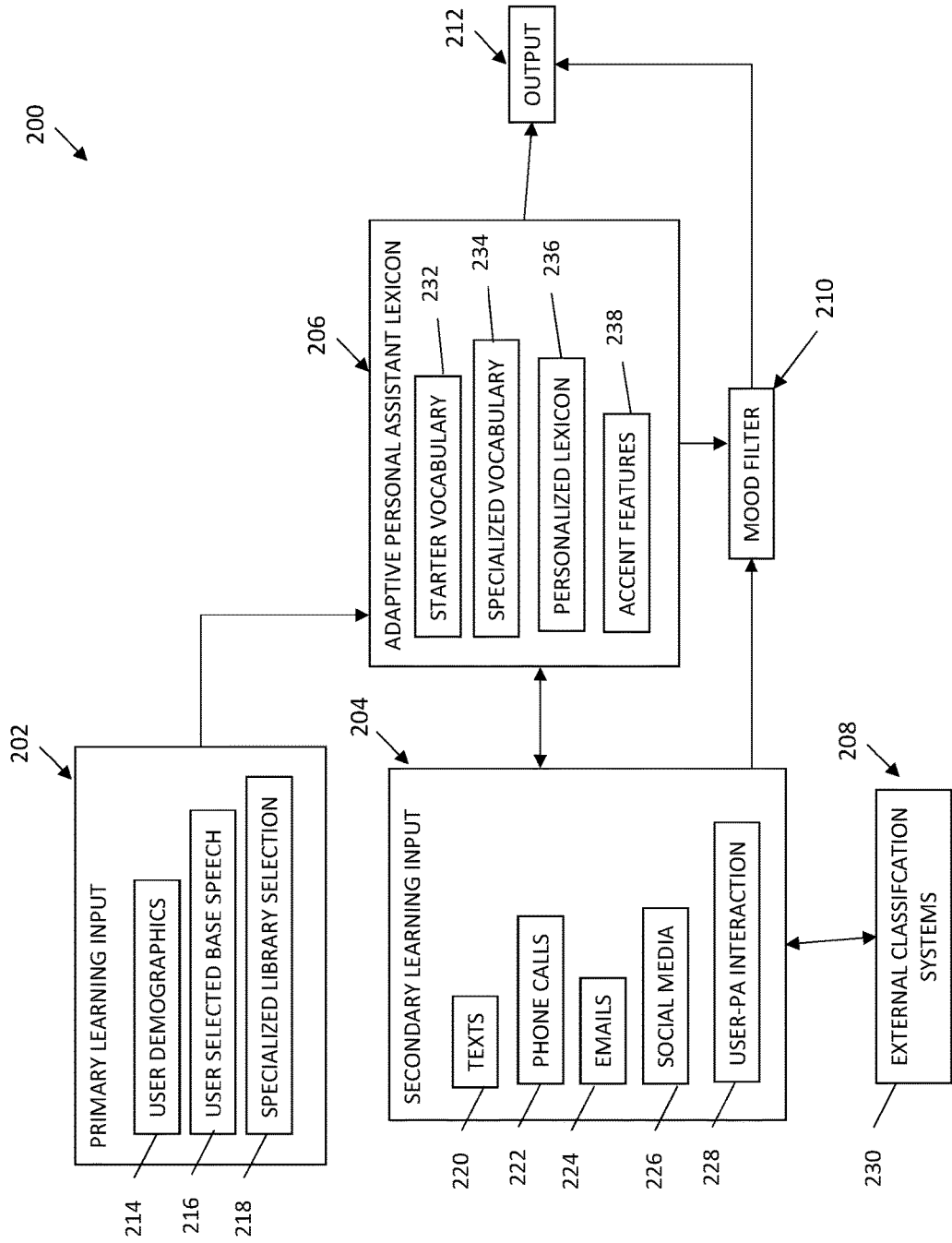


FIG. 4



FIG. 5

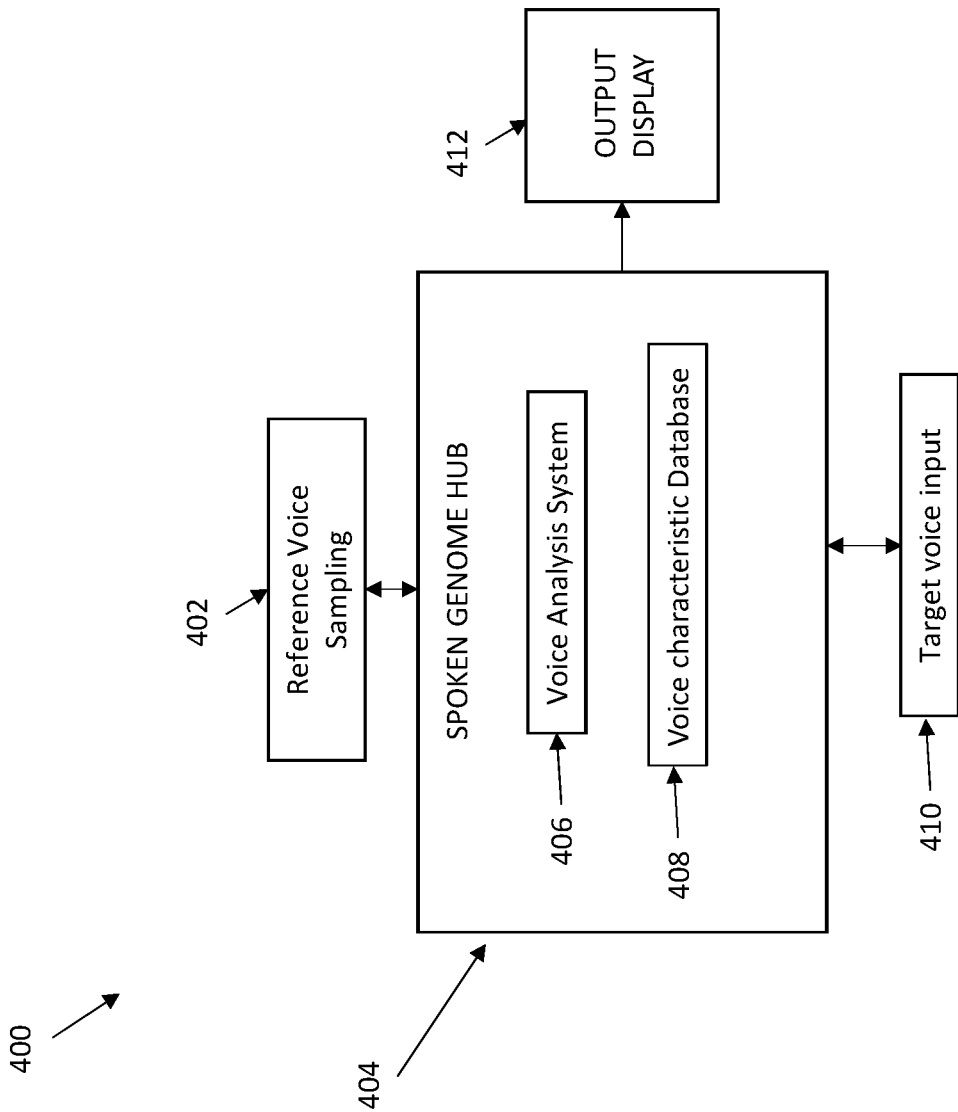


FIG. 6

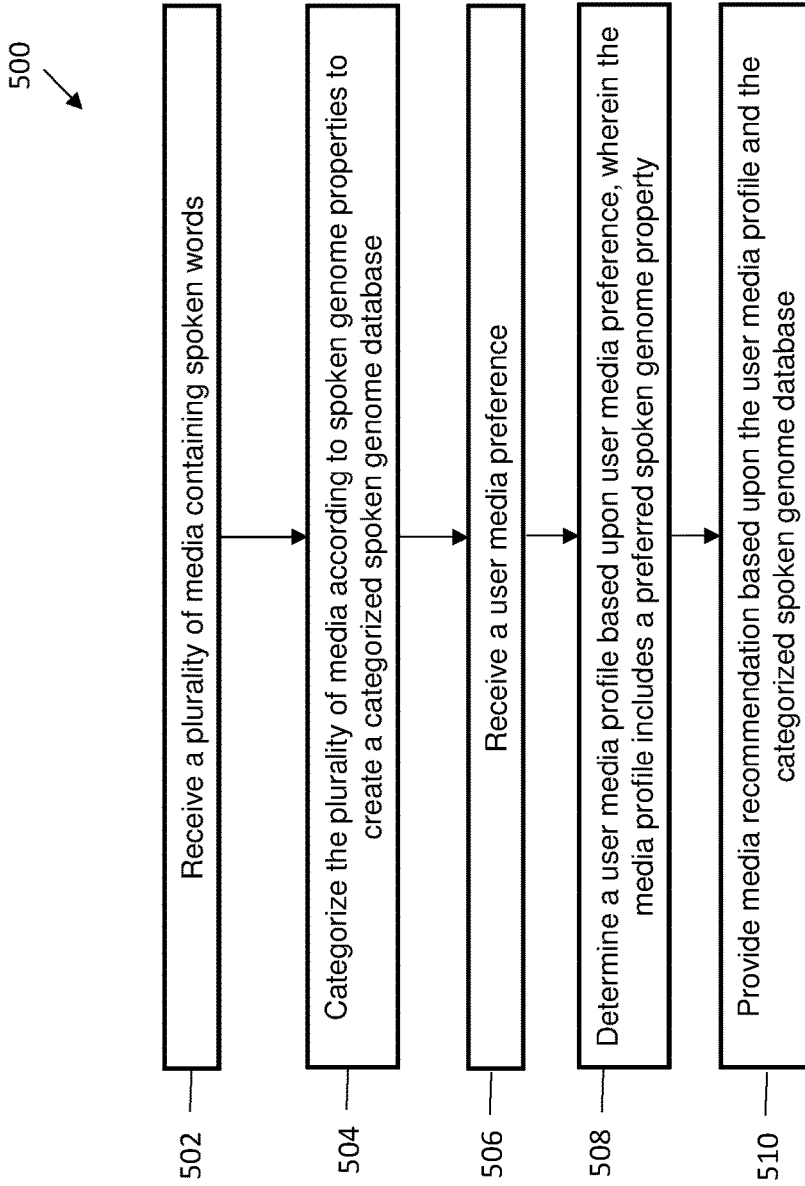


FIG. 7

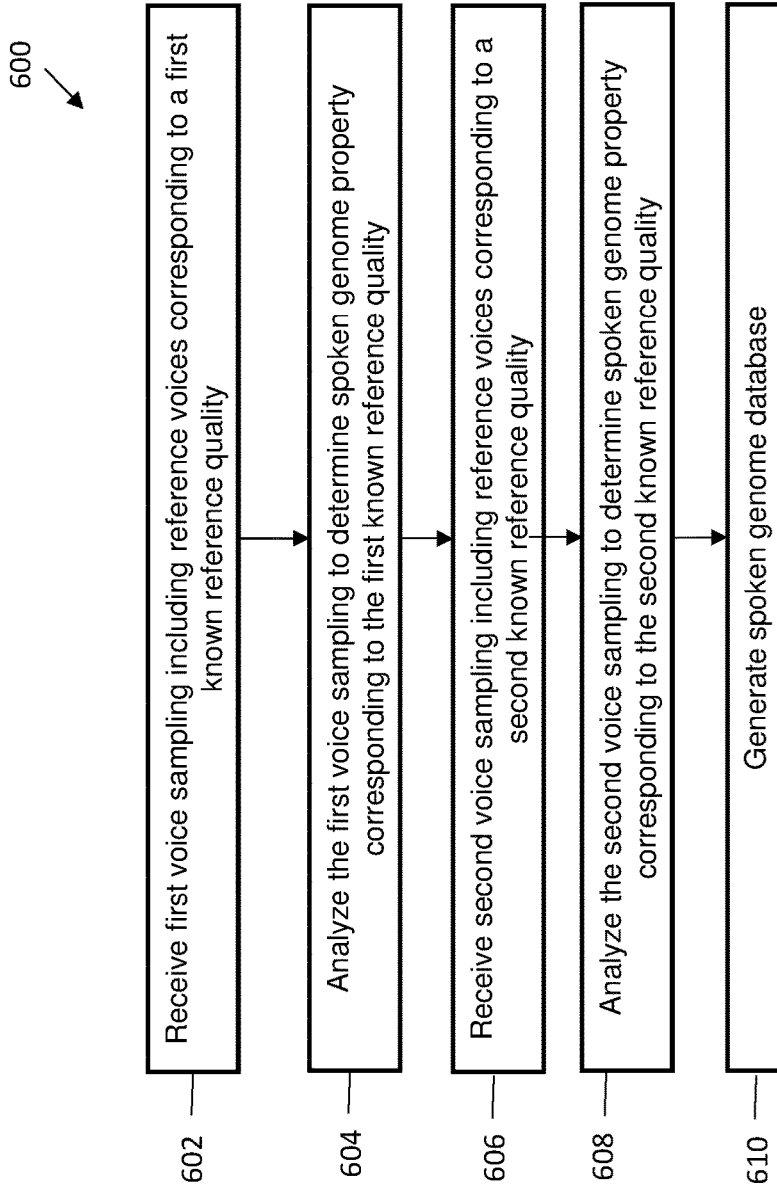


FIG. 8

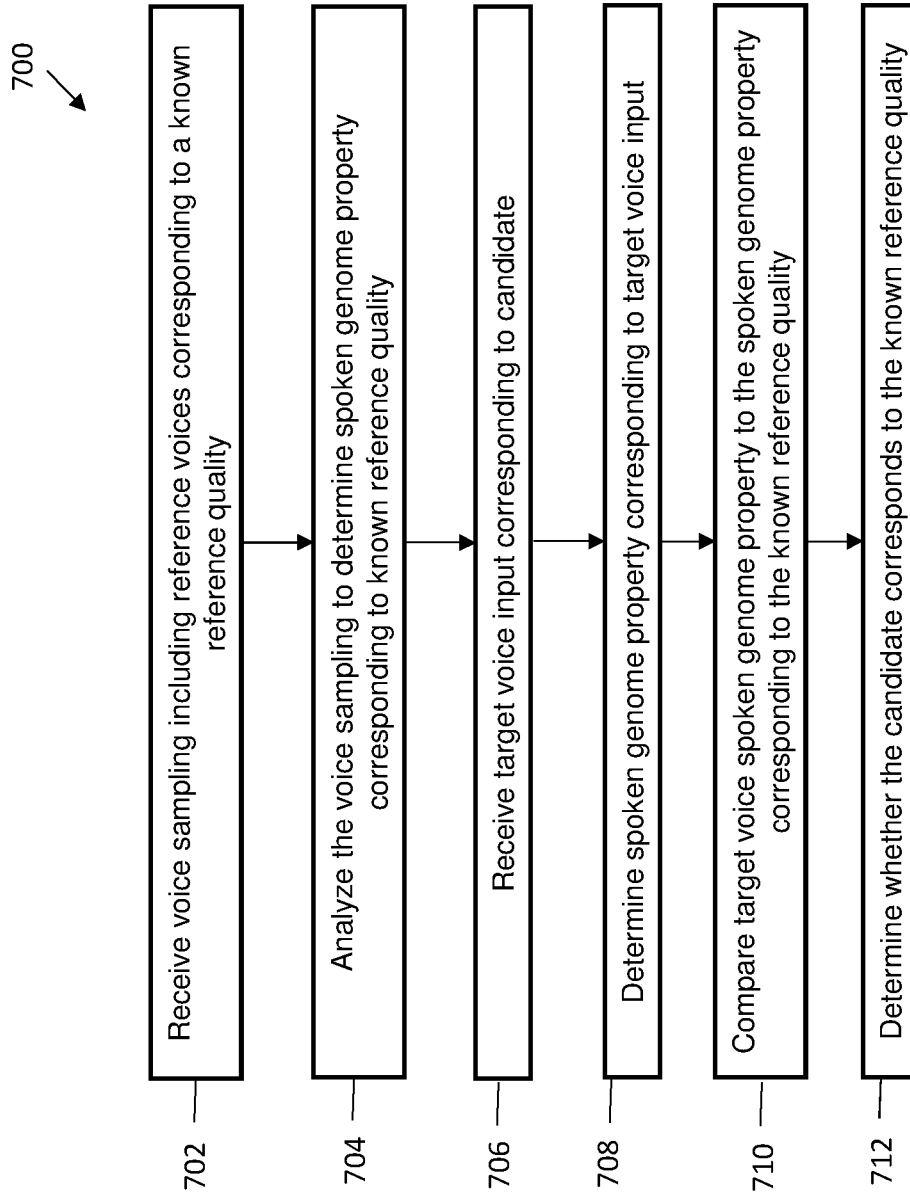


FIG. 9

ADAPTIVE DIGITAL ASSISTANT AND SPOKEN GENOME

BACKGROUND

[0001] The present invention relates generally to a digital assistant and spoken genome, and more specifically to a context sensitive adaptive digital assistant for personalized interaction and spoken genome for characterization and analysis of human voice.

[0002] Personal digital assistants, including digital assistants associated with smart devices, are increasingly becoming integrated into the daily lives of the general population. Such personal digital assistants can readily facilitate looking up information, scheduling appointments, setting task lists, and the like through vocal requests by the user. Conventional personal digital assistants, although increasingly popular with smart device users, lack personalization. In addition, although people can enjoy music and the sound their own voice, they may yet resort to silent texting and internet posting for communication. Qualification of sound and voice characteristics could aid not only the consumer, but also the provider of consumer content.

SUMMARY

[0003] In accordance with embodiments of the invention, a computer-implemented method for personalized digital interaction includes selecting a starter vocabulary from a starter vocabulary set. The method also includes receiving a plurality of user voice communications from a user. The method also includes generating a frequent word list based at