IN THE UNITED STATES DISTRICT COURT
FOR THE EASTERN DISTRICT OF VIRGINIA
ALEXANDRIA DIVISION

DIALECT, LLC 133 E. Tyler St. Longview, Texas 75601-7216, *Plaintiff*, v. AMAZON.COM, INC., 410 Terry Avenue North Seattle, Washington 98109, and AMAZON WEB SERVICES, INC. 410 Terry Avenue North Seattle, Washington 98109, *Defendants*.	Civil Action No. _____ **JURY TRIAL DEMANDED**

COMPLAINT FOR PATENT INFRINGEMENT AND DAMAGES
AND DEMAND FOR JURY TRIAL

Plaintiff Dialect, LLC ("Dialect" or "Plaintiff") files this Complaint for patent infringement and damages against Amazon.com, Inc. and Amazon Web Services, Inc. ("AWS") (collectively, "Amazon" or "Defendants") and alleges as follows:

INTRODUCTION

1. The novel inventions disclosed in U.S. Patent Nos. 7,693,720 (the "'720 Patent"); 8,015,006 (the "'006 Patent"); 8,140,327 (the "'327 Patent"); 8,195,468 (the "'468 Patent"); 9,031,845 (the "'845 Patent"); 9,263,039 (the "'039 Patent"); 9,495,957 (the "'957 Patent") (collectively, the "Asserted Patents") in this matter were invented by VoiceBox Technologies ("VoiceBox"). VoiceBox was a key pioneer in the fields of voice recognition technology and

Meta Platforms, Inc. - IPR2025-01332
Meta Platforms, Inc. v. Dialect, LLC
Ex. 1043 - Page 1

natural language understanding ("NLU"). These technologies power a wide variety of consumer electronics and provide key functionality for smart phones, tablets, TVs, and Internet of Things ("IoT") devices. VoiceBox spent more than a decade developing and building key early NLU inventions producing one of the most valuable portfolios of technology according to the Institute of Electrical and Electronics Engineers ("IEEE"), including the Asserted Patents. In recognition of its many innovations, the U.S. Patent & Trademark Office awarded and issued the Asserted Patents to VoiceBox. The Asserted Patents in this case are the result of this substantial investment and research and were fundamental to the development of voice commerce technology.

2. Over the years, the inventions claimed in the Asserted Patents have been licensed to key companies in the industry.

3. Despite opportunities to do so and its knowledge of the Asserted Patents, Amazon has never licensed the Asserted Patents. VoiceBox's opportunities to promote and build a business based on these patents were thwarted when Amazon introduced the infringing Echo and other Alexa Products[1] and used its enormous size and clout to poach dozens of VoiceBox Technologies' engineers and scientists.

[1] "Alexa Products" collectively refers to Amazon's Alexa virtual assistant and offerings that include Alexa, including the Echo product line (such as Echo 1st Gen., Echo 2nd Gen., Echo Dot 1st Gen., Echo Dot 2nd Gen., Echo Dot 3rd Gen., Echo Dot Kids Edition, Echo Show 1st Gen., Echo Show 2nd Gen., Echo Show 5, Echo Spot, Echo Plus 1st Gen., Echo Plus 2nd Gen., Echo Auto, and Echo Look), Amazon's Alexa apps, Music apps, and Shopping apps on a smartphone or other mobile device, Amazon's Alexa cloud, Alexa Voice Services, and Amazon.com website, and any other device, app, or instrumentality that includes, provides access to, or works with Alexa (such as Amazon Tap, Amazon Dash Wand, Echo Wall Clock, AmazonBasics Microwave, Amazon SmartPlug, Amazon Fire TV Sticks, Amazon Fire TVs, Amazon Fire TV Cubes, and Amazon Fire and Fire HD tablets) as well as software, hardware, and cloud infrastructure associated with any of the foregoing. Plaintiff reserves the right to expand upon or otherwise modify the above list during the discovery process in this case and is in no way limiting the scope of the accused products to what is currently listed.

THE PARTIES

4. Plaintiff is the current owner and assignee of the Asserted Patents.

5. Plaintiff is a Texas limited liability company with its principal place of business located at 133 E. Tyler St., Longview, TX 75601-7216.

