UNITED STATES DISTRICT COURT
EASTERN DISTRICT OF TEXAS
MARSHALL DIVISION

KAIFI LLC, Plaintiff, v. AMAZON.COM, INC.; AMAZON.COM SERVICES LLC; and AMAZON WEB SERVICES, INC. Defendants.	Case No. 2:24-cv-00542 JURY TRIAL DEMANDED

PLAINTIFF KAIFI LLC'S P.R. 3-1 AND 3-2 DISCLOSURE OF ASSERTED CLAIMS AND INFRINGEMENT CONTENTIONS AND ACCOMPANYING DOCUMENT PRODUCTION

Pursuant to Local Patent Rules 3-1 and 3-2 and the Court's Scheduling Order, Plaintiff KAIFI LLC ("KAIFI") hereby provides its disclosure of asserted claims and infringement contentions against Defendants Amazon.com, Inc., ("Amazon.com"), Amazon.com Services LLC ("Amazon Services"), and Amazon Web Services, Inc. ("Amazon AWS"), including by reference the entirety of the attached appendices and exhibits, and accompanying document production.

KAIFI makes these disclosures based on the information that is presently known and publicly available, without the benefit of discovery. KAIFI has not completed its preparation for trial and has not received discovery from Defendants or third parties, particularly of confidential, proprietary and/or non-public documents and information that describe the accused instrumentalities, or discovery on past, discontinued or other instrumentalities reasonably similar to the presently accused instrumentalities.

KAIFI's investigations are ongoing. KAIFI reserves the right to amend, supplement or otherwise revise its asserted claims and infringement contentions, including but not limited to identification of infringed patents and claims, and infringing instrumentalities.

I. Asserted claims

Pursuant to P.R. 3-1(a), KAIFI presently identifies each claim of each patent in suit that is infringed by Defendants as follows:

U.S. Patent No. 8,930,196 (the "'196 Patent"): claims 1-20;

U.S. Patent No. 8,040,232 (the "'232 Patent"): claims 1-12;

U.S. Patent No. 11,082,518 (the "'518 patent"): claims 1-20; and

U.S. Patent No. 7,689,001 (the "'001 Patent"): claims 1-3, 5, 7-8, 10-11.

II. Accused Instrumentalities

Pursuant to P.R. 3-1(b), KAIFI presently identifies the accused apparatus, product, device, process, method, act, or other instrumentality of which it is aware. This identification is

Ex.1029 / Page 2 of 347
AMAZON.COM, INC.

as specific as possible. Each product, device, and apparatus is identified by name or model number, if known. Each method or process is identified by name, if known.

The accused instrumentalities for the '196 Patent include at least methods and apparatuses for speech recognition, including but not limited to wake word detection. The accused instrumentalities for the '232 Patent include at least sensor node and network methods, apparatuses, and systems, including but not limited to Ubiquitous Sensor Network (USN) middleware. The accused instrumentalities for the '518 Patent include at least systems for multiple machines, including but not limited to a relation server and relation profile, such as for tasks and routines. The accused instrumentalities for the '001 Patent include at least methods and devices for location recognition, including but not limited to autonomous mobile robotics.

Appendix A provides an exemplary list of accused instrumentalities, as corresponding to each of the '196 Patent, '232 Patent, '518 Patent, and '001 Patent.

The accused instrumentalities may be provided in a variety of configurations, embodiments, or combinations of the claimed components and steps. The accused instrumentalities may incorporate various devices, servers, chips, circuit elements, modules, code, and the like. Different models, versions, or iterations may incorporate or embody the same or substantially similar components or configurations.

KAIFI reserves the right to identify additional accused instrumentalities as discovery commences and its investigation continues. KAIFI anticipates the full scope of configurations or combinations of the accused and reasonably similar instrumentalities will be further identified as discovery commences and investigation continues.

III. Claim Charts

Pursuant to P.R. 3-1(c), KAIFI provides claim charts, which are incorporated by reference as if fully set forth herein, identifying where each element of each asserted claim is

found within each accused instrumentality:

Appendix B – the '196 Patent;

Appendix C – the '232 Patent;

Appendix D – the '518 Patent; and

Appendix E – the '001 Patent.

The accused instrumentalities utilize the same or substantially similar technology as presented in the charts, such that the charts are representative of the manner of infringement by each accused instrumentality. On information and belief, the charting of elements of the asserted claims to the accused instrumentalities is substantially the same. The comparisons made in these contentions are representative of, and apply to, all accused and reasonably similar instrumentalities and their configurations and implementations.

The charts subdivide the asserted claims to better illustrate where each claim element or step may be found within the accused instrumentalities. The annotations to figures are provided solely for illustrative purposes. References and descriptions provided in conjunction with one claim within a chart are applicable to other claims within the chart with respect to the same or similar elements and limitations. These subdivisions or annotations should not be taken as an indication of the boundaries of claim elements or steps with respect to the doctrine of equivalents or any other issue. For example, particular claim structures or functionalities may be embodied by or distributed over one or more elements or functional blocks, including or in addition to those that have been annotated. KAIFI reserves the right to provide alternate claim mapping, and to amend, supplement or otherwise revise these charts as discovery and its investigation continues.

IV. Contentions

A. Direct Infringement

KAIFI contends that Defendants directly infringe the asserted claims by making, using, selling, and/or offering for sale in the United States, and/or importing into the United States, this State, and this District the accused instrumentalities, without authority. KAIFI also contends that Defendants' infringement is literal and/or under the doctrine of equivalents.

For example, on information and belief, and without the benefit of discovery, KAIFI presently contends: Amazon.com directly infringes the '196 Patent, the '232 Patent, the '518 Patent, and the '001 Patent; Amazon Services directly infringes the '196 Patent, the '232 Patent, the '518 Patent, and the '001 Patent; and Amazon AWS directly infringes at least the '196 Patent and the '232 Patent.

Additionally or alternatively, each Defendant directs or controls the performance by others (either another Defendant or third party) of one or more claim limitation(s), such that the performance or use of each claim limitation can be attributed to the Defendant. In particular, Defendants condition participation in an activity or receipt of a benefit upon performance of the step(s) and establish the manner or timing of that performance. For example, Amazon Services directs or controls Amazon AWS's performance of certain steps of the claimed methods. Furthermore, Amazon AWS directs or controls Amazon Service's performance of certain steps of the claimed methods.

Additionally or alternatively, Amazon.com, Amazon Services and/or Amazon AWS act as a joint enterprise whose members can be charged with each other's acts in the enterprise. For example, a joint enterprise exists between or among Defendants by virtue of agreement, express or implied, between or among Defendants with a common purpose to be carried out. There exists a community of pecuniary interest in that there is a common purpose among the members with an equal right among all members to a voice in the direction of the enterprise, giving equal

right to control.

Additionally or alternatively, each Defendant has, at least, used the claimed system or apparatus by putting the system or apparatus as a whole into service, *i.e.*, controlled the system and obtained benefit from it. For example, at least Amazon Services, including its employees, have put systems for multiple devices, machines, or components, including but not limited to local devices and back-end servers, as a whole into service, *i.e.*, controlled the system and obtained benefit from it.

Additionally or alternatively, Defendants have directly infringed by combining all elements of the infringing systems, devices, and other instrumentalities together, including via Defendants' software connecting the claimed devices. Further, Defendants directly infringe because they are the single supplier of the accused system where it comprises multiple devices and software.

B. Indirect Infringement

KAIFI contends that Defendants indirectly infringe by actively, knowingly, and/or intentionally inducing and/or contributing to the direct infringement of the asserted claims.

Defendants have induced and continued to induce infringement of the asserted claims by others (either other Defendants or third parties, including end-users and customers). By way of non-limiting examples, (i) Amazon Services induces Amazon AWS's performance of the claimed method or use of a claimed system or apparatus; (ii) Amazon AWS induces Amazon Service's performance of the claimed method or use of a claimed system or apparatus; (iii) Amazon Services or Amazon AWS induces direct infringement by an end-user or customer; and/or (iv) Amazon.com induces direct infringement by Amazon AWS and/or Amazon Services.

Additionally or alternatively, Defendants' customers and/or end-users have used the claimed systems by putting the claimed system as a whole into service, i.e., controlled the system and obtained benefit from it and were induced to do so by Defendants.

Defendants had knowledge or were on notice of the asserted patents by virtue of pre-suit communications between Amazon and PowerVoice, citations to one or more of the asserted patents during the prosecution of patent applications by Defendants and/or their affiliates, and the filing of the complaint in this case.

The acts of inducement may include making, selling, offering to sell, and/or importing the Accused Instrumentalities to or for others, and providing others with encouragement, instructions and training on how to use the accused instrumentalities in a manner that directly infringes the claims, all while Defendants have knowledge of the asserted patents and asserted claims, and knowledge that such use of the accused instrumentalities would infringe the asserted claims. Defendants affirmatively act with knowledge and specific intent to encourage and facilitate such infringing acts by others. Defendants' tutorials, user guides, and technical and marketing documentation for the accused instrumentalities provide specific instruction for using, and actively encourage others to use, the accused instrumentalities in an infringing manner.

Affirmative acts of inducement include, but are not limited to, any one or combination of encouraging and/or facilitating the infringement of others through advertisement, marketing, and dissemination of the accused instrumentalities and their components; creating and publishing promotional and marketing materials, supporting materials, product manuals, and/or technical support and information relating to the accused instrumentalities; describing to, training, and instructing users on the implementation of the accused instrumentalities and/or its components; and providing product manuals and other technical information that cause other Defendants, Defendants' customers and other third parties to use and/or operate the accused instrumentalities for their ordinary and customary use.

Defendants took active steps to encourage end users to use and operate the accused

instrumentalities despite knowing of the asserted patents, in a manner they knew directly infringements each element of the asserted claims. Defendants specifically intended and were aware that the ordinary and customary use of the accused instrumentalities and/or their components would infringe one or more of the asserted claims.

Defendants knew that the induced conduct would constitute infringement and intended said infringement at the time of committing the aforementioned acts, such that those acts and conduct have been and continue to be committed with the specific intent to induce infringement, or to deliberately avoid learning of the infringing circumstances at the time these acts were committed, so as to be willfully blind to the induced infringement.

Defendants have and continue to contributorily infringe the asserted claims by using, offering to sell, selling, or importing the Accused Instrumentalities. The accused instrumentalities in operation form a component of a machine, manufacture, combination, or composition, constituting a material part of the invention for each of the asserted patents. Defendants know that the Accused Instrumentalities are especially made or adapted for use in an infringement of the asserted claims, and that the Accused Instrumentalities are not a staple article or other system capable of substantial non-infringing use.

C. Doctrine of Equivalents

KAIFI contends that Defendants infringe under the doctrine of equivalents, to the extent that Defendants contend any claim elements or limitations are not found literally. Defendants have not specifically identified which claim elements or limitations they contend are not literally met. Should Defendants contend that any element, step or limitation is literally absent, KAIFI contends and reserves the right to prove that at least the equivalent of the allegedly missing element, step or limitation is nevertheless present and that accused instrumentalities at least infringe under the doctrine of equivalents. KAIFI anticipates that discovery will enable it to set

forth the additional scope of equivalents, and therefore reserves the right to amend, supplement or otherwise revise these contentions, and specifically as to the doctrine of equivalents.

V. Priority Dates

Pursuant to P.R. 3-1(e), Plaintiff identifies the priority date to which each asserted claim is entitled for asserted patents that claim priority to an earlier application.

- The asserted claims of '196 claim priority to PCT Appl. PCT/KR2009/002118 filed April 22, 2009.
- The asserted claims of '232 claim priority to PCT Appl. PCT/KR2006/005715 filed December 26, 2006.
- The asserted claims of '518 claim priority to US Appl. 14/931,213 filed November 3, 2015.
- The asserted claims of '001 claim priority to KR Appl. 10-2005-0114022 filed November 28, 2005.

KAIFI's investigation is continuing. KAIFI reserves the right to supplement its priority claim date (and to assert any applicable prior invention dates).

VI. Plaintiff's Practicing Instrumentalities

Pursuant to P.R. 3-1(f), at this time, KAIFI is not asserting that any apparatus, product, device, process, method, act, or other instrumentality of its own practices any of the claimed inventions. KAIFI's investigation is continuing. To the extent that additional responsive, discoverable information is found, including with respect to third parties, KAIFI reserves the right to and will supplement its response concerning P.R. 3-1(f) as appropriate.

VII. Document Production

Pursuant to P.R. 3-2(a), KAIFI is producing the following non-privileged documents sufficient to evidence each discussion with, disclosure to, or other manner of providing to a third party, or sale of or offer to sell, the claimed invention prior to the date of application for the patent in suit:

Bates Nos. KAIFI0007448 - KAIFI0007495; KAIFI0007509 - KAIFI0007579; KAIFI0007673 - KAIFI0007691.

Pursuant to P.R. 3-2(b), KAIFI is producing herewith non-privileged documents evidencing the conception, reduction to practice, design, and development of each claimed invention, which were created on or before the date of application for the patent in suit or the priority date identified pursuant to P. R. 3-1(e):

Bates Nos. KAIFI0007448 - KAIFI0007691.

Pursuant to P.R. 3-2(c), KAIFI is producing herewith a copy of the file history for each of the Asserted Patents:

Bates Nos. KAIFI0000001 - KAIFI0001071.

KAIFI reserves the right to amend or supplement this production based on further investigation.

Date: November 14, 2024

/s/ Andrew Choung

Jennifer L. Truelove
Texas State Bar No. 24012906
jtruelove@mckoolsmith.com
MCKOOL SMITH, P.C.
104 East Houston Street Suite 300
Marshall, TX 75670
Telephone: (903) 923-9000
Facsimile: (903) 923-9099

Jason G. Sheasby
(*pro hac vice* to be filed)
jsheasby@irell.com
IRELL & MANELLA LLP
1800 Avenue of the Stars, Suite 900
Los Angeles, CA 90067-4276
Telephone: (310) 277-1010
Facsimile: (310) 203-7199

Andrew Choung
(admitted in E.D. Texas)
achoung@nixonpeabody.com
Jennifer Hayes
(admitted *Pro Hac Vice*)
jhayes@nixonpeabody.com
NIXON PEABODY LLP
300 S. Grand Avenue, Suite 4100
Los Angeles, CA 90071
Telephone: (213) 629-6000
Facsimile: (213) 629-6001

Attorneys for Plaintiff KAIFI LLC

CERTIFICATE OF SERVICE

I certify that the foregoing document was served by electronic mail to counsel for Defendants.

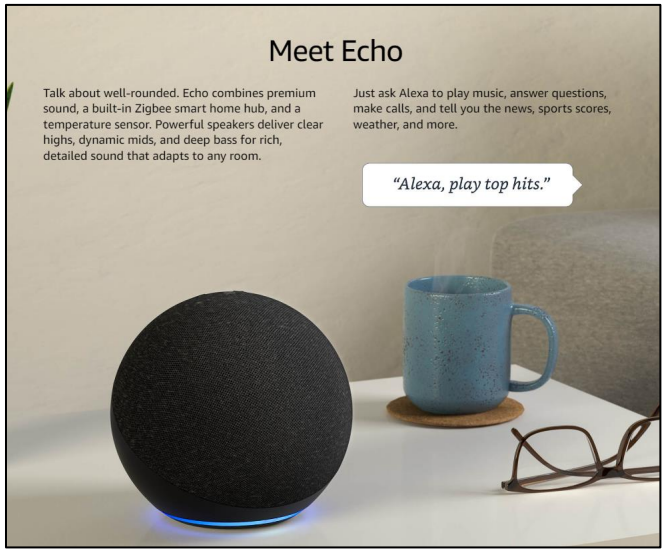
Date: November 14, 2024

/s/ Andrew Choung

Ex.1029 / Page 12 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

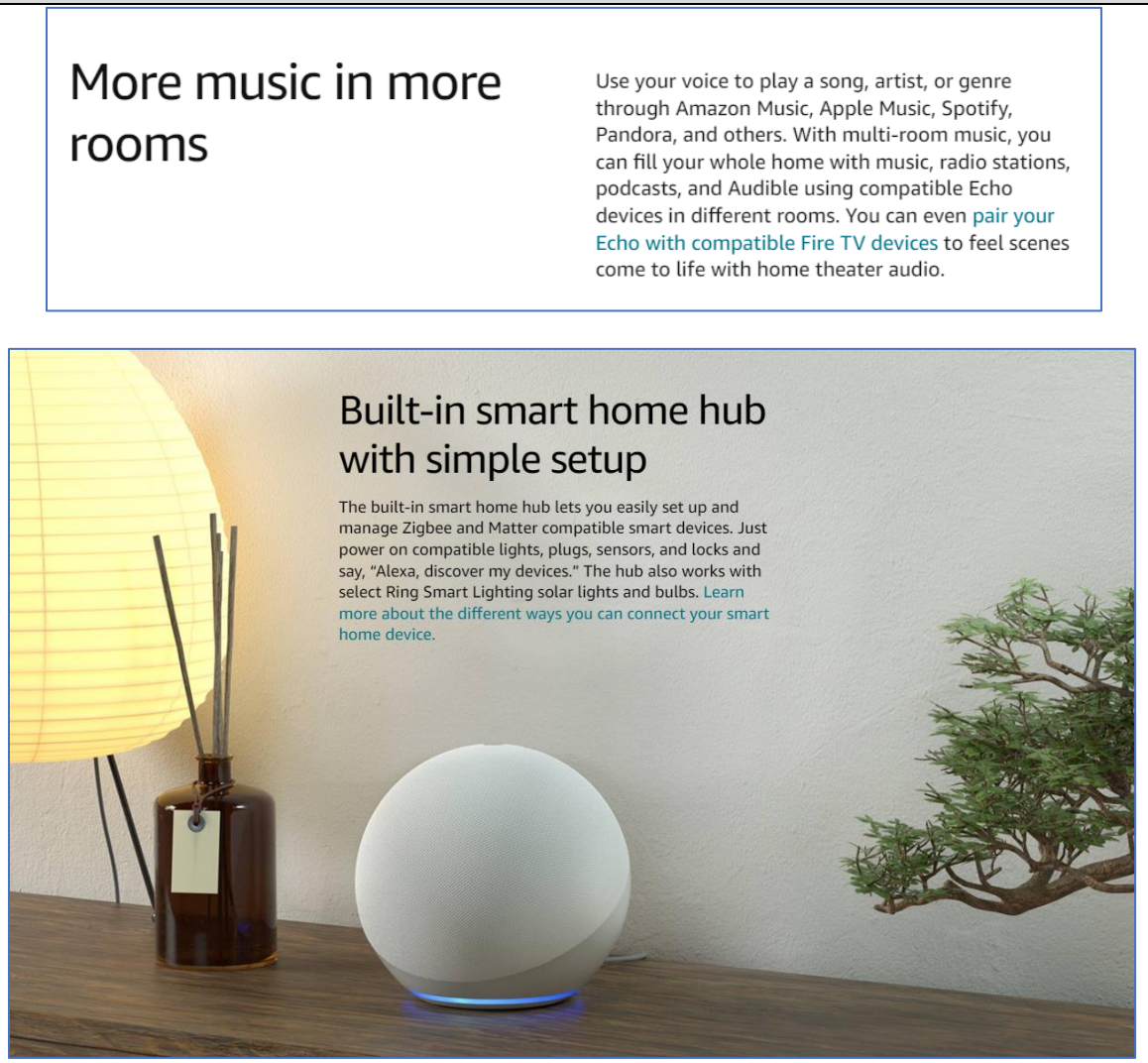
Claim 1 of the ’196 Patent

Claim 1	Accused Instrumentality(ies)
A method for recognizing a speech of a speaker by a device, the method comprising:	Amazon practices a method of recognizing a speech of a speaker by a device, for example, in its Echo and Alexa devices, such as the Echo, Echo Pop, Echo Auto, Echo Show, Echo Flex, Echo Dot, Echo Studio, Echo Buds, and Echo Hub, as well as in Alexa Built-in Devices, and on Amazon’s cloud systems where it provides wake word detection, Automatic Speech Recognition (ASR) and Alexa Voice Service (AVS). For example, Amazon provides Echo (4th Gen) with premium sound, smart home hub, and Alexa.[1] 

[1] *See* KAIFI0001202 (*Echo (4th Gen) | With premium sound, smart home hub, and Alexa | Charcoal*, https://www.amazon.com/dp/B07XKF5RM3?ref=MARS_NAV_desktop_shop_dvc_smart_spkrs_Echo_Spkr).

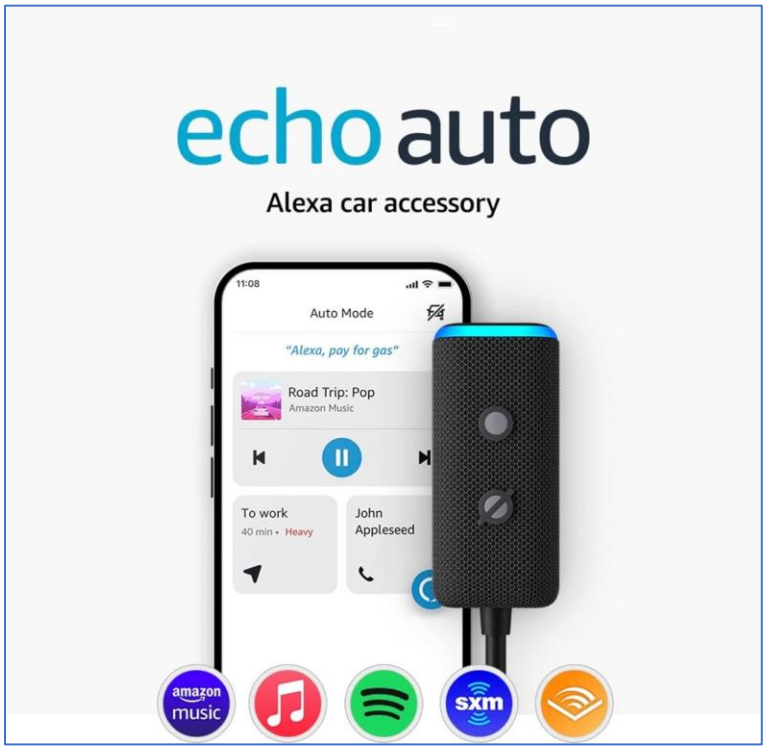
Ex.1029 / Page 13 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	

Ex.1029 / Page 14 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	For example, Amazon also provides the Echo Auto.[2] 

[2] *See* KAIFI0001979 *(Echo Auto (2nd Gen, 2022 release)*, https://www.amazon.com/All-New-Echo-Auto-release-Alexa/dp/B09X27YPS1?th=1).

Ex.1029 / Page 15 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

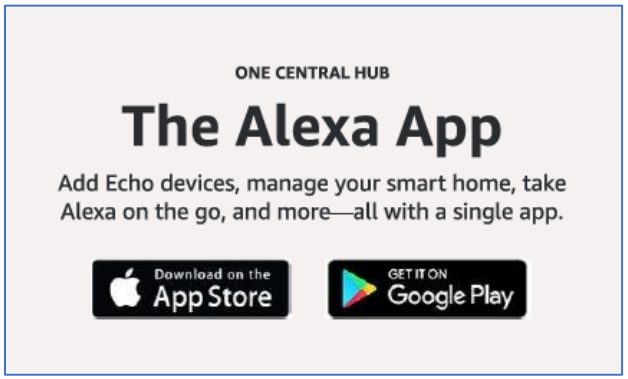
Claim 1	Accused Instrumentality(ies)
	According to Amazon, the Echo Auto provides speech recognition and synthesis.[3] Alexa Auto SDK Capabilities The Alexa Auto SDK includes core Alexa functionality, such as speech recognition and synthesis, and other capabilities such as streaming media, controlling smart home devices, notifications, weather reports, and tens-of-thousands of custom skills. Additionally, the SDK provides the hooks required to connect to a wake word engine, local media player, local phone, and local navigation system, and the ability to access certain Alexa features and capabilities even when internet connectivity is intermittent or not available. Amazon also provides Alexa.[4] What Is Alexa? Alexa is Amazon's cloud-based voice service available on hundreds of millions of devices from Amazon and third-party device manufacturers. With Alexa, you can build natural voice experiences that offer customers a more intuitive way to interact with the technology they use every day. We offer a collection of tools, APIs, reference solutions, and documentation to make it easier to build for Alexa. Start building for voice today by creating Alexa skills, connecting Alexa to devices, or integrating Alexa directly into your products. You can engage our Alexa Solution Provider network for a range of services including strategy, pre-tested reference architectures and hardware, hardware and software development, manufacturing, and go-to-market support. 

[3] *See* KAIFI0001144 *(Alexa Auto Software Development Kit*, https://developer.amazon.com/en-US/alexa/alexa-auto/sdk).

[4] *See* KAIFI0007144 (*What Is Alexa*, https://developer.amazon.com/en-US/alexa).

Ex.1029 / Page 16 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

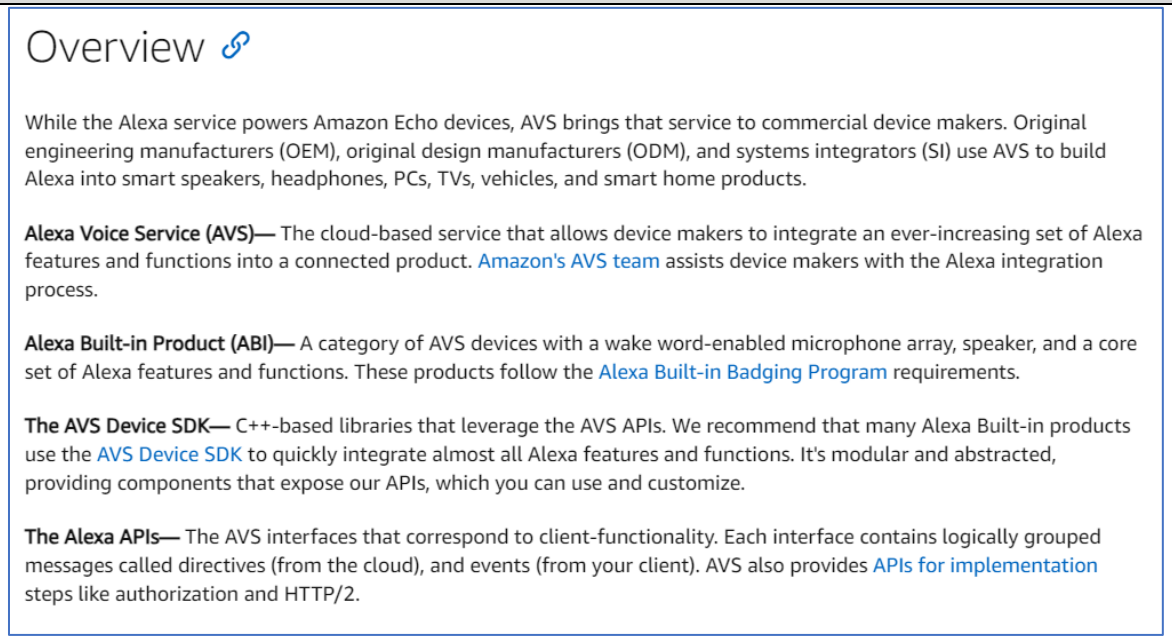
Claim 1	Accused Instrumentality(ies)
	Amazon also provides the Alexa App.[5]  Amazon also provides Alexa Voice Service (AVS).[6]

[5] *See* KAIFI0007048 (*The Alexa App*, https://www.amazon.com/Alexa-App/b?ie=UTF8&node=18354642011).

[6] *See* KAIFI0007371 *(What is the Alexa Voice Service?*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/get-started-with-alexa-voice-service.html).

Ex.1029 / Page 17 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	 Amazon provides speech processing as both cloud-based and on-device.[7] At Amazon, we always look to invent new technology for improving customer experience. One technology we have been working on at Alexa is on-device speech processing, which has multiple benefits: a reduction in latency, or the time it takes Alexa to respond to queries; lowered bandwidth consumption, which is important on portable devices; and increased availability in in-car units and other applications where Internet connectivity is intermittent. On-device processing also enables the

[7] *See* KAIFI0006791(Ariya Rastrow, Shehzad Mevawalla, *On-device speech processing makes Alexa faster, lower-bandwidth*, AMAZON SCIENCE (Jan. 25, 2022), https://www.amazon.science/blog/on-device-speech-processing-makes-alexa-faster-lower-bandwidth).

Ex.1029 / Page 18 of 347
AMAZON.COM, INC.

Claim 1	Accused Instrumentality(ies)
	fusion of the speech signal with other modalities, like vision, for features such as Alexa's natural turn-taking. In the last year, we've continued to build upon Alexa's on-device speech-processing capabilities. As a result of these inventions, we are launching a new setting that gives customers the option of having the audio of their Alexa voice requests processed locally, without being sent to the cloud. In the cloud, storage space and computational capacity are effectively unconstrained. To ensure accuracy, our cloud models can be large and computationally demanding. Executing the same functions on-device means compressing our models into less than 1% as much space — with minimal loss in accuracy. Moreover, in the cloud, the separate components of Alexa's speech-processing stack — automatic speech recognition (ASR), whisper detection, and speaker identification — run on separate server nodes with their own powerful processors. On-device, those functions have to share hardware not only with each other but with Alexa's other core device functions, such as music playback. Re-creating Alexa's speech-processing stack on-device was a massive undertaking. New methods for training small-footprint ASR models were part of the solution, but so were innovations in system design and hardware-software codesign. It was a joint effort across science and engineering teams over a span of years. Here's a quick overview of how it works. For example, Amazon describes the technology supporting a voice request.[8]

[8] *See* KAIFI0002023 *(Follow the journey of a voice request*, https://www.amazon.com/b/?node=23608618011).

Ex.1029 / Page 19 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	You can ask Alexa any question and get a response in a matter of seconds. If you're interested in the technology behind all of that magic, you've come to the right place. Here's a more detailed look at how Alexa works and the ways in which Echo devices and Alexa are designed to protect your privacy every step of the way. **The wake word** Think of the wake word as a verbal cue that makes things happen. When you have a request for Alexa, you first need to say your chosen wake word, which by default is "Alexa." Only after your Echo device detects the wake word is Alexa listening to your requests. When it comes to privacy, there should be no surprises. You'll always be able to tell when Alexa is listening to your request because a light indicator will appear on your Echo device or an audible tone will sound. Think of the "On the Air" signs that light up in television studios during a broadcast. These indicators notify you that your device has detected the wake word and Alexa is now processing your request. **Keyword spotting** Echo devices use built-in technology called 'keyword spotting' that matches spoken audio to the acoustic patterns of the wake word. Simply put, Echo devices are designed by default to detect only the sound waves of your chosen wake word, and everything else is ignored. Like water through a strainer, all other audio (people talking, taps running, birds chirping) passes through the device until the wake word is 'caught' and sent to Amazon's secure cloud, where meaning is assigned to your request. **Cloud verification and encryption** When your Echo device detects the wake word, it sends your request to Amazon's secure cloud where audio is reanalysed to verify the wake word was spoken. If this cloud software verification is unable to confirm the wake word was spoken, the Alexa system stops processing the audio. If the wake word is verified (or Alexa is activated using the Action button), your request will be processed using several sophisticated algorithms that allow Alexa to respond appropriately. All of your interactions with Alexa are encrypted in transit throughout this process. **Automatic Speech Recognition (ASR)** After wake word verification, your voice request undergoes Automatic Speech Recognition (ASR) which transcribes audio into text. Using our example request (***"Alexa, what's the weather in sydney?"***), transcriptions could include one or several of the following:

Ex.1029 / Page 20 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	For example, Amazon provides and uses Alexa Built-in Devices, which are powered by Alexa Voice Service (AVS).[9]

What are Alexa Built-in Devices?



The world's leading OEMs offer Alexa Built-in devices that let customers talk directly with Alexa through a microphone and speaker. Alexa turns your device into a smart device, and expands its capabilities. Customers can ask Alexa for help with everyday tasks and entertainment, and control Alexa connected devices or their own device.

Learn About the Underlying Technology

Alexa Built-in devices are powered by **Alexa Voice Service (AVS)**. The AVS team manages the services and infrastructure required for Alexa experiences, and provides a suite of device APIs, SDKs, and documentation. Developers may also work with third-party solution providers to utilize a qualified AVS solution.

Start Developing with Solution Providers

Our global network of **Alexa Solution Providers** work together to help you develop high-quality experiences, accelerate your development, and reduce your overall development cost. The earlier you engage a Solution Provider, the better they can assist you.

[9] *See* KAIFI0007733 *(Alexa Built-in Devices, https://developer.amazon.com/en-US/alexa/devices/alexa-built-in).*

Ex.1029 / Page 21 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	 For example, Amazon provides AVS Integration, which is a feature of AWS IoT Core that enables Alexa Voice to be produced on any type of connected device, and shifts various Alexa Voice functions from a physical device to a new virtual Built-in device in the cloud.[10]

[10] *See* KAIFI0007818 *(New Alexa Voice Service Integration for AWS IoT Core cost-effectively brings Alexa Voice to any type of connected device*, https://aws.amazon.com/about-aws/whats-new/2019/11/new-alexa-voice-integration-for-aws-iot-core-cost-effectively-brings-alexa-voice-to-any-connected-device/*)*.

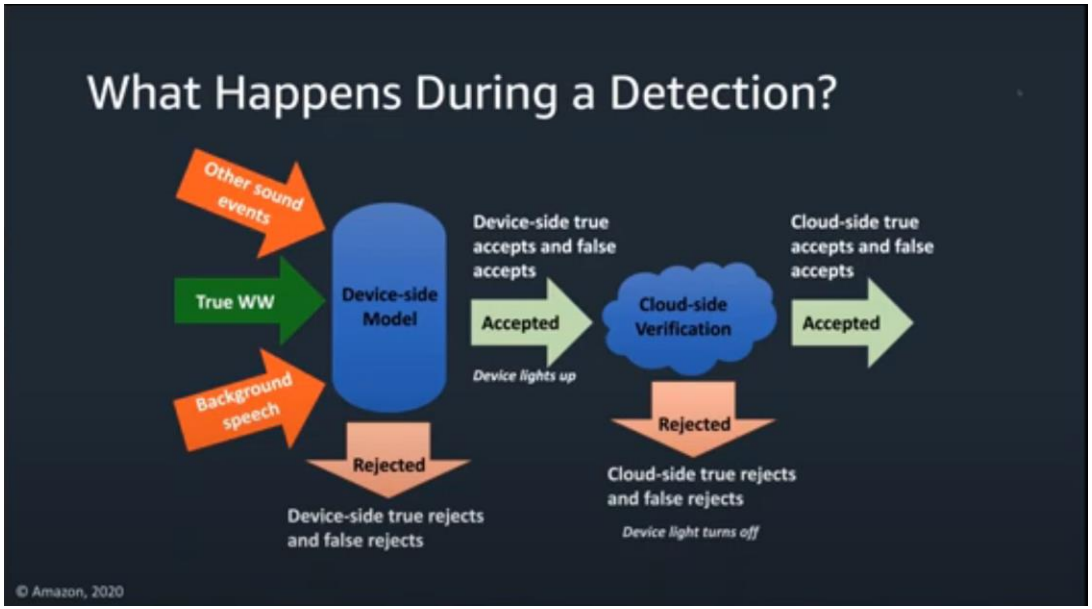
Ex.1029 / Page 22 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	New Alexa Voice Service Integration for AWS IoT Core cost-effectively brings Alexa Voice to any type of connected device Posted On: Nov 25, 2019 Alexa Voice Service (AVS) Integration is a new feature of AWS IoT Core that enables Alexa Voice to be produced on any type of connected device. AVS Integration for IoT Core reduces the cost of producing Alexa Built-in devices by up to 50% by offloading compute and memory intensive audio workloads to the cloud. With this reduction in production cost, customers can now cost-effectively build new categories of differentiated voice-enabled products such as light switches, thermostats, and small appliances. This allows consumers to talk directly to Alexa in new parts of their home, office, or hotel rooms for a truly ambient experience. Alexa Voice Service (AVS) Integration for IoT Core shifts media retrieval, audio decoding, audio mixing, and state management from a physical device to a new virtual Alexa Built-in device in the cloud. This lowers the hardware requirements of integrating AVS on a device from 50 MB of RAM to 1MB of RAM and ARM Cortex 'A' class microprocessors to ARM Cortex 'M' class microcontrollers. Once a device is connected to AWS IoT Core, customers can use a new set of AWS IoT-reserved MQTT topics to transfer audio messages between their devices and AVS. The AVS Integration for AWS IoT Core is available in all AWS regions where AWS IoT Core is available other than China (Beijing and Ningxia), Asia Pacific (Hong Kong) and Middle East (Bahrain). See the AWS Region Table for the current list of regions for AWS IoT Core. To learn more about how to get started, view the AVS Integration for IoT Core Blog Post.
receiving, by a preset call command recognition unit, an input speech spoken by the speaker;	Amazon receives, through a preset call command recognition unit, an input speech spoken by the speaker.

Ex.1029 / Page 23 of 347
AMAZON.COM, INC.

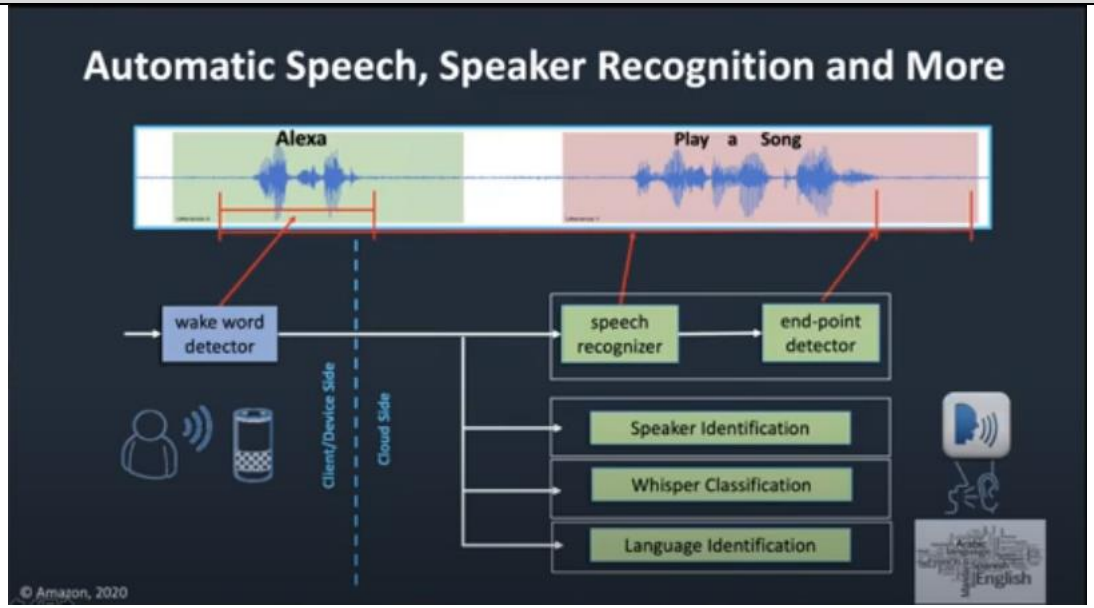
Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	For example, Amazon receives a wake word through a combination of hardware, processing, and software functions both on devices and on the cloud server.[11]  This is further illustrated below.[12]

[11] *See* KAIFI0007814 *(The Science Behind Alexa's Wake Word Detection*, https://www.youtube.com/watch?v=kYrw2cC9dvs).
[12] *See* KAIFI0007816 *(How Alexa Uses Automatic Speech Recognition*, https://www.youtube.com/watch?v=8drK6Oyv1a0&t=296s).

Ex.1029 / Page 24 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	 For example, Amazon's wake word engine is configured to receive an input speech spoken by the speaker.[13] Voice-initiated devices begin to stream user speech to AVS when the wake word engine detects a spoken wake word, such as "Alexa."

[13] *See* KAIFI0007134 *(Wake Word Verification Requirements*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/implement-ww-verification.html).

Ex.1029 / Page 25 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	4. **Wake Word Engine** - The component that reads the audio samples written to the shared memory ring buffer and analyzes the audio samples for occurrences of the wake word. The procedure of reading and processing audio samples occurs in a loop. For example, the loop reads N samples starting at location `0`; then analyzes the samples for an occurrence of the wake word. In the next iteration, it reads and analyzes the next N samples. When an occurrence of the wake word is detected, the wake word engine must identify two locations: 1) the start of the wake word, and 2) the end of the wake word. These values are stored as `startIndexInSamples` and `endIndexInSamples`. For example, according to Amazon, its speech recognizer captures and interacts with user speech.[14] The `SpeechRecognizer` interface is the core interface of the Alexa Voice Service (AVS) and exposes directives and events for capturing and interacting with user speech. This page discusses the concepts and process flows for the `SpeechRecognizer` interface.

[14] *See* KAIFI0007093 *(Understand the AVS SpeechRecognizer*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/avs-speechrecognizer-concepts.html).

Ex.1029 / Page 26 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	SpeechRecognizer functionality Every user utterance leverages `SpeechRecognizer`, including the following interaction types between an Alexa Built-in device and AVS: • **User Speech Capture**: Captures user speech from an Alexa Built-in device. • **User Speech Prompting**: Prompts a user for more speech input when needed by Alexa to deliver an appropriate response. • **Interaction Initiation Communication**: Enables a device to inform AVS of how a user initiated an Alexa interaction, such as press-and-hold, tap-and-release, or voice-initiated/wake word enabled. See Device Form Factor and Alexa Interaction. • **ASR Profile Selection**: SpeechRecognizers chooses the appropriate Automatic Speech Recognition (ASR) profile for your product, which allows Alexa to understand user speech and respond with precision. See Automatic Speech Recognition (ASR) profile • **RECOGNIZING:** When a user begins interacting with your client, specifically when the client streams captured audio to AVS, `SpeechRecognizer` should transition from the idle state to the `Recognize` state. `SpeechRecognizer` should remain in the `Recognize` until the client stops recording speech or finishes streaming, at which point your `SpeechRecognizer` component should transition from the `Recognize` state to the "busy" state. • **BUSY:** While processing the speech request, `SpeechRecognizer` should be in the "busy" state. You cannot start another speech request until `SpeechRecognizer` transitions out of the busy state. From the *busy* state, `SpeechRecognizer` transitions to the idle state if Alexa processes and completes the request or to the `ExpectSpeech` state, if Alexa requires more speech input from the user. • **EXPECTING SPEECH:** `SpeechRecognizer` should be in the `ExpectSpeech` state when Alexa requires more audio input from a user. From `ExpectSpeech`, `SpeechRecognizer` should transition to the `Recognize` when a user interaction occurs, or the interaction is automatically started on behalf of the user. It should transition to the idle state if Alexa detects no user interaction within the specified timeout window.

Ex.1029 / Page 27 of 347
AMAZON.COM, INC.

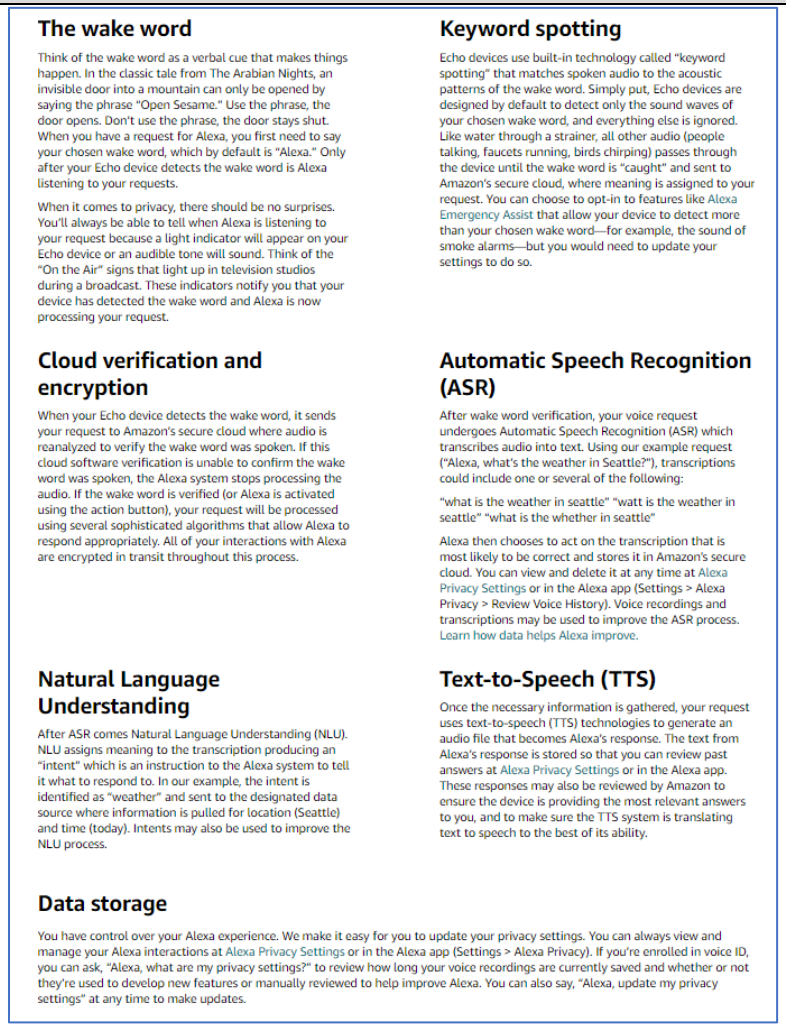
Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	For example, Amazon uses wake word, keyword spotting, cloud verification and encryption, automatic speech recognition (ASR), natural language understanding (NLU), text-to-speech (TTS), and data storage technologies.[15]

[15] *See* KAIFI0002023 *(Follow the journey of a voice request*, https://www.amazon.com/b/?node=23608618011).

Ex.1029 / Page 28 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	

The wake word

Think of the wake word as a verbal cue that makes things happen. In the classic tale from The Arabian Nights, an invisible door into a mountain can only be opened by saying the phrase "Open Sesame." Use the phrase, the door opens. Don't use the phrase, the door stays shut. When you have a request for Alexa, you first need to say your chosen wake word, which by default is "Alexa." Only after your Echo device detects the wake word is Alexa listening to your requests.

When it comes to privacy, there should be no surprises. You'll always be able to tell when Alexa is listening to your request because a light indicator will appear on your Echo device or an audible tone will sound. Think of the "On the Air" signs that light up in television studios during a broadcast. These indicators notify you that your device has detected the wake word and Alexa is now processing your request.

Keyword spotting

Echo devices use built-in technology called "keyword spotting" that matches spoken audio to the acoustic patterns of the wake word. Simply put, Echo devices are designed by default to detect only the sound waves of your chosen wake word, and everything else is ignored. Like water through a strainer, all other audio (people talking, faucets running, birds chirping) passes through the device until the wake word is "caught" and sent to Amazon's secure cloud, where meaning is assigned to your request. You can choose to opt-in to features like Alexa Emergency Assist that allow your device to detect more than your chosen wake word—for example, the sound of smoke alarms—but you would need to update your settings to do so.

Cloud verification and encryption

When your Echo device detects the wake word, it sends your request to Amazon's secure cloud where audio is reanalyzed to verify the wake word was spoken. If this cloud software verification is unable to confirm the wake word was spoken, the Alexa system stops processing the audio. If the wake word is verified (or Alexa is activated using the action button), your request will be processed using several sophisticated algorithms that allow Alexa to respond appropriately. All of your interactions with Alexa are encrypted in transit throughout this process.

Automatic Speech Recognition (ASR)

After wake word verification, your voice request undergoes Automatic Speech Recognition (ASR) which transcribes audio into text. Using our example request ("Alexa, what's the weather in Seattle?"), transcriptions could include one or several of the following:

"what is the weather in seattle" "watt is the weather in seattle" "what is the whether in seattle"

Alexa then chooses to act on the transcription that is most likely to be correct and stores it in Amazon's secure cloud. You can view and delete it at any time at Alexa Privacy Settings or in the Alexa app (Settings > Alexa Privacy > Review Voice History). Voice recordings and transcriptions may be used to improve the ASR process. Learn how data helps Alexa improve.

Natural Language Understanding

After ASR comes Natural Language Understanding (NLU). NLU assigns meaning to the transcription producing an "intent" which is an instruction to the Alexa system to tell it what to respond to. In our example, the intent is identified as "weather" and sent to the designated data source where information is pulled for location (Seattle) and time (today). Intents may also be used to improve the NLU process.

Text-to-Speech (TTS)

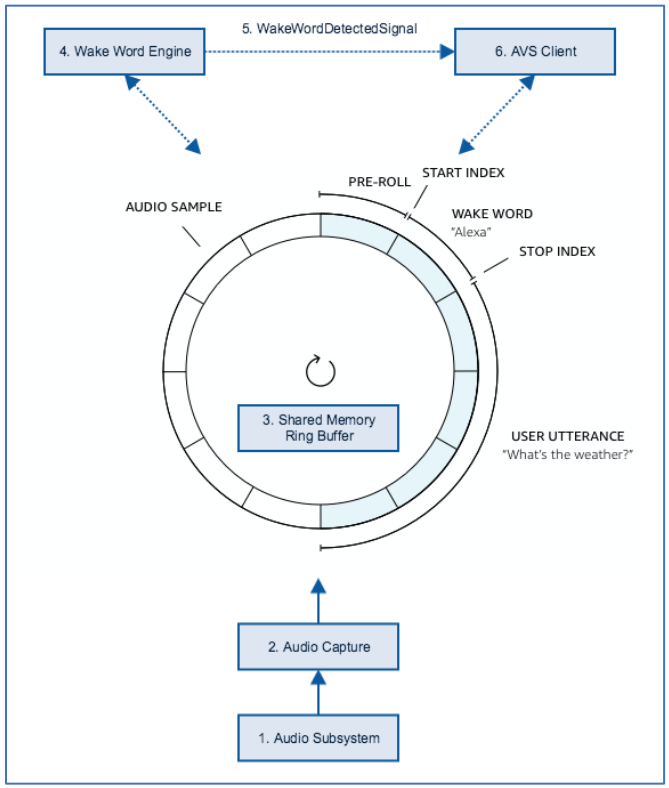
Once the necessary information is gathered, your request uses text-to-speech (TTS) technologies to generate an audio file that becomes Alexa's response. The text from Alexa's response is stored so that you can review past answers at Alexa Privacy Settings or in the Alexa app. These responses may also be reviewed by Amazon to ensure the device is providing the most relevant answers to you, and to make sure the TTS system is translating text to speech to the best of its ability.

Data storage

You have control over your Alexa experience. We make it easy for you to update your privacy settings. You can always view and manage your Alexa interactions at Alexa Privacy Settings or in the Alexa app (Settings > Alexa Privacy). If you're enrolled in voice ID, you can ask, "Alexa, what are my privacy settings?" to review how long your voice recordings are currently saved and whether or not they're used to develop new features or manually reviewed to help improve Alexa. You can also say, "Alexa, update my privacy settings" at any time to make updates.

Ex.1029 / Page 29 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	For example, the wake word engine is configured to receive an input speech spoken by the speaker.[16] 

[16] *See* KAIFI0007141 *(Wake Word Verification Requirements*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/implement-ww-verification.html).

Ex.1029 / Page 30 of 347
AMAZON.COM, INC.

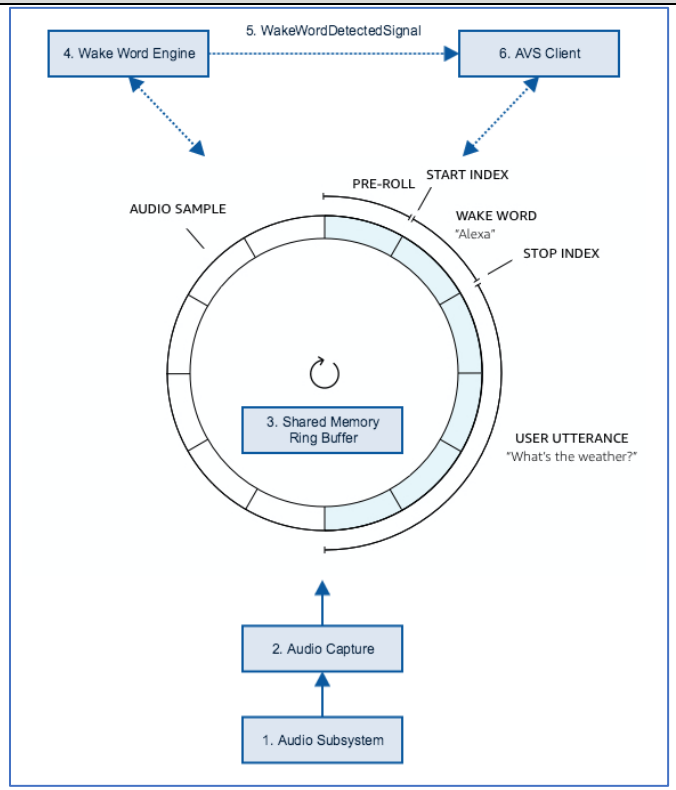
Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	3. **Shared Memory Ring Buffer** - A shared memory ring buffer object that is thread safe, and accessible by multiple processes. Access to this memory block should support at least one writer and two readers simultaneously. You may choose to implement these executables as a shared library. • **Write API** - Used by the recording process to write audio samples to the shared memory ring buffer. Write samples should move data so that the oldest audio samples are overwritten. This example shows a typical API signature: `WriteSamples(BufferIn, BufferInLength)`. • **Read API** - Used by the wake word engine and the AVS client to read from specific locations in the shared memory ring buffer. A typical implementation represents location as a 64-bit integer. For example, reading 200 samples from the location 100 is reading the 100th to 300th samples that were written to the shared memory ring buffer. This is an example of a typical API signature: `ReadSamples(BufferIn, BufferInLength)`. **Note**: Reading from an old location that was overwritten with new audio samples will result in an unintended sample read. 4. **Wake Word Engine** - The component that reads the audio samples written to the shared memory ring buffer and analyzes the audio samples for occurrences of the wake word. The procedure of reading and processing audio samples occurs in a loop. For example, the loop reads N samples starting at location `0`; then analyzes the samples for an occurrence of the wake word. In the next iteration, it reads and analyzes the next N samples. When an occurrence of the wake word is detected, the wake word engine must identify two locations: 1) the start of the wake word, and 2) the end of the wake word. These values are stored as `startIndexInSamples` and `endIndexInSamples`. For example, according to Amazon, the AVS Client is also configured to receive an input speech spoken by the speaker.[17]

[17] *See Id.*

Ex.1029 / Page 31 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	 6. **AVS Client** - The client receives the `WakeWordDetectedSignal`, extracts the `startIndexInSamples`, then reads and streams audio samples to AVS including 500 milliseconds of pre-roll. • The pre-roll is the number of samples corresponding to 500 milliseconds. For example, if the audio samples were recorded at a rate of 16 kilohertz (kHz), then pre-roll is 8,000 samples.

Ex.1029 / Page 32 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	For example, according to Amazon, the Alexa Voice Service speech recognizer captures and interacts with user speech.[18]

> The `SpeechRecognizer` interface is the core interface of the Alexa Voice Service (AVS) and exposes directives and events for capturing and interacting with user speech. This page discusses the concepts and process flows for the `SpeechRecognizer` interface.

SpeechRecognizer functionality

Every user utterance leverages `SpeechRecognizer`, including the following interaction types between an Alexa Built-in device and AVS:

- **User Speech Capture**: Captures user speech from an Alexa Built-in device.
- **User Speech Prompting**: Prompts a user for more speech input when needed by Alexa to deliver an appropriate response.
- **Interaction Initiation Communication**: Enables a device to inform AVS of how a user initiated an Alexa interaction, such as press-and-hold, tap-and-release, or voice-initiated/wake word enabled. See Device Form Factor and Alexa Interaction.
- **ASR Profile Selection**: SpeechRecognizers chooses the appropriate Automatic Speech Recognition (ASR) profile for your product, which allows Alexa to understand user speech and respond with precision. See Automatic Speech Recognition (ASR) profile

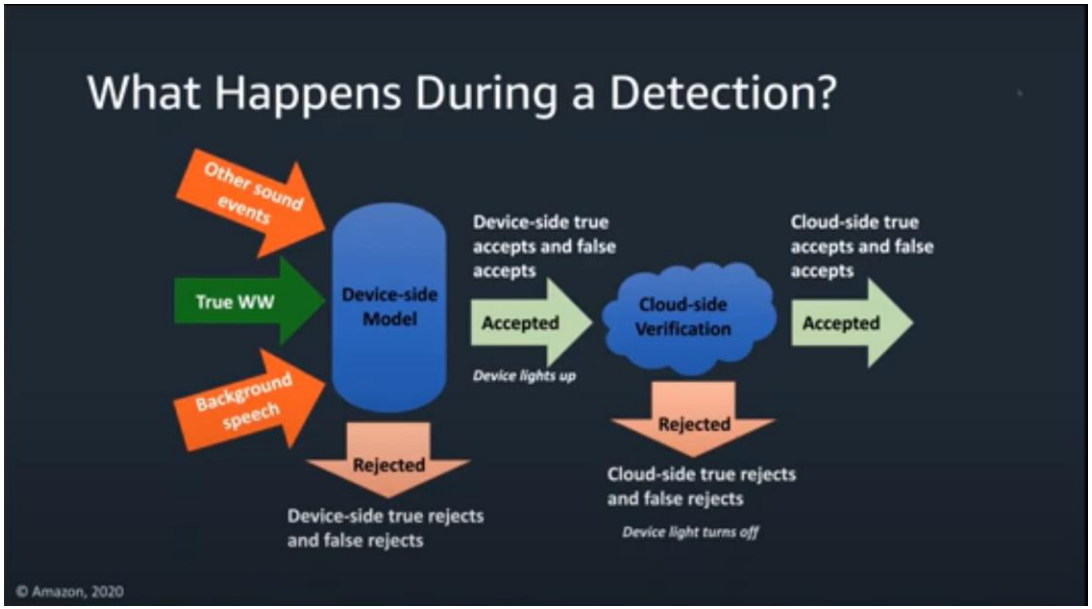
[18] *See* KAIFI0007093 *(Understand the AVS SpeechRecognizer*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/avs-speechrecognizer-concepts.html).

Ex.1029 / Page 33 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	• **RECOGNIZING:** When a user begins interacting with your client, specifically when the client streams captured audio to AVS, `SpeechRecognizer` should transition from the idle state to the `Recognize` state. `SpeechRecognizer` should remain in the `Recognize` until the client stops recording speech or finishes streaming, at which point your `SpeechRecognizer` component should transition from the `Recognize` state to the "busy" state. • **BUSY:** While processing the speech request, `SpeechRecognizer` should be in the "busy" state. You cannot start another speech request until `SpeechRecognizer` transitions out of the busy state. From the *busy* state, `SpeechRecognizer` transitions to the idle state if Alexa processes and completes the request or to the `ExpectSpeech` state, if Alexa requires more speech input from the user. • **EXPECTING SPEECH:** `SpeechRecognizer` should be in the `ExpectSpeech` state when Alexa requires more audio input from a user. From `ExpectSpeech`, `SpeechRecognizer` should transition to the `Recognize` when a user interaction occurs, or the interaction is automatically started on behalf of the user. It should transition to the idle state if Alexa detects no user interaction within the specified timeout window. This claim element is literally present, as described herein, but is also satisfied under the doctrine of equivalents because the structure and functionality provided by Amazon is being used to satisfy this claim limitation and therefore there is no vitiation of this element, and the Accused Instrumentalities incorporate structure and functionality that is not substantially different from the requirements of this limitation because they achieve substantially or exactly the same function, in substantially or exactly the same way as set forth in the claim element (as described herein), and achieves substantially or exactly the same result.
receiving, by the preset call command recognition unit, a subsequent input speech spoken, by the	Amazon receives, through the preset call command recognition unit, a subsequent input speech spoken, by the speaker, subsequently to the input speech.

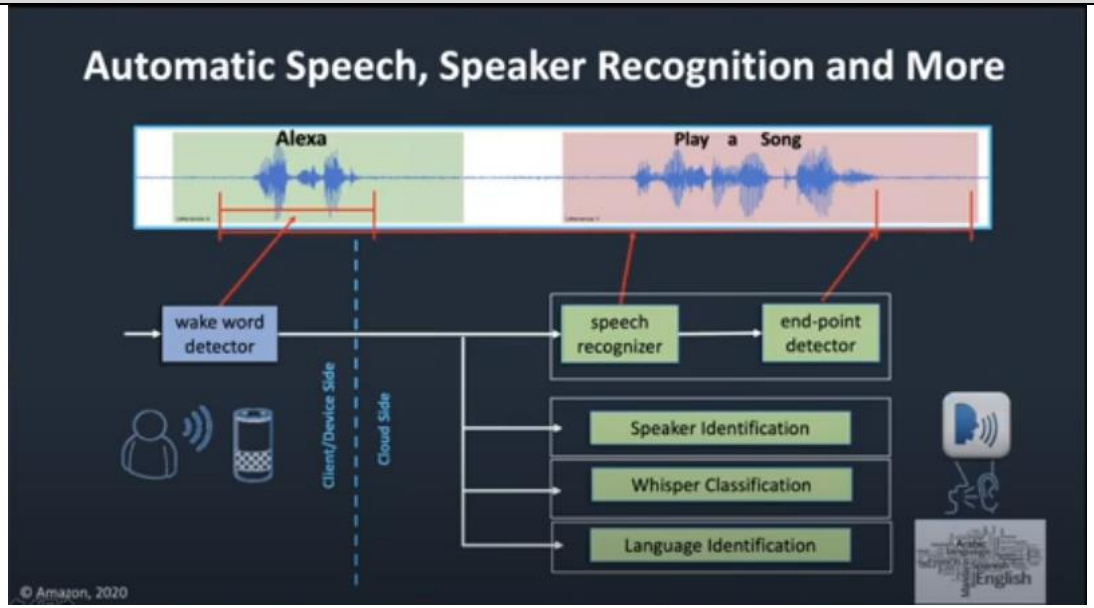
Ex.1029 / Page 34 of 347
AMAZON.COM, INC.

Claim 1	Accused Instrumentality(ies)
speaker, subsequently to the input speech;	For example, Amazon receives wake word and a subsequent speech, through a combination of hardware, processing, and software functions both on devices and on the cloud server.[19]  This is further illustrated below.[20]

[19] *See* KAIFI0007814 *(The Science Behind Alexa's Wake Word Detection*, https://www.youtube.com/watch?v=kYrw2cC9dvs).
[20] *See* KAIFI0007816 *(How Alexa Uses Automatic Speech Recognition*, https://www.youtube.com/watch?v=8drK6Oyv1a0&t=296s).

Ex.1029 / Page 35 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	 For example, Amazon describes a voice request.[21]

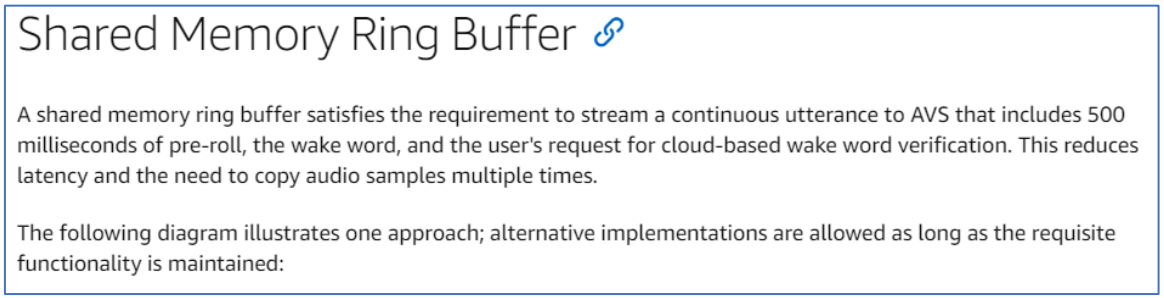
[21] *See* KAIFI0002023 *(Follow the journey of a voice request*, https://www.amazon.com/b/?node=23608618011).

Ex.1029 / Page 36 of 347
AMAZON.COM, INC.

Ex.1029 / Page 37 of 347
AMAZON.COM, INC.

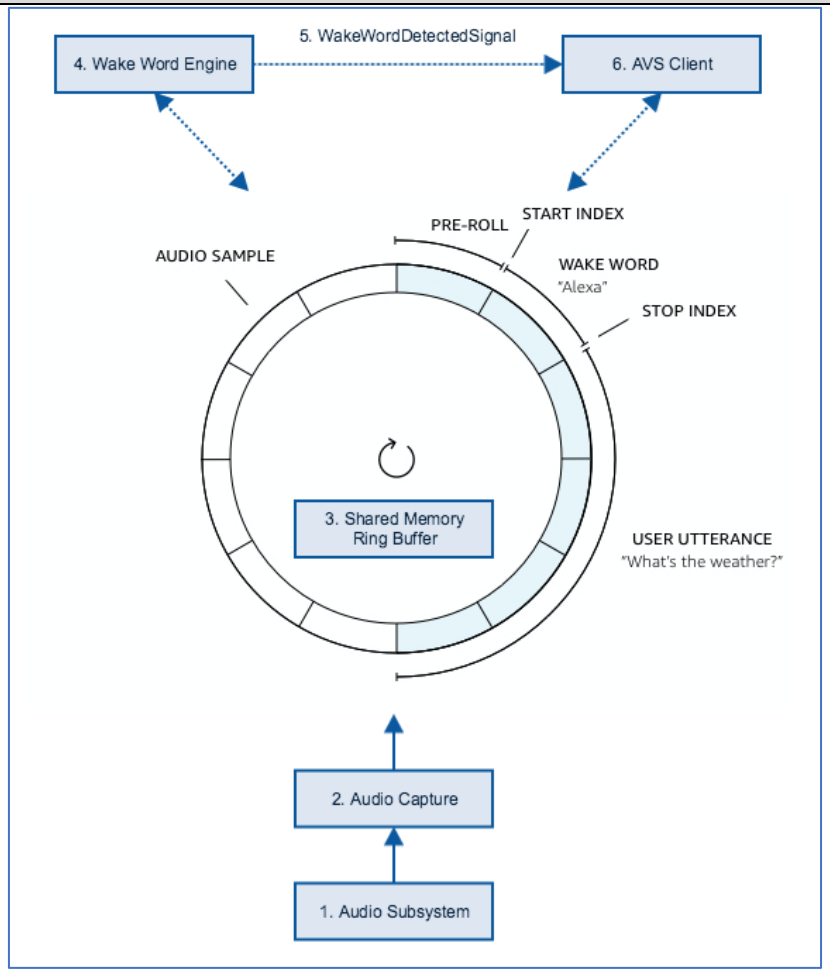
Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	For example, Amazon provides a shared memory ring buffer.[22]  Shared Memory Ring Buffer A shared memory ring buffer satisfies the requirement to stream a continuous utterance to AVS that includes 500 milliseconds of pre-roll, the wake word, and the user's request for cloud-based wake word verification. This reduces latency and the need to copy audio samples multiple times. The following diagram illustrates one approach; alternative implementations are allowed as long as the requisite functionality is maintained:

[22] *See* KAIFI0007134 *(Wake Word Verification Requirements*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/implement-ww-verification.html).

Ex.1029 / Page 38 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	

Ex.1029 / Page 39 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	1. **Audio Subsystem** - Middleware, such as Advanced Linux Sound Architecture (ALSA), that provides APIs to open a stream from a recording device, and receive an audio sample stream. 2. **Audio Capture** - A process that opens a recording device, and writes audio samples into the shared memory ring buffer. 3. **Shared Memory Ring Buffer** - A shared memory ring buffer object that is thread safe, and accessible by multiple processes. Access to this memory block should support at least one writer and two readers simultaneously. You may choose to implement these executables as a shared library. • **Write API** - Used by the recording process to write audio samples to the shared memory ring buffer. Write samples should move data so that the oldest audio samples are overwritten. This example shows a typical API signature: `WriteSamples(BufferIn, BufferInLength)`. • **Read API** - Used by the wake word engine and the AVS client to read from specific locations in the shared memory ring buffer. A typical implementation represents location as a 64-bit integer. For example, reading 200 samples from the location 100 is reading the 100th to 300th samples that were written to the shared memory ring buffer. This is an example of a typical API signature: `ReadSamples(BufferIn, BufferInLength)`. **Note**: Reading from an old location that was overwritten with new audio samples will result in an unintended sample read. 4. **Wake Word Engine** - The component that reads the audio samples written to the shared memory ring buffer and analyzes the audio samples for occurrences of the wake word. The procedure of reading and processing audio samples occurs in a loop. For example, the loop reads N samples starting at location `0`; then analyzes the samples for an occurrence of the wake word. In the next iteration, it reads and analyzes the next N samples. When an occurrence of the wake word is detected, the wake word engine must identify two locations: 1) the start of the wake word, and 2) the end of the wake word. These values are stored as `startIndexInSamples` and `endIndexInSamples`. 5. **WakeWordDetectedSignal** - A signal sent from the wake word engine to the AVS client when an occurrence of the wake word is detected. This signal includes two locations, the `startIndexInSamples` and `endIndexInSamples`. 6. **AVS Client** - The client receives the `WakeWordDetectedSignal`, extracts the `startIndexInSamples`, then reads and streams audio samples to AVS including 500 milliseconds of pre-roll. • The pre-roll is the number of samples corresponding to 500 milliseconds. For example, if the audio samples were recorded at a rate of 16 kilohertz (kHz), then pre-roll is 8,000 samples.

Ex.1029 / Page 40 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	For example, according to Amazon, its speech recognizer captures and interacts with user speech.[23]

> The `SpeechRecognizer` interface is the core interface of the Alexa Voice Service (AVS) and exposes directives and events for capturing and interacting with user speech. This page discusses the concepts and process flows for the `SpeechRecognizer` interface.

SpeechRecognizer functionality

Every user utterance leverages `SpeechRecognizer`, including the following interaction types between an Alexa Built-in device and AVS:

- **User Speech Capture**: Captures user speech from an Alexa Built-in device.
- **User Speech Prompting**: Prompts a user for more speech input when needed by Alexa to deliver an appropriate response.
- **Interaction Initiation Communication**: Enables a device to inform AVS of how a user initiated an Alexa interaction, such as press-and-hold, tap-and-release, or voice-initiated/wake word enabled. See Device Form Factor and Alexa Interaction.
- **ASR Profile Selection**: SpeechRecognizers chooses the appropriate Automatic Speech Recognition (ASR) profile for your product, which allows Alexa to understand user speech and respond with precision. See Automatic Speech Recognition (ASR) profile

[23] *See* KAIFI0007093 *(Understand the AVS SpeechRecognizer*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/avs-speechrecognizer-concepts.html).

Ex.1029 / Page 41 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	• **RECOGNIZING:** When a user begins interacting with your client, specifically when the client streams captured audio to AVS, `SpeechRecognizer` should transition from the idle state to the `Recognize` state. `SpeechRecognizer` should remain in the `Recognize` until the client stops recording speech or finishes streaming, at which point your `SpeechRecognizer` component should transition from the `Recognize` state to the "busy" state. • **BUSY:** While processing the speech request, `SpeechRecognizer` should be in the "busy" state. You cannot start another speech request until `SpeechRecognizer` transitions out of the busy state. From the *busy* state, `SpeechRecognizer` transitions to the idle state if Alexa processes and completes the request or to the `ExpectSpeech` state, if Alexa requires more speech input from the user. • **EXPECTING SPEECH:** `SpeechRecognizer` should be in the `ExpectSpeech` state when Alexa requires more audio input from a user. From `ExpectSpeech`, `SpeechRecognizer` should transition to the `Recognize` when a user interaction occurs, or the interaction is automatically started on behalf of the user. It should transition to the idle state if Alexa detects no user interaction within the specified timeout window. For example, Amazon uses wake word, keyword spotting, cloud verification and encryption, automatic speech recognition (ASR), natural language understanding (NLU), text-to-speech (TTS), and data storage technologies. [24]

[24] *See* KAIFI0002023 *(Follow the journey of a voice request*, https://www.amazon.com/b/?node=23608618011).

Ex.1029 / Page 42 of 347
AMAZON.COM, INC.

The wake word

Think of the wake word as a verbal cue that makes things happen. In the classic tale from The Arabian Nights, an invisible door into a mountain can only be opened by saying the phrase "Open Sesame." Use the phrase, the door opens. Don't use the phrase, the door stays shut. When you have a request for Alexa, you first need to say your chosen wake word, which by default is "Alexa." Only after your Echo device detects the wake word is Alexa listening to your requests.

When it comes to privacy, there should be no surprises. You'll always be able to tell when Alexa is listening to your request because a light indicator will appear on your Echo device or an audible tone will sound. Think of the "On the Air" signs that light up in television studios during a broadcast. These indicators notify you that your device has detected the wake word and Alexa is now processing your request.

Keyword spotting

Echo devices use built-in technology called "keyword spotting" that matches spoken audio to the acoustic patterns of the wake word. Simply put, Echo devices are designed by default to detect only the sound waves of your chosen wake word, and everything else is ignored. Like water through a strainer, all other audio (people talking, faucets running, birds chirping) passes through the device until the wake word is "caught" and sent to Amazon's secure cloud, where meaning is assigned to your request. You can choose to opt-in to features like Alexa Emergency Assist that allow your device to detect more than your chosen wake word—for example, the sound of smoke alarms—but you would need to update your settings to do so.

Cloud verification and encryption

When your Echo device detects the wake word, it sends your request to Amazon's secure cloud where audio is reanalyzed to verify the wake word was spoken. If this cloud software verification is unable to confirm the wake word was spoken, the Alexa system stops processing the audio. If the wake word is verified (or Alexa is activated using the action button), your request will be processed using several sophisticated algorithms that allow Alexa to respond appropriately. All of your interactions with Alexa are encrypted in transit throughout this process.

Automatic Speech Recognition (ASR)

After wake word verification, your voice request undergoes Automatic Speech Recognition (ASR) which transcribes audio into text. Using our example request ("Alexa, what's the weather in Seattle?"), transcriptions could include one or several of the following:

"what is the weather in seattle" "watt is the weather in seattle" "what is the whether in seattle"

Alexa then chooses to act on the transcription that is most likely to be correct and stores it in Amazon's secure cloud. You can view and delete it at any time at Alexa Privacy Settings or in the Alexa app (Settings > Alexa Privacy > Review Voice History). Voice recordings and transcriptions may be used to improve the ASR process. Learn how data helps Alexa improve.

Natural Language Understanding

After ASR comes Natural Language Understanding (NLU). NLU assigns meaning to the transcription producing an "intent" which is an instruction to the Alexa system to tell it what to respond to. In our example, the intent is identified as "weather" and sent to the designated data source where information is pulled for location (Seattle) and time (today). Intents may also be used to improve the NLU process.

Text-to-Speech (TTS)

Once the necessary information is gathered, your request uses text-to-speech (TTS) technologies to generate an audio file that becomes Alexa's response. The text from Alexa's response is stored so that you can review past answers at Alexa Privacy Settings or in the Alexa app. These responses may also be reviewed by Amazon to ensure the device is providing the most relevant answers to you, and to make sure the TTS system is translating text to speech to the best of its ability.

Data storage

You have control over your Alexa experience. We make it easy for you to update your privacy settings. You can always view and manage your Alexa interactions at Alexa Privacy Settings or in the Alexa app (Settings > Alexa Privacy). If you're enrolled in voice ID, you can ask, "Alexa, what are my privacy settings?" to review how long your voice recordings are currently saved and whether or not they're used to develop new features or manually reviewed to help improve Alexa. You can also say, "Alexa, update my privacy settings" at any time to make updates.

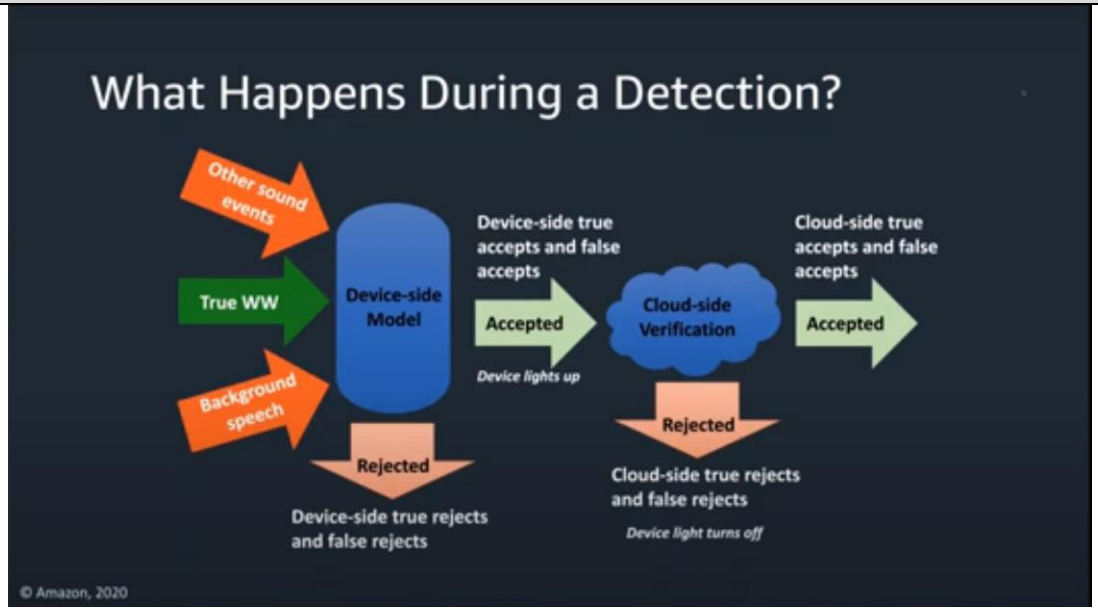
Ex.1029 / Page 43 of 347
AMAZON.COM, INC.