least in part on the plurality of user voice communications. The method also includes modifying the starter vocabulary with a plurality of words from the frequent word list to generate a personalized lexicon. The method also includes generating a personalized verbal output based at least in part on the personalized lexicon.

[0004] In accordance with embodiments of the invention, a computer program product for characterization and analysis of human voice includes a computer readable storage medium readable by a processing circuit and storing program instructions for execution by the processing circuit for performing a method. The method includes receiving a plurality of media files, wherein the plurality of media files include spoken words. The method also includes categorizing the plurality of media files according to spoken genome properties to create a categorized spoken genome database. The method also includes receiving a user media preference. The method also includes determining a user media profile based at least in part on the user media preference, wherein the media profile includes a spoken genome property. The method also includes generating a media recommendation based at least in part on the user media profile and the categorized spoken genome database.

[0005] In accordance with embodiments of the invention, a processing system for characterization and analysis of human voice includes a processor in communication with one or more types of memory. The processor is configured to receive a first reference voice sampling, wherein the first reference voice sampling includes a plurality of reference voices corresponding to a first known reference quality. The processor is also configured to analyze the first voice sampling to determine a spoken genome property corresponding to the first known reference quality.

BRIEF DESCRIPTION OF THE DRAWINGS

[0006] The subject matter of embodiments of the invention is particularly pointed out and distinctly claimed in the claims at the conclusion of the specification. The foregoing and other features and advantages of the one or more embodiments described herein are apparent from the following detailed description taken in conjunction with the accompanying drawings in which:

[0007] FIG. 1 depicts a cloud computing environment according to an embodiment of the present invention.

[0008] FIG. 2 depicts abstraction model layers according to an embodiment of the present invention.

[0009] FIG. 3 depicts a computing node according to one or more embodiments of the present invention.

[0010] FIG. 4 depicts a diagram illustrating an exemplary adaptive digital assistant system according to one or more embodiments of the present invention.

[0011] FIG. 5 depicts a flow diagram illustrating a method for personalized digital interaction according to one or more embodiments of the present invention.

[0012] FIG. 6 depicts a diagram illustrating an exemplary voice characterization system according to one or more embodiments of the present invention.

[0013] FIG. 7 depicts a flow diagram illustrating a method for characterization and analysis of human voice according to one or more embodiments of the present invention.

[0014] FIG. 8 depicts a flow diagram illustrating a method for characterization and analysis of human voice according to one or more embodiments of the present invention.