6. Amazon.com, Inc. is a corporation incorporated in Delaware and has a principal place of business at 410 Terry Avenue North, Seattle, WA, 98109. Amazon.com, Inc. is the ultimate parent company of various subsidiaries, including AWS, which are responsible for making, using, selling, offering for sale and/or importing Alexa Products. Amazon maintains a regular and established place of business in this District through multiple permanent physical facilities, including in particular Amazon's second headquarters that is located in this District. Amazon's second headquarters is located in the National Landing neighborhood of Arlington, Virginia, which includes at least the leased properties at 241 18th St. S., 1770 Crystal Dr., 2100 Crystal Dr., 2345 Crystal Dr., and 1800 S. Bell St., Arlington, Virginia and included at least two additional office buildings that were developed for and have been delivered to Amazon for occupancy.[2]

7. AWS is a corporation incorporated in Delaware and has a principal place of business at 410 Terry Avenue North, Seattle, WA, 98109. AWS provides the cloud computing platforms that power one or more Alexa Products or are for use with one or more Alexa Products.[3] AWS is registered to do business in Virginia.[4]

---

[2] *See* Presentation by Joe Chapman, Director, Amazon Global Real Estate & Facilities, PenPlace, Amazon Arlington HQ, Site Plan Review Committee: SPRC #4, February 10, 2022 at 3, available at https://www.arlingtonva.us/files/sharedassets/public/projects/documents/site-planprojects/penplace/2022_02-07-sprc-4-deck.pdf.

[3] https://aws.amazon.com/blogs/aws/majority-of-alexa-now-running-on-faster-more-cost-effective-amazon-ec2-inf1-instances/

[4] https://cis.scc.virginia.gov/EntitySearch/Index

Entity ID	Entity Name	Name Type	Entity Type	Series LLC	Principal Office Address	RA Name	Status
T0395139	AMAZON WEB SERVICES LLC	Legal Name	Limited Liability Company	No	1200 12TH AVE STE 1200, SEATTLE, WA, 98144 - 0000, USA	CORPORATION SERVICE COMPANY	Inactive
F1918947	AMAZON WEB SERVICES, INC.	Legal Name	Stock Corporation		410 Terry Ave N, Seattle, WA, 98109 - 5210, USA	CORPORATION SERVICE COMPANY	Active

8. On information and belief, Amazon directly and/or indirectly develops, designs, manufactures, distributes, markets, offers to sell and/or sells infringing products and services in the United States, including in the Commonwealth of Virginia and within the Eastern District of Virginia, and otherwise direct infringing activities to this District in connection with their products and services as set forth in this Complaint.

JURISDICTION

9. This civil action arises under the Patent Laws of the United States, 35 U.S.C. § 1 *et seq.*, including without limitation 35 U.S.C. §§ 271, 281, 283, 284, and 285. Accordingly, this Court has subject matter jurisdiction under, *inter alia*, 28 U.S.C. §§ 1331 and 1338(a).

10. This District has general and specific personal jurisdiction over Amazon because Amazon has committed acts, directly or through intermediaries, in this District, giving rise to this action; is present in and transacts and conducts business in this District and the Commonwealth of Virginia; and transacts and conducts business with residents of this District and the Commonwealth of Virginia.

11. Plaintiff's causes of action arise, at least in part, from Amazon's contacts with and activities in this District and the Commonwealth of Virginia.

12. Amazon has infringed the Asserted Patents within this District and the Commonwealth of Virginia by making, using, selling, offering for sale, and/or importing in or into this District and elsewhere in the Commonwealth of Virginia, products that infringe the Asserted Patents, including, without limitation, products that practice the claimed methods of the Asserted Patents. Amazon, directly and through intermediaries, makes, uses, sells, offers for sale, imports, ships, distributes, advertises, promotes, and/or otherwise commercializes such infringing products

in or into this District and the Commonwealth of Virginia. Amazon regularly conducts and solicits business in, engages in other persistent courses of conduct in, and/or derives substantial revenue from goods and services provided to residents of this District and the Commonwealth of Virginia.

13. This Court has general and specific personal jurisdiction over Amazon. Amazon has substantial contacts with the forum as a consequence of actively working on and establishing its second headquarters in Virginia and in this District, and Amazon conducts substantial business in Virginia. Amazon states in its most recent 10-K filing to the SEC for fiscal year ended December 31, 2022, dated February 3, 2023, that "[w]e own and lease our corporate headquarters in Washington's Puget Sound region and Arlington, Virginia."

14. Amazon sells, makes, uses, and offers for sale its products and services, including products and services that infringe the Plaintiff's Asserted Patents, within the Commonwealth of Virginia, including to customers in Virginia. Amazon further operates data centers in Ashburn, Virginia[5] and has announced $35 billion in additional investment in data centers in Virginia through 2040.[6]

15. A Court in this District has observed in April 2020:

> It must be said that Amazon is nothing if not ubiquitous in the United States. Furthermore, after considering 238 cities, Amazon chose Arlington in the Eastern District of Virginia as the location for its HQ2 and will invest $2.5 billion and 25,000 jobs in the undertaking. As such, Amazon cannot in good faith represent to the Court that E.D. Va. is an undesirable or inconvenient location to operate and do business. Litigating should not be an additional significant strain.