Claim 1	Accused Instrumentality(ies)
	This claim element is literally present, as described herein, but is also satisfied under the doctrine of equivalents because the structure and functionality provided by Amazon is being used to satisfy this claim limitation and therefore there is no vitiation of this element, and the Accused Instrumentalities incorporate structure and functionality that is not substantially different from the requirements of this limitation because they achieve substantially or exactly the same function, in substantially or exactly the same way as set forth in the claim element (as described herein), and achieves substantially or exactly the same result.
comparing, by the preset call command recognition unit, the input speech with a preset call command;	Amazon compares, through the preset call command recognition unit, the input speech with a preset call command. For example, Amazon compares, through a combination of hardware, processing, and software functions both on devices and on the cloud server, an input speech with a wake word.[25]

[25] *See* KAIFI0007814 *(The Science Behind Alexa's Wake Word Detection*, https://www.youtube.com/watch?v=kYrw2cC9dvs).

Ex.1029 / Page 44 of 347
AMAZON.COM, INC.

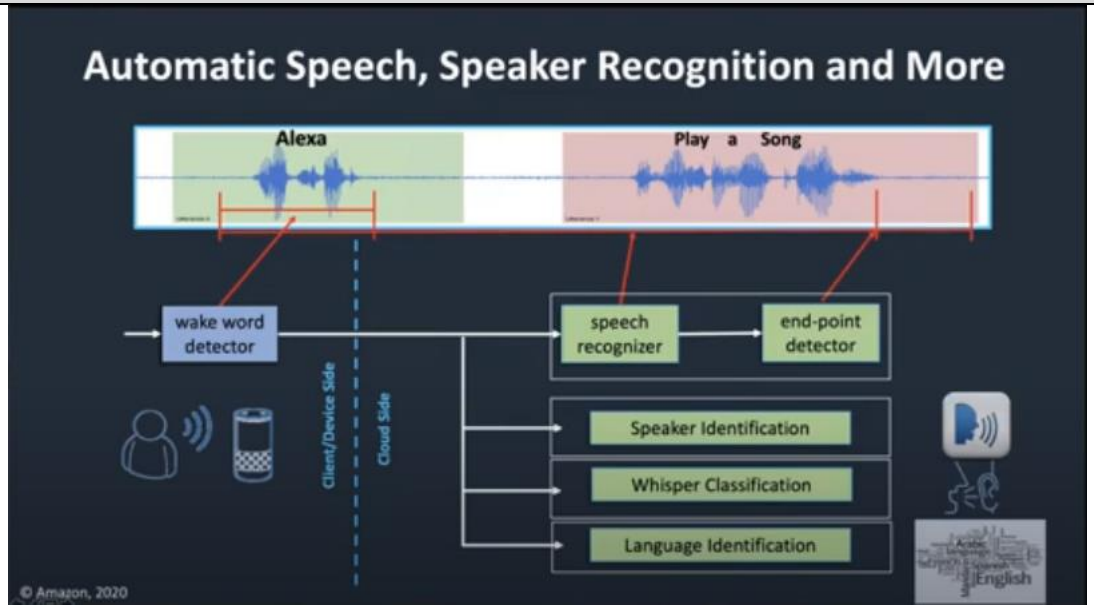
Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	 This is further illustrated below.[26]

[26] *See* KAIFI0007816 *(How Alexa Uses Automatic Speech Recognition*, https://www.youtube.com/watch?v=8drK6Oyv1a0&t=296s).

Ex.1029 / Page 45 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	 For example, Amazon uses keyword spotting.[27] Echo devices use built-in technology called "keyword spotting" that matches spoken audio to the acoustic patterns of the wake word. Simply put, Echo devices are designed by default to detect only the sound waves of your chosen wake word, and everything else is ignored.

[27] *See* KAIFI0002092 *(How Alexa works: Wake word*, https://www.amazon.com/b?ie=UTF8&node=23608571011).

Ex.1029 / Page 46 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	For example, Amazon uses wake words.[28] Wake words The list of wake words informs Alexa of the possible valid wake words that a device might be set to listen for through the `SetWakeWords`, `WakeWordsReport`, and `WakeWordsChanged` messages. Currently, the only wake word available for Alexa Built-in devices is `ALEXA`, which applies to every possible locale for a device. Therefore, specify `ALEXA` in the global `DEFAULT` scope. For example, Amazon provides a wake word verification engine.[29] 4. **Wake Word Engine** - The component that reads the audio samples written to the shared memory ring buffer and analyzes the audio samples for occurrences of the wake word. The procedure of reading and processing audio samples occurs in a loop. For example, the loop reads N samples starting at location `0`; then analyzes the samples for an occurrence of the wake word. In the next iteration, it reads and analyzes the next N samples. When an occurrence of the wake word is detected, the wake word engine must identify two locations: 1) the start of the wake word, and 2) the end of the wake word. These values are stored as `startIndexInSamples` and `endIndexInSamples`. According to Amazon, the speech recognizer transitions are illustrated in the below diagram.[30]

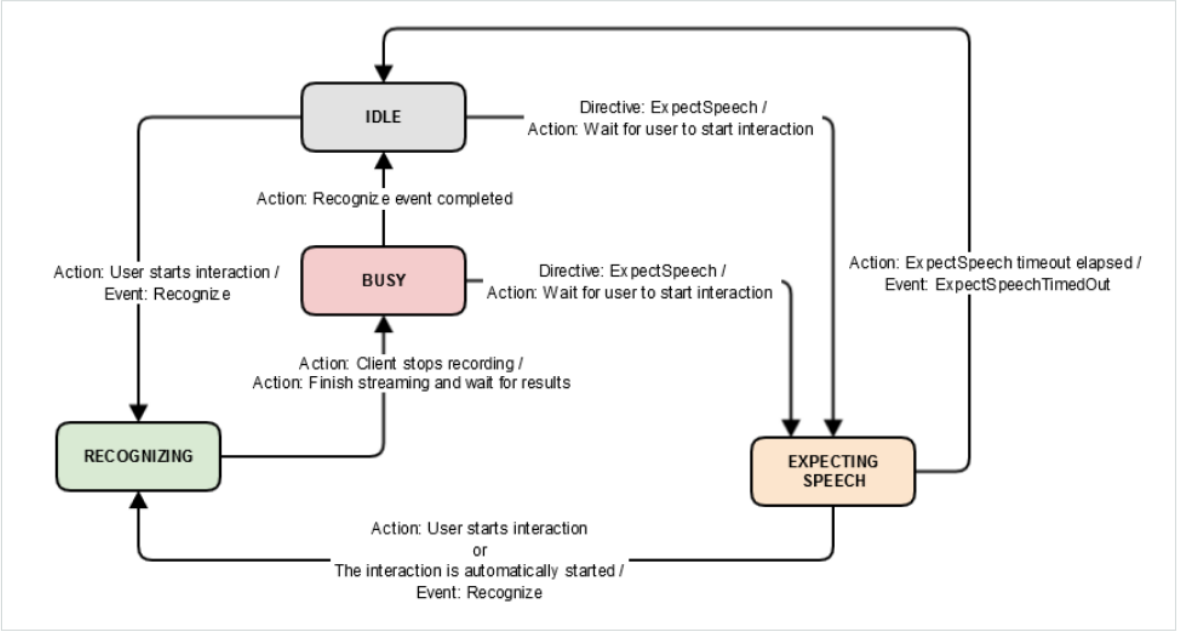
[28] *See* KAIFI0004342 *(How Our Scientists are Making Alexa Smarter* (Mar. 29, 2018), https://www.aboutamazon.com/news/devices/how-our-scientists-are-making-alexa-smarter).

[29] *See* KAIFI0007141 *(Wake Word Verification Requirements*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/implement-ww-verification.html).

[30] *See* KAIFI0007093 *(Understand the AVS SpeechRecognizer*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-

Ex.1029 / Page 47 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	The following diagram illustrates the expected transitions among the four `SpeechRecognizer` states:  Amazon also provides wake word verification.[31]

service/avs-speechrecognizer-concepts.html).

[31] *See* KAIFI0002012 *(Enable Wake Word Verification,* https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/enable-cloud-based-wake-word-verification.html).

Ex.1029 / Page 48 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)

Cloud-based wake word verification improves wake word accuracy for Alexa Built-in products by reducing false wakes caused by words that sound similar to the wake word. Examples of words that might cause a false wake for "Alexa" include "Alex", "election", "Alexis". Cloud-based wake word verification also detects media mentions of the wake word, such as, the mention of "Alexa" in an Amazon commercial.

The wake word engine on a device performs the initial wake word detection, and then the cloud verifies the wake word. If the cloud detects a false wake, the Alexa Voice Service (AVS) sends a `StopCapture` directive to the device in the downchannel that instructs it to close the audio stream, and if applicable, to turn off the blue LEDs to indicate that Alexa has stopped listening.

Voice-initiated devices begin to stream user speech to AVS when the wake word engine detects a spoken wake word, such as "Alexa." The stream closes when the user stops speaking or when AVS identifies a user intent, and AVS returns a `StopCapture` directive to the device. Cloud-based wake word verification has the following requirements:

- **Wake word** – Include the wake word in the stream so that AVS can perform cloud-based wake word verification, which reduces false wakes. If AVS can't detect the wake word detected during cloud-based wake word verification, AVS discards the utterance.
- **500 milliseconds of pre-roll** – Pre-roll is the audio captured before AVS detects the wake word and helps calibrate the ambient noise level of the recording to enhance speech recognition.
- **User speech** – Any user speech that the device captures until receiving a `StopCapture` directive. This allows AVS to verify the wake word included in the stream, reducing the number of erroneous responses due to false wakes.

To learn how to implement a shared memory ring buffer for writing and reading audio samples and to include the start and stop indices for wake word detection in each `Recognize` event sent to AVS, see the Requirements for Cloud-Based Wake Word Verification.

For example, Amazon uses wake word, keyword spotting, cloud verification and encryption, automatic speech recognition (ASR), natural language understanding (NLU), text-to-speech (TTS), and data storage technologies. [32]

[32] *See* KAIFI0002023 *(Follow the journey of a voice request*, https://www.amazon.com/b/?node=23608618011).

Ex.1029 / Page 49 of 347
AMAZON.COM, INC.

The wake word

Think of the wake word as a verbal cue that makes things happen. In the classic tale from The Arabian Nights, an invisible door into a mountain can only be opened by saying the phrase "Open Sesame." Use the phrase, the door opens. Don't use the phrase, the door stays shut. When you have a request for Alexa, you first need to say your chosen wake word, which by default is "Alexa." Only after your Echo device detects the wake word is Alexa listening to your requests.

When it comes to privacy, there should be no surprises. You'll always be able to tell when Alexa is listening to your request because a light indicator will appear on your Echo device or an audible tone will sound. Think of the "On the Air" signs that light up in television studios during a broadcast. These indicators notify you that your device has detected the wake word and Alexa is now processing your request.

Keyword spotting

Echo devices use built-in technology called "keyword spotting" that matches spoken audio to the acoustic patterns of the wake word. Simply put, Echo devices are designed by default to detect only the sound waves of your chosen wake word, and everything else is ignored. Like water through a strainer, all other audio (people talking, faucets running, birds chirping) passes through the device until the wake word is "caught" and sent to Amazon's secure cloud, where meaning is assigned to your request. You can choose to opt-in to features like Alexa Emergency Assist that allow your device to detect more than your chosen wake word—for example, the sound of smoke alarms—but you would need to update your settings to do so.

Cloud verification and encryption

When your Echo device detects the wake word, it sends your request to Amazon's secure cloud where audio is reanalyzed to verify the wake word was spoken. If this cloud software verification is unable to confirm the wake word was spoken, the Alexa system stops processing the audio. If the wake word is verified (or Alexa is activated using the action button), your request will be processed using several sophisticated algorithms that allow Alexa to respond appropriately. All of your interactions with Alexa are encrypted in transit throughout this process.

Automatic Speech Recognition (ASR)

After wake word verification, your voice request undergoes Automatic Speech Recognition (ASR) which transcribes audio into text. Using our example request ("Alexa, what's the weather in Seattle?"), transcriptions could include one or several of the following:

"what is the weather in seattle" "watt is the weather in seattle" "what is the whether in seattle"

Alexa then chooses to act on the transcription that is most likely to be correct and stores it in Amazon's secure cloud. You can view and delete it at any time at Alexa Privacy Settings or in the Alexa app (Settings > Alexa Privacy > Review Voice History). Voice recordings and transcriptions may be used to improve the ASR process. Learn how data helps Alexa improve.

Natural Language Understanding

After ASR comes Natural Language Understanding (NLU). NLU assigns meaning to the transcription producing an "intent" which is an instruction to the Alexa system to tell it what to respond to. In our example, the intent is identified as "weather" and sent to the designated data source where information is pulled for location (Seattle) and time (today). Intents may also be used to improve the NLU process.

Text-to-Speech (TTS)

Once the necessary information is gathered, your request uses text-to-speech (TTS) technologies to generate an audio file that becomes Alexa's response. The text from Alexa's response is stored so that you can review past answers at Alexa Privacy Settings or in the Alexa app. These responses may also be reviewed by Amazon to ensure the device is providing the most relevant answers to you, and to make sure the TTS system is translating text to speech to the best of its ability.

Data storage

You have control over your Alexa experience. We make it easy for you to update your privacy settings. You can always view and manage your Alexa interactions at Alexa Privacy Settings or in the Alexa app (Settings > Alexa Privacy). If you're enrolled in voice ID, you can ask, "Alexa, what are my privacy settings?" to review how long your voice recordings are currently saved and whether or not they're used to develop new features or manually reviewed to help improve Alexa. You can also say, "Alexa, update my privacy settings" at any time to make updates.

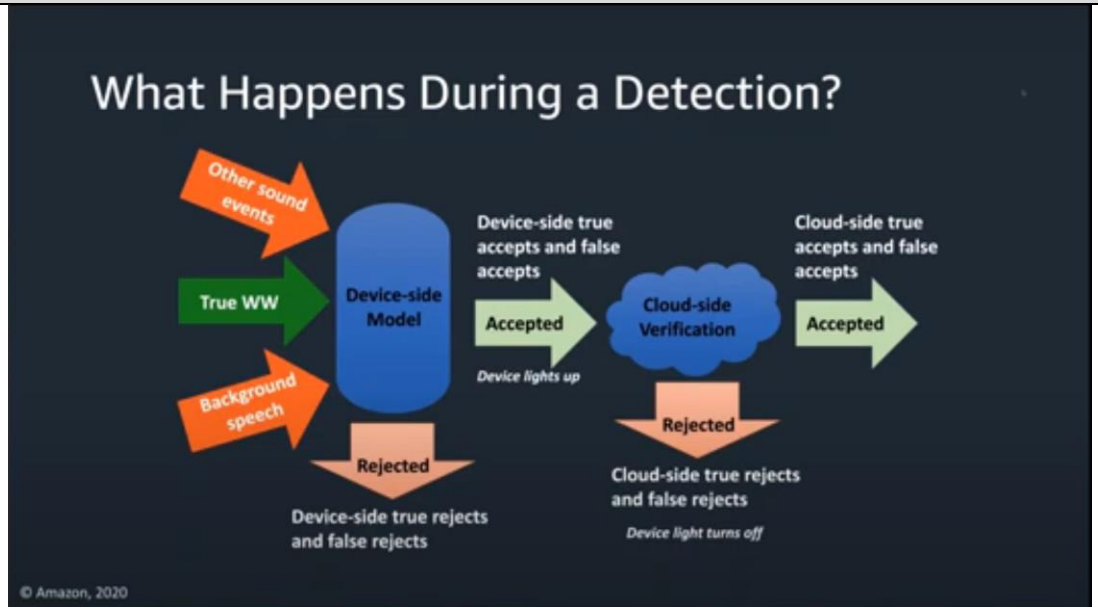
Ex.1029 / Page 50 of 347
AMAZON.COM, INC.

Claim 1	Accused Instrumentality(ies)
	This claim element is literally present, as described herein, but is also satisfied under the doctrine of equivalents because the structure and functionality provided by Amazon is being used to satisfy this claim limitation and therefore there is no vitiation of this element, and the Accused Instrumentalities incorporate structure and functionality that is not substantially different from the requirements of this limitation because they achieve substantially or exactly the same function, in substantially or exactly the same way as set forth in the claim element (as described herein), and achieves substantially or exactly the same result.
determining, by the preset call command recognition unit, whether the input speech is recognized as the preset call command according to a result of the comparison; and	Amazon determines, through the preset call command recognition unit, whether the input speech is recognized as the preset call command according to a result of the comparison. For example, Amazon determines, through a combination of hardware, processing, and software functions both on devices and on the cloud server, whether the input speech is a true accept or a true accept for a wake word.[33]

[33] *See* KAIFI0007814 *(The Science Behind Alexa's Wake Word Detection*, https://www.youtube.com/watch?v=kYrw2cC9dvs).

Ex.1029 / Page 51 of 347
AMAZON.COM, INC.

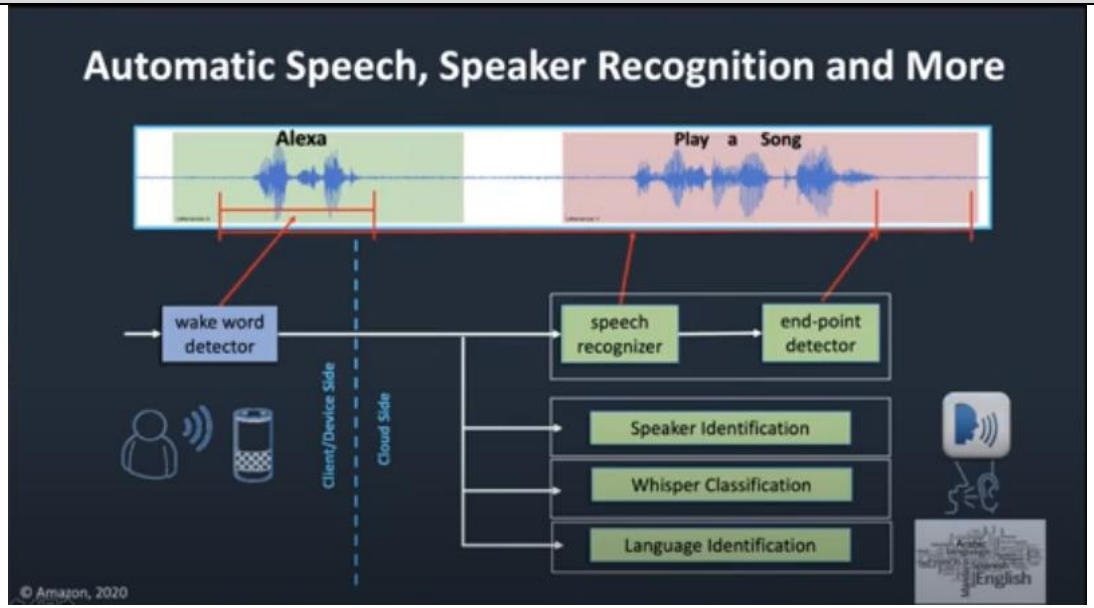
Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	 This is further illustrated below.[34]

[34] *See* KAIFI0007816 *(How Alexa Uses Automatic Speech Recognition*, https://www.youtube.com/watch?v=8drK6Oyv1a0&t=296s).

Ex.1029 / Page 52 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	 For example, Amazon describes a voice request.[35]

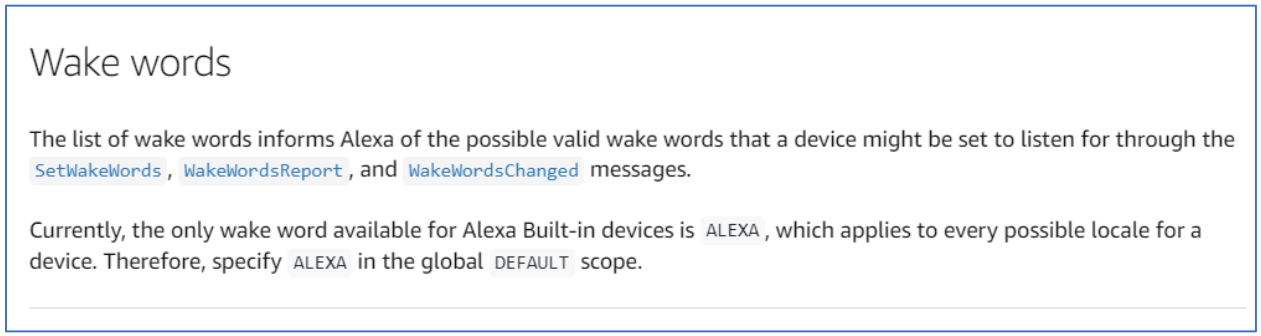
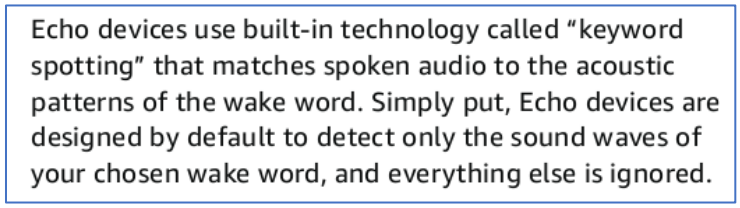
[35] *See* KAIFI0002023 *(Follow the journey of a voice request*, https://www.amazon.com/b/?node=23608618011).

Ex.1029 / Page 53 of 347
AMAZON.COM, INC.

Ex.1029 / Page 54 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	For example, Amazon uses wake words.[36]  Wake words The list of wake words informs Alexa of the possible valid wake words that a device might be set to listen for through the `SetWakeWords`, `WakeWordsReport`, and `WakeWordsChanged` messages. Currently, the only wake word available for Alexa Built-in devices is `ALEXA`, which applies to every possible locale for a device. Therefore, specify `ALEXA` in the global `DEFAULT` scope. For example, Amazon uses keyword spotting.[37]  Echo devices use built-in technology called "keyword spotting" that matches spoken audio to the acoustic patterns of the wake word. Simply put, Echo devices are designed by default to detect only the sound waves of your chosen wake word, and everything else is ignored. For example, Amazon uses speech recognition.[38]

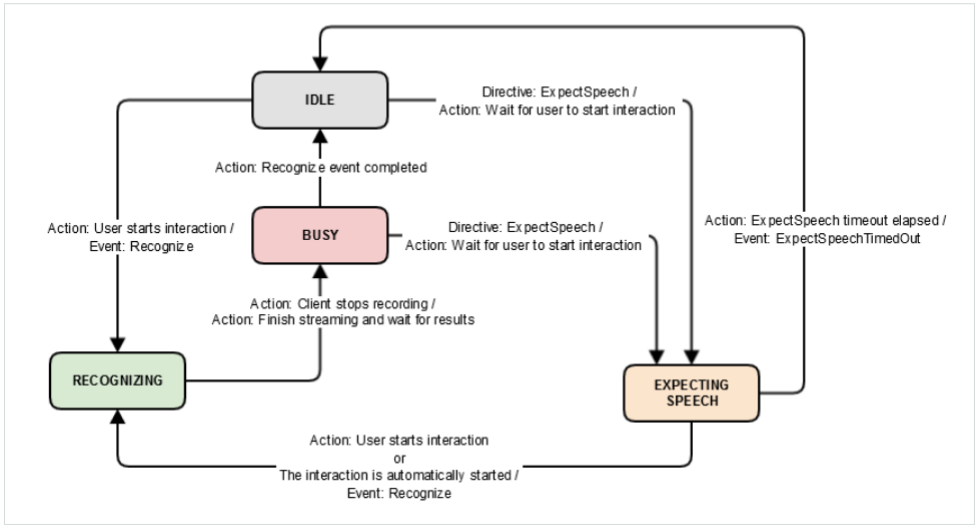
[36] *See* KAIFI0004342 *(How Our Scientists are Making Alexa Smarter* (Mar. 29, 2018), https://www.aboutamazon.com/news/devices/how-our-scientists-are-making-alexa-smarter).

[37] *See* KAIFI0002092 *(How Alexa works: Wake word*, https://www.amazon.com/b?ie=UTF8&node=23608571011).

[38] *See* KAIFI0002023 *(Follow the journey of a voice request*, https://www.amazon.com/b/?node=23608618011).

Ex.1029 / Page 55 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	The following diagram illustrates the expected transitions among the four `SpeechRecognizer` states:  For example, Amazon provides wake word verification.[39] Cloud-based wake word verification improves wake word accuracy for Alexa Built-in products by reducing false wakes caused by words that sound similar to the wake word. Examples of words that might cause a false wake for "Alexa" include "Alex", "election", "Alexis". Cloud-based wake word verification also detects media mentions of the wake word, such as, the mention of "Alexa" in an Amazon commercial.

[39] *See* KAIFI0002012 *(Enable Wake Word Verification,* https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/enable-cloud-based-wake-word-verification.html).

Ex.1029 / Page 56 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	The wake word engine on a device performs the initial wake word detection, and then the cloud verifies the wake word. If the cloud detects a false wake, the Alexa Voice Service (AVS) sends a `StopCapture` directive to the device in the downchannel that instructs it to close the audio stream, and if applicable, to turn off the blue LEDs to indicate that Alexa has stopped listening. Voice-initiated devices begin to stream user speech to AVS when the wake word engine detects a spoken wake word, such as "Alexa." The stream closes when the user stops speaking or when AVS identifies a user intent, and AVS returns a `StopCapture` directive to the device. Cloud-based wake word verification has the following requirements: • **Wake word** – Include the wake word in the stream so that AVS can perform cloud-based wake word verification, which reduces false wakes. If AVS can't detect the wake word detected during cloud-based wake word verification, AVS discards the utterance. • **500 milliseconds of pre-roll** – Pre-roll is the audio captured before AVS detects the wake word and helps calibrate the ambient noise level of the recording to enhance speech recognition. • **User speech** – Any user speech that the device captures until receiving a `StopCapture` directive. This allows AVS to verify the wake word included in the stream, reducing the number of erroneous responses due to false wakes. To learn how to implement a shared memory ring buffer for writing and reading audio samples and to include the start and stop indices for wake word detection in each `Recognize` event sent to AVS, see the Requirements for Cloud-Based Wake Word Verification. For example, Amazon uses wake word, keyword spotting, cloud verification and encryption, automatic speech recognition (ASR), natural language understanding (NLU), text-to-speech (TTS), and data storage technologies.[40]

[40] *See* KAIFI0002023 *(Follow the journey of a voice request*, https://www.amazon.com/b/?node=23608618011).

Ex.1029 / Page 57 of 347
AMAZON.COM, INC.

The wake word

Think of the wake word as a verbal cue that makes things happen. In the classic tale from The Arabian Nights, an invisible door into a mountain can only be opened by saying the phrase "Open Sesame." Use the phrase, the door opens. Don't use the phrase, the door stays shut. When you have a request for Alexa, you first need to say your chosen wake word, which by default is "Alexa." Only after your Echo device detects the wake word is Alexa listening to your requests.

When it comes to privacy, there should be no surprises. You'll always be able to tell when Alexa is listening to your request because a light indicator will appear on your Echo device or an audible tone will sound. Think of the "On the Air" signs that light up in television studios during a broadcast. These indicators notify you that your device has detected the wake word and Alexa is now processing your request.

Keyword spotting

Echo devices use built-in technology called "keyword spotting" that matches spoken audio to the acoustic patterns of the wake word. Simply put, Echo devices are designed by default to detect only the sound waves of your chosen wake word, and everything else is ignored. Like water through a strainer, all other audio (people talking, faucets running, birds chirping) passes through the device until the wake word is "caught" and sent to Amazon's secure cloud, where meaning is assigned to your request. You can choose to opt-in to features like Alexa Emergency Assist that allow your device to detect more than your chosen wake word—for example, the sound of smoke alarms—but you would need to update your settings to do so.

Cloud verification and encryption

When your Echo device detects the wake word, it sends your request to Amazon's secure cloud where audio is reanalyzed to verify the wake word was spoken. If this cloud software verification is unable to confirm the wake word was spoken, the Alexa system stops processing the audio. If the wake word is verified (or Alexa is activated using the action button), your request will be processed using several sophisticated algorithms that allow Alexa to respond appropriately. All of your interactions with Alexa are encrypted in transit throughout this process.

Automatic Speech Recognition (ASR)

After wake word verification, your voice request undergoes Automatic Speech Recognition (ASR) which transcribes audio into text. Using our example request ("Alexa, what's the weather in Seattle?"), transcriptions could include one or several of the following:

"what is the weather in seattle" "watt is the weather in seattle" "what is the whether in seattle"

Alexa then chooses to act on the transcription that is most likely to be correct and stores it in Amazon's secure cloud. You can view and delete it at any time at Alexa Privacy Settings or in the Alexa app (Settings > Alexa Privacy > Review Voice History). Voice recordings and transcriptions may be used to improve the ASR process. Learn how data helps Alexa improve.

Natural Language Understanding

After ASR comes Natural Language Understanding (NLU). NLU assigns meaning to the transcription producing an "intent" which is an instruction to the Alexa system to tell it what to respond to. In our example, the intent is identified as "weather" and sent to the designated data source where information is pulled for location (Seattle) and time (today). Intents may also be used to improve the NLU process.

Text-to-Speech (TTS)

Once the necessary information is gathered, your request uses text-to-speech (TTS) technologies to generate an audio file that becomes Alexa's response. The text from Alexa's response is stored so that you can review past answers at Alexa Privacy Settings or in the Alexa app. These responses may also be reviewed by Amazon to ensure the device is providing the most relevant answers to you, and to make sure the TTS system is translating text to speech to the best of its ability.

Data storage

You have control over your Alexa experience. We make it easy for you to update your privacy settings. You can always view and manage your Alexa interactions at Alexa Privacy Settings or in the Alexa app (Settings > Alexa Privacy). If you're enrolled in voice ID, you can ask, "Alexa, what are my privacy settings?" to review how long your voice recordings are currently saved and whether or not they're used to develop new features or manually reviewed to help improve Alexa. You can also say, "Alexa, update my privacy settings" at any time to make updates.

Ex.1029 / Page 58 of 347
AMAZON.COM, INC.

Claim 1	Accused Instrumentality(ies)
	This claim element is literally present, as described herein, but is also satisfied under the doctrine of equivalents because the structure and functionality provided by Amazon is being used to satisfy this claim limitation and therefore there is no vitiation of this element, and the Accused Instrumentalities incorporate structure and functionality that is not substantially different from the requirements of this limitation because they achieve substantially or exactly the same function, in substantially or exactly the same way as set forth in the claim element (as described herein), and achieves substantially or exactly the same result.
based on determining that the input speech is recognized as the preset call command, activating a continuous speech recognition unit configured to determine whether the subsequent input speech is recognized as an actual command, in order to reduce computational load on the device; and	Amazon, based on determining that the input speech is recognized as the preset call command, activates a continuous speech recognition unit configured to determine whether the subsequent input speech is recognized as an actual command, in order to reduce computational load on the device. For example, Amazon describes a voice request.[41]

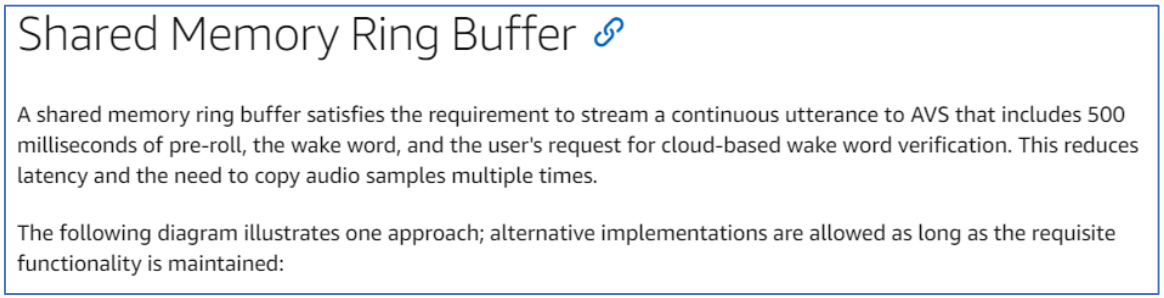
[41] *See* KAIFI0002023 *(Follow the journey of a voice request*, https://www.amazon.com/b/?node=23608618011).

Ex.1029 / Page 59 of 347
AMAZON.COM, INC.

Ex.1029 / Page 60 of 347
AMAZON.COM, INC.

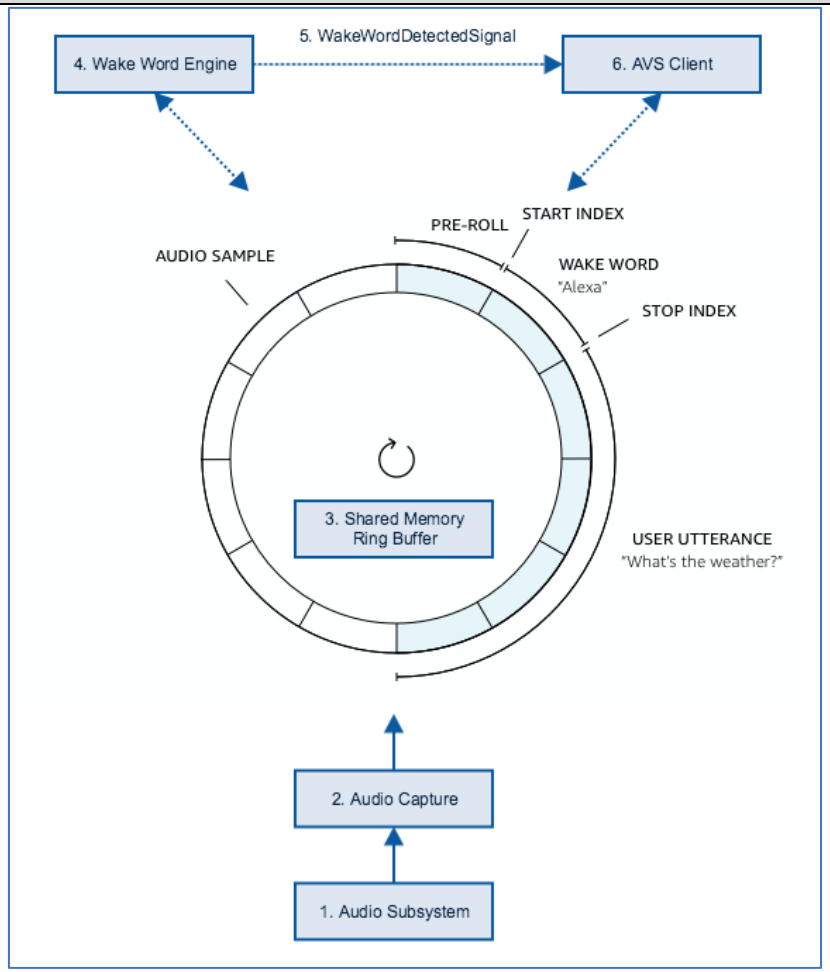
Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	For example, Amazon provides a shared memory ring buffer.[42]  Shared Memory Ring Buffer A shared memory ring buffer satisfies the requirement to stream a continuous utterance to AVS that includes 500 milliseconds of pre-roll, the wake word, and the user's request for cloud-based wake word verification. This reduces latency and the need to copy audio samples multiple times. The following diagram illustrates one approach; alternative implementations are allowed as long as the requisite functionality is maintained:

[42] *See* KAIFI0007141 *(Wake Word Verification Requirements*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/implement-ww-verification.html).

Ex.1029 / Page 61 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	

Ex.1029 / Page 62 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	1. **Audio Subsystem** - Middleware, such as Advanced Linux Sound Architecture (ALSA), that provides APIs to open a stream from a recording device, and receive an audio sample stream. 2. **Audio Capture** - A process that opens a recording device, and writes audio samples into the shared memory ring buffer. 3. **Shared Memory Ring Buffer** - A shared memory ring buffer object that is thread safe, and accessible by multiple processes. Access to this memory block should support at least one writer and two readers simultaneously. You may choose to implement these executables as a shared library. • **Write API** - Used by the recording process to write audio samples to the shared memory ring buffer. Write samples should move data so that the oldest audio samples are overwritten. This example shows a typical API signature: `WriteSamples(BufferIn, BufferInLength)`. • **Read API** - Used by the wake word engine and the AVS client to read from specific locations in the shared memory ring buffer. A typical implementation represents location as a 64-bit integer. For example, reading 200 samples from the location 100 is reading the 100th to 300th samples that were written to the shared memory ring buffer. This is an example of a typical API signature: `ReadSamples(BufferIn, BufferInLength)`. **Note**: Reading from an old location that was overwritten with new audio samples will result in an unintended sample read. 4. **Wake Word Engine** - The component that reads the audio samples written to the shared memory ring buffer and analyzes the audio samples for occurrences of the wake word. The procedure of reading and processing audio samples occurs in a loop. For example, the loop reads N samples starting at location `0`; then analyzes the samples for an occurrence of the wake word. In the next iteration, it reads and analyzes the next N samples. When an occurrence of the wake word is detected, the wake word engine must identify two locations: 1) the start of the wake word, and 2) the end of the wake word. These values are stored as `startIndexInSamples` and `endIndexInSamples`. 5. **WakeWordDetectedSignal** - A signal sent from the wake word engine to the AVS client when an occurrence of the wake word is detected. This signal includes two locations, the `startIndexInSamples` and `endIndexInSamples`. 6. **AVS Client** - The client receives the `WakeWordDetectedSignal`, extracts the `startIndexInSamples`, then reads and streams audio samples to AVS including 500 milliseconds of pre-roll. • The pre-roll is the number of samples corresponding to 500 milliseconds. For example, if the audio samples were recorded at a rate of 16 kilohertz (kHz), then pre-roll is 8,000 samples.

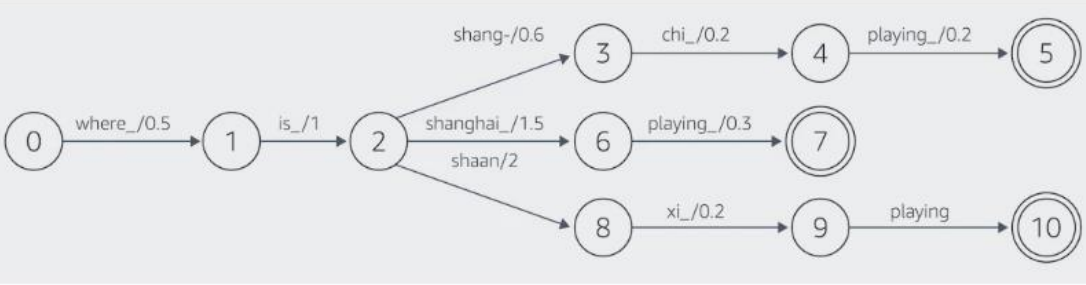
Ex.1029 / Page 63 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	For example, Amazon describes its ASR model.[43]

System architecture

Our on-device ASR model takes in an acoustic speech signal and outputs a set of hypotheses about what the speaker said, ranked according to probability. We represent those hypotheses as a *lattice* — a graph whose edges represent recognized words and the probability that a given word follows from the previous one.



An example of a lattice representing ASR hypotheses.

[43] *See* KAIFI0006791 (Ariya Rastrow, *On-device speech processing makes Alexa faster, lower-bandwidth* (Jan. 25, 2022), https://www.amazon.science/blog/on-device-speech-processing-makes-alexa-faster-lower-bandwidth).

Ex.1029 / Page 64 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	For example, Amazon uses cloud verification and encryption, automatic speech recognition (ASR), natural language understanding (NLU), text-to-speech (TTS), and data storage technologies.[44]

[44] *See* KAIFI0002023 *(Follow the journey of a voice request*, https://www.amazon.com/b/?node=23608618011).

Ex.1029 / Page 65 of 347
AMAZON.COM, INC.

The wake word

Think of the wake word as a verbal cue that makes things happen. In the classic tale from The Arabian Nights, an invisible door into a mountain can only be opened by saying the phrase "Open Sesame." Use the phrase, the door opens. Don't use the phrase, the door stays shut. When you have a request for Alexa, you first need to say your chosen wake word, which by default is "Alexa." Only after your Echo device detects the wake word is Alexa listening to your requests.

When it comes to privacy, there should be no surprises. You'll always be able to tell when Alexa is listening to your request because a light indicator will appear on your Echo device or an audible tone will sound. Think of the "On the Air" signs that light up in television studios during a broadcast. These indicators notify you that your device has detected the wake word and Alexa is now processing your request.

Keyword spotting

Echo devices use built-in technology called "keyword spotting" that matches spoken audio to the acoustic patterns of the wake word. Simply put, Echo devices are designed by default to detect only the sound waves of your chosen wake word, and everything else is ignored. Like water through a strainer, all other audio (people talking, faucets running, birds chirping) passes through the device until the wake word is "caught" and sent to Amazon's secure cloud, where meaning is assigned to your request. You can choose to opt-in to features like Alexa Emergency Assist that allow your device to detect more than your chosen wake word—for example, the sound of smoke alarms—but you would need to update your settings to do so.

Cloud verification and encryption

When your Echo device detects the wake word, it sends your request to Amazon's secure cloud where audio is reanalyzed to verify the wake word was spoken. If this cloud software verification is unable to confirm the wake word was spoken, the Alexa system stops processing the audio. If the wake word is verified (or Alexa is activated using the action button), your request will be processed using several sophisticated algorithms that allow Alexa to respond appropriately. All of your interactions with Alexa are encrypted in transit throughout this process.

Automatic Speech Recognition (ASR)

After wake word verification, your voice request undergoes Automatic Speech Recognition (ASR) which transcribes audio into text. Using our example request ("Alexa, what's the weather in Seattle?"), transcriptions could include one or several of the following:

"what is the weather in seattle" "watt is the weather in seattle" "what is the whether in seattle"

Alexa then chooses to act on the transcription that is most likely to be correct and stores it in Amazon's secure cloud. You can view and delete it at any time at Alexa Privacy Settings or in the Alexa app (Settings > Alexa Privacy > Review Voice History). Voice recordings and transcriptions may be used to improve the ASR process. Learn how data helps Alexa improve.

Natural Language Understanding

After ASR comes Natural Language Understanding (NLU). NLU assigns meaning to the transcription producing an "intent" which is an instruction to the Alexa system to tell it what to respond to. In our example, the intent is identified as "weather" and sent to the designated data source where information is pulled for location (Seattle) and time (today). Intents may also be used to improve the NLU process.

Text-to-Speech (TTS)

Once the necessary information is gathered, your request uses text-to-speech (TTS) technologies to generate an audio file that becomes Alexa's response. The text from Alexa's response is stored so that you can review past answers at Alexa Privacy Settings or in the Alexa app. These responses may also be reviewed by Amazon to ensure the device is providing the most relevant answers to you, and to make sure the TTS system is translating text to speech to the best of its ability.

Data storage

You have control over your Alexa experience. We make it easy for you to update your privacy settings. You can always view and manage your Alexa interactions at Alexa Privacy Settings or in the Alexa app (Settings > Alexa Privacy). If you're enrolled in voice ID, you can ask, "Alexa, what are my privacy settings?" to review how long your voice recordings are currently saved and whether or not they're used to develop new features or manually reviewed to help improve Alexa. You can also say, "Alexa, update my privacy settings" at any time to make updates.

Ex.1029 / Page 66 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	For example, Amazon uses a speech recognizer, including a "Recognize" state and an "ExpectSpeech" state.[45]

The `SpeechRecognizer` interface is the core interface of the Alexa Voice Service (AVS) and exposes directives and events for capturing and interacting with user speech. This page discusses the concepts and process flows for the `SpeechRecognizer` interface.

SpeechRecognizer functionality

Every user utterance leverages `SpeechRecognizer`, including the following interaction types between an Alexa Built-in device and AVS:

- **User Speech Capture**: Captures user speech from an Alexa Built-in device.
- **User Speech Prompting**: Prompts a user for more speech input when needed by Alexa to deliver an appropriate response.
- **Interaction Initiation Communication**: Enables a device to inform AVS of how a user initiated an Alexa interaction, such as press-and-hold, tap-and-release, or voice-initiated/wake word enabled. See Device Form Factor and Alexa Interaction.
- **ASR Profile Selection**: SpeechRecognizers chooses the appropriate Automatic Speech Recognition (ASR) profile for your product, which allows Alexa to understand user speech and respond with precision. See Automatic Speech Recognition (ASR) profile

[45] *See* KAIFI0007093 *(Understand the AVS SpeechRecognizer*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/avs-speechrecognizer-concepts.html).

Ex.1029 / Page 67 of 347
AMAZON.COM, INC.

State diagram

The following diagram illustrates state changes driven by `SpeechRecognizer` components. Boxes represent `SpeechRecognizer` states, and the connectors represent state transitions.

`SpeechRecognizer` has the following states:

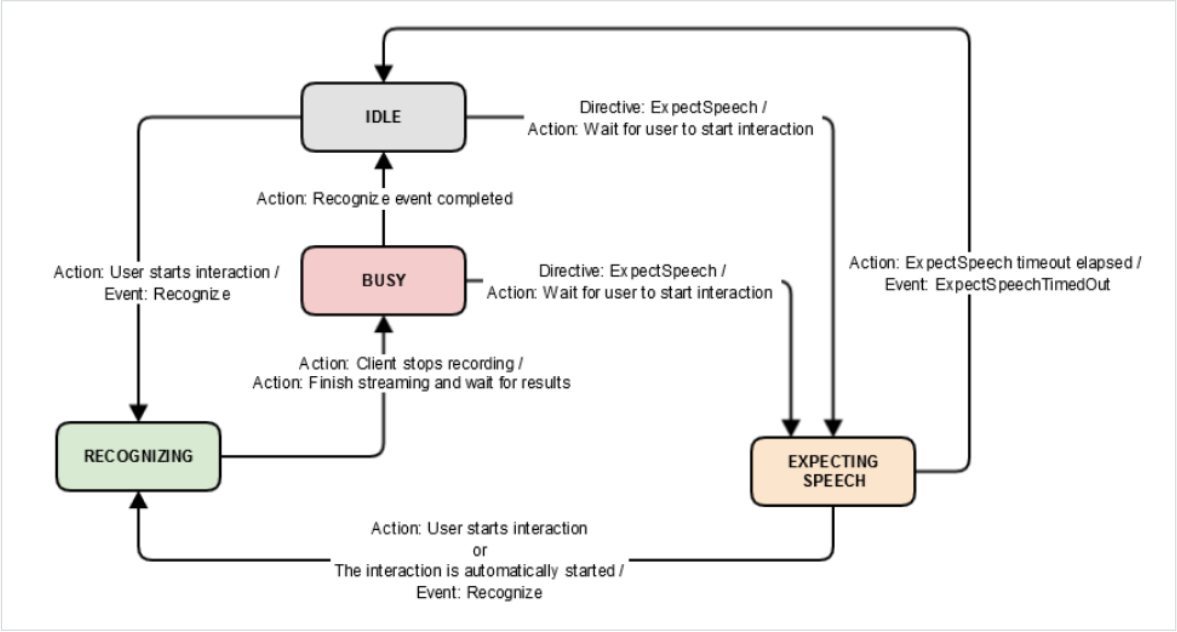
- **IDLE:** When not actively processing speech, the `SpeechRecognizer` is in an "idle" state. The idle state occurs under the following conditions:
  - Before capturing user speech.
  - Returning to an idle state after concluding a speech interaction with Alexa.
  - When an `ExpectSpeechTimedOut` event elapses.

  Note: In a multi-turn Alexa interaction, if Alexa requires more user speech input, `SpeechRecognizer` should transition from the idle state to the `ExpectSpeech` state without the user starting a new interaction.

- **RECOGNIZING:** When a user begins interacting with your client, specifically when the client streams captured audio to AVS, `SpeechRecognizer` should transition from the idle state to the `Recognize` state. `SpeechRecognizer` should remain in the `Recognize` until the client stops recording speech or finishes streaming, at which point your `SpeechRecognizer` component should transition from the `Recognize` state to the "busy" state.
- **BUSY:** While processing the speech request, `SpeechRecognizer` should be in the "busy" state. You cannot start another speech request until `SpeechRecognizer` transitions out of the busy state. From the *busy* state, `SpeechRecognizer` transitions to the idle state if Alexa processes and completes the request or to the `ExpectSpeech` state, if Alexa requires more speech input from the user.
- **EXPECTING SPEECH:** `SpeechRecognizer` should be in the `ExpectSpeech` state when Alexa requires more audio input from a user. From `ExpectSpeech`, `SpeechRecognizer` should transition to the `Recognize` when a user interaction occurs, or the interaction is automatically started on behalf of the user. It should transition to the idle state if Alexa detects no user interaction within the specified timeout window.

Ex.1029 / Page 68 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	The following diagram illustrates the expected transitions among the four `SpeechRecognizer` states:  For example, Amazon uses a wake word engine.[46]

[46] *See* KAIFI0007141 *(Wake Word Verification Requirements*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/implement-ww-verification.html).

Ex.1029 / Page 69 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

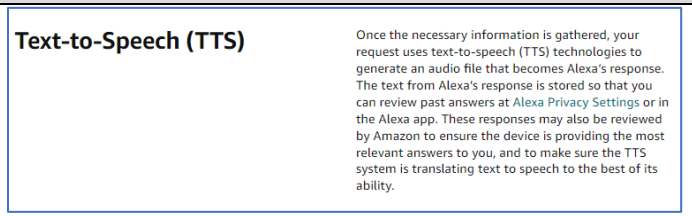
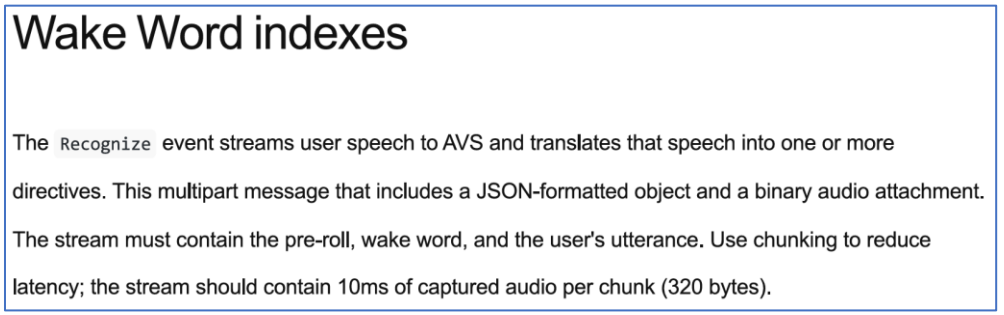
Claim 1	Accused Instrumentality(ies)
	For example, Amazon provides wake word verification.[47] For example, Amazon uses cloud verification and encryption, automatic speech recognition (ASR), natural language understanding (NLU), text-to-speech (TTS), and data storage technologies on the recognized actual subsequent input speech command.[48] **Automatic Speech Recognition (ASR)** After wake word verification, your voice request undergoes Automatic Speech Recognition (ASR) which transcribes audio into text. Using our example request (*"Alexa, what's the weather in Seattle?"*), transcriptions could include one or several of the following: • *"what is the weather in seattle"* • *"watt is the weather in seattle"* • *"what is the whether in seattle"* Alexa then chooses to act on the transcription that is most likely to be correct and stores it in Amazon's secure cloud. You can view and delete it at any time at Alexa Privacy Settings or in the Alexa app (Settings > Alexa Privacy > Review Voice History). Voice recordings and transcriptions may be used to improve the ASR process. Learn how data helps Alexa improve. **Natural Language Understanding** After ASR comes Natural Language Understanding (NLU). NLU assigns meaning to the transcription producing an "intent" which is an instruction to the Alexa system to tell it what to respond to. In our example, the intent is identified as "weather" and sent to the designated data source where information is pulled for location (Seattle) and time (today). Intents may also be used to improve the NLU process.

[47] *See* KAIFI0002012 *(Enable Wake Word Verification*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/enable-cloud-based-wake-word-verification.html).

[48] *See* KAIFI0002023 *(Follow the journey of a voice request*, https://www.amazon.com/b/?node=23608618011).

Ex.1029 / Page 70 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	 For example, Amazon uses chunking to reduce latency.[49]  For example, Amazon Alexa's wake word detection on the device works with limited CPU power.[50]

[49] *See* KAIFI0007141 *(Wake Word Verification Requirements*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/implement-ww-verification.html).
[50] *See* KAIFI0002097 (How Amazon Alexa works? Your Guide to Natural Language Processing (AI), https://towardsdatascience.com/how-amazon-alexa-works-your-guide-to-natural-language-processing-ai-7506004709d3).

Ex.1029 / Page 71 of 347
AMAZON.COM, INC.

Claim 1	Accused Instrumentality(ies)
	> The next task is "Wake Word Detection". It determines whether the user says one of the words the device is programmed to need to turn on, such as "Alexa". This is needed to minimize false positives and false negatives, which could lead to accidental purchases and angry customers. This is really complicated as it needs to identify pronunciation differences, and it needs to do so on the device, which has limited CPU power. For example, when a customer says the wake word, subsequent phrases are processed and stored by Amazon in the cloud to respond to the customer's request.[51] > Alexa is designed to get smarter every day—this is accomplished through the power of machine learning and the cloud. The more a customer uses Alexa, the more it adapts to their speech patterns, vocabulary, and personal preferences. When the customer says the wake word, their subsequent phrases are processed and stored in the cloud to respond to the customer's request and to improve the customer's experience and our services, including training our speech recognition and natural language understanding systems so Alexa can better understand customers' requests. This claim element is literally present, as described herein, but is also satisfied under the doctrine of equivalents because the structure and functionality provided by Amazon is being used to satisfy this claim limitation and therefore there is no vitiation of this element, and the Accused Instrumentalities incorporate structure and functionality that is not substantially different from the requirements of this limitation because they achieve substantially or exactly the same function, in substantially or exactly the same way as set forth in the claim element (as described herein), and achieves substantially or exactly the same result.

[51] *See* KAIFI0007380 *(White paper – Alexa Privacy and Data Handling Overview*, https://d1.awsstatic.com/product-marketing/A4B/White%20Paper%20-%20Alexa%20Privacy%20and%20Data%20Handling%20Overview.pdf)

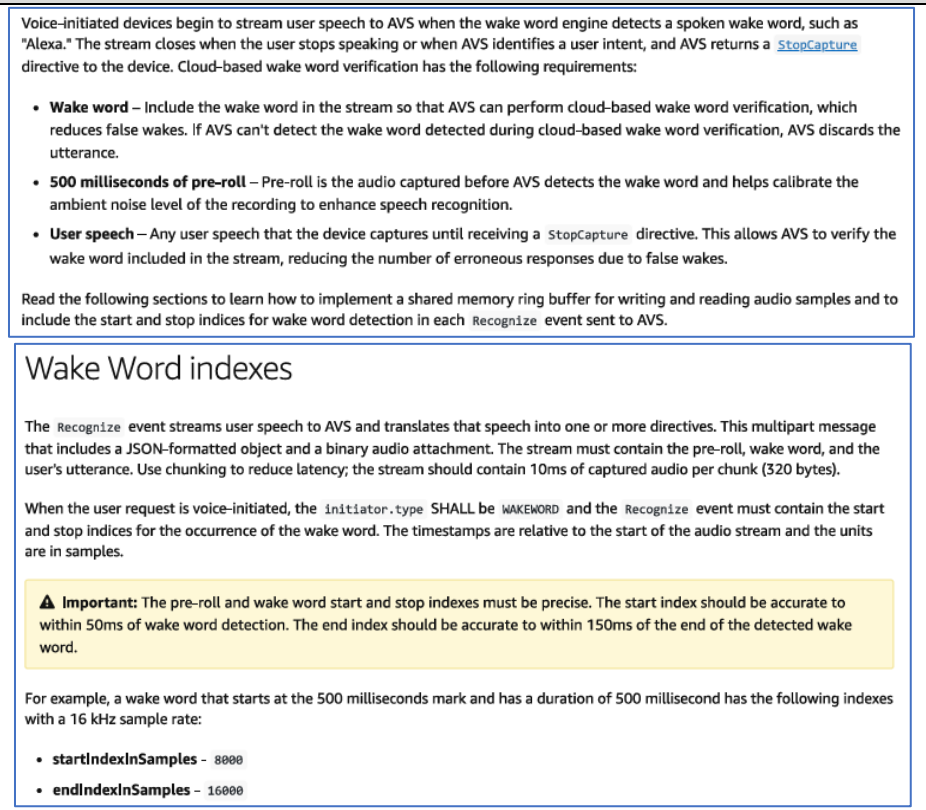
Ex.1029 / Page 72 of 347
AMAZON.COM, INC.

Claim 1	Accused Instrumentality(ies)
wherein the preset call command recognition unit comprises a token passing unit configured to assume the input speech to be the preset call command according to the result of the comparison using a minimum recognition network composed of a silence interval accompanied by noise and the preset call command.	In the Accused Instrumentalities, the preset call command recognition unit comprises a token passing unit. The token passing unit is configured to assume the input speech to be the preset call command according to the result of the comparison using a minimum recognition network composed of a silence interval accompanied by noise and the preset call command. For example, Amazon uses a user speech stream. [52]

[52] *See* KAIFI0007141 *(Wake Word Verification Requirements*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/implement-ww-verification.html).

Ex.1029 / Page 73 of 347
AMAZON.COM, INC.

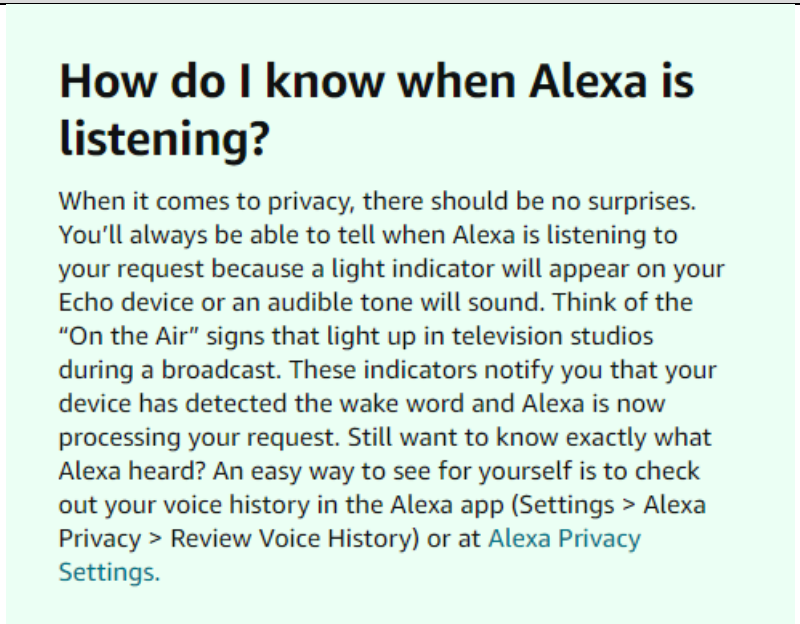
Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	 For example, Amazon's Alexa indicators notify that a device has detected the wake word and is now processing a request. [53]

[53] *See* KAIFI0002092 *(How Alexa works: Wake word*, https://www.amazon.com/b?ie=UTF8&node=23608571011).

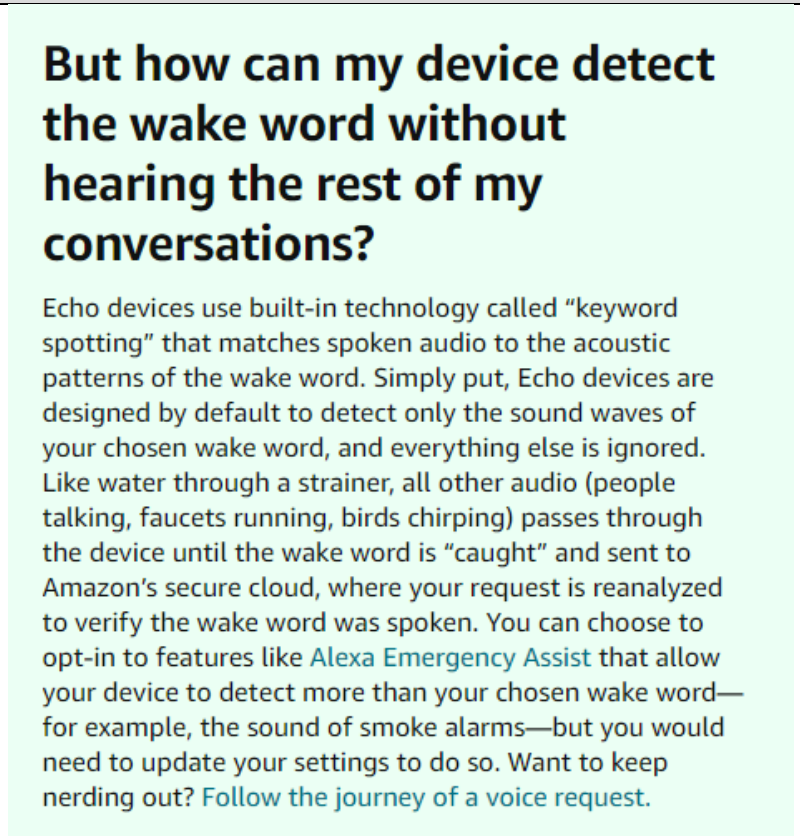
Ex.1029 / Page 74 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	 For example, Amazon detects only the sound waves of a chosen wake word, and everything else is ignored.[54]

[54] *See Id*

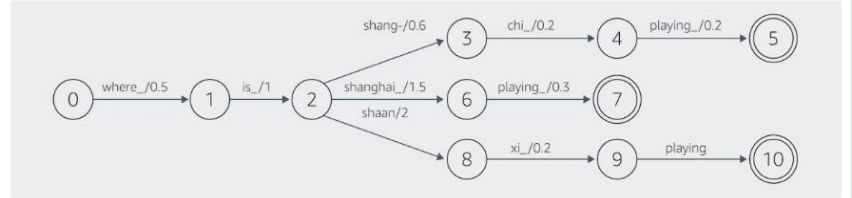
Ex.1029 / Page 75 of 347
AMAZON.COM, INC.

Claim 1	Accused Instrumentality(ies)
	 For example, Amazon provides an ASR model.[55]

[55] *See* KAIFI0006791 (Ariya Rastrow, *On-device speech processing makes Alexa faster, lower-bandwidth* (Jan. 25, 2022), https://www.amazon.science/blog/on-device-speech-processing-makes-alexa-faster-lower-bandwidth).

Ex.1029 / Page 76 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	System architecture Our on-device ASR model takes in an acoustic speech signal and outputs a set of hypotheses about what the speaker said, ranked according to probability. We represent those hypotheses as a *lattice* — a graph whose edges represent recognized words and the probability that a given word follows from the previous one.  An example of a lattice representing ASR hypotheses.

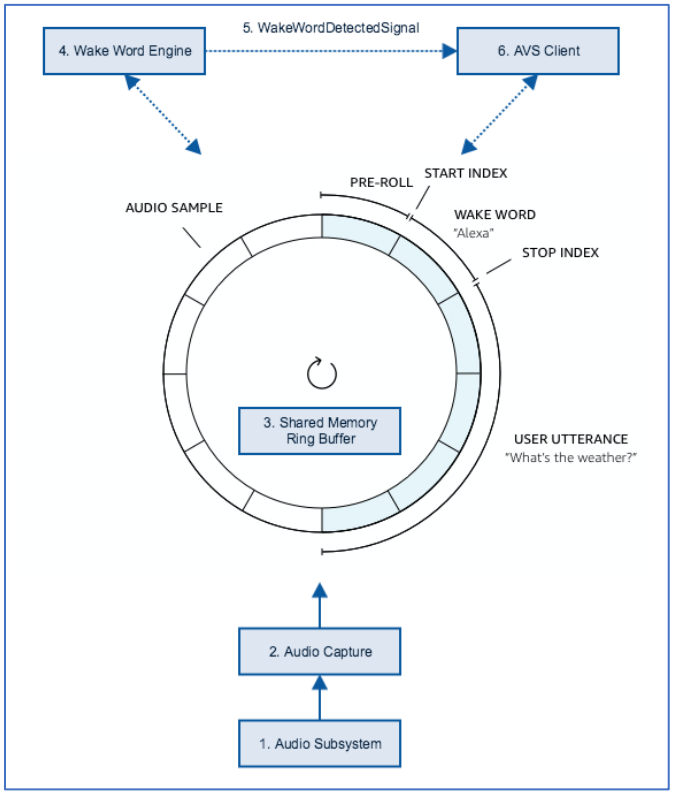
Ex.1029 / Page 77 of 347
AMAZON.COM, INC.

Claim 1	Accused Instrumentality(ies)
	With cloud-based ASR, encrypted audio streams to the cloud in small snippets called "frames". With on-device ASR, only the lattice is sent to the cloud, where a large and powerful neural language model reranks the hypotheses. The lattice can't be sent until the customer has finished speaking, as words later in a sequence can dramatically change the overall probability of a hypothesis. The model that determines when the customer has finished speaking is called an end-pointer. End-pointers offer a natural trade-off between accuracy and latency: an aggressive end-pointer will initiate speech processing earlier, but it might cut the speaker off prematurely, resulting in a poor customer experience. On the device, we in fact run two end-pointers: One is a speculative end-pointer that we have tuned to be about 200 milliseconds faster than the final end-pointer, so we can initiate downstream processing — such as natural-language understanding (NLU) — ahead of the final end-pointed ASR result. In exchange for speed, however, we trade off a little accuracy. The final end-pointer takes longer to make a decision but is more accurate. In cases in which the first end-pointer cuts speech off too early, the final end-pointer sends a revised lattice and the instruction to reset downstream processing. In the large majority of cases, however, the aggressive end-pointer is correct, which reduces user-perceived latency, since downstream tasks are initiated earlier. For example, Amazon uses a shared memory ring buffer.[56]

[56] *See* KAIFI0007141 *(Wake Word Verification Requirements*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/implement-ww-verification.html).

Ex.1029 / Page 78 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	Shared Memory Ring Buffer A shared memory ring buffer satisfies the requirement to stream a continuous utterance to AVS that includes 500 milliseconds of pre-roll, the wake word, and the user's request for cloud-based wake word verification. This reduces latency and the need to copy audio samples multiple times. The following diagram illustrates one approach; alternative implementations are allowed as long as the requisite functionality is maintained: 

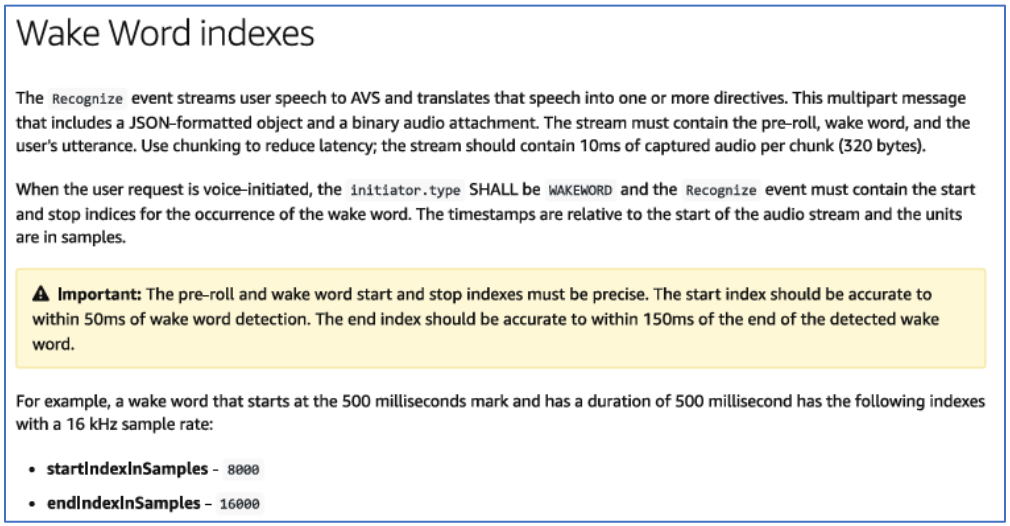
Ex.1029 / Page 79 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	1. **Audio Subsystem** - Middleware, such as Advanced Linux Sound Architecture (ALSA), that provides APIs to open a stream from a recording device, and receive an audio sample stream. 2. **Audio Capture** - A process that opens a recording device, and writes audio samples into the shared memory ring buffer. 3. **Shared Memory Ring Buffer** - A shared memory ring buffer object that is thread safe, and accessible by multiple processes. Access to this memory block should support at least one writer and two readers simultaneously. You may choose to implement these executables as a shared library. • **Write API** - Used by the recording process to write audio samples to the shared memory ring buffer. Write samples should move data so that the oldest audio samples are overwritten. This example shows a typical API signature: `WriteSamples(BufferIn, BufferInLength)`. • **Read API** - Used by the wake word engine and the AVS client to read from specific locations in the shared memory ring buffer. A typical implementation represents location as a 64-bit integer. For example, reading 200 samples from the location 100 is reading the 100th to 300th samples that were written to the shared memory ring buffer. This is an example of a typical API signature: `ReadSamples(BufferIn, BufferInLength)`. **Note**: Reading from an old location that was overwritten with new audio samples will result in an unintended sample read. 4. **Wake Word Engine** - The component that reads the audio samples written to the shared memory ring buffer and analyzes the audio samples for occurrences of the wake word. The procedure of reading and processing audio samples occurs in a loop. For example, the loop reads N samples starting at location `0`; then analyzes the samples for an occurrence of the wake word. In the next iteration, it reads and analyzes the next N samples. When an occurrence of the wake word is detected, the wake word engine must identify two locations: 1) the start of the wake word, and 2) the end of the wake word. These values are stored as `startIndexInSamples` and `endIndexInSamples`. 5. **WakeWordDetectedSignal** - A signal sent from the wake word engine to the AVS client when an occurrence of the wake word is detected. This signal includes two locations, the `startIndexInSamples` and `endIndexInSamples`. 6. **AVS Client** - The client receives the `WakeWordDetectedSignal`, extracts the `startIndexInSamples`, then reads and streams audio samples to AVS including 500 milliseconds of pre-roll. • The pre-roll is the number of samples corresponding to 500 milliseconds. For example, if the audio samples were recorded at a rate of 16 kilohertz (kHz), then pre-roll is 8,000 samples.

Ex.1029 / Page 80 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	For example, Amazon uses wake word indexes.[57]  Wake Word indexes The `Recognize` event streams user speech to AVS and translates that speech into one or more directives. This multipart message that includes a JSON-formatted object and a binary audio attachment. The stream must contain the pre-roll, wake word, and the user's utterance. Use chunking to reduce latency; the stream should contain 10ms of captured audio per chunk (320 bytes). When the user request is voice-initiated, the `initiator.type` SHALL be `WAKEWORD` and the `Recognize` event must contain the start and stop indices for the occurrence of the wake word. The timestamps are relative to the start of the audio stream and the units are in samples. **Important:** The pre-roll and wake word start and stop indexes must be precise. The start index should be accurate to within 50ms of wake word detection. The end index should be accurate to within 150ms of the end of the detected wake word. For example, a wake word that starts at the 500 milliseconds mark and has a duration of 500 millisecond has the following indexes with a 16 kHz sample rate: • **startIndexInSamples** - `8000` • **endIndexInSamples** - `16000` For example, Amazon describes a voice request. [58]

[57] *See id.*

[58] *See* KAIFI0002023 *(Follow the journey of a voice request*, https://www.amazon.com/b/?node=23608618011).

Ex.1029 / Page 81 of 347
AMAZON.COM, INC.



Ex.1029 / Page 82 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	For example, Amazon uses speech recognition.[59] Near-field and far-field tests are conducted in various locations, which mimic real-world user-device interaction patterns. The device's performance is evaluated under various conditions such as in silence, during audio playback from the device, or in environments with background noise. Such noise could be from music playing on another device, stationary noise such as an air conditioner, or other ambient noise. For example, Amazon uses a user speech stream.[60] Voice-initiated devices begin to stream user speech to AVS when the wake word engine detects a spoken wake word, such as "Alexa." The stream closes when the user stops speaking or when AVS identifies a user intent, and AVS returns a `StopCapture` directive to the device. Cloud-based wake word verification has the following requirements: For example, Amazon uses ASR.[61]

[59] *See* KAIFI0001353 (Felipe Pinzon, *An Amazon technical expert provides three acoustic certification-related tips for AVS device makers* (Aug. 16, 2023) https://developer.amazon.com/en-US/blogs/alexa/device-makers/2023/08/acoustic-certification-alexa-august-2023).

[60] *See* KAIFI0007141 *(Wake Word Verification Requirements*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/implement-ww-verification.html).

[61] *See* KAIFI0006791 (Ariya Rastrow, Shehzad Mevawalla, *On-device speech processing makes Alexa faster, lower-bandwidth*, AMAZON SCIENCE (Jan. 25, 2022), https://www.amazon.science/blog/on-device-speech-processing-makes-alexa-faster-lower-bandwidth).

Ex.1029 / Page 83 of 347
AMAZON.COM, INC.

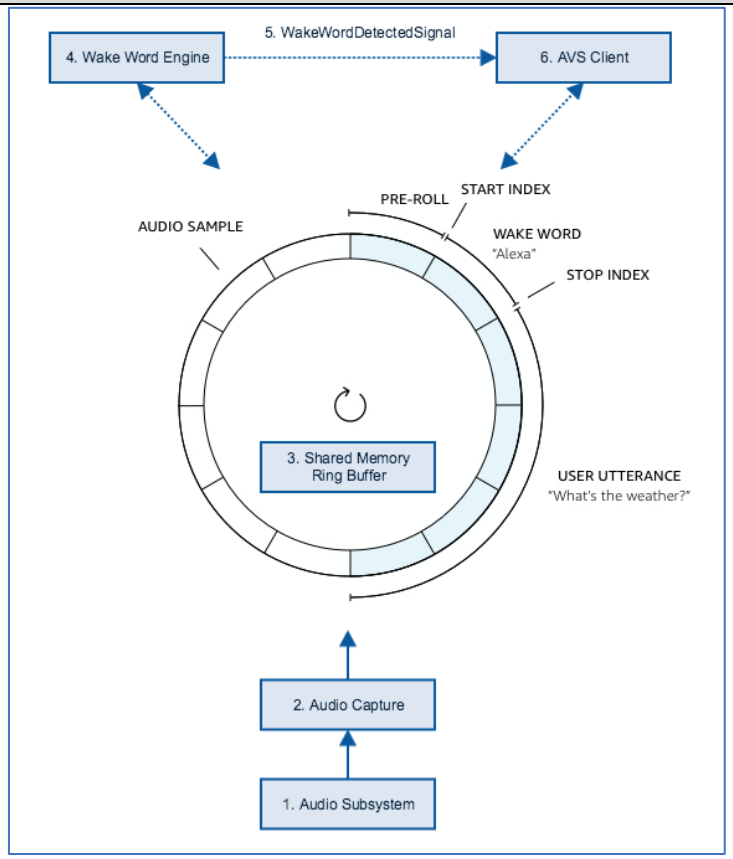
Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	Moreover, in the cloud, the separate components of Alexa's speech-processing stack — automatic speech recognition (ASR), whisper detection, and speaker identification — run on separate server nodes with their own powerful processors. On-device, those functions have to share hardware not only with each other but with Alexa's other core device functions, such as music playback. With cloud-based ASR, encrypted audio streams to the cloud in small snippets called "frames". With on-device ASR, only the lattice is sent to the cloud, where a large and powerful neural language model reranks the hypotheses. The lattice can't be sent until the customer has finished speaking, as words later in a sequence can dramatically change the overall probability of a hypothesis. With cloud-based ASR, encrypted audio streams to the cloud in small snippets called "frames". With on-device ASR, only the lattice is sent to the cloud, where a large and powerful neural language model reranks the hypotheses. The lattice can't be sent until the customer has finished speaking, as words later in a sequence can dramatically change the overall probability of a hypothesis. For example, Amazon uses wake word and speech recognition.[62]

[62] *See* KAIFI0007141 *(Wake Word Verification Requirements*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/implement-ww-verification.html).