[0015] FIG. 9 depicts a flow diagram illustrating a method for characterization and analysis of human voice according to one or more embodiments of the present invention.

DETAILED DESCRIPTION

[0016] Embodiments of the invention relate to systems and methods for characterization, analysis, and personalization of voice.

[0017] A large appeal of personal digital assistants is that they desirably reduce or even eliminate the need to type information into search engines and scheduling programs, and in some instances, they can provide shortcuts to commands associated with a smart device. Conventional personal digital assistants, although increasingly popular with smart device users, lack personalization. Traditional personal digital assistants have the same characteristics for each user, including, for example, speaking with the same lexicon and same degree of politeness for every person serviced by the personal digital assistant. This lack of personalization, however, is undesirable for a user of such assistants. For example, a user who interacts multiple times daily with a personal digital assistants could prefer characteristics that are less automated to increase the quality of and frequency of interaction and to achieve a communication that feels more like a human-human interaction than a human-machine interaction.

[0018] Embodiments of the invention include a personalized digital assistant that can adapt its lexicon and vocal mannerisms for an individual user. In some embodiments of the invention, personal digital assistants are personalized to an individual based at least in part on one or more characteristics and can further adapt with continued use.

[0019] With increased emphasis on texting for communication and sharing thoughts and feelings through posting of

text and pictures on social media, the emphasis and focus on sound and the human voice has decreased. Although people can enjoy music and the sound their own voice, they may yet resort to silent texting and internet posting for communication. However, wording expressed by people vocally, including aspects such as cadence, tonality, dynamics and wording, can have a potentially large impact on human emotions and actions. For example, characteristics of sound and voice can impact an individual's decision of whether to watch a commercial or television program. Although actions and emotions can be guided by how much people like what they hear, individuals can lack an ability to pre-determine whether what individuals hear will be appealing or influential.

[0020] Various aspects of sound characteristics, such as cadence, tonality, dynamics, and wording choice, can have a potentially dramatic impact on user emotion and activity. Voice and sound characteristics can affect how listeners perceive, appreciate, and pay attention to a speaker. For example, listening to a voice with a low volume and slow cadence can be more likely to induce sleepiness and feelings of relaxation than listening to a voice with a high volume and fast cadence. Moreover, individual listeners or categories of listeners can have their own sound preferences, which can be known or unknown to a given individual. For example, a female voice at a given pitch could be more appealing or influential to a 30-year old male than a 60-year old female and, thus, could be a better choice for a truck advertisement. Similarly, speaking cadence can be an important characteristic for listening ability but also as an indicator of content of a viewable program. Voice and sound characteristics, if known and understood, can be used for targeted advertising, tailored content perception, for example in educational contexts, and a variety of other purposes.

[0021] Embodiments of the invention include a spoken genome that categorizes voices based at least in part on pitch, cadence, tonality, rhythm, accent, timing, elocution, and related or similar characteristics that can affect how listeners perceive, appreciate, or pay attention to a speaker. Embodiments of the invention can guide listeners and content providers to content they will like. Embodiments of the invention can facilitate gaining the interest, of target audiences, for example, consumers, television viewers, or students. Thus, embodiments of the invention can provide benefits to advertisers, educators, and television program providers, who can desire to provide subtly targeted content. In some embodiments of the invention, a user can use a spoken genome to identify content that is likely to be of interest. For example, a fan of television programming that provides characteristic rapid-fire clever dialogue, can use the spoken genome to find other television programming with similar rapid-fire dialogue. In some embodiments of the invention, a spoken genome can qualify advertisements and television shows. In some embodiments of the invention, a spoken genome can guide marketers and producers as they set out to target certain audiences.

[0022] It is understood in advance that although this description includes a detailed description on cloud computing, implementation of the teachings recited herein are not limited to a cloud computing environment. Rather, embodiments of the present invention are capable of being implemented in conjunction with any other type of computing environment now known or later developed.

[0023] Cloud computing is a model of service delivery for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g. networks, network bandwidth, servers, processing, memory, storage, applications, virtual machines, and services) that can be rapidly provisioned and released with minimal management effort or interaction with a provider of the service. This cloud model can include at least five characteristics, at least three service models, and at least four deployment models.

[0024] Characteristics are as follows:

[0025] On-demand self-service: a cloud consumer can unilaterally provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with the service's provider.

[0026] Broad network access: capabilities are available over a network and accessed through standard mechanisms that promote use by heterogeneous thin or thick client platforms (e.g., mobile phones, laptops, and PDAs).

[0027] Resource pooling: the provider's computing resources are pooled to serve multiple consumers using a multi-tenant model, with different physical and virtual resources dynamically assigned and reassigned according to demand. There is a sense of location independence in that the consumer generally has no control or knowledge over the exact location of the provided resources but can be able to specify location at a higher level of abstraction (e.g., country, state, or datacenter).

[0028] Rapid elasticity: capabilities can be rapidly and elastically provisioned, in some cases automatically, to quickly scale out and rapidly released to quickly scale in. To the consumer, the capabilities available for provisioning often appear to be unlimited and can be purchased in any quantity at any time.

[0029] Measured service: cloud systems automatically control and optimize resource use by leveraging a metering capability at some level of abstraction appropriate to the type of service (e.g., storage, processing, bandwidth, and active user accounts). Resource usage can be monitored, controlled, and reported providing transparency for both the provider and consumer of the utilized service.

[0030] Service Models are as follows:

[0031] Software as a Service (SaaS): the capability provided to the consumer is to use the provider's applications running on a cloud infrastructure. The applications are accessible from various client devices through a thin client interface such as a web browser (e.g., web-based e-mail). The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, storage, or even individual application capabilities, with the possible exception of limited user-specific application configuration settings.

[0032] Platform as a Service (PaaS): the capability provided to the consumer is to deploy onto the cloud infrastructure consumer-created or acquired applications created using programming languages and tools supported by the provider. The consumer does not manage or control the underlying cloud infrastructure including networks, servers, operating systems, or storage, but has control over the deployed applications and possibly application hosting environment configurations.

[0033] Infrastructure as a Service (IaaS): the capability provided to the consumer is to provision processing, storage, networks, and other fundamental computing resources where the consumer is able to deploy and run arbitrary

software, which can include operating systems and applications. The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, deployed applications, and possibly limited control of select networking components (e.g., host firewalls).

[0034] Deployment Models are as follows:

[0035] Private cloud: the cloud infrastructure is operated solely for an organization. It can be managed by the organization or a third party and can exist on-premises or off-premises.

[0036] Community cloud: the cloud infrastructure is shared by several organizations and supports a specific community that has shared concerns (e.g., mission, security requirements, policy, and compliance considerations). It can be managed by the organizations or a third party and can exist on-premises or off-premises.

[0037] Public cloud: the cloud infrastructure is made available to the general public or a large industry group and is owned by an organization selling cloud services.