---

[5] https://www.datacenters.com/amazon-aws-ashburn

[6] https://apnews.com/article/technology-data-management-and-storage-amazoncom-inc-virginia-business-c75df1f34069b09549fe15c99335b8fb

Maglula, Ltd. v. Amazon.com, Inc., No. 1:19-cv-01570, ECF No. 52 at 32-33 (E.D. Va. Apr. 9, 2020).

16. Amazon has further stated that it is specifically hiring for its Alexa teams at HQ2 located in Alexandria, Virginia. At least as early as September 2019, Amazon listed 25 open positions for the Alexa team at its HQ2.[7] Amazon's website stated that as of September 2021, Amazon was still growing its teams at HQ2, "including: software development engineers, technical sales representatives, program managers, and solutions architects on teams across" divisions including "Alexa."[8] In 2022, Amazon publicly promoted the "Alexa Teams in HQ2," including multiple "Software Development Managers" that moved to HQ2 in Alexandria.[9] Amazon's description of this video describes taking "a dive into the teams that make up Alexa at Amazon's new second headquarters in Arlington, Virginia. In this video, meet different managers leading the Alexa teams in HQ2 to learn more about what they're building and their favorite things about living in Northern Virginia."[10] Amazon further stated in the same description that "Alexa is hiring on all teams."[11] As of December 2022, Amazon reported that students from the University of the District of Columbia and the University of Maryland College Park "presented their Alexa Skills final projects" to "Alexa employees, and Amazon leaders at the company's headquarters in Virginia (HQ2)."[12]

---

[7] https://www.spglobal.com/marketintelligence/en/news-insights/latest-news-headlines/amazon-hiring-for-cloud-services-alexa-products-at-hq2-in-arlington-va-53798578

[8] https://www.aboutamazon.com/news/amazon-offices/the-next-chapter-for-hq2-sustainable-buildings-surrounded-by-nature

[9] https://www.youtube.com/watch?v=vU2szgs2M7c

[10] *Id.*

[11] *Id.*

[12] https://www.amazon.science/academic-engagements/amazonnext-program-hosts-final-project-presentations-at-virginia-hq2

17. This Court also has personal jurisdiction over Amazon, in part, because Amazon does continuous and systematic business in this District, including by providing infringing products and services to the residents of this District that Amazon knew would be used within this District, and by soliciting business from the residents of this District.

18. Amazon has pled in legal filings that venue in this District is proper. In a 2020 action filed by Amazon in this District, Amazon asserted that venue was proper under 28 U.S.C. § 1391(b) because, among other things "it is a district in which Plaintiff [Amazon] maintains headquarters and/or substantial business operations". *Amazon.com, Inc. v. WDC Holdings LLC*, No. 1:20-cv-484, ECF No. 1, ¶ 26 (E.D. Va. Apr. 27, 2020).

19. Venue is proper in this Court under 28 U.S.C. §§ 1391 and 1400(b).

THE ASSERTED PATENTS

20. The VoiceBox inventions contained in the Asserted Patents in this case relate to groundbreaking improvements to voice recognition and NLU and have particular application in consumer electronics such as smart phones, tablets, and IoT devices.

U.S. PATENT NO. 7,693,720

21. On April 6, 2010, the U.S. Patent Office duly and legally issued the '720 Patent, entitled "Mobile Systems And Methods For Responding To Natural Language Speech Utterance." A true and correct copy of the '720 Patent is attached hereto as **Exhibit 1**.

22. Dialect is the owner and assignee of all right, title, and interest in and to the '720 Patent, including the right to assert all causes of action arising under the '720 Patent and the right to sue and obtain any remedies for past, present, or future infringement.

23. The '720 Patent describes, among other things, a complete speech-based information query, retrieval, presentation and local or remote command environment. '720 Patent,

Abstract. The invention can be used in dynamic environments such as those of mobile vehicles to control and communicate with both vehicle systems and remote systems and devices. *Id*. As the ’720 Patent explains, prior to its inventions, “creating a natural language speech interface that is suitable for use in the vehicular environment has proved difficult. A general-purpose telematics system must accommodate commands and queries from a wide range of domains and from many users with diverse preferences and needs. Further, multiple vehicle occupants may want to use such systems, often simultaneously. Finally, most vehicle environments are relatively noisy, making accurate speech recognition inherently difficult.” *Id*. at 1:34–42. “Managing and evaluating complex and uncertain queries while maintaining real-time performance is a significant challenge.” *Id*. at 2:40–42. The inventions described and claimed in the ’720 Patent overcome these challenges in various embodiments, for example by providing “a complete speech-based information query, retrieval, presentation and command environment,” which “makes significant use of context, prior information, domain knowledge, and user specific profile data to achieve a natural environment for one or more users making queries or commands in multiple domains. Through this integrated approach, a speech-based natural language query, response and command environment is created. Further, at each step in the process, accommodation may be made for full or partial failure and graceful recovery.” *Id.* at 2:52–61.