Ex.1029 / Page 84 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	

Ex.1029 / Page 85 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	4. **Wake Word Engine** - The component that reads the audio samples written to the shared memory ring buffer and analyzes the audio samples for occurrences of the wake word. The procedure of reading and processing audio samples occurs in a loop. For example, the loop reads N samples starting at location `0`; then analyzes the samples for an occurrence of the wake word. In the next iteration, it reads and analyzes the next N samples. When an occurrence of the wake word is detected, the wake word engine must identify two locations: 1) the start of the wake word, and 2) the end of the wake word. These values are stored as `startIndexInSamples` and `endIndexInSamples`. For example, Amazon uses an ASR decoder.[63] In addition to these two classifiers, a third model component comes into play: meta-features emitted by the ASR decoder. These meta-features describe for how long (if at all) the ASR decoder is running "idle." In other words, how meaningful are the incoming audio chunks (actual content vs. silence/noise)? For example, according to Amazon, performance conditions include noise.[64] Near-field and far-field tests are conducted in various locations, which mimic real-world user-device interaction patterns. The device's performance is evaluated under various conditions such as in silence, during audio playback from the device, or in environments with background noise. Such noise could be from music playing on another device, stationary noise such as an air conditioner, or other ambient noise.

[63] *See* KAIFI0001184 (Roland Maas, *Alexa scientists address challenges of end-pointing* (Apr. 10, 2018), https://www.amazon.science/blog/alexa-scientists-address-challenges-of-end-pointing).

[64] *See* KAIFI0001353 (Felipe Pinzon, *An Amazon technical expert provides three acoustic certification-related tips for AVS device makers* (Aug. 16, 2023) https://developer.amazon.com/en-US/blogs/alexa/device-makers/2023/08/acoustic-certification-alexa-august-2023).

Ex.1029 / Page 86 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	For example, according to Amazon, ASR is trained for silence and noise.[65] To train our acoustic models, we used connectionist temporal classification (CTC), followed by multiple passes of state-level minimum-Bayes-risk (sMBR) training. To make the acoustic model more robust, we also mixed noise into the training set, enabling the model to focus on characteristics of the input signal that vary less under different acoustic conditions. Adapting to conversational speech also required modification of Alexa's end-pointer, which determines when a customer has finished speaking. The end-pointer already distinguishes between pauses at the ends of sentences, indicating that the customer has stopped speaking and that Alexa needs to follow up, and mid-sentence pauses, which may be permitted to go on a little longer. For Live Translation, we modified the end-pointer to tolerate longer pauses at the ends of sentences, as speakers engaged in long conversations will often take time between sentences to formulate their thoughts. For example, according to Amazon, ASR adds a silence label at start and end of the label sequence and in between words.[66]

[65] *See* KAIFI0002096 (Shirin Saleem, Roland Maas, *How Alexa's new Live Translation for conversations works* (Dec. 14, 2020), https://www.amazon.science/blog/how-alexas-new-live-translation-for-conversations-works).

[66] *See* KAIFI0004444 (Tulsiani, et. al, *Improved training strategies for end-to-end speech recognition in digital voice assistants*, https://assets.amazon.science/ab/b4/2f3e5aec4506930e172478586bf4/improved-training-strategies-for-end-to-end-speech-recognition-in-digital-voice-assistants.pdf).

Ex.1029 / Page 87 of 347
AMAZON.COM, INC.

Claim 1	Accused Instrumentality(ies)

3.3. Label smoothing via probabilistic modelling of silence and use of multiple pronunciation

Label smoothing was proposed as a regularization technique to improve generalization [28, 29]. It proposes to penalize low entropy output distribution by adding noise to the labels [19, 28]. In this work, we introduce a simple yet powerful technique for changing label distribution. To penalize over-confident predictions by network, while training we: (i) randomly use one of the multiple pronunciations from the lexicon (as also done in [10, 22]) (ii) randomly add silence label at start and end of the label sequence and in between words. This has the effect of changing label distribution for a fixed word sequence, thereby providing the regularization effect. For this work, we use probability of 0.8 for inserting silence label at start and end and 0.2 for inserting a silence label between words.

For example, according to Amazon, ASR uses end pointers.[67]

The model that determines when the customer has finished speaking is called an end-pointer. End-pointers offer a natural trade-off between accuracy and latency: an aggressive end-pointer will initiate speech processing earlier, but it might cut the speaker off prematurely, resulting in a poor customer experience.

[67] *See* KAIFI0006791 (Ariya Rastrow, Shehzad Mevawalla, *On-device speech processing makes Alexa faster, lower-bandwidth*, AMAZON SCIENCE (Jan. 25, 2022), https://www.amazon.science/blog/on-device-speech-processing-makes-alexa-faster-lower-bandwidth).

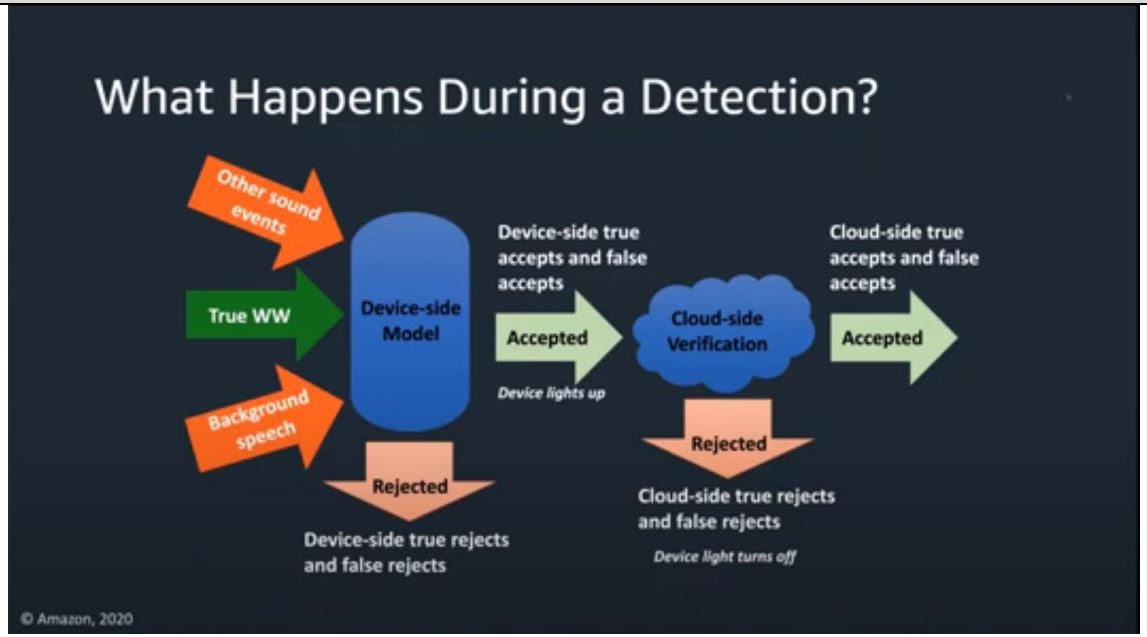
Ex.1029 / Page 88 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 1	Accused Instrumentality(ies)
	In a process called "end-pointing", however, Alexa recognizes the end of an utterance by the short period of silence that follows end of speech, and that silence is part of the input to the whisper detector. When we apply the detector to live data, we typically see that its confidence increases across most of the duration of an utterance then falls off precipitously in the final 50 or so frames. For example, after receiving a wake word together with other sound events and background speech, Amazon's device-side model and cloud-side verification perform "true accepts" of the wake word.[68]

[68] *See* KAIFI0007814 (*The Science Behind Alexa's Wake Word Detection*, https://www.youtube.com/watch?v=kYrw2cC9dvs).

Ex.1029 / Page 89 of 347
AMAZON.COM, INC.

Claim 1	Accused Instrumentality(ies)
	 This claim element is literally present, as described herein, but is also satisfied under the doctrine of equivalents because the structure and functionality provided by Amazon is being used to satisfy this claim limitation and therefore there is no vitiation of this element, and the Accused Instrumentalities incorporate structure and functionality that is not substantially different from the requirements of this limitation because they achieve substantially or exactly the same function, in substantially or exactly the same way as set forth in the claim element (as described herein), and achieves substantially or exactly the same result.

Claim 2 of the ’196 Patent

Ex.1029 / Page 90 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 2	Accused Instrumentality(ies)
The method of claim 1, wherein determining whether the input speech is recognized as the preset call command further comprises: computing, by a confidence detection unit, a confidence rate of the input speech; and recognizing, by the confidence detection unit, the input speech as the preset call command based on the computed confidence rate.	Amazon's determining whether the input speech is recognized as the preset call command further comprises: computing, by a confidence detection unit, a confidence rate of the input speech; and recognizing, by the confidence detection unit, the input speech as the preset call command based on the computed confidence rate. For example, Amazon only listens to requests after the wake word is detected.[69] **How Alexa works: Wake word** Think of the wake word as a verbal cue that makes things happen. In the classic tale from *The Arabian Nights*, an invisible door into a mountain can only be opened by saying the phrase "Open Sesame." Use the phrase, the door opens. Don't use the phrase, the door stays shut. While it's not magic, because there's a lot of technology behind Alexa, the wake word does open the door for Alexa to begin fulfilling your request. When you have a request for Alexa, you first need to say your chosen wake word, which by default is "Alexa." Only after your Echo device detects the wake word is Alexa listening to your requests. For example, Amazon uses built-in technology called "keyword spotting" that matches spoken audio to the acoustic patterns of the wake word and are designed to detect only the sound waves of the chosen wake word, which is sent to Amazon's secure cloud, where the request is reanalyzed to verify the wake word was spoken.[70]

[69] *See* KAIFI0002092 *(How Alexa works: Wake word*, https://www.amazon.com/b?ie=UTF8&node=23608571011).
[70] *See id.*

Ex.1029 / Page 91 of 347
AMAZON.COM, INC.

Claim 2	Accused Instrumentality(ies)
	But how can my device detect the wake word without hearing the rest of my conversations? Echo devices use built-in technology called "keyword spotting" that matches spoken audio to the acoustic patterns of the wake word. Simply put, Echo devices are designed by default to detect only the sound waves of your chosen wake word, and everything else is ignored. Like water through a strainer, all other audio (people talking, faucets running, birds chirping) passes through the device until the wake word is "caught" and sent to Amazon's secure cloud, where your request is reanalyzed to verify the wake word was spoken. You can choose to opt-in to features like Alexa Emergency Assist that allow your device to detect more than your chosen wake word—for example, the sound of smoke alarms—but you would need to update your settings to do so. Want to keep nerding out? Follow the journey of a voice request. For example, Amazon achieves far-field speech recognition in noisy environments with high accuracy through a combination of machine learning algorithms, data, and immense computing power.[71] By launching Echo in November 2014, we showed that far-field speech recognition in noisy environments can be done with high accuracy through a combination of machine learning algorithms, data, and immense computing power.

[71] *See* KAIFI0004342 *(How Our Scientists are Making Alexa Smarter* (Mar. 29, 2018), https://www.aboutamazon.com/news/devices/how-our-scientists-are-making-alexa-smarter).

Ex.1029 / Page 92 of 347
AMAZON.COM, INC.

Claim 2	Accused Instrumentality(ies)
	For example, Amazon's wake word detection employs deep learning technology running on-device to recognize the wake word the user has chosen.[72] It starts with detecting the "wake word" that triggers Alexa to start listening to the subsequent spoken words from the user. Wake word detection employs deep learning technology running on-device to recognize the wake word the user has chosen. Far-field automatic speech recognition (ASR) in the Amazon Web Services (AWS) cloud then converts into text the audio following the wake word, and determines when the user has stopped talking to Alexa. For example, Amazon uses a suite of learning techniques, including supervised, semi-supervised, and unsupervised learning.[73]

[72] *See* KAIFI0004342 *(How Our Scientists are Making Alexa Smarter* (Mar. 29, 2018), https://www.aboutamazon.com/news/devices/how-our-scientists-are-making-alexa-smarter).
[73] *See id.*

Ex.1029 / Page 93 of 347
AMAZON.COM, INC.

Claim 2	Accused Instrumentality(ies)
	A. Since Alexa's brain is mostly in the cloud, she gets smarter with every interaction. Alexa uses a suite of learning techniques: supervised, semi-supervised, and unsupervised learning. While supervised learning is still most powerful, it does not scale since we cannot generate manual labels at the pace required to continually improve Alexa for our customers. Therefore, our scientists and engineers are continually applying and inventing new learning techniques to reduce reliance on manual labels for training our statistical models. For instance, active learning as a class of semi-supervised learning techniques, where the system itself posits which part of the interactions it needs input from a human expert, is pervasive across our technologies. Unsupervised learning without any labeled responses is also being applied to make Alexa smarter, especially for speech recognition. Finally, we also use transfer learning, allowing Alexa to learn from one skill to another or even one language to another. For example, Amazon uses a machine learning model trained from many occurrences of detected wake words to score a match with the input voice data.[74] **Wake Word Model** A WWD uses a machine learning model trained from many occurrences of detected wake words to score a match with the input voice data. If the confidence score exceeds the model's threshold, the wake word is considered detected and the WWD triggers. The wake word model may incorporate multiple wake words and also be trained to support multiple languages.

[74] *See* KAIFI0007120 (*Voice Interoperability Initiative Architecture Series Whitepapers*, https://m.media-amazon.com/images/G/01/vii/VII_Architecture_Series_Whitepapers_-_Multi-Agents_Wake_Word.pdf).

Ex.1029 / Page 94 of 347
AMAZON.COM, INC.

Claim 2	Accused Instrumentality(ies)
	For example, Amazon's ASR system outputs a confidence score to verify the presence of the wake word by comparing to a fixed threshold.[75]

2.2. Wakeword Verification

Unlike detection, which operates in a streaming manner, wake-word verification performs a single pass on the input localized context frame. To validate our proposed CRA architecture, we consider several baseline systems that include both whole-word and phonetic based models.

2.2.1. Baselines

Phonetically modeled systems that we consider are ASR [21] and 2-stage DNN-HMM [11]. The ASR system we consider is a large vocabulary DNN-HMM system, with an n-gram language model to recognize the spoken words in the audio stream. The system outputs a confidence score for each recognized word in

[75] *See* KAIFI0001937 (Rajath Kumar, et. al, *Building a Robust Word-Level Wakeword Verification Network,* https://assets.amazon.science/ba/c5/bd48ce11445ba0368d2d9191600d/building-a-robust-word-level-wakeword-verification-network.pdf).

Ex.1029 / Page 95 of 347
AMAZON.COM, INC.

Claim 2	Accused Instrumentality(ies)
	the range (0,1). Similarly, for the defined wakeword, the ASR model outputs a confidence score, which we use to verify the presence of the wakeword by comparing to a fixed threshold. Since every stream is initiated with the wakeword, we set the confidence score for the wakeword to be 0 when the wakeword is not detected by ASR. Note that this ASR model is not specifically trained for the wakeword verification task, i.e., is not biased towards the wakeword. In the 2-stage DNN-HMM system, the first stage is similar to the acoustic part of the ASR system. However, unlike ASR, the 2-stage system is modeled only on the phonemes representing the wakeword and 2 additional outputs to distinguish background and silence [22]. A fixed length feature vector is extracted from the first stage DNN-HMM. This vector is engineered to aggregate information over all the frames in the hypothesized wakeword segment of the stream [22]. We pass the extracted vector as input to the second stage, which consist of multiple feed-forward layers. The second stage is a binary classification system trained to verify the presence of the wakeword. The two stages are trained separately, and the resulting score from the second stage is used for verification. This claim element is literally present, as described herein, but is also satisfied under the doctrine of equivalents because the structure and functionality provided by Amazon is being used to satisfy this claim limitation and therefore there is no vitiation of this element, and the Accused Instrumentalities incorporate structure and functionality that is not substantially different from the requirements of this limitation because they achieve substantially or exactly the same function, in substantially or exactly the same way as set forth in the claim element (as described herein), and achieves substantially or exactly the same result.

Ex.1029 / Page 96 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 3 of the '196 Patent

Claim 3	Accused Instrumentality(ies)
The method of claim 2, wherein recognizing the input speech comprises: recognizing the input speech as the preset call command if the computed confidence rate is higher than a threshold value.	Amazon's recognizing the input speech comprises recognizing the input speech as the preset call command if the computed confidence rate is higher than a threshold value. For example, Amazon uses a machine learning model trained from many occurrences of detected wake words to score a match with the input voice data.[76] **Wake Word Model** A WWD uses a machine learning model trained from many occurrences of detected wake words to score a match with the input voice data. If the confidence score exceeds the model's threshold, the wake word is considered detected and the WWD triggers. The wake word model may incorporate multiple wake words and also be trained to support multiple languages. For example, Amazon's ASR system outputs a confidence score to verify the presence of the wake word by comparing to a fixed threshold.[77]

[76] *See* KAIFI0007120 (*Voice Interoperability Initiative Architecture Series Whitepapers*, https://m.media-amazon.com/images/G/01/vii/VII_Architecture_Series_Whitepapers_-_Multi-Agents_Wake_Word.pdf).
[77] *See* KAIFI0001937 (Rajath Kumar, et. al, *Building a Robust Word-Level Wakeword Verification Network*, https://assets.amazon.science/ba/c5/bd48ce11445ba0368d2d9191600d/building-a-robust-word-level-wakeword-verification-network.pdf).

Ex.1029 / Page 97 of 347
AMAZON.COM, INC.

Claim 3	Accused Instrumentality(ies)
	2.2. Wakeword Verification Unlike detection, which operates in a streaming manner, wake-word verification performs a single pass on the input localized context frame. To validate our proposed CRA architecture, we consider several baseline systems that include both whole-word and phonetic based models. *2.2.1. Baselines* Phonetically modeled systems that we consider are ASR [21] and 2-stage DNN-HMM [11]. The ASR system we consider is a large vocabulary DNN-HMM system, with an n-gram language model to recognize the spoken words in the audio stream. The system outputs a confidence score for each recognized word in

Ex.1029 / Page 98 of 347
AMAZON.COM, INC.

Claim 3	Accused Instrumentality(ies)
	the range (0,1). Similarly, for the defined wakeword, the ASR model outputs a confidence score, which we use to verify the presence of the wakeword by comparing to a fixed threshold. Since every stream is initiated with the wakeword, we set the confidence score for the wakeword to be 0 when the wakeword is not detected by ASR. Note that this ASR model is not specifically trained for the wakeword verification task, i.e., is not biased towards the wakeword. In the 2-stage DNN-HMM system, the first stage is similar to the acoustic part of the ASR system. However, unlike ASR, the 2-stage system is modeled only on the phonemes representing the wakeword and 2 additional outputs to distinguish background and silence [22]. A fixed length feature vector is extracted from the first stage DNN-HMM. This vector is engineered to aggregate information over all the frames in the hypothesized wakeword segment of the stream [22]. We pass the extracted vector as input to the second stage, which consist of multiple feed-forward layers. The second stage is a binary classification system trained to verify the presence of the wakeword. The two stages are trained separately, and the resulting score from the second stage is used for verification. For example, Amazon employs a general far-filed ASR model to automatically transcribe the unlabeled speech corpus. Since the ASR transcript may not be accurate, only the utterances with WW confidence score higher than a threshold θ_P is selected.[78]

[78] *See* KAIFI0007074 (Yixin Gao, et. al, *TOWARDS DATA-EFFICIENT MODELING FOR WAKE WORD SPOTTING*, https://assets.amazon.science/7c/b2/5e3e6a164920bfc167fb5586d3f2/scipub-1260.pdf).

Ex.1029 / Page 99 of 347
AMAZON.COM, INC.

Claim 3	Accused Instrumentality(ies)
	In this paper we develop a simple but efficient semi-supervised learning strategy to mine useful WW-training data from large un-transcribed general-speech corpora. We employ a general far-field ASR model to automatically transcribe the unlabeled speech corpus, then select WW-specific utterances and add them to our training set. The DNN targets are obtained by force aligning the speech to the automatic transcripts. Since the ASR transcript may not be accurate, only the utterances with WW confidence score higher than a thresh-old θ_P are selected. It should be noted that, in the utterances mined here the WW may appear in the middle of the utterance, therefore the phonetic context may be different from the utterances initiated by the WW. In practice, our experiments show that such context mismatch have unobservable effects on WW spotting. This claim element is literally present, as described herein, but is also satisfied under the doctrine of equivalents because the structure and functionality provided by Amazon is being used to satisfy this claim limitation and therefore there is no vitiation of this element, and the Accused Instrumentalities incorporate structure and functionality that is not substantially different from the requirements of this limitation because they achieve substantially or exactly the same function, in substantially or exactly the same way as set forth in the claim element (as described herein), and achieves substantially or exactly the same result.

Ex.1029 / Page 100 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 4 of the '196 Patent

Claim 4	Accused Instrumentality(ies)
The method of claim 2, wherein recognizing the input speech comprises: determining whether or not the input speech is an authentic call command based on the computed confidence rate; and recognizing the input speech as the preset call command if it is determined that the input speech is the authentic call command.	Amazon's recognizing the input speech comprises: determining whether or not the input speech is an authentic call command based on the computed confidence rate; and recognizing the input speech as the preset call command if it is determined that the input speech is the authentic call command. For example, Amazon uses a machine learning model trained from many occurrences of detected wake words to score a match with the input voice data.[79] **Wake Word Model** A WWD uses a machine learning model trained from many occurrences of detected wake words to score a match with the input voice data. If the confidence score exceeds the model's threshold, the wake word is considered detected and the WWD triggers. The wake word model may incorporate multiple wake words and also be trained to support multiple languages.

[79] *See* KAIFI0007120 (*Voice Interoperability Initiative Architecture Series Whitepapers*, https://m.media-amazon.com/images/G/01/vii/VII_Architecture_Series_Whitepapers_-_Multi-Agents_Wake_Word.pdf).

Ex.1029 / Page 101 of 347
AMAZON.COM, INC.

Claim 4	Accused Instrumentality(ies)
	For example, Amazon's ASR system outputs a confidence score to verify the presence of the wake word by comparing to a fixed threshold.[80]

2.2. Wakeword Verification

Unlike detection, which operates in a streaming manner, wake-word verification performs a single pass on the input localized context frame. To validate our proposed CRA architecture, we consider several baseline systems that include both whole-word and phonetic based models.

2.2.1. Baselines

Phonetically modeled systems that we consider are ASR [21] and 2-stage DNN-HMM [11]. The ASR system we consider is a large vocabulary DNN-HMM system, with an n-gram language model to recognize the spoken words in the audio stream. The system outputs a confidence score for each recognized word in

[80] *See* KAIFI0001937 (Rajath Kumar, et. al, *Building a Robust Word-Level Wakeword Verification Network*, https://assets.amazon.science/ba/c5/bd48ce11445ba0368d2d9191600d/building-a-robust-word-level-wakeword-verification-network.pdf).

Ex.1029 / Page 102 of 347
AMAZON.COM, INC.

Claim 4	Accused Instrumentality(ies)
	the range (0,1). Similarly, for the defined wakeword, the ASR model outputs a confidence score, which we use to verify the presence of the wakeword by comparing to a fixed threshold. Since every stream is initiated with the wakeword, we set the confidence score for the wakeword to be 0 when the wakeword is not detected by ASR. Note that this ASR model is not specifically trained for the wakeword verification task, i.e., is not biased towards the wakeword. In the 2-stage DNN-HMM system, the first stage is similar to the acoustic part of the ASR system. However, unlike ASR, the 2-stage system is modeled only on the phonemes representing the wakeword and 2 additional outputs to distinguish background and silence [22]. A fixed length feature vector is extracted from the first stage DNN-HMM. This vector is engineered to aggregate information over all the frames in the hypothesized wakeword segment of the stream [22]. We pass the extracted vector as input to the second stage, which consist of multiple feed-forward layers. The second stage is a binary classification system trained to verify the presence of the wakeword. The two stages are trained separately, and the resulting score from the second stage is used for verification. For example, Amazon employs a general far-field ASR model to automatically transcribe the unlabeled speech corpus. Since the ASR transcript may not be accurate, only the utterances with WW confidence score higher than a threshold θ_P is selected.[81]

[81] *See* KAIFI0007074 (Yixin Gao, et. al, *TOWARDS DATA-EFFICIENT MODELING FOR WAKE WORD SPOTTING*, https://assets.amazon.science/7c/b2/5e3e6a164920bfc167fb5586d3f2/scipub-1260.pdf).

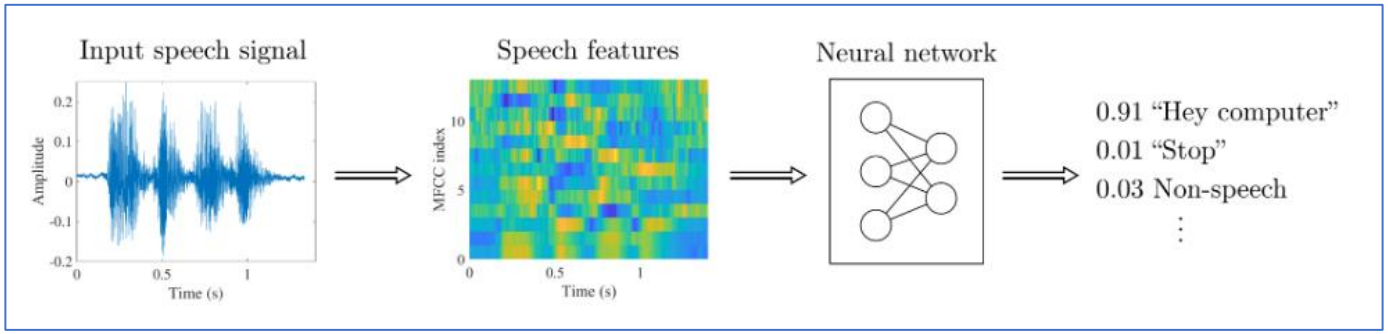
Ex.1029 / Page 103 of 347
AMAZON.COM, INC.

Claim 4	Accused Instrumentality(ies)
	In this paper we develop a simple but efficient semi-supervised learning strategy to mine useful WW-training data from large un-transcribed general-speech corpora. We employ a general far-field ASR model to automatically transcribe the unlabeled speech corpus, then select WW-specific utterances and add them to our training set. The DNN targets are obtained by force aligning the speech to the automatic transcripts. Since the ASR transcript may not be accurate, only the utterances with WW confidence score higher than a thresh-old θ_P are selected. It should be noted that, in the utterances mined here the WW may appear in the middle of the utterance, therefore the phonetic context may be different from the utterances initiated by the WW. In practice, our experiments show that such context mismatch have unobservable effects on WW spotting. This claim element is literally present, as described herein, but is also satisfied under the doctrine of equivalents because the structure and functionality provided by Amazon is being used to satisfy this claim limitation and therefore there is no vitiation of this element, and the Accused Instrumentalities incorporate structure and functionality that is not substantially different from the requirements of this limitation because they achieve substantially or exactly the same function, in substantially or exactly the same way as set forth in the claim element (as described herein), and achieves substantially or exactly the same result.

Ex.1029 / Page 104 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 5 of the '196 Patent

Claim 5	Accused Instrumentality(ies)
The method of claim 1, wherein determining the input speech is recognized as the preset call command further comprises: computing a log likelihood ratio of the input speech; and assuming the input speech to be the preset call command based on the computed log likelihood ratio.	Amazon's determining the input speech is recognized as the preset call command further comprises: computing a log likelihood ratio of the input speech; and assuming the input speech to be the preset call command based on the computed log likelihood ratio. For example, Amazon uses keyword-spotting algorithms as illustrated below. The input speech signal is first converted to a feature representation, such as MFCCs, which are fed to a neural network, and the output is the likelihood of each keyword.[82] 

[82] *See* KAIFI0001092 (*Introduction to Speech Processing*, https://speechprocessingbook.aalto.fi/Recognition/Wake-word_and_keyword_spotting.html).

Ex.1029 / Page 105 of 347
AMAZON.COM, INC.

Claim 5	Accused Instrumentality(ies)
	This claim element is literally present, as described herein, but is also satisfied under the doctrine of equivalents because the structure and functionality provided by Amazon is being used to satisfy this claim limitation and therefore there is no vitiation of this element, and the Accused Instrumentalities incorporate structure and functionality that is not substantially different from the requirements of this limitation because they achieve substantially or exactly the same function, in substantially or exactly the same way as set forth in the claim element (as described herein), and achieves substantially or exactly the same result.

Claim 6 of the '196 Patent

Claim 6	Accused Instrumentality(ies)
The method of claim 5, wherein the computed log likelihood ratio is a natural logarithm of a probability on the input speech for a phoneme model of the preset call command, and the computed confidence rate is another log likelihood ratio.	Amazon's computed log likelihood ratio is a natural logarithm of a probability on the input speech for a phoneme model of the preset call command, and the computed confidence rate is another log likelihood ratio. For example, Amazon employs Natural Language Understanding, which involves, for example, finding phonemes and statistically determining the most likely words and sentences.[83] Think the Alexa Voice Service as the brain of Alexa enabled devices and perform all the complex operations such as Automatic Speech Recognition (ASR) and Natural Language Understanding (NLU).

[83] *See* KAIFI0002097 (How Amazon Alexa works? Your Guide to Natural Language Processing (AI), https://towardsdatascience.com/how-amazon-alexa-works-your-guide-to-natural-language-processing-ai-7506004709d3).

Ex.1029 / Page 106 of 347
AMAZON.COM, INC.

Claim 6	Accused Instrumentality(ies)
	For Trevor Jackins, Marketing Specialist at NeoSpeech, "It tries to understand what you said by taking the voice data and breaking it down to a small sample of particular time duration mostly 10–20 ms. These data sets are further compared to pre-fed speech to decode what you said in each unit of your speech. The purpose here is to find **phoneme (the smallest unit of speech).** Then, the machine looks at the series of such phonemes and statistically determine the most likely words and sentences to spoke." For example, Amazon employs ASR systems that include both an acoustic model and a language model. The acoustic model converts audio into phonemes, the smallest units of speech; the language model encodes the probabilities of particular strings of words, which helps the ASR system decide between alternative interpretations of the same sequence of phonemes.[84] Like most ASR systems, the ones we use for live translation include both an acoustic model and a language model. The acoustic model converts audio into phonemes, the smallest units of speech; the language model encodes the probabilities of particular strings of words, which helps the ASR system decide between alternative interpretations of the same sequence of phonemes. This claim element is literally present, as described herein, but is also satisfied under the doctrine of equivalents because the structure and functionality provided by Amazon is being used to satisfy this claim limitation and therefore there is no vitiation of this element, and the Accused Instrumentalities incorporate structure and functionality that is not substantially different from the requirements of this limitation

[84] *See* KAIFI0002096 (Shirin Saleem, Roland Maas, *How Alexa's new Live Translation for conversations works* (Dec. 14, 2020), https://www.amazon.science/blog/how-alexas-new-live-translation-for-conversations-works)..

Ex.1029 / Page 107 of 347
AMAZON.COM, INC.

Claim 6	Accused Instrumentality(ies)
	because they achieve substantially or exactly the same function, in substantially or exactly the same way as set forth in the claim element (as described herein), and achieves substantially or exactly the same result.

Claim 7 of the '196 Patent

Claim 7	Accused Instrumentality(ies)
The method of claim 6, wherein said another log likelihood ratio is computed by subtracting, from the log likelihood ratio, a natural logarithm of a probability on the input speech for an anti-phoneme model of the preset call command.	Amazon's another log likelihood ratio is computed by subtracting, from the log likelihood ratio, a natural logarithm of a probability on the input speech for an anti-phoneme model of the preset call command. For example, Amazon uses Hidden Markov Models (HMMs). HMMs are statistical models widely used in ASR. They represent the probabilistic relationship between hidden states (phonemes or words) and observed acoustic features. HMM-based ASR systems align audio features with phonetic units to generate transcriptions.[85] Techniques used in ASR include: • **Hidden Markov Models (HMMs):** HMMs are statistical models widely used in ASR. They represent the probabilistic relationship between hidden states (phonemes or words) and observed acoustic features. HMM-based ASR systems align audio features with phonetic units to generate transcriptions.

[85] *See* KAIFI0007060 (Skill Upwards, *Top 5 AI technologies behind voice assistants like Siri and Alexa*, https://skillupwards.com/blog/ai-technologies-behind-voice-assistants-like-siri-and-alexa).

Ex.1029 / Page 108 of 347
AMAZON.COM, INC.

Claim 7	Accused Instrumentality(ies)
	This claim element is literally present, as described herein, but is also satisfied under the doctrine of equivalents because the structure and functionality provided by Amazon is being used to satisfy this claim limitation and therefore there is no vitiation of this element, and the Accused Instrumentalities incorporate structure and functionality that is not substantially different from the requirements of this limitation because they achieve substantially or exactly the same function, in substantially or exactly the same way as set forth in the claim element (as described herein), and achieves substantially or exactly the same result.

Claim 8 of the ’196 Patent

Claim 8	Accused Instrumentality(ies)
The method of claim 1, further comprising: determining, using a rule check unit, whether or not the recognized subsequent input speech corresponds to a target sentence.	Amazon further determines, using a rule check unit, whether or not the recognized subsequent input speech corresponds to a target sentence. For example, Amazon’s rule check unit is used to determine whether or not the recognized subsequent input speech corresponds to a target sentence. For example, Amazon describes its ASR model.[86]

[86] *See* KAIFI0006791 (Ariya Rastrow, *On-device speech processing makes Alexa faster, lower-bandwidth* (Jan. 25, 2022), https://www.amazon.science/blog/on-device-speech-processing-makes-alexa-faster-lower-bandwidth).

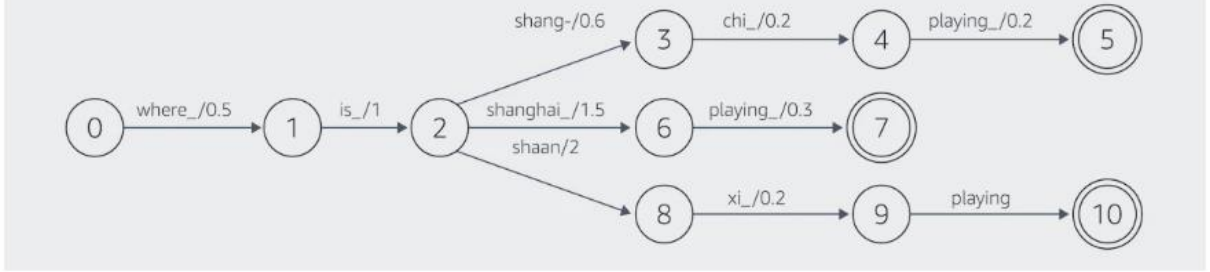
Ex.1029 / Page 109 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 8	Accused Instrumentality(ies)

System architecture

Our on-device ASR model takes in an acoustic speech signal and outputs a set of hypotheses about what the speaker said, ranked according to probability. We represent those hypotheses as a *lattice* — a graph whose edges represent recognized words and the probability that a given word follows from the previous one.



An example of a lattice representing ASR hypotheses.

For example, Amazon uses a speech recognizer.[87]

[87] *See* KAIFI0007093 *(Understand the AVS SpeechRecognizer*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/avs-speechrecognizer-concepts.html).

Ex.1029 / Page 110 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 8	Accused Instrumentality(ies)
	The `SpeechRecognizer` interface is the core interface of the Alexa Voice Service (AVS) and exposes directives and events for capturing and interacting with user speech. This page discusses the concepts and process flows for the `SpeechRecognizer` interface. SpeechRecognizer functionality Every user utterance leverages `SpeechRecognizer`, including the following interaction types between an Alexa Built-in device and AVS: • **User Speech Capture**: Captures user speech from an Alexa Built-in device. • **User Speech Prompting**: Prompts a user for more speech input when needed by Alexa to deliver an appropriate response. • **Interaction Initiation Communication**: Enables a device to inform AVS of how a user initiated an Alexa interaction, such as press-and-hold, tap-and-release, or voice-initiated/wake word enabled. See Device Form Factor and Alexa Interaction. • **ASR Profile Selection**: SpeechRecognizers chooses the appropriate Automatic Speech Recognition (ASR) profile for your product, which allows Alexa to understand user speech and respond with precision. See Automatic Speech Recognition (ASR) profile

Ex.1029 / Page 111 of 347
AMAZON.COM, INC.

State diagram

The following diagram illustrates state changes driven by `SpeechRecognizer` components. Boxes represent `SpeechRecognizer` states, and the connectors represent state transitions.

`SpeechRecognizer` has the following states:

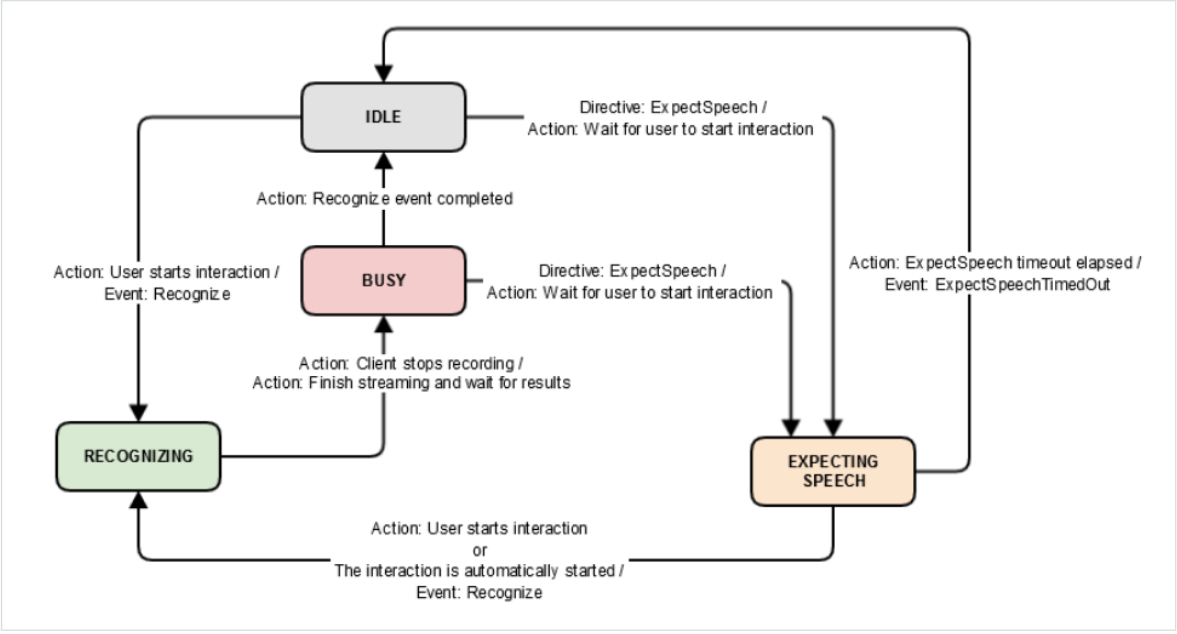
- **IDLE:** When not actively processing speech, the `SpeechRecognizer` is in an "idle" state. The idle state occurs under the following conditions:
  - Before capturing user speech.
  - Returning to an idle state after concluding a speech interaction with Alexa.
  - When an `ExpectSpeechTimedOut` event elapses.

  Note: In a multi-turn Alexa interaction, if Alexa requires more user speech input, `SpeechRecognizer` should transition from the idle state to the `ExpectSpeech` state without the user starting a new interaction.

- **RECOGNIZING:** When a user begins interacting with your client, specifically when the client streams captured audio to AVS, `SpeechRecognizer` should transition from the idle state to the `Recognize` state. `SpeechRecognizer` should remain in the `Recognize` until the client stops recording speech or finishes streaming, at which point your `SpeechRecognizer` component should transition from the `Recognize` state to the "busy" state.
- **BUSY:** While processing the speech request, `SpeechRecognizer` should be in the "busy" state. You cannot start another speech request until `SpeechRecognizer` transitions out of the busy state. From the *busy* state, `SpeechRecognizer` transitions to the idle state if Alexa processes and completes the request or to the `ExpectSpeech` state, if Alexa requires more speech input from the user.
- **EXPECTING SPEECH:** `SpeechRecognizer` should be in the `ExpectSpeech` state when Alexa requires more audio input from a user. From `ExpectSpeech`, `SpeechRecognizer` should transition to the `Recognize` when a user interaction occurs, or the interaction is automatically started on behalf of the user. It should transition to the idle state if Alexa detects no user interaction within the specified timeout window.

Ex.1029 / Page 112 of 347
AMAZON.COM, INC.

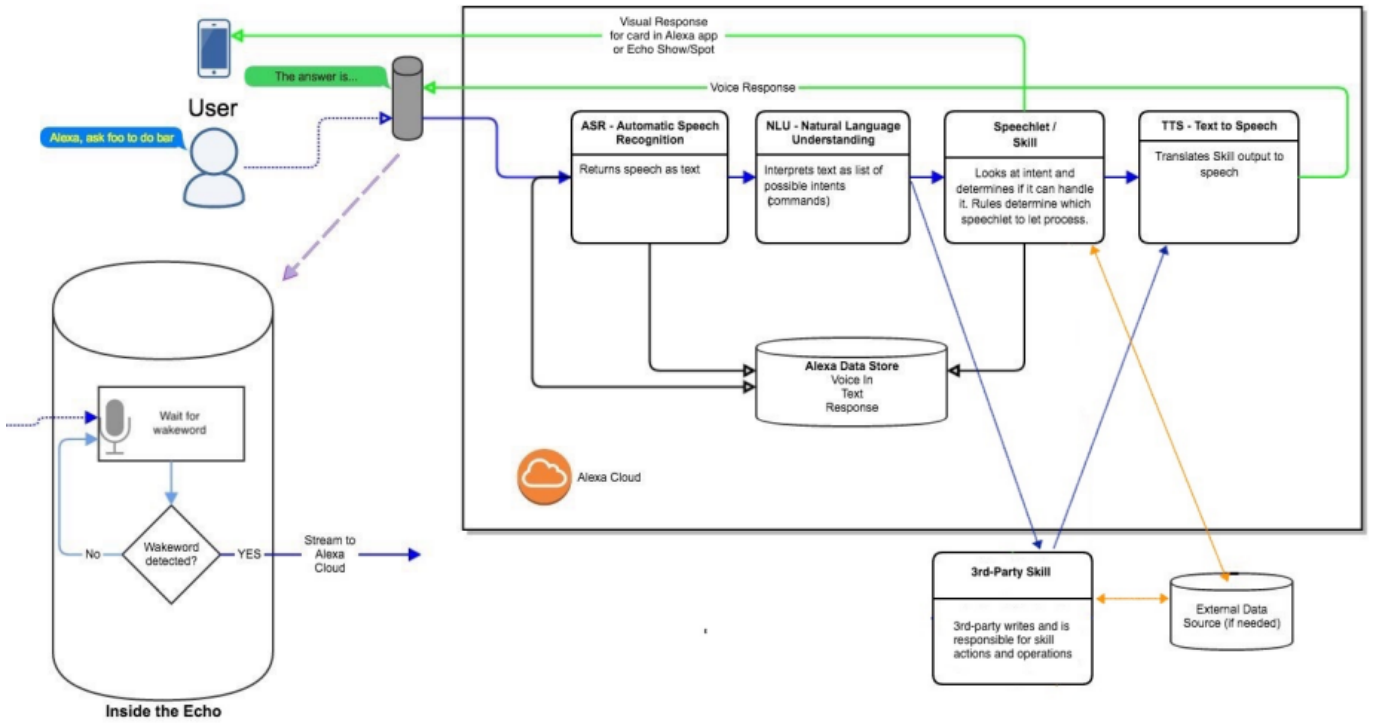
Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 8	Accused Instrumentality(ies)
	The following diagram illustrates the expected transitions among the four `SpeechRecognizer` states:  For example, Amazon uses a Speechlet/Skill that looks at intent and determines if it can handle a request.[88]

[88] *See* KAIFI0007380 *(White paper – Alexa Privacy and Data Handling Overview*, https://d1.awsstatic.com/product-marketing/A4B/White%20Paper%20-%20Alexa%20Privacy%20and%20Data%20Handling%20Overview.pdf)

Ex.1029 / Page 113 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 8	Accused Instrumentality(ies)
	 **Figure: Overview of Echo and the Alexa System**

Ex.1029 / Page 114 of 347
AMAZON.COM, INC.

Claim 8	Accused Instrumentality(ies)
	For example, Amazon uses cloud verification and encryption, automatic speech recognition (ASR), natural language understanding (NLU), text-to-speech (TTS), and data storage technologies on the recognized actual subsequent input speech command.[89]

Automatic Speech Recognition (ASR)

After wake word verification, your voice request undergoes Automatic Speech Recognition (ASR) which transcribes audio into text. Using our example request (*"Alexa, what's the weather in Seattle?"*), transcriptions could include one or several of the following:

- *"what is the weather in seattle"*
- *"watt is the weather in seattle"*
- *"what is the whether in seattle"*

Alexa then chooses to act on the transcription that is most likely to be correct and stores it in Amazon's secure cloud. You can view and delete it at any time at Alexa Privacy Settings or in the Alexa app (Settings > Alexa Privacy > Review Voice History). Voice recordings and transcriptions may be used to improve the ASR process. Learn how data helps Alexa improve.

Natural Language Understanding

After ASR comes Natural Language Understanding (NLU). NLU assigns meaning to the transcription producing an "intent" which is an instruction to the Alexa system to tell it what to respond to. In our example, the intent is identified as "weather" and sent to the designated data source where information is pulled for location (Seattle) and time (today). Intents may also be used to improve the NLU process.

Text-to-Speech (TTS)

Once the necessary information is gathered, your request uses text-to-speech (TTS) technologies to generate an audio file that becomes Alexa's response. The text from Alexa's response is stored so that you can review past answers at Alexa Privacy Settings or in the Alexa app. These responses may also be reviewed by Amazon to ensure the device is providing the most relevant answers to you, and to make sure the TTS system is translating text to speech to the best of its ability.

[89] *See* KAIFI0002023 *(Follow the journey of a voice request*, https://www.amazon.com/b/?node=23608618011).

Ex.1029 / Page 115 of 347
AMAZON.COM, INC.

Claim 8	Accused Instrumentality(ies)
	This claim element is literally present, as described herein, but is also satisfied under the doctrine of equivalents because the structure and functionality provided by Amazon is being used to satisfy this claim limitation and therefore there is no vitiation of this element, and the Accused Instrumentalities incorporate structure and functionality that is not substantially different from the requirements of this limitation because they achieve substantially or exactly the same function, in substantially or exactly the same way as set forth in the claim element (as described herein), and achieves substantially or exactly the same result.

Claim 9 of the '196 Patent

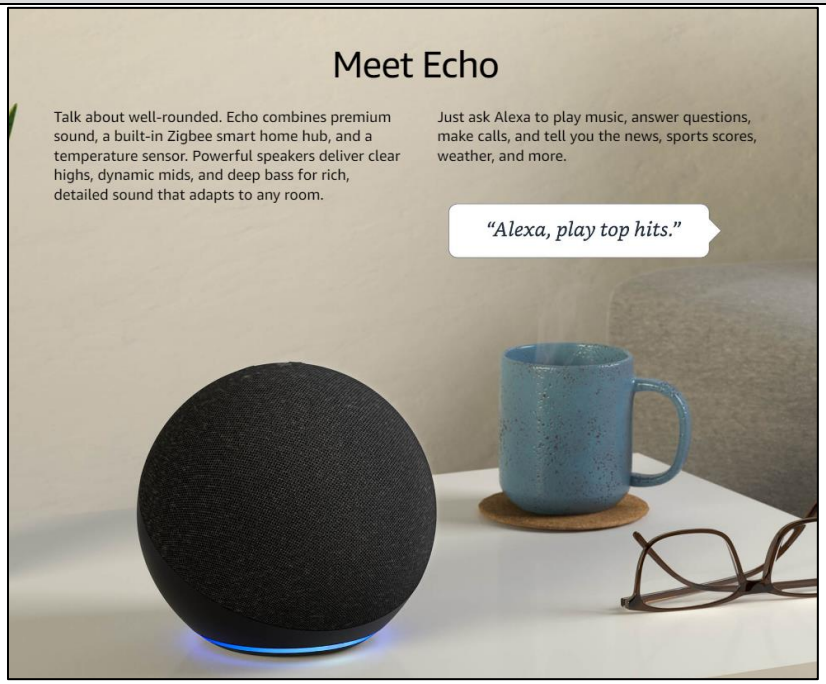
Claim 9	Accused Instrumentality(ies)
An apparatus for recognizing a speech of a speaker, the apparatus comprising:	Amazon provides an apparatus for recognizing a speech of a speaker, including the Accused Instrumentalities which include, for example, the Echo and Alexa devices (such as the Echo, Echo Pop, Echo Auto, Echo Show, Echo Flex, Echo Dot, Echo Studio, Echo Buds, and Echo Hub), Alexa Built-in Devices, and Amazon's cloud-side systems for performing wake word detection, Automatic Speech Recognition (ASR) and Alexa Voice Service (AVS). For example, Amazon provides Echo (4th Gen) with premium sound, smart home hub, and Alexa.[90]

[90] *See* KAIFI0001202 (*Echo (4th Gen) | With premium sound, smart home hub, and Alexa | Charcoal*, https://www.amazon.com/dp/B07XKF5RM3?ref=MARS_NAV_desktop_shop_dvc_smart_spkrs_Echo_Spkr).

Ex.1029 / Page 116 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)



More music in more rooms

Use your voice to play a song, artist, or genre through Amazon Music, Apple Music, Spotify, Pandora, and others. With multi-room music, you can fill your whole home with music, radio stations, podcasts, and Audible using compatible Echo devices in different rooms. You can even pair your Echo with compatible Fire TV devices to feel scenes come to life with home theater audio.

Ex.1029 / Page 117 of 347
AMAZON.COM, INC.

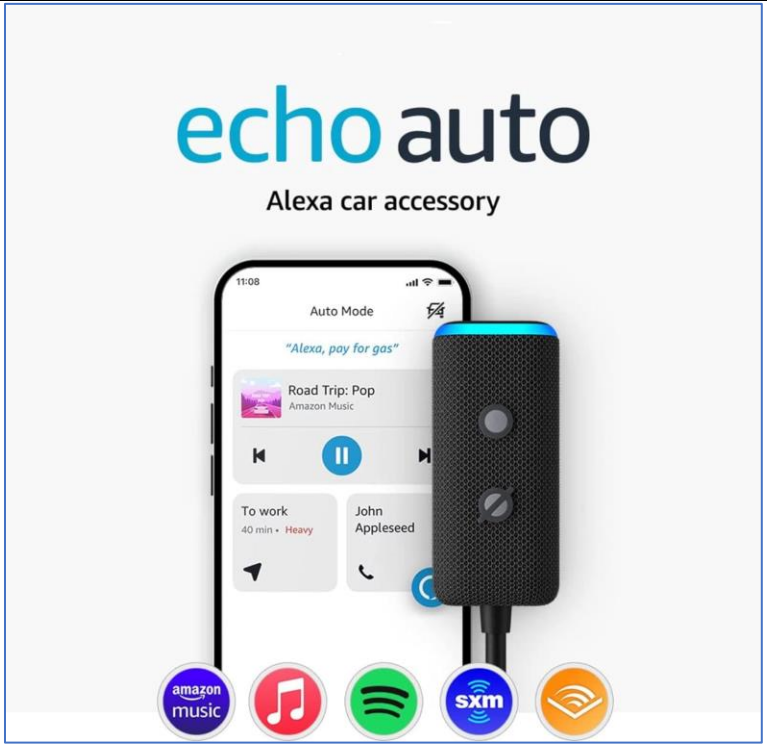
Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	 For example, Amazon also provides the Echo Auto.[91]

[91] *See* KAIFI0001979 *(Echo Auto (2nd Gen, 2022 release)*, https://www.amazon.com/All-New-Echo-Auto-release-Alexa/dp/B09X27YPS1?th=1).

Ex.1029 / Page 118 of 347
AMAZON.COM, INC.

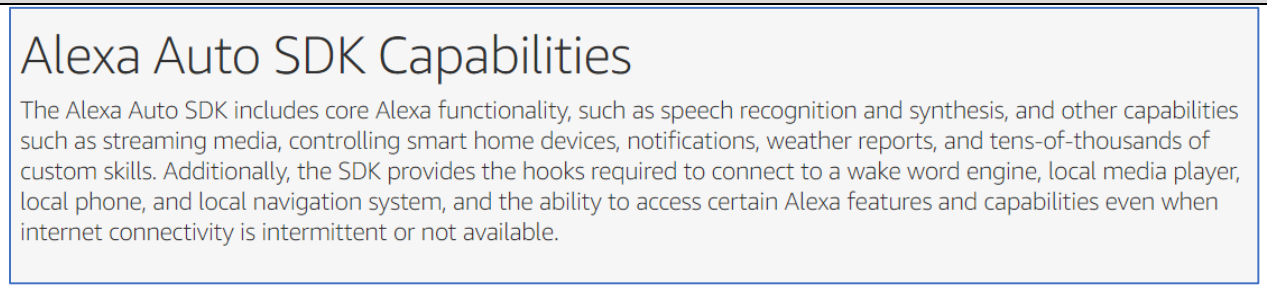
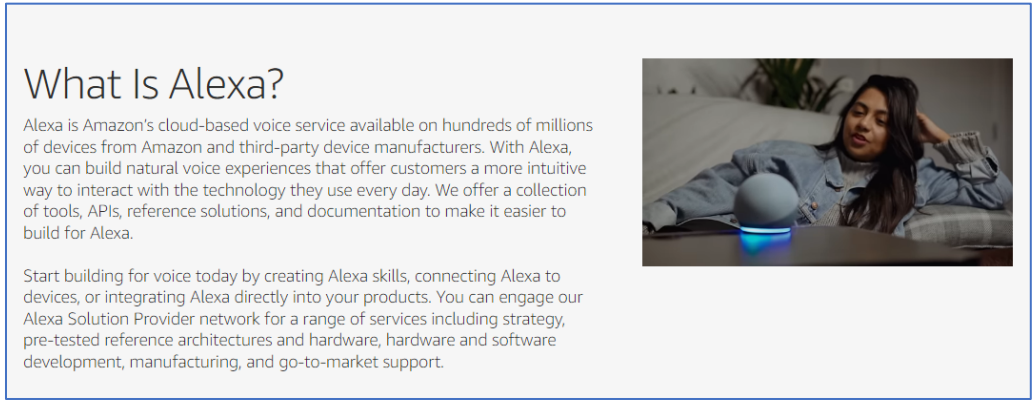
Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	 According to Amazon, the Echo Auto provides speech recognition and synthetis.[92]

[92] *See* KAIFI0001144 *(Alexa Auto Software Development Kit*, https://developer.amazon.com/en-US/alexa/alexa-auto/sdk).

Ex.1029 / Page 119 of 347
AMAZON.COM, INC.

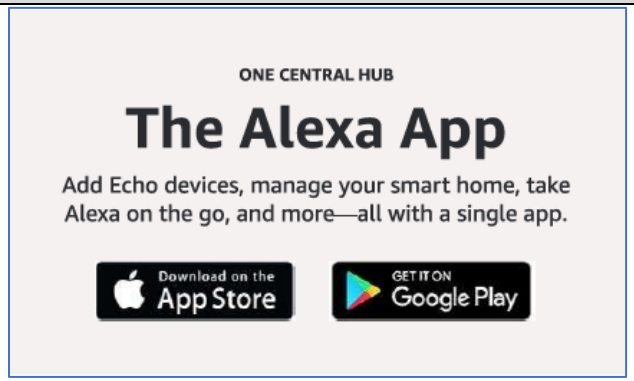
Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	 Amazon also provides Alexa.[93]  Amazon also provides the Alexa App.[94]

[93] *See* KAIFI0007144 (*What Is Alexa*, https://developer.amazon.com/en-US/alexa).

[94] *See* KAIFI0007048 (*The Alexa App*, https://www.amazon.com/Alexa-App/b?ie=UTF8&node=18354642011).

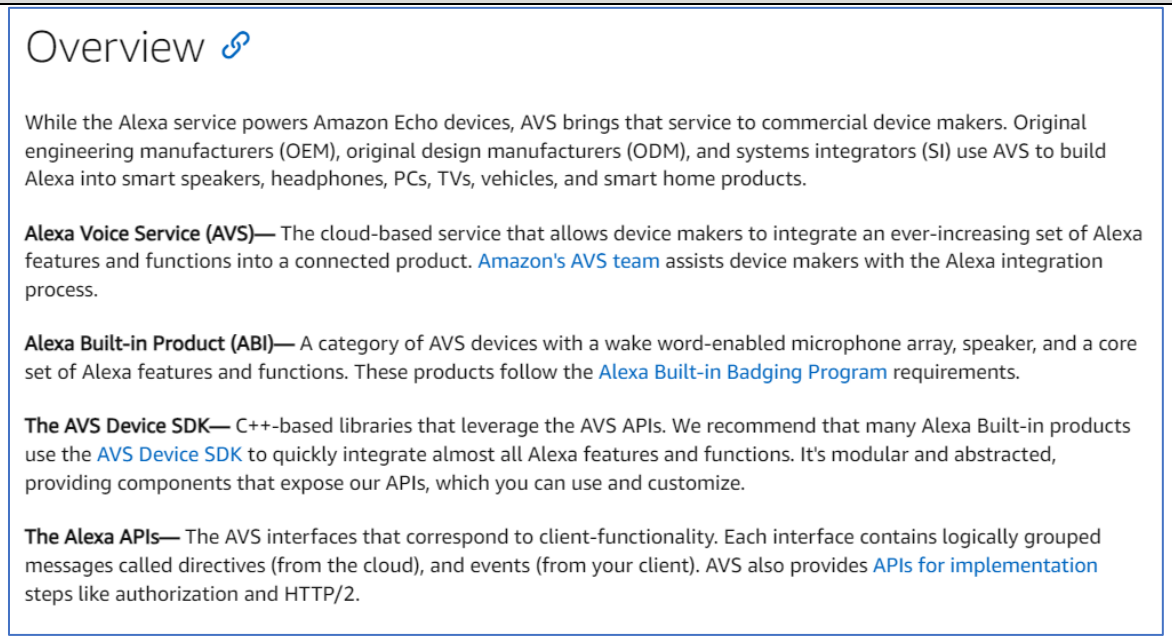
Ex.1029 / Page 120 of 347
AMAZON.COM, INC.

Claim 9	Accused Instrumentality(ies)
	 Amazon also provides Alexa Voice Service (AVS).[95]

[95] *See* KAIFI0007371 *(What is the Alexa Voice Service?*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/get-started-with-alexa-voice-service.html).

Ex.1029 / Page 121 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	 Amazon provides speech processing as both cloud-based and on-device.[96] At Amazon, we always look to invent new technology for improving customer experience. One technology we have been working on at Alexa is on-device speech processing, which has multiple benefits: a reduction in latency, or the time it takes Alexa to respond to queries; lowered bandwidth consumption, which is important on portable devices; and increased availability in in-car units and other applications where Internet connectivity is intermittent. On-device processing also enables the

[96] *See* KAIFI0006791 (Ariya Rastrow, Shehzad Mevawalla, *On-device speech processing makes Alexa faster, lower-bandwidth*, AMAZON SCIENCE (Jan. 25, 2022), https://www.amazon.science/blog/on-device-speech-processing-makes-alexa-faster-lower-bandwidth).

Ex.1029 / Page 122 of 347
AMAZON.COM, INC.

Claim 9	Accused Instrumentality(ies)
	fusion of the speech signal with other modalities, like vision, for features such as Alexa's natural turn-taking. In the last year, we've continued to build upon Alexa's on-device speech-processing capabilities. As a result of these inventions, we are launching a new setting that gives customers the option of having the audio of their Alexa voice requests processed locally, without being sent to the cloud. In the cloud, storage space and computational capacity are effectively unconstrained. To ensure accuracy, our cloud models can be large and computationally demanding. Executing the same functions on-device means compressing our models into less than 1% as much space — with minimal loss in accuracy. Moreover, in the cloud, the separate components of Alexa's speech-processing stack — automatic speech recognition (ASR), whisper detection, and speaker identification — run on separate server nodes with their own powerful processors. On-device, those functions have to share hardware not only with each other but with Alexa's other core device functions, such as music playback. Re-creating Alexa's speech-processing stack on-device was a massive undertaking. New methods for training small-footprint ASR models were part of the solution, but so were innovations in system design and hardware-software codesign. It was a joint effort across science and engineering teams over a span of years. Here's a quick overview of how it works. For example, Amazon describes the journey of a voice request.[97]

[97] *See* KAIFI0002023 *(Follow the journey of a voice request*, https://www.amazon.com/b/?node=23608618011).

Ex.1029 / Page 123 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	You can ask Alexa any question and get a response in a matter of seconds. If you're interested in the technology behind all of that magic, you've come to the right place. Here's a more detailed look at how Alexa works and the ways in which Echo devices and Alexa are designed to protect your privacy every step of the way. **The wake word** Think of the wake word as a verbal cue that makes things happen. When you have a request for Alexa, you first need to say your chosen wake word, which by default is "Alexa." Only after your Echo device detects the wake word is Alexa listening to your requests. When it comes to privacy, there should be no surprises. You'll always be able to tell when Alexa is listening to your request because a light indicator will appear on your Echo device or an audible tone will sound. Think of the "On the Air" signs that light up in television studios during a broadcast. These indicators notify you that your device has detected the wake word and Alexa is now processing your request. **Keyword spotting** Echo devices use built-in technology called 'keyword spotting' that matches spoken audio to the acoustic patterns of the wake word. Simply put, Echo devices are designed by default to detect only the sound waves of your chosen wake word, and everything else is ignored. Like water through a strainer, all other audio (people talking, taps running, birds chirping) passes through the device until the wake word is 'caught' and sent to Amazon's secure cloud, where meaning is assigned to your request. **Cloud verification and encryption** When your Echo device detects the wake word, it sends your request to Amazon's secure cloud where audio is reanalysed to verify the wake word was spoken. If this cloud software verification is unable to confirm the wake word was spoken, the Alexa system stops processing the audio. If the wake word is verified (or Alexa is activated using the Action button), your request will be processed using several sophisticated algorithms that allow Alexa to respond appropriately. All of your interactions with Alexa are encrypted in transit throughout this process. **Automatic Speech Recognition (ASR)** After wake word verification, your voice request undergoes Automatic Speech Recognition (ASR) which transcribes audio into text. Using our example request (***"Alexa, what's the weather in sydney?"***), transcriptions could include one or several of the following:

Ex.1029 / Page 124 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	For example, Amazon provides and uses Alexa Built-in Devices, which are powered by Alexa Voice Service (AVS).[98]

What are Alexa Built-in Devices?



The world's leading OEMs offer Alexa Built-in devices that let customers talk directly with Alexa through a microphone and speaker. Alexa turns your device into a smart device, and expands its capabilities. Customers can ask Alexa for help with everyday tasks and entertainment, and control Alexa connected devices or their own device.

Learn About the Underlying Technology

Alexa Built-in devices are powered by **Alexa Voice Service (AVS)**. The AVS team manages the services and infrastructure required for Alexa experiences, and provides a suite of device APIs, SDKs, and documentation. Developers may also work with third-party solution providers to utilize a qualified AVS solution.

Start Developing with Solution Providers

Our global network of **Alexa Solution Providers** work together to help you develop high-quality experiences, accelerate your development, and reduce your overall development cost. The earlier you engage a Solution Provider, the better they can assist you.

[98] *See* KAIFI0007733 *(Alexa Built-in Devices, https://developer.amazon.com/en-US/alexa/devices/alexa-built-in).*

Ex.1029 / Page 125 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	 For example, Amazon's AVS Integration is a feature of AWS IoT Core that enables Alexa Voice to be produced on any type of connected device, and shifts various Alexa Voice functions from a physical device to a new virtual Built-in device in the cloud.[99]

[99] *See* KAIFI0007818 *(New Alexa Voice Service Integration for AWS IoT Core cost-effectively brings Alexa Voice to any type of connected device*, https://aws.amazon.com/about-aws/whats-new/2019/11/new-alexa-voice-integration-for-aws-iot-core-cost-effectively-brings-alexa-voice-to-any-connected-device/*)*.

Ex.1029 / Page 126 of 347
AMAZON.COM, INC.

Claim 9	Accused Instrumentality(ies)
	New Alexa Voice Service Integration for AWS IoT Core cost-effectively brings Alexa Voice to any type of connected device Posted On: Nov 25, 2019 Alexa Voice Service (AVS) Integration is a new feature of AWS IoT Core that enables Alexa Voice to be produced on any type of connected device. AVS Integration for IoT Core reduces the cost of producing Alexa Built-in devices by up to 50% by offloading compute and memory intensive audio workloads to the cloud. With this reduction in production cost, customers can now cost-effectively build new categories of differentiated voice-enabled products such as light switches, thermostats, and small appliances. This allows consumers to talk directly to Alexa in new parts of their home, office, or hotel rooms for a truly ambient experience. Alexa Voice Service (AVS) Integration for IoT Core shifts media retrieval, audio decoding, audio mixing, and state management from a physical device to a new virtual Alexa Built-in device in the cloud. This lowers the hardware requirements of integrating AVS on a device from 50 MB of RAM to 1MB of RAM and ARM Cortex 'A' class microprocessors to ARM Cortex 'M' class microcontrollers. Once a device is connected to AWS IoT Core, customers can use a new set of AWS IoT-reserved MQTT topics to transfer audio messages between their devices and AVS. The AVS Integration for AWS IoT Core is available in all AWS regions where AWS IoT Core is available other than China (Beijing and Ningxia), Asia Pacific (Hong Kong) and Middle East (Bahrain). See the AWS Region Table for the current list of regions for AWS IoT Core. To learn more about how to get started, view the AVS Integration for IoT Core Blog Post.
a processor,	The Accused Instrumentalities comprise a processor. For example, the Echo is powered by Amazon's Neural Edge processor. The new Echo is powered by Amazon's first-generation AZ1 Neural Edge processor—an all-new silicon module that's purpose-built for accelerating machine learning applications. With AZ1,

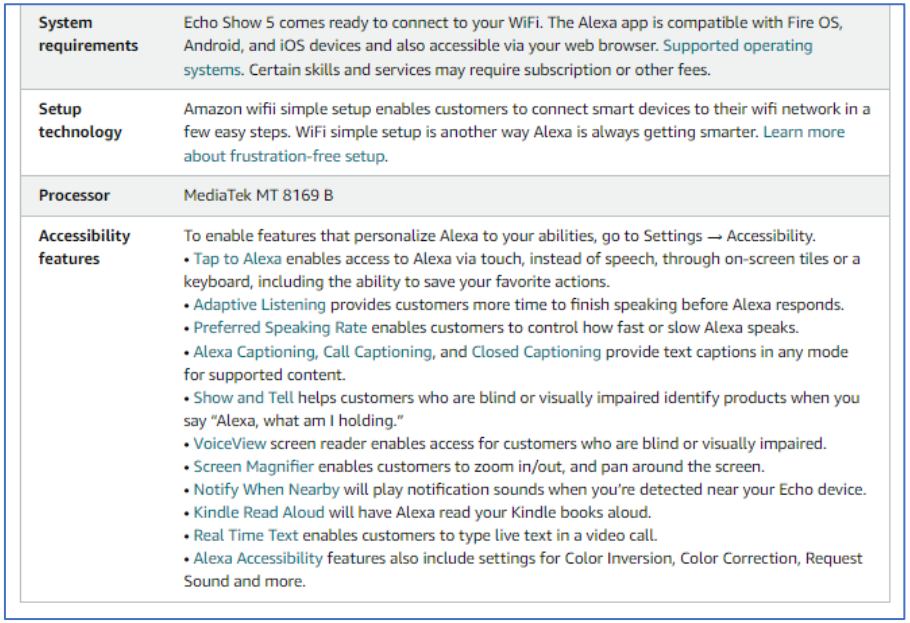
Ex.1029 / Page 127 of 347
AMAZON.COM, INC.

Claim 9	Accused Instrumentality(ies)
	powerful inference engines can run quickly on the edge—starting with an all-neural speech recognition model that will process requests faster, making Alexa even more responsive.[100] For example, the Echo Pop and Echo Show are powered by Amazon's Neural Edge processor. Echo Pop is a new addition to the Echo family with a loveable semi-sphere form factor, and new Lavender Bloom and Midnight Teal color options for just $39.99. It features a custom-designed front-facing directional speaker to deliver full sound that's great for bedrooms, dorm rooms, apartments, or any small space in the home. Simply ask Alexa to read audiobooks, track the record of a favorite sports team, control smart lights and plugs, or reorder household essentials. Echo Pop is powered by the Amazon AZ2 Neural Edge processor and comes with eero Built-in, which enables the device to add up to 1,000 square feet of coverage to an existing eero wifi network, packing even more value into its compact size.[101] Echo Show 5 combines the utility of Alexa, with the convenience of a compact screen so customers can watch news clips, check their Ring doorbell camera, view shopping lists, or easily make video calls to friends and family. The new Echo Show 5 is 20 percent faster than the previous generation, and features an entirely new speaker system, doubling the bass and delivering even clearer sound when listening to music, Audible, video content, or a podcast. It comes with a completely reengineered microphone array and the AZ2 Neural Edge processor. Echo Show 5 and Echo Pop both

[100] *See* KAIFI0004450 *(Introducing the All-New Echo Family—Reimagined, Inside and Out* (Sep. 24, 2020), https://press.aboutamazon.com/2020/9/introducing-the-all-new-echo-family-reimagined-inside-and-out).

[101] *See* KAIFI0001352 *(Amazon Introduces Four All-New Echo Devices; Sales of Alexa-Enabled Devices Surpass Half a Billion* (May 17, 2023), https://press.aboutamazon.com/2023/5/amazon-introduces-four-all-new-echo-devices-sales-of-alexa-enabled-devices-surpass-half-a-billion).

Ex.1029 / Page 128 of 347
AMAZON.COM, INC.

Claim 9	Accused Instrumentality(ies)
	support Matter, making it easy to connect and control compatible smart home products across various brands.[102] For example, the Echo Show 5 (3rd Gen) comprises a MediaTek MT 8169 B processor.[103] 

System requirements	Echo Show 5 comes ready to connect to your WiFi. The Alexa app is compatible with Fire OS, Android, and iOS devices and also accessible via your web browser. Supported operating systems. Certain skills and services may require subscription or other fees.
Setup technology	Amazon wifii simple setup enables customers to connect smart devices to their wifi network in a few easy steps. WiFi simple setup is another way Alexa is always getting smarter. Learn more about frustration-free setup.
Processor	MediaTek MT 8169 B
Accessibility features	To enable features that personalize Alexa to your abilities, go to Settings → Accessibility. • Tap to Alexa enables access to Alexa via touch, instead of speech, through on-screen tiles or a keyboard, including the ability to save your favorite actions. • Adaptive Listening provides customers more time to finish speaking before Alexa responds. • Preferred Speaking Rate enables customers to control how fast or slow Alexa speaks. • Alexa Captioning, Call Captioning, and Closed Captioning provide text captions in any mode for supported content. • Show and Tell helps customers who are blind or visually impaired identify products when you say "Alexa, what am I holding." • VoiceView screen reader enables access for customers who are blind or visually impaired. • Screen Magnifier enables customers to zoom in/out, and pan around the screen. • Notify When Nearby will play notification sounds when you're detected near your Echo device. • Kindle Read Aloud will have Alexa read your Kindle books aloud. • Real Time Text enables customers to type live text in a video call. • Alexa Accessibility features also include settings for Color Inversion, Color Correction, Request Sound and more.

[102] *See id.*

[103] *See* KAIFI0001322 *(All-new Echo Show 5 (3rd Gen, 2023 release) | Smart display with 2x the bass and clearer sound | Charcoal,* https://www.amazon.com/All-new-Echo-Show-5/dp/B09B2SBHQK).

Ex.1029 / Page 129 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	For example, the Echo Show 15 comprises an Amlogic Pop1 octa-core SoC with Amazon AZ2 neural network engine processor.[104]

System requirements	Echo Show 15 comes ready to connect to your wifi. The Alexa app is compatible with Fire OS, Android, and iOS devices and also accessible via your web browser. Learn more about supported operating systems. Certain skills and services may require subscription or other fees.
Set up technology	Amazon wifi simple setup enables customers to connect smart devices to their wifi network in a few easy steps. Wifi simple setup is another way Alexa is always getting smarter. Learn more about frustration-free setup.
Processor	Amlogic Pop1 octa-core SoC with Amazon AZ2 neural network engine
Sensor	ALS RGB + Accelerometer
Included in the box	Echo Show 15, power adapter (30W) and 5 ft. cord (white), wall mount, anchors, screws, mounting template, and quick start guide. Setup requires wifi. Device can be wall mounted (electric drill required) or placed on a stand (sold separately).
Warranty and service	1-year limited warranty and service included. Optional 1-year, 2-year, and 3-year extended warranty available for U.S. customers sold separately. Use of Echo Show 15 is subject to the terms found here. Use of Fire TV on Echo Show 15 is subject to all the terms found here.
Privacy features	Wake word technology, streaming indicators, microphone/camera off button, camera shutter, the ability to view and delete your voice recordings, and more. Visit the Alexa Privacy Hub to explore how Alexa and Echo devices are designed to protect your privacy.

[104] *See* KAIFI0001990 *(Echo Show 15 | Full HD 15.6" smart display with Alexa and Fire TV built in | Remote included*, https://www.amazon.com/dp/B0BHZX3TYV).

Ex.1029 / Page 130 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	For example, the Echo Auto comprises a MediaTek MT7697 processor.[105] The Echo Auto is powered by a Mediatek MT7697 processor and has a Bluetooth radio for connecting to your car's audio system and to your phone and the mobile Alexa app. The device itself doesn't contain any cellular circuitry; instead it taps into your phone's data plan to access the Amazon cloud servers. This won't be an issue for users that have an unlimited mobile data plan, but if you have a fixed plan you'll want to keep an eye on your data usage. You can connect the Echo Auto directly to your car's USB port or use the 12-volt adapter. If your car's audio system doesn't support Bluetooth, use the 3.5mm audio cable. For example, Alexa's ASR speech-processing stack uses a processor for speech recognition.[106] Moreover, in the cloud, the separate components of Alexa's speech-processing stack — automatic speech recognition (ASR), whisper detection, and speaker identification — run on separate server nodes with their own powerful processors. On-device, those functions have to share hardware not only with each other but with Alexa's other core device functions, such as music playback.

[105] *See* KAIFI0001231 (John R. Delaney, *Amazon Echo Auto Review* (Nov. 7, 2019), https://www.pcmag.com/reviews/amazon-echo-auto).

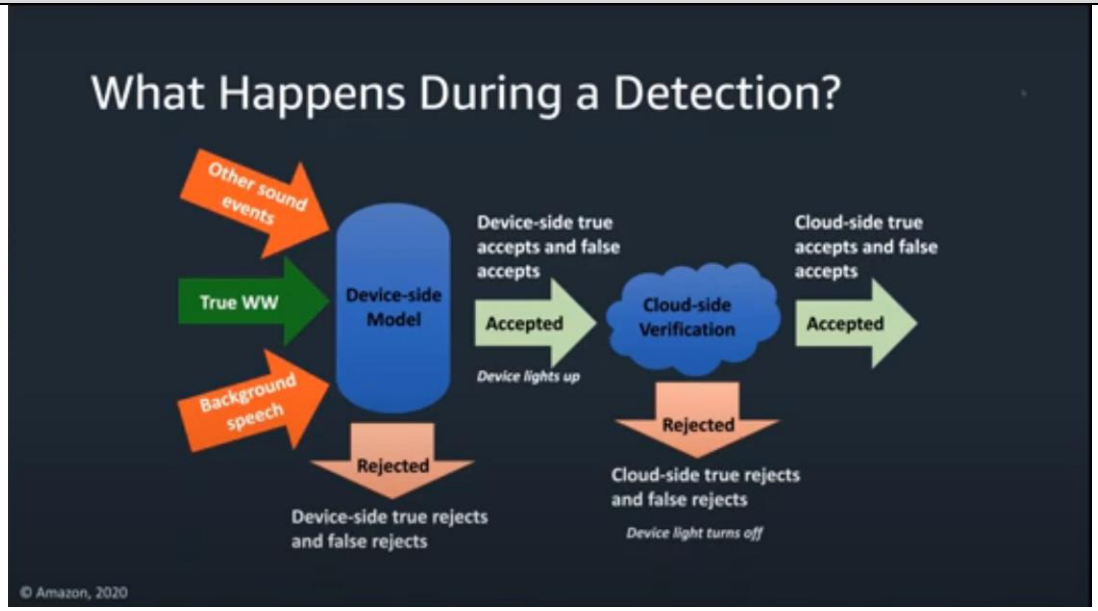
[106] *See* KAIFI0006791 (Ariya Rastrow, Shehzad Mevawalla, *On-device speech processing makes Alexa faster, lower-bandwidth*, AMAZON SCIENCE (Jan. 25, 2022), https://www.amazon.science/blog/on-device-speech-processing-makes-alexa-faster-lower-bandwidth).