[0038] Hybrid cloud: the cloud infrastructure is a composition of two or more clouds (private, community, or public) that remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load-balancing between clouds).

[0039] A cloud computing environment is service oriented with a focus on statelessness, low coupling, modularity, and semantic interoperability. At the heart of cloud computing is an infrastructure including a network of interconnected nodes.

[0040] Referring now to FIG. 1, illustrative cloud computing environment 50 is depicted. As shown, cloud computing environment 50 includes one or more cloud computing nodes 10 with which local computing devices used by cloud consumers, such as, for example, personal digital assistant (PDA) or cellular telephone 54A, desktop computer 54B, laptop computer 54C, and/or automobile computer system 54N can communicate. Nodes 10 can communicate with one another. They can be grouped (not shown) physically or virtually, in one or more networks, such as Private, Community, Public, or Hybrid clouds as described hereinabove, or a combination thereof. This allows cloud computing environment 50 to offer infrastructure, platforms and/or software as services for which a cloud consumer does not need to maintain resources on a local computing device. It is understood that the types of computing devices 54A-N shown in FIG. 1 are intended to be illustrative only and that computing nodes 10 and cloud computing environment 50 can communicate with any type of computerized device over any type of network and/or network addressable connection (e.g., using a web browser).

[0041] Referring now to FIG. 2, a set of functional abstraction layers provided by cloud computing environment 50 (FIG. 1) is shown. It should be understood in advance that the components, layers, and functions shown in FIG. 2 are intended to be illustrative only and embodiments of the invention are not limited thereto. As depicted, the following layers and corresponding functions are provided:

[0042] Hardware and software layer 60 includes hardware and software components. Examples of hardware components include: mainframes 61; RISC (Reduced Instruction Set Computer) architecture based servers 62; servers 63; blade servers 64; storage devices 65; and networks and

networking components 66. In some embodiments, software components include network application server software 67 and database software 68.

[0043] Virtualization layer 70 provides an abstraction layer from which the following examples of virtual entities can be provided: virtual servers 71; virtual storage 72; virtual networks 73, including virtual private networks; virtual applications and operating systems 74; and virtual clients 75.

[0044] In one example, management layer 80 can provide the functions described below. Resource provisioning 81 provides dynamic procurement of computing resources and other resources that are utilized to perform tasks within the cloud computing environment. Metering and Pricing 82 provide cost tracking as resources are utilized within the cloud computing environment, and billing or invoicing for consumption of these resources. In one example, these resources can include application software licenses. Security provides identity verification for cloud consumers and tasks, as well as protection for data and other resources. User portal 83 provides access to the cloud computing environment for consumers and system administrators. Service level management 84 provides cloud computing resource allocation and management such that required service levels are met. Service Level Agreement (SLA) planning and fulfillment 85 provide pre-arrangement for, and procurement of, cloud computing resources for which a future requirement is anticipated in accordance with an SLA.

[0045] Workloads layer 90 provides examples of functionality for which the cloud computing environment can be utilized. Examples of workloads and functions which can be provided from this layer include: mapping and navigation 91; software development and lifecycle management 92; virtual classroom education delivery 93; data analytics processing 94; transaction processing 95; and voice categorization and personalization 96.

[0046] Referring now to FIG. 3, a schematic of a cloud computing node 100 included in a distributed cloud environment or cloud service network is shown according to a non-limiting embodiment. The cloud computing node 100 is only one example of a suitable cloud computing node and is not intended to suggest any limitation as to the scope of use or functionality of embodiments of the invention described herein. Regardless, cloud computing node 100 is capable of being implemented and/or performing any of the functionality set forth hereinabove.

[0047] In cloud computing node 100 there is a computer system/server 12, which is operational with numerous other general purpose or special purpose computing system environments or configurations. Examples of well-known computing systems, environments, and/or configurations that can be suitable for use with computer system/server 12 include, but are not limited to, personal computer systems, server computer systems, thin clients, thick clients, handheld or laptop devices, multiprocessor systems, microprocessor-based systems, set top boxes, programmable consumer electronics, network PCs, minicomputer systems, mainframe computer systems, and distributed cloud computing environments that include any of the above systems or devices, and the like.

[0048] Computer system/server 12 can be described in the general context of computer system-executable instructions, such as program modules, being executed by a computer system. Generally, program modules can include routines,

programs, objects, components, logic, data structures, and so on that perform particular tasks or implement particular abstract data types. Computer system/server 12 can be practiced in distributed cloud computing environments where tasks are performed by remote processing devices that are linked through a communications network. In a distributed cloud computing environment, program modules can be located in both local and remote computer system storage media including memory storage devices.

[0049] As shown in FIG. 3, computer system/server 12 in cloud computing node 100 is shown in the form of a general-purpose computing device. The components of computer system/server 12 can include, but are not limited to, one or more processors or processing units 16, a system memory 28, and a bus 18 that couples various system components including system memory 28 to processor 16.

[0050] Bus 18 represents one or more of any of several types of bus structures, including a memory bus or memory controller, a peripheral bus, an accelerated graphics port, and a processor or local bus using any of a variety of bus architectures. By way of example, and not limitation, such architectures include Industry Standard Architecture (ISA) bus, Micro Channel Architecture (MCA) bus, Enhanced ISA (EISA) bus, Video Electronics Standards Association (VESA) local bus, and Peripheral Component Interconnect (PCI) bus.

[0051] Computer system/server 12 typically includes a variety of computer system readable media. Such media can be any available media that is accessible by computer system/server 12, and it includes both volatile and non-volatile media, removable and non-removable media.

[0052] System memory 28 can include computer system readable media in the form of volatile memory, such as random access memory (RAM) 30 and/or cache memory 32. Computer system/server 12 can further include other removable/non-removable, volatile/non-volatile computer system storage media. By way of example only, storage system 34 can be provided for reading from and writing to a non-removable, non-volatile magnetic media (not shown and typically called a “hard drive”). Although not shown, a magnetic disk drive for reading from and writing to a removable, non-volatile magnetic disk (e.g., a “floppy disk”), and an optical disk drive for reading from or writing to a removable, non-volatile optical disk such as a CD-ROM, DVD-ROM or other optical media can be provided. In such instances, each can be connected to bus 18 by one or more data media interfaces. As will be further depicted and described below, memory 28 can include at least one program product having a set (e.g., at least one) of program modules that are configured to carry out the functions of embodiments of the invention.

[0053] Program/utility 40, having a set (at least one) of program modules 42, can be stored in memory 28 by way of example, and not limitation, as well as an operating system, one or more application programs, other program modules, and program data. Each of the operating system, one or more application programs, other program modules, and program data or some combination thereof, can include an implementation of a networking environment. Program modules 42 generally carry out the functions and/or methodologies of embodiments of the invention as described herein.