24. The novel features of the invention are recited in the claims. For example, Claim 1 of the ’720 Patent recites:

> A mobile system responsive to a user generated natural language speech utterance, comprising:
>
> a speech unit connected to a computer device on a vehicle, wherein the speech unit receives a natural language speech utterance from a user and converts the received natural language speech utterance into an electronic signal; and
>
> a natural language speech processing system connected to the computer device on the vehicle, wherein the natural language speech processing system receives, processes, and responds to the electronic signal using data

received from a plurality of domain agents, wherein the natural language speech processing system includes:

a speech recognition engine that recognizes at least one of words or phrases from the electronic signal using at least the data received from the plurality of domain agents, wherein the data used by the speech recognition engine includes a plurality of dictionary and phrase entries that are dynamically updated based on at least a history of a current dialog and one or more prior dialogs associated with the user;

a parser that interprets the recognized words or phrases, wherein the parser uses at least the data received from the plurality of domain agents to interpret the recognized words or phrases, wherein the parser interprets the recognized words or phrases by:

determining a context for the natural language speech utterance;

selecting at least one of the plurality of domain agents based on the determined context; and

transforming the recognized words or phrases into at least one of a question or a command, wherein the at least one question or command is formulated in a grammar that the selected domain agent uses to process the formulated question or command; and

an agent architecture that communicatively couples services of each of an agent manager, a system agent, the plurality of domain agents, and an agent library that includes one or more utilities that can be used by the system agent and the plurality of domain agents, wherein the selected domain agent uses the communicatively coupled services to create a response to the formulated question or command and format the response for presentation to the user.

'720 Patent at Claim 1.

U.S. PATENT NO. 8,015,006

25. On September 6, 2011, the U.S. Patent Office duly and legally issued the '006 Patent, entitled "Systems And Methods For Processing Natural Language Speech Utterances With Context-Specific Domain Agents." A true and correct copy of the '006 Patent is attached hereto as **Exhibit 2**.

26. Dialect is the owner and assignee of all right, title, and interest in and to the '006 Patent, including the right to assert all causes of action arising under the '006 Patent and the right to sue and obtain any remedies for past, present, or future infringement.

27. The ’006 Patent describes, among other things, novel systems and methods for receiving natural language queries and/or commands and executing the queries and/or commands. ’006 Patent, Abstract. The claimed invention makes significant use of context, prior information, domain knowledge, and user specific profile data to achieve a natural environment for one or more users. *Id*. As the ’006 Patent explains, prior to its inventions, a machine’s ability to communicate with humans in a natural manner was a difficult technical problem in need of a technical solution. As described in the specification, in the prior art “human questions and machine processing of queries may be fundamentally incompatible,” because “a person asking a question or giving a command typically relies heavily on context and the domain knowledge of the person answering,” whereas “machine-based queries” are “highly structured and are not inherently natural to the human user.” *Id.* at 1:27–41. The inventions described and claimed in the ’209 Patent overcome these challenges in various embodiments, for example by providing a system that uses domain agents to organize domain specific behavior and information. *Id.* at 2:53–3:7. The inventions in various embodiments further include a system that can “determine the user’s identity by voice and name for each utterance,” so that “[r]ecognized words and phrases may be tagged with this identity in all further processing” for security and other purposes. *Id.* at 16:60–17:4.

28. The novel features of the invention are recited in the claims. For example, Claim 5 of the ’006 Patent recites:

> A method for processing natural language speech utterances with context-specific domain agents, comprising:
>
> receiving, at a speech unit coupled to a processing device, a natural language speech utterance that contains a request;
>
> recognizing, at a speech recognition engine coupled to the processing device, one or more words or phrases contained in the utterance using information in one or more dictionary and phrase tables;
>
> parsing, at a parser coupled to the processing device, information relating to the utterance to determine a meaning associated with the utterance and a context

associated with the request contained in the utterance, wherein the parsed information includes the one or more recognized words or phrases;

formulating, at the parser, the request contained in the utterance in accordance with a grammar used by a domain agent associated with the determined context, wherein formulating the request in accordance with the grammar used by the domain agent includes:

determining one or more required values and one or more optional values associated with formulating the request in the grammar used by the domain agent;

extracting one or more criteria and one or more parameters from one or more keywords contained in the one or more recognized words or phrases, wherein the parser extracts the one or more criteria and the one or more parameters using procedures sensitive to the determined context;

inferring one or more further criteria and one or more further parameters associated with the request using a dynamic set of prior probabilities or fuzzy possibilities; and

transforming the one or more extracted criteria, the one or more extracted parameters, the one or more inferred criteria, and the one or more inferred parameters into one or more tokens having a format compatible with the grammar used by the domain agent, wherein the one or more tokens include all the required values and one or more of the optional values associated with formulating the request in the grammar used by the domain agent;

processing the formulated request with the domain agent associated with the determined context to generate a response to the utterance; and

presenting the generated response to the utterance via the speech unit.

'006 Patent at Claim 5.

U.S. PATENT NO. 8,140,327

29. On March 20, 2012, the U.S. Patent Office duly and legally issued the '327 Patent, entitled "System And Method For Filtering And Eliminating Noise From Natural Language Utterances To Improve Speech Recognition And Parsing." A true and correct copy of the '327 Patent is attached hereto as **Exhibit 3**.

30. Dialect is the owner and assignee of all right, title, and interest in and to the '327 Patent, including the right to assert all causes of action arising under the '327 Patent and the right to sue and obtain any remedies for past, present, or future infringement.

31. The ’327 Patent describes, among other things, a novel mobile system and method that filters and eliminates noise from natural language utterances to improve accuracy associated with speech recognition and parsing. ’327 Patent, Abstract. The claimed invention uses a microphone array and a filter to process a speech signal before it is sent to an encoder. *Id*. As the ’327 Patent explains, existing systems “are generally unable to provide a complete environment for users to make natural language speech queries and receive natural sounding responses.” *Id*. at 1:56-59. According to the ’327 Patent, prior to its inventions, a machine’s ability to communicate with humans in a natural manner was a difficult technical problem in need of a technical solution. As described in the specification, there “remains a number of significant barriers to creation of a complete natural language speech-based query and response environment.” *Id*. at 1:59-61. The inventions described and claimed in the ’327 Patent overcome some of these challenges in various embodiments, “including environments with background noise, point noise sources and people holding conversations, for example by providing a method that uses “one-dimensional or two-dimensional array microphones to receive human speech” and “filtering of speech input.” *Id.* 7:35-51.

32. The novel features of the invention are recited in the claims. For example, claim 14 of the ’327 Patent recites:

> A system for filtering and eliminating noise from natural language speech utterances, comprising:
>
> a microphone array configured to add one or more nulls to a beam pattern steered to point in a direction associated with a user speaking a natural language utterance to capture an input speech signal corresponding to the natural language utterance, wherein the one or more nulls notch out point or limited area noise sources from the input speech signal;
>
> an adaptive filter coupled to the microphone array, wherein the adaptive filter is configured to:
>
> receive the input speech signal corresponding to the natural language utterance from the microphone array and compare environmental, noise to

the input speech signal to set one or more parameters associated with the adaptive filter;

use band shaping and notch filtering to remove narrow-band noise from the input speech signal received from the microphone array according to the one or more parameters; and

suppress cross-talk and environmentally caused echoes in the input speech signal received from the microphone array using adaptive echo cancellation;

a speech coder arranged between the adaptive filter and a speech recognition engine, wherein the speech coder is configured to receive the input speech signal passed through the adaptive filter and use adaptive lossy audio compression to remove momentary gaps from the input speech signal and variable rate sampling to compress and digitize the input speech signal, wherein the speech coder optimizes the adaptive lossy audio compression and the variable rate sampling to only preserve components in the input speech signal that will be input to the speech recognition engine; and

a transceiver configured to communicate the digitized input speech signal from a buffer in the speech coder to the speech recognition engine at a rate that depends on available bandwidth associated with a communication link that connects the transceiver and the speech recognition engine.

’327 Patent at claim 14.

33. Figure 1 of the ’327 Patent, reproduced below, shows an overall diagrammatic view, including an array microphone, according to one embodiment of the invention.