Ex.1029 / Page 131 of 347
AMAZON.COM, INC.

Claim 9	Accused Instrumentality(ies)
	For example, Amazon uses cloud verification and encryption. [107] **Cloud verification and encryption** When your Echo device detects the wake word, it sends your request to Amazon's secure cloud where audio is reanalysed to verify the wake word was spoken. If this cloud software verification is unable to confirm the wake word was spoken, the Alexa system stops processing the audio. If the wake word is verified (or Alexa is activated using the Action button), your request will be processed using several sophisticated algorithms that allow Alexa to respond appropriately. All of your interactions with Alexa are encrypted in transit throughout this process. This claim element is literally present, as described herein, but is also satisfied under the doctrine of equivalents because the structure and functionality provided by Amazon is being used to satisfy this claim limitation and therefore there is no vitiation of this element, and the Accused Instrumentalities incorporate structure and functionality that is not substantially different from the requirements of this limitation because they achieve substantially or exactly the same function, in substantially or exactly the same way as set forth in the claim element (as described herein), and achieves substantially or exactly the same result.
a preset call command recognition unit using the processor; and	Amazon's Accused Instrumentalities comprise a preset call command recognition unit. For example, the Accused Instrumentalities include a combination of hardware, processing, and software functions both on devices and on the cloud server to process speech.[108]

[107] *See* KAIFI0002023 *(Follow the journey of a voice request*, https://www.amazon.com/b/?node=23608618011).
[108] *See* KAIFI0007814 *(The Science Behind Alexa's Wake Word Detection*, https://www.youtube.com/watch?v=kYrw2cC9dvs).

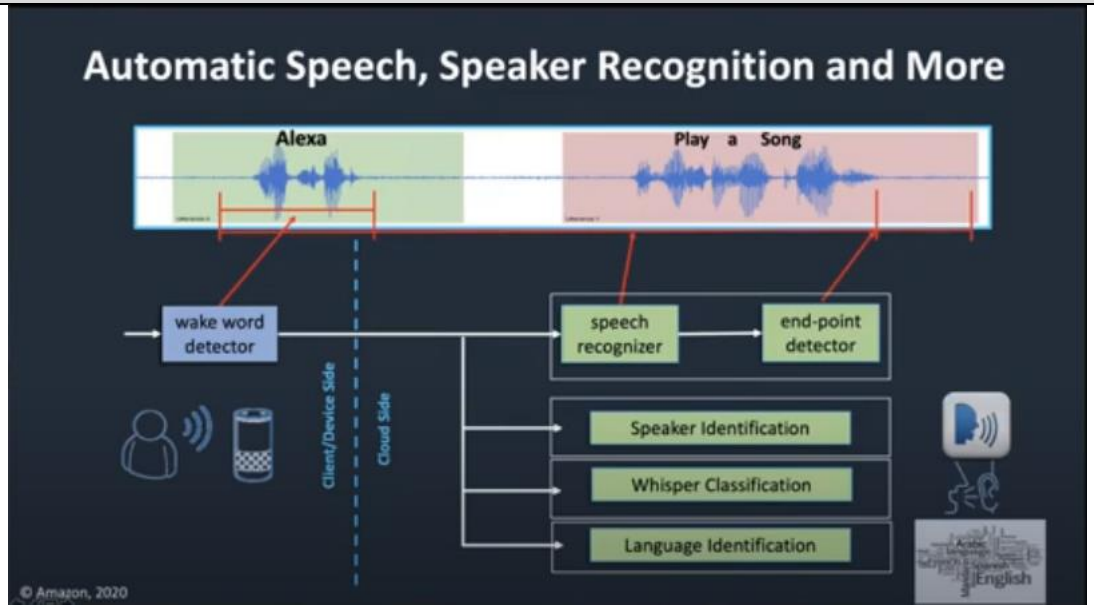
Ex.1029 / Page 132 of 347
AMAZON.COM, INC.

Claim 9	Accused Instrumentality(ies)
	 This is further illustrated below.[109]

[109] *See* KAIFI0007816 *(How Alexa Uses Automatic Speech Recognition*, https://www.youtube.com/watch?v=8drK6Oyv1a0&t=296s).

Ex.1029 / Page 133 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	 For example, according to Amazon, it provides on-device and cloud-based automatic speech recognition (ASR).[110] Our on-device ASR model takes in an acoustic speech signal and outputs a set of hypotheses about what the speaker said, ranked according to probability. We represent those hypotheses as a *lattice* —

[110] *See* KAIFI0006791 (Ariya Rastrow, Shehzad Mevawalla, *On-device speech processing makes Alexa faster, lower-bandwidth*, AMAZON SCIENCE (Jan. 25, 2022), https://www.amazon.science/blog/on-device-speech-processing-makes-alexa-faster-lower-bandwidth).

Ex.1029 / Page 134 of 347
AMAZON.COM, INC.

Claim 9	Accused Instrumentality(ies)
	a graph whose edges represent recognized words and the probability that a given word follows from the previous one. With cloud-based ASR, encrypted audio streams to the cloud in small snippets called "frames". With on-device ASR, only the lattice is sent to the cloud, where a large and powerful neural language model reranks the hypotheses. The lattice can't be sent until the customer has finished speaking, as words later in a sequence can dramatically change the overall probability of a hypothesis. The model that determines when the customer has finished speaking is called an end-pointer. End-pointers offer a natural trade-off between accuracy and latency: an aggressive end-pointer will initiate speech processing earlier, but it might cut the speaker off prematurely, resulting in a poor customer experience. On the device, we in fact run two end-pointers: One is a speculative end-pointer that we have tuned to be about 200 milliseconds faster than the final end-pointer, so we can initiate downstream processing — such as natural-language understanding (NLU) — ahead of the final end-pointed ASR result. In exchange for speed, however, we trade off a little accuracy. The final end-pointer takes longer to make a decision but is more accurate. In cases in which the first end-pointer cuts speech off too early, the final end-pointer sends a revised lattice and the instruction to reset downstream processing. In the large majority of cases, however, the aggressive end-pointer is correct, which reduces user-perceived latency, since downstream tasks are initiated earlier. Another aspect of ASR that had to move on-device is *context awareness*. When computing the probabilities in a lattice, the ASR model should, for instance, give added weight to otherwise uncommon names that happen to be in the customer's address book or the names the customer has assigned to household devices.

Ex.1029 / Page 135 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	For example, Amazon uses built-in technology for keyword spotting.[111] Echo devices use built-in technology called "keyword spotting" that matches spoken audio to the acoustic patterns of the wake word. Simply put, Echo devices are designed by default to detect only the sound waves of your chosen wake word, and everything else is ignored. Like water through a strainer, all other audio (people talking, faucets running, birds chirping) passes through the device until the wake word is "caught" and sent to Amazon's secure cloud, where your request is reanalyzed to verify the wake word was spoken. You can choose to opt-in to features like Alexa Emergency Assist that allow your device to detect more than your chosen wake word—for example, the sound of smoke alarms—but you would need to update your settings to do so. Want to keep nerding out? Follow the journey of a voice request. For example, according to Amazon, Echo devices use wake words.[112] It starts with detecting the "wake word" that triggers Alexa to start listening to the subsequent spoken words from the user. Wake word detection employs deep learning technology running on-device to recognize the wake word the user has chosen. Far-field automatic speech recognition (ASR) in the Amazon Web Services (AWS) cloud then converts into text the audio following the wake word, and determines when the user has stopped talking to Alexa.

[111] *See* KAIFI0002092 *(How Alexa works: Wake word*, https://www.amazon.com/b?ie=UTF8&node=23608571011).

[112] *See* KAIFI0004342 *(How Our Scientists are Making Alexa Smarter* (Mar. 29, 2018), https://www.aboutamazon.com/news/devices/how-our-scientists-are-making-alexa-smarter).

Ex.1029 / Page 136 of 347
AMAZON.COM, INC.

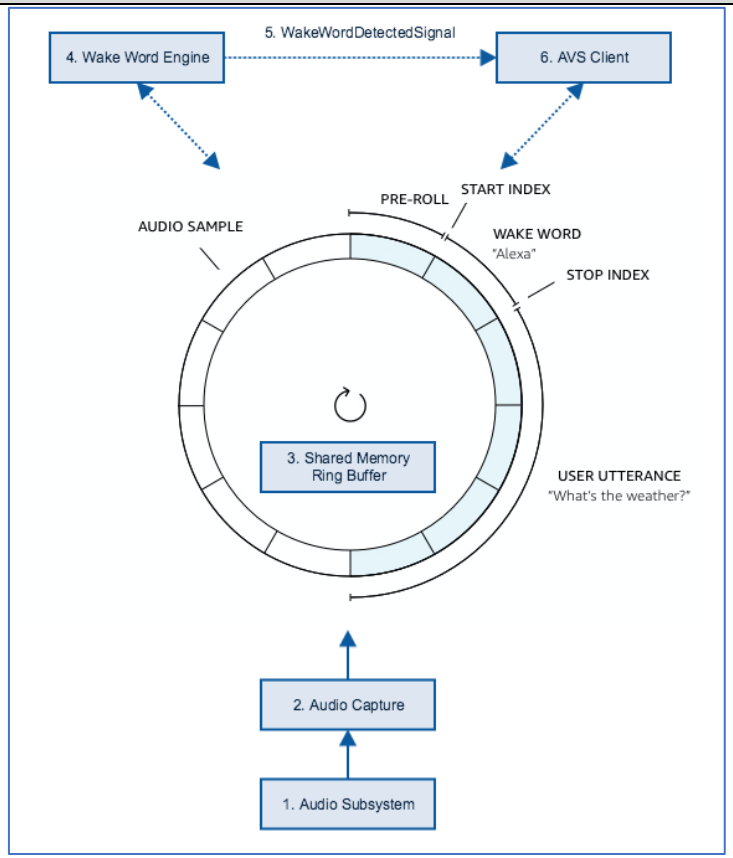
Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	For example, according to Amazon, it uses a wake word verification engine.[113] Voice-initiated devices begin to stream user speech to AVS when the wake word engine detects a spoken wake word, such as "Alexa."

[113] *See* KAIFI0007141 *(Wake Word Verification Requirements*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/implement-ww-verification.html).

Ex.1029 / Page 137 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	

Ex.1029 / Page 138 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	4. **Wake Word Engine** - The component that reads the audio samples written to the shared memory ring buffer and analyzes the audio samples for occurrences of the wake word. The procedure of reading and processing audio samples occurs in a loop. For example, the loop reads N samples starting at location `0`; then analyzes the samples for an occurrence of the wake word. In the next iteration, it reads and analyzes the next N samples. When an occurrence of the wake word is detected, the wake word engine must identify two locations: 1) the start of the wake word, and 2) the end of the wake word. These values are stored as `startIndexInSamples` and `endIndexInSamples`. For example, according to Amazon, the wake word engine uses wake words.[114] Wake words The list of wake words informs Alexa of the possible valid wake words that a device might be set to listen for through the `SetWakeWords`, `WakeWordsReport`, and `WakeWordsChanged` messages. Currently, the only wake word available for Alexa Built-in devices is `ALEXA`, which applies to every possible locale for a device. Therefore, specify `ALEXA` in the global `DEFAULT` scope. According to Amazon, the wake word engine uses a processor for speech recognition.[115]

[114] *See* KAIFI0007093 *(Understand the AVS SpeechRecognizer*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/avs-speechrecognizer-concepts.html).

[115] *See* KAIFI0006993 (Bob Mars, *Sizing Up CPU, Memory, and Storage for Your Alexa Built-in Device* (Sept. 13, 2018), https://developer.amazon.com/en-US/blogs/alexa/post/2a32d792-d471-4136-8262-79962a2b4d72/cpu-memory-and-storage-for-alexa-built-in-device.html).

Ex.1029 / Page 139 of 347
AMAZON.COM, INC.

Claim 9	Accused Instrumentality(ies)

Each solution has varying degrees of optimization as illustrated in the Level of Optimization (LoO) column. It also shows an example of a very optimized solution on a considerably less powerful ARM Cortex-R4 microcontroller for illustration purposes, but where the Alexa wake word engine runs exclusively on the digital signal processor front-end of the device to offload the host processor.

Table 2 and Figure 1 below illustrate how effectively the various solutions utilize the host processor and memory.

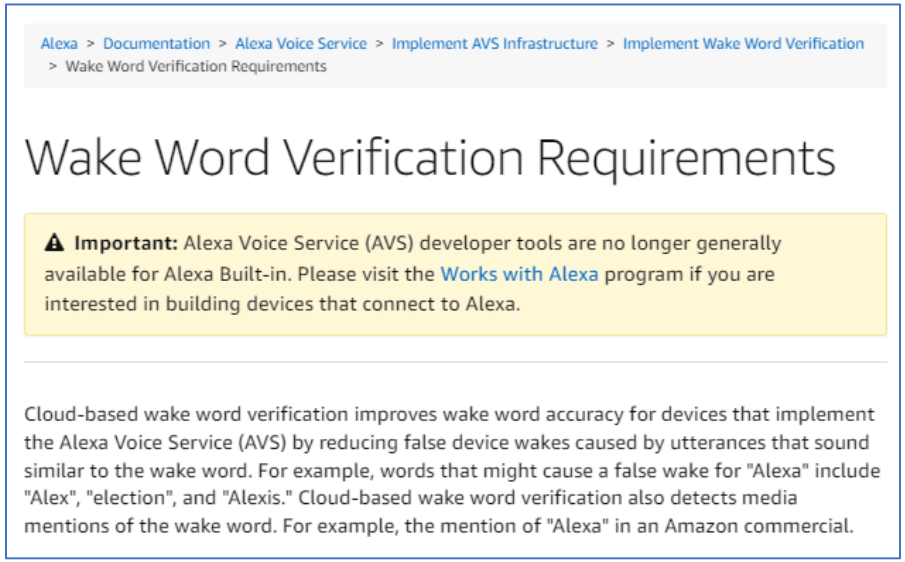
Arch	Host Processor(s)	Vendor	LoO 1-5	Gain	eCycles 1GHz Alexa	eCycles 1GHz Music
MIPS32	MIPS 24KEc	SI_1	4	2	31	24
MIPS32	MIPS 24KEc	SI_1	4	2	24	24
MIPS32	MIPS 24KEc	SI_2	4	2	266	183
MIPS32	Xburst	SI_2	4	2	172	128
MIPS32	Xburst	SI_3	3	2	90	90
ARMv7-A	Cortex-A7	V_2	2	1	50	46
ARMv7-A	Cortex-A7+Neon+M4	Dev Kit	1	1	78	75
ARMv7-A	Cortex-A7+Neon/VFPU	SI_1	4	1.5	106	104
ARMv8-A	Cortex A53	SI_3	3	1.5	70	65
ARMv8-A	Cortex A53	Dev Kit	1	1	52	48

eCycle: effective Cycles; MCPS normalized to 1 GHz with Gain multiplier
Gain: Cycle gain multiplier due to additional memory; 1.5=50%; 2=100%

Table 2- Normalized Headroom on Alexa Built-in Devices

Ex.1029 / Page 140 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	For example, according to Amazon, the AVS wake word engine verification has an associated ASR profile.[116] 

[116] *See* KAIFI0007141 *(Wake Word Verification Requirements*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/implement-ww-verification.html).

Ex.1029 / Page 141 of 347
AMAZON.COM, INC.

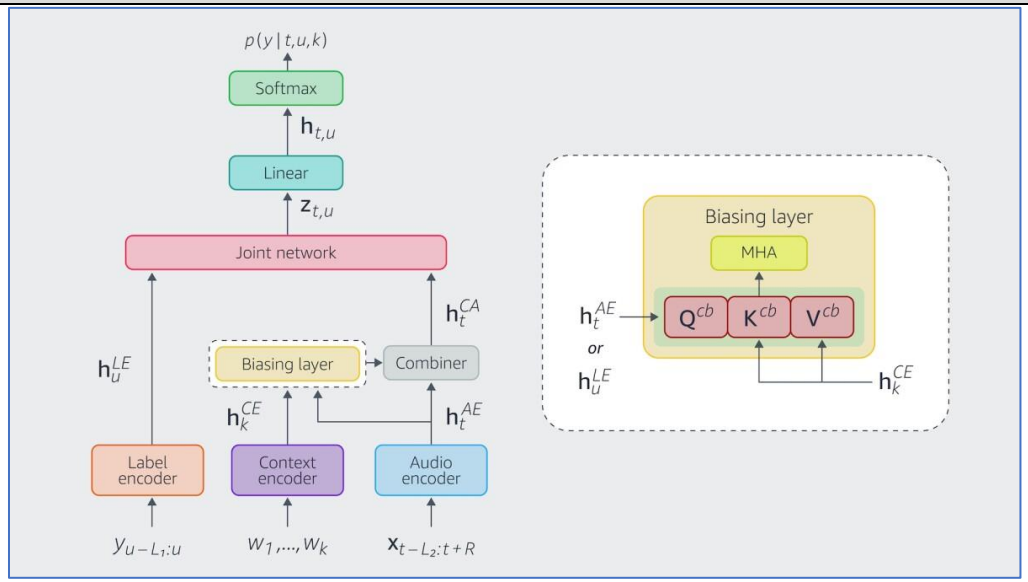
Claim 9	Accused Instrumentality(ies)
	Payload Parameters Parameter / Description / Type profile — Identifies the Automatic Speech Recognition (ASR) profile associated with your product. AVS supports three distinct ASR profiles optimized for user speech from varying distances. Accepted values: `CLOSE_TALK`, `NEAR_FIELD`, `FAR_FIELD`. — string For example, according to Amazon, its ASR network is shown in the below diagram.[117]

[117] *See* KAIFI0006791 (Ariya Rastrow, Shehzad Mevawalla, *On-device speech processing makes Alexa faster, lower-bandwidth*, AMAZON SCIENCE (Jan. 25, 2022), https://www.amazon.science/blog/on-device-speech-processing-makes-alexa-faster-lower-bandwidth).

Ex.1029 / Page 142 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)



For example, Amazon provides automatic speech recognition, whisper detection, and speaker identification.[118]

> Moreover, in the cloud, the separate components of Alexa's speech-processing stack — automatic speech recognition (ASR), whisper detection, and speaker identification — run on separate server nodes with their own powerful processors. On-device, those functions have to share hardware not only with each other but with Alexa's other core device functions, such as music playback.

[118] *See id.*

Ex.1029 / Page 143 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

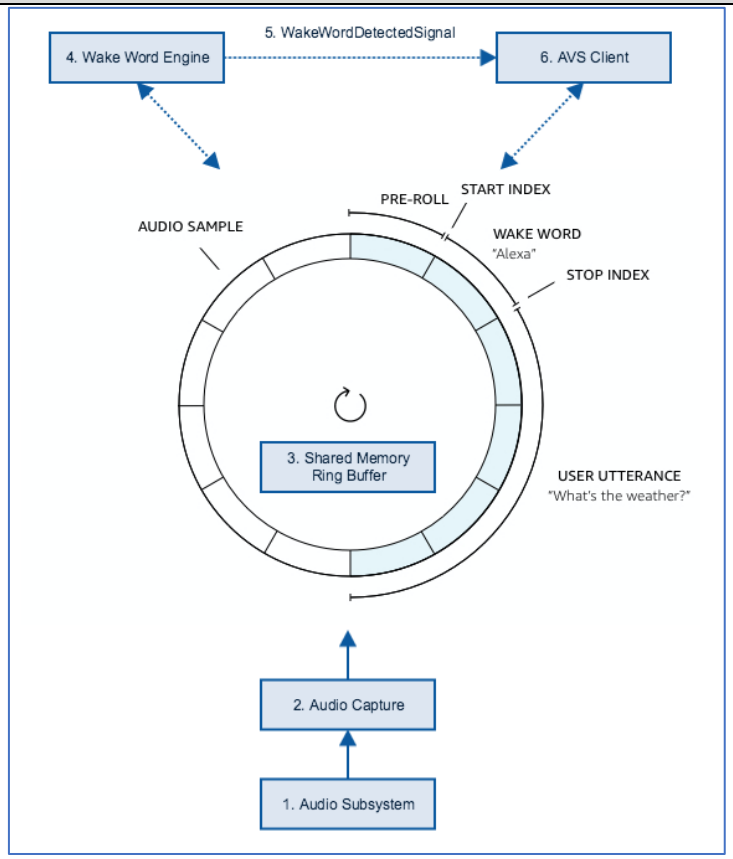
Claim 9	Accused Instrumentality(ies)
	For example, Amazon uses built-in technology for keyword spotting.[119] Echo devices use built-in technology called "keyword spotting" that matches spoken audio to the acoustic patterns of the wake word. Simply put, Echo devices are designed by default to detect only the sound waves of your chosen wake word, and everything else is ignored. Like water through a strainer, all other audio (people talking, faucets running, birds chirping) passes through the device until the wake word is "caught" and sent to Amazon's secure cloud, where your request is reanalyzed to verify the wake word was spoken. You can choose to opt-in to features like Alexa Emergency Assist that allow your device to detect more than your chosen wake word—for example, the sound of smoke alarms—but you would need to update your settings to do so. Want to keep nerding out? Follow the journey of a voice request. For example, according to Amazon, it uses a wake word verification engine.[120] Voice-initiated devices begin to stream user speech to AVS when the wake word engine detects a spoken wake word, such as "Alexa."

[119] *See* KAIFI0002092 *(How Alexa works: Wake word*, https://www.amazon.com/b?ie=UTF8&node=23608571011).

[120] *See* KAIFI0007141 *(Wake Word Verification Requirements*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/implement-ww-verification.html).

Ex.1029 / Page 144 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	

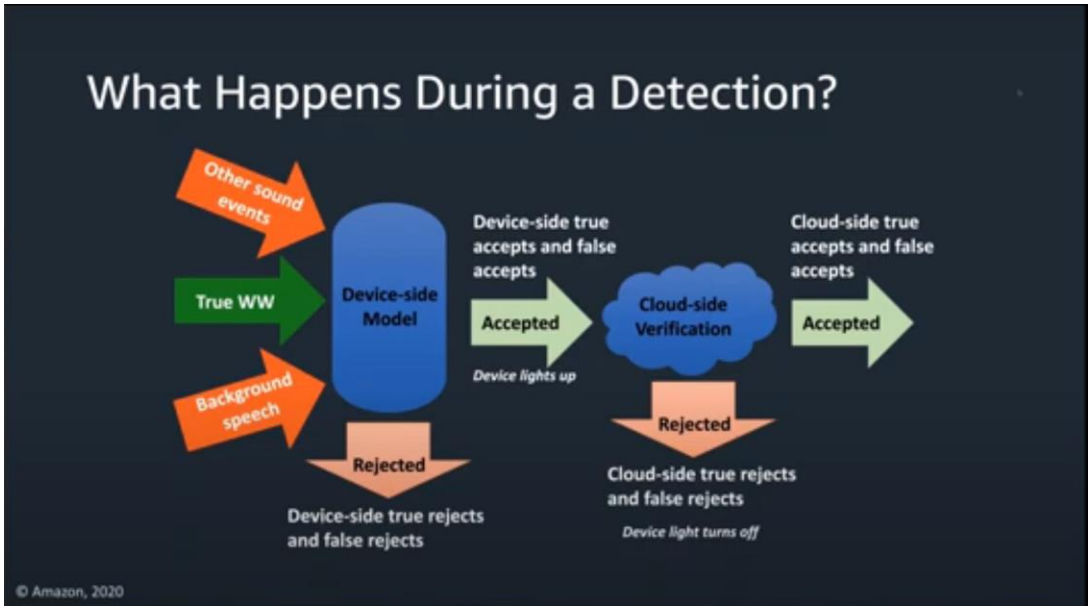
Ex.1029 / Page 145 of 347
AMAZON.COM, INC.

Claim 9	Accused Instrumentality(ies)
	4. **Wake Word Engine** - The component that reads the audio samples written to the shared memory ring buffer and analyzes the audio samples for occurrences of the wake word. The procedure of reading and processing audio samples occurs in a loop. For example, the loop reads N samples starting at location `0`; then analyzes the samples for an occurrence of the wake word. In the next iteration, it reads and analyzes the next N samples. When an occurrence of the wake word is detected, the wake word engine must identify two locations: 1) the start of the wake word, and 2) the end of the wake word. These values are stored as `startIndexInSamples` and `endIndexInSamples`. For example, according to Amazon, the wake word engine uses wake words. [121] Wake words The list of wake words informs Alexa of the possible valid wake words that a device might be set to listen for through the `SetWakeWords`, `WakeWordsReport`, and `WakeWordsChanged` messages. Currently, the only wake word available for Alexa Built-in devices is `ALEXA`, which applies to every possible locale for a device. Therefore, specify `ALEXA` in the global `DEFAULT` scope. This claim element is literally present, as described herein, but is also satisfied under the doctrine of equivalents because the structure and functionality provided by Amazon is being used to satisfy this claim limitation and therefore there is no vitiation of this element, and the Accused Instrumentalities incorporate structure and functionality that is not substantially different from the requirements of this limitation because they achieve substantially or exactly the same function, in substantially or exactly the same way as set forth in the claim element (as described herein), and achieves substantially or exactly the same result.

[121] *See* KAIFI0007093 *(Understand the AVS SpeechRecognizer*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/avs-speechrecognizer-concepts.html).

Ex.1029 / Page 146 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

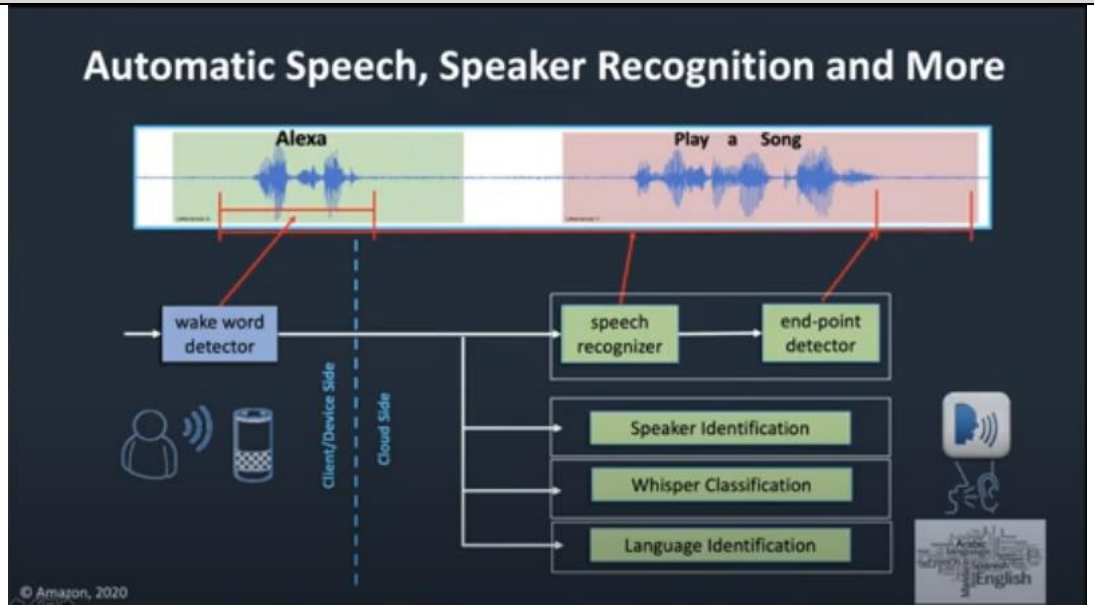
Claim 9	Accused Instrumentality(ies)
a continuous speech recognition unit using the processor,	The Accused Instrumentalities comprise a continuous speech recognition unit. For example, the Accused Instrumentalities include a combination of hardware, processing, and software functions both on devices and on the cloud server to process speech.[122]  This is further illustrated below.[123]

[122] *See* KAIFI0007814 *(The Science Behind Alexa's Wake Word Detection*, https://www.youtube.com/watch?v=kYrw2cC9dvs).
[123] *See* KAIFI0007816 *(How Alexa Uses Automatic Speech Recognition*, https://www.youtube.com/watch?v=8drK6Oyv1a0&t=296s).

Ex.1029 / Page 147 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)



For example, Amazon uses automatic speech recognition and natural language understanding.[124]

> It starts with detecting the "wake word" that triggers Alexa to start listening to the subsequent spoken words from the user. Wake word detection employs deep learning technology running on-device to recognize the wake word the user has chosen. Far-field automatic speech recognition (ASR) in the Amazon Web Services (AWS) cloud then converts into text the audio following the wake word, and determines when the user has stopped talking to Alexa.

[124] *See* KAIFI0004342 *(How Our Scientists are Making Alexa Smarter* (Mar. 29, 2018), https://www.aboutamazon.com/news/devices/how-our-scientists-are-making-alexa-smarter).

Ex.1029 / Page 148 of 347
AMAZON.COM, INC.

Claim 9	Accused Instrumentality(ies)
	Once the spoken audio has been converted to text, Alexa uses natural language understanding (NLU) to convert the words into a structured interpretation of intent that can be used to respond to the user from the more than 30,000 Alexa skills built by first- and third-party developers. This structured interpretation is used in combination with different forms of context, such as which type of device the user is interacting with, what the most likely skills are that can provide a response, or who is speaking. This context helps determine the best next action Alexa should take. The possible outcomes are to either respond with the best action from a skill or to ask for more information from the user. How Alexa responds or sounds is also critical for natural dialog. This is accomplished via text-to-speech synthesis (TTS) which converts arbitrary sequences of words into natural sounding, intelligible audio. What's common across all the technologies mentioned above is the emphasis on data-driven machine learning and fast inference at runtime to deliver an accurate response in as short a time as possible. As scientists and engineers we're always battling this healthy tension between accuracy and latency from when the user stops speaking to Alexa to when she responds. For example, Amazon uses a shared memory ring buffer.[125] Shared Memory Ring Buffer A shared memory ring buffer satisfies the requirement to stream a continuous utterance to AVS that includes 500 milliseconds of pre-roll, the wake word, and the user's request for cloud-based wake word verification. This reduces latency and the need to copy audio samples multiple times. The following diagram illustrates one approach; alternative implementations are allowed as long as the requisite functionality is maintained:

[125] *See* KAIFI0007141 *(Wake Word Verification Requirements*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/implement-ww-verification.html).

Ex.1029 / Page 149 of 347
AMAZON.COM, INC.

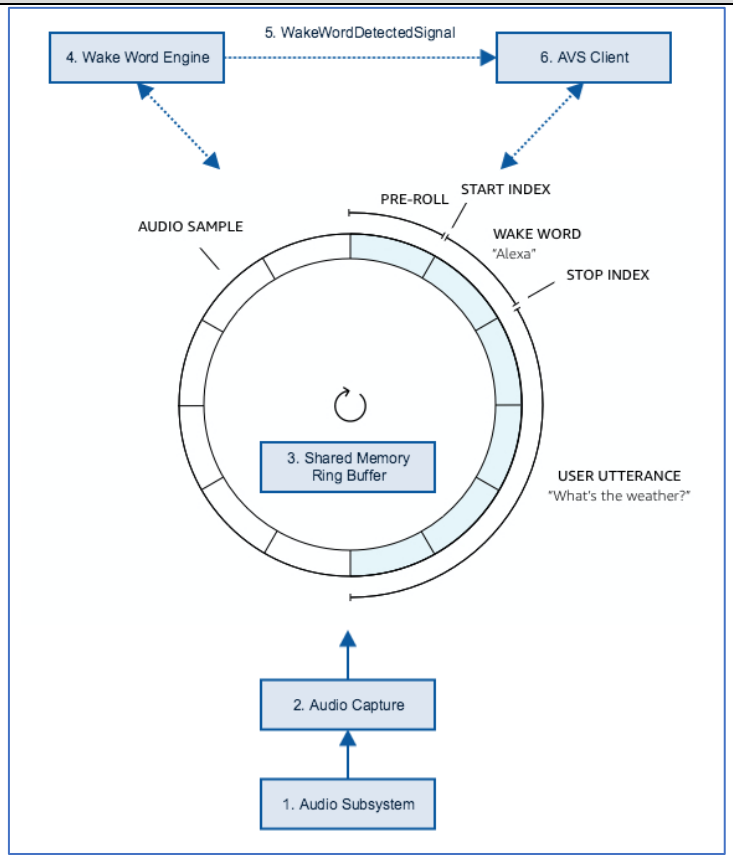
Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	For example, Amazon's wake word engine continuously recognizes speech.[126] Voice-initiated devices begin to stream user speech to AVS when the wake word engine detects a spoken wake word, such as "Alexa."

[126] *See id.*

Ex.1029 / Page 150 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	

Ex.1029 / Page 151 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	4. **Wake Word Engine** - The component that reads the audio samples written to the shared memory ring buffer and analyzes the audio samples for occurrences of the wake word. The procedure of reading and processing audio samples occurs in a loop. For example, the loop reads N samples starting at location `0`; then analyzes the samples for an occurrence of the wake word. In the next iteration, it reads and analyzes the next N samples. When an occurrence of the wake word is detected, the wake word engine must identify two locations: 1) the start of the wake word, and 2) the end of the wake word. These values are stored as `startIndexInSamples` and `endIndexInSamples`. For example, according to Amazon, the Wake Word Engine comprises a processor for speech recognition.[127] Each solution has varying degrees of optimization as illustrated in the Level of Optimization (LoO) column. It also shows an example of a very optimized solution on a considerably less powerful ARM Cortex-R4 microcontroller for illustration purposes, but where the Alexa wake word engine runs exclusively on the digital signal processor front-end of the device to offload the host processor. Table 2 and Figure 1 below illustrate how effectively the various solutions utilize the host processor and memory.

[127] *See* KAIFI0006993 (Bob Mars, *Sizing Up CPU, Memory, and Storage for Your Alexa Built-in Device* (Sept. 13, 2018), https://developer.amazon.com/en-US/blogs/alexa/post/2a32d792-d471-4136-8262-79962a2b4d72/cpu-memory-and-storage-for-alexa-built-in-device.html).

Ex.1029 / Page 152 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)

Arch	Host Processor(s)	Vendor	LoO 1-5	Gain	eCycles 1GHz Alexa	eCycles 1GHz Music
MIPS32	MIPS 24KEc	SI_1	4	2	31	24
MIPS32	MIPS 24KEc	SI_1	4	2	24	24
MIPS32	MIPS 24KEc	SI_2	4	2	266	183
MIPS32	Xburst	SI_2	4	2	172	128
MIPS32	Xburst	SI_3	3	2	90	90
ARMv7-A	Cortex-A7	V_2	2	1	50	46
ARMv7-A	Cortex-A7+Neon+M4	Dev Kit	1	1	78	75
ARMv7-A	Cortex-A7+Neon/VFPU	SI_1	4	1.5	106	104
ARMv8-A	Cortex A53	SI_3	3	1.5	70	65
ARMv8-A	Cortex A53	Dev Kit	1	1	52	48

eCycle: effective Cycles; MCPS normalized to 1 GHz with Gain multiplier
Gain: Cycle gain multiplier due to additional memory; 1.5=50%; 2=100%

Table 2- Normalized Headroom on Alexa Built-in Devices

For example, Amazon uses a speech recognizer.[128]

> The `SpeechRecognizer` interface is the core interface of the Alexa Voice Service (AVS) and exposes directives and events for capturing and interacting with user speech. This page discusses the concepts and process flows for the `SpeechRecognizer` interface.

[128] *See* KAIFI0007093 *(Understand the AVS SpeechRecognizer*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/avs-speechrecognizer-concepts.html).

Ex.1029 / Page 153 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	SpeechRecognizer functionality Every user utterance leverages `SpeechRecognizer`, including the following interaction types between an Alexa Built-in device and AVS: • **User Speech Capture**: Captures user speech from an Alexa Built-in device. • **User Speech Prompting**: Prompts a user for more speech input when needed by Alexa to deliver an appropriate response. • **Interaction Initiation Communication**: Enables a device to inform AVS of how a user initiated an Alexa interaction, such as press-and-hold, tap-and-release, or voice-initiated/wake word enabled. See Device Form Factor and Alexa Interaction. • **ASR Profile Selection**: SpeechRecognizers chooses the appropriate Automatic Speech Recognition (ASR) profile for your product, which allows Alexa to understand user speech and respond with precision. See Automatic Speech Recognition (ASR) profile

Ex.1029 / Page 154 of 347
AMAZON.COM, INC.

State diagram

The following diagram illustrates state changes driven by `SpeechRecognizer` components. Boxes represent `SpeechRecognizer` states, and the connectors represent state transitions.

`SpeechRecognizer` has the following states:

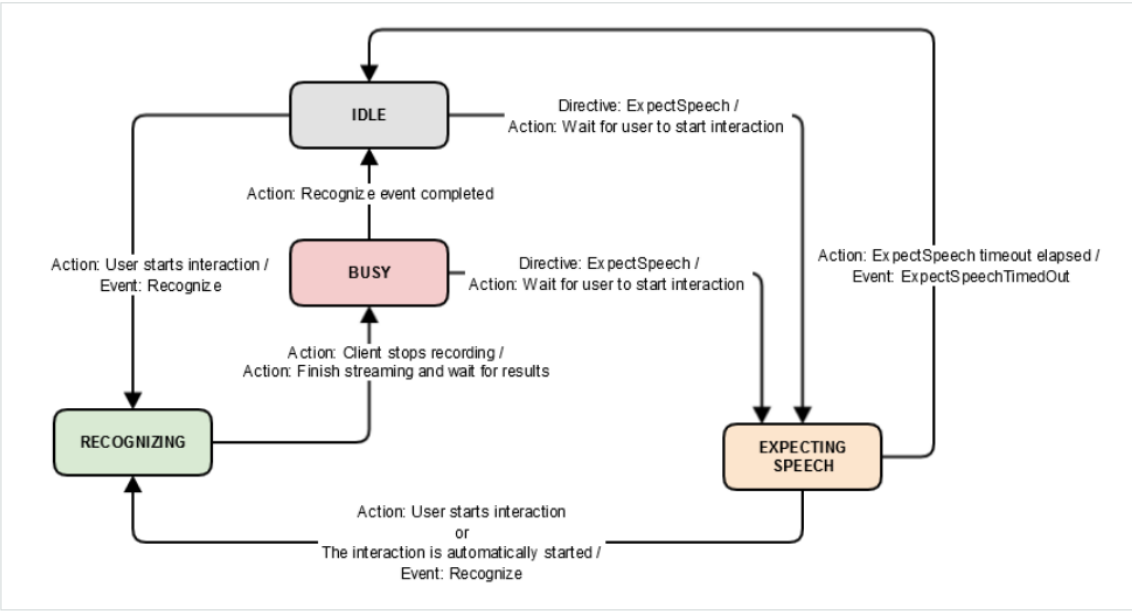
- **IDLE:** When not actively processing speech, the `SpeechRecognizer` is in an "idle" state. The idle state occurs under the following conditions:
  - Before capturing user speech.
  - Returning to an idle state after concluding a speech interaction with Alexa.
  - When an `ExpectSpeechTimedOut` event elapses.

  Note: In a multi-turn Alexa interaction, if Alexa requires more user speech input, `SpeechRecognizer` should transition from the idle state to the `ExpectSpeech` state without the user starting a new interaction.

- **RECOGNIZING:** When a user begins interacting with your client, specifically when the client streams captured audio to AVS, `SpeechRecognizer` should transition from the idle state to the `Recognize` state. `SpeechRecognizer` should remain in the `Recognize` until the client stops recording speech or finishes streaming, at which point your `SpeechRecognizer` component should transition from the `Recognize` state to the "busy" state.
- **BUSY:** While processing the speech request, `SpeechRecognizer` should be in the "busy" state. You cannot start another speech request until `SpeechRecognizer` transitions out of the busy state. From the *busy* state, `SpeechRecognizer` transitions to the idle state if Alexa processes and completes the request or to the `ExpectSpeech` state, if Alexa requires more speech input from the user.
- **EXPECTING SPEECH:** `SpeechRecognizer` should be in the `ExpectSpeech` state when Alexa requires more audio input from a user. From `ExpectSpeech`, `SpeechRecognizer` should transition to the `Recognize` when a user interaction occurs, or the interaction is automatically started on behalf of the user. It should transition to the idle state if Alexa detects no user interaction within the specified timeout window.

Ex.1029 / Page 155 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	The following diagram illustrates the expected transitions among the four `SpeechRecognizer` states:  For example, Amazon uses ASR.[129]

[129] *See* KAIFI0006791 (Ariya Rastrow, Shehzad Mevawalla, *On-device speech processing makes Alexa faster, lower-bandwidth*, AMAZON SCIENCE (Jan. 25, 2022), https://www.amazon.science/blog/on-device-speech-processing-makes-alexa-faster-lower-bandwidth).

Ex.1029 / Page 156 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	Moreover, in the cloud, the separate components of Alexa's speech-processing stack — automatic speech recognition (ASR), whisper detection, and speaker identification — run on separate server nodes with their own powerful processors. On-device, those functions have to share hardware not only with each other but with Alexa's other core device functions, such as music playback. For example, Amazon uses a shared memory ring buffer. [130] Shared Memory Ring Buffer A shared memory ring buffer satisfies the requirement to stream a continuous utterance to AVS that includes 500 milliseconds of pre-roll, the wake word, and the user's request for cloud-based wake word verification. This reduces latency and the need to copy audio samples multiple times. The following diagram illustrates one approach; alternative implementations are allowed as long as the requisite functionality is maintained: For example, Amazon uses conversational AI and machine learning technologies.[131]

[130] *See* KAIFI0007141 *(Wake Word Verification Requirements*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/implement-ww-verification.html).

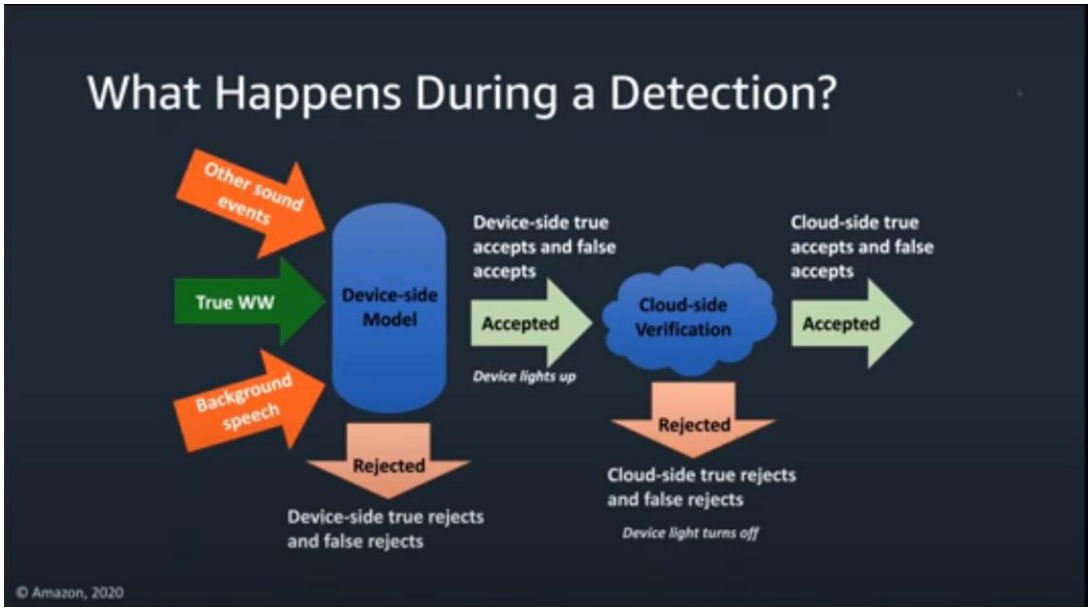
[131] *See* KAIFI0004342 *(How Our Scientists are Making Alexa Smarter* (Mar. 29, 2018), https://www.aboutamazon.com/news/devices/how-our-scientists-are-making-alexa-smarter).

Ex.1029 / Page 157 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	It starts with detecting the "wake word" that triggers Alexa to start listening to the subsequent spoken words from the user. Wake word detection employs deep learning technology running on-device to recognize the wake word the user has chosen. Far-field automatic speech recognition (ASR) in the Amazon Web Services (AWS) cloud then converts into text the audio following the wake word, and determines when the user has stopped talking to Alexa. Once the spoken audio has been converted to text, Alexa uses natural language understanding (NLU) to convert the words into a structured interpretation of intent that can be used to respond to the user from the more than 30,000 Alexa skills built by first- and third-party developers. This structured interpretation is used in combination with different forms of context, such as which type of device the user is interacting with, what the most likely skills are that can provide a response, or who is speaking. This context helps determine the best next action Alexa should take. The possible outcomes are to either respond with the best action from a skill or to ask for more information from the user. How Alexa responds or sounds is also critical for natural dialog. This is accomplished via text-to-speech synthesis (TTS) which converts arbitrary sequences of words into natural sounding, intelligible audio. What's common across all the technologies mentioned above is the emphasis on data-driven machine learning and fast inference at runtime to deliver an accurate response in as short a time as possible. As scientists and engineers we're always battling this healthy tension between accuracy and latency from when the user stops speaking to Alexa to when she responds. This claim element is literally present, as described herein, but is also satisfied under the doctrine of equivalents because the structure and functionality provided by Amazon is being used to satisfy this claim limitation and therefore there is no vitiation of this element, and the Accused Instrumentalities incorporate structure and functionality that is not substantially different from the requirements of this limitation because they achieve substantially or exactly the same function, in substantially or exactly the same way as

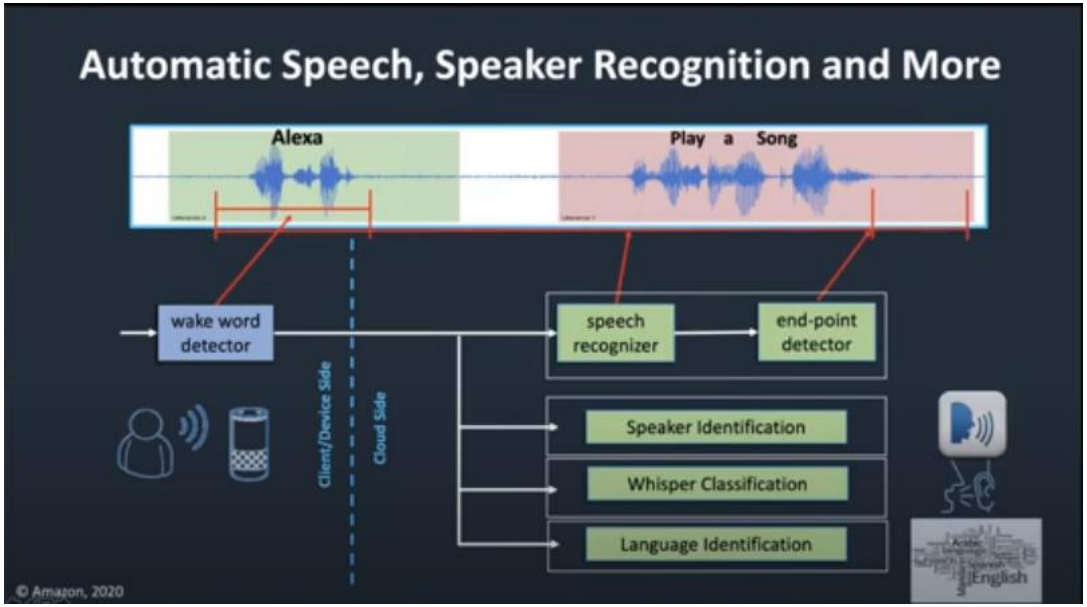
Ex.1029 / Page 158 of 347
AMAZON.COM, INC.

Claim 9	Accused Instrumentality(ies)
	set forth in the claim element (as described herein), and achieves substantially or exactly the same result.
wherein the preset call command recognition unit is configured to receive an input speech spoken by the speaker;	In the Accused Instrumentalities, the preset call command recognition unit is configured to receive an input speech spoken by the speaker. For example, the Accused Instrumentalities receive a wake word through a combination of hardware, processing, and software functions both on devices and on the cloud server.[132] 

[132] *See* KAIFI0007814 *(The Science Behind Alexa's Wake Word Detection*, https://www.youtube.com/watch?v=kYrw2cC9dvs).

Ex.1029 / Page 159 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	This is further illustrated below.[133]  For example, the wake word engine is configured to receive an input speech spoken by the speaker. [134] Voice-initiated devices begin to stream user speech to AVS when the wake word engine detects a spoken wake word, such as "Alexa."

[133] *See* KAIFI0007816 *(How Alexa Uses Automatic Speech Recognition*, https://www.youtube.com/watch?v=8drK6Oyv1a0&t=296s).

[134] *See* KAIFI0007141 *(Wake Word Verification Requirements*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/implement-ww-verification.html).

Ex.1029 / Page 160 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	4. **Wake Word Engine** - The component that reads the audio samples written to the shared memory ring buffer and analyzes the audio samples for occurrences of the wake word. The procedure of reading and processing audio samples occurs in a loop. For example, the loop reads N samples starting at location `0`; then analyzes the samples for an occurrence of the wake word. In the next iteration, it reads and analyzes the next N samples. When an occurrence of the wake word is detected, the wake word engine must identify two locations: 1) the start of the wake word, and 2) the end of the wake word. These values are stored as `startIndexInSamples` and `endIndexInSamples`. For example, according to Amazon, its speech recognizer captures and interacts with user speech.[135] The `SpeechRecognizer` interface is the core interface of the Alexa Voice Service (AVS) and exposes directives and events for capturing and interacting with user speech. This page discusses the concepts and process flows for the `SpeechRecognizer` interface.

[135] *See* KAIFI0007093 *(Understand the AVS SpeechRecognizer*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/avs-speechrecognizer-concepts.html).

Ex.1029 / Page 161 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	## SpeechRecognizer functionality Every user utterance leverages `SpeechRecognizer`, including the following interaction types between an Alexa Built-in device and AVS: • **User Speech Capture**: Captures user speech from an Alexa Built-in device. • **User Speech Prompting**: Prompts a user for more speech input when needed by Alexa to deliver an appropriate response. • **Interaction Initiation Communication**: Enables a device to inform AVS of how a user initiated an Alexa interaction, such as press-and-hold, tap-and-release, or voice-initiated/wake word enabled. See Device Form Factor and Alexa Interaction. • **ASR Profile Selection**: SpeechRecognizers chooses the appropriate Automatic Speech Recognition (ASR) profile for your product, which allows Alexa to understand user speech and respond with precision. See Automatic Speech Recognition (ASR) profile • **RECOGNIZING:** When a user begins interacting with your client, specifically when the client streams captured audio to AVS, `SpeechRecognizer` should transition from the idle state to the `Recognize` state. `SpeechRecognizer` should remain in the `Recognize` until the client stops recording speech or finishes streaming, at which point your `SpeechRecognizer` component should transition from the `Recognize` state to the "busy" state. • **BUSY:** While processing the speech request, `SpeechRecognizer` should be in the "busy" state. You cannot start another speech request until `SpeechRecognizer` transitions out of the busy state. From the *busy* state, `SpeechRecognizer` transitions to the idle state if Alexa processes and completes the request or to the `ExpectSpeech` state, if Alexa requires more speech input from the user. • **EXPECTING SPEECH:** `SpeechRecognizer` should be in the `ExpectSpeech` state when Alexa requires more audio input from a user. From `ExpectSpeech`, `SpeechRecognizer` should transition to the `Recognize` when a user interaction occurs, or the interaction is automatically started on behalf of the user. It should transition to the idle state if Alexa detects no user interaction within the specified timeout window.

Ex.1029 / Page 162 of 347
AMAZON.COM, INC.

Claim 9	Accused Instrumentality(ies)
	For example, Amazon uses wake word, keyword spotting, cloud verification and encryption, automatic speech recognition (ASR), natural language understanding (NLU), text-to-speech (TTS), and data storage technologies.[136]

[136] *See* KAIFI0002023 *(Follow the journey of a voice request*, https://www.amazon.com/b/?node=23608618011).

Ex.1029 / Page 163 of 347
AMAZON.COM, INC.

The wake word

Think of the wake word as a verbal cue that makes things happen. In the classic tale from The Arabian Nights, an invisible door into a mountain can only be opened by saying the phrase "Open Sesame." Use the phrase, the door opens. Don't use the phrase, the door stays shut. When you have a request for Alexa, you first need to say your chosen wake word, which by default is "Alexa." Only after your Echo device detects the wake word is Alexa listening to your requests.

When it comes to privacy, there should be no surprises. You'll always be able to tell when Alexa is listening to your request because a light indicator will appear on your Echo device or an audible tone will sound. Think of the "On the Air" signs that light up in television studios during a broadcast. These indicators notify you that your device has detected the wake word and Alexa is now processing your request.

Keyword spotting

Echo devices use built-in technology called "keyword spotting" that matches spoken audio to the acoustic patterns of the wake word. Simply put, Echo devices are designed by default to detect only the sound waves of your chosen wake word, and everything else is ignored. Like water through a strainer, all other audio (people talking, faucets running, birds chirping) passes through the device until the wake word is "caught" and sent to Amazon's secure cloud, where meaning is assigned to your request. You can choose to opt-in to features like Alexa Emergency Assist that allow your device to detect more than your chosen wake word—for example, the sound of smoke alarms—but you would need to update your settings to do so.

Cloud verification and encryption

When your Echo device detects the wake word, it sends your request to Amazon's secure cloud where audio is reanalyzed to verify the wake word was spoken. If this cloud software verification is unable to confirm the wake word was spoken, the Alexa system stops processing the audio. If the wake word is verified (or Alexa is activated using the action button), your request will be processed using several sophisticated algorithms that allow Alexa to respond appropriately. All of your interactions with Alexa are encrypted in transit throughout this process.

Automatic Speech Recognition (ASR)

After wake word verification, your voice request undergoes Automatic Speech Recognition (ASR) which transcribes audio into text. Using our example request ("Alexa, what's the weather in Seattle?"), transcriptions could include one or several of the following:

"what is the weather in seattle" "watt is the weather in seattle" "what is the whether in seattle"

Alexa then chooses to act on the transcription that is most likely to be correct and stores it in Amazon's secure cloud. You can view and delete it at any time at Alexa Privacy Settings or in the Alexa app (Settings > Alexa Privacy > Review Voice History). Voice recordings and transcriptions may be used to improve the ASR process. Learn how data helps Alexa improve.

Natural Language Understanding

After ASR comes Natural Language Understanding (NLU). NLU assigns meaning to the transcription producing an "intent" which is an instruction to the Alexa system to tell it what to respond to. In our example, the intent is identified as "weather" and sent to the designated data source where information is pulled for location (Seattle) and time (today). Intents may also be used to improve the NLU process.

Text-to-Speech (TTS)

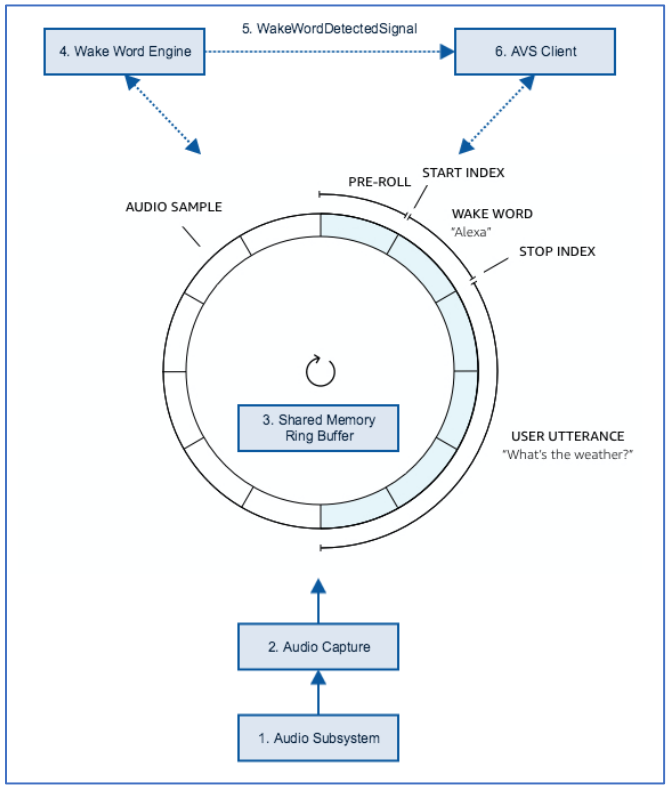
Once the necessary information is gathered, your request uses text-to-speech (TTS) technologies to generate an audio file that becomes Alexa's response. The text from Alexa's response is stored so that you can review past answers at Alexa Privacy Settings or in the Alexa app. These responses may also be reviewed by Amazon to ensure the device is providing the most relevant answers to you, and to make sure the TTS system is translating text to speech to the best of its ability.

Data storage

You have control over your Alexa experience. We make it easy for you to update your privacy settings. You can always view and manage your Alexa interactions at Alexa Privacy Settings or in the Alexa app (Settings > Alexa Privacy). If you're enrolled in voice ID, you can ask, "Alexa, what are my privacy settings?" to review how long your voice recordings are currently saved and whether or not they're used to develop new features or manually reviewed to help improve Alexa. You can also say, "Alexa, update my privacy settings" at any time to make updates.

Ex.1029 / Page 164 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	For example, the wake word engine is configured to receive an input speech spoken by the speaker.[137] 

[137] *See* KAIFI0007141 *(Wake Word Verification Requirements*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/implement-ww-verification.html).

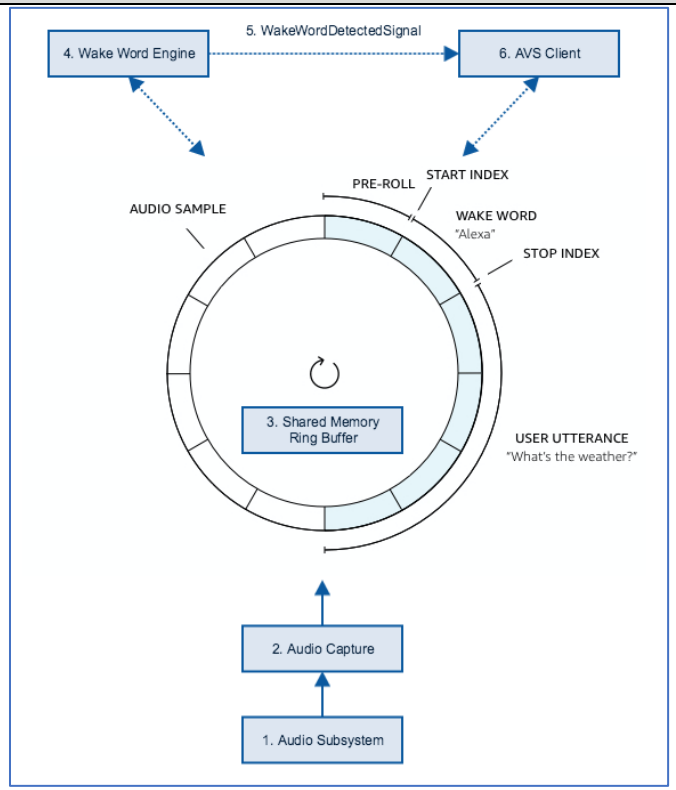
Ex.1029 / Page 165 of 347
AMAZON.COM, INC.

Claim 9	Accused Instrumentality(ies)
	3. **Shared Memory Ring Buffer** - A shared memory ring buffer object that is thread safe, and accessible by multiple processes. Access to this memory block should support at least one writer and two readers simultaneously. You may choose to implement these executables as a shared library. • **Write API** - Used by the recording process to write audio samples to the shared memory ring buffer. Write samples should move data so that the oldest audio samples are overwritten. This example shows a typical API signature: `WriteSamples(BufferIn, BufferInLength)`. • **Read API** - Used by the wake word engine and the AVS client to read from specific locations in the shared memory ring buffer. A typical implementation represents location as a 64-bit integer. For example, reading 200 samples from the location 100 is reading the 100th to 300th samples that were written to the shared memory ring buffer. This is an example of a typical API signature: `ReadSamples(BufferIn, BufferInLength)`. **Note**: Reading from an old location that was overwritten with new audio samples will result in an unintended sample read. 4. **Wake Word Engine** - The component that reads the audio samples written to the shared memory ring buffer and analyzes the audio samples for occurrences of the wake word. The procedure of reading and processing audio samples occurs in a loop. For example, the loop reads N samples starting at location `0`; then analyzes the samples for an occurrence of the wake word. In the next iteration, it reads and analyzes the next N samples. When an occurrence of the wake word is detected, the wake word engine must identify two locations: 1) the start of the wake word, and 2) the end of the wake word. These values are stored as `startIndexInSamples` and `endIndexInSamples`. For example, according to Amazon, the AVS Client is also configured to receive a user utterance.[138] .

[138] *See id.*

Ex.1029 / Page 166 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	 6. **AVS Client** - The client receives the `WakeWordDetectedSignal`, extracts the `startIndexInSamples`, then reads and streams audio samples to AVS including 500 milliseconds of pre-roll. • The pre-roll is the number of samples corresponding to 500 milliseconds. For example, if the audio samples were recorded at a rate of 16 kilohertz (kHz), then pre-roll is 8,000 samples.

Ex.1029 / Page 167 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	For example, according to Amazon, the Alexa Voice Service speech recognizer captures and interacts with user speech.[139]

> The `SpeechRecognizer` interface is the core interface of the Alexa Voice Service (AVS) and exposes directives and events for capturing and interacting with user speech. This page discusses the concepts and process flows for the `SpeechRecognizer` interface.

SpeechRecognizer functionality

Every user utterance leverages `SpeechRecognizer`, including the following interaction types between an Alexa Built-in device and AVS:

- **User Speech Capture**: Captures user speech from an Alexa Built-in device.
- **User Speech Prompting**: Prompts a user for more speech input when needed by Alexa to deliver an appropriate response.
- **Interaction Initiation Communication**: Enables a device to inform AVS of how a user initiated an Alexa interaction, such as press-and-hold, tap-and-release, or voice-initiated/wake word enabled. See Device Form Factor and Alexa Interaction.
- **ASR Profile Selection**: SpeechRecognizers chooses the appropriate Automatic Speech Recognition (ASR) profile for your product, which allows Alexa to understand user speech and respond with precision. See Automatic Speech Recognition (ASR) profile

[139] *See* KAIFI0007093 *(Understand the AVS SpeechRecognizer*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/avs-speechrecognizer-concepts.html).

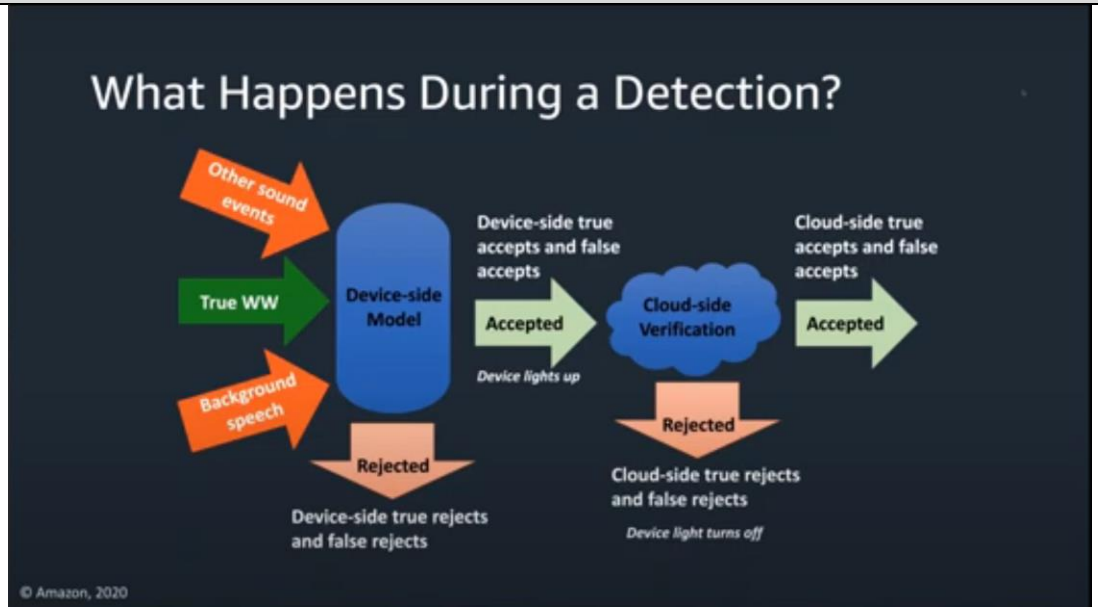
Ex.1029 / Page 168 of 347
AMAZON.COM, INC.

Claim 9	Accused Instrumentality(ies)
	• **RECOGNIZING:** When a user begins interacting with your client, specifically when the client streams captured audio to AVS, `SpeechRecognizer` should transition from the idle state to the `Recognize` state. `SpeechRecognizer` should remain in the `Recognize` until the client stops recording speech or finishes streaming, at which point your `SpeechRecognizer` component should transition from the `Recognize` state to the "busy" state. • **BUSY:** While processing the speech request, `SpeechRecognizer` should be in the "busy" state. You cannot start another speech request until `SpeechRecognizer` transitions out of the busy state. From the *busy* state, `SpeechRecognizer` transitions to the idle state if Alexa processes and completes the request or to the `ExpectSpeech` state, if Alexa requires more speech input from the user. • **EXPECTING SPEECH:** `SpeechRecognizer` should be in the `ExpectSpeech` state when Alexa requires more audio input from a user. From `ExpectSpeech`, `SpeechRecognizer` should transition to the `Recognize` when a user interaction occurs, or the interaction is automatically started on behalf of the user. It should transition to the idle state if Alexa detects no user interaction within the specified timeout window. This claim element is literally present, as described herein, but is also satisfied under the doctrine of equivalents because the structure and functionality provided by Amazon is being used to satisfy this claim limitation and therefore there is no vitiation of this element, and the Accused Instrumentalities incorporate structure and functionality that is not substantially different from the requirements of this limitation because they achieve substantially or exactly the same function, in substantially or exactly the same way as set forth in the claim element (as described herein), and achieves substantially or exactly the same result.
receive a subsequent input speech spoken, by the speaker, subsequently to the input speech;	The Accused Instrumentalities are configured to receive a subsequent input speech spoken, by the speaker, subsequently to the input speech. For example, the Accused Instrumentalities receive a wake word and subsequent speech, through a combination of hardware, processing, and software functions both on devices and on the cloud server.[140]

[140] *See* KAIFI0007814 *(The Science Behind Alexa's Wake Word Detection*, https://www.youtube.com/watch?v=kYrw2cC9dvs).

Ex.1029 / Page 169 of 347
AMAZON.COM, INC.

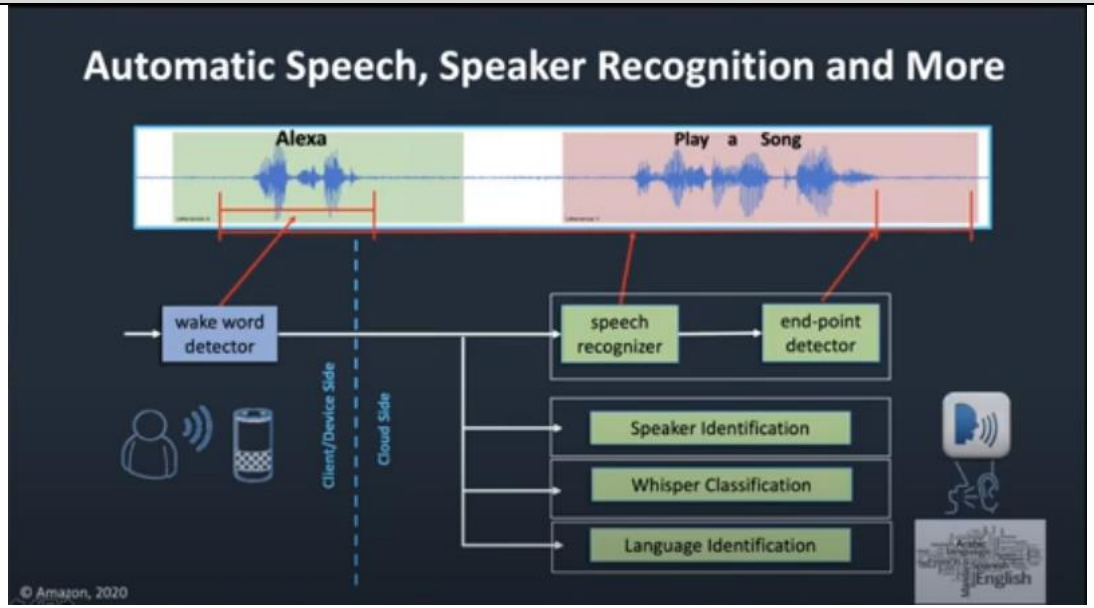
Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	 This is further illustrated below.[141]

[141] *See* KAIFI0007816 *(How Alexa Uses Automatic Speech Recognition*, https://www.youtube.com/watch?v=8drK6Oyv1a0&t=296s).

Ex.1029 / Page 170 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	 For example, Amazon describes a voice request.[142]

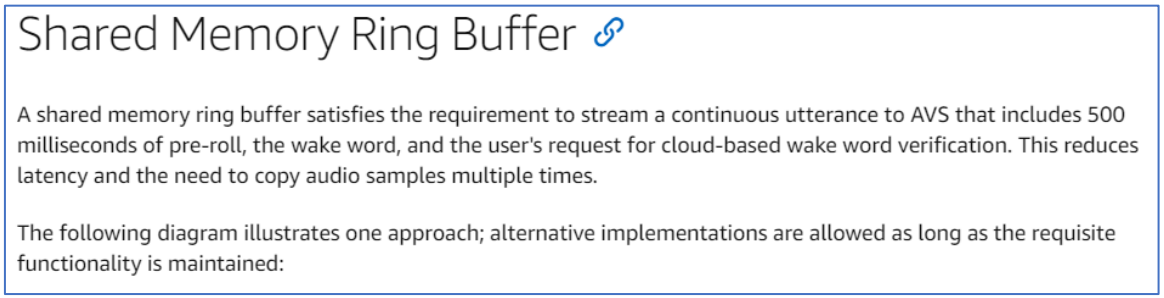
[142] *See* KAIFI0002023 *(Follow the journey of a voice request*, https://www.amazon.com/b/?node=23608618011).

Ex.1029 / Page 171 of 347
AMAZON.COM, INC.



For example, Amazon provides a shared memory

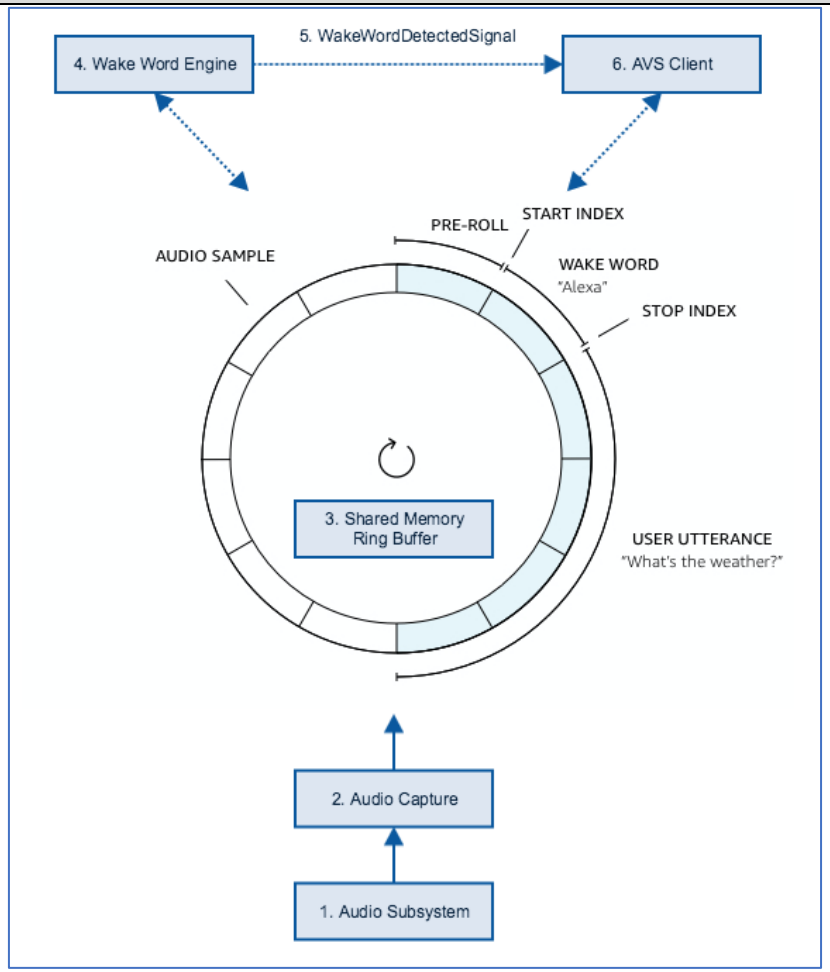
Ex.1029 / Page 172 of 347
AMAZON.COM, INC.

Claim 9	Accused Instrumentality(ies)
	ring buffer.[143]  Shared Memory Ring Buffer A shared memory ring buffer satisfies the requirement to stream a continuous utterance to AVS that includes 500 milliseconds of pre-roll, the wake word, and the user's request for cloud-based wake word verification. This reduces latency and the need to copy audio samples multiple times. The following diagram illustrates one approach; alternative implementations are allowed as long as the requisite functionality is maintained:

[143] *See* KAIFI0007141 *(Wake Word Verification Requirements*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/implement-ww-verification.html). .

Ex.1029 / Page 173 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	

Ex.1029 / Page 174 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	1. **Audio Subsystem** - Middleware, such as Advanced Linux Sound Architecture (ALSA), that provides APIs to open a stream from a recording device, and receive an audio sample stream. 2. **Audio Capture** - A process that opens a recording device, and writes audio samples into the shared memory ring buffer. 3. **Shared Memory Ring Buffer** - A shared memory ring buffer object that is thread safe, and accessible by multiple processes. Access to this memory block should support at least one writer and two readers simultaneously. You may choose to implement these executables as a shared library. • **Write API** - Used by the recording process to write audio samples to the shared memory ring buffer. Write samples should move data so that the oldest audio samples are overwritten. This example shows a typical API signature: `WriteSamples(BufferIn, BufferInLength)`. • **Read API** - Used by the wake word engine and the AVS client to read from specific locations in the shared memory ring buffer. A typical implementation represents location as a 64-bit integer. For example, reading 200 samples from the location 100 is reading the 100th to 300th samples that were written to the shared memory ring buffer. This is an example of a typical API signature: `ReadSamples(BufferIn, BufferInLength)`. **Note**: Reading from an old location that was overwritten with new audio samples will result in an unintended sample read. 4. **Wake Word Engine** - The component that reads the audio samples written to the shared memory ring buffer and analyzes the audio samples for occurrences of the wake word. The procedure of reading and processing audio samples occurs in a loop. For example, the loop reads N samples starting at location `0`; then analyzes the samples for an occurrence of the wake word. In the next iteration, it reads and analyzes the next N samples. When an occurrence of the wake word is detected, the wake word engine must identify two locations: 1) the start of the wake word, and 2) the end of the wake word. These values are stored as `startIndexInSamples` and `endIndexInSamples`. 5. **WakeWordDetectedSignal** - A signal sent from the wake word engine to the AVS client when an occurrence of the wake word is detected. This signal includes two locations, the `startIndexInSamples` and `endIndexInSamples`. 6. **AVS Client** - The client receives the `WakeWordDetectedSignal`, extracts the `startIndexInSamples`, then reads and streams audio samples to AVS including 500 milliseconds of pre-roll. • The pre-roll is the number of samples corresponding to 500 milliseconds. For example, if the audio samples were recorded at a rate of 16 kilohertz (kHz), then pre-roll is 8,000 samples.

Ex.1029 / Page 175 of 347
AMAZON.COM, INC.