[0054] Computer system/server 12 can also communicate with one or more external devices 14 such as a keyboard, a pointing device, a display 24, etc., one or more devices that

enable a user to interact with computer system/server 12, and/or any devices (e.g., network card, modem, etc.) that enable computer system/server 12 to communicate with one or more other computing devices. Such communication can occur via Input/Output (I/O) interfaces 22. Still yet, computer system/server 12 can communicate with one or more networks such as a local area network (LAN), a general wide area network (WAN), and/or a public network (e.g., the Internet) via network adapter 20. As depicted, network adapter 20 communicates with the other components of computer system/server 12 via bus 18. It should be understood that although not shown, other hardware and/or software components could be used in conjunction with computer system/server 12. Examples, include, but are not limited to: microcode, device drivers, redundant processing units, external disk drive arrays, RAID systems, tape drives, and data archival storage systems, etc.

[0055] Turning now to a more detailed description of embodiments of the present invention, FIG. 4 depicts a diagram illustrating an exemplary adaptive digital assistant system 200. The system 200 includes a primary learning input 202. The primary learning input 202 includes components that can be used to establish a starter vocabulary 232. For example, the primary learning input 202 can include any general data that can be used to select from a defined subset of starter vocabularies. For example, the primary learning input can include user demographics 214, such as age, gender, geographic location, marital status, ethnicity, or educational level. In some embodiments of the invention, the primary learning input 202 can include a user selected base speech 216. For example, a user can select their own base speech vocabulary, upon system start up, from a list of predefined starter vocabularies, such as teenage girl, middle aged woman, man aged 60 to 70, and the like. The primary learning input 202 can include, in some embodiments of the invention, a specialized library selection 218. For example, a user can be prompted, upon system start up, to optionally select a specialized library corresponding to a specialized interest or occupation, such as a music specialization or an engineering, medical or other occupation. A specialized library selection 218, for instance, can aid in enhancement or supplementation of the base vocabulary, providing greater understanding of the user commands by the personal assistant and, thereby, enhanced functionality.

[0056] In some embodiments of the invention, the system 200 includes a secondary learning input 204. The secondary learning input 204 can include a variety of sources of vocabulary and grammatical expression involving the user or others in frequent contact with the user, such as user friends, family, and colleagues. In some embodiments of the invention, the system 200 can prompt a user for permission to access secondary sources. Secondary learning input 204 can include text messages 220, including text messages sent by the user or text messages sent to the user. The system 200 can, for example, analyze text messages for frequent words or phrases. In some embodiments of the invention, text messages 220 include a subset of all text messages received by the user, such as text messages sent by a pre-determined list of individuals or text messages from frequent senders. Secondary learning input 204 can include phone calls 222. For example, the system 200 could receive user permission to access the input and output of a smartphone and can monitor conversations for frequently used words or phrases, for voice accents, vocal mannerisms, and specialized

vocabulary. Secondary learning input 204 can include email messages 224, including email messages sent by the user or email messages sent to the user. The system 200 can, for example, analyze email messages for frequent words or phrases. In some embodiments of the invention, email messages 224 include a subset of all email messages received by the user, such as text messages sent by a pre-determined list of individuals or text messages from frequent senders. Secondary learning input 204 can include social media 226 access. For example, the system 200 could receive user permission to access social media postings by a user or the user's friends or family and can monitor posts for frequently used words or phrases, or specialized vocabulary.

[0057] In some embodiments of the invention, secondary learning input 204 includes user-personal assistant interaction 228. User personal assistant interaction 228 includes all voice communications a user provides to the personal assistant, including verbal commands and statements, such as scheduling requests, meeting requests, research requests, requests to perform tasks such as writing emails, text messages, or placing telephone calls, comments on results from the personal assistant, responses to personal assistant output, such as responses to requests for clarification and follow up questions by the personal assistant, and the like.

[0058] In some embodiments of the invention, the system 200 includes external classification systems 230. For example, when the system 200 encounters an unknown word or phrase, the system 200 can consult external classification systems 230, such as web-based dictionaries or urban dictionaries, to determine a meaning or characterization for the word or phrase.

[0059] As is shown, the system 200 can include an adaptive personal assistant lexicon 206. The adaptive personal assistant lexicon can analyze the primary learning input and the secondary learning input to provide a personalized verbal output. In some embodiments of the invention, the adaptive personal assistant lexicon 206 includes a starter vocabulary 232. The adaptive personal assistant lexicon 206 optionally includes a specialized vocabulary 234. The adaptive personal assistant lexicon 206 includes a personalized lexicon 236. The personalized lexicon 236 can be generated based at least in part on the primary learning input 202, the secondary learning input 204, and external classification systems 230. The personalized lexicon 236 can be continuously or periodically modified or updated, for example, through continuous or periodic receipt of secondary learning input 204. In some embodiments of the invention, the adaptive personal assistant lexicon 206 includes accent features 238. Accent features 238 can include pre-defined vocabulary or pronunciation features, such as features associated with a known geographic region (southern United States, Australia, Canada, etc.). In some embodiments of the invention, accent features 238 can include vocabulary or pronunciation features derived from analyzing speech from the user and optionally the user's friends and family.

[0060] The system 200 includes an output 212 including vocal digital output from the personal assistant. The output 212 can include any output requested by a user of a personal digital assistant, such as responses to request for information, such as requests or directions, schedules, internet searches, weather, and the like; responses for requests to perform tasks, such as calendaring, telephoning, sending text messages or emails, and the like; requests for clarifi-

cation; reminders of upcoming events, etc. In some embodiments of the invention, the adaptive personal assistant lexicon 206 can adapt the personal assistant output to include words and vocal mannerisms of the user. For example, the personal digital assistant can adapt its speech in order to approximate a relationship of a friend or trusted advisor.

[0061] FIG. 5 depicts a flow diagram illustrating a method 300 for personalized digital interaction according to one or more embodiments of the present invention. The method 300 includes, as shown at block 302, selecting a starter vocabulary based at least in part on user characteristics. The method 300 also includes, as shown at block 304, receiving a plurality of user voice communications from a user to a personal assistant. The method 300 also includes, as shown at block 306, optionally receiving a plurality of verbal and nonverbal third-party communications between a user and a third-party. For example, the third-party communications can include text messages, social media posts, telephone conversations, or emails between a user and a third-party, such as a friend, colleague, acquaintance, co-worker, or family member. The method 300 also includes, as shown at block 308, generating a frequent word list based at least in part on a plurality of user voice communications and optional third-party communications. The method 300 also includes, as shown at block 310, generating a personalized vocabulary based at least in part on the starter vocabulary and frequent word list. The method 300 also includes, as shown at block 312, delivering a personalized verbal output to a user based at least in part on the personalized vocabulary.

[0062] In some embodiments of the invention, methods for personalized digital interaction include receiving a secondary learning input from a smart device and modifying the personalized vocabulary based at least in part on the secondary learning input.