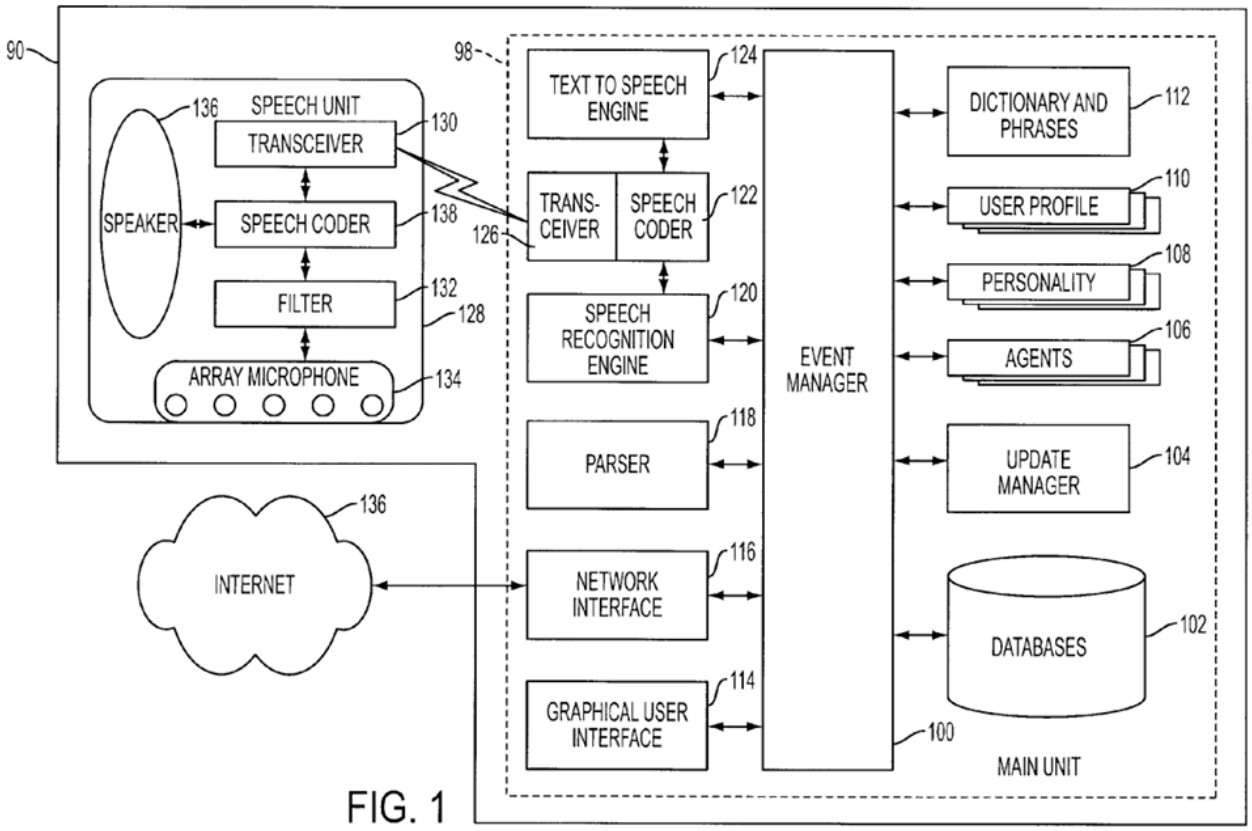


FIG. 1

'327 Patent, Fig. 1.

U.S. PATENT NO. 8,195,468

34. On June 5, 2012, the U.S. Patent Office duly and legally issued the '468 Patent, entitled "Mobile Systems And Methods Of Supporting Natural Language Human-Machine Interactions." A true and correct copy of the '468 Patent is attached hereto as **Exhibit 4**.

35. Dialect is the owner and assignee of all right, title, and interest in and to the '468 Patent, including the right to assert all causes of action arising under the '468 Patent and the right to sue and obtain any remedies for past, present, or future infringement.

36. The '468 Patent describes, among other things, a novel mobile system that identifies and uses context, prior information, domain knowledge, and user specific profile data to achieve a natural environment for users to submit natural language requests. '468 Patent, Abstract.

The claimed invention creates, stores and uses extensive personal profile information for each user to improve the reliability of determining the context of a request and presenting the expected results. *Id*. As the ’468 Patent explains, prior to its inventions, a machine’s ability to communicate with humans in a natural manner was a difficult technical problem in need of a technical solution. As described in the specification, under the existing systems and devices “verbal communications and machine processing of requests that are extracted from the verbal communications may be fundamentally incompatible,” because the existing systems and devices use requests that are “highly structured and may not be inherently natural to the human user.” *Id*. at 1:53-58. “Cognitive research on human interaction,” however, “shows that verbal communication, such as a person asking a question or giving a command, typically relies heavily on context and domain knowledge of the target person.” *Id*. at 1:49-52. The inventions described and claimed in the ’468 Patent overcome these challenges in various embodiments, for example by providing a system that uses “multi-modal communications that enable displaying of non-speech search results on a graphical interface” in conjunction with “speech commands” to execute requests. *Id.* at 21:49-60.