Claim 9	Accused Instrumentality(ies)
	For example, according to Amazon, its speech recognizer captures and interacts with user speech.[144] The `SpeechRecognizer` interface is the core interface of the Alexa Voice Service (AVS) and exposes directives and events for capturing and interacting with user speech. This page discusses the concepts and process flows for the `SpeechRecognizer` interface. SpeechRecognizer functionality Every user utterance leverages `SpeechRecognizer`, including the following interaction types between an Alexa Built-in device and AVS: • **User Speech Capture**: Captures user speech from an Alexa Built-in device. • **User Speech Prompting**: Prompts a user for more speech input when needed by Alexa to deliver an appropriate response. • **Interaction Initiation Communication**: Enables a device to inform AVS of how a user initiated an Alexa interaction, such as press-and-hold, tap-and-release, or voice-initiated/wake word enabled. See Device Form Factor and Alexa Interaction. • **ASR Profile Selection**: SpeechRecognizers chooses the appropriate Automatic Speech Recognition (ASR) profile for your product, which allows Alexa to understand user speech and respond with precision. See Automatic Speech Recognition (ASR) profile

[144] *See* KAIFI0007093 *(Understand the AVS SpeechRecognizer*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/avs-speechrecognizer-concepts.html).

Ex.1029 / Page 176 of 347
AMAZON.COM, INC.

Claim 9	Accused Instrumentality(ies)
	• **RECOGNIZING:** When a user begins interacting with your client, specifically when the client streams captured audio to AVS, `SpeechRecognizer` should transition from the idle state to the `Recognize` state. `SpeechRecognizer` should remain in the `Recognize` until the client stops recording speech or finishes streaming, at which point your `SpeechRecognizer` component should transition from the `Recognize` state to the "busy" state. • **BUSY:** While processing the speech request, `SpeechRecognizer` should be in the "busy" state. You cannot start another speech request until `SpeechRecognizer` transitions out of the busy state. From the *busy* state, `SpeechRecognizer` transitions to the idle state if Alexa processes and completes the request or to the `ExpectSpeech` state, if Alexa requires more speech input from the user. • **EXPECTING SPEECH:** `SpeechRecognizer` should be in the `ExpectSpeech` state when Alexa requires more audio input from a user. From `ExpectSpeech`, `SpeechRecognizer` should transition to the `Recognize` when a user interaction occurs, or the interaction is automatically started on behalf of the user. It should transition to the idle state if Alexa detects no user interaction within the specified timeout window. For example, Amazon uses wake word, keyword spotting, cloud verification and encryption, automatic speech recognition (ASR), natural language understanding (NLU), text-to-speech (TTS), and data storage technologies. [145]

[145] *See* KAIFI0002023 *(Follow the journey of a voice request*, https://www.amazon.com/b/?node=23608618011).

Ex.1029 / Page 177 of 347
AMAZON.COM, INC.

The wake word

Think of the wake word as a verbal cue that makes things happen. In the classic tale from The Arabian Nights, an invisible door into a mountain can only be opened by saying the phrase "Open Sesame." Use the phrase, the door opens. Don't use the phrase, the door stays shut. When you have a request for Alexa, you first need to say your chosen wake word, which by default is "Alexa." Only after your Echo device detects the wake word is Alexa listening to your requests.

When it comes to privacy, there should be no surprises. You'll always be able to tell when Alexa is listening to your request because a light indicator will appear on your Echo device or an audible tone will sound. Think of the "On the Air" signs that light up in television studios during a broadcast. These indicators notify you that your device has detected the wake word and Alexa is now processing your request.

Keyword spotting

Echo devices use built-in technology called "keyword spotting" that matches spoken audio to the acoustic patterns of the wake word. Simply put, Echo devices are designed by default to detect only the sound waves of your chosen wake word, and everything else is ignored. Like water through a strainer, all other audio (people talking, faucets running, birds chirping) passes through the device until the wake word is "caught" and sent to Amazon's secure cloud, where meaning is assigned to your request. You can choose to opt-in to features like Alexa Emergency Assist that allow your device to detect more than your chosen wake word—for example, the sound of smoke alarms—but you would need to update your settings to do so.

Cloud verification and encryption

When your Echo device detects the wake word, it sends your request to Amazon's secure cloud where audio is reanalyzed to verify the wake word was spoken. If this cloud software verification is unable to confirm the wake word was spoken, the Alexa system stops processing the audio. If the wake word is verified (or Alexa is activated using the action button), your request will be processed using several sophisticated algorithms that allow Alexa to respond appropriately. All of your interactions with Alexa are encrypted in transit throughout this process.

Automatic Speech Recognition (ASR)

After wake word verification, your voice request undergoes Automatic Speech Recognition (ASR) which transcribes audio into text. Using our example request ("Alexa, what's the weather in Seattle?"), transcriptions could include one or several of the following:

"what is the weather in seattle" "watt is the weather in seattle" "what is the whether in seattle"

Alexa then chooses to act on the transcription that is most likely to be correct and stores it in Amazon's secure cloud. You can view and delete it at any time at Alexa Privacy Settings or in the Alexa app (Settings > Alexa Privacy > Review Voice History). Voice recordings and transcriptions may be used to improve the ASR process. Learn how data helps Alexa improve.

Natural Language Understanding

After ASR comes Natural Language Understanding (NLU). NLU assigns meaning to the transcription producing an "intent" which is an instruction to the Alexa system to tell it what to respond to. In our example, the intent is identified as "weather" and sent to the designated data source where information is pulled for location (Seattle) and time (today). Intents may also be used to improve the NLU process.

Text-to-Speech (TTS)

Once the necessary information is gathered, your request uses text-to-speech (TTS) technologies to generate an audio file that becomes Alexa's response. The text from Alexa's response is stored so that you can review past answers at Alexa Privacy Settings or in the Alexa app. These responses may also be reviewed by Amazon to ensure the device is providing the most relevant answers to you, and to make sure the TTS system is translating text to speech to the best of its ability.

Data storage

You have control over your Alexa experience. We make it easy for you to update your privacy settings. You can always view and manage your Alexa interactions at Alexa Privacy Settings or in the Alexa app (Settings > Alexa Privacy). If you're enrolled in voice ID, you can ask, "Alexa, what are my privacy settings?" to review how long your voice recordings are currently saved and whether or not they're used to develop new features or manually reviewed to help improve Alexa. You can also say, "Alexa, update my privacy settings" at any time to make updates.

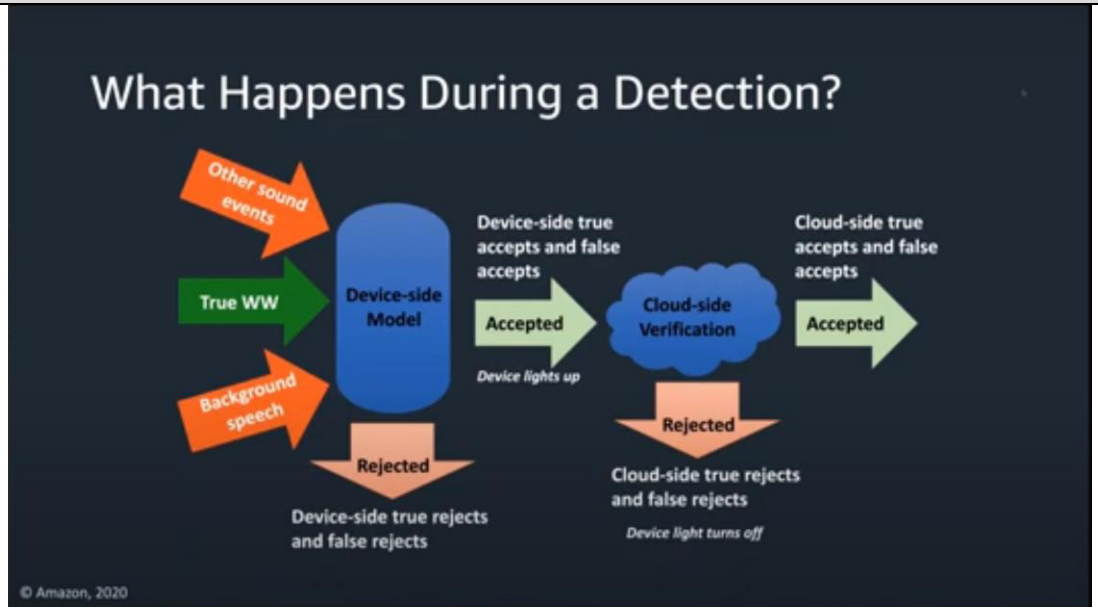
Ex.1029 / Page 178 of 347
AMAZON.COM, INC.

Claim 9	Accused Instrumentality(ies)
	This claim element is literally present, as described herein, but is also satisfied under the doctrine of equivalents because the structure and functionality provided by Amazon is being used to satisfy this claim limitation and therefore there is no vitiation of this element, and the Accused Instrumentalities incorporate structure and functionality that is not substantially different from the requirements of this limitation because they achieve substantially or exactly the same function, in substantially or exactly the same way as set forth in the claim element (as described herein), and achieves substantially or exactly the same result.
compare the input speech with a preset call command; and	The Accused Instrumentalities are configured to compare the input speech with a preset call command. For example, the Accused Instrumentalities compare through a combination of hardware, processing, and software functions both on devices and on the cloud server, an input speech with a wake word.[146]

[146] *See* KAIFI0007814 *(The Science Behind Alexa's Wake Word Detection*, https://www.youtube.com/watch?v=kYrw2cC9dvs).

Ex.1029 / Page 179 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

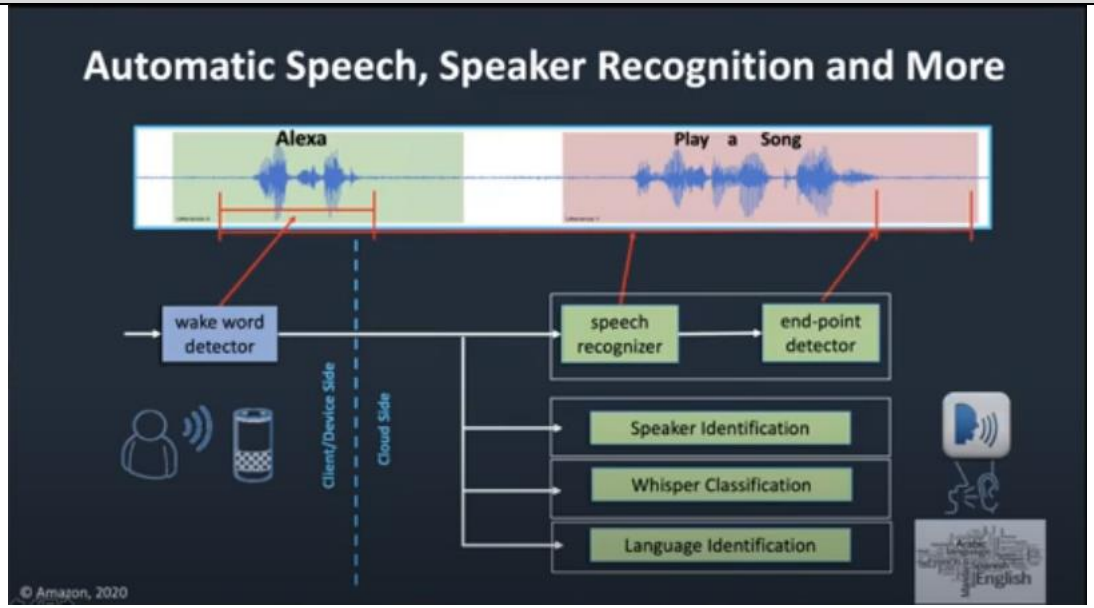
Claim 9	Accused Instrumentality(ies)
	 This is further illustrated below.[147]

[147] *See* KAIFI0007816 *(How Alexa Uses Automatic Speech Recognition*, https://www.youtube.com/watch?v=8drK6Oyv1a0&t=296s)..

Ex.1029 / Page 180 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)



For example, Amazon uses keyword spotting.[148]

> Echo devices use built-in technology called "keyword spotting" that matches spoken audio to the acoustic patterns of the wake word. Simply put, Echo devices are designed by default to detect only the sound waves of your chosen wake word, and everything else is ignored.

[148] *See* KAIFI0002092 *(How Alexa works: Wake word*, https://www.amazon.com/b?ie=UTF8&node=23608571011).

Ex.1029 / Page 181 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	For example, Amazon uses wake words.[149] Wake words The list of wake words informs Alexa of the possible valid wake words that a device might be set to listen for through the `SetWakeWords`, `WakeWordsReport`, and `WakeWordsChanged` messages. Currently, the only wake word available for Alexa Built-in devices is `ALEXA`, which applies to every possible locale for a device. Therefore, specify `ALEXA` in the global `DEFAULT` scope. For example, Amazon provides a wake word verification engine.[150] 4. **Wake Word Engine** - The component that reads the audio samples written to the shared memory ring buffer and analyzes the audio samples for occurrences of the wake word. The procedure of reading and processing audio samples occurs in a loop. For example, the loop reads N samples starting at location `0`; then analyzes the samples for an occurrence of the wake word. In the next iteration, it reads and analyzes the next N samples. When an occurrence of the wake word is detected, the wake word engine must identify two locations: 1) the start of the wake word, and 2) the end of the wake word. These values are stored as `startIndexInSamples` and `endIndexInSamples`. According to Amazon, the speech recognizer transitions are illustrated in the below diagram.[151]

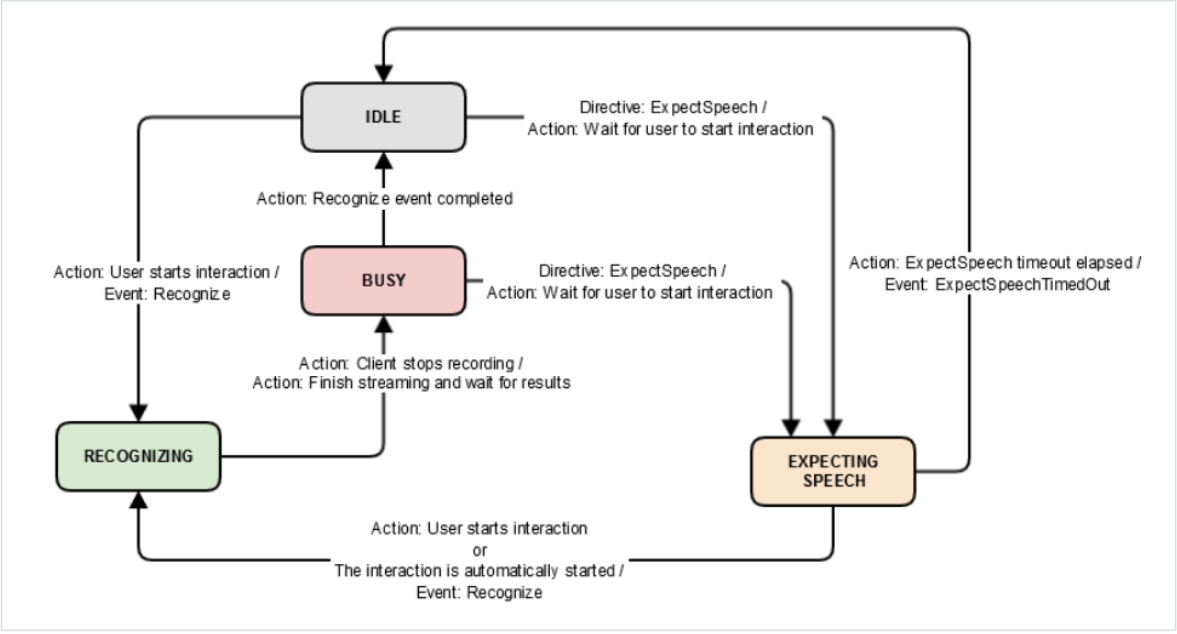
[149] *See* KAIFI0004342 *(How Our Scientists are Making Alexa Smarter* (Mar. 29, 2018), https://www.aboutamazon.com/news/devices/how-our-scientists-are-making-alexa-smarter).

[150] *See* KAIFI0007141 *(Wake Word Verification Requirements*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/implement-ww-verification.html).

[151] *See* KAIFI0007093 *(Understand the AVS SpeechRecognizer*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-

Ex.1029 / Page 182 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	The following diagram illustrates the expected transitions among the four `SpeechRecognizer` states:  Amazon also provides wake word verification.[152]

service/avs-speechrecognizer-concepts.html).

[152] *See* KAIFI0002012 *(Enable Wake Word Verification*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/enable-cloud-based-wake-word-verification.html).

Ex.1029 / Page 183 of 347
AMAZON.COM, INC.

Claim 9	Accused Instrumentality(ies)
	Cloud-based wake word verification improves wake word accuracy for Alexa Built-in products by reducing false wakes caused by words that sound similar to the wake word. Examples of words that might cause a false wake for "Alexa" include "Alex", "election", "Alexis". Cloud-based wake word verification also detects media mentions of the wake word, such as, the mention of "Alexa" in an Amazon commercial. The wake word engine on a device performs the initial wake word detection, and then the cloud verifies the wake word. If the cloud detects a false wake, the Alexa Voice Service (AVS) sends a `StopCapture` directive to the device in the downchannel that instructs it to close the audio stream, and if applicable, to turn off the blue LEDs to indicate that Alexa has stopped listening. Voice-initiated devices begin to stream user speech to AVS when the wake word engine detects a spoken wake word, such as "Alexa." The stream closes when the user stops speaking or when AVS identifies a user intent, and AVS returns a `StopCapture` directive to the device. Cloud-based wake word verification has the following requirements: • **Wake word** – Include the wake word in the stream so that AVS can perform cloud-based wake word verification, which reduces false wakes. If AVS can't detect the wake word detected during cloud-based wake word verification, AVS discards the utterance. • **500 milliseconds of pre-roll** – Pre-roll is the audio captured before AVS detects the wake word and helps calibrate the ambient noise level of the recording to enhance speech recognition. • **User speech** – Any user speech that the device captures until receiving a `StopCapture` directive. This allows AVS to verify the wake word included in the stream, reducing the number of erroneous responses due to false wakes. To learn how to implement a shared memory ring buffer for writing and reading audio samples and to include the start and stop indices for wake word detection in each `Recognize` event sent to AVS, see the Requirements for Cloud-Based Wake Word Verification. For example, Amazon uses wake word, keyword spotting, cloud verification and encryption, automatic speech recognition (ASR), natural language understanding (NLU), text-to-speech (TTS), and data storage technologies. [153]

[153] *See* KAIFI0002023 *(Follow the journey of a voice request*, https://www.amazon.com/b/?node=23608618011).

Ex.1029 / Page 184 of 347
AMAZON.COM, INC.

The wake word

Think of the wake word as a verbal cue that makes things happen. In the classic tale from The Arabian Nights, an invisible door into a mountain can only be opened by saying the phrase "Open Sesame." Use the phrase, the door opens. Don't use the phrase, the door stays shut. When you have a request for Alexa, you first need to say your chosen wake word, which by default is "Alexa." Only after your Echo device detects the wake word is Alexa listening to your requests.

When it comes to privacy, there should be no surprises. You'll always be able to tell when Alexa is listening to your request because a light indicator will appear on your Echo device or an audible tone will sound. Think of the "On the Air" signs that light up in television studios during a broadcast. These indicators notify you that your device has detected the wake word and Alexa is now processing your request.

Keyword spotting

Echo devices use built-in technology called "keyword spotting" that matches spoken audio to the acoustic patterns of the wake word. Simply put, Echo devices are designed by default to detect only the sound waves of your chosen wake word, and everything else is ignored. Like water through a strainer, all other audio (people talking, faucets running, birds chirping) passes through the device until the wake word is "caught" and sent to Amazon's secure cloud, where meaning is assigned to your request. You can choose to opt-in to features like Alexa Emergency Assist that allow your device to detect more than your chosen wake word—for example, the sound of smoke alarms—but you would need to update your settings to do so.

Cloud verification and encryption

When your Echo device detects the wake word, it sends your request to Amazon's secure cloud where audio is reanalyzed to verify the wake word was spoken. If this cloud software verification is unable to confirm the wake word was spoken, the Alexa system stops processing the audio. If the wake word is verified (or Alexa is activated using the action button), your request will be processed using several sophisticated algorithms that allow Alexa to respond appropriately. All of your interactions with Alexa are encrypted in transit throughout this process.

Automatic Speech Recognition (ASR)

After wake word verification, your voice request undergoes Automatic Speech Recognition (ASR) which transcribes audio into text. Using our example request ("Alexa, what's the weather in Seattle?"), transcriptions could include one or several of the following:

"what is the weather in seattle" "watt is the weather in seattle" "what is the whether in seattle"

Alexa then chooses to act on the transcription that is most likely to be correct and stores it in Amazon's secure cloud. You can view and delete it at any time at Alexa Privacy Settings or in the Alexa app (Settings > Alexa Privacy > Review Voice History). Voice recordings and transcriptions may be used to improve the ASR process. Learn how data helps Alexa improve.

Natural Language Understanding

After ASR comes Natural Language Understanding (NLU). NLU assigns meaning to the transcription producing an "intent" which is an instruction to the Alexa system to tell it what to respond to. In our example, the intent is identified as "weather" and sent to the designated data source where information is pulled for location (Seattle) and time (today). Intents may also be used to improve the NLU process.

Text-to-Speech (TTS)

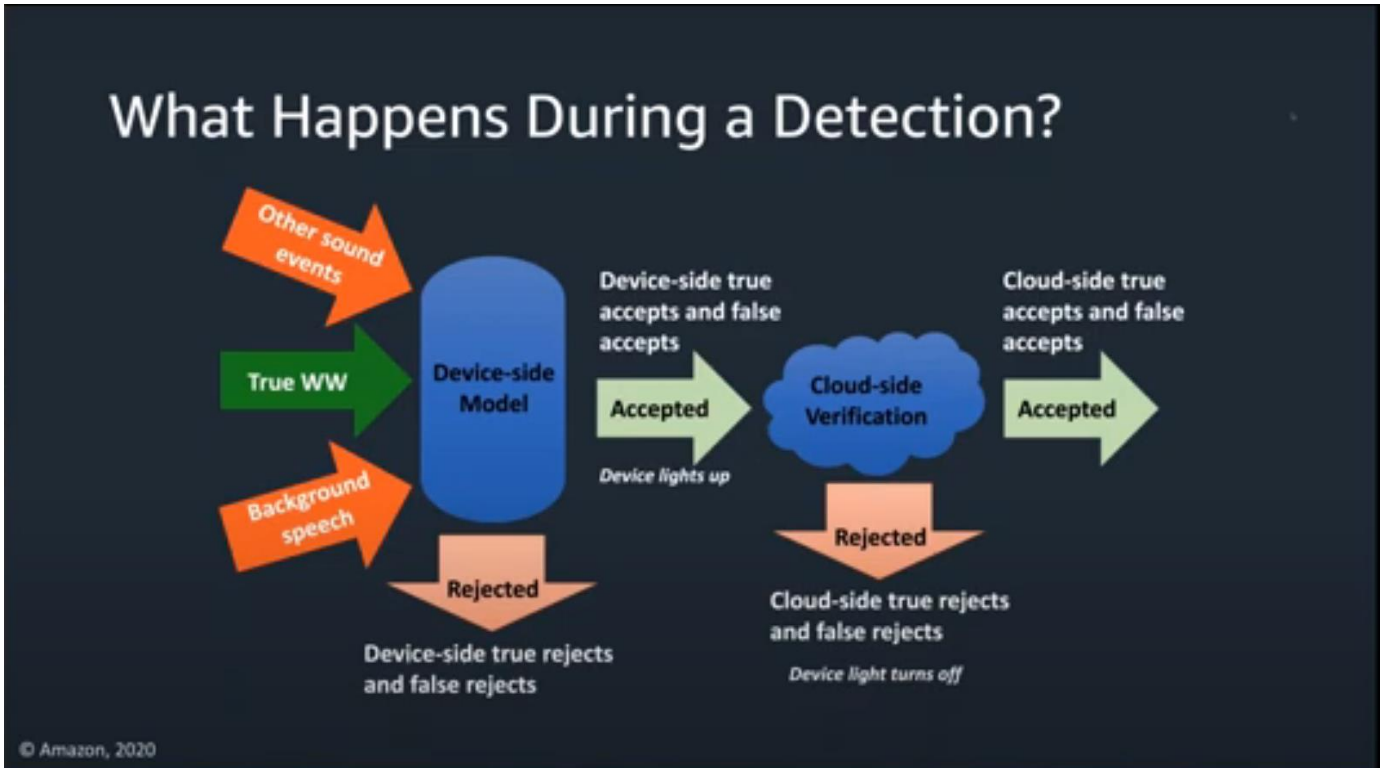
Once the necessary information is gathered, your request uses text-to-speech (TTS) technologies to generate an audio file that becomes Alexa's response. The text from Alexa's response is stored so that you can review past answers at Alexa Privacy Settings or in the Alexa app. These responses may also be reviewed by Amazon to ensure the device is providing the most relevant answers to you, and to make sure the TTS system is translating text to speech to the best of its ability.

Data storage

You have control over your Alexa experience. We make it easy for you to update your privacy settings. You can always view and manage your Alexa interactions at Alexa Privacy Settings or in the Alexa app (Settings > Alexa Privacy). If you're enrolled in voice ID, you can ask, "Alexa, what are my privacy settings?" to review how long your voice recordings are currently saved and whether or not they're used to develop new features or manually reviewed to help improve Alexa. You can also say, "Alexa, update my privacy settings" at any time to make updates.

Ex.1029 / Page 185 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	For example, Amazon's device-side model and cloud-side verification perform "true accepts" of a wake word.[154]  This claim element is literally present, as described herein, but is also satisfied under the doctrine of equivalents because the structure and functionality provided by Amazon is being used to satisfy this claim limitation and therefore there is no vitiation of this element, and the Accused Instrumentalities incorporate

[154] *See* KAIFI0007814 *(The Science Behind Alexa's Wake Word Detection*, https://www.youtube.com/watch?v=kYrw2cC9dvs).

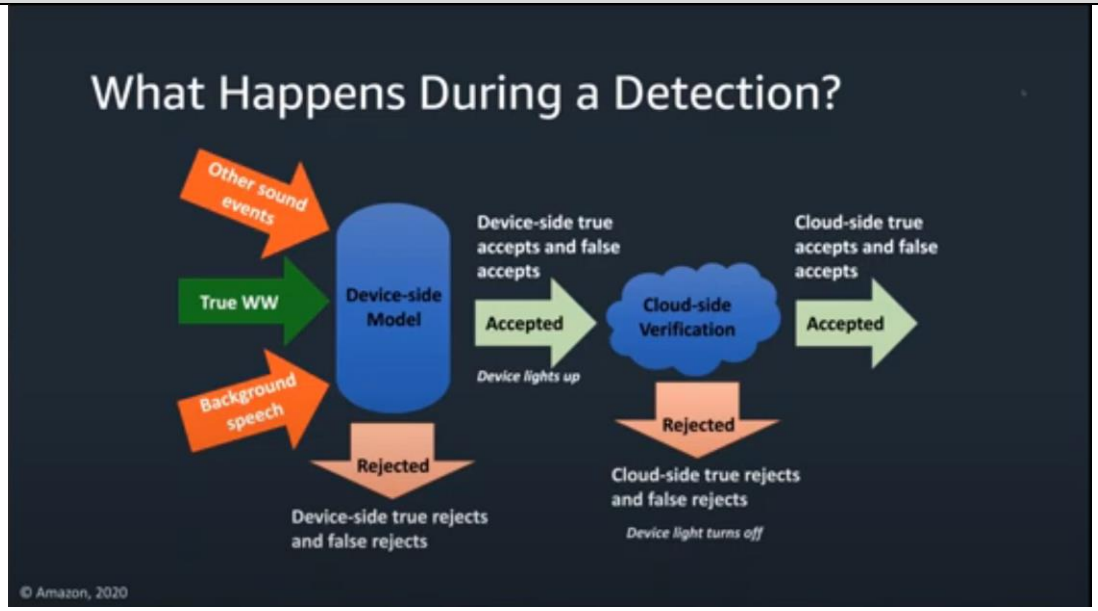
Ex.1029 / Page 186 of 347
AMAZON.COM, INC.

Claim 9	Accused Instrumentality(ies)
	structure and functionality that is not substantially different from the requirements of this limitation because they achieve substantially or exactly the same function, in substantially or exactly the same way as set forth in the claim element (as described herein), and achieves substantially or exactly the same result.
recognize the input speech as the preset call command according to a result of the comparison,	The Accused Instrumentalities are configured to recognize the input speech as the preset call command according to a result of the comparison. For example, the Accused Instrumentalities recognize, through a combination of hardware, processing, and software functions both on devices and on the cloud server, whether the input speech or wake word is a true accept or a true accept.[155]

[155] *See* KAIFI0007814 *(The Science Behind Alexa's Wake Word Detection*, https://www.youtube.com/watch?v=kYrw2cC9dvs).

Ex.1029 / Page 187 of 347
AMAZON.COM, INC.

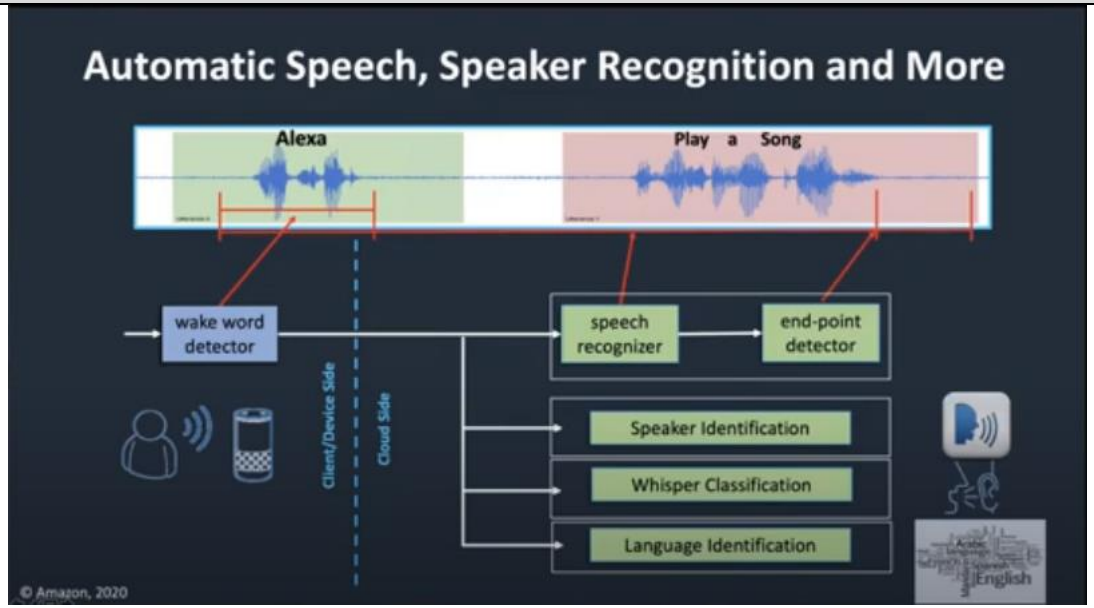
Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	 This is further illustrated below.[156]

[156] *See* KAIFI0007816 *(How Alexa Uses Automatic Speech Recognition*, https://www.youtube.com/watch?v=8drK6Oyv1a0&t=296s).

Ex.1029 / Page 188 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	 For example, Amazon describes a voice request.[157]

[157] *See* KAIFI0002023 *(Follow the journey of a voice request*, https://www.amazon.com/b/?node=23608618011).

Ex.1029 / Page 189 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Ex.1029 / Page 190 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	For example, Amazon uses keyword spotting.[158] Echo devices use built-in technology called "keyword spotting" that matches spoken audio to the acoustic patterns of the wake word. Simply put, Echo devices are designed by default to detect only the sound waves of your chosen wake word, and everything else is ignored. For example, Amazon uses wake words.[159] Wake words The list of wake words informs Alexa of the possible valid wake words that a device might be set to listen for through the `SetWakeWords`, `WakeWordsReport`, and `WakeWordsChanged` messages. Currently, the only wake word available for Alexa Built-in devices is `ALEXA`, which applies to every possible locale for a device. Therefore, specify `ALEXA` in the global `DEFAULT` scope. For example, Amazon uses speech recognition.[160]

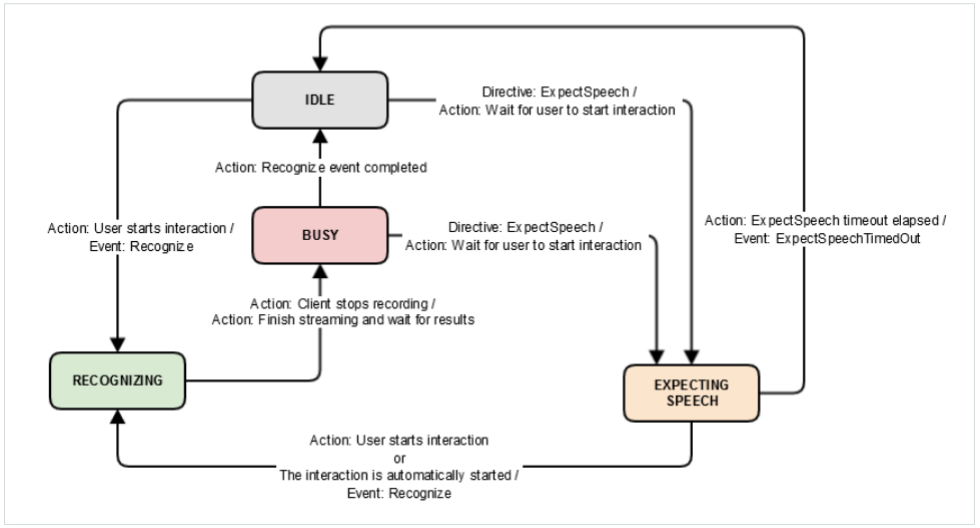
[158] *See* KAIFI0002092 *(How Alexa works: Wake word*, https://www.amazon.com/b?ie=UTF8&node=23608571011).

[159] *See* KAIFI0004342 *(How Our Scientists are Making Alexa Smarter* (Mar. 29, 2018), https://www.aboutamazon.com/news/devices/how-our-scientists-are-making-alexa-smarter).

[160] *See* KAIFI0002023 *(Follow the journey of a voice request*, https://www.amazon.com/b/?node=23608618011).

Ex.1029 / Page 191 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	The following diagram illustrates the expected transitions among the four `SpeechRecognizer` states:  For example, Amazon provides wake word verification.[161] Cloud-based wake word verification improves wake word accuracy for Alexa Built-in products by reducing false wakes caused by words that sound similar to the wake word. Examples of words that might cause a false wake for "Alexa" include "Alex", "election", "Alexis". Cloud-based wake word verification also detects media mentions of the wake word, such as, the mention of "Alexa" in an Amazon commercial.

[161] *See* KAIFI0002012 *(Enable Wake Word Verification*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/enable-cloud-based-wake-word-verification.html).

Ex.1029 / Page 192 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	The wake word engine on a device performs the initial wake word detection, and then the cloud verifies the wake word. If the cloud detects a false wake, the Alexa Voice Service (AVS) sends a `StopCapture` directive to the device in the downchannel that instructs it to close the audio stream, and if applicable, to turn off the blue LEDs to indicate that Alexa has stopped listening. Voice-initiated devices begin to stream user speech to AVS when the wake word engine detects a spoken wake word, such as "Alexa." The stream closes when the user stops speaking or when AVS identifies a user intent, and AVS returns a `StopCapture` directive to the device. Cloud-based wake word verification has the following requirements: • **Wake word** – Include the wake word in the stream so that AVS can perform cloud-based wake word verification, which reduces false wakes. If AVS can't detect the wake word detected during cloud-based wake word verification, AVS discards the utterance. • **500 milliseconds of pre-roll** – Pre-roll is the audio captured before AVS detects the wake word and helps calibrate the ambient noise level of the recording to enhance speech recognition. • **User speech** – Any user speech that the device captures until receiving a `StopCapture` directive. This allows AVS to verify the wake word included in the stream, reducing the number of erroneous responses due to false wakes. To learn how to implement a shared memory ring buffer for writing and reading audio samples and to include the start and stop indices for wake word detection in each `Recognize` event sent to AVS, see the Requirements for Cloud-Based Wake Word Verification. For example, Amazon uses wake word, keyword spotting, cloud verification and encryption, automatic speech recognition (ASR), natural language understanding (NLU), text-to-speech (TTS), and data storage technologies.[162]

[162] *See* KAIFI0002023 *(Follow the journey of a voice request*, https://www.amazon.com/b/?node=23608618011).

Ex.1029 / Page 193 of 347
AMAZON.COM, INC.

The wake word

Think of the wake word as a verbal cue that makes things happen. In the classic tale from The Arabian Nights, an invisible door into a mountain can only be opened by saying the phrase "Open Sesame." Use the phrase, the door opens. Don't use the phrase, the door stays shut. When you have a request for Alexa, you first need to say your chosen wake word, which by default is "Alexa." Only after your Echo device detects the wake word is Alexa listening to your requests.

When it comes to privacy, there should be no surprises. You'll always be able to tell when Alexa is listening to your request because a light indicator will appear on your Echo device or an audible tone will sound. Think of the "On the Air" signs that light up in television studios during a broadcast. These indicators notify you that your device has detected the wake word and Alexa is now processing your request.

Keyword spotting

Echo devices use built-in technology called "keyword spotting" that matches spoken audio to the acoustic patterns of the wake word. Simply put, Echo devices are designed by default to detect only the sound waves of your chosen wake word, and everything else is ignored. Like water through a strainer, all other audio (people talking, faucets running, birds chirping) passes through the device until the wake word is "caught" and sent to Amazon's secure cloud, where meaning is assigned to your request. You can choose to opt-in to features like Alexa Emergency Assist that allow your device to detect more than your chosen wake word—for example, the sound of smoke alarms—but you would need to update your settings to do so.

Cloud verification and encryption

When your Echo device detects the wake word, it sends your request to Amazon's secure cloud where audio is reanalyzed to verify the wake word was spoken. If this cloud software verification is unable to confirm the wake word was spoken, the Alexa system stops processing the audio. If the wake word is verified (or Alexa is activated using the action button), your request will be processed using several sophisticated algorithms that allow Alexa to respond appropriately. All of your interactions with Alexa are encrypted in transit throughout this process.

Automatic Speech Recognition (ASR)

After wake word verification, your voice request undergoes Automatic Speech Recognition (ASR) which transcribes audio into text. Using our example request ("Alexa, what's the weather in Seattle?"), transcriptions could include one or several of the following:

"what is the weather in seattle" "watt is the weather in seattle" "what is the whether in seattle"

Alexa then chooses to act on the transcription that is most likely to be correct and stores it in Amazon's secure cloud. You can view and delete it at any time at Alexa Privacy Settings or in the Alexa app (Settings > Alexa Privacy > Review Voice History). Voice recordings and transcriptions may be used to improve the ASR process. Learn how data helps Alexa improve.

Natural Language Understanding

After ASR comes Natural Language Understanding (NLU). NLU assigns meaning to the transcription producing an "intent" which is an instruction to the Alexa system to tell it what to respond to. In our example, the intent is identified as "weather" and sent to the designated data source where information is pulled for location (Seattle) and time (today). Intents may also be used to improve the NLU process.

Text-to-Speech (TTS)

Once the necessary information is gathered, your request uses text-to-speech (TTS) technologies to generate an audio file that becomes Alexa's response. The text from Alexa's response is stored so that you can review past answers at Alexa Privacy Settings or in the Alexa app. These responses may also be reviewed by Amazon to ensure the device is providing the most relevant answers to you, and to make sure the TTS system is translating text to speech to the best of its ability.

Data storage

You have control over your Alexa experience. We make it easy for you to update your privacy settings. You can always view and manage your Alexa interactions at Alexa Privacy Settings or in the Alexa app (Settings > Alexa Privacy). If you're enrolled in voice ID, you can ask, "Alexa, what are my privacy settings?" to review how long your voice recordings are currently saved and whether or not they're used to develop new features or manually reviewed to help improve Alexa. You can also say, "Alexa, update my privacy settings" at any time to make updates.

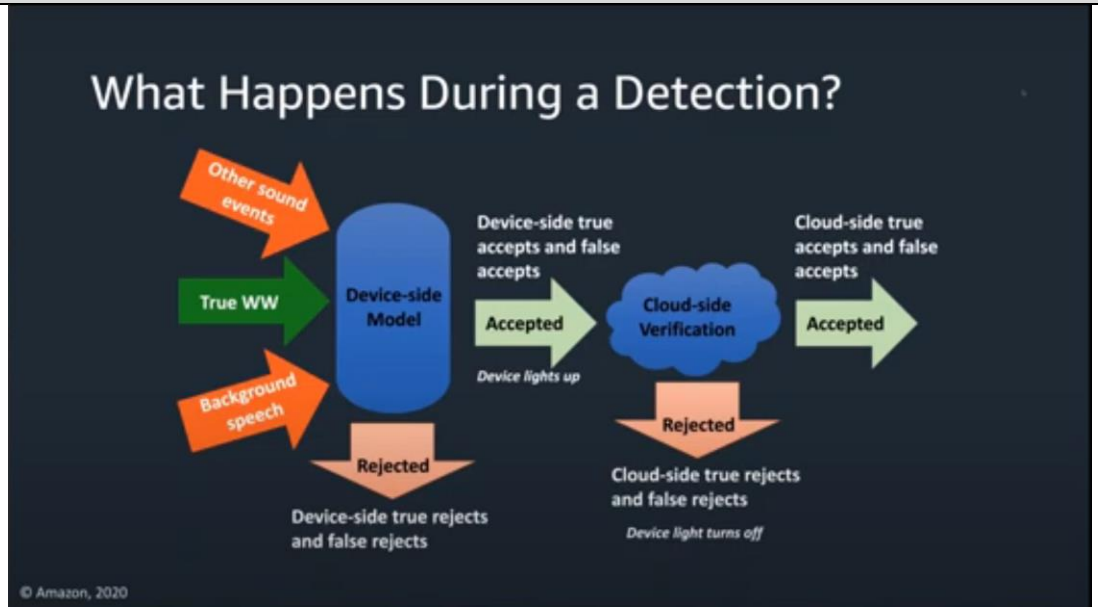
Ex.1029 / Page 194 of 347
AMAZON.COM, INC.

Claim 9	Accused Instrumentality(ies)
	This claim element is literally present, as described herein, but is also satisfied under the doctrine of equivalents because the structure and functionality provided by Amazon is being used to satisfy this claim limitation and therefore there is no vitiation of this element, and the Accused Instrumentalities incorporate structure and functionality that is not substantially different from the requirements of this limitation because they achieve substantially or exactly the same function, in substantially or exactly the same way as set forth in the claim element (as described herein), and achieves substantially or exactly the same result.
wherein the continuous speech recognition unit is configured to recognize the subsequent input speech as an actual command if the input speech is recognized as the preset call command by the preset call command recognition unit, and	In the Accused Instrumentalities, the continuous speech recognition unit is configured to recognize the subsequent input speech as an actual command if the input speech is recognized as the preset call command by the preset call command recognition unit. For example, the Accused Instrumentalities recognize, through a combination of hardware, processing, and software functions both on devices and on the cloud server, whether the input speech or wake word is a true accept or a true accept.[163]

[163] *See* KAIFI0007814 *(The Science Behind Alexa's Wake Word Detection*, https://www.youtube.com/watch?v=kYrw2cC9dvs).

Ex.1029 / Page 195 of 347
AMAZON.COM, INC.

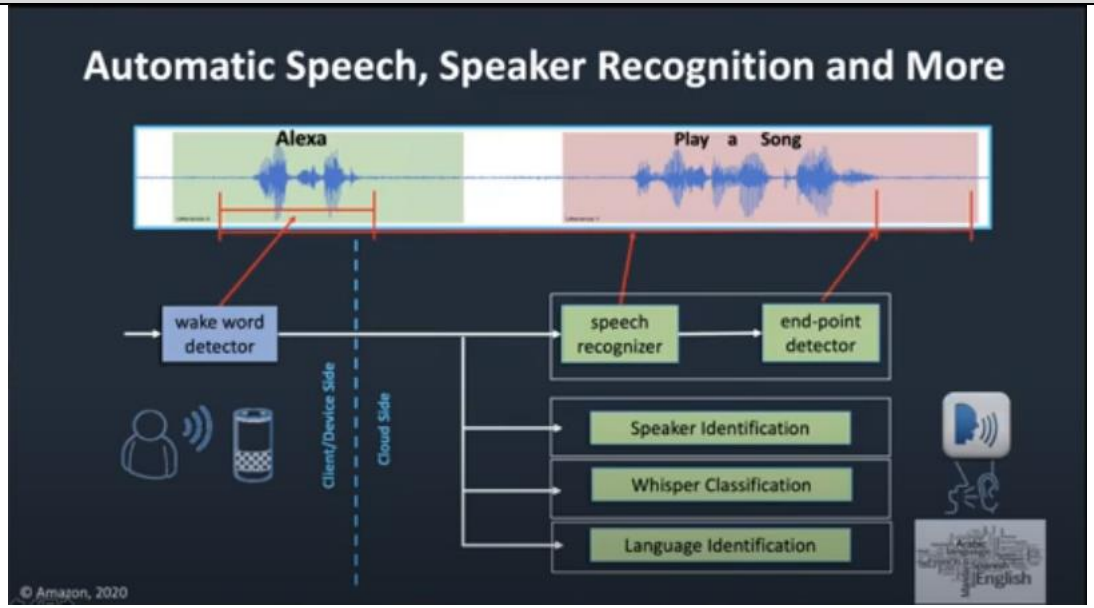
Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	 This is further illustrated below.[164]

[164] *See* KAIFI0007816 *(How Alexa Uses Automatic Speech Recognition*, https://www.youtube.com/watch?v=8drK6Oyv1a0&t=296s).

Ex.1029 / Page 196 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	 For example, Amazon describes a voice request.[165]

[165] *See* KAIFI0002023 *(Follow the journey of a voice request*, https://www.amazon.com/b/?node=23608618011).

Ex.1029 / Page 197 of 347
AMAZON.COM, INC.

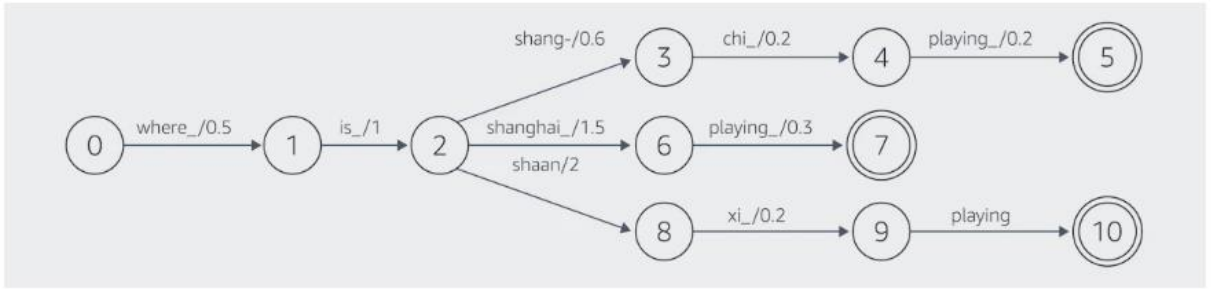
Ex.1029 / Page 198 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	For example, Amazon describes its ASR model.[166]

System architecture

Our on-device ASR model takes in an acoustic speech signal and outputs a set of hypotheses about what the speaker said, ranked according to probability. We represent those hypotheses as a *lattice* — a graph whose edges represent recognized words and the probability that a given word follows from the previous one.



An example of a lattice representing ASR hypotheses.

[166] *See* KAIFI0006791 (Ariya Rastrow, *On-device speech processing makes Alexa faster, lower-bandwidth* (Jan. 25, 2022), https://www.amazon.science/blog/on-device-speech-processing-makes-alexa-faster-lower-bandwidth).

Ex.1029 / Page 199 of 347
AMAZON.COM, INC.

Claim 9	Accused Instrumentality(ies)
	For example, Amazon uses cloud verification and encryption, automatic speech recognition (ASR), natural language understanding (NLU), text-to-speech (TTS), and data storage technologies.[167]

[167] *See* KAIFI0002023 *(Follow the journey of a voice request*, https://www.amazon.com/b/?node=23608618011).

Ex.1029 / Page 200 of 347
AMAZON.COM, INC.

The wake word

Think of the wake word as a verbal cue that makes things happen. In the classic tale from The Arabian Nights, an invisible door into a mountain can only be opened by saying the phrase "Open Sesame." Use the phrase, the door opens. Don't use the phrase, the door stays shut. When you have a request for Alexa, you first need to say your chosen wake word, which by default is "Alexa." Only after your Echo device detects the wake word is Alexa listening to your requests.

When it comes to privacy, there should be no surprises. You'll always be able to tell when Alexa is listening to your request because a light indicator will appear on your Echo device or an audible tone will sound. Think of the "On the Air" signs that light up in television studios during a broadcast. These indicators notify you that your device has detected the wake word and Alexa is now processing your request.

Keyword spotting

Echo devices use built-in technology called "keyword spotting" that matches spoken audio to the acoustic patterns of the wake word. Simply put, Echo devices are designed by default to detect only the sound waves of your chosen wake word, and everything else is ignored. Like water through a strainer, all other audio (people talking, faucets running, birds chirping) passes through the device until the wake word is "caught" and sent to Amazon's secure cloud, where meaning is assigned to your request. You can choose to opt-in to features like Alexa Emergency Assist that allow your device to detect more than your chosen wake word—for example, the sound of smoke alarms—but you would need to update your settings to do so.

Cloud verification and encryption

When your Echo device detects the wake word, it sends your request to Amazon's secure cloud where audio is reanalyzed to verify the wake word was spoken. If this cloud software verification is unable to confirm the wake word was spoken, the Alexa system stops processing the audio. If the wake word is verified (or Alexa is activated using the action button), your request will be processed using several sophisticated algorithms that allow Alexa to respond appropriately. All of your interactions with Alexa are encrypted in transit throughout this process.

Automatic Speech Recognition (ASR)

After wake word verification, your voice request undergoes Automatic Speech Recognition (ASR) which transcribes audio into text. Using our example request ("Alexa, what's the weather in Seattle?"), transcriptions could include one or several of the following:

"what is the weather in seattle" "watt is the weather in seattle" "what is the whether in seattle"

Alexa then chooses to act on the transcription that is most likely to be correct and stores it in Amazon's secure cloud. You can view and delete it at any time at Alexa Privacy Settings or in the Alexa app (Settings > Alexa Privacy > Review Voice History). Voice recordings and transcriptions may be used to improve the ASR process. Learn how data helps Alexa improve.

Natural Language Understanding

After ASR comes Natural Language Understanding (NLU). NLU assigns meaning to the transcription producing an "intent" which is an instruction to the Alexa system to tell it what to respond to. In our example, the intent is identified as "weather" and sent to the designated data source where information is pulled for location (Seattle) and time (today). Intents may also be used to improve the NLU process.

Text-to-Speech (TTS)

Once the necessary information is gathered, your request uses text-to-speech (TTS) technologies to generate an audio file that becomes Alexa's response. The text from Alexa's response is stored so that you can review past answers at Alexa Privacy Settings or in the Alexa app. These responses may also be reviewed by Amazon to ensure the device is providing the most relevant answers to you, and to make sure the TTS system is translating text to speech to the best of its ability.

Data storage

You have control over your Alexa experience. We make it easy for you to update your privacy settings. You can always view and manage your Alexa interactions at Alexa Privacy Settings or in the Alexa app (Settings > Alexa Privacy). If you're enrolled in voice ID, you can ask, "Alexa, what are my privacy settings?" to review how long your voice recordings are currently saved and whether or not they're used to develop new features or manually reviewed to help improve Alexa. You can also say, "Alexa, update my privacy settings" at any time to make updates.

Ex.1029 / Page 201 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	For example, Amazon uses a speech recognizer.[168] The `SpeechRecognizer` interface is the core interface of the Alexa Voice Service (AVS) and exposes directives and events for capturing and interacting with user speech. This page discusses the concepts and process flows for the `SpeechRecognizer` interface. SpeechRecognizer functionality Every user utterance leverages `SpeechRecognizer`, including the following interaction types between an Alexa Built-in device and AVS: • **User Speech Capture**: Captures user speech from an Alexa Built-in device. • **User Speech Prompting**: Prompts a user for more speech input when needed by Alexa to deliver an appropriate response. • **Interaction Initiation Communication**: Enables a device to inform AVS of how a user initiated an Alexa interaction, such as press-and-hold, tap-and-release, or voice-initiated/wake word enabled. See Device Form Factor and Alexa Interaction. • **ASR Profile Selection**: SpeechRecognizers chooses the appropriate Automatic Speech Recognition (ASR) profile for your product, which allows Alexa to understand user speech and respond with precision. See Automatic Speech Recognition (ASR) profile

[168] *See* KAIFI0007093 *(Understand the AVS SpeechRecognizer*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/avs-speechrecognizer-concepts.html).

Ex.1029 / Page 202 of 347
AMAZON.COM, INC.

State diagram

The following diagram illustrates state changes driven by `SpeechRecognizer` components. Boxes represent `SpeechRecognizer` states, and the connectors represent state transitions.

`SpeechRecognizer` has the following states:

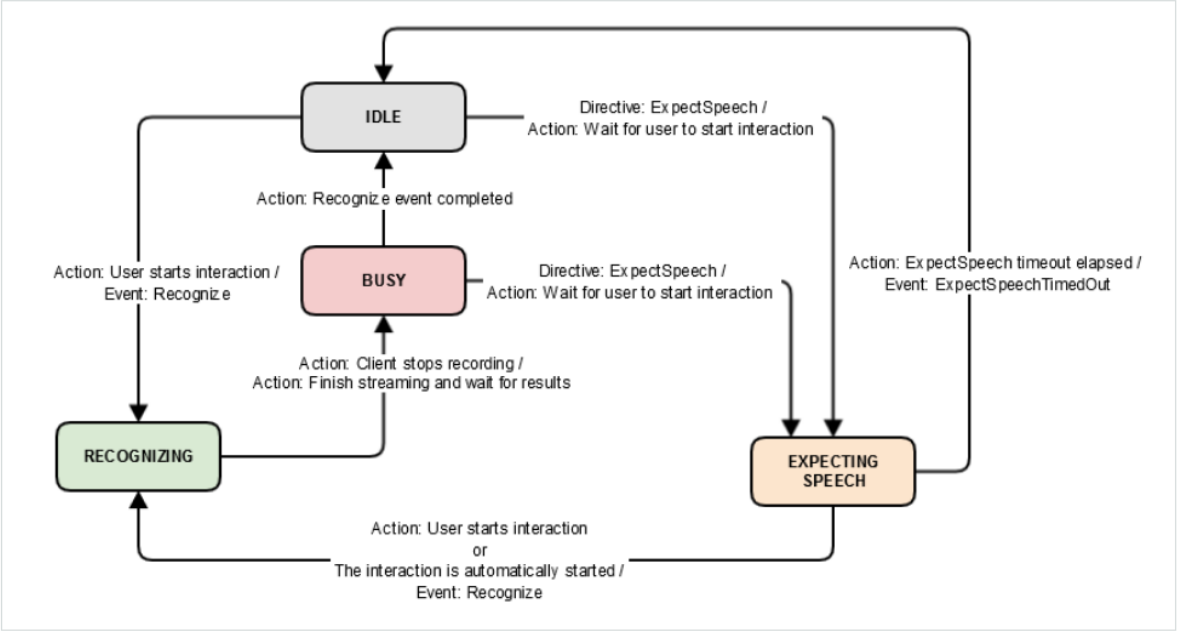
- **IDLE:** When not actively processing speech, the `SpeechRecognizer` is in an "idle" state. The idle state occurs under the following conditions:
  - Before capturing user speech.
  - Returning to an idle state after concluding a speech interaction with Alexa.
  - When an `ExpectSpeechTimedOut` event elapses.

  Note: In a multi-turn Alexa interaction, if Alexa requires more user speech input, `SpeechRecognizer` should transition from the idle state to the `ExpectSpeech` state without the user starting a new interaction.

- **RECOGNIZING:** When a user begins interacting with your client, specifically when the client streams captured audio to AVS, `SpeechRecognizer` should transition from the idle state to the `Recognize` state. `SpeechRecognizer` should remain in the `Recognize` until the client stops recording speech or finishes streaming, at which point your `SpeechRecognizer` component should transition from the `Recognize` state to the "busy" state.
- **BUSY:** While processing the speech request, `SpeechRecognizer` should be in the "busy" state. You cannot start another speech request until `SpeechRecognizer` transitions out of the busy state. From the *busy* state, `SpeechRecognizer` transitions to the idle state if Alexa processes and completes the request or to the `ExpectSpeech` state, if Alexa requires more speech input from the user.
- **EXPECTING SPEECH:** `SpeechRecognizer` should be in the `ExpectSpeech` state when Alexa requires more audio input from a user. From `ExpectSpeech`, `SpeechRecognizer` should transition to the `Recognize` when a user interaction occurs, or the interaction is automatically started on behalf of the user. It should transition to the idle state if Alexa detects no user interaction within the specified timeout window.

Ex.1029 / Page 203 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	The following diagram illustrates the expected transitions among the four `SpeechRecognizer` states:  For example, Amazon uses a wake word engine.[169]

[169] *See* KAIFI0007141 *(Wake Word Verification Requirements*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/implement-ww-verification.html).

Ex.1029 / Page 204 of 347
AMAZON.COM, INC.

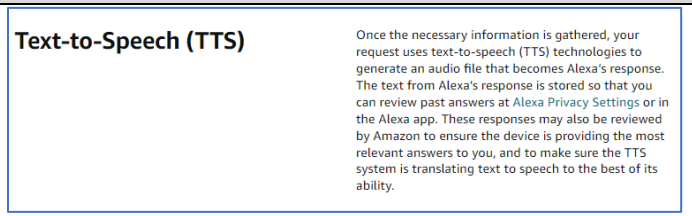
Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	For example, Amazon provides wake word verification.[170] For example, Amazon uses cloud verification and encryption, automatic speech recognition (ASR), natural language understanding (NLU), text-to-speech (TTS), and data storage technologies on the recognized actual subsequent input speech command.[171] **Automatic Speech Recognition (ASR)** After wake word verification, your voice request undergoes Automatic Speech Recognition (ASR) which transcribes audio into text. Using our example request (*"Alexa, what's the weather in Seattle?"*), transcriptions could include one or several of the following: • *"what is the weather in seattle"* • *"watt is the weather in seattle"* • *"what is the whether in seattle"* Alexa then chooses to act on the transcription that is most likely to be correct and stores it in Amazon's secure cloud. You can view and delete it at any time at Alexa Privacy Settings or in the Alexa app (Settings > Alexa Privacy > Review Voice History). Voice recordings and transcriptions may be used to improve the ASR process. Learn how data helps Alexa improve. **Natural Language Understanding** After ASR comes Natural Language Understanding (NLU). NLU assigns meaning to the transcription producing an "intent" which is an instruction to the Alexa system to tell it what to respond to. In our example, the intent is identified as "weather" and sent to the designated data source where information is pulled for location (Seattle) and time (today). Intents may also be used to improve the NLU process.

[170] *See* KAIFI0002012 *(Enable Wake Word Verification*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/enable-cloud-based-wake-word-verification.html).

[171] *See* KAIFI0002023 *(Follow the journey of a voice request*, https://www.amazon.com/b/?node=23608618011).

Ex.1029 / Page 205 of 347
AMAZON.COM, INC.

Claim 9	Accused Instrumentality(ies)
	 For example, when a customer says the wake word, subsequent phrases are processed and stored by Amazon in the cloud to respond to the customer's request.[172] Alexa is designed to get smarter every day—this is accomplished through the power of machine learning and the cloud. The more a customer uses Alexa, the more it adapts to their speech patterns, vocabulary, and personal preferences. When the customer says the wake word, their subsequent phrases are processed and stored in the cloud to respond to the customer's request and to improve the customer's experience and our services, including training our speech recognition and natural language understanding systems so Alexa can better understand customers' requests. This claim element is literally present, as described herein, but is also satisfied under the doctrine of equivalents because the structure and functionality provided by Amazon is being used to satisfy this claim limitation and therefore there is no vitiation of this element, and the Accused Instrumentalities incorporate structure and functionality that is not substantially different from the requirements of this limitation because they achieve substantially or exactly the same function, in substantially or exactly the same way as set forth in the claim element (as described herein), and achieves substantially or exactly the same result.
wherein the preset call command recognition unit	In the Accused Instrumentalities, the preset call command recognition unit comprises a token passing unit.

[172] *See* KAIFI0007380 *(White paper – Alexa Privacy and Data Handling Overview*, https://d1.awsstatic.com/product-marketing/A4B/White%20Paper%20-%20Alexa%20Privacy%20and%20Data%20Handling%20Overview.pdf)

Ex.1029 / Page 206 of 347
AMAZON.COM, INC.

Claim 9	Accused Instrumentality(ies)
comprises a token passing unit configured to assume the input speech to be the preset call command according to the result of the comparison using a minimum recognition network composed of a silence interval accompanied by noise and the preset call command.	The token passing unit is configured to assume the input speech to be the preset call command according to the result of the comparison using a minimum recognition network composed of a silence interval accompanied by noise and the preset call command. For example, Amazon uses a user speech stream. [173]

[173] *See* KAIFI0007141 *(Wake Word Verification Requirements*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/implement-ww-verification.html).

Ex.1029 / Page 207 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)

Voice-initiated devices begin to stream user speech to AVS when the wake word engine detects a spoken wake word, such as "Alexa." The stream closes when the user stops speaking or when AVS identifies a user intent, and AVS returns a `StopCapture` directive to the device. Cloud-based wake word verification has the following requirements:

- **Wake word** – Include the wake word in the stream so that AVS can perform cloud-based wake word verification, which reduces false wakes. If AVS can't detect the wake word detected during cloud-based wake word verification, AVS discards the utterance.
- **500 milliseconds of pre-roll** – Pre-roll is the audio captured before AVS detects the wake word and helps calibrate the ambient noise level of the recording to enhance speech recognition.
- **User speech** – Any user speech that the device captures until receiving a `StopCapture` directive. This allows AVS to verify the wake word included in the stream, reducing the number of erroneous responses due to false wakes.

Read the following sections to learn how to implement a shared memory ring buffer for writing and reading audio samples and to include the start and stop indices for wake word detection in each `Recognize` event sent to AVS.

Wake Word indexes

The `Recognize` event streams user speech to AVS and translates that speech into one or more directives. This multipart message that includes a JSON-formatted object and a binary audio attachment. The stream must contain the pre-roll, wake word, and the user's utterance. Use chunking to reduce latency; the stream should contain 10ms of captured audio per chunk (320 bytes).

When the user request is voice-initiated, the `initiator.type` SHALL be `WAKEWORD` and the `Recognize` event must contain the start and stop indices for the occurrence of the wake word. The timestamps are relative to the start of the audio stream and the units are in samples.

⚠ **Important:** The pre-roll and wake word start and stop indexes must be precise. The start index should be accurate to within 50ms of wake word detection. The end index should be accurate to within 150ms of the end of the detected wake word.

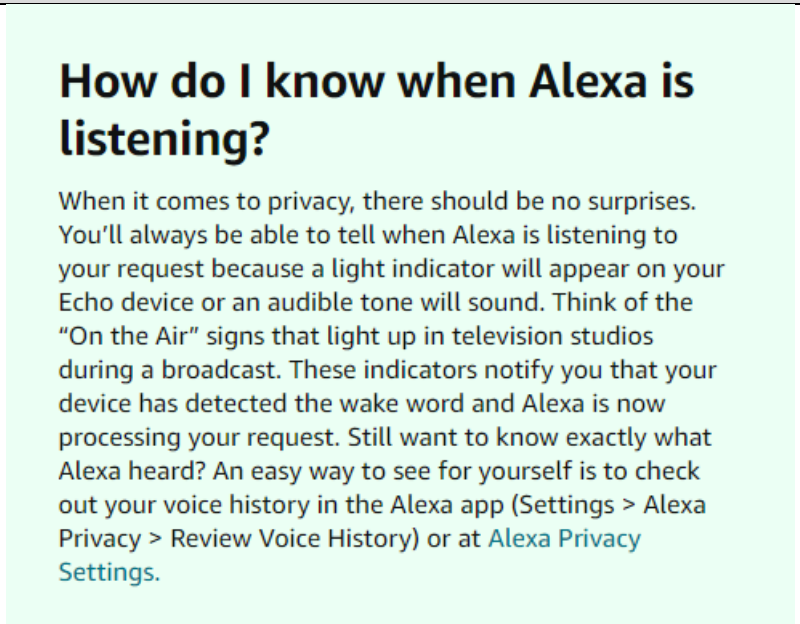
For example, a wake word that starts at the 500 milliseconds mark and has a duration of 500 millisecond has the following indexes with a 16 kHz sample rate:

- **startIndexInSamples** - `8000`
- **endIndexInSamples** - `16000`

For example, Amazon's Alexa indicators notify that a device has detected the wake word and is now processing a request. [174]

[174] *See* KAIFI0002092 *(How Alexa works: Wake word*, https://www.amazon.com/b?ie=UTF8&node=23608571011).

Ex.1029 / Page 208 of 347
AMAZON.COM, INC.

Claim 9	Accused Instrumentality(ies)
	 For example, Amazon detects only the sound waves of a chosen wake word, and everything else is ignored.[175]

[175] *See id.*

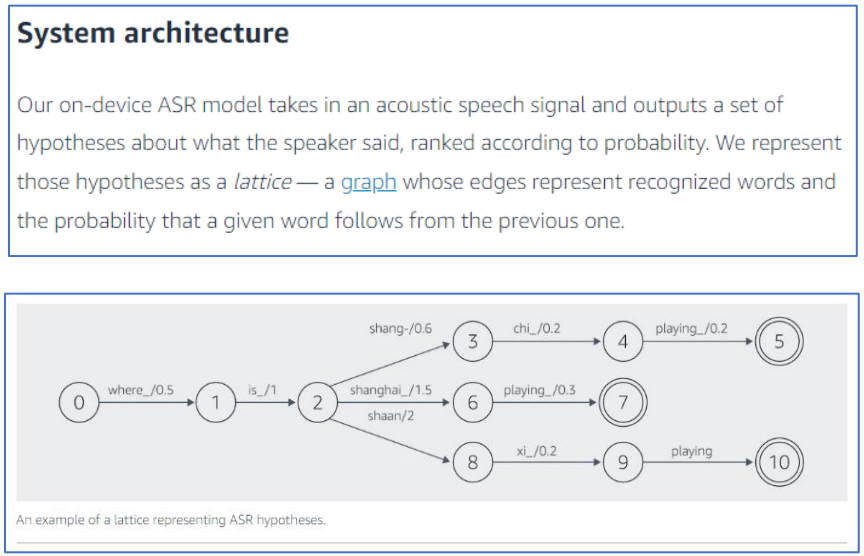
Ex.1029 / Page 209 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	But how can my device detect the wake word without hearing the rest of my conversations? Echo devices use built-in technology called "keyword spotting" that matches spoken audio to the acoustic patterns of the wake word. Simply put, Echo devices are designed by default to detect only the sound waves of your chosen wake word, and everything else is ignored. Like water through a strainer, all other audio (people talking, faucets running, birds chirping) passes through the device until the wake word is "caught" and sent to Amazon's secure cloud, where your request is reanalyzed to verify the wake word was spoken. You can choose to opt-in to features like Alexa Emergency Assist that allow your device to detect more than your chosen wake word—for example, the sound of smoke alarms—but you would need to update your settings to do so. Want to keep nerding out? Follow the journey of a voice request.

Ex.1029 / Page 210 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	For example, Amazon provides an ASR model.[176] 

[176] *See* KAIFI0006791 (Ariya Rastrow, *On-device speech processing makes Alexa faster, lower-bandwidth* (Jan. 25, 2022), https://www.amazon.science/blog/on-device-speech-processing-makes-alexa-faster-lower-bandwidth).

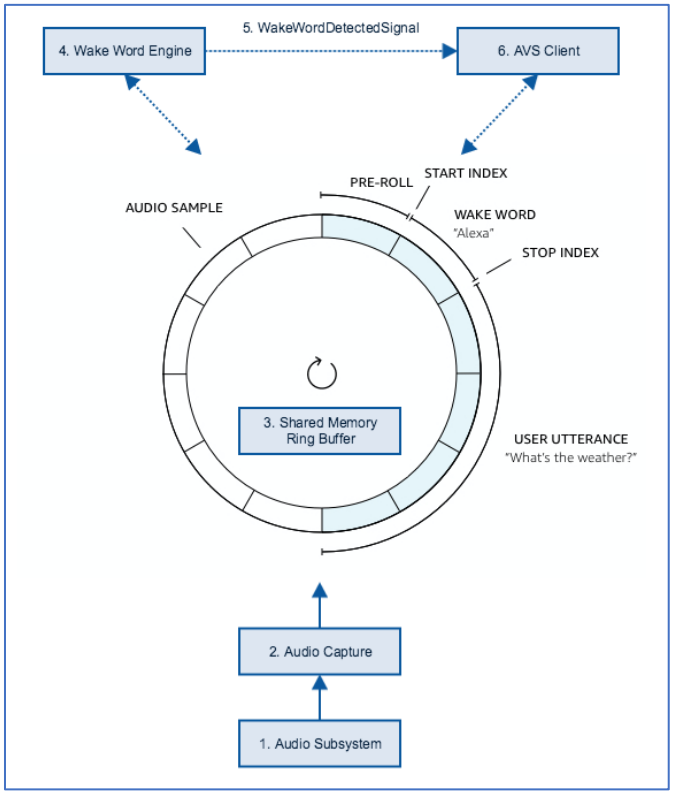
Ex.1029 / Page 211 of 347
AMAZON.COM, INC.

Claim 9	Accused Instrumentality(ies)
	With cloud-based ASR, encrypted audio streams to the cloud in small snippets called "frames". With on-device ASR, only the lattice is sent to the cloud, where a large and powerful neural language model reranks the hypotheses. The lattice can't be sent until the customer has finished speaking, as words later in a sequence can dramatically change the overall probability of a hypothesis. The model that determines when the customer has finished speaking is called an end-pointer. End-pointers offer a natural trade-off between accuracy and latency: an aggressive end-pointer will initiate speech processing earlier, but it might cut the speaker off prematurely, resulting in a poor customer experience. On the device, we in fact run two end-pointers: One is a speculative end-pointer that we have tuned to be about 200 milliseconds faster than the final end-pointer, so we can initiate downstream processing — such as natural-language understanding (NLU) — ahead of the final end-pointed ASR result. In exchange for speed, however, we trade off a little accuracy. The final end-pointer takes longer to make a decision but is more accurate. In cases in which the first end-pointer cuts speech off too early, the final end-pointer sends a revised lattice and the instruction to reset downstream processing. In the large majority of cases, however, the aggressive end-pointer is correct, which reduces user-perceived latency, since downstream tasks are initiated earlier. For example, Amazon uses a shared memory ring buffer.[177]

[177] *See* KAIFI0007141 *(Wake Word Verification Requirements*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/implement-ww-verification.html).

Ex.1029 / Page 212 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

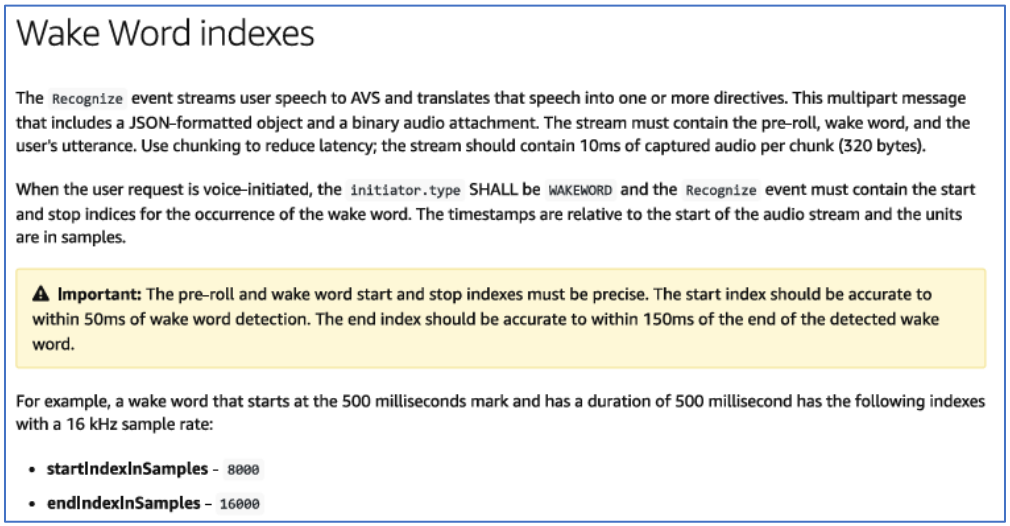
Claim 9	Accused Instrumentality(ies)
	Shared Memory Ring Buffer A shared memory ring buffer satisfies the requirement to stream a continuous utterance to AVS that includes 500 milliseconds of pre-roll, the wake word, and the user's request for cloud-based wake word verification. This reduces latency and the need to copy audio samples multiple times. The following diagram illustrates one approach; alternative implementations are allowed as long as the requisite functionality is maintained: 

Ex.1029 / Page 213 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	1. **Audio Subsystem** - Middleware, such as Advanced Linux Sound Architecture (ALSA), that provides APIs to open a stream from a recording device, and receive an audio sample stream. 2. **Audio Capture** - A process that opens a recording device, and writes audio samples into the shared memory ring buffer. 3. **Shared Memory Ring Buffer** - A shared memory ring buffer object that is thread safe, and accessible by multiple processes. Access to this memory block should support at least one writer and two readers simultaneously. You may choose to implement these executables as a shared library. • **Write API** - Used by the recording process to write audio samples to the shared memory ring buffer. Write samples should move data so that the oldest audio samples are overwritten. This example shows a typical API signature: `WriteSamples(BufferIn, BufferInLength)`. • **Read API** - Used by the wake word engine and the AVS client to read from specific locations in the shared memory ring buffer. A typical implementation represents location as a 64-bit integer. For example, reading 200 samples from the location 100 is reading the 100th to 300th samples that were written to the shared memory ring buffer. This is an example of a typical API signature: `ReadSamples(BufferIn, BufferInLength)`. **Note**: Reading from an old location that was overwritten with new audio samples will result in an unintended sample read. 4. **Wake Word Engine** - The component that reads the audio samples written to the shared memory ring buffer and analyzes the audio samples for occurrences of the wake word. The procedure of reading and processing audio samples occurs in a loop. For example, the loop reads N samples starting at location `0`; then analyzes the samples for an occurrence of the wake word. In the next iteration, it reads and analyzes the next N samples. When an occurrence of the wake word is detected, the wake word engine must identify two locations: 1) the start of the wake word, and 2) the end of the wake word. These values are stored as `startIndexInSamples` and `endIndexInSamples`. 5. **WakeWordDetectedSignal** - A signal sent from the wake word engine to the AVS client when an occurrence of the wake word is detected. This signal includes two locations, the `startIndexInSamples` and `endIndexInSamples`. 6. **AVS Client** - The client receives the `WakeWordDetectedSignal`, extracts the `startIndexInSamples`, then reads and streams audio samples to AVS including 500 milliseconds of pre-roll. • The pre-roll is the number of samples corresponding to 500 milliseconds. For example, if the audio samples were recorded at a rate of 16 kilohertz (kHz), then pre-roll is 8,000 samples.

Ex.1029 / Page 214 of 347
AMAZON.COM, INC.

Claim 9	Accused Instrumentality(ies)
	For example, Amazon uses wake word indexes.[178]  Wake Word indexes The `Recognize` event streams user speech to AVS and translates that speech into one or more directives. This multipart message that includes a JSON-formatted object and a binary audio attachment. The stream must contain the pre-roll, wake word, and the user's utterance. Use chunking to reduce latency; the stream should contain 10ms of captured audio per chunk (320 bytes). When the user request is voice-initiated, the `initiator.type` SHALL be `WAKEWORD` and the `Recognize` event must contain the start and stop indices for the occurrence of the wake word. The timestamps are relative to the start of the audio stream and the units are in samples. **Important:** The pre-roll and wake word start and stop indexes must be precise. The start index should be accurate to within 50ms of wake word detection. The end index should be accurate to within 150ms of the end of the detected wake word. For example, a wake word that starts at the 500 milliseconds mark and has a duration of 500 millisecond has the following indexes with a 16 kHz sample rate: • **startIndexInSamples** - `8000` • **endIndexInSamples** - `16000` For example, Amazon describes a voice request. [179]

[178] *See* KAIFI0007141 *(Wake Word Verification Requirements*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/implement-ww-verification.html).