[0063] In some embodiments of the invention, selecting the starter vocabulary includes receiving a user demographic data set. The user demographic data set can include any information helpful to determining an initial vocabulary, such as age, gender, geographic location, and the like. In some embodiments of the invention, selecting the starter vocabulary includes comparing a user demographic data set to a plurality of characteristics of a starter vocabulary set. The starter vocabulary set can include base vocabularies of designated age groups, genders, and the like. Selecting the starter vocabulary can also include determining based at least in part on the comparison, a preferred starter vocabulary.

[0064] In some embodiments of the invention, methods for personalized digital interaction include analyzing a voice command from a user to determine the user's mood and adjusting a characteristic of the personalized verbal output based at least in part on the mood. For example, a mood filter can analyze the timbre, tenor, and inflection of words and identify the mood of the speaker. The adaptive digital assistant can optionally modify the verbal output delivery to better match the mood of the user. For instance, if a speaker is very dynamic and loud and laughing, the adaptive digital assistant can be dynamic and funny. If a speaker is sober and speaks slowly with little inflection or dynamics, the adaptive digital assistant can provide a relatively muted delivery output. In some embodiments of the invention, a user can set mood filter preferences, for example a user can set the mood

filter to operate only upon request or during specified time intervals. In some embodiments of the invention, methods for personalized digital interaction include selecting a set of accent features and adjusting a characteristic of the personalized verbal output based at least in part on the accent features.

[0065] Embodiments of the invention provide a voice characterization system. Voice characterization systems and methods according to embodiments of the invention can aid advertisers, television producers, and other providers of spoken content in reaching an appealing to target audiences.

[0066] Characterization of music, such as in the case of the music genome project, can include analysis, characterization, and grouping of songs. Characterization of speech, on the other hand, presents a number of complexities not present in the case of music. For example, in speech, pitch variation can form part of a more complex set of modulations known as prosody. Prosody includes speech parameters that can apply on a variety of levels, including the level of a syllable, word, phrase, or sentence, and moreover, can involve distinguishing word meanings in tone languages, distinguishing questions from statements, signaling emotion, such as irony and sarcasm, and other similar nuances of speech. Such characteristics and nuances, however, can play a role in a listener's overall preference or dislike for an instance of spoken word.

[0067] FIG. 6 depicts a diagram illustrating an exemplary voice characterization system 400 according to one or more embodiments of the present invention. The system 400 can include a reference voice sampling 402. The system 400 can also include a spoken genome hub 404. The spoken genome hub 404 can include a voice analysis system 406 and a voice characteristic database 408. In some embodiments of the invention, the system 400 also includes a target voice input 410. Target voice input 410 can include a voice input that is desired to be analyzed or compared to reference voice samplings or characteristics in the voice characteristic database 408. In some embodiments of the invention, the system 400 includes an output display 412.

[0068] Reference voice sampling 402 can include a plurality of voice data files associated with a variety of known characteristics. For example, the voice sampling 402 can include a plurality of voices having different genders, ages, socio-economic characteristics, educational levels, or any other quality that could be useful for categorization in a consumer or media context, an advertising context, an educational context, and the like.

[0069] Spoken genome hub 404 can include a voice analysis system 406 that can analyze human voice and categorize voices based at least in part on gender, pitch, cadence, tonality, rhythm, accent, timing, slurring, elocution, and other similar or related characteristics that can affect how a listener perceives, appreciates, or pays attention to a speaker. Spoken genome hub 404 can include a voice characterization database 408 that can include a plurality of reference qualities and associated voice characteristics.

[0070] FIG. 7 depicts a flow diagram illustrating a method for characterization and analysis of human voice 500 according to one or more embodiments of the present invention. The method 500 includes, as shown at block 502, receiving a plurality of media containing spoken words. The method 500 also includes, as shown at block 504, categorizing the plurality of media according to spoken genome properties to create a categorized spoken genome database.

The method 500 also includes, as shown at block 506, receiving a user media preference. The method 500 also includes, as shown at block 508, determining a user media profile based at least in part on a user media preference, wherein the user media profile includes a preferred spoken genome property. The method 500 also includes, as shown at block 510, providing a media recommendation based at least in part on the user media profile and the categorized spoken genome database.

[0071] FIG. 8 depicts a flow diagram illustrating another method for characterization and analysis of human voice 600 according to one or more embodiments of the present invention. The method 600 includes, as shown at block 602, receiving a first voice sampling including reference voices corresponding to a first known reference quality. The method 600 also includes, as shown at block 604, analyzing the first voice sampling to determine a spoken genome property corresponding to the first known reference quality. The method 600 also includes, as shown at block 606, receiving a second voice sampling including reference voices corresponding to a second known reference quality. The method 600 also includes, as shown at block 608, analyzing the second voice sampling to determine spoken genome properties corresponding to the second known reference quality. The method 600 also includes, as shown at block 610, generating a spoken genome database.

[0072] FIG. 7 depicts a flow diagram illustrating yet another method for characterization and analysis of human voice 700 according to one or more embodiments of the present invention. The method 700 includes, as shown at block 702, receiving a voice sampling including reference voices corresponding to a known reference quality. The method 700 also includes, as shown at block 704, analyzing the voice sampling to determine spoken genome properties corresponding to known reference qualities. The method 700 also includes, as shown at block 706, receiving a target voice input corresponding to a candidate. The candidate can be, for example, an advertising candidate. The method 700 also includes, as shown at block 708, determining a spoken genome property corresponding to a target voice input. The method 700 also includes, as shown at block 710, comparing a target voice spoken genome property to the spoken genome property corresponding to the known reference quality. The method 700 also includes, as shown at block 712, determining whether the candidate corresponds to the known reference quality.

[0073] For example, a couple may be fans of a television show having characteristic clever dialogue that is quickly delivered. The couple could have been searching for similar fast-paced dead-pan delivery programming and could desire to be alerted to shows that have that dynamism and spacing. Embodiments of the invention can identify other shows that the couple is likely to enjoy.

[0074] Exemplary embodiments of the invention can aid advertisers in appealing to their target demographic. For instance, a voiceover for an advertisement or the preeminent vocal quality for a television show can be scored for spoken genome categories, such as gender, pitch, cadence, tonality, rhythm, accent, timing, slurring, and elocution. A variety of demographic groups can indicate their preference for various types of content that has been scored. The preferences of such demographic groups can enable an advertiser or producer to select or coach a narrator or actors in the content being presented according to desired attributes.

[0075] For example, a 32 year old male, who is single and living in a large city in state X could be searching for a new car and a girlfriend. Embodiments of the invention could determine that males aged **25** to **35** in state X respond to a firm voice from a man that speaks with a steady cadence or to a woman speaking with a faster cadence and that has a relatively low-pitched voice. Embodiments of the invention could also identify that in state Y, the same gender and age group respond most favorably to a woman with a high-pitched voice. Based at least in part on this data, car manufacturers can tailor their ads to meet these needs.