37. The novel features of the invention are recited in the claims. For example, claim 19 of the ’468 Patent recites:

> A method for processing multi-modal natural language inputs, comprising:
>
> > receiving a multi-modal natural language input at a conversational voice user interface, the multi-modal input including a natural language utterance and a non-speech input provided by a user, wherein a transcription module coupled to the conversational voice user interface transcribes the non-speech input to create a non-speech-based transcription;
> >
> > identifying the user that provided the multi-modal input;
> >
> > creating a speech-based transcription of the natural language utterance using a speech recognition engine and a semantic knowledge-based model, wherein the semantic knowledge-based model includes a personalized cognitive model derived from one or more prior interactions between the identified user and the conversational voice user interface, a general cognitive model derived from one or more prior interactions between a

> plurality of users and the conversational voice user interface, and an environmental model derived from an environment of the identified user and the conversational voice user interface;
>
> merging the speech-based transcription and the non-speech-based transcription to create a merged transcription;
>
> identifying one or more entries in a context stack matching information contained in the merged transcription;
>
> determining a most likely context for the multi-modal input based on the identified entries;
>
> identifying a domain agent associated with the most likely context for the multi-modal input;
>
> communicating a request to the identified domain agent; and
>
> generating a response to the user from content provided by the identified domain agent as a result of processing the request.

'468 Patent at claim 19.

U.S. PATENT NO. 9,031,845

38. On May 12, 2015, the U.S. Patent Office duly and legally issued the '845 Patent, entitled "Mobile Systems And Methods For Responding To Natural Language Speech Utterance." A true and correct copy of the '845 Patent is attached hereto as **Exhibit 5**.

39. Dialect is the owner and assignee of all right, title, and interest in and to the '845 Patent, including the right to assert all causes of action arising under the '845 Patent and the right to sue and obtain any remedies for past, present, or future infringement.

40. The '845 Patent describes, among other things, a complete speech-based information query, retrieval, presentation and local or remote command environment. '845 Patent, Abstract. The invention can be used in dynamic environments such as those of mobile vehicles to control and communicate with both vehicle systems and remote systems and devices. *Id*. As the '845 Patent explains, prior to its inventions, "creating a natural language speech interface that is suitable for use in the vehicular environment has proved difficult. A general-purpose telematics system must accommodate commands and queries from a wide range of domains and from many

users with diverse preferences and needs. Further, multiple vehicle occupants may want to use such systems, often simultaneously. Finally, most vehicle environments are relatively noisy, making accurate speech recognition inherently difficult." *Id*. at 1:38–46. In some cases, "multiple queries, perhaps with several parts, need to be made to multiple data sources, which can be both local or on a network." *Id*. at 2:25–27. The inventions described and claimed in the '845 Patent overcome these challenges in various embodiments, for example by providing "a complete speech-based information query, retrieval, presentation and command environment," which "makes significant use of context, prior information, domain knowledge, and user specific profile data to achieve a natural environment for one or more users making queries or commands in multiple domains. Through this integrated approach, a speech-based natural language query, response and command environment is created. Further, at each step in the process, accommodation may be made for full or partial failure and graceful recovery." *Id.* at 2:52–61. "The invention can be used for generalized local or network information query, retrieval and presentation in a mobile environment." *Id.* at 6:43–45.

41. The novel features of the invention are recited in the claims. For example, Claim 1 of the '845 Patent recites:

> A mobile system for processing natural language utterances, comprising:
>
> one or more physical processors at a vehicle that are programmed to execute one or more computer program instructions which, when executed, cause the one or more physical processors to:
>
> receive a natural language utterance associated with a user;
>
> perform speech recognition on the natural language utterance;
>
> parse and interpret the speech recognized natural language utterance;
>
> determine a domain and a context that are associated with the parsed and interpreted natural language utterance;
>
> formulate a command or query based on the domain and the context; determine whether the command or query is to be executed on-board or off-board the vehicle;

execute the command or query at the vehicle in response to a determination that the command or query is to be executed on-board the vehicle; and

invoke a device that communicates wirelessly over a wide area network to process the command or query such that the command or query is executed off-board the vehicle in response to a determination that the command or query is to be executed off-board the vehicle.

'845 Patent at Claim 1.

U.S. PATENT NO. 9,263,039

42. On February 16, 2016, the U.S. Patent Office duly and legally issued the '039 Patent, entitled "Systems And Methods For Responding To Natural Language Speech Utterance." A true and correct copy of the '039 Patent is attached hereto as **Exhibit 6**.