[179] *See* KAIFI0002023 *(Follow the journey of a voice request*, https://www.amazon.com/b/?node=23608618011).

Ex.1029 / Page 215 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Ex.1029 / Page 216 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	For example, Amazon uses speech recognition.[180] Near-field and far-field tests are conducted in various locations, which mimic real-world user-device interaction patterns. The device's performance is evaluated under various conditions such as in silence, during audio playback from the device, or in environments with background noise. Such noise could be from music playing on another device, stationary noise such as an air conditioner, or other ambient noise. For example, Amazon uses a user speech stream.[181] Voice-initiated devices begin to stream user speech to AVS when the wake word engine detects a spoken wake word, such as "Alexa." The stream closes when the user stops speaking or when AVS identifies a user intent, and AVS returns a `StopCapture` directive to the device. Cloud-based wake word verification has the following requirements: For example, Amazon uses ASR.[182]

[180] *See* KAIFI0001353 (Felipe Pinzon, *An Amazon technical expert provides three acoustic certification-related tips for AVS device makers* (Aug. 16, 2023) https://developer.amazon.com/en-US/blogs/alexa/device-makers/2023/08/acoustic-certification-alexa-august-2023).

[181] *See* KAIFI0007141 *(Wake Word Verification Requirements*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/implement-ww-verification.html).

[182] *See* KAIFI0006791 (Ariya Rastrow, Shehzad Mevawalla, *On-device speech processing makes Alexa faster, lower-bandwidth*, AMAZON SCIENCE (Jan. 25, 2022), https://www.amazon.science/blog/on-device-speech-processing-makes-alexa-faster-lower-bandwidth).

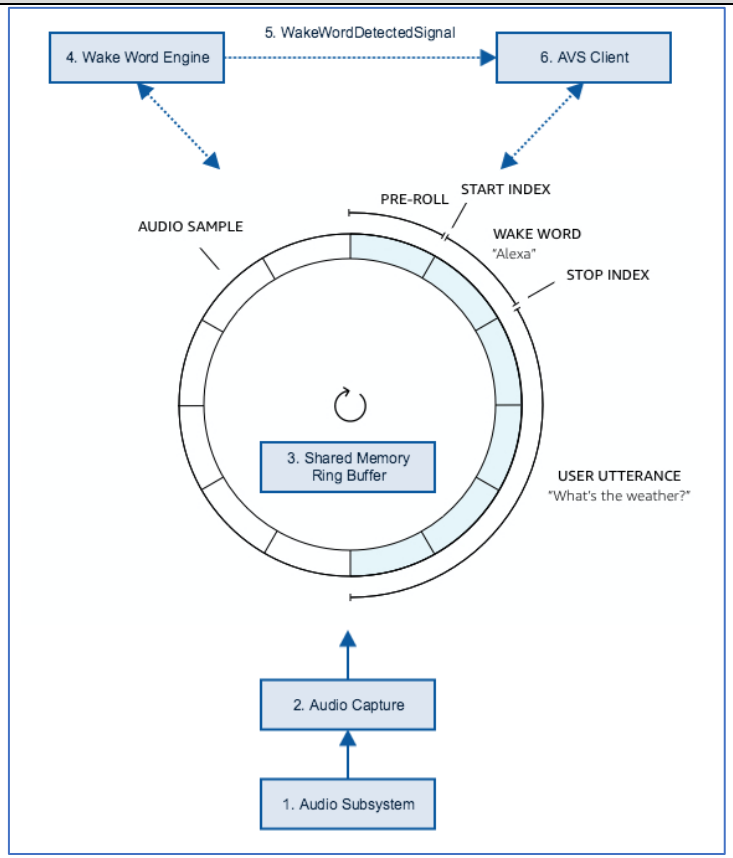
Ex.1029 / Page 217 of 347
AMAZON.COM, INC.

Claim 9	Accused Instrumentality(ies)
	Moreover, in the cloud, the separate components of Alexa's speech-processing stack — automatic speech recognition (ASR), whisper detection, and speaker identification — run on separate server nodes with their own powerful processors. On-device, those functions have to share hardware not only with each other but with Alexa's other core device functions, such as music playback. With cloud-based ASR, encrypted audio streams to the cloud in small snippets called "frames". With on-device ASR, only the lattice is sent to the cloud, where a large and powerful neural language model reranks the hypotheses. The lattice can't be sent until the customer has finished speaking, as words later in a sequence can dramatically change the overall probability of a hypothesis. With cloud-based ASR, encrypted audio streams to the cloud in small snippets called "frames". With on-device ASR, only the lattice is sent to the cloud, where a large and powerful neural language model reranks the hypotheses. The lattice can't be sent until the customer has finished speaking, as words later in a sequence can dramatically change the overall probability of a hypothesis. For example, Amazon uses wake word and speech recognition.[183]

[183] *See* KAIFI0007141 *(Wake Word Verification Requirements*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/implement-ww-verification.html).

Ex.1029 / Page 218 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	

Ex.1029 / Page 219 of 347
AMAZON.COM, INC.

Claim 9	Accused Instrumentality(ies)
	4. **Wake Word Engine** - The component that reads the audio samples written to the shared memory ring buffer and analyzes the audio samples for occurrences of the wake word. The procedure of reading and processing audio samples occurs in a loop. For example, the loop reads N samples starting at location `0`; then analyzes the samples for an occurrence of the wake word. In the next iteration, it reads and analyzes the next N samples. When an occurrence of the wake word is detected, the wake word engine must identify two locations: 1) the start of the wake word, and 2) the end of the wake word. These values are stored as `startIndexInSamples` and `endIndexInSamples`. For example, Amazon uses an ASR decoder.[184] In addition to these two classifiers, a third model component comes into play: meta-features emitted by the ASR decoder. These meta-features describe for how long (if at all) the ASR decoder is running "idle." In other words, how meaningful are the incoming audio chunks (actual content vs. silence/noise)? For example, according to Amazon, performance conditions include noise.[185] Near-field and far-field tests are conducted in various locations, which mimic real-world user-device interaction patterns. The device's performance is evaluated under various conditions such as in silence, during audio playback from the device, or in environments with background noise. Such noise could be from music playing on another device, stationary noise such as an air conditioner, or other ambient noise.

[184] *See* KAIFI0001184 (Roland Maas, *Alexa scientists address challenges of end-pointing* (Apr. 10, 2018), https://www.amazon.science/blog/alexa-scientists-address-challenges-of-end-pointing).

[185] *See* KAIFI0001353 (Felipe Pinzon, *An Amazon technical expert provides three acoustic certification-related tips for AVS device makers* (Aug. 16, 2023) https://developer.amazon.com/en-US/blogs/alexa/device-makers/2023/08/acoustic-certification-alexa-august-2023).

Ex.1029 / Page 220 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	For example, according to Amazon, ASR is trained for silence and noise.[186] To train our acoustic models, we used connectionist temporal classification (CTC), followed by multiple passes of state-level minimum-Bayes-risk (sMBR) training. To make the acoustic model more robust, we also mixed noise into the training set, enabling the model to focus on characteristics of the input signal that vary less under different acoustic conditions. Adapting to conversational speech also required modification of Alexa's end-pointer, which determines when a customer has finished speaking. The end-pointer already distinguishes between pauses at the ends of sentences, indicating that the customer has stopped speaking and that Alexa needs to follow up, and mid-sentence pauses, which may be permitted to go on a little longer. For Live Translation, we modified the end-pointer to tolerate longer pauses at the ends of sentences, as speakers engaged in long conversations will often take time between sentences to formulate their thoughts. For example, according to Amazon, ASR adds a silence label at start and end of the label sequence and in between words.[187]

[186] *See* KAIFI0002096 (Shirin Saleem, Roland Maas, *How Alexa's new Live Translation for conversations works* (Dec. 14, 2020), https://www.amazon.science/blog/how-alexas-new-live-translation-for-conversations-works).

[187] *See* KAIFI0004444 (Tulsiani, et. al, *Improved training strategies for end-to-end speech recognition in digital voice assistants*, https://assets.amazon.science/ab/b4/2f3e5aec4506930e172478586bf4/improved-training-strategies-for-end-to-end-speech-recognition-in-digital-voice-assistants.pdf).

Ex.1029 / Page 221 of 347
AMAZON.COM, INC.

Claim 9	Accused Instrumentality(ies)
	3.3. Label smoothing via probabilistic modelling of silence and use of multiple pronunciation Label smoothing was proposed as a regularization technique to improve generalization [28, 29]. It proposes to penalize low entropy output distribution by adding noise to the labels [19, 28]. In this work, we introduce a simple yet powerful technique for changing label distribution. To penalize over-confident predictions by network, while training we: (i) randomly use one of the multiple pronunciations from the lexicon (as also done in [10, 22]) (ii) randomly add silence label at start and end of the label sequence and in between words. This has the effect of changing label distribution for a fixed word sequence, thereby providing the regularization effect. For this work, we use probability of 0.8 for inserting silence label at start and end and 0.2 for inserting a silence label between words. For example, according to Amazon, ASR uses end pointers.[188] The model that determines when the customer has finished speaking is called an end-pointer. End-pointers offer a natural trade-off between accuracy and latency: an aggressive end-pointer will initiate speech processing earlier, but it might cut the speaker off prematurely, resulting in a poor customer experience.

[188] *See* KAIFI0006791 (Ariya Rastrow, Shehzad Mevawalla, *On-device speech processing makes Alexa faster, lower-bandwidth*, AMAZON SCIENCE (Jan. 25, 2022), https://www.amazon.science/blog/on-device-speech-processing-makes-alexa-faster-lower-bandwidth).

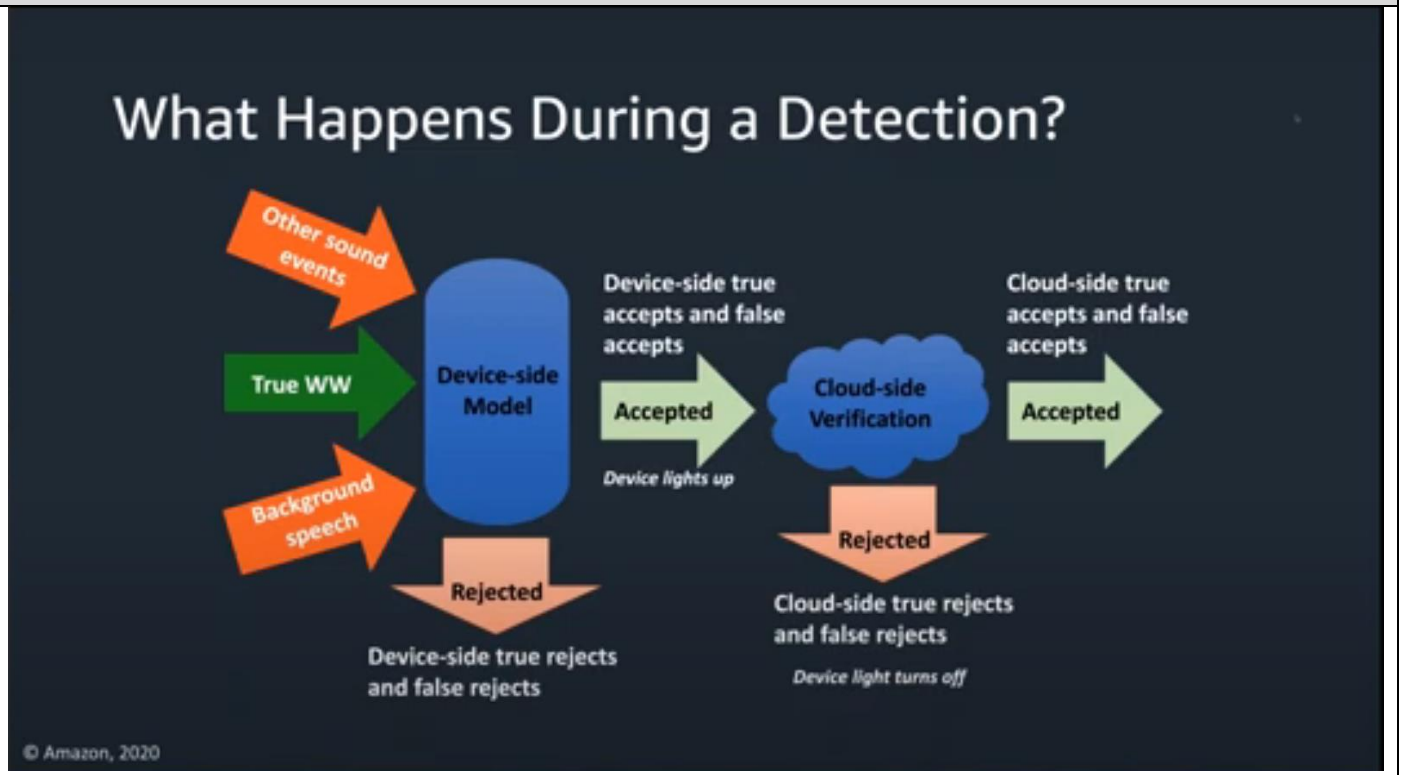
Ex.1029 / Page 222 of 347
AMAZON.COM, INC.

Claim 9	Accused Instrumentality(ies)
	In a process called "end-pointing", however, Alexa recognizes the end of an utterance by the short period of silence that follows end of speech, and that silence is part of the input to the whisper detector. When we apply the detector to live data, we typically see that its confidence increases across most of the duration of an utterance then falls off precipitously in the final 50 or so frames. For example, after receiving a wake word together with other sound events and background speech, Amazon's device-side model and cloud-side verification perform "true accepts" of the wake word.[189]

[189] *See* KAIFI0007814 *(The Science Behind Alexa's Wake Word Detection*, https://www.youtube.com/watch?v=kYrw2cC9dvs).

Ex.1029 / Page 223 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 9	Accused Instrumentality(ies)
	 This claim element is literally present, as described herein, but is also satisfied under the doctrine of equivalents because the structure and functionality provided by Amazon is being used to satisfy this claim limitation and therefore there is no vitiation of this element, and the Accused Instrumentalities incorporate structure and functionality that is not substantially different from the requirements of this limitation because they achieve substantially or exactly the same function, in substantially or exactly the same way as set forth in the claim element (as described herein), and achieves substantially or exactly the same result.

Ex.1029 / Page 224 of 347
AMAZON.COM, INC.

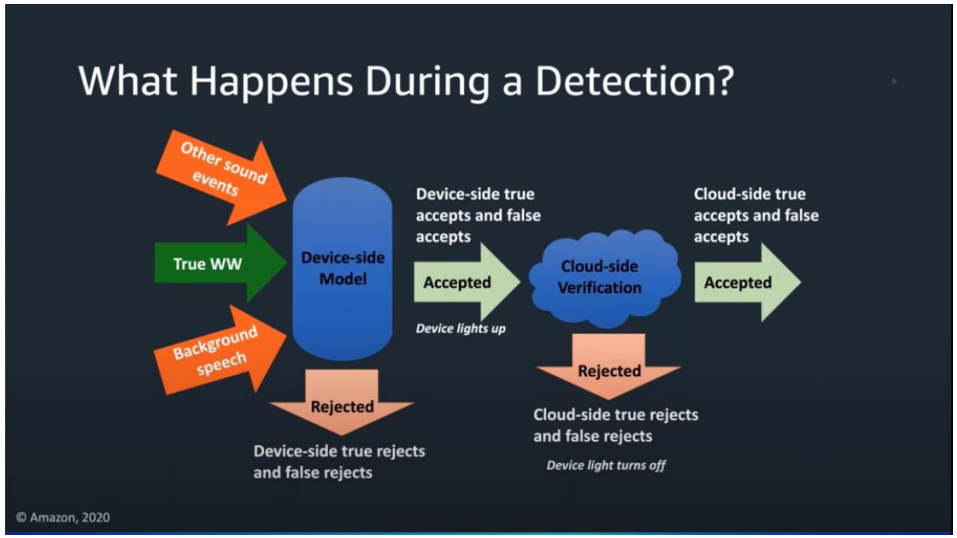
Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 10 of the '196 Patent

Claim 10	Accused Instrumentality(ies)
The apparatus of claim 9, wherein the preset call command recognition unit further comprises: an adaptive filter configured to receive the input speech and the subsequent input speech; and	The Accused Instrumentalities' preset call command recognition unit comprises an adaptive filter configured to receive the input speech and the subsequent input speech. For example, Amazon improves ASR accuracy by adapting models, such as filtering out background voices during the utterance of the wake word "Alexa."[190] > But one of the ways the Alexa AI team improves ASR accuracy is by adapting models, on the fly, to customer *context*. For instance, Alexa can use acoustic properties of the speaker's voice during the utterance of the wake word "Alexa" to filter out background voices when processing the customer's request. For example, the Accused Instrumentalities take input from one or more microphones with noise cancellation ability.[191] > The AVS Client is the main component that implements communication with the Amazon cloud using the Alexa Voice Services (AVS) APIs. It handles the audio path, taking input from one or more microphones (typically with noise cancellation ability) and playing the audio result returned by AVS via the car speakers. This component is responsible for mixing and/or prioritizing the audio with other local sources that the user might be listening to (radio, media player, etc.). The AVS client can be activated using a push-to-talk button (e.g. on the steering wheel) or by a voice command that is recognized by the Wake Word Engine.

[190] *See* KAIFI0007049 (*The Engineering Behind Alexa's Contextual Speech Recognition*, https://www.amazon.science/latest-news/the-engineering-behind-alexas-contextual-speech-recognition).
[191] *See* KAIFI0007111 *(Understanding the Amazon Alexa APIs for In-Vehicle Use: Part II*, https://abaltatech.com/blog/2018/05/understanding-the-amazon-alexa-apis-for-in-vehicle-use-part-ii/).

Ex.1029 / Page 225 of 347
AMAZON.COM, INC.

Claim 10	Accused Instrumentality(ies)
	For example, the Accused Instrumentalities comprise 500 milliseconds of pre-roll to help calibrate the ambient noise level of the recording to enhance speech recognition.[192] **500 milliseconds of pre-roll** – Pre-roll is the audio captured before AVS detects the wake word and helps calibrate the ambient noise level of the recording to enhance speech recognition. For example, the Accused Instrumentalities receive background speech and other sound events, in addition to the wake word, which may be rejected on the device-side, or rejected during cloud-side verification.[193] 

[192] *See* KAIFI0007141 *(Wake Word Verification Requirements*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/implement-ww-verification.html).
[193] *See* KAIFI0007814 *(The Science Behind Alexa's Wake Word Detection*, https://www.youtube.com/watch?v=kYrw2cC9dvs).

Ex.1029 / Page 226 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 10	Accused Instrumentality(ies)
	For example, the Accused Instrumentalities clean the audio signal during signal processing, identify ambient noise and minimize them, and employ acoustic echo cancellation so only the remaining important signal remains.[194] According to Trupti Behera, "It starts with signal processing, which gives Alexa as many chances as possible to make sense of the audio by cleaning the signal. Signal processing is one of the most important challenges in far-field audio. The idea is to improve the target signal, which means being able to identify ambient noise like the TV and minimize them. To resolve these issues, seven microphones are used to identify roughly where the signal is coming from so the device can focus on it. Acoustic echo cancellation can subtract that signal so only the remaining important signal remains. For example, the Accused Instrumentalities are designed to detect only the sound waves of a chosen wake word, and everything else is ignored.[195]

[194] *See* KAIFI0002097 (*How Amazon Alexa works? Your Guide to Natural Language Processing (AI)*, https://towardsdatascience.com/how-amazon-alexa-works-your-guide-to-natural-language-processing-ai-7506004709d3).
[195] *See* KAIFI0002092 *(How Alexa works: Wake word*, https://www.amazon.com/b?ie=UTF8&node=23608571011).

Ex.1029 / Page 227 of 347
AMAZON.COM, INC.

Claim 10	Accused Instrumentality(ies)
	Echo devices use built-in technology called "keyword spotting" that matches spoken audio to the acoustic patterns of the wake word. Simply put, Echo devices are designed by default to detect only the sound waves of your chosen wake word, and everything else is ignored. Like water through a strainer, all other audio (people talking, faucets running, birds chirping) passes through the device until the wake word is "caught" and sent to Amazon's secure cloud, where your request is reanalyzed to verify the wake word was spoken. You can choose to opt-in to features like Alexa Emergency Assist that allow your device to detect more than your chosen wake word—for example, the sound of smoke alarms—but you would need to update your settings to do so. Want to keep nerding out? Follow the journey of a voice request. For example, when the Accused Instrumentalities encounter external noise, certain techniques can prevent it from interfering with the device's speech recognition capabilities. Techniques like beamforming enhance the device's ability to capture clear speech in the presence of background noise. [196]

[196] *See* KAIFI0001353 (Felipe Pinzon, *An Amazon technical expert provides three acoustic certification-related tips for AVS device makers* (Aug. 16, 2023) https://developer.amazon.com/en-US/blogs/alexa/device-makers/2023/08/acoustic-certification-alexa-august-2023).

Ex.1029 / Page 228 of 347
AMAZON.COM, INC.

Claim 10	Accused Instrumentality(ies)
	Pay attention to noise sources: When an Alexa device encounters external noise, certain techniques can prevent it from interfering with the device's speech recognition capabilities. "I always recommend using at least two or more microphones to improve the signal-to-noise ratio so that the device can cancel the external noise and capture the user's speech," Pinzon says. He recommends using techniques like beamforming to enhance the device's ability to capture clear speech in the presence of background noise. Beamforming is a signal processing technique for multi-microphone arrays that emphasizes user speech from a desired direction when suppressing audio interference from other directions. These algorithms result in an increase in Signal-to-Noise ratio, and a reduction in reverberation in the customer's speech from the desired direction that improves the accuracy of speech recognition systems, especially for far-field. For example, when the Accused Instrumentalities have multiple, built-in microphones that help the device ignore background noise and decipher wake words spoken from a distance. [197] All Amazon Echo devices have multiple, built-in microphones that help the device ignore background noise and decipher wake words spoken from a distance. To prevent Alexa from being woken accidentally, companies can submit audio samples to Amazon to have specific instances of the wake word ignored. For example, a company making a television commercial about "Alexa" can submit the audio to Amazon.

[197] *See* KAIFI0007045 (Tech Target Contributor, *Definition wake word*, https://www.techtarget.com/whatis/definition/wake-word).

Ex.1029 / Page 229 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 10	Accused Instrumentality(ies)
a confidence detection unit configured to compute a confidence rate of the assumed input speech and recognize the assumed input speech as the preset call command based on the computed confidence rate.	The Accused Instrumentalities' preset call command recognition unit comprises a confidence detection unit configured to compute a confidence rate of the assumed input speech and recognize the assumed input speech as the preset call command based on the computed confidence rate. For example, the Accused Instrumentalities only listen to requests after the wake word is detected.[198] **How Alexa works: Wake word** Think of the wake word as a verbal cue that makes things happen. In the classic tale from *The Arabian Nights*, an invisible door into a mountain can only be opened by saying the phrase "Open Sesame." Use the phrase, the door opens. Don't use the phrase, the door stays shut. While it's not magic, because there's a lot of technology behind Alexa, the wake word does open the door for Alexa to begin fulfilling your request. When you have a request for Alexa, you first need to say your chosen wake word, which by default is "Alexa." Only after your Echo device detects the wake word is Alexa listening to your requests. For example, Amazon uses built-in technology called "keyword spotting" that matches spoken audio to the acoustic patterns of the wake word and are designed to detect only the sound waves of the chosen wake word, which is sent to Amazon's secure cloud, where the request is reanalyzed to verify the wake word was spoken.[199]

[198] *See* KAIFI0002092 *(How Alexa works: Wake word*, https://www.amazon.com/b?ie=UTF8&node=23608571011).
[199] *See id.*

Ex.1029 / Page 230 of 347
AMAZON.COM, INC.

Claim 10	Accused Instrumentality(ies)
	But how can my device detect the wake word without hearing the rest of my conversations? Echo devices use built-in technology called "keyword spotting" that matches spoken audio to the acoustic patterns of the wake word. Simply put, Echo devices are designed by default to detect only the sound waves of your chosen wake word, and everything else is ignored. Like water through a strainer, all other audio (people talking, faucets running, birds chirping) passes through the device until the wake word is "caught" and sent to Amazon's secure cloud, where your request is reanalyzed to verify the wake word was spoken. You can choose to opt-in to features like Alexa Emergency Assist that allow your device to detect more than your chosen wake word—for example, the sound of smoke alarms—but you would need to update your settings to do so. Want to keep nerding out? Follow the journey of a voice request. For example, Amazon achieves far-field speech recognition in noisy environments with high accuracy through a combination of machine learning algorithms, data, and immense computing power.[200] By launching Echo in November 2014, we showed that far-field speech recognition in noisy environments can be done with high accuracy through a combination of machine learning algorithms, data, and immense computing power.

[200] *See* KAIFI0004342 *(How Our Scientists are Making Alexa Smarter* (Mar. 29, 2018), https://www.aboutamazon.com/news/devices/how-our-scientists-are-making-alexa-smarter).

Ex.1029 / Page 231 of 347
AMAZON.COM, INC.

Claim 10	Accused Instrumentality(ies)
	For example, Amazon's wake word detection employs deep learning technology running on-device to recognize the wake word the user has chosen.[201] It starts with detecting the "wake word" that triggers Alexa to start listening to the subsequent spoken words from the user. Wake word detection employs deep learning technology running on-device to recognize the wake word the user has chosen. Far-field automatic speech recognition (ASR) in the Amazon Web Services (AWS) cloud then converts into text the audio following the wake word, and determines when the user has stopped talking to Alexa. For example, Amazon uses a suite of learning techniques, including supervised, semi-supervised, and unsupervised learning.[202]

[201] *See* KAIFI0004342 *(How Our Scientists are Making Alexa Smarter* (Mar. 29, 2018), https://www.aboutamazon.com/news/devices/how-our-scientists-are-making-alexa-smarter).
[202] *See id.*

Ex.1029 / Page 232 of 347
AMAZON.COM, INC.

Claim 10	Accused Instrumentality(ies)
	A. Since Alexa's brain is mostly in the cloud, she gets smarter with every interaction. Alexa uses a suite of learning techniques: supervised, semi-supervised, and unsupervised learning. While supervised learning is still most powerful, it does not scale since we cannot generate manual labels at the pace required to continually improve Alexa for our customers. Therefore, our scientists and engineers are continually applying and inventing new learning techniques to reduce reliance on manual labels for training our statistical models. For instance, active learning as a class of semi-supervised learning techniques, where the system itself posits which part of the interactions it needs input from a human expert, is pervasive across our technologies. Unsupervised learning without any labeled responses is also being applied to make Alexa smarter, especially for speech recognition. Finally, we also use transfer learning, allowing Alexa to learn from one skill to another or even one language to another. For example, Amazon uses machine learning model trained from many occurrences of detected wake words to score a match with the input voice data.[203] **Wake Word Model** A WWD uses a machine learning model trained from many occurrences of detected wake words to score a match with the input voice data. If the confidence score exceeds the model's threshold, the wake word is considered detected and the WWD triggers. The wake word model may incorporate multiple wake words and also be trained to support multiple languages.

[203] *See* KAIFI0007120 (*Voice Interoperability Initiative Architecture Series Whitepapers*, https://m.media-amazon.com/images/G/01/vii/VII_Architecture_Series_Whitepapers_-_Multi-Agents_Wake_Word.pdf).

Ex.1029 / Page 233 of 347
AMAZON.COM, INC.

Claim 10	Accused Instrumentality(ies)
	For example, Amazon's ASR system outputs a confidence score to verify the presence of the wake word by comparing to a fixed threshold.[204]

2.2. Wakeword Verification

Unlike detection, which operates in a streaming manner, wake-word verification performs a single pass on the input localized context frame. To validate our proposed CRA architecture, we consider several baseline systems that include both whole-word and phonetic based models.

2.2.1. Baselines

Phonetically modeled systems that we consider are ASR [21] and 2-stage DNN-HMM [11]. The ASR system we consider is a large vocabulary DNN-HMM system, with an n-gram language model to recognize the spoken words in the audio stream. The system outputs a confidence score for each recognized word in

[204] *See* KAIFI0001937 (Rajath Kumar, et. al, *Building a Robust Word-Level Wakeword Verification Network,* https://assets.amazon.science/ba/c5/bd48ce11445ba0368d2d9191600d/building-a-robust-word-level-wakeword-verification-network.pdf).

Ex.1029 / Page 234 of 347
AMAZON.COM, INC.

Claim 10	Accused Instrumentality(ies)
	the range (0,1). Similarly, for the defined wakeword, the ASR model outputs a confidence score, which we use to verify the presence of the wakeword by comparing to a fixed threshold. Since every stream is initiated with the wakeword, we set the confidence score for the wakeword to be 0 when the wakeword is not detected by ASR. Note that this ASR model is not specifically trained for the wakeword verification task, i.e., is not biased towards the wakeword. In the 2-stage DNN-HMM system, the first stage is similar to the acoustic part of the ASR system. However, unlike ASR, the 2-stage system is modeled only on the phonemes representing the wakeword and 2 additional outputs to distinguish background and silence [22]. A fixed length feature vector is extracted from the first stage DNN-HMM. This vector is engineered to aggregate information over all the frames in the hypothesized wakeword segment of the stream [22]. We pass the extracted vector as input to the second stage, which consist of multiple feed-forward layers. The second stage is a binary classification system trained to verify the presence of the wakeword. The two stages are trained separately, and the resulting score from the second stage is used for verification. This claim element is literally present, as described herein, but is also satisfied under the doctrine of equivalents because the structure and functionality provided by Amazon is being used to satisfy this claim limitation and therefore there is no vitiation of this element, and the Accused Instrumentalities incorporate structure and functionality that is not substantially different from the requirements of this limitation because they achieve substantially or exactly the same function, in substantially or exactly the same way as set forth in the claim element (as described herein), and achieves substantially or exactly the same result.

Ex.1029 / Page 235 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 11 of the '196 Patent

Claim 11	Accused Instrumentality(ies)
The apparatus of claim 10, wherein the confidence detection unit is further configured to recognize the assumed input speech as the preset call command if the computed confidence rate is higher than a threshold value.	The Accused Instrumentalities' confidence detection unit is further configured to recognize the assumed input speech as the preset call command if the computed confidence rate is higher than a threshold value. For example, Amazon uses a machine learning model trained from many occurrences of detected wake words to score a match with the input voice data.[205] **Wake Word Model** A WWD uses a machine learning model trained from many occurrences of detected wake words to score a match with the input voice data. If the confidence score exceeds the model's threshold, the wake word is considered detected and the WWD triggers. The wake word model may incorporate multiple wake words and also be trained to support multiple languages. For example, Amazon uses an ASR system that outputs a confidence score to verify the presence of the wake word by comparing to a fixed threshold.[206]

[205] *See* KAIFI0007120 (*Voice Interoperability Initiative Architecture Series Whitepapers*, https://m.media-amazon.com/images/G/01/vii/VII_Architecture_Series_Whitepapers_-_Multi-Agents_Wake_Word.pdf).
[206] *See* KAIFI0001937 (Rajath Kumar, et. al, *Building a Robust Word-Level Wakeword Verification Network,* https://assets.amazon.science/ba/c5/bd48ce11445ba0368d2d9191600d/building-a-robust-word-level-wakeword-verification-network.pdf).

Ex.1029 / Page 236 of 347
AMAZON.COM, INC.

Claim 11	Accused Instrumentality(ies)
	2.2. Wakeword Verification Unlike detection, which operates in a streaming manner, wake-word verification performs a single pass on the input localized context frame. To validate our proposed CRA architecture, we consider several baseline systems that include both whole-word and phonetic based models. *2.2.1. Baselines* Phonetically modeled systems that we consider are ASR [21] and 2-stage DNN-HMM [11]. The ASR system we consider is a large vocabulary DNN-HMM system, with an n-gram language model to recognize the spoken words in the audio stream. The system outputs a confidence score for each recognized word in

Ex.1029 / Page 237 of 347
AMAZON.COM, INC.

Claim 11	Accused Instrumentality(ies)
	the range (0,1). Similarly, for the defined wakeword, the ASR model outputs a confidence score, which we use to verify the presence of the wakeword by comparing to a fixed threshold. Since every stream is initiated with the wakeword, we set the confidence score for the wakeword to be 0 when the wakeword is not detected by ASR. Note that this ASR model is not specifically trained for the wakeword verification task, i.e., is not biased towards the wakeword. In the 2-stage DNN-HMM system, the first stage is similar to the acoustic part of the ASR system. However, unlike ASR, the 2-stage system is modeled only on the phonemes representing the wakeword and 2 additional outputs to distinguish background and silence [22]. A fixed length feature vector is extracted from the first stage DNN-HMM. This vector is engineered to aggregate information over all the frames in the hypothesized wakeword segment of the stream [22]. We pass the extracted vector as input to the second stage, which consist of multiple feed-forward layers. The second stage is a binary classification system trained to verify the presence of the wakeword. The two stages are trained separately, and the resulting score from the second stage is used for verification. For example, Amazon employs a general far-field ASR model to automatically transcribe the unlabeled speech corpus. Since ASR transcript may not be accurate, only the utterances with WW confidence score higher than a threshold θ_P are selected.[207]

[207] *See* KAIFI0007074 (Yixin Gao, et. al, *TOWARDS DATA-EFFICIENT MODELING FOR WAKE WORD SPOTTING*, https://assets.amazon.science/7c/b2/5e3e6a164920bfc167fb5586d3f2/scipub-1260.pdf).

Ex.1029 / Page 238 of 347
AMAZON.COM, INC.

Claim 11	Accused Instrumentality(ies)
	In this paper we develop a simple but efficient semi-supervised learning strategy to mine useful WW-training data from large un-transcribed general-speech corpora. We employ a general far-field ASR model to automatically transcribe the unlabeled speech corpus, then select WW-specific utterances and add them to our training set. The DNN targets are obtained by force aligning the speech to the automatic transcripts. Since the ASR transcript may not be accurate, only the utterances with WW confidence score higher than a thresh-old θ_P are selected. It should be noted that, in the utterances mined here the WW may appear in the middle of the utterance, therefore the phonetic context may be different from the utterances initiated by the WW. In practice, our experiments show that such context mismatch have unobservable effects on WW spotting. This claim element is literally present, as described herein, but is also satisfied under the doctrine of equivalents because the structure and functionality provided by Amazon is being used to satisfy this claim limitation and therefore there is no vitiation of this element, and the Accused Instrumentalities incorporate structure and functionality that is not substantially different from the requirements of this limitation because they achieve substantially or exactly the same function, in substantially or exactly the same way as set forth in the claim element (as described herein), and achieves substantially or exactly the same result.

Ex.1029 / Page 239 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 12 of the '196 Patent

Claim 12	Accused Instrumentality(ies)
The apparatus of claim 10, wherein the confidence detection unit is further configured to determine whether or not the assumed input speech is an authentic call command based on the computed confidence rate and recognize the assumed input speech as the preset call command if it is determined that the assumed input speech is the authentic call command.	The Accused Instrumentalities' confidence detection unit is further configured to determine whether or not the assumed input speech is an authentic call command based on the computed confidence rate and recognize the assumed input speech as the preset call command if it is determined that the assumed input speech is the authentic call command. For example, Amazon uses a machine learning model trained from many occurrences of detected wake words to score a match with the input voice data.[208] **Wake Word Model** A WWD uses a machine learning model trained from many occurrences of detected wake words to score a match with the input voice data. If the confidence score exceeds the model's threshold, the wake word is considered detected and the WWD triggers. The wake word model may incorporate multiple wake words and also be trained to support multiple languages. For example, Amazon uses an ASR system that outputs a confidence score to verify the presence of the wake word by comparing to a fixed threshold.[209]

[208] *See* KAIFI0007120 (*Voice Interoperability Initiative Architecture Series Whitepapers*, https://m.media-amazon.com/images/G/01/vii/VII_Architecture_Series_Whitepapers_-_Multi-Agents_Wake_Word.pdf).
[209] *See* KAIFI0001937 (Rajath Kumar, et. al, *Building a Robust Word-Level Wakeword Verification Network*, https://assets.amazon.science/ba/c5/bd48ce11445ba0368d2d9191600d/building-a-robust-word-level-wakeword-verification-network.pdf).

Ex.1029 / Page 240 of 347
AMAZON.COM, INC.

Claim 12	Accused Instrumentality(ies)
	2.2. Wakeword Verification Unlike detection, which operates in a streaming manner, wake-word verification performs a single pass on the input localized context frame. To validate our proposed CRA architecture, we consider several baseline systems that include both whole-word and phonetic based models. *2.2.1. Baselines* Phonetically modeled systems that we consider are ASR [21] and 2-stage DNN-HMM [11]. The ASR system we consider is a large vocabulary DNN-HMM system, with an n-gram language model to recognize the spoken words in the audio stream. The system outputs a confidence score for each recognized word in

Ex.1029 / Page 241 of 347
AMAZON.COM, INC.

Claim 12	Accused Instrumentality(ies)
	the range (0,1). Similarly, for the defined wakeword, the ASR model outputs a confidence score, which we use to verify the presence of the wakeword by comparing to a fixed threshold. Since every stream is initiated with the wakeword, we set the confidence score for the wakeword to be 0 when the wakeword is not detected by ASR. Note that this ASR model is not specifically trained for the wakeword verification task, i.e., is not biased towards the wakeword. In the 2-stage DNN-HMM system, the first stage is similar to the acoustic part of the ASR system. However, unlike ASR, the 2-stage system is modeled only on the phonemes representing the wakeword and 2 additional outputs to distinguish background and silence [22]. A fixed length feature vector is extracted from the first stage DNN-HMM. This vector is engineered to aggregate information over all the frames in the hypothesized wakeword segment of the stream [22]. We pass the extracted vector as input to the second stage, which consist of multiple feed-forward layers. The second stage is a binary classification system trained to verify the presence of the wakeword. The two stages are trained separately, and the resulting score from the second stage is used for verification. For example, Amazon employs a general far-field ASR model to automatically transcribe the unlabeled speech corpus. Since ASR transcript may not be accurate, only the utterances with WW confidence score higher than a threshold θ_P is selected.[210]

[210] *See* KAIFI0007074 (Yixin Gao, et. al, *TOWARDS DATA-EFFICIENT MODELING FOR WAKE WORD SPOTTING*, https://assets.amazon.science/7c/b2/5e3e6a164920bfc167fb5586d3f2/scipub-1260.pdf).

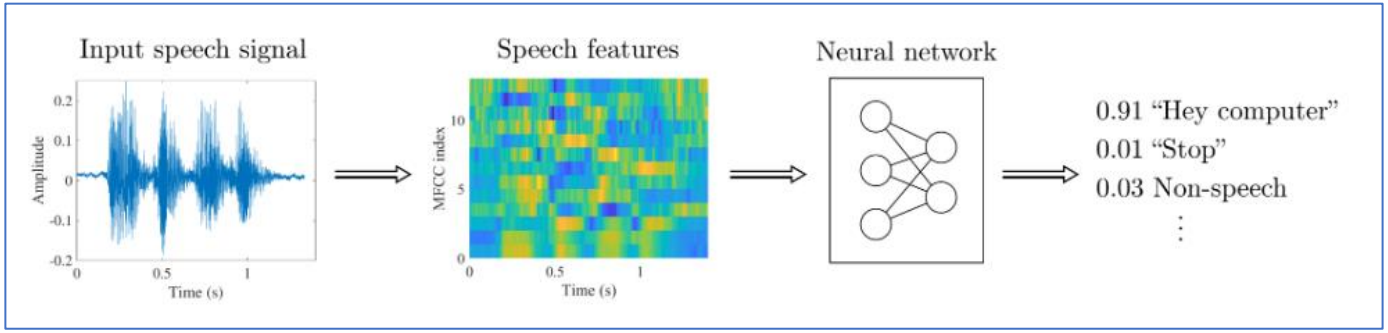
Ex.1029 / Page 242 of 347
AMAZON.COM, INC.

Claim 12	Accused Instrumentality(ies)
	In this paper we develop a simple but efficient semi-supervised learning strategy to mine useful WW-training data from large un-transcribed general-speech corpora. We employ a general far-field ASR model to automatically transcribe the unlabeled speech corpus, then select WW-specific utterances and add them to our training set. The DNN targets are obtained by force aligning the speech to the automatic transcripts. Since the ASR transcript may not be accurate, only the utterances with WW confidence score higher than a thresh-old θ_P are selected. It should be noted that, in the utterances mined here the WW may appear in the middle of the utterance, therefore the phonetic context may be different from the utterances initiated by the WW. In practice, our experiments show that such context mismatch have unobservable effects on WW spotting. This claim element is literally present, as described herein, but is also satisfied under the doctrine of equivalents because the structure and functionality provided by Amazon is being used to satisfy this claim limitation and therefore there is no vitiation of this element, and the Accused Instrumentalities incorporate structure and functionality that is not substantially different from the requirements of this limitation because they achieve substantially or exactly the same function, in substantially or exactly the same way as set forth in the claim element (as described herein), and achieves substantially or exactly the same result.

Ex.1029 / Page 243 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 13 of the '196 Patent

Claim 13	Accused Instrumentality(ies)
The apparatus of claim 10, wherein the token passing unit is further configured to compute a log likelihood ratio of the input speech and assume the input speech to be the preset call command based on the computed log likelihood ratio.	The Accused Instrumentalities' token passing unit is further configured to compute a log likelihood ratio of the input speech and assume the input speech to be the preset call command based on the computed log likelihood ratio. For example, Amazon uses keyword-spotting algorithms as illustrated below. The input speech signal is first converted to a feature representation, such as MFCCs, which are fed to a neural network, and the output is the likelihood of each keyword.[211]  This claim element is literally present, as described herein, but is also satisfied under the doctrine of equivalents because the structure and functionality provided by Amazon is being used to satisfy this claim limitation and therefore there is no vitiation of this element, and the Accused Instrumentalities incorporate structure and functionality that is not substantially different from the requirements of this limitation

[211] *See* KAIFI0001092 (*Introduction to Speech Processing*, https://speechprocessingbook.aalto.fi/Recognition/Wake-word_and_keyword_spotting.html).

Ex.1029 / Page 244 of 347
AMAZON.COM, INC.

Claim 13	Accused Instrumentality(ies)
	because they achieve substantially or exactly the same function, in substantially or exactly the same way as set forth in the claim element (as described herein), and achieves substantially or exactly the same result.

Claim 14 of the '196 Patent

Claim 14	Accused Instrumentality(ies)
The apparatus of claim 13, wherein the computed log likelihood ratio is a natural logarithm of a probability on the input speech for a phoneme model of the preset call command, and the computed confidence rate is another log likelihood ratio.	The Accused Instrumentalities employ Natural Language Understanding, which involves, for example, finding phonemes and statistically determining the most likely words and sentences.[212] Think the Alexa Voice Service as the brain of Alexa enabled devices and perform all the complex operations such as Automatic Speech Recognition (ASR) and Natural Language Understanding (NLU). For Trevor Jackins, Marketing Specialist at NeoSpeech, "It tries to understand what you said by taking the voice data and breaking it down to a small sample of particular time duration mostly 10–20 ms. These data sets are further compared to pre-fed speech to decode what you said in each unit of your speech. The purpose here is to find **phoneme (the smallest unit of speech).** Then, the machine looks at the series of such phonemes and statistically determine the most likely words and sentences to spoke."

[212] *See* KAIFI0002097 (How Amazon Alexa works? Your Guide to Natural Language Processing (AI), https://towardsdatascience.com/how-amazon-alexa-works-your-guide-to-natural-language-processing-ai-7506004709d3).

Ex.1029 / Page 245 of 347
AMAZON.COM, INC.

Claim 14	Accused Instrumentality(ies)
	For example, Amazon employs ASR systems that include both an acoustic model and a language model. The acoustic model converts audio into phonemes, the smallest units of speech; the language model encodes the probabilities of particular strings of words, which helps the ASR system decide between alternative interpretations of the same sequence of phonemes.[213] Like most ASR systems, the ones we use for live translation include both an acoustic model and a language model. The acoustic model converts audio into phonemes, the smallest units of speech; the language model encodes the probabilities of particular strings of words, which helps the ASR system decide between alternative interpretations of the same sequence of phonemes. This claim element is literally present, as described herein, but is also satisfied under the doctrine of equivalents because the structure and functionality provided by Amazon is being used to satisfy this claim limitation and therefore there is no vitiation of this element, and the Accused Instrumentalities incorporate structure and functionality that is not substantially different from the requirements of this limitation because they achieve substantially or exactly the same function, in substantially or exactly the same way as set forth in the claim element (as described herein), and achieves substantially or exactly the same result.

Claim 15 of the ’196 Patent

[213] *See* KAIFI0002096 (Shirin Saleem, Roland Maas, *How Alexa's new Live Translation for conversations works* (Dec. 14, 2020), https://www.amazon.science/blog/how-alexas-new-live-translation-for-conversations-works).

Ex.1029 / Page 246 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 15	Accused Instrumentality(ies)
The apparatus of claim 14, wherein said another log likelihood ratio is computed by subtracting, from the log likelihood ratio, a natural logarithm of a probability on the input speech for an anti-phoneme model of the preset call command.	The Accused Instrumentalities' another log likelihood ratio is computed by subtracting, from the log likelihood ratio, a natural logarithm of a probability on the input speech for an anti-phoneme model of the preset call command. For example, Amazon uses Hidden Markov Models (HMMs). HMMs are statistical models widely used in ASR. They represent the probabilistic relationship between hidden states (phonemes or words) and observed acoustic features. HMM-based ASR systems align audio features with phonetic units to generate transcriptions.[214] Techniques used in ASR include: • **Hidden Markov Models (HMMs):** HMMs are statistical models widely used in ASR. They represent the probabilistic relationship between hidden states (phonemes or words) and observed acoustic features. HMM-based ASR systems align audio features with phonetic units to generate transcriptions. This claim element is literally present, as described herein, but is also satisfied under the doctrine of equivalents because the structure and functionality provided by Amazon is being used to satisfy this claim limitation and therefore there is no vitiation of this element, and the Accused Instrumentalities incorporate structure and functionality that is not substantially different from the requirements of this limitation because they achieve substantially or exactly the same function, in substantially or exactly the same way as set forth in the claim element (as described herein), and achieves substantially or exactly the same result.

[214] *See* KAIFI0007060 *(*Skill Upwards, *Top 5 AI technologies behind voice assistants like Siri and Alexa,* https://skillupwards.com/blog/ai-technologies-behind-voice-assistants-like-siri-and-alexa).

Ex.1029 / Page 247 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 16 of the '196 Patent

Claim 16	Accused Instrumentality(ies)
The apparatus of claim 9, wherein the continuous speech recognition unit comprises: a keyword spotting unit configured to recognize the subsequent input speech as the actual command if the input speech is recognized as the preset call command; and	The Accused Instrumentalities' continuous speech recognition unit comprises: a keyword spotting unit configured to recognize the subsequent input speech as the actual command if the input speech is recognized as the preset call command. For example, Amazon uses on-device keyword spotting designed to detect when a customer says the wake word. [215] **Echo Devices** Echo devices are the input and output devices for Alexa, and we designed and built them, from the beginning, with multiple layers of security and privacy protections and controls. Echo devices use on-device keyword spotting designed to detect when a customer says the wake word. This technology inspects acoustic patterns in the room to detect when the wake word has been spoken using a short, on-device buffer that is continuously overwritten. This on-device buffer exists in temporary memory (RAM); audio is not recorded to any on-device storage. The device does not stream audio to the cloud until the wake word is detected or the action button on the device is pressed. If it does not detect the wake word or if the action button is not pressed, no audio is sent to the cloud. The customer can turn the microphone off at any time by pressing the microphone button on the device. When the button is pressed to turn the microphones off, the microphones are electrically disconnected and a dedicated red LED on the microphone button is illuminated to indicate the microphones are off. As a safeguard, we designed the circuitry of Echo devices so that power can only be provided either to this dedicated red LED or to the device microphones, not to both at the same time. As a result, if the dedicated red LED is illuminated, the microphones are off and cannot stream audio to the cloud.

[215] *See* KAIFI0007380 *(White paper – Alexa Privacy and Data Handling Overview*, https://d1.awsstatic.com/product-marketing/A4B/White%20Paper%20-%20Alexa%20Privacy%20and%20Data%20Handling%20Overview.pdf).

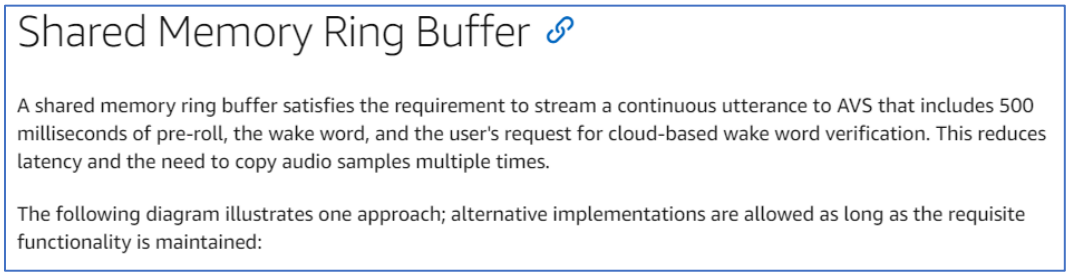
Ex.1029 / Page 248 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 16	Accused Instrumentality(ies)
	For example, Amazon uses automatic speech recognition and natural language understanding.[216] It starts with detecting the "wake word" that triggers Alexa to start listening to the subsequent spoken words from the user. Wake word detection employs deep learning technology running on-device to recognize the wake word the user has chosen. Far-field automatic speech recognition (ASR) in the Amazon Web Services (AWS) cloud then converts into text the audio following the wake word, and determines when the user has stopped talking to Alexa. Once the spoken audio has been converted to text, Alexa uses natural language understanding (NLU) to convert the words into a structured interpretation of intent that can be used to respond to the user from the more than 30,000 Alexa skills built by first- and third-party developers. This structured interpretation is used in combination with different forms of context, such as which type of device the user is interacting with, what the most likely skills are that can provide a response, or who is speaking. This context helps determine the best next action Alexa should take. The possible outcomes are to either respond with the best action from a skill or to ask for more information from the user. How Alexa responds or sounds is also critical for natural dialog. This is accomplished via text-to-speech synthesis (TTS) which converts arbitrary sequences of words into natural sounding, intelligible audio. What's common across all the technologies mentioned above is the emphasis on data-driven machine learning and fast inference at runtime to deliver an accurate response in as short a time as possible. As scientists and engineers we're always battling this healthy tension between accuracy and latency from when the user stops speaking to Alexa to when she responds.

[216] *See* KAIFI0004342 *(How Our Scientists are Making Alexa Smarter* (Mar. 29, 2018), https://www.aboutamazon.com/news/devices/how-our-scientists-are-making-alexa-smarter).

Ex.1029 / Page 249 of 347
AMAZON.COM, INC.

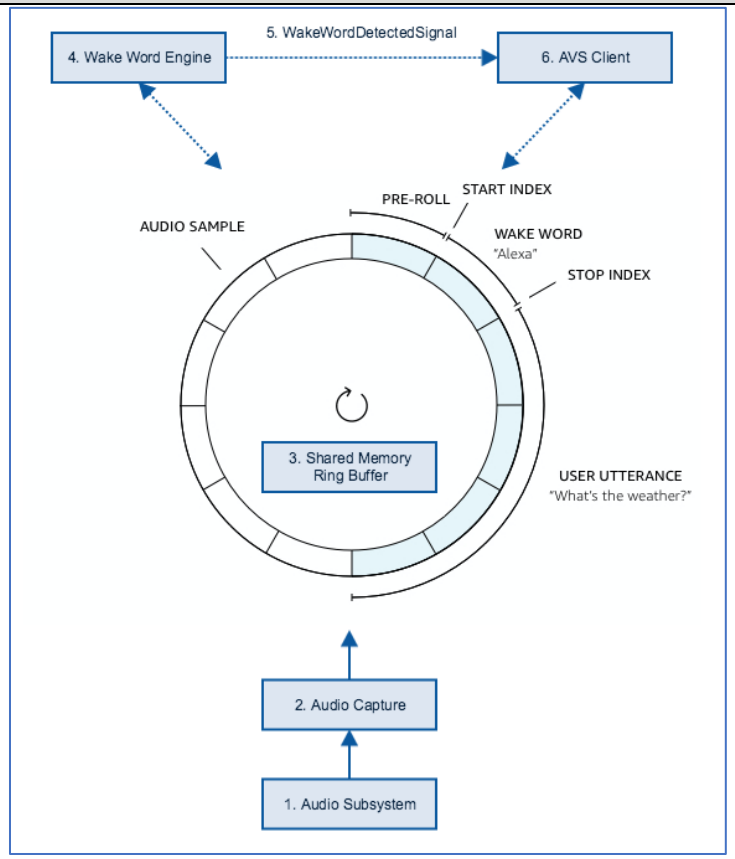
Claim 16	Accused Instrumentality(ies)
	For example, Amazon uses a shared memory ring buffer.[217]  Shared Memory Ring Buffer A shared memory ring buffer satisfies the requirement to stream a continuous utterance to AVS that includes 500 milliseconds of pre-roll, the wake word, and the user's request for cloud-based wake word verification. This reduces latency and the need to copy audio samples multiple times. The following diagram illustrates one approach; alternative implementations are allowed as long as the requisite functionality is maintained: For example, Amazon's wake word engine continuously recognizes speech.[218] Voice-initiated devices begin to stream user speech to AVS when the wake word engine detects a spoken wake word, such as "Alexa."

[217] *See* KAIFI0007141 *(Wake Word Verification Requirements*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/implement-ww-verification.html).

[218] *See Id.*

Ex.1029 / Page 250 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 16	Accused Instrumentality(ies)
	

Ex.1029 / Page 251 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 16	Accused Instrumentality(ies)
	4. **Wake Word Engine** - The component that reads the audio samples written to the shared memory ring buffer and analyzes the audio samples for occurrences of the wake word. The procedure of reading and processing audio samples occurs in a loop. For example, the loop reads N samples starting at location `0`; then analyzes the samples for an occurrence of the wake word. In the next iteration, it reads and analyzes the next N samples. When an occurrence of the wake word is detected, the wake word engine must identify two locations: 1) the start of the wake word, and 2) the end of the wake word. These values are stored as `startIndexInSamples` and `endIndexInSamples`. For example, Amazon's Wake Word Engine comprises a processor for speech recognition.[219] Each solution has varying degrees of optimization as illustrated in the Level of Optimization (LoO) column. It also shows an example of a very optimized solution on a considerably less powerful ARM Cortex-R4 microcontroller for illustration purposes, but where the Alexa wake word engine runs exclusively on the digital signal processor front-end of the device to offload the host processor. Table 2 and Figure 1 below illustrate how effectively the various solutions utilize the host processor and memory.

[219] *See* KAIFI0006993 (Bob Mars, *Sizing Up CPU, Memory, and Storage for Your Alexa Built-in Device* (Sept. 13, 2018), https://developer.amazon.com/en-US/blogs/alexa/post/2a32d792-d471-4136-8262-79962a2b4d72/cpu-memory-and-storage-for-alexa-built-in-device.html).

Ex.1029 / Page 252 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 16	Accused Instrumentality(ies)

Arch	Host Processor(s)	Vendor	LoO 1-5	Gain	eCycles 1GHz Alexa	eCycles 1GHz Music
MIPS32	MIPS 24KEc	SI_1	4	2	31	24
MIPS32	MIPS 24KEc	SI_1	4	2	24	24
MIPS32	MIPS 24KEc	SI_2	4	2	266	183
MIPS32	Xburst	SI_2	4	2	172	128
MIPS32	Xburst	SI_3	3	2	90	90
ARMv7-A	Cortex-A7	V_2	2	1	50	46
ARMv7-A	Cortex-A7+Neon+M4	Dev Kit	1	1	78	75
ARMv7-A	Cortex-A7+Neon/VFPU	SI_1	4	1.5	106	104
ARMv8-A	Cortex A53	SI_3	3	1.5	70	65
ARMv8-A	Cortex A53	Dev Kit	1	1	52	48

eCycle: effective Cycles; MCPS normalized to 1 GHz with Gain multiplier
Gain: Cycle gain multiplier due to additional memory; 1.5=50%; 2=100%

Table 2- Normalized Headroom on Alexa Built-in Devices

For example, Amazon uses a speech recognizer.[220]

> The `SpeechRecognizer` interface is the core interface of the Alexa Voice Service (AVS) and exposes directives and events for capturing and interacting with user speech. This page discusses the concepts and process flows for the `SpeechRecognizer` interface.

[220] *See* KAIFI0007093 *(Understand the AVS SpeechRecognizer*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/avs-speechrecognizer-concepts.html).

Ex.1029 / Page 253 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 16	Accused Instrumentality(ies)

SpeechRecognizer functionality

Every user utterance leverages `SpeechRecognizer`, including the following interaction types between an Alexa Built-in device and AVS:

- **User Speech Capture**: Captures user speech from an Alexa Built-in device.
- **User Speech Prompting**: Prompts a user for more speech input when needed by Alexa to deliver an appropriate response.
- **Interaction Initiation Communication**: Enables a device to inform AVS of how a user initiated an Alexa interaction, such as press-and-hold, tap-and-release, or voice-initiated/wake word enabled. See Device Form Factor and Alexa Interaction.
- **ASR Profile Selection**: SpeechRecognizers chooses the appropriate Automatic Speech Recognition (ASR) profile for your product, which allows Alexa to understand user speech and respond with precision. See Automatic Speech Recognition (ASR) profile

State diagram

The following diagram illustrates state changes driven by `SpeechRecognizer` components. Boxes represent `SpeechRecognizer` states, and the connectors represent state transitions.

`SpeechRecognizer` has the following states:

- **IDLE:** When not actively processing speech, the `SpeechRecognizer` is in an "idle" state. The idle state occurs under the following conditions:
  - Before capturing user speech.
  - Returning to an idle state after concluding a speech interaction with Alexa.
  - When an `ExpectSpeechTimedOut` event elapses.

> **Note:** In a multi-turn Alexa interaction, if Alexa requires more user speech input, `SpeechRecognizer` should transition from the idle state to the `ExpectSpeech` state without the user starting a new interaction.

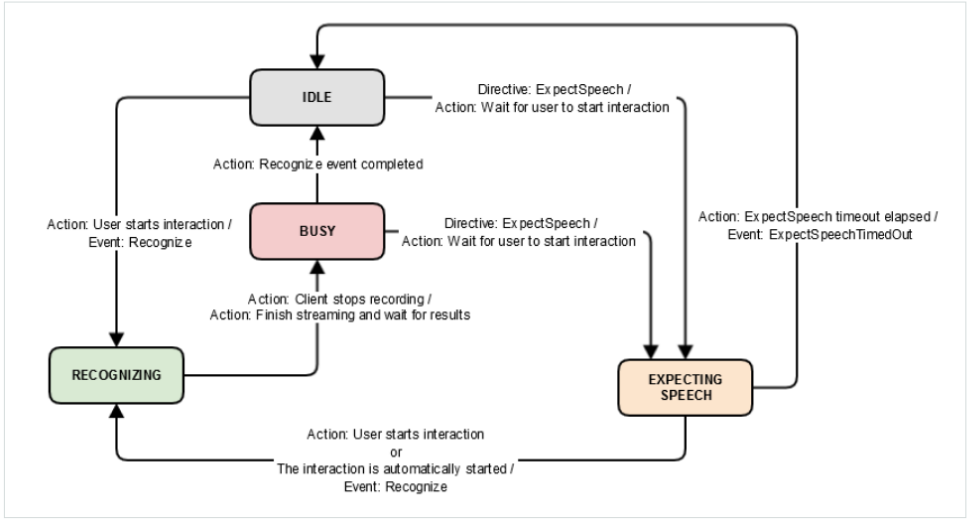
Ex.1029 / Page 254 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 16	Accused Instrumentality(ies)

- **RECOGNIZING:** When a user begins interacting with your client, specifically when the client streams captured audio to AVS, `SpeechRecognizer` should transition from the idle state to the `Recognize` state. `SpeechRecognizer` should remain in the `Recognize` until the client stops recording speech or finishes streaming, at which point your `SpeechRecognizer` component should transition from the `Recognize` state to the "busy" state.
- **BUSY:** While processing the speech request, `SpeechRecognizer` should be in the "busy" state. You cannot start another speech request until `SpeechRecognizer` transitions out of the busy state. From the *busy* state, `SpeechRecognizer` transitions to the idle state if Alexa processes and completes the request or to the `ExpectSpeech` state, if Alexa requires more speech input from the user.
- **EXPECTING SPEECH:** `SpeechRecognizer` should be in the `ExpectSpeech` state when Alexa requires more audio input from a user. From `ExpectSpeech`, `SpeechRecognizer` should transition to the `Recognize` when a user interaction occurs, or the interaction is automatically started on behalf of the user. It should transition to the idle state if Alexa detects no user interaction within the specified timeout window.

The following diagram illustrates the expected transitions among the four `SpeechRecognizer` states:



Ex.1029 / Page 255 of 347
AMAZON.COM, INC.

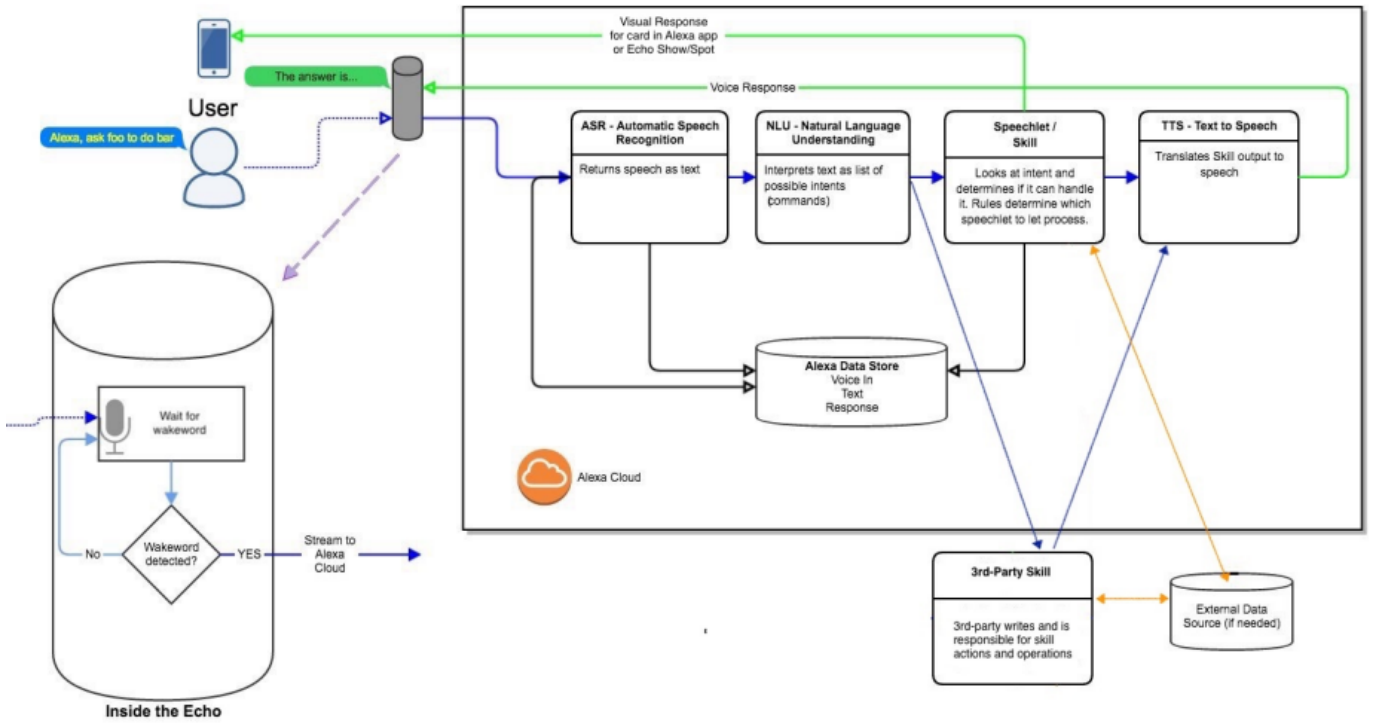
Claim 16	Accused Instrumentality(ies)
	For example, Amazon uses automatic speech recognition (ASR).[221] Moreover, in the cloud, the separate components of Alexa's speech-processing stack — automatic speech recognition (ASR), whisper detection, and speaker identification — run on separate server nodes with their own powerful processors. On-device, those functions have to share hardware not only with each other but with Alexa's other core device functions, such as music playback.
a rule check unit configured to determine whether the recognized subsequent input speech corresponds to a target sentence.	The Accused Instrumentalities' continuous speech recognition unit comprises a rule check unit configured to determine whether the recognized subsequent input speech corresponds to a target sentence. For example, the Accused Instrumentalities use a Speechlet/Skill that looks at intent and determines if it can handle a request.[222]

[221] *See* KAIFI0006791 (Ariya Rastrow, Shehzad Mevawalla, *On-device speech processing makes Alexa faster, lower-bandwidth*, AMAZON SCIENCE (Jan. 25, 2022), https://www.amazon.science/blog/on-device-speech-processing-makes-alexa-faster-lower-bandwidth).

[222] *See* KAIFI0007380 *(White paper – Alexa Privacy and Data Handling Overview*, https://d1.awsstatic.com/product-marketing/A4B/White%20Paper%20-%20Alexa%20Privacy%20and%20Data%20Handling%20Overview.pdf)

Ex.1029 / Page 256 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 16	Accused Instrumentality(ies)
	 **Figure: Overview of Echo and the Alexa System** This claim element is literally present, as described herein, but is also satisfied under the doctrine of equivalents because the structure and functionality provided by Amazon is being used to satisfy this claim limitation and therefore there is no vitiation of this element, and the Accused Instrumentalities incorporate structure and functionality that is not substantially different from the requirements of this limitation

Ex.1029 / Page 257 of 347
AMAZON.COM, INC.

Claim 16	Accused Instrumentality(ies)
	because they achieve substantially or exactly the same function, in substantially or exactly the same way as set forth in the claim element (as described herein), and achieves substantially or exactly the same result.

Claim 17 of the ’196 Patent

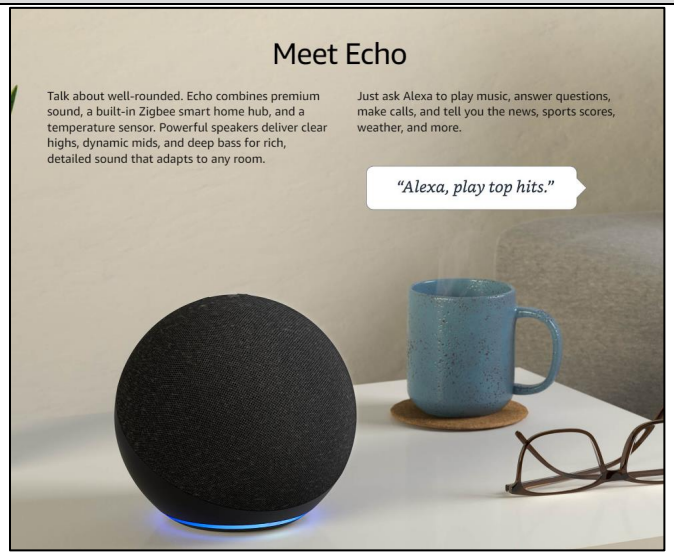
Claim 17	Accused Instrumentality(ies)
A method for recognizing a speech of a speaker by a single device, the method comprising:	Amazon practices a method of recognizing a speech of a speaker by a single device, for example, in its Echo and Alexa devices, such as the Echo, Echo Pop, Echo Auto, Echo Show, Echo Flex, Echo Dot, Echo Studio, Echo Buds, and Echo Hub, as well as in Alexa Built-in Devices, and on Amazon’s cloud system where it provides wake word detection, Automatic Speech Recognition (ASR) and Alexa Voice Service (AVS). For example, Amazon provides Echo (4th Gen) with premium sound, smart home hub, and Alexa.[223]

[223] *See* KAIFI0001202 (*Echo (4th Gen) | With premium sound, smart home hub, and Alexa | Charcoal*, https://www.amazon.com/dp/B07XKF5RM3?ref=MARS_NAV_desktop_shop_dvc_smart_spkrs_Echo_Spkr).

Ex.1029 / Page 258 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 17	Accused Instrumentality(ies)



More music in more rooms

Use your voice to play a song, artist, or genre through Amazon Music, Apple Music, Spotify, Pandora, and others. With multi-room music, you can fill your whole home with music, radio stations, podcasts, and Audible using compatible Echo devices in different rooms. You can even pair your Echo with compatible Fire TV devices to feel scenes come to life with home theater audio.

Ex.1029 / Page 259 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 17	Accused Instrumentality(ies)
	 For example, Amazon also provides the Echo Auto.[224]

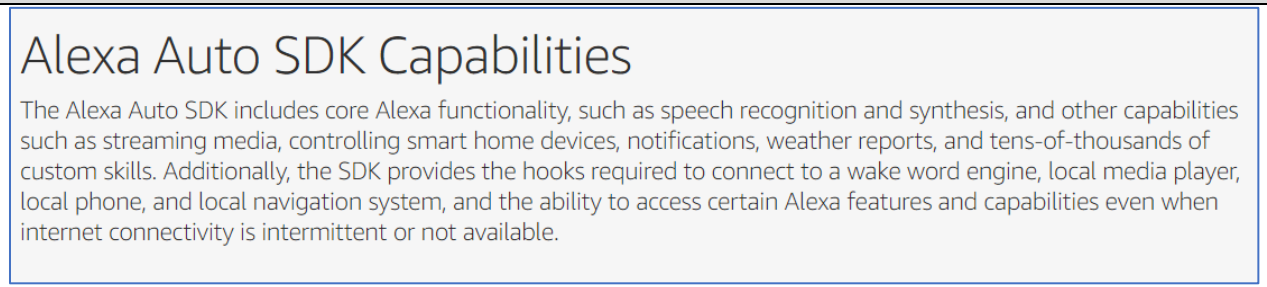
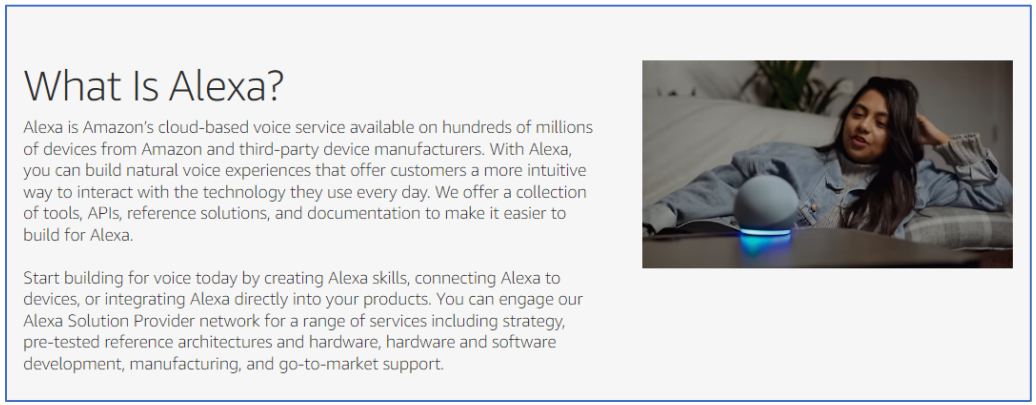
[224] *See* KAIFI0001979 *(Echo Auto (2nd Gen, 2022 release)*, https://www.amazon.com/All-New-Echo-Auto-release-Alexa/dp/B09X27YPS1?th=1).

Ex.1029 / Page 260 of 347
AMAZON.COM, INC.

Claim 17	Accused Instrumentality(ies)
	 According to Amazon, the Echo Auto provides speech recognition and synthesis.[225]

[225] *See* KAIFI0001144 *(Alexa Auto Software Development Kit*, https://developer.amazon.com/en-US/alexa/alexa-auto/sdk).

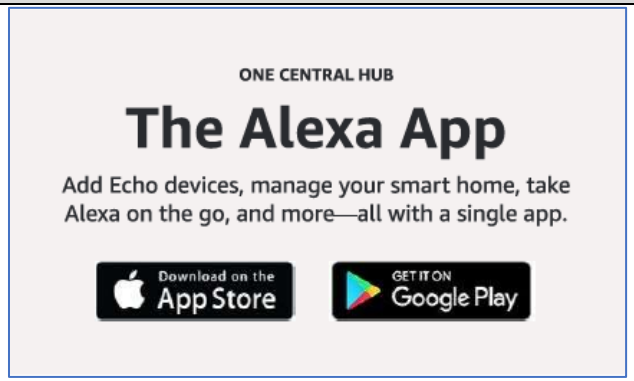
Ex.1029 / Page 261 of 347
AMAZON.COM, INC.

Claim 17	Accused Instrumentality(ies)
	 Amazon also provides Alexa.[226]  Amazon also provides the Alexa App.[227]

[226] *See* KAIFI0007144 (*What Is Alexa*, https://developer.amazon.com/en-US/alexa).

[227] *See* KAIFI0007048 (*The Alexa App*, https://www.amazon.com/Alexa-App/b?ie=UTF8&node=18354642011).

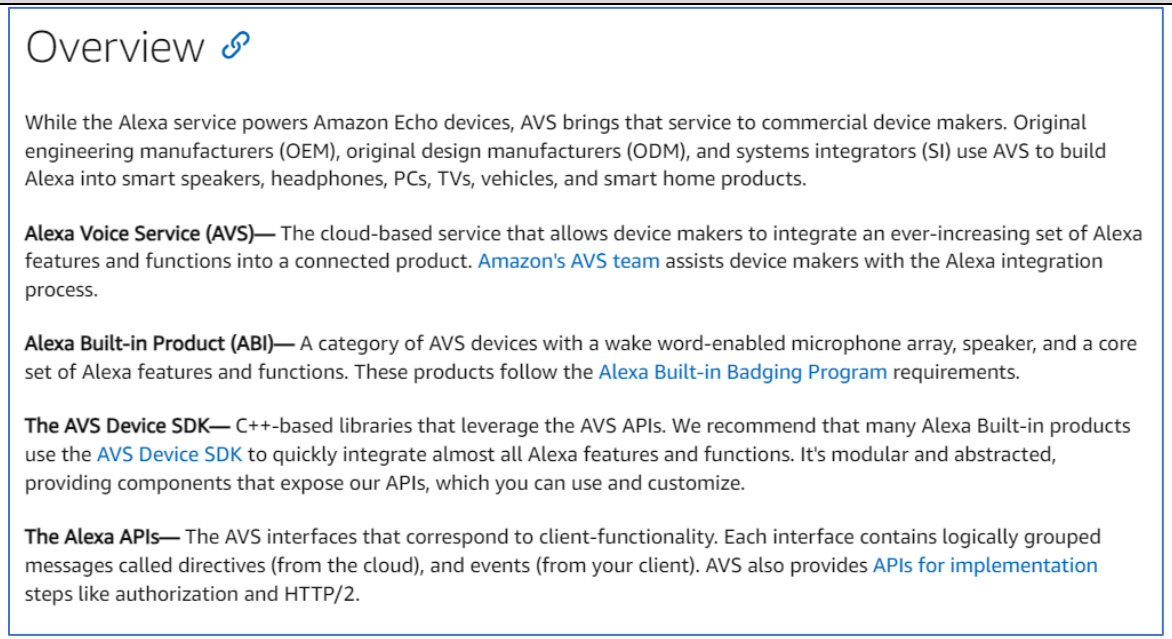
Ex.1029 / Page 262 of 347
AMAZON.COM, INC.

Claim 17	Accused Instrumentality(ies)
	 Amazon also provides Alexa Voice Service (AVS).[228]

[228] *See* KAIFI0007371 *(What is the Alexa Voice Service?*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/get-started-with-alexa-voice-service.html).

Ex.1029 / Page 263 of 347
AMAZON.COM, INC.

Claim 17	Accused Instrumentality(ies)



Overview

While the Alexa service powers Amazon Echo devices, AVS brings that service to commercial device makers. Original engineering manufacturers (OEM), original design manufacturers (ODM), and systems integrators (SI) use AVS to build Alexa into smart speakers, headphones, PCs, TVs, vehicles, and smart home products.

Alexa Voice Service (AVS)— The cloud-based service that allows device makers to integrate an ever-increasing set of Alexa features and functions into a connected product. Amazon's AVS team assists device makers with the Alexa integration process.

Alexa Built-in Product (ABI)— A category of AVS devices with a wake word-enabled microphone array, speaker, and a core set of Alexa features and functions. These products follow the Alexa Built-in Badging Program requirements.