[0076] As another example, a couple traveling cross-country by car could be in search of an audiobook that they are likely to enjoy. Embodiments of the present invention can direct the couple to audiobooks they are likely to enjoy. Or, for instance, an internet web-surfer can search the internet repeatedly for interesting content. Embodiments of the invention can aid the web-surfer in locating content she is more likely to find enjoyable.

[0077] The present invention may be a system, a method, and/or a computer program product at any possible technical detail level of integration. 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 invention.

[0078] 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 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.

[0079] 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.

[0080] Computer readable program instructions for carrying out operations of the present invention may be assembler instructions, instruction-set-architecture (ISA) instructions, machine instructions, machine dependent instructions, microcode, firmware instructions, state-setting data, configuration data for integrated circuitry, 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 Smalltalk, C++, or the like, and 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 invention.

[0081] Aspects of the present invention 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 invention. 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.

[0082] 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.

[0083] 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.

[0084] The flowchart and block diagrams 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 invention. In this regard, each block in the flowchart or block diagrams may represent a module, segment, or portion of instructions, which comprises one or more executable instructions for implementing the specified logical function(s). In some alternative implementations, the functions noted in the blocks 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 carry out combinations of special purpose hardware and computer instructions.

[0085] The terminology used herein is for the purpose of describing particular embodiments only and is not intended to be limiting of the invention. 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, element components, and/or groups thereof.

[0086] The corresponding structures, materials, acts, and equivalents of all means or step plus function elements in the claims below are intended to include any structure, material, or act for performing the function in combination with other claimed elements as specifically claimed. The description of the present invention has been presented for purposes of illustration and description, but is not intended to be exhaustive or limited to the invention in the form described. Many modifications and variations will be apparent to those of ordinary skill in the art without departing from the scope and spirit of the invention. The embodiment was chosen and described in order to best explain the principles of the invention and the practical application, and to enable others of ordinary skill in the art to understand the invention for various embodiments with various modifications as are suited to the particular use contemplated.

[0087] The flow diagrams depicted herein are just one example. There can be many variations to this diagram or the steps (or operations) described therein without departing from the spirit of embodiments of the invention. For instance, the steps can be performed in a differing order or steps can be added, deleted or modified. All of these variations are considered a part of the claimed invention.

[0088] The descriptions of the various embodiments of the present invention have been presented for purposes of illustration, but are not intended to be exhaustive or limited to the embodiments described. Many modifications and variations will be apparent to those of ordinary skill in the art without departing from the scope and spirit of the

described embodiments. The terminology used herein was chosen to best explain the principles of the embodiments, the practical application or technical improvement over technologies found in the marketplace, or to enable others of ordinary skill in the art to understand the embodiments described herein.

What is claimed is:

1. A computer-implemented method for personalized digital interaction comprising:

selecting a starter vocabulary from a starter vocabulary set;

receiving a plurality of user voice communications from a user;

generating a frequent word list based at least in part on the plurality of user voice communications;

modifying the starter vocabulary with a plurality of words from the frequent word list to generate a personalized lexicon; and

generating a personalized verbal output based at least in part on the personalized lexicon.

2. The method of claim 1, further comprising receiving a secondary learning input from a smart device and modifying the personalized vocabulary based at least in part on the secondary learning input.

3. The method of claim 1, wherein selecting the starter vocabulary comprises:

receiving a user demographic data set,

comparing the user demographic data set to a plurality of characteristics of the starter vocabulary set, and determining, based at least in part on the comparison, a preferred starter vocabulary.

4. The method of claim 1, further comprising receiving a voice command from the user.

5. The method of claim 4, further comprising:

analyzing the voice command to determine a mood; and adjusting a characteristic of the personalized verbal output based at least in part on the mood.

6. The method of claim 1, further comprising selecting a specialized vocabulary and augmenting the personalized lexicon with the specialized vocabulary.

7. The method of claim 1, further comprising selecting a set of accent features and adjusting a characteristic of the personalized verbal output based at least in part on the accent features.

8. The method of claim 1, wherein modifying the starter vocabulary comprises replacing a word from the starter vocabulary with a frequent word from the frequent word list or augmenting the starter vocabulary with a new word from the frequent word list.

9. The method of claim 1, wherein the personalized verbal output comprises a personalized gender, pitch, cadence, tonality, rhythm, accent, timing, slurring, elocution, or combination thereof.

10. The method of claim 1, further comprising communicating with an external classification system.

11. A computer program product for characterization and analysis of human voice, wherein the computer program product comprises:

a computer readable storage medium readable by a processing circuit and storing program instructions for execution by the processing circuit for performing a method comprising:

receiving a plurality of media files, wherein the plurality of media files comprise spoken words;

categorizing the plurality of media files according to spoken genome properties to create a categorized spoken genome database;

receiving a user media preference;

determining a user media profile based at least in part on the user media preference, wherein the media profile comprises a spoken genome property; and
 providing a media recommendation based at least in part on the user media profile and the categorized spoken genome database.

12. The computer program product of claim **11**, wherein the spoken genome property is selected from the group consisting of gender, pitch, cadence, tonality, rhythm, accent, timing, slurring, and elocution.

13. The computer program product of claim **11**, wherein the media recommendation comprises a recommended television program, a recommended movie, or a recommended audio book.

14. A processing system for characterization and analysis of human voice comprising:

a processor in communication with one or more types of memory, the processor configured to:

receive a first reference voice sampling, wherein the first reference voice sampling comprises a plurality of reference voices corresponding to a first known reference quality; and

analyze the first voice sampling to determine a spoken genome property corresponding to the first known reference quality.

15. The processing system of claim **14**, wherein the processor is configured to output the spoken genome property corresponding to the first known reference quality.

16. The processing system of claim **14**, wherein the processor is configured to:

receive a second reference voice sampling, wherein the second reference voice sampling comprises a plurality of reference voices corresponding to a second known reference quality; and

analyze the second voice sampling to determine a spoken genome property corresponding to the second known reference quality.

17. The processing system of claim **16**, wherein the processor is further configured to generate a spoken genome database, wherein the spoken genome database comprises:

a first bin comprising the first known reference quality and the spoken genome property corresponding to the first known reference quality, and

a second bin comprising the second known reference quality and the spoken genome property corresponding to the second known reference quality.

18. The processing system of claim **14**, wherein the processor is further configured to:

receive a target voice input corresponding to a candidate; determine a spoken genome property corresponding to the target voice input;

compare the target voice input spoken genome property to the spoken genome property corresponding to the first known reference quality; and

determine whether the candidate corresponds to the first known reference quality.

19. The processing system of claim **18**, wherein the processor is configured to output the determination of whether the candidate corresponds the first known reference quality to a display.

20. The processing system of claim **14**, wherein the spoken genome property is selected from the group consisting of pitch, cadence, tonality, rhythm, and timing.

* * * * *