The AVS Device SDK— C++-based libraries that leverage the AVS APIs. We recommend that many Alexa Built-in products use the AVS Device SDK to quickly integrate almost all Alexa features and functions. It's modular and abstracted, providing components that expose our APIs, which you can use and customize.

The Alexa APIs— The AVS interfaces that correspond to client-functionality. Each interface contains logically grouped messages called directives (from the cloud), and events (from your client). AVS also provides APIs for implementation steps like authorization and HTTP/2.

Amazon provides speech processing as both cloud-based and on-device.[229]

> At Amazon, we always look to invent new technology for improving customer experience. One technology we have been working on at Alexa is on-device speech processing, which has multiple benefits: a reduction in latency, or the time it takes Alexa to respond to queries; lowered bandwidth consumption, which is important on portable devices; and increased availability in in-car units and other applications where Internet connectivity is intermittent. On-device processing also enables the

[229] *See* KAIFI0006791 (Ariya Rastrow, Shehzad Mevawalla, *On-device speech processing makes Alexa faster, lower-bandwidth*, AMAZON SCIENCE (Jan. 25, 2022), https://www.amazon.science/blog/on-device-speech-processing-makes-alexa-faster-lower-bandwidth).

Ex.1029 / Page 264 of 347
AMAZON.COM, INC.

Claim 17	Accused Instrumentality(ies)
	fusion of the speech signal with other modalities, like vision, for features such as Alexa's natural turn-taking. In the last year, we've continued to build upon Alexa's on-device speech-processing capabilities. As a result of these inventions, we are launching a new setting that gives customers the option of having the audio of their Alexa voice requests processed locally, without being sent to the cloud. In the cloud, storage space and computational capacity are effectively unconstrained. To ensure accuracy, our cloud models can be large and computationally demanding. Executing the same functions on-device means compressing our models into less than 1% as much space — with minimal loss in accuracy. Moreover, in the cloud, the separate components of Alexa's speech-processing stack — automatic speech recognition (ASR), whisper detection, and speaker identification — run on separate server nodes with their own powerful processors. On-device, those functions have to share hardware not only with each other but with Alexa's other core device functions, such as music playback. Re-creating Alexa's speech-processing stack on-device was a massive undertaking. New methods for training small-footprint ASR models were part of the solution, but so were innovations in system design and hardware-software codesign. It was a joint effort across science and engineering teams over a span of years. Here's a quick overview of how it works. For example, Amazon describes the technology supporting a voice request.[230]

[230] *See* KAIFI0002023 *(Follow the journey of a voice request*, https://www.amazon.com/b/?node=23608618011).

Ex.1029 / Page 265 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 17	Accused Instrumentality(ies)
	You can ask Alexa any question and get a response in a matter of seconds. If you're interested in the technology behind all of that magic, you've come to the right place. Here's a more detailed look at how Alexa works and the ways in which Echo devices and Alexa are designed to protect your privacy every step of the way. **The wake word** Think of the wake word as a verbal cue that makes things happen. When you have a request for Alexa, you first need to say your chosen wake word, which by default is "Alexa." Only after your Echo device detects the wake word is Alexa listening to your requests. When it comes to privacy, there should be no surprises. You'll always be able to tell when Alexa is listening to your request because a light indicator will appear on your Echo device or an audible tone will sound. Think of the "On the Air" signs that light up in television studios during a broadcast. These indicators notify you that your device has detected the wake word and Alexa is now processing your request. **Keyword spotting** Echo devices use built-in technology called 'keyword spotting' that matches spoken audio to the acoustic patterns of the wake word. Simply put, Echo devices are designed by default to detect only the sound waves of your chosen wake word, and everything else is ignored. Like water through a strainer, all other audio (people talking, taps running, birds chirping) passes through the device until the wake word is 'caught' and sent to Amazon's secure cloud, where meaning is assigned to your request. **Cloud verification and encryption** When your Echo device detects the wake word, it sends your request to Amazon's secure cloud where audio is reanalysed to verify the wake word was spoken. If this cloud software verification is unable to confirm the wake word was spoken, the Alexa system stops processing the audio. If the wake word is verified (or Alexa is activated using the Action button), your request will be processed using several sophisticated algorithms that allow Alexa to respond appropriately. All of your interactions with Alexa are encrypted in transit throughout this process. **Automatic Speech Recognition (ASR)** After wake word verification, your voice request undergoes Automatic Speech Recognition (ASR) which transcribes audio into text. Using our example request (***"Alexa, what's the weather in sydney?"***), transcriptions could include one or several of the following:

Ex.1029 / Page 266 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 17	Accused Instrumentality(ies)
	For example, Amazon provides and uses Alexa Built-in Devices, which are marketed through Amazon.com, and powered by Alexa Voice Service (AVS).[231]

What are Alexa Built-in Devices?



The world's leading OEMs offer Alexa Built-in devices that let customers talk directly with Alexa through a microphone and speaker. Alexa turns your device into a smart device, and expands its capabilities. Customers can ask Alexa for help with everyday tasks and entertainment, and control Alexa connected devices or their own device.

Learn About the Underlying Technology

Alexa Built-in devices are powered by **Alexa Voice Service (AVS)**. The AVS team manages the services and infrastructure required for Alexa experiences, and provides a suite of device APIs, SDKs, and documentation. Developers may also work with third-party solution providers to utilize a qualified AVS solution.

Start Developing with Solution Providers

Our global network of **Alexa Solution Providers** work together to help you develop high-quality experiences, accelerate your development, and reduce your overall development cost. The earlier you engage a Solution Provider, the better they can assist you.

[231] *See* KAIFI0007733 *(Alexa Built-in Devices, https://developer.amazon.com/en-US/alexa/devices/alexa-built-in).*

Ex.1029 / Page 267 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 17	Accused Instrumentality(ies)
	 For example, AVS Integration is a feature of AWS IoT Core that enables Alexa Voice to be produced on any type of connected device, and shifts various Alexa Voice functions from a physical device to a new virtual Built-in device in the cloud.[232]

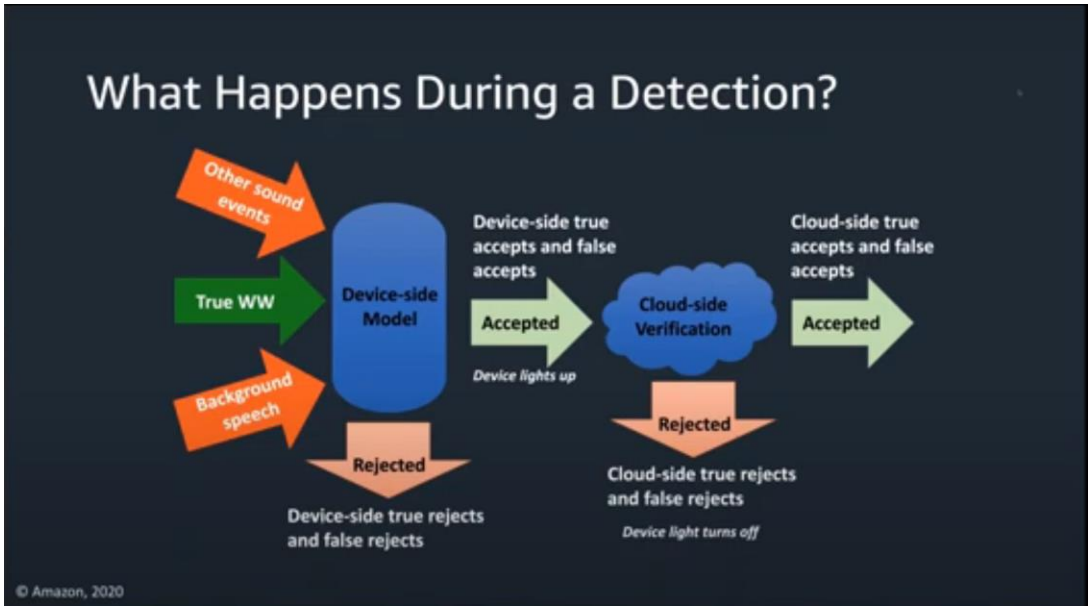
[232] *See* KAIFI0007818 *(New Alexa Voice Service Integration for AWS IoT Core cost-effectively brings Alexa Voice to any type of connected device*, https://aws.amazon.com/about-aws/whats-new/2019/11/new-alexa-voice-integration-for-aws-iot-core-cost-effectively-brings-alexa-voice-to-any-connected-device/*)*.

Ex.1029 / Page 268 of 347
AMAZON.COM, INC.

Claim 17	Accused Instrumentality(ies)
	New Alexa Voice Service Integration for AWS IoT Core cost-effectively brings Alexa Voice to any type of connected device Posted On: Nov 25, 2019 Alexa Voice Service (AVS) Integration is a new feature of AWS IoT Core that enables Alexa Voice to be produced on any type of connected device. AVS Integration for IoT Core reduces the cost of producing Alexa Built-in devices by up to 50% by offloading compute and memory intensive audio workloads to the cloud. With this reduction in production cost, customers can now cost-effectively build new categories of differentiated voice-enabled products such as light switches, thermostats, and small appliances. This allows consumers to talk directly to Alexa in new parts of their home, office, or hotel rooms for a truly ambient experience. Alexa Voice Service (AVS) Integration for IoT Core shifts media retrieval, audio decoding, audio mixing, and state management from a physical device to a new virtual Alexa Built-in device in the cloud. This lowers the hardware requirements of integrating AVS on a device from 50 MB of RAM to 1MB of RAM and ARM Cortex 'A' class microprocessors to ARM Cortex 'M' class microcontrollers. Once a device is connected to AWS IoT Core, customers can use a new set of AWS IoT-reserved MQTT topics to transfer audio messages between their devices and AVS. The AVS Integration for AWS IoT Core is available in all AWS regions where AWS IoT Core is available other than China (Beijing and Ningxia), Asia Pacific (Hong Kong) and Middle East (Bahrain). See the AWS Region Table for the current list of regions for AWS IoT Core. To learn more about how to get started, view the AVS Integration for IoT Core Blog Post.
receiving, by a preset call command recognition unit of the single device, an input speech spoken by the speaker;	Amazon receives, by a preset call command recognition unit of the single device, an input speech spoken by the speaker.

Ex.1029 / Page 269 of 347
AMAZON.COM, INC.

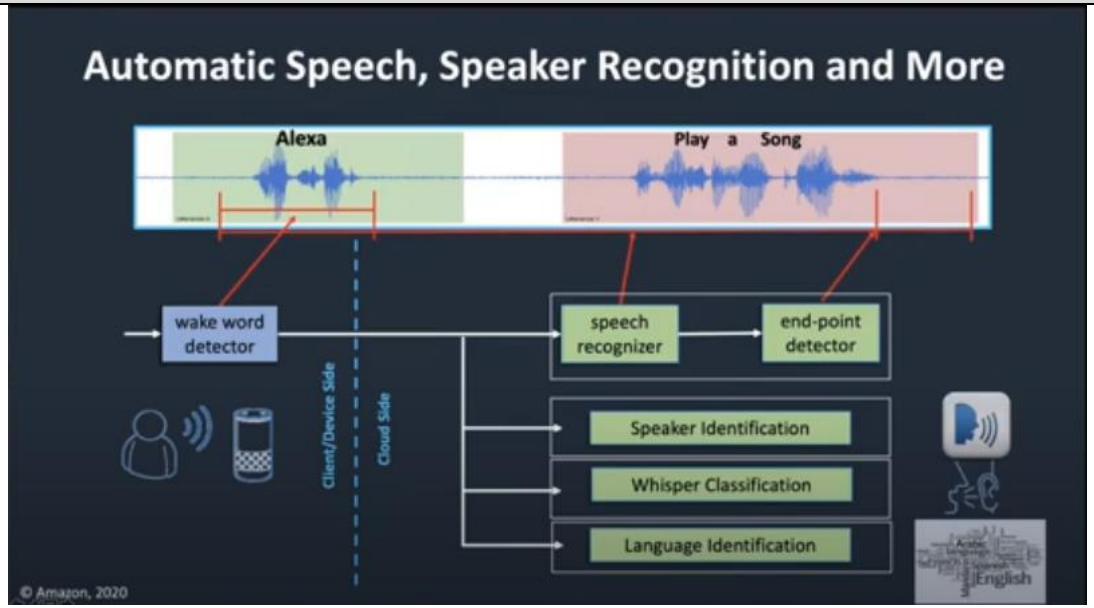
Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 17	Accused Instrumentality(ies)
	For example, Amazon receives a wake word through a combination of hardware, processing, and software functions both on devices and on the cloud server.[233]  This is further illustrated below.[234]

[233] *See* KAIFI0007814 *(The Science Behind Alexa's Wake Word Detection*, https://www.youtube.com/watch?v=kYrw2cC9dvs).
[234] *See* KAIFI0007816 *(How Alexa Uses Automatic Speech Recognition*, https://www.youtube.com/watch?v=8drK6Oyv1a0&t=296s).

Ex.1029 / Page 270 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 17	Accused Instrumentality(ies)
	 For example, the wake word engine is configured to receive an input speech spoken by the speaker. [235] Voice-initiated devices begin to stream user speech to AVS when the wake word engine detects a spoken wake word, such as "Alexa."

[235] *See* KAIFI0007141 *(Wake Word Verification Requirements*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/implement-ww-verification.html).

Ex.1029 / Page 271 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 17	Accused Instrumentality(ies)
	4. **Wake Word Engine** - The component that reads the audio samples written to the shared memory ring buffer and analyzes the audio samples for occurrences of the wake word. The procedure of reading and processing audio samples occurs in a loop. For example, the loop reads N samples starting at location `0`; then analyzes the samples for an occurrence of the wake word. In the next iteration, it reads and analyzes the next N samples. When an occurrence of the wake word is detected, the wake word engine must identify two locations: 1) the start of the wake word, and 2) the end of the wake word. These values are stored as `startIndexInSamples` and `endIndexInSamples`. For example, according to Amazon, its speech recognizer captures and interacts with user speech.[236] The `SpeechRecognizer` interface is the core interface of the Alexa Voice Service (AVS) and exposes directives and events for capturing and interacting with user speech. This page discusses the concepts and process flows for the `SpeechRecognizer` interface.

[236] *See* KAIFI0007093 *(Understand the AVS SpeechRecognizer*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/avs-speechrecognizer-concepts.html).

Ex.1029 / Page 272 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 17	Accused Instrumentality(ies)
	SpeechRecognizer functionality Every user utterance leverages `SpeechRecognizer`, including the following interaction types between an Alexa Built-in device and AVS: • **User Speech Capture**: Captures user speech from an Alexa Built-in device. • **User Speech Prompting**: Prompts a user for more speech input when needed by Alexa to deliver an appropriate response. • **Interaction Initiation Communication**: Enables a device to inform AVS of how a user initiated an Alexa interaction, such as press-and-hold, tap-and-release, or voice-initiated/wake word enabled. See Device Form Factor and Alexa Interaction. • **ASR Profile Selection**: SpeechRecognizers chooses the appropriate Automatic Speech Recognition (ASR) profile for your product, which allows Alexa to understand user speech and respond with precision. See Automatic Speech Recognition (ASR) profile • **RECOGNIZING:** When a user begins interacting with your client, specifically when the client streams captured audio to AVS, `SpeechRecognizer` should transition from the idle state to the `Recognize` state. `SpeechRecognizer` should remain in the `Recognize` until the client stops recording speech or finishes streaming, at which point your `SpeechRecognizer` component should transition from the `Recognize` state to the "busy" state. • **BUSY:** While processing the speech request, `SpeechRecognizer` should be in the "busy" state. You cannot start another speech request until `SpeechRecognizer` transitions out of the busy state. From the *busy* state, `SpeechRecognizer` transitions to the idle state if Alexa processes and completes the request or to the `ExpectSpeech` state, if Alexa requires more speech input from the user. • **EXPECTING SPEECH:** `SpeechRecognizer` should be in the `ExpectSpeech` state when Alexa requires more audio input from a user. From `ExpectSpeech`, `SpeechRecognizer` should transition to the `Recognize` when a user interaction occurs, or the interaction is automatically started on behalf of the user. It should transition to the idle state if Alexa detects no user interaction within the specified timeout window.

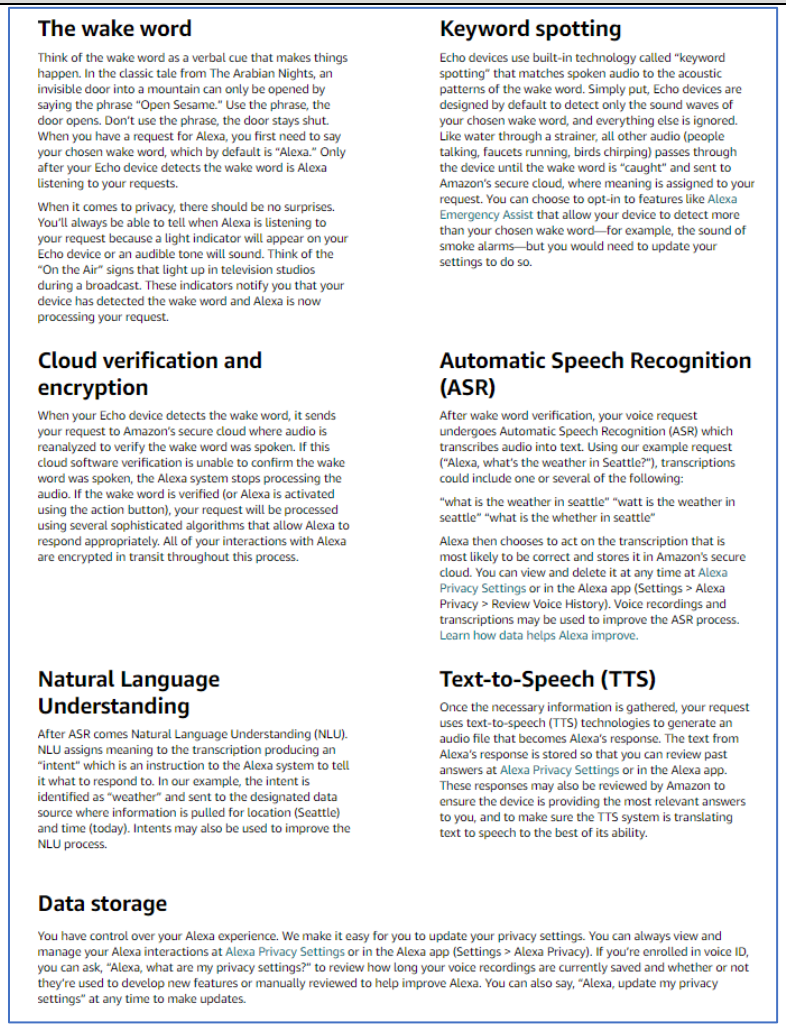
Ex.1029 / Page 273 of 347
AMAZON.COM, INC.

Claim 17	Accused Instrumentality(ies)
	For example, Amazon uses wake word, keyword spotting, cloud verification and encryption, automatic speech recognition (ASR), natural language understanding (NLU), text-to-speech (TTS), and data storage technologies.[237]

[237] *See* KAIFI0002023 *(Follow the journey of a voice request*, https://www.amazon.com/b/?node=23608618011).

Ex.1029 / Page 274 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 17	Accused Instrumentality(ies)
	

The wake word

Think of the wake word as a verbal cue that makes things happen. In the classic tale from The Arabian Nights, an invisible door into a mountain can only be opened by saying the phrase "Open Sesame." Use the phrase, the door opens. Don't use the phrase, the door stays shut. When you have a request for Alexa, you first need to say your chosen wake word, which by default is "Alexa." Only after your Echo device detects the wake word is Alexa listening to your requests.

When it comes to privacy, there should be no surprises. You'll always be able to tell when Alexa is listening to your request because a light indicator will appear on your Echo device or an audible tone will sound. Think of the "On the Air" signs that light up in television studios during a broadcast. These indicators notify you that your device has detected the wake word and Alexa is now processing your request.

Keyword spotting

Echo devices use built-in technology called "keyword spotting" that matches spoken audio to the acoustic patterns of the wake word. Simply put, Echo devices are designed by default to detect only the sound waves of your chosen wake word, and everything else is ignored. Like water through a strainer, all other audio (people talking, faucets running, birds chirping) passes through the device until the wake word is "caught" and sent to Amazon's secure cloud, where meaning is assigned to your request. You can choose to opt-in to features like Alexa Emergency Assist that allow your device to detect more than your chosen wake word—for example, the sound of smoke alarms—but you would need to update your settings to do so.

Cloud verification and encryption

When your Echo device detects the wake word, it sends your request to Amazon's secure cloud where audio is reanalyzed to verify the wake word was spoken. If this cloud software verification is unable to confirm the wake word was spoken, the Alexa system stops processing the audio. If the wake word is verified (or Alexa is activated using the action button), your request will be processed using several sophisticated algorithms that allow Alexa to respond appropriately. All of your interactions with Alexa are encrypted in transit throughout this process.

Automatic Speech Recognition (ASR)

After wake word verification, your voice request undergoes Automatic Speech Recognition (ASR) which transcribes audio into text. Using our example request ("Alexa, what's the weather in Seattle?"), transcriptions could include one or several of the following:

"what is the weather in seattle" "watt is the weather in seattle" "what is the whether in seattle"

Alexa then chooses to act on the transcription that is most likely to be correct and stores it in Amazon's secure cloud. You can view and delete it at any time at Alexa Privacy Settings or in the Alexa app (Settings > Alexa Privacy > Review Voice History). Voice recordings and transcriptions may be used to improve the ASR process. Learn how data helps Alexa improve.

Natural Language Understanding

After ASR comes Natural Language Understanding (NLU). NLU assigns meaning to the transcription producing an "intent" which is an instruction to the Alexa system to tell it what to respond to. In our example, the intent is identified as "weather" and sent to the designated data source where information is pulled for location (Seattle) and time (today). Intents may also be used to improve the NLU process.

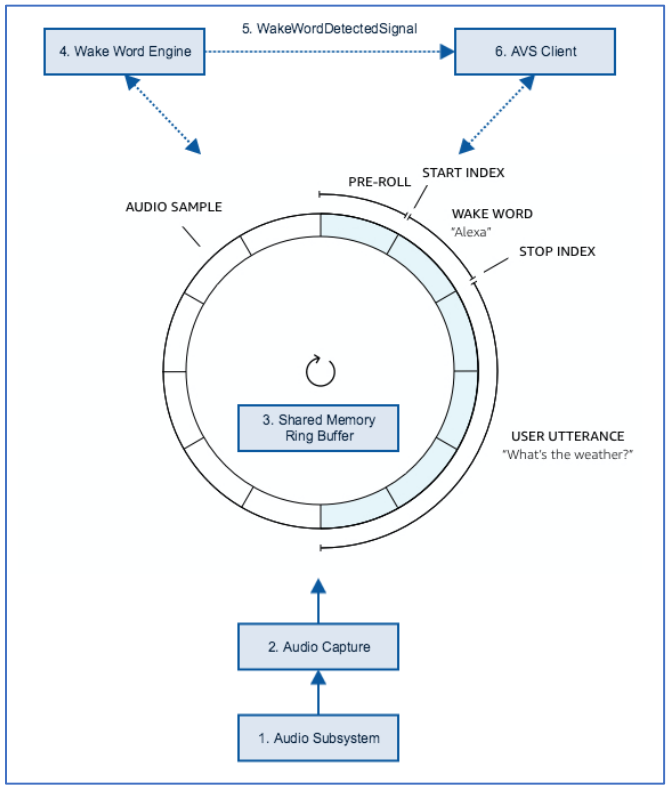
Text-to-Speech (TTS)

Once the necessary information is gathered, your request uses text-to-speech (TTS) technologies to generate an audio file that becomes Alexa's response. The text from Alexa's response is stored so that you can review past answers at Alexa Privacy Settings or in the Alexa app. These responses may also be reviewed by Amazon to ensure the device is providing the most relevant answers to you, and to make sure the TTS system is translating text to speech to the best of its ability.

Data storage

You have control over your Alexa experience. We make it easy for you to update your privacy settings. You can always view and manage your Alexa interactions at Alexa Privacy Settings or in the Alexa app (Settings > Alexa Privacy). If you're enrolled in voice ID, you can ask, "Alexa, what are my privacy settings?" to review how long your voice recordings are currently saved and whether or not they're used to develop new features or manually reviewed to help improve Alexa. You can also say, "Alexa, update my privacy settings" at any time to make updates.

Ex.1029 / Page 275 of 347
AMAZON.COM, INC.

Claim 17	Accused Instrumentality(ies)
	For example, the wake word engine is configured to receive an input speech spoken by the speaker.[238] 

[238] *See* KAIFI0007141 *(Wake Word Verification Requirements*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/implement-ww-verification.html).

Ex.1029 / Page 276 of 347
AMAZON.COM, INC.

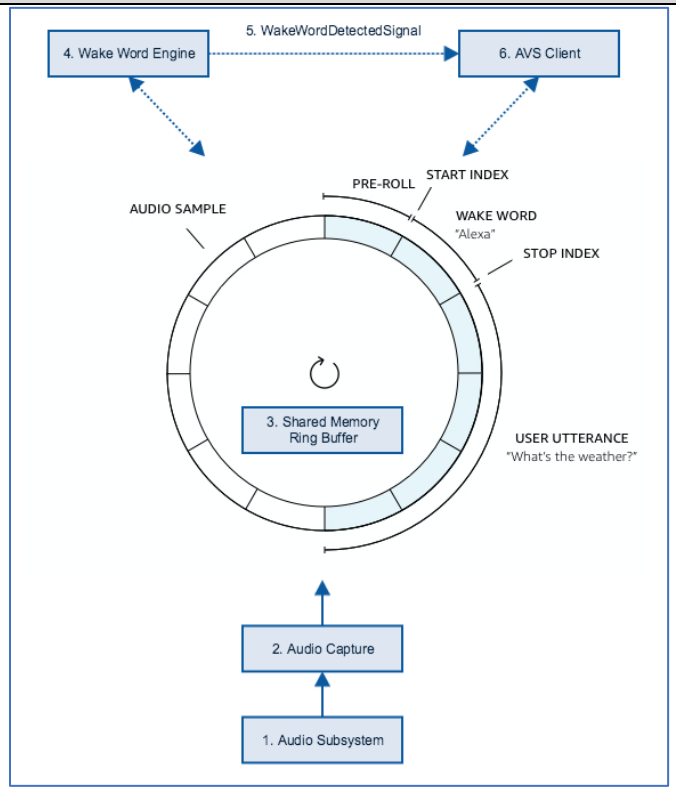
Claim 17	Accused Instrumentality(ies)
	3. **Shared Memory Ring Buffer** - A shared memory ring buffer object that is thread safe, and accessible by multiple processes. Access to this memory block should support at least one writer and two readers simultaneously. You may choose to implement these executables as a shared library. • **Write API** - Used by the recording process to write audio samples to the shared memory ring buffer. Write samples should move data so that the oldest audio samples are overwritten. This example shows a typical API signature: `WriteSamples(BufferIn, BufferInLength)`. • **Read API** - Used by the wake word engine and the AVS client to read from specific locations in the shared memory ring buffer. A typical implementation represents location as a 64-bit integer. For example, reading 200 samples from the location 100 is reading the 100th to 300th samples that were written to the shared memory ring buffer. This is an example of a typical API signature: `ReadSamples(BufferIn, BufferInLength)`. **Note**: Reading from an old location that was overwritten with new audio samples will result in an unintended sample read. 4. **Wake Word Engine** - The component that reads the audio samples written to the shared memory ring buffer and analyzes the audio samples for occurrences of the wake word. The procedure of reading and processing audio samples occurs in a loop. For example, the loop reads N samples starting at location `0`; then analyzes the samples for an occurrence of the wake word. In the next iteration, it reads and analyzes the next N samples. When an occurrence of the wake word is detected, the wake word engine must identify two locations: 1) the start of the wake word, and 2) the end of the wake word. These values are stored as `startIndexInSamples` and `endIndexInSamples`. For example, according to Amazon, the AVS Client is also configured to receive an input speech spoken by the speaker.[239]

[239] *See Id.*

Ex.1029 / Page 277 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 17	Accused Instrumentality(ies)



6. **AVS Client** - The client receives the `WakeWordDetectedSignal`, extracts the `startIndexInSamples`, then reads and streams audio samples to AVS including 500 milliseconds of pre-roll.

- The pre-roll is the number of samples corresponding to 500 milliseconds. For example, if the audio samples were recorded at a rate of 16 kilohertz (kHz), then pre-roll is 8,000 samples.

Ex.1029 / Page 278 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 17	Accused Instrumentality(ies)
	For example, according to Amazon, the Alexa Voice Service speech recognizer captures and interacts with user speech.[240]

> The `SpeechRecognizer` interface is the core interface of the Alexa Voice Service (AVS) and exposes directives and events for capturing and interacting with user speech. This page discusses the concepts and process flows for the `SpeechRecognizer` interface.

SpeechRecognizer functionality

Every user utterance leverages `SpeechRecognizer`, including the following interaction types between an Alexa Built-in device and AVS:

- **User Speech Capture**: Captures user speech from an Alexa Built-in device.
- **User Speech Prompting**: Prompts a user for more speech input when needed by Alexa to deliver an appropriate response.
- **Interaction Initiation Communication**: Enables a device to inform AVS of how a user initiated an Alexa interaction, such as press-and-hold, tap-and-release, or voice-initiated/wake word enabled. See Device Form Factor and Alexa Interaction.
- **ASR Profile Selection**: SpeechRecognizers chooses the appropriate Automatic Speech Recognition (ASR) profile for your product, which allows Alexa to understand user speech and respond with precision. See Automatic Speech Recognition (ASR) profile

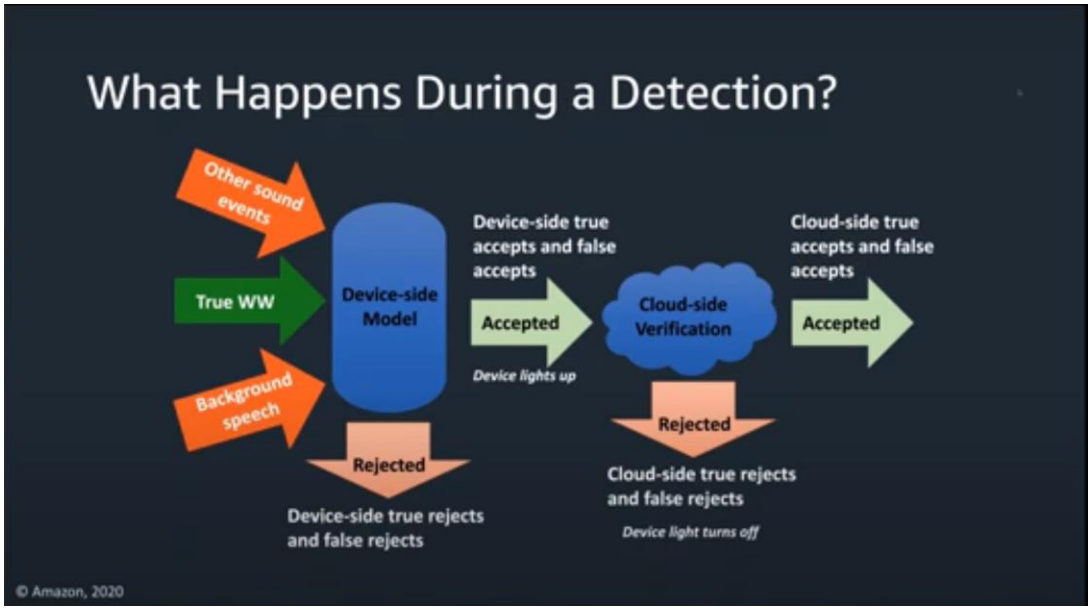
[240] *See* KAIFI0007093 *(Understand the AVS SpeechRecognizer*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/avs-speechrecognizer-concepts.html).

Ex.1029 / Page 279 of 347
AMAZON.COM, INC.

Claim 17	Accused Instrumentality(ies)
	• **RECOGNIZING:** When a user begins interacting with your client, specifically when the client streams captured audio to AVS, `SpeechRecognizer` should transition from the idle state to the `Recognize` state. `SpeechRecognizer` should remain in the `Recognize` until the client stops recording speech or finishes streaming, at which point your `SpeechRecognizer` component should transition from the `Recognize` state to the "busy" state. • **BUSY:** While processing the speech request, `SpeechRecognizer` should be in the "busy" state. You cannot start another speech request until `SpeechRecognizer` transitions out of the busy state. From the *busy* state, `SpeechRecognizer` transitions to the idle state if Alexa processes and completes the request or to the `ExpectSpeech` state, if Alexa requires more speech input from the user. • **EXPECTING SPEECH:** `SpeechRecognizer` should be in the `ExpectSpeech` state when Alexa requires more audio input from a user. From `ExpectSpeech`, `SpeechRecognizer` should transition to the `Recognize` when a user interaction occurs, or the interaction is automatically started on behalf of the user. It should transition to the idle state if Alexa detects no user interaction within the specified timeout window. This claim element is literally present, as described herein, but is also satisfied under the doctrine of equivalents because the structure and functionality provided by Amazon is being used to satisfy this claim limitation and therefore there is no vitiation of this element, and the Accused Instrumentalities incorporate structure and functionality that is not substantially different from the requirements of this limitation because they achieve substantially or exactly the same function, in substantially or exactly the same way as set forth in the claim element (as described herein), and achieves substantially or exactly the same result.
receiving, by the preset call command recognition unit, a subsequent input speech spoken, by the speaker, subsequently to the input speech;	Amazon receives, by the preset call command recognition unit, a subsequent input speech spoken, by the speaker, subsequently to the input speech.

Ex.1029 / Page 280 of 347
AMAZON.COM, INC.

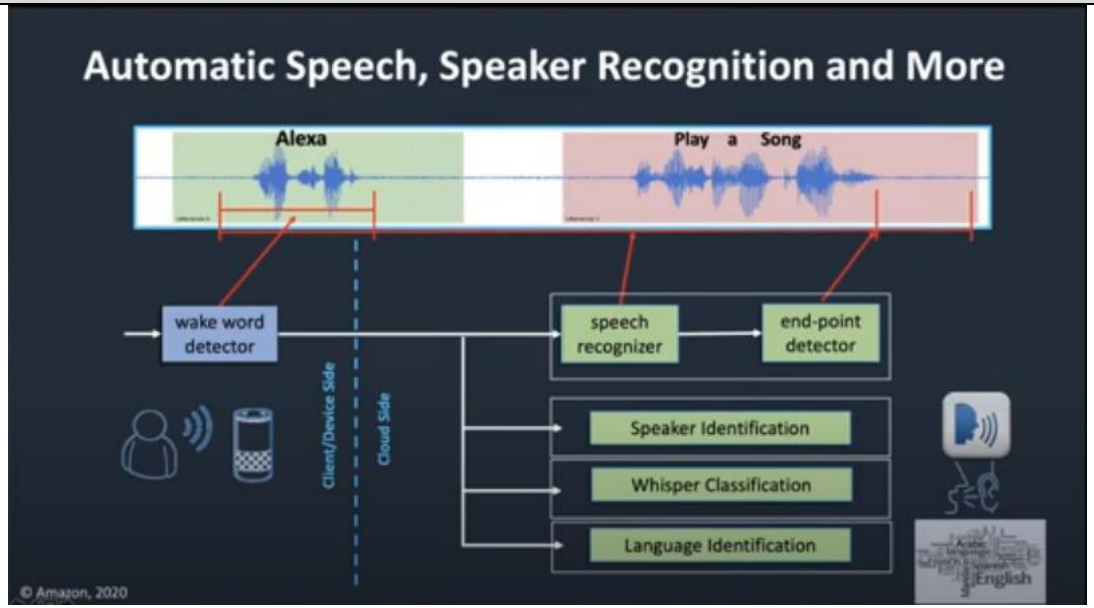
Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 17	Accused Instrumentality(ies)
	For example, Amazon receives wake word and a subsequent speech, through a combination of hardware, processing, and software functions both on devices and on the cloud server.[241]  This is further illustrated below.[242]

[241] *See* KAIFI0007814 *(The Science Behind Alexa's Wake Word Detection*, https://www.youtube.com/watch?v=kYrw2cC9dvs).
[242] *See* KAIFI0007816 *(How Alexa Uses Automatic Speech Recognition*, https://www.youtube.com/watch?v=8drK6Oyv1a0&t=296s).

Ex.1029 / Page 281 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

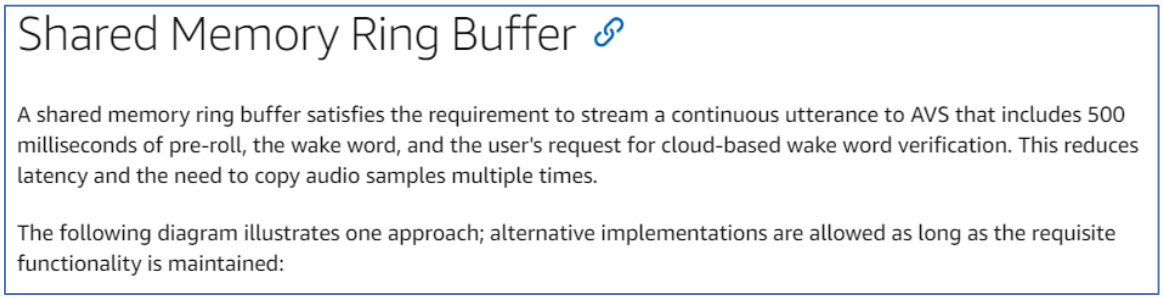
Claim 17	Accused Instrumentality(ies)
	 For example, Amazon describes a voice request.[243]

[243] *See* KAIFI0002023 *(Follow the journey of a voice request*, https://www.amazon.com/b/?node=23608618011).

Ex.1029 / Page 282 of 347
AMAZON.COM, INC.

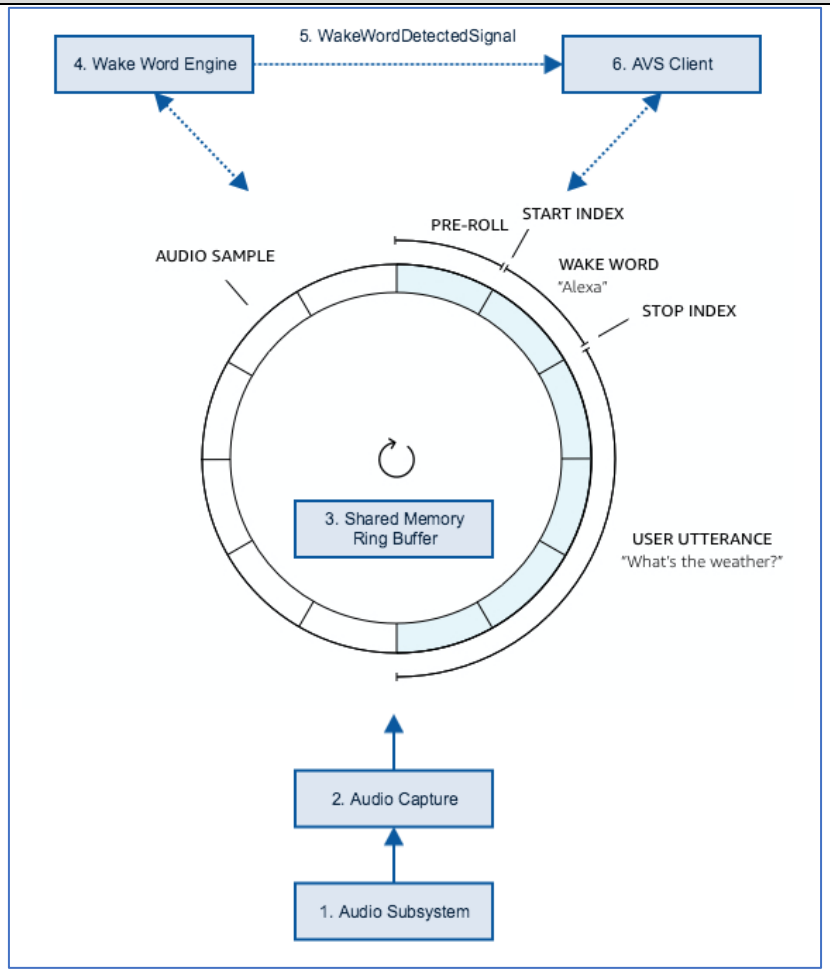
Ex.1029 / Page 283 of 347
AMAZON.COM, INC.

Claim 17	Accused Instrumentality(ies)
	For example, Amazon provides a shared memory ring buffer.[244]  Shared Memory Ring Buffer A shared memory ring buffer satisfies the requirement to stream a continuous utterance to AVS that includes 500 milliseconds of pre-roll, the wake word, and the user's request for cloud-based wake word verification. This reduces latency and the need to copy audio samples multiple times. The following diagram illustrates one approach; alternative implementations are allowed as long as the requisite functionality is maintained:

[244] *See* KAIFI0007141 *(Wake Word Verification Requirements*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/implement-ww-verification.html).

Ex.1029 / Page 284 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 17	Accused Instrumentality(ies)
	

Ex.1029 / Page 285 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 17	Accused Instrumentality(ies)
	1. **Audio Subsystem** - Middleware, such as Advanced Linux Sound Architecture (ALSA), that provides APIs to open a stream from a recording device, and receive an audio sample stream. 2. **Audio Capture** - A process that opens a recording device, and writes audio samples into the shared memory ring buffer. 3. **Shared Memory Ring Buffer** - A shared memory ring buffer object that is thread safe, and accessible by multiple processes. Access to this memory block should support at least one writer and two readers simultaneously. You may choose to implement these executables as a shared library. • **Write API** - Used by the recording process to write audio samples to the shared memory ring buffer. Write samples should move data so that the oldest audio samples are overwritten. This example shows a typical API signature: `WriteSamples(BufferIn, BufferInLength)`. • **Read API** - Used by the wake word engine and the AVS client to read from specific locations in the shared memory ring buffer. A typical implementation represents location as a 64-bit integer. For example, reading 200 samples from the location 100 is reading the 100th to 300th samples that were written to the shared memory ring buffer. This is an example of a typical API signature: `ReadSamples(BufferIn, BufferInLength)`. **Note**: Reading from an old location that was overwritten with new audio samples will result in an unintended sample read. 4. **Wake Word Engine** - The component that reads the audio samples written to the shared memory ring buffer and analyzes the audio samples for occurrences of the wake word. The procedure of reading and processing audio samples occurs in a loop. For example, the loop reads N samples starting at location `0`; then analyzes the samples for an occurrence of the wake word. In the next iteration, it reads and analyzes the next N samples. When an occurrence of the wake word is detected, the wake word engine must identify two locations: 1) the start of the wake word, and 2) the end of the wake word. These values are stored as `startIndexInSamples` and `endIndexInSamples`. 5. **WakeWordDetectedSignal** - A signal sent from the wake word engine to the AVS client when an occurrence of the wake word is detected. This signal includes two locations, the `startIndexInSamples` and `endIndexInSamples`. 6. **AVS Client** - The client receives the `WakeWordDetectedSignal`, extracts the `startIndexInSamples`, then reads and streams audio samples to AVS including 500 milliseconds of pre-roll. • The pre-roll is the number of samples corresponding to 500 milliseconds. For example, if the audio samples were recorded at a rate of 16 kilohertz (kHz), then pre-roll is 8,000 samples.

Ex.1029 / Page 286 of 347
AMAZON.COM, INC.

Claim 17	Accused Instrumentality(ies)
	For example, according to Amazon, its speech recognizer captures and interacts with user speech.[245] The `SpeechRecognizer` interface is the core interface of the Alexa Voice Service (AVS) and exposes directives and events for capturing and interacting with user speech. This page discusses the concepts and process flows for the `SpeechRecognizer` interface. SpeechRecognizer functionality Every user utterance leverages `SpeechRecognizer`, including the following interaction types between an Alexa Built-in device and AVS: • **User Speech Capture**: Captures user speech from an Alexa Built-in device. • **User Speech Prompting**: Prompts a user for more speech input when needed by Alexa to deliver an appropriate response. • **Interaction Initiation Communication**: Enables a device to inform AVS of how a user initiated an Alexa interaction, such as press-and-hold, tap-and-release, or voice-initiated/wake word enabled. See Device Form Factor and Alexa Interaction. • **ASR Profile Selection**: SpeechRecognizers chooses the appropriate Automatic Speech Recognition (ASR) profile for your product, which allows Alexa to understand user speech and respond with precision. See Automatic Speech Recognition (ASR) profile

[245] *See* KAIFI0007093 *(Understand the AVS SpeechRecognizer*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/avs-speechrecognizer-concepts.html).

Ex.1029 / Page 287 of 347
AMAZON.COM, INC.

Claim 17	Accused Instrumentality(ies)
	• **RECOGNIZING:** When a user begins interacting with your client, specifically when the client streams captured audio to AVS, `SpeechRecognizer` should transition from the idle state to the `Recognize` state. `SpeechRecognizer` should remain in the `Recognize` until the client stops recording speech or finishes streaming, at which point your `SpeechRecognizer` component should transition from the `Recognize` state to the "busy" state. • **BUSY:** While processing the speech request, `SpeechRecognizer` should be in the "busy" state. You cannot start another speech request until `SpeechRecognizer` transitions out of the busy state. From the *busy* state, `SpeechRecognizer` transitions to the idle state if Alexa processes and completes the request or to the `ExpectSpeech` state, if Alexa requires more speech input from the user. • **EXPECTING SPEECH:** `SpeechRecognizer` should be in the `ExpectSpeech` state when Alexa requires more audio input from a user. From `ExpectSpeech`, `SpeechRecognizer` should transition to the `Recognize` when a user interaction occurs, or the interaction is automatically started on behalf of the user. It should transition to the idle state if Alexa detects no user interaction within the specified timeout window. For example, Amazon uses wake word, keyword spotting, cloud verification and encryption, automatic speech recognition (ASR), natural language understanding (NLU), text-to-speech (TTS), and data storage technologies. [246]

[246] *See* KAIFI0002023 *(Follow the journey of a voice request*, https://www.amazon.com/b/?node=23608618011).

Ex.1029 / Page 288 of 347
AMAZON.COM, INC.

The wake word

Think of the wake word as a verbal cue that makes things happen. In the classic tale from The Arabian Nights, an invisible door into a mountain can only be opened by saying the phrase "Open Sesame." Use the phrase, the door opens. Don't use the phrase, the door stays shut. When you have a request for Alexa, you first need to say your chosen wake word, which by default is "Alexa." Only after your Echo device detects the wake word is Alexa listening to your requests.

When it comes to privacy, there should be no surprises. You'll always be able to tell when Alexa is listening to your request because a light indicator will appear on your Echo device or an audible tone will sound. Think of the "On the Air" signs that light up in television studios during a broadcast. These indicators notify you that your device has detected the wake word and Alexa is now processing your request.

Keyword spotting

Echo devices use built-in technology called "keyword spotting" that matches spoken audio to the acoustic patterns of the wake word. Simply put, Echo devices are designed by default to detect only the sound waves of your chosen wake word, and everything else is ignored. Like water through a strainer, all other audio (people talking, faucets running, birds chirping) passes through the device until the wake word is "caught" and sent to Amazon's secure cloud, where meaning is assigned to your request. You can choose to opt-in to features like Alexa Emergency Assist that allow your device to detect more than your chosen wake word—for example, the sound of smoke alarms—but you would need to update your settings to do so.

Cloud verification and encryption

When your Echo device detects the wake word, it sends your request to Amazon's secure cloud where audio is reanalyzed to verify the wake word was spoken. If this cloud software verification is unable to confirm the wake word was spoken, the Alexa system stops processing the audio. If the wake word is verified (or Alexa is activated using the action button), your request will be processed using several sophisticated algorithms that allow Alexa to respond appropriately. All of your interactions with Alexa are encrypted in transit throughout this process.

Automatic Speech Recognition (ASR)

After wake word verification, your voice request undergoes Automatic Speech Recognition (ASR) which transcribes audio into text. Using our example request ("Alexa, what's the weather in Seattle?"), transcriptions could include one or several of the following:

"what is the weather in seattle" "watt is the weather in seattle" "what is the whether in seattle"

Alexa then chooses to act on the transcription that is most likely to be correct and stores it in Amazon's secure cloud. You can view and delete it at any time at Alexa Privacy Settings or in the Alexa app (Settings > Alexa Privacy > Review Voice History). Voice recordings and transcriptions may be used to improve the ASR process. Learn how data helps Alexa improve.

Natural Language Understanding

After ASR comes Natural Language Understanding (NLU). NLU assigns meaning to the transcription producing an "intent" which is an instruction to the Alexa system to tell it what to respond to. In our example, the intent is identified as "weather" and sent to the designated data source where information is pulled for location (Seattle) and time (today). Intents may also be used to improve the NLU process.

Text-to-Speech (TTS)

Once the necessary information is gathered, your request uses text-to-speech (TTS) technologies to generate an audio file that becomes Alexa's response. The text from Alexa's response is stored so that you can review past answers at Alexa Privacy Settings or in the Alexa app. These responses may also be reviewed by Amazon to ensure the device is providing the most relevant answers to you, and to make sure the TTS system is translating text to speech to the best of its ability.

Data storage

You have control over your Alexa experience. We make it easy for you to update your privacy settings. You can always view and manage your Alexa interactions at Alexa Privacy Settings or in the Alexa app (Settings > Alexa Privacy). If you're enrolled in voice ID, you can ask, "Alexa, what are my privacy settings?" to review how long your voice recordings are currently saved and whether or not they're used to develop new features or manually reviewed to help improve Alexa. You can also say, "Alexa, update my privacy settings" at any time to make updates.

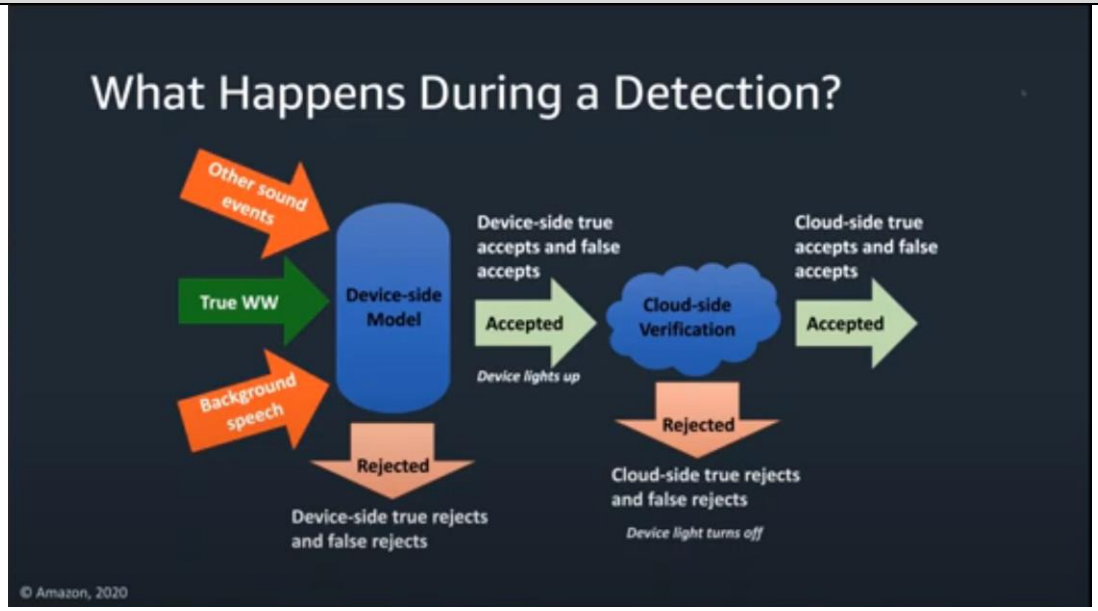
Ex.1029 / Page 289 of 347
AMAZON.COM, INC.

Claim 17	Accused Instrumentality(ies)
	This claim element is literally present, as described herein, but is also satisfied under the doctrine of equivalents because the structure and functionality provided by Amazon is being used to satisfy this claim limitation and therefore there is no vitiation of this element, and the Accused Instrumentalities incorporate structure and functionality that is not substantially different from the requirements of this limitation because they achieve substantially or exactly the same function, in substantially or exactly the same way as set forth in the claim element (as described herein), and achieves substantially or exactly the same result.
comparing, by the preset call command recognition unit, the input speech with a preset call command;	Amazon compares, by the preset call command recognition unit, the input speech with a preset call command. For example, Amazon compares, through a combination of hardware, processing, and software functions both on devices and on the cloud server, an input speech with a wake word.[247]

[247] *See* KAIFI0007814 *(The Science Behind Alexa's Wake Word Detection*, https://www.youtube.com/watch?v=kYrw2cC9dvs).

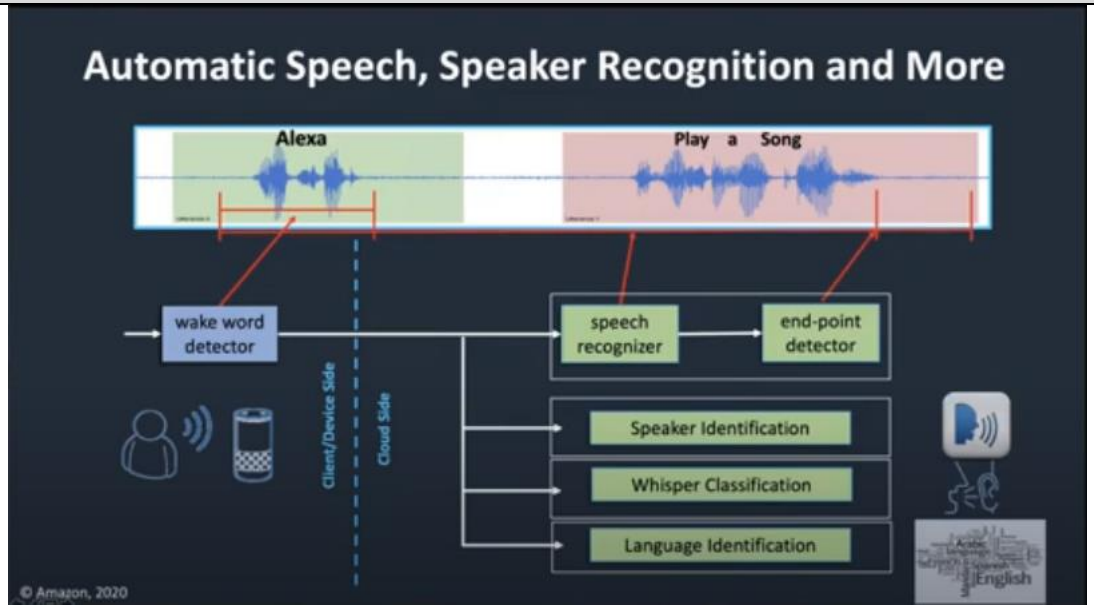
Ex.1029 / Page 290 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 17	Accused Instrumentality(ies)
	 This is further illustrated below.[248]

[248] *See* KAIFI0007816 *(How Alexa Uses Automatic Speech Recognition*, https://www.youtube.com/watch?v=8drK6Oyv1a0&t=296s).

Ex.1029 / Page 291 of 347
AMAZON.COM, INC.

Claim 17	Accused Instrumentality(ies)
	 For example, Amazon uses keyword spotting.[249] Echo devices use built-in technology called "keyword spotting" that matches spoken audio to the acoustic patterns of the wake word. Simply put, Echo devices are designed by default to detect only the sound waves of your chosen wake word, and everything else is ignored.

[249] *See* KAIFI0002092 *(How Alexa works: Wake word*, https://www.amazon.com/b?ie=UTF8&node=23608571011).

Ex.1029 / Page 292 of 347
AMAZON.COM, INC.

Claim 17	Accused Instrumentality(ies)
	For example, Amazon uses wake words.[250] Wake words The list of wake words informs Alexa of the possible valid wake words that a device might be set to listen for through the `SetWakeWords`, `WakeWordsReport`, and `WakeWordsChanged` messages. Currently, the only wake word available for Alexa Built-in devices is `ALEXA`, which applies to every possible locale for a device. Therefore, specify `ALEXA` in the global `DEFAULT` scope. For example, Amazon provides a wake word verification engine.[251] 4. **Wake Word Engine** - The component that reads the audio samples written to the shared memory ring buffer and analyzes the audio samples for occurrences of the wake word. The procedure of reading and processing audio samples occurs in a loop. For example, the loop reads N samples starting at location `0`; then analyzes the samples for an occurrence of the wake word. In the next iteration, it reads and analyzes the next N samples. When an occurrence of the wake word is detected, the wake word engine must identify two locations: 1) the start of the wake word, and 2) the end of the wake word. These values are stored as `startIndexInSamples` and `endIndexInSamples`. According to Amazon, the speech recognizer transitions are illustrated in the below diagram.[252]

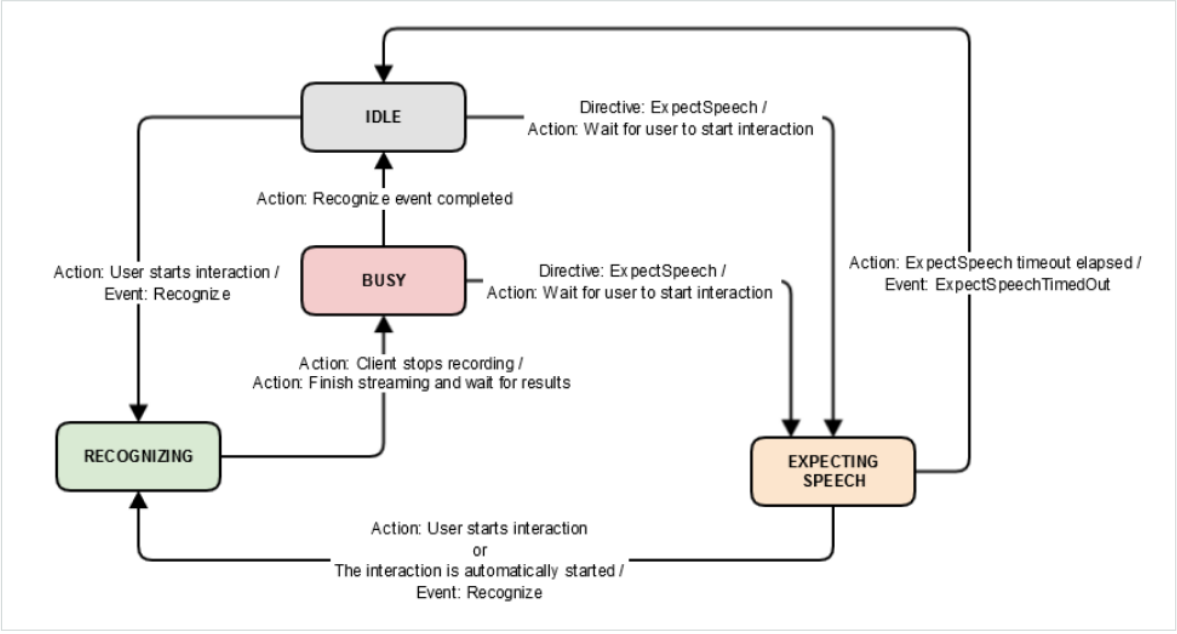
[250] *See* KAIFI0004342 *(How Our Scientists are Making Alexa Smarter* (Mar. 29, 2018), https://www.aboutamazon.com/news/devices/how-our-scientists-are-making-alexa-smarter).

[251] *See* KAIFI0007141 *(Wake Word Verification Requirements*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/implement-ww-verification.html).

[252] *See* KAIFI0007093 *(Understand the AVS SpeechRecognizer*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-

Ex.1029 / Page 293 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 17	Accused Instrumentality(ies)
	The following diagram illustrates the expected transitions among the four `SpeechRecognizer` states:  Amazon also provides wake word verification.[253]

service/avs-speechrecognizer-concepts.html).

[253] *See* KAIFI0002012 *(Enable Wake Word Verification*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/enable-cloud-based-wake-word-verification.html).

Ex.1029 / Page 294 of 347
AMAZON.COM, INC.

Claim 17	Accused Instrumentality(ies)

Cloud-based wake word verification improves wake word accuracy for Alexa Built-in products by reducing false wakes caused by words that sound similar to the wake word. Examples of words that might cause a false wake for "Alexa" include "Alex", "election", "Alexis". Cloud-based wake word verification also detects media mentions of the wake word, such as, the mention of "Alexa" in an Amazon commercial.

The wake word engine on a device performs the initial wake word detection, and then the cloud verifies the wake word. If the cloud detects a false wake, the Alexa Voice Service (AVS) sends a `StopCapture` directive to the device in the downchannel that instructs it to close the audio stream, and if applicable, to turn off the blue LEDs to indicate that Alexa has stopped listening.

Voice-initiated devices begin to stream user speech to AVS when the wake word engine detects a spoken wake word, such as "Alexa." The stream closes when the user stops speaking or when AVS identifies a user intent, and AVS returns a `StopCapture` directive to the device. Cloud-based wake word verification has the following requirements:

- **Wake word** – Include the wake word in the stream so that AVS can perform cloud-based wake word verification, which reduces false wakes. If AVS can't detect the wake word detected during cloud-based wake word verification, AVS discards the utterance.
- **500 milliseconds of pre-roll** – Pre-roll is the audio captured before AVS detects the wake word and helps calibrate the ambient noise level of the recording to enhance speech recognition.
- **User speech** – Any user speech that the device captures until receiving a `StopCapture` directive. This allows AVS to verify the wake word included in the stream, reducing the number of erroneous responses due to false wakes.

To learn how to implement a shared memory ring buffer for writing and reading audio samples and to include the start and stop indices for wake word detection in each `Recognize` event sent to AVS, see the Requirements for Cloud-Based Wake Word Verification.

For example, Amazon uses wake word, keyword spotting, cloud verification and encryption, automatic speech recognition (ASR), natural language understanding (NLU), text-to-speech (TTS), and data storage technologies. [254]

[254] *See* KAIFI0002023 *(Follow the journey of a voice request*, https://www.amazon.com/b/?node=23608618011).

Ex.1029 / Page 295 of 347
AMAZON.COM, INC.

The wake word

Think of the wake word as a verbal cue that makes things happen. In the classic tale from The Arabian Nights, an invisible door into a mountain can only be opened by saying the phrase "Open Sesame." Use the phrase, the door opens. Don't use the phrase, the door stays shut. When you have a request for Alexa, you first need to say your chosen wake word, which by default is "Alexa." Only after your Echo device detects the wake word is Alexa listening to your requests.

When it comes to privacy, there should be no surprises. You'll always be able to tell when Alexa is listening to your request because a light indicator will appear on your Echo device or an audible tone will sound. Think of the "On the Air" signs that light up in television studios during a broadcast. These indicators notify you that your device has detected the wake word and Alexa is now processing your request.

Keyword spotting

Echo devices use built-in technology called "keyword spotting" that matches spoken audio to the acoustic patterns of the wake word. Simply put, Echo devices are designed by default to detect only the sound waves of your chosen wake word, and everything else is ignored. Like water through a strainer, all other audio (people talking, faucets running, birds chirping) passes through the device until the wake word is "caught" and sent to Amazon's secure cloud, where meaning is assigned to your request. You can choose to opt-in to features like Alexa Emergency Assist that allow your device to detect more than your chosen wake word—for example, the sound of smoke alarms—but you would need to update your settings to do so.

Cloud verification and encryption

When your Echo device detects the wake word, it sends your request to Amazon's secure cloud where audio is reanalyzed to verify the wake word was spoken. If this cloud software verification is unable to confirm the wake word was spoken, the Alexa system stops processing the audio. If the wake word is verified (or Alexa is activated using the action button), your request will be processed using several sophisticated algorithms that allow Alexa to respond appropriately. All of your interactions with Alexa are encrypted in transit throughout this process.

Automatic Speech Recognition (ASR)

After wake word verification, your voice request undergoes Automatic Speech Recognition (ASR) which transcribes audio into text. Using our example request ("Alexa, what's the weather in Seattle?"), transcriptions could include one or several of the following:

"what is the weather in seattle" "watt is the weather in seattle" "what is the whether in seattle"

Alexa then chooses to act on the transcription that is most likely to be correct and stores it in Amazon's secure cloud. You can view and delete it at any time at Alexa Privacy Settings or in the Alexa app (Settings > Alexa Privacy > Review Voice History). Voice recordings and transcriptions may be used to improve the ASR process. Learn how data helps Alexa improve.

Natural Language Understanding

After ASR comes Natural Language Understanding (NLU). NLU assigns meaning to the transcription producing an "intent" which is an instruction to the Alexa system to tell it what to respond to. In our example, the intent is identified as "weather" and sent to the designated data source where information is pulled for location (Seattle) and time (today). Intents may also be used to improve the NLU process.

Text-to-Speech (TTS)

Once the necessary information is gathered, your request uses text-to-speech (TTS) technologies to generate an audio file that becomes Alexa's response. The text from Alexa's response is stored so that you can review past answers at Alexa Privacy Settings or in the Alexa app. These responses may also be reviewed by Amazon to ensure the device is providing the most relevant answers to you, and to make sure the TTS system is translating text to speech to the best of its ability.

Data storage

You have control over your Alexa experience. We make it easy for you to update your privacy settings. You can always view and manage your Alexa interactions at Alexa Privacy Settings or in the Alexa app (Settings > Alexa Privacy). If you're enrolled in voice ID, you can ask, "Alexa, what are my privacy settings?" to review how long your voice recordings are currently saved and whether or not they're used to develop new features or manually reviewed to help improve Alexa. You can also say, "Alexa, update my privacy settings" at any time to make updates.

Ex.1029 / Page 296 of 347
AMAZON.COM, INC.

Claim 17	Accused Instrumentality(ies)
	This claim element is literally present, as described herein, but is also satisfied under the doctrine of equivalents because the structure and functionality provided by Amazon is being used to satisfy this claim limitation and therefore there is no vitiation of this element, and the Accused Instrumentalities incorporate structure and functionality that is not substantially different from the requirements of this limitation because they achieve substantially or exactly the same function, in substantially or exactly the same way as set forth in the claim element (as described herein), and achieves substantially or exactly the same result.
determining, by the preset call command recognition unit of the single device, whether the input speech is recognized as the preset call command according to a result of the comparison; and	Amazon determines, by the preset call command recognition unit of the single device, whether the input speech is recognized as the preset call command according to a result of the comparison. For example, Amazon describes a voice request.[255]

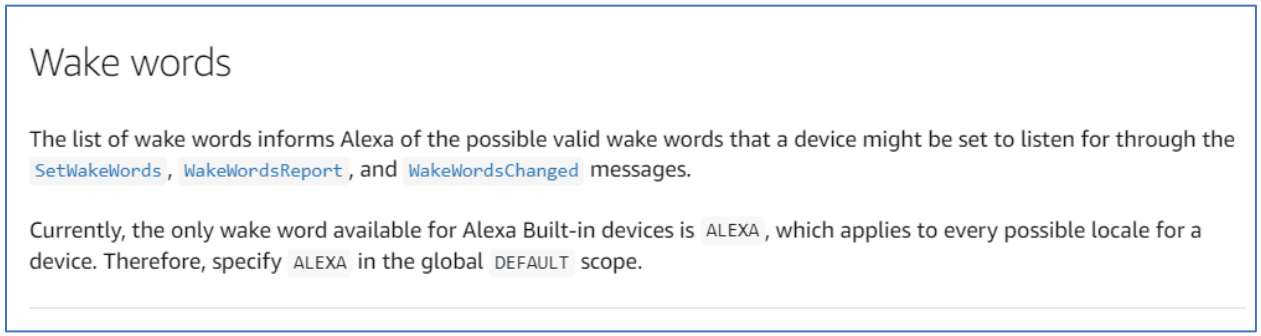
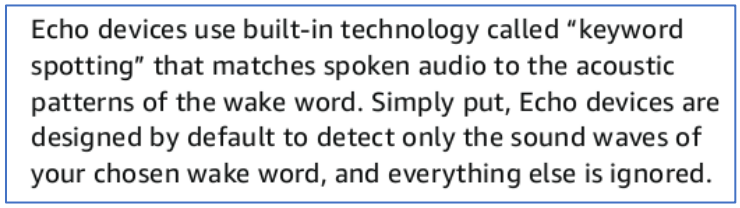
[255] *See* KAIFI0002023 *(Follow the journey of a voice request*, https://www.amazon.com/b/?node=23608618011).

Ex.1029 / Page 297 of 347
AMAZON.COM, INC.

Ex.1029 / Page 298 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 17	Accused Instrumentality(ies)
	For example, Amazon uses wake words.[256]  Wake words The list of wake words informs Alexa of the possible valid wake words that a device might be set to listen for through the `SetWakeWords`, `WakeWordsReport`, and `WakeWordsChanged` messages. Currently, the only wake word available for Alexa Built-in devices is `ALEXA`, which applies to every possible locale for a device. Therefore, specify `ALEXA` in the global `DEFAULT` scope. For example, Amazon uses keyword spotting.[257]  Echo devices use built-in technology called "keyword spotting" that matches spoken audio to the acoustic patterns of the wake word. Simply put, Echo devices are designed by default to detect only the sound waves of your chosen wake word, and everything else is ignored. For example, Amazon uses speech recognition.[258]

[256] *See* KAIFI0004342 *(How Our Scientists are Making Alexa Smarter* (Mar. 29, 2018), https://www.aboutamazon.com/news/devices/how-our-scientists-are-making-alexa-smarter).

[257] *See* KAIFI0002092 *(How Alexa works: Wake word*, https://www.amazon.com/b?ie=UTF8&node=23608571011).

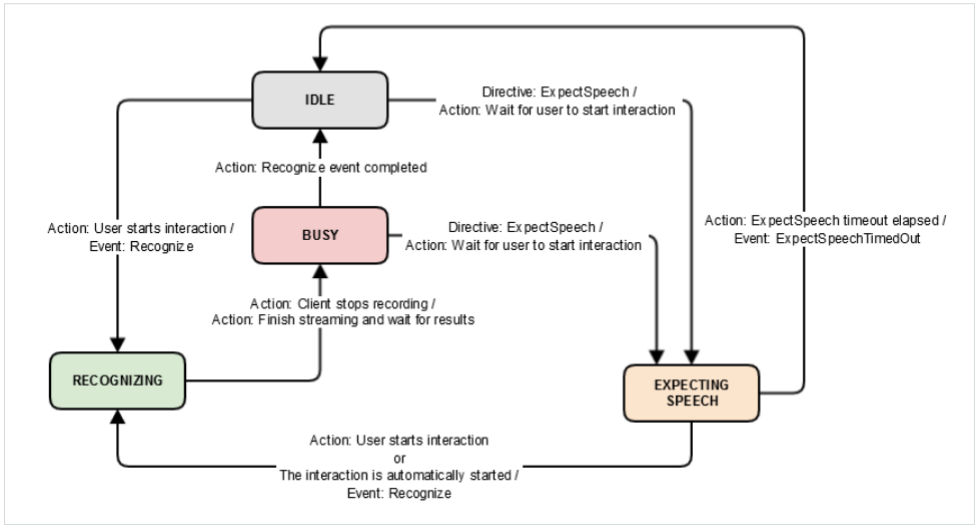
[258] *See* KAIFI0002023 *(Follow the journey of a voice request*, https://www.amazon.com/b/?node=23608618011).

Ex.1029 / Page 299 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 17	Accused Instrumentality(ies)

The following diagram illustrates the expected transitions among the four `SpeechRecognizer` states:



For example, Amazon provides wake word verification.[259]

> Cloud-based wake word verification improves wake word accuracy for Alexa Built-in products by reducing false wakes caused by words that sound similar to the wake word. Examples of words that might cause a false wake for "Alexa" include "Alex", "election", "Alexis". Cloud-based wake word verification also detects media mentions of the wake word, such as, the mention of "Alexa" in an Amazon commercial.

[259] *See* KAIFI0002012 *(Enable Wake Word Verification*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/enable-cloud-based-wake-word-verification.html).

Ex.1029 / Page 300 of 347
AMAZON.COM, INC.

Claim 17	Accused Instrumentality(ies)
	The wake word engine on a device performs the initial wake word detection, and then the cloud verifies the wake word. If the cloud detects a false wake, the Alexa Voice Service (AVS) sends a `StopCapture` directive to the device in the downchannel that instructs it to close the audio stream, and if applicable, to turn off the blue LEDs to indicate that Alexa has stopped listening. Voice-initiated devices begin to stream user speech to AVS when the wake word engine detects a spoken wake word, such as "Alexa." The stream closes when the user stops speaking or when AVS identifies a user intent, and AVS returns a `StopCapture` directive to the device. Cloud-based wake word verification has the following requirements: • **Wake word** – Include the wake word in the stream so that AVS can perform cloud-based wake word verification, which reduces false wakes. If AVS can't detect the wake word detected during cloud-based wake word verification, AVS discards the utterance. • **500 milliseconds of pre-roll** – Pre-roll is the audio captured before AVS detects the wake word and helps calibrate the ambient noise level of the recording to enhance speech recognition. • **User speech** – Any user speech that the device captures until receiving a `StopCapture` directive. This allows AVS to verify the wake word included in the stream, reducing the number of erroneous responses due to false wakes. To learn how to implement a shared memory ring buffer for writing and reading audio samples and to include the start and stop indices for wake word detection in each `Recognize` event sent to AVS, see the Requirements for Cloud-Based Wake Word Verification. For example, Amazon uses cloud verification and encryption, automatic speech recognition (ASR), natural language understanding (NLU), text-to-speech (TTS), and data storage technologies.[260]

[260] *See* KAIFI0002023 *(Follow the journey of a voice request*, https://www.amazon.com/b/?node=23608618011).

Ex.1029 / Page 301 of 347
AMAZON.COM, INC.

The wake word

Think of the wake word as a verbal cue that makes things happen. In the classic tale from The Arabian Nights, an invisible door into a mountain can only be opened by saying the phrase "Open Sesame." Use the phrase, the door opens. Don't use the phrase, the door stays shut. When you have a request for Alexa, you first need to say your chosen wake word, which by default is "Alexa." Only after your Echo device detects the wake word is Alexa listening to your requests.

When it comes to privacy, there should be no surprises. You'll always be able to tell when Alexa is listening to your request because a light indicator will appear on your Echo device or an audible tone will sound. Think of the "On the Air" signs that light up in television studios during a broadcast. These indicators notify you that your device has detected the wake word and Alexa is now processing your request.

Keyword spotting

Echo devices use built-in technology called "keyword spotting" that matches spoken audio to the acoustic patterns of the wake word. Simply put, Echo devices are designed by default to detect only the sound waves of your chosen wake word, and everything else is ignored. Like water through a strainer, all other audio (people talking, faucets running, birds chirping) passes through the device until the wake word is "caught" and sent to Amazon's secure cloud, where meaning is assigned to your request. You can choose to opt-in to features like Alexa Emergency Assist that allow your device to detect more than your chosen wake word—for example, the sound of smoke alarms—but you would need to update your settings to do so.

Cloud verification and encryption

When your Echo device detects the wake word, it sends your request to Amazon's secure cloud where audio is reanalyzed to verify the wake word was spoken. If this cloud software verification is unable to confirm the wake word was spoken, the Alexa system stops processing the audio. If the wake word is verified (or Alexa is activated using the action button), your request will be processed using several sophisticated algorithms that allow Alexa to respond appropriately. All of your interactions with Alexa are encrypted in transit throughout this process.

Automatic Speech Recognition (ASR)

After wake word verification, your voice request undergoes Automatic Speech Recognition (ASR) which transcribes audio into text. Using our example request ("Alexa, what's the weather in Seattle?"), transcriptions could include one or several of the following:

"what is the weather in seattle" "watt is the weather in seattle" "what is the whether in seattle"

Alexa then chooses to act on the transcription that is most likely to be correct and stores it in Amazon's secure cloud. You can view and delete it at any time at Alexa Privacy Settings or in the Alexa app (Settings > Alexa Privacy > Review Voice History). Voice recordings and transcriptions may be used to improve the ASR process. Learn how data helps Alexa improve.

Natural Language Understanding

After ASR comes Natural Language Understanding (NLU). NLU assigns meaning to the transcription producing an "intent" which is an instruction to the Alexa system to tell it what to respond to. In our example, the intent is identified as "weather" and sent to the designated data source where information is pulled for location (Seattle) and time (today). Intents may also be used to improve the NLU process.

Text-to-Speech (TTS)

Once the necessary information is gathered, your request uses text-to-speech (TTS) technologies to generate an audio file that becomes Alexa's response. The text from Alexa's response is stored so that you can review past answers at Alexa Privacy Settings or in the Alexa app. These responses may also be reviewed by Amazon to ensure the device is providing the most relevant answers to you, and to make sure the TTS system is translating text to speech to the best of its ability.

Data storage

You have control over your Alexa experience. We make it easy for you to update your privacy settings. You can always view and manage your Alexa interactions at Alexa Privacy Settings or in the Alexa app (Settings > Alexa Privacy). If you're enrolled in voice ID, you can ask, "Alexa, what are my privacy settings?" to review how long your voice recordings are currently saved and whether or not they're used to develop new features or manually reviewed to help improve Alexa. You can also say, "Alexa, update my privacy settings" at any time to make updates.

Ex.1029 / Page 302 of 347
AMAZON.COM, INC.

Claim 17	Accused Instrumentality(ies)
	This claim element is literally present, as described herein, but is also satisfied under the doctrine of equivalents because the structure and functionality provided by Amazon is being used to satisfy this claim limitation and therefore there is no vitiation of this element, and the Accused Instrumentalities incorporate structure and functionality that is not substantially different from the requirements of this limitation because they achieve substantially or exactly the same function, in substantially or exactly the same way as set forth in the claim element (as described herein), and achieves substantially or exactly the same result.
only if the input speech is recognized as the preset call command, activating a continuous speech recognition unit of the single device that is configured to perform continuous speech recognition to determine whether the subsequent input speech is recognized as an actual command, in order to reduce overall computational load on the single device; and	Amazon activates, only if the input speech is recognized as the preset call command, a continuous speech recognition unit of the single device that is configured to perform continuous speech recognition to determine whether the subsequent input speech is recognized as an actual command, in order to reduce overall computational load on the single device. For example, Amazon describes a voice request.[261]

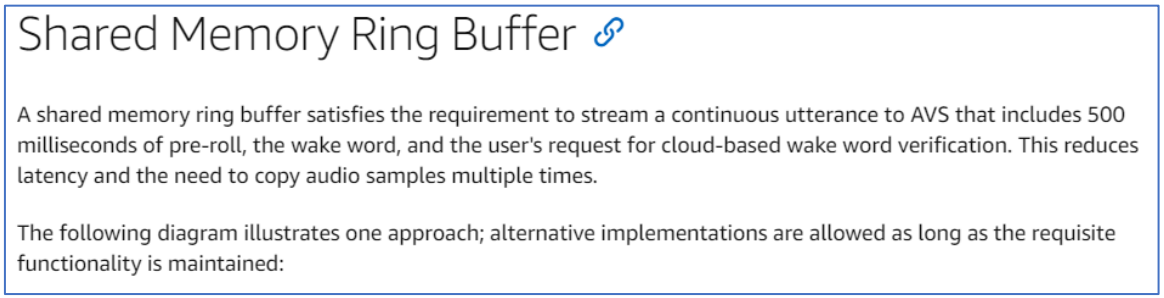
[261] *See* KAIFI0002023 *(Follow the journey of a voice request*, https://www.amazon.com/b/?node=23608618011).

Ex.1029 / Page 303 of 347
AMAZON.COM, INC.

Ex.1029 / Page 304 of 347
AMAZON.COM, INC.

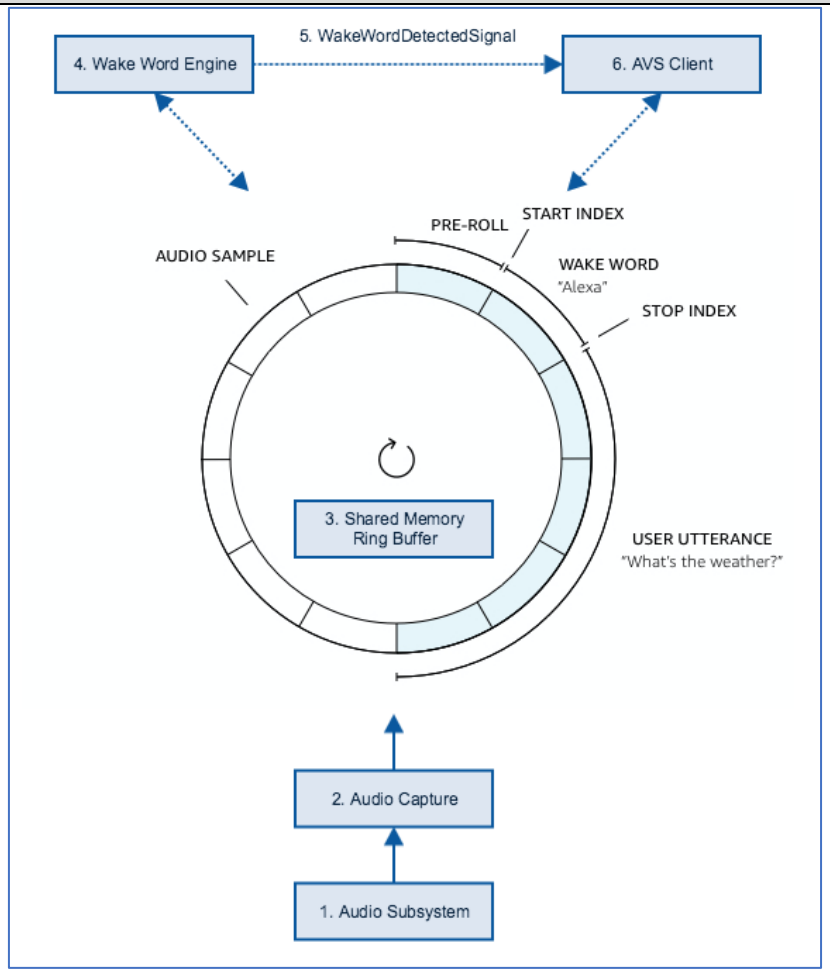
Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 17	Accused Instrumentality(ies)
	For example, Amazon provides a shared memory ring buffer.[262]  Shared Memory Ring Buffer A shared memory ring buffer satisfies the requirement to stream a continuous utterance to AVS that includes 500 milliseconds of pre-roll, the wake word, and the user's request for cloud-based wake word verification. This reduces latency and the need to copy audio samples multiple times. The following diagram illustrates one approach; alternative implementations are allowed as long as the requisite functionality is maintained:

[262] *See* KAIFI0007141 *(Wake Word Verification Requirements*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/implement-ww-verification.html).

Ex.1029 / Page 305 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 17	Accused Instrumentality(ies)
	

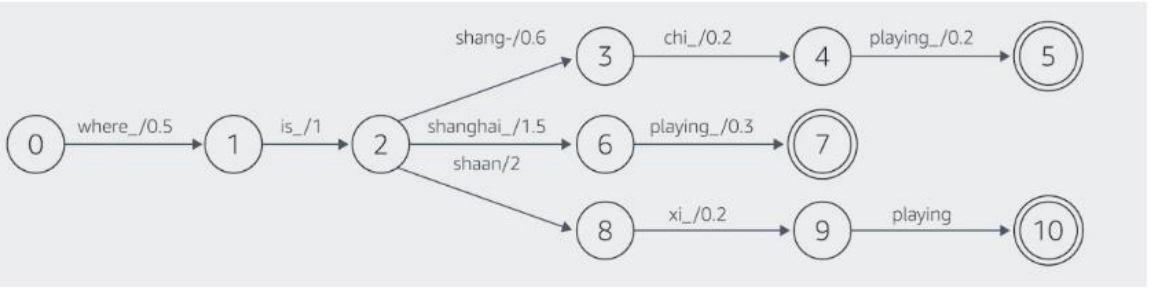
Ex.1029 / Page 306 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 17	Accused Instrumentality(ies)
	1. **Audio Subsystem** - Middleware, such as Advanced Linux Sound Architecture (ALSA), that provides APIs to open a stream from a recording device, and receive an audio sample stream. 2. **Audio Capture** - A process that opens a recording device, and writes audio samples into the shared memory ring buffer. 3. **Shared Memory Ring Buffer** - A shared memory ring buffer object that is thread safe, and accessible by multiple processes. Access to this memory block should support at least one writer and two readers simultaneously. You may choose to implement these executables as a shared library. • **Write API** - Used by the recording process to write audio samples to the shared memory ring buffer. Write samples should move data so that the oldest audio samples are overwritten. This example shows a typical API signature: `WriteSamples(BufferIn, BufferInLength)`. • **Read API** - Used by the wake word engine and the AVS client to read from specific locations in the shared memory ring buffer. A typical implementation represents location as a 64-bit integer. For example, reading 200 samples from the location 100 is reading the 100th to 300th samples that were written to the shared memory ring buffer. This is an example of a typical API signature: `ReadSamples(BufferIn, BufferInLength)`. **Note**: Reading from an old location that was overwritten with new audio samples will result in an unintended sample read. 4. **Wake Word Engine** - The component that reads the audio samples written to the shared memory ring buffer and analyzes the audio samples for occurrences of the wake word. The procedure of reading and processing audio samples occurs in a loop. For example, the loop reads N samples starting at location `0`; then analyzes the samples for an occurrence of the wake word. In the next iteration, it reads and analyzes the next N samples. When an occurrence of the wake word is detected, the wake word engine must identify two locations: 1) the start of the wake word, and 2) the end of the wake word. These values are stored as `startIndexInSamples` and `endIndexInSamples`. 5. **WakeWordDetectedSignal** - A signal sent from the wake word engine to the AVS client when an occurrence of the wake word is detected. This signal includes two locations, the `startIndexInSamples` and `endIndexInSamples`. 6. **AVS Client** - The client receives the `WakeWordDetectedSignal`, extracts the `startIndexInSamples`, then reads and streams audio samples to AVS including 500 milliseconds of pre-roll. • The pre-roll is the number of samples corresponding to 500 milliseconds. For example, if the audio samples were recorded at a rate of 16 kilohertz (kHz), then pre-roll is 8,000 samples.

Ex.1029 / Page 307 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 17	Accused Instrumentality(ies)
	For example, Amazon describes its ASR model.[263] **System architecture** Our on-device ASR model takes in an acoustic speech signal and outputs a set of hypotheses about what the speaker said, ranked according to probability. We represent those hypotheses as a *lattice* — a graph whose edges represent recognized words and the probability that a given word follows from the previous one.  An example of a lattice representing ASR hypotheses.

[263] *See* KAIFI0006791 (Ariya Rastrow, *On-device speech processing makes Alexa faster, lower-bandwidth* (Jan. 25, 2022), https://www.amazon.science/blog/on-device-speech-processing-makes-alexa-faster-lower-bandwidth).

Ex.1029 / Page 308 of 347
AMAZON.COM, INC.

Claim 17	Accused Instrumentality(ies)
	For example, Amazon uses cloud verification and encryption, automatic speech recognition (ASR), natural language understanding (NLU), text-to-speech (TTS), and data storage technologies.[264]

[264] *See* KAIFI0002023 *(Follow the journey of a voice request*, https://www.amazon.com/b/?node=23608618011).

Ex.1029 / Page 309 of 347
AMAZON.COM, INC.

The wake word

Think of the wake word as a verbal cue that makes things happen. In the classic tale from The Arabian Nights, an invisible door into a mountain can only be opened by saying the phrase "Open Sesame." Use the phrase, the door opens. Don't use the phrase, the door stays shut. When you have a request for Alexa, you first need to say your chosen wake word, which by default is "Alexa." Only after your Echo device detects the wake word is Alexa listening to your requests.

When it comes to privacy, there should be no surprises. You'll always be able to tell when Alexa is listening to your request because a light indicator will appear on your Echo device or an audible tone will sound. Think of the "On the Air" signs that light up in television studios during a broadcast. These indicators notify you that your device has detected the wake word and Alexa is now processing your request.

Keyword spotting

Echo devices use built-in technology called "keyword spotting" that matches spoken audio to the acoustic patterns of the wake word. Simply put, Echo devices are designed by default to detect only the sound waves of your chosen wake word, and everything else is ignored. Like water through a strainer, all other audio (people talking, faucets running, birds chirping) passes through the device until the wake word is "caught" and sent to Amazon's secure cloud, where meaning is assigned to your request. You can choose to opt-in to features like Alexa Emergency Assist that allow your device to detect more than your chosen wake word—for example, the sound of smoke alarms—but you would need to update your settings to do so.

Cloud verification and encryption

When your Echo device detects the wake word, it sends your request to Amazon's secure cloud where audio is reanalyzed to verify the wake word was spoken. If this cloud software verification is unable to confirm the wake word was spoken, the Alexa system stops processing the audio. If the wake word is verified (or Alexa is activated using the action button), your request will be processed using several sophisticated algorithms that allow Alexa to respond appropriately. All of your interactions with Alexa are encrypted in transit throughout this process.

Automatic Speech Recognition (ASR)

After wake word verification, your voice request undergoes Automatic Speech Recognition (ASR) which transcribes audio into text. Using our example request ("Alexa, what's the weather in Seattle?"), transcriptions could include one or several of the following:

"what is the weather in seattle" "watt is the weather in seattle" "what is the whether in seattle"

Alexa then chooses to act on the transcription that is most likely to be correct and stores it in Amazon's secure cloud. You can view and delete it at any time at Alexa Privacy Settings or in the Alexa app (Settings > Alexa Privacy > Review Voice History). Voice recordings and transcriptions may be used to improve the ASR process. Learn how data helps Alexa improve.

Natural Language Understanding

After ASR comes Natural Language Understanding (NLU). NLU assigns meaning to the transcription producing an "intent" which is an instruction to the Alexa system to tell it what to respond to. In our example, the intent is identified as "weather" and sent to the designated data source where information is pulled for location (Seattle) and time (today). Intents may also be used to improve the NLU process.

Text-to-Speech (TTS)

Once the necessary information is gathered, your request uses text-to-speech (TTS) technologies to generate an audio file that becomes Alexa's response. The text from Alexa's response is stored so that you can review past answers at Alexa Privacy Settings or in the Alexa app. These responses may also be reviewed by Amazon to ensure the device is providing the most relevant answers to you, and to make sure the TTS system is translating text to speech to the best of its ability.

Data storage

You have control over your Alexa experience. We make it easy for you to update your privacy settings. You can always view and manage your Alexa interactions at Alexa Privacy Settings or in the Alexa app (Settings > Alexa Privacy). If you're enrolled in voice ID, you can ask, "Alexa, what are my privacy settings?" to review how long your voice recordings are currently saved and whether or not they're used to develop new features or manually reviewed to help improve Alexa. You can also say, "Alexa, update my privacy settings" at any time to make updates.

Ex.1029 / Page 310 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 17	Accused Instrumentality(ies)
	For example, Amazon uses a speech recognizer, including a "Recognize" state and an "ExpectSpeech" state.[265]

The `SpeechRecognizer` interface is the core interface of the Alexa Voice Service (AVS) and exposes directives and events for capturing and interacting with user speech. This page discusses the concepts and process flows for the `SpeechRecognizer` interface.

SpeechRecognizer functionality

Every user utterance leverages `SpeechRecognizer`, including the following interaction types between an Alexa Built-in device and AVS:

- **User Speech Capture**: Captures user speech from an Alexa Built-in device.
- **User Speech Prompting**: Prompts a user for more speech input when needed by Alexa to deliver an appropriate response.
- **Interaction Initiation Communication**: Enables a device to inform AVS of how a user initiated an Alexa interaction, such as press-and-hold, tap-and-release, or voice-initiated/wake word enabled. See Device Form Factor and Alexa Interaction.
- **ASR Profile Selection**: SpeechRecognizers chooses the appropriate Automatic Speech Recognition (ASR) profile for your product, which allows Alexa to understand user speech and respond with precision. See Automatic Speech Recognition (ASR) profile

[265] *See* KAIFI0007093 *(Understand the AVS SpeechRecognizer*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/avs-speechrecognizer-concepts.html).

Ex.1029 / Page 311 of 347
AMAZON.COM, INC.

State diagram

The following diagram illustrates state changes driven by `SpeechRecognizer` components. Boxes represent `SpeechRecognizer` states, and the connectors represent state transitions.

`SpeechRecognizer` has the following states:

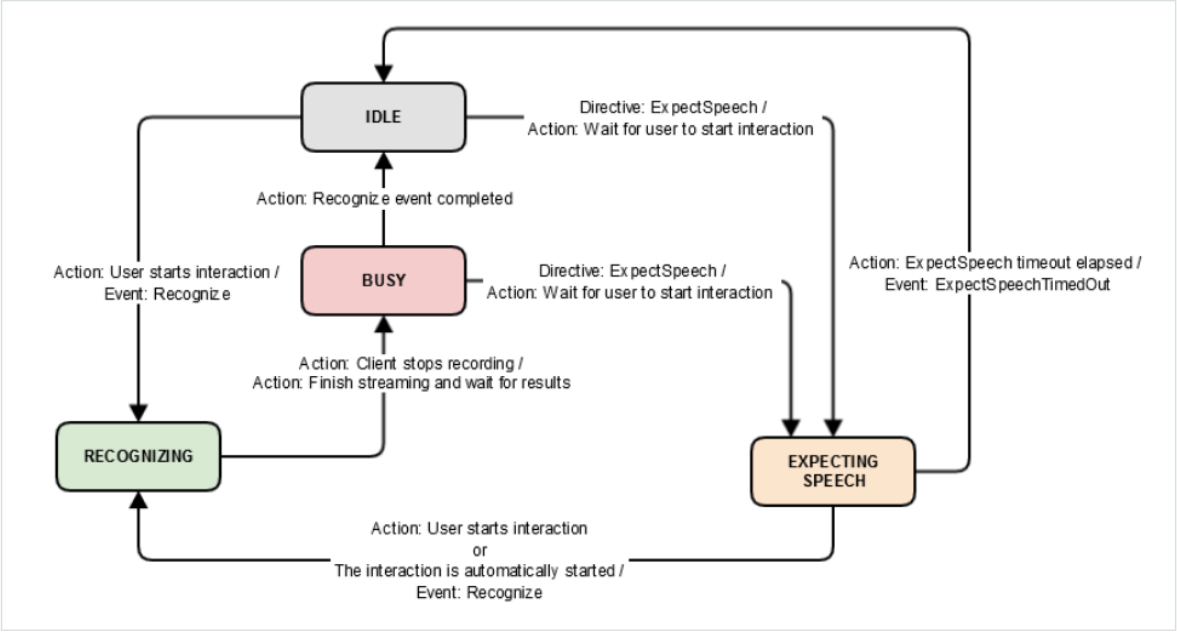
- **IDLE:** When not actively processing speech, the `SpeechRecognizer` is in an "idle" state. The idle state occurs under the following conditions:
  - Before capturing user speech.
  - Returning to an idle state after concluding a speech interaction with Alexa.
  - When an `ExpectSpeechTimedOut` event elapses.

  Note: In a multi-turn Alexa interaction, if Alexa requires more user speech input, `SpeechRecognizer` should transition from the idle state to the `ExpectSpeech` state without the user starting a new interaction.

- **RECOGNIZING:** When a user begins interacting with your client, specifically when the client streams captured audio to AVS, `SpeechRecognizer` should transition from the idle state to the `Recognize` state. `SpeechRecognizer` should remain in the `Recognize` until the client stops recording speech or finishes streaming, at which point your `SpeechRecognizer` component should transition from the `Recognize` state to the "busy" state.
- **BUSY:** While processing the speech request, `SpeechRecognizer` should be in the "busy" state. You cannot start another speech request until `SpeechRecognizer` transitions out of the busy state. From the *busy* state, `SpeechRecognizer` transitions to the idle state if Alexa processes and completes the request or to the `ExpectSpeech` state, if Alexa requires more speech input from the user.
- **EXPECTING SPEECH:** `SpeechRecognizer` should be in the `ExpectSpeech` state when Alexa requires more audio input from a user. From `ExpectSpeech`, `SpeechRecognizer` should transition to the `Recognize` when a user interaction occurs, or the interaction is automatically started on behalf of the user. It should transition to the idle state if Alexa detects no user interaction within the specified timeout window.

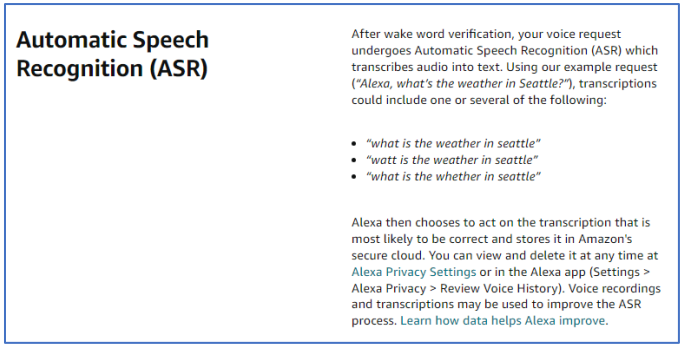
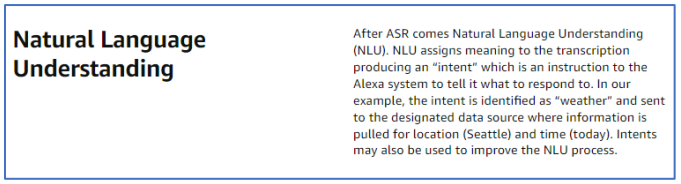
Ex.1029 / Page 312 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 17	Accused Instrumentality(ies)
	The following diagram illustrates the expected transitions among the four `SpeechRecognizer` states:  For example, Amazon uses a wake word engine.[266]

[266] *See* KAIFI0007141 *(Wake Word Verification Requirements*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/implement-ww-verification.html).

Ex.1029 / Page 313 of 347
AMAZON.COM, INC.

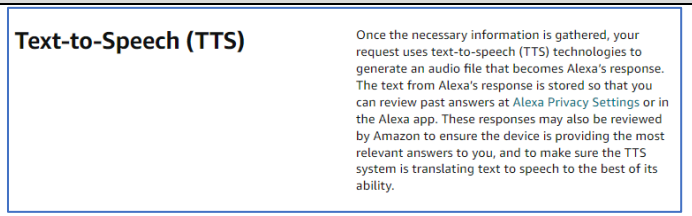
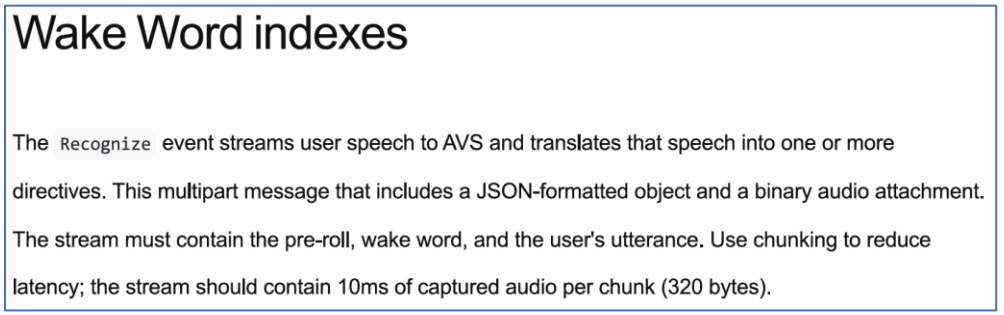
Claim 17	Accused Instrumentality(ies)
	For example, Amazon provides wake word verification.[267] For example, Amazon uses cloud verification and encryption, automatic speech recognition (ASR), natural language understanding (NLU), text-to-speech (TTS), and data storage technologies on the recognized actual subsequent input speech command.[268]  

[267] *See* KAIFI0002012 *(Enable Wake Word Verification*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/enable-cloud-based-wake-word-verification.html).

[268] *See* KAIFI0002023 *(Follow the journey of a voice request*, https://www.amazon.com/b/?node=23608618011).

Ex.1029 / Page 314 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 17	Accused Instrumentality(ies)
	 For example, Amazon uses chunking to reduce latency.[269]  For example, Amazon Alexa's wake word detection on the device works with limited CPU power.[270]

[269] *See* KAIFI0007141 *(Wake Word Verification Requirements*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/implement-ww-verification.html).
[270] *See* KAIFI0002097 (How Amazon Alexa works? Your Guide to Natural Language Processing (AI), https://towardsdatascience.com/how-amazon-alexa-works-your-guide-to-natural-language-processing-ai-7506004709d3).

Ex.1029 / Page 315 of 347
AMAZON.COM, INC.

Claim 17	Accused Instrumentality(ies)
	> The next task is "Wake Word Detection". It determines whether the user says one of the words the device is programmed to need to turn on, such as "Alexa". This is needed to minimize false positives and false negatives, which could lead to accidental purchases and angry customers. This is really complicated as it needs to identify pronunciation differences, and it needs to do so on the device, which has limited CPU power. For example, when a customer says the wake word, subsequent phrases are processed and stored by Amazon in the cloud to respond to the customer's request.[271] > Alexa is designed to get smarter every day—this is accomplished through the power of machine learning and the cloud. The more a customer uses Alexa, the more it adapts to their speech patterns, vocabulary, and personal preferences. When the customer says the wake word, their subsequent phrases are processed and stored in the cloud to respond to the customer's request and to improve the customer's experience and our services, including training our speech recognition and natural language understanding systems so Alexa can better understand customers' requests. This claim element is literally present, as described herein, but is also satisfied under the doctrine of equivalents because the structure and functionality provided by Amazon is being used to satisfy this claim limitation and therefore there is no vitiation of this element, and the Accused Instrumentalities incorporate structure and functionality that is not substantially different from the requirements of this limitation because they achieve substantially or exactly the same function, in substantially or exactly the same way as set forth in the claim element (as described herein), and achieves substantially or exactly the same result.

[271] *See* KAIFI0007380 *(White paper – Alexa Privacy and Data Handling Overview*, https://d1.awsstatic.com/product-marketing/A4B/White%20Paper%20-%20Alexa%20Privacy%20and%20Data%20Handling%20Overview.pdf).

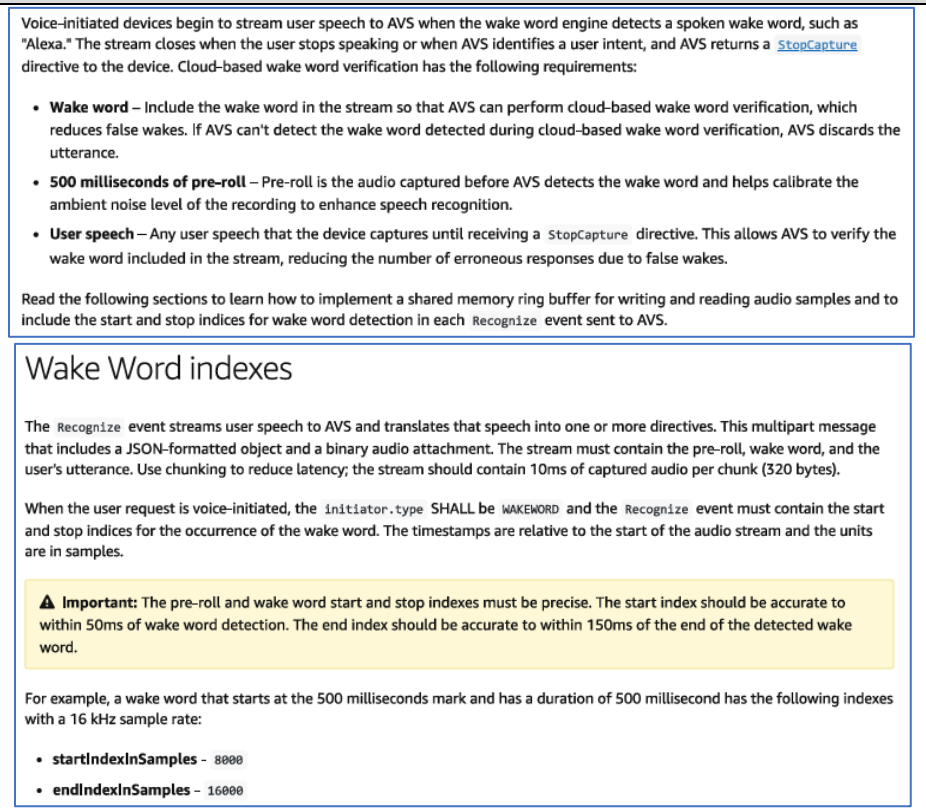
Ex.1029 / Page 316 of 347
AMAZON.COM, INC.

Claim 17	Accused Instrumentality(ies)
wherein the preset call command recognition unit comprises a token passing unit configured to assume the input speech to be the preset call command according to the result of the comparison using a minimum recognition network composed of a silence interval accompanied by noise and the preset call command.	Amazon's preset call command recognition unit comprises a token passing unit configured to assume the input speech to be the preset call command according to the result of the comparison using a minimum recognition network composed of a silence interval accompanied by noise and the preset call command. For example, Amazon uses a user speech stream.[272]

[272] *See* KAIFI0007141 *(Wake Word Verification Requirements*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/implement-ww-verification.html).

Ex.1029 / Page 317 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 17	Accused Instrumentality(ies)
	

Voice-initiated devices begin to stream user speech to AVS when the wake word engine detects a spoken wake word, such as "Alexa." The stream closes when the user stops speaking or when AVS identifies a user intent, and AVS returns a `StopCapture` directive to the device. Cloud-based wake word verification has the following requirements:

- **Wake word** – Include the wake word in the stream so that AVS can perform cloud-based wake word verification, which reduces false wakes. If AVS can't detect the wake word detected during cloud-based wake word verification, AVS discards the utterance.
- **500 milliseconds of pre-roll** – Pre-roll is the audio captured before AVS detects the wake word and helps calibrate the ambient noise level of the recording to enhance speech recognition.
- **User speech** – Any user speech that the device captures until receiving a `StopCapture` directive. This allows AVS to verify the wake word included in the stream, reducing the number of erroneous responses due to false wakes.

Read the following sections to learn how to implement a shared memory ring buffer for writing and reading audio samples and to include the start and stop indices for wake word detection in each `Recognize` event sent to AVS.

Wake Word indexes

The `Recognize` event streams user speech to AVS and translates that speech into one or more directives. This multipart message that includes a JSON-formatted object and a binary audio attachment. The stream must contain the pre-roll, wake word, and the user's utterance. Use chunking to reduce latency; the stream should contain 10ms of captured audio per chunk (320 bytes).

When the user request is voice-initiated, the `initiator.type` SHALL be `WAKEWORD` and the `Recognize` event must contain the start and stop indices for the occurrence of the wake word. The timestamps are relative to the start of the audio stream and the units are in samples.

⚠ **Important:** The pre-roll and wake word start and stop indexes must be precise. The start index should be accurate to within 50ms of wake word detection. The end index should be accurate to within 150ms of the end of the detected wake word.

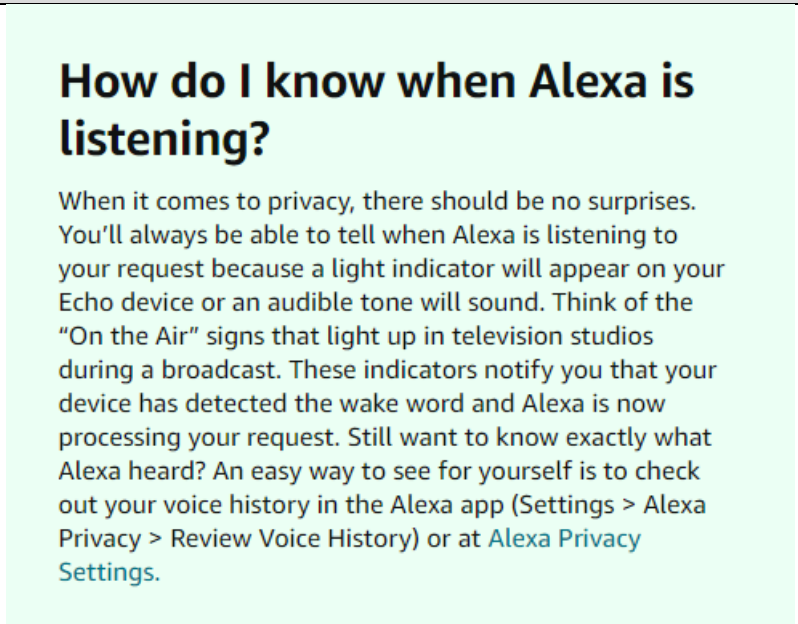
For example, a wake word that starts at the 500 milliseconds mark and has a duration of 500 millisecond has the following indexes with a 16 kHz sample rate:

- **startIndexInSamples** - `8000`
- **endIndexInSamples** - `16000`

For example, Amazon's Alexa indicators notify that a device has detected the wake word and is now processing a request. [273]

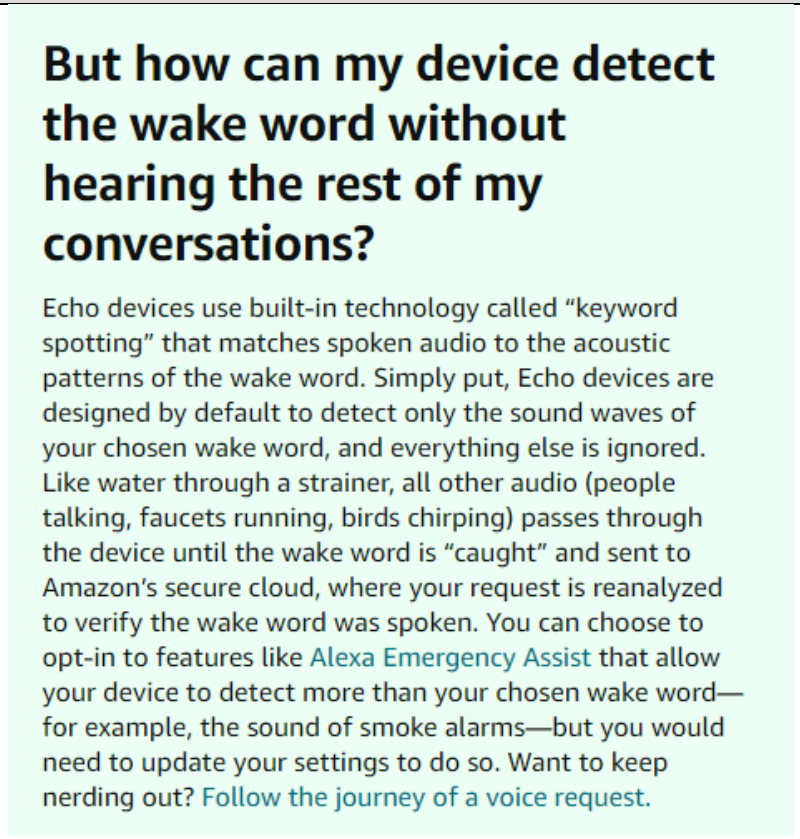
[273] *See* KAIFI0002092 *(How Alexa works: Wake word*, https://www.amazon.com/b?ie=UTF8&node=23608571011).

Ex.1029 / Page 318 of 347
AMAZON.COM, INC.

Claim 17	Accused Instrumentality(ies)
	 For example, Amazon detects only the sound waves of a chosen wake word, and everything else is ignored.[274]

[274] *See id.*

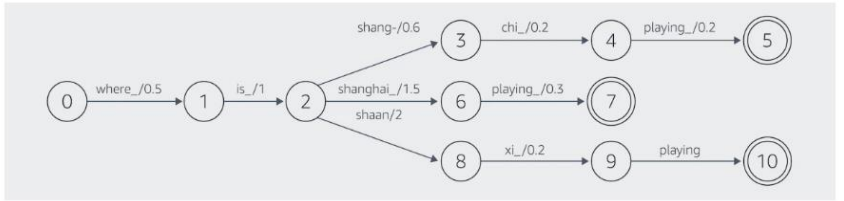
Ex.1029 / Page 319 of 347
AMAZON.COM, INC.

Claim 17	Accused Instrumentality(ies)
	 For example, Amazon provides an ASR model.[275]

[275] *See* KAIFI0006791 (Ariya Rastrow, *On-device speech processing makes Alexa faster, lower-bandwidth* (Jan. 25, 2022), https://www.amazon.science/blog/on-device-speech-processing-makes-alexa-faster-lower-bandwidth).

Ex.1029 / Page 320 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 17	Accused Instrumentality(ies)
	System architecture Our on-device ASR model takes in an acoustic speech signal and outputs a set of hypotheses about what the speaker said, ranked according to probability. We represent those hypotheses as a *lattice* — a graph whose edges represent recognized words and the probability that a given word follows from the previous one.  An example of a lattice representing ASR hypotheses.

Ex.1029 / Page 321 of 347
AMAZON.COM, INC.

Claim 17	Accused Instrumentality(ies)
	With cloud-based ASR, encrypted audio streams to the cloud in small snippets called "frames". With on-device ASR, only the lattice is sent to the cloud, where a large and powerful neural language model reranks the hypotheses. The lattice can't be sent until the customer has finished speaking, as words later in a sequence can dramatically change the overall probability of a hypothesis. The model that determines when the customer has finished speaking is called an end-pointer. End-pointers offer a natural trade-off between accuracy and latency: an aggressive end-pointer will initiate speech processing earlier, but it might cut the speaker off prematurely, resulting in a poor customer experience. On the device, we in fact run two end-pointers: One is a speculative end-pointer that we have tuned to be about 200 milliseconds faster than the final end-pointer, so we can initiate downstream processing — such as natural-language understanding (NLU) — ahead of the final end-pointed ASR result. In exchange for speed, however, we trade off a little accuracy. The final end-pointer takes longer to make a decision but is more accurate. In cases in which the first end-pointer cuts speech off too early, the final end-pointer sends a revised lattice and the instruction to reset downstream processing. In the large majority of cases, however, the aggressive end-pointer is correct, which reduces user-perceived latency, since downstream tasks are initiated earlier. For example, Amazon uses a shared memory ring buffer.[276]

[276] *See* KAIFI0007141 *(Wake Word Verification Requirements*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/implement-ww-verification.html).

Ex.1029 / Page 322 of 347
AMAZON.COM, INC.

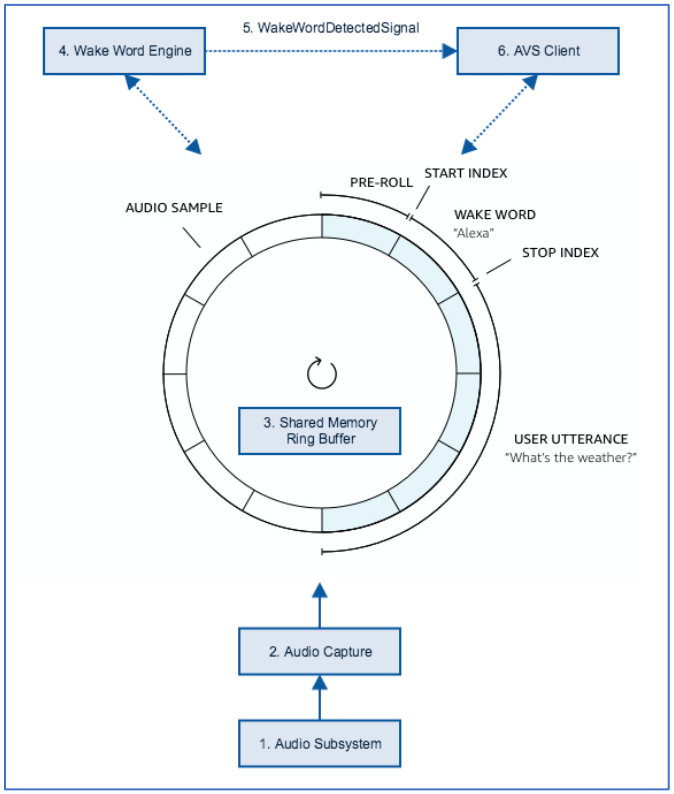
Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 17	Accused Instrumentality(ies)

Shared Memory Ring Buffer

A shared memory ring buffer satisfies the requirement to stream a continuous utterance to AVS that includes 500 milliseconds of pre-roll, the wake word, and the user's request for cloud-based wake word verification. This reduces latency and the need to copy audio samples multiple times.

The following diagram illustrates one approach; alternative implementations are allowed as long as the requisite functionality is maintained:



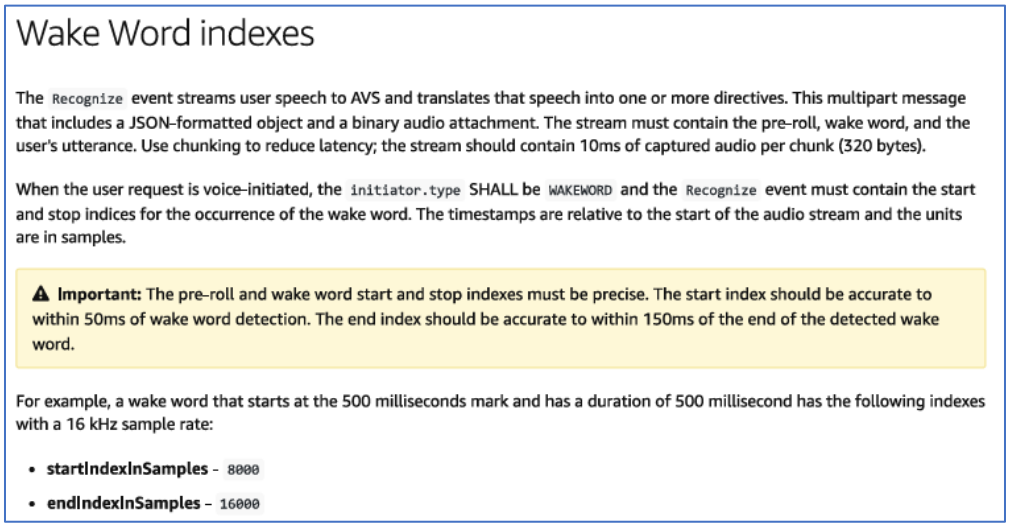
Ex.1029 / Page 323 of 347
AMAZON.COM, INC.

Claim 17	Accused Instrumentality(ies)

1. **Audio Subsystem** - Middleware, such as Advanced Linux Sound Architecture (ALSA), that provides APIs to open a stream from a recording device, and receive an audio sample stream.
2. **Audio Capture** - A process that opens a recording device, and writes audio samples into the shared memory ring buffer.
3. **Shared Memory Ring Buffer** - A shared memory ring buffer object that is thread safe, and accessible by multiple processes. Access to this memory block should support at least one writer and two readers simultaneously. You may choose to implement these executables as a shared library.
   - **Write API** - Used by the recording process to write audio samples to the shared memory ring buffer. Write samples should move data so that the oldest audio samples are overwritten. This example shows a typical API signature: `WriteSamples(BufferIn, BufferInLength)`.
   - **Read API** - Used by the wake word engine and the AVS client to read from specific locations in the shared memory ring buffer. A typical implementation represents location as a 64-bit integer. For example, reading 200 samples from the location 100 is reading the 100th to 300th samples that were written to the shared memory ring buffer. This is an example of a typical API signature: `ReadSamples(BufferIn, BufferInLength)`. **Note**: Reading from an old location that was overwritten with new audio samples will result in an unintended sample read.

4. **Wake Word Engine** - The component that reads the audio samples written to the shared memory ring buffer and analyzes the audio samples for occurrences of the wake word. The procedure of reading and processing audio samples occurs in a loop. For example, the loop reads N samples starting at location `0`; then analyzes the samples for an occurrence of the wake word. In the next iteration, it reads and analyzes the next N samples. When an occurrence of the wake word is detected, the wake word engine must identify two locations: 1) the start of the wake word, and 2) the end of the wake word. These values are stored as `startIndexInSamples` and `endIndexInSamples`.
5. **WakeWordDetectedSignal** - A signal sent from the wake word engine to the AVS client when an occurrence of the wake word is detected. This signal includes two locations, the `startIndexInSamples` and `endIndexInSamples`.
6. **AVS Client** - The client receives the `WakeWordDetectedSignal`, extracts the `startIndexInSamples`, then reads and streams audio samples to AVS including 500 milliseconds of pre-roll.
   - The pre-roll is the number of samples corresponding to 500 milliseconds. For example, if the audio samples were recorded at a rate of 16 kilohertz (kHz), then pre-roll is 8,000 samples.

Ex.1029 / Page 324 of 347
AMAZON.COM, INC.

Claim 17	Accused Instrumentality(ies)
	For example, Amazon uses wake word indexes.[277]  Wake Word indexes The `Recognize` event streams user speech to AVS and translates that speech into one or more directives. This multipart message that includes a JSON-formatted object and a binary audio attachment. The stream must contain the pre-roll, wake word, and the user's utterance. Use chunking to reduce latency; the stream should contain 10ms of captured audio per chunk (320 bytes). When the user request is voice-initiated, the `initiator.type` SHALL be `WAKEWORD` and the `Recognize` event must contain the start and stop indices for the occurrence of the wake word. The timestamps are relative to the start of the audio stream and the units are in samples. **Important:** The pre-roll and wake word start and stop indexes must be precise. The start index should be accurate to within 50ms of wake word detection. The end index should be accurate to within 150ms of the end of the detected wake word. For example, a wake word that starts at the 500 milliseconds mark and has a duration of 500 millisecond has the following indexes with a 16 kHz sample rate: • **startIndexInSamples** - `8000` • **endIndexInSamples** - `16000` For example, Amazon describes a voice request. [278]

[277] *See id.*

[278] *See* KAIFI0002023 *(Follow the journey of a voice request*, https://www.amazon.com/b/?node=23608618011).

Ex.1029 / Page 325 of 347
AMAZON.COM, INC.

Ex.1029 / Page 326 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 17	Accused Instrumentality(ies)
	For example, Amazon uses speech recognition.[279] Near-field and far-field tests are conducted in various locations, which mimic real-world user-device interaction patterns. The device's performance is evaluated under various conditions such as in silence, during audio playback from the device, or in environments with background noise. Such noise could be from music playing on another device, stationary noise such as an air conditioner, or other ambient noise. For example, Amazon uses a user speech stream.[280] Voice-initiated devices begin to stream user speech to AVS when the wake word engine detects a spoken wake word, such as "Alexa." The stream closes when the user stops speaking or when AVS identifies a user intent, and AVS returns a `StopCapture` directive to the device. Cloud-based wake word verification has the following requirements: For example, Amazon uses ASR.[281]

[279] *See* KAIFI0001353 (Felipe Pinzon, *An Amazon technical expert provides three acoustic certification-related tips for AVS device makers* (Aug. 16, 2023) https://developer.amazon.com/en-US/blogs/alexa/device-makers/2023/08/acoustic-certification-alexa-august-2023).

[280] *See* KAIFI0007141 *(Wake Word Verification Requirements*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/implement-ww-verification.html).

[281] *See* KAIFI0006791 (Ariya Rastrow, Shehzad Mevawalla, *On-device speech processing makes Alexa faster, lower-bandwidth*, AMAZON SCIENCE (Jan. 25, 2022), https://www.amazon.science/blog/on-device-speech-processing-makes-alexa-faster-lower-bandwidth).

Ex.1029 / Page 327 of 347
AMAZON.COM, INC.

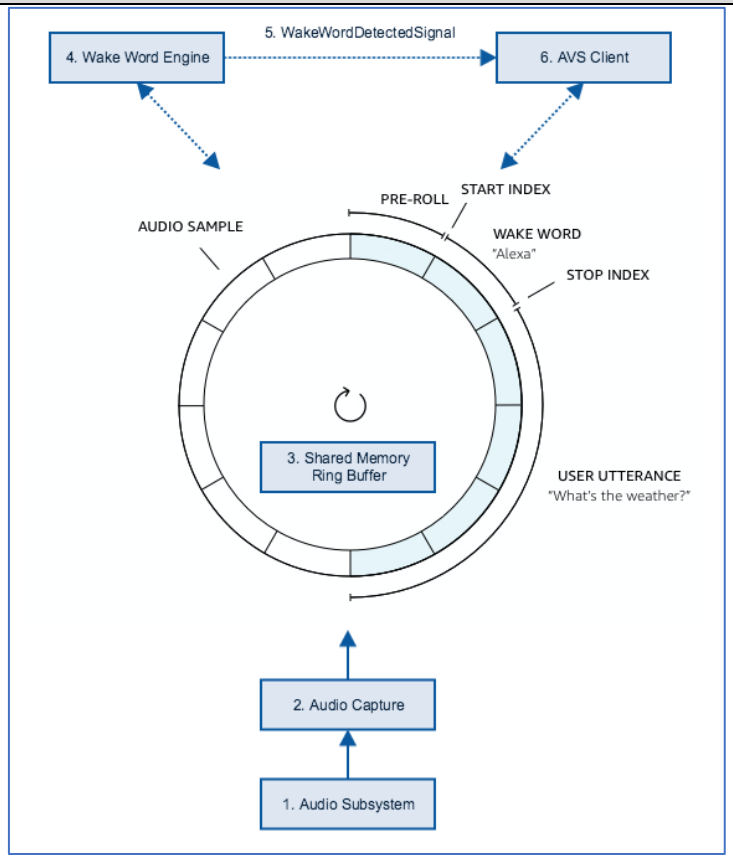
Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 17	Accused Instrumentality(ies)
	Moreover, in the cloud, the separate components of Alexa's speech-processing stack — automatic speech recognition (ASR), whisper detection, and speaker identification — run on separate server nodes with their own powerful processors. On-device, those functions have to share hardware not only with each other but with Alexa's other core device functions, such as music playback. With cloud-based ASR, encrypted audio streams to the cloud in small snippets called "frames". With on-device ASR, only the lattice is sent to the cloud, where a large and powerful neural language model reranks the hypotheses. The lattice can't be sent until the customer has finished speaking, as words later in a sequence can dramatically change the overall probability of a hypothesis. With cloud-based ASR, encrypted audio streams to the cloud in small snippets called "frames". With on-device ASR, only the lattice is sent to the cloud, where a large and powerful neural language model reranks the hypotheses. The lattice can't be sent until the customer has finished speaking, as words later in a sequence can dramatically change the overall probability of a hypothesis. For example, Amazon uses wake word and speech recognition.[282]

[282] *See* KAIFI0007141 *(Wake Word Verification Requirements*, https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/implement-ww-verification.html).

Ex.1029 / Page 328 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 17	Accused Instrumentality(ies)
	

Ex.1029 / Page 329 of 347
AMAZON.COM, INC.

Claim 17	Accused Instrumentality(ies)

> 4. **Wake Word Engine** - The component that reads the audio samples written to the shared memory ring buffer and analyzes the audio samples for occurrences of the wake word. The procedure of reading and processing audio samples occurs in a loop. For example, the loop reads N samples starting at location `0`; then analyzes the samples for an occurrence of the wake word. In the next iteration, it reads and analyzes the next N samples. When an occurrence of the wake word is detected, the wake word engine must identify two locations: 1) the start of the wake word, and 2) the end of the wake word. These values are stored as `startIndexInSamples` and `endIndexInSamples`.

For example, Amazon uses an ASR decoder.[283]

> In addition to these two classifiers, a third model component comes into play: meta-features emitted by the ASR decoder. These meta-features describe for how long (if at all) the ASR decoder is running "idle." In other words, how meaningful are the incoming audio chunks (actual content vs. silence/noise)?

For example, according to Amazon, performance conditions include noise.[284]

> Near-field and far-field tests are conducted in various locations, which mimic real-world user-device interaction patterns. The device's performance is evaluated under various conditions such as in silence, during audio playback from the device, or in environments with background noise. Such noise could be from music playing on another device, stationary noise such as an air conditioner, or other ambient noise.

[283] *See* KAIFI0001184 (Roland Maas, *Alexa scientists address challenges of end-pointing* (Apr. 10, 2018), https://www.amazon.science/blog/alexa-scientists-address-challenges-of-end-pointing).

[284] *See* KAIFI0001353 (Felipe Pinzon, *An Amazon technical expert provides three acoustic certification-related tips for AVS device makers* (Aug. 16, 2023) https://developer.amazon.com/en-US/blogs/alexa/device-makers/2023/08/acoustic-certification-alexa-august-2023).

Ex.1029 / Page 330 of 347
AMAZON.COM, INC.

Claim 17	Accused Instrumentality(ies)
	For example, according to Amazon, ASR is trained for silence and noise.[285] To train our acoustic models, we used connectionist temporal classification (CTC), followed by multiple passes of state-level minimum-Bayes-risk (sMBR) training. To make the acoustic model more robust, we also mixed noise into the training set, enabling the model to focus on characteristics of the input signal that vary less under different acoustic conditions. Adapting to conversational speech also required modification of Alexa's end-pointer, which determines when a customer has finished speaking. The end-pointer already distinguishes between pauses at the ends of sentences, indicating that the customer has stopped speaking and that Alexa needs to follow up, and mid-sentence pauses, which may be permitted to go on a little longer. For Live Translation, we modified the end-pointer to tolerate longer pauses at the ends of sentences, as speakers engaged in long conversations will often take time between sentences to formulate their thoughts. For example, according to Amazon, ASR adds a silence label at start and end of the label sequence and in between words.[286]

[285] *See* KAIFI0002096 (Shirin Saleem, Roland Maas, *How Alexa's new Live Translation for conversations works* (Dec. 14, 2020), https://www.amazon.science/blog/how-alexas-new-live-translation-for-conversations-works).

[286] *See* KAIFI0004444 (Tulsiani, et. al, *Improved training strategies for end-to-end speech recognition in digital voice assistants*, https://assets.amazon.science/ab/b4/2f3e5aec4506930e172478586bf4/improved-training-strategies-for-end-to-end-speech-recognition-in-digital-voice-assistants.pdf).

Ex.1029 / Page 331 of 347
AMAZON.COM, INC.

Claim 17	Accused Instrumentality(ies)

3.3. Label smoothing via probabilistic modelling of silence and use of multiple pronunciation

Label smoothing was proposed as a regularization technique to improve generalization [28, 29]. It proposes to penalize low entropy output distribution by adding noise to the labels [19, 28]. In this work, we introduce a simple yet powerful technique for changing label distribution. To penalize over-confident predictions by network, while training we: (i) randomly use one of the multiple pronunciations from the lexicon (as also done in [10, 22]) (ii) randomly add silence label at start and end of the label sequence and in between words. This has the effect of changing label distribution for a fixed word sequence, thereby providing the regularization effect. For this work, we use probability of 0.8 for inserting silence label at start and end and 0.2 for inserting a silence label between words.

For example, according to Amazon, ASR uses end pointers.[287]

The model that determines when the customer has finished speaking is called an end-pointer. End-pointers offer a natural trade-off between accuracy and latency: an aggressive end-pointer will initiate speech processing earlier, but it might cut the speaker off prematurely, resulting in a poor customer experience.

[287] *See* KAIFI0006791 (Ariya Rastrow, Shehzad Mevawalla, *On-device speech processing makes Alexa faster, lower-bandwidth*, AMAZON SCIENCE (Jan. 25, 2022), https://www.amazon.science/blog/on-device-speech-processing-makes-alexa-faster-lower-bandwidth).

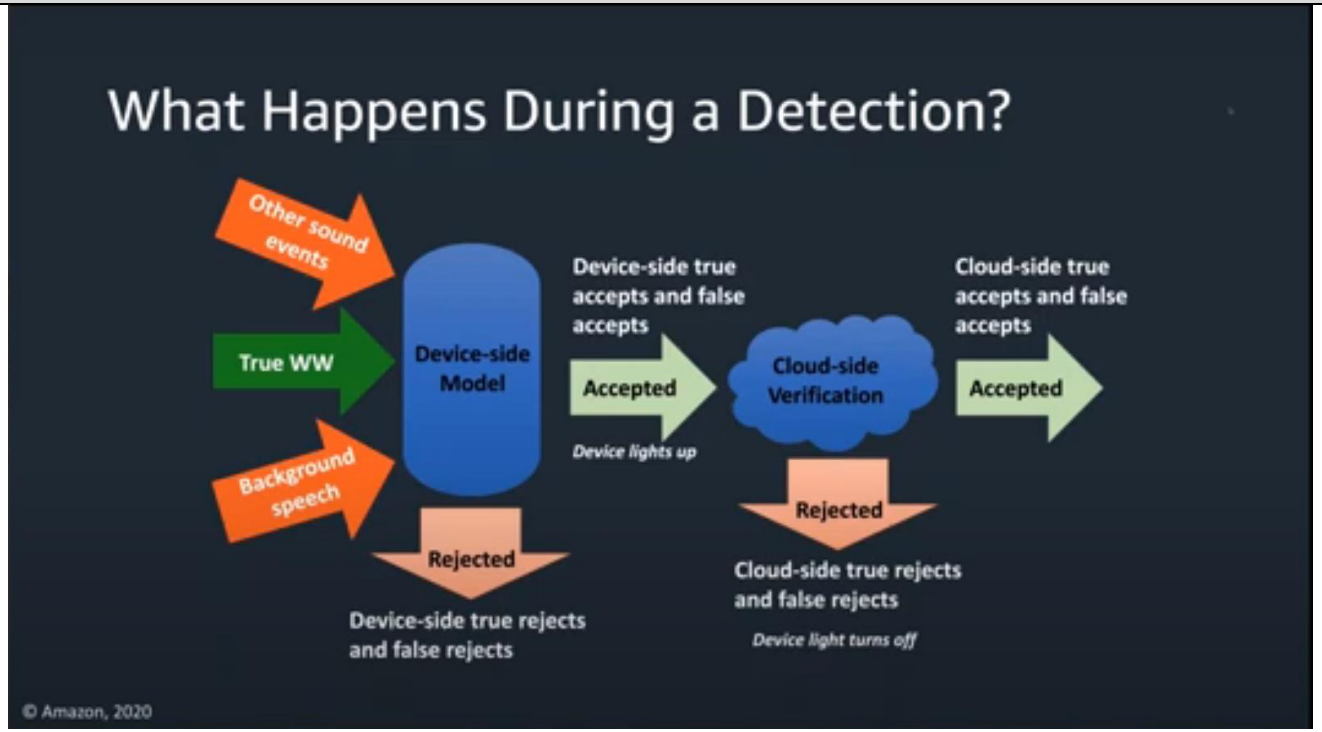
Ex.1029 / Page 332 of 347
AMAZON.COM, INC.

Claim 17	Accused Instrumentality(ies)
	> In a process called "end-pointing", however, Alexa recognizes the end of an utterance by the short period of silence that follows end of speech, and that silence is part of the input to the whisper detector. When we apply the detector to live data, we typically see that its confidence increases across most of the duration of an utterance then falls off precipitously in the final 50 or so frames. As a further example, after receiving a wake word together with other sound events and background speech, Amazon's device-side model and cloud-side verification perform "true accepts" of the wake word.[288]

[288] *See* KAIFI0007814 *(The Science Behind Alexa's Wake Word Detection*, https://www.youtube.com/watch?v=kYrw2cC9dvs).

Ex.1029 / Page 333 of 347
AMAZON.COM, INC.

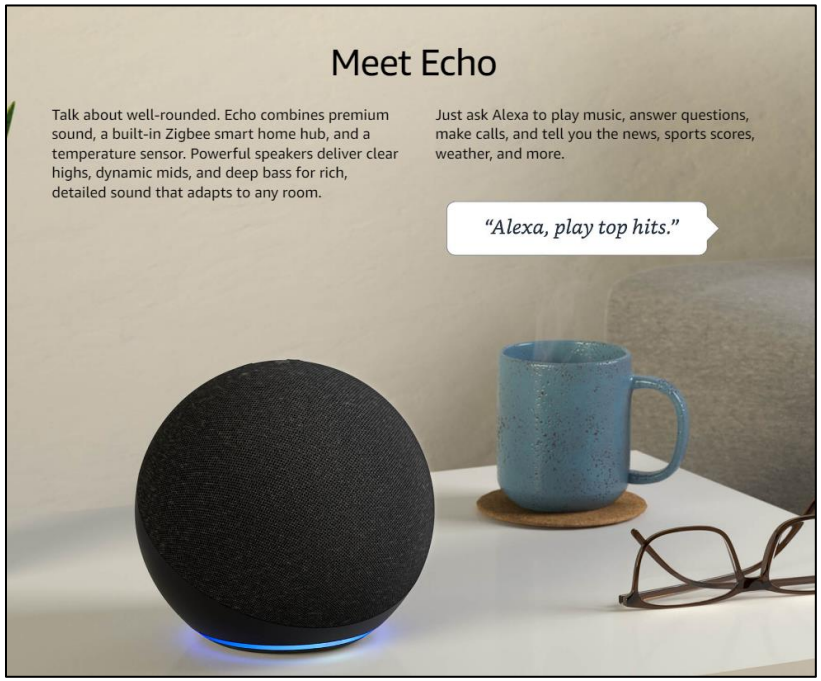
Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 17	Accused Instrumentality(ies)
	

Claim 18 of the '196 Patent

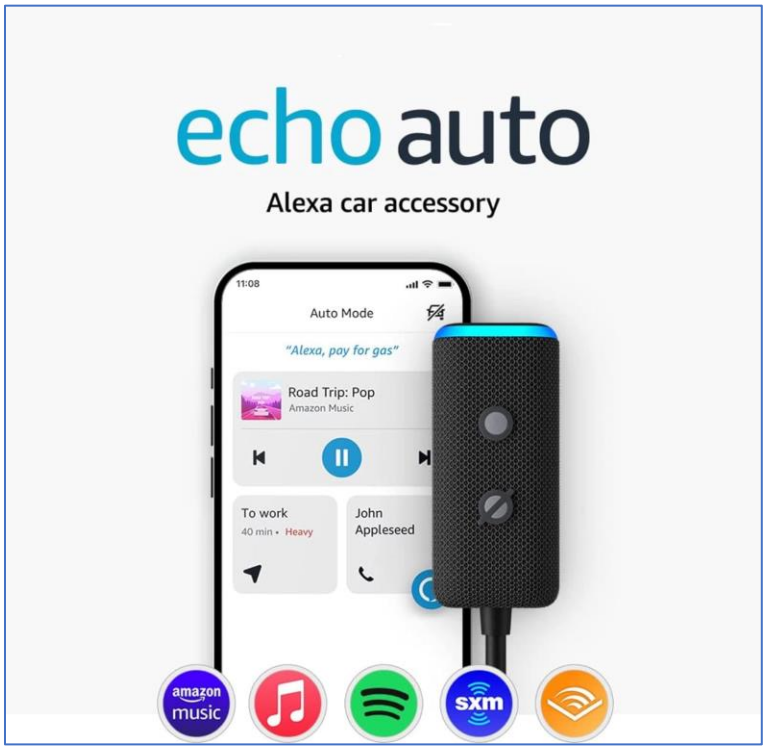
Claim 18	Accused Instrumentality(ies)
The method of claim 17, wherein the single device	Amazon's single device is a portable device or an embedded system.

Ex.1029 / Page 334 of 347
AMAZON.COM, INC.

Claim 18	Accused Instrumentality(ies)
is a portable device or an embedded system.	For example, Amazon's single device can be an Echo (4th Gen).[289] 

[289] *See* KAIFI0001202 (*Echo (4th Gen) | With premium sound, smart home hub, and Alexa | Charcoal*, https://www.amazon.com/dp/B07XKF5RM3?ref=MARS_NAV_desktop_shop_dvc_smart_spkrs_Echo_Spkr).

Ex.1029 / Page 335 of 347
AMAZON.COM, INC.

Claim 18	Accused Instrumentality(ies)
	For example, Amazon's single device can be an Echo Auto. [290]  For example, Amazon's single device can be an Alexa Built-in Device, including smart speakers, smart headphones, and personal computers.[291]

[290] *See* KAIFI0001979 *(Echo Auto (2nd Gen, 2022 release)*, https://www.amazon.com/All-New-Echo-Auto-release-Alexa/dp/B09X27YPS1?th=1).
[291] *See* KAIFI0007733 *(Alexa Built-in Devices, https://developer.amazon.com/en-US/alexa/devices/alexa-built-in).*

Ex.1029 / Page 336 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 18	Accused Instrumentality(ies)
	What are Alexa Built-in Devices?  The world's leading OEMs offer Alexa Built-in devices that let customers talk directly with Alexa through a microphone and speaker. Alexa turns your device into a smart device, and expands its capabilities. Customers can ask Alexa for help with everyday tasks and entertainment, and control Alexa connected devices or their own device. **Learn About the Underlying Technology** Alexa Built-in devices are powered by **Alexa Voice Service (AVS)**. The AVS team manages the services and infrastructure required for Alexa experiences, and provides a suite of device APIs, SDKs, and documentation. Developers may also work with third-party solution providers to utilize a qualified AVS solution. **Start Developing with Solution Providers** Our global network of **Alexa Solution Providers** work together to help you develop high-quality experiences, accelerate your development, and reduce your overall development cost. The earlier you engage a Solution Provider, the better they can assist you.

Ex.1029 / Page 337 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 18	Accused Instrumentality(ies)

Why Build Alexa Built-in Devices?

Delight Your Customers

Alexa makes your devices easier to use through voice, the most natural user interface. And, you can provide customers the choice and flexibility to interact with multiple voice services through your device.

Increase Customer Loyalty

Alexa makes your device more useful throughout the day. Customers can access information and entertainment, set timers and alarms, control Alexa connected devices, and more – all from your device.

Market to More Customers

Alexa differentiates your device, and helps you take advantage of the increased demand for smart devices. Certified Alexa Built-in devices receive increased exposure through Amazon.com.

What You Can Build



Smart Speakers



Smart Headphones



Personal Computers



Smart Vehicles

Learn More



Smart TVs and Screens

Ex.1029 / Page 338 of 347
AMAZON.COM, INC.

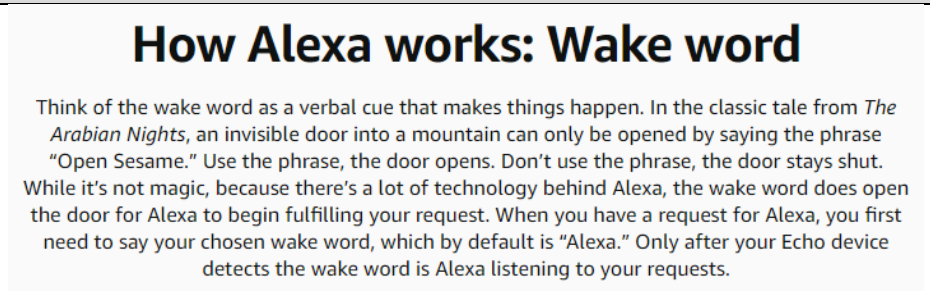
Claim 18	Accused Instrumentality(ies)
	This claim element is literally present, as described herein, but is also satisfied under the doctrine of equivalents because the structure and functionality provided by Amazon is being used to satisfy this claim limitation and therefore there is no vitiation of this element, and the Accused Instrumentalities incorporate structure and functionality that is not substantially different from the requirements of this limitation because they achieve substantially or exactly the same function, in substantially or exactly the same way as set forth in the claim element (as described herein), and achieves substantially or exactly the same result.

Claim 19 of the ’196 Patent

Claim 19	Accused Instrumentality(ies)
The method of claim 17, wherein the single device, wherein determining whether the input speech is recognized as the preset call command further comprises: computing, by a confidence detection unit, a confidence rate of the input speech; and recognizing, by the confidence detection unit, the input speech as the preset call command	Amazon’s determining whether the input speech is recognized as the preset call command further comprises: computing, by a confidence detection unit, a confidence rate of the input speech; and recognizing, by the confidence detection unit, the input speech as the preset call command based on the computed confidence rate. For example, the Accused Instrumentalities only listen to requests after the wake word is detected.[292]

[292] *See* KAIFI0002092 *(How Alexa works: Wake word*, https://www.amazon.com/b?ie=UTF8&node=23608571011).

Ex.1029 / Page 339 of 347
AMAZON.COM, INC.

Claim 19	Accused Instrumentality(ies)
based on the computed confidence rate.	 For example, the Accused Instrumentalities use built-in technology called "keyword spotting" that matches spoken audio to the acoustic patterns of the wake word and are designed to detect only the sound waves of the chosen wake word, which is sent to Amazon's secure cloud, where the request is reanalyzed to verify the wake word was spoken.[293]

[293] *See id.*

Ex.1029 / Page 340 of 347
AMAZON.COM, INC.

Claim 19	Accused Instrumentality(ies)
	But how can my device detect the wake word without hearing the rest of my conversations? Echo devices use built-in technology called "keyword spotting" that matches spoken audio to the acoustic patterns of the wake word. Simply put, Echo devices are designed by default to detect only the sound waves of your chosen wake word, and everything else is ignored. Like water through a strainer, all other audio (people talking, faucets running, birds chirping) passes through the device until the wake word is "caught" and sent to Amazon's secure cloud, where your request is reanalyzed to verify the wake word was spoken. You can choose to opt-in to features like Alexa Emergency Assist that allow your device to detect more than your chosen wake word—for example, the sound of smoke alarms—but you would need to update your settings to do so. Want to keep nerding out? Follow the journey of a voice request. For example, the Accused Instrumentalities achieve far-field speech recognition in noisy environments with high accuracy through a combination of machine learning algorithms, data, and immense computing power.[294] By launching Echo in November 2014, we showed that far-field speech recognition in noisy environments can be done with high accuracy through a combination of machine learning algorithms, data, and immense computing power.

[294] *See* KAIFI0004342 *(How Our Scientists are Making Alexa Smarter* (Mar. 29, 2018), https://www.aboutamazon.com/news/devices/how-our-scientists-are-making-alexa-smarter).

Ex.1029 / Page 341 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 19	Accused Instrumentality(ies)
	For example, Amazon's wake word detection employs deep learning technology running on-device to recognize the wake word the user has chosen.[295] It starts with detecting the "wake word" that triggers Alexa to start listening to the subsequent spoken words from the user. Wake word detection employs deep learning technology running on-device to recognize the wake word the user has chosen. Far-field automatic speech recognition (ASR) in the Amazon Web Services (AWS) cloud then converts into text the audio following the wake word, and determines when the user has stopped talking to Alexa. For example, Amazon uses a suite of learning techniques, including supervised, semi-supervised, and unsupervised learning.[296]

[295] *See id.*
[296] *See id.*

Ex.1029 / Page 342 of 347
AMAZON.COM, INC.

Claim 19	Accused Instrumentality(ies)
	A. Since Alexa's brain is mostly in the cloud, she gets smarter with every interaction. Alexa uses a suite of learning techniques: supervised, semi-supervised, and unsupervised learning. While supervised learning is still most powerful, it does not scale since we cannot generate manual labels at the pace required to continually improve Alexa for our customers. Therefore, our scientists and engineers are continually applying and inventing new learning techniques to reduce reliance on manual labels for training our statistical models. For instance, active learning as a class of semi-supervised learning techniques, where the system itself posits which part of the interactions it needs input from a human expert, is pervasive across our technologies. Unsupervised learning without any labeled responses is also being applied to make Alexa smarter, especially for speech recognition. Finally, we also use transfer learning, allowing Alexa to learn from one skill to another or even one language to another. For example, Amazon uses machine learning model trained from many occurrences of detected wake words to score a match with the input voice data.[297] **Wake Word Model** A WWD uses a machine learning model trained from many occurrences of detected wake words to score a match with the input voice data. If the confidence score exceeds the model's threshold, the wake word is considered detected and the WWD triggers. The wake word model may incorporate multiple wake words and also be trained to support multiple languages.

[297] *See* KAIFI0007120 (*Voice Interoperability Initiative Architecture Series Whitepapers*, https://m.media-amazon.com/images/G/01/vii/VII_Architecture_Series_Whitepapers_-_Multi-Agents_Wake_Word.pdf).

Ex.1029 / Page 343 of 347
AMAZON.COM, INC.

Claim 19	Accused Instrumentality(ies)
	For example, Amazon's ASR system outputs a confidence score to verify the presence of the wake word by comparing to a fixed threshold.[298]

2.2. Wakeword Verification

Unlike detection, which operates in a streaming manner, wake-word verification performs a single pass on the input localized context frame. To validate our proposed CRA architecture, we consider several baseline systems that include both whole-word and phonetic based models.

2.2.1. Baselines

Phonetically modeled systems that we consider are ASR [21] and 2-stage DNN-HMM [11]. The ASR system we consider is a large vocabulary DNN-HMM system, with an n-gram language model to recognize the spoken words in the audio stream. The system outputs a confidence score for each recognized word in

[298] *See* KAIFI0001937 (Rajath Kumar, et. al, *Building a Robust Word-Level Wakeword Verification Network,* https://assets.amazon.science/ba/c5/bd48ce11445ba0368d2d9191600d/building-a-robust-word-level-wakeword-verification-network.pdf).

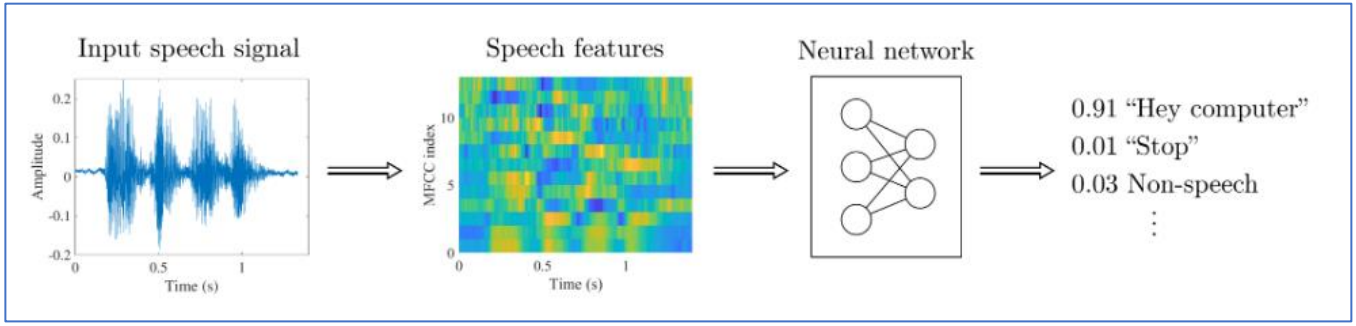
Ex.1029 / Page 344 of 347
AMAZON.COM, INC.

Claim 19	Accused Instrumentality(ies)
	the range (0,1). Similarly, for the defined wakeword, the ASR model outputs a confidence score, which we use to verify the presence of the wakeword by comparing to a fixed threshold. Since every stream is initiated with the wakeword, we set the confidence score for the wakeword to be 0 when the wakeword is not detected by ASR. Note that this ASR model is not specifically trained for the wakeword verification task, i.e., is not biased towards the wakeword. In the 2-stage DNN-HMM system, the first stage is similar to the acoustic part of the ASR system. However, unlike ASR, the 2-stage system is modeled only on the phonemes representing the wakeword and 2 additional outputs to distinguish background and silence [22]. A fixed length feature vector is extracted from the first stage DNN-HMM. This vector is engineered to aggregate information over all the frames in the hypothesized wakeword segment of the stream [22]. We pass the extracted vector as input to the second stage, which consist of multiple feed-forward layers. The second stage is a binary classification system trained to verify the presence of the wakeword. The two stages are trained separately, and the resulting score from the second stage is used for verification. This claim element is literally present, as described herein, but is also satisfied under the doctrine of equivalents because the structure and functionality provided by Amazon is being used to satisfy this claim limitation and therefore there is no vitiation of this element, and the Accused Instrumentalities incorporate structure and functionality that is not substantially different from the requirements of this limitation because they achieve substantially or exactly the same function, in substantially or exactly the same way as set forth in the claim element (as described herein), and achieves substantially or exactly the same result.

Ex.1029 / Page 345 of 347
AMAZON.COM, INC.

Appendix B
Claim Chart for U.S. Patent No. 8,930,196

Claim 20 of the '196 Patent

Claim 20	Accused Instrumentality(ies)
The method of claim 17, wherein determining whether the input speech is recognized as the preset call command further comprises: computing a log likelihood ratio of the input speech; and assuming the input speech to be the preset call command based on the computed log likelihood ratio.	Amazon's determining whether the input speech is recognized as the preset call command further comprises: computing a log likelihood ratio of the input speech; and assuming the input speech to be the preset call command based on the computed log likelihood ratio. For example, Amazon uses keyword-spotting algorithms as illustrated below. The input speech signal is first converted to a feature representation, such as MFCCs, which are fed to a neural network, and the output is the likelihood of each keyword.[299]  This claim element is literally present, as described herein, but is also satisfied under the doctrine of equivalents because the structure and functionality provided by Amazon is being used to satisfy this claim limitation and therefore there is no vitiation of this element, and the Accused Instrumentalities incorporate structure and functionality that is not substantially different from the requirements of this limitation

[299] *See* KAIFI0001092 (*Introduction to Speech Processing*, https://speechprocessingbook.aalto.fi/Recognition/Wake-word_and_keyword_spotting.html).

Ex.1029 / Page 346 of 347
AMAZON.COM, INC.

Claim 20	Accused Instrumentality(ies)
	because they achieve substantially or exactly the same function, in substantially or exactly the same way as set forth in the claim element (as described herein), and achieves substantially or exactly the same result.

Ex.1029 / Page 347 of 347
AMAZON.COM, INC.