43. Dialect is the owner and assignee of all right, title, and interest in and to the '039 Patent, including the right to assert all causes of action arising under the '039 Patent and the right to sue and obtain any remedies for past, present, or future infringement.

44. The '039 Patent describes, among other things, novel systems and methods for receiving speech and non-speech communications and executing the questions and/or commands. '039 Patent, Abstract. The systems and methods improve the reliability of determining the context of speech and non-speech communications and presenting the expected results for a particular question or command. *Id*. As the '039 Patent explains, prior to its inventions, a machine's ability to communicate with humans in a natural manner was a difficult technical problem in need of a technical solution. As described in the specification, in the prior art a "machine's ability to communicate with humans in a natural manner remains a difficult problem," because "a person asking a question or giving a command[] typically relies heavily on context and the domain knowledge of the target person," whereas "machine-based queries" may be "highly structured and may not be inherently natural to the human user." *Id.* at 1:39-51. The inventions described and claimed in the '039 Patent overcome these challenges in various embodiments, for example by

obtaining information and presenting results in a natural manner, accommodating full or partial failure through the use of probabilistic and fuzzy reasoning. *Id.* at 2:12-32. This is true even when utterances "include[e] imperfect information such as, incomplete thoughts, incomplete sentences, incomplete phrases, slang terminology, repeated words, word variations, synonyms, or other imperfect information." *Id.* at 2:1-11.

45. The novel features of the invention are recited in the claims. For example, Claim 13 of the '039 Patent recites:

> A method of processing speech and non-speech communications, comprising:
>
> receiving the speech and non-speech communications;
>
> transcribing the speech and non-speech communications to create a speech-based textual message and a non-speech-based textual message;
>
> merging the speech-based textual message and the non-speech-based textual message to generate a query;
>
> searching the query for text combinations;
>
> comparing the text combinations to entries in a context description grammar;
>
> accessing a plurality of domain agents that are associated with the context description grammar;
>
> generating a relevance score based on results from comparing the text combinations to entries in the context description grammar;
>
> selecting one or more domain agents based on results from the relevance score;
>
> obtaining content that is gathered by the selected domain agents; and
>
> generating a response from the content, wherein the content is arranged in a selected order based on results from the relevance score.

'039 Patent at Claim 13.

U.S. PATENT NO. 9,495,957

46. On November 15, 2016, the U.S. Patent Office duly and legally issued the '957 Patent, entitled "Mobile Systems And Methods Of Supporting Natural Language Human-Machine Interactions." A true and correct copy of the '957 Patent is attached hereto as **Exhibit 7**.

47. Dialect is the owner and assignee of all right, title, and interest in and to the ’957 Patent, including the right to assert all causes of action arising under the ’957 Patent and the right to sue and obtain any remedies for past, present, or future infringement.

48. The ’957 Patent describes, among other things, a novel mobile system that identifies and uses context, prior information, domain knowledge, and user specific profile data to achieve a natural environment for users to submit natural language requests. ’957 Patent, Abstract. The claimed invention creates, stores and uses extensive personal profile information for each user to improve the reliability of determining the context of a request and presenting the expected results. *Id*. As the ’957 Patent explains, prior to its inventions, a machine’s ability to communicate with humans in a natural manner was a difficult technical problem in need of a technical solution. As described in the specification, under the existing systems and devices “verbal communications and machine processing of requests that are extracted from the verbal communications may be fundamentally incompatible,” because the existing systems and devices use requests that are “highly structured and may not be inherently natural to the human user.” *Id*. at 1:60-66. “Cognitive research on human interaction,” however, “shows that verbal communication, such as a person asking a question or giving a command, typically relies heavily on context and domain knowledge of the target person.” *Id*. at 1:57-60. The inventions described and claimed in the ’957 Patent overcome these challenges in various embodiments, for example by providing a system that uses context information determined from a command or request, comparing it against one or more words to create a score for context entries and generating a context stack to enable future requests. *Id.* 4:5-55.

49. The novel features of the invention are recited in the claims. For example, claim 1 of the ’957 Patent recites:

A system for processing a natural language utterance, the system including one or more processors executing one or more computer program modules which, when executed, cause the one or more processors to:

generate a context stack comprising context information that corresponds to a plurality of prior utterances, wherein the context stack includes a plurality of context entries;

receive the natural language utterance, wherein the natural language utterance is associated with a command or is associated with a request;

determine one or more words of the natural language utterance by performing speech recognition on the natural language utterance;

identify, from among the plurality of context entries, one or more context entries that correspond to the one or more words, wherein the context information includes the one or more context entries, wherein identifying the one or more context entries comprises:

comparing the plurality of context entries to the one or more words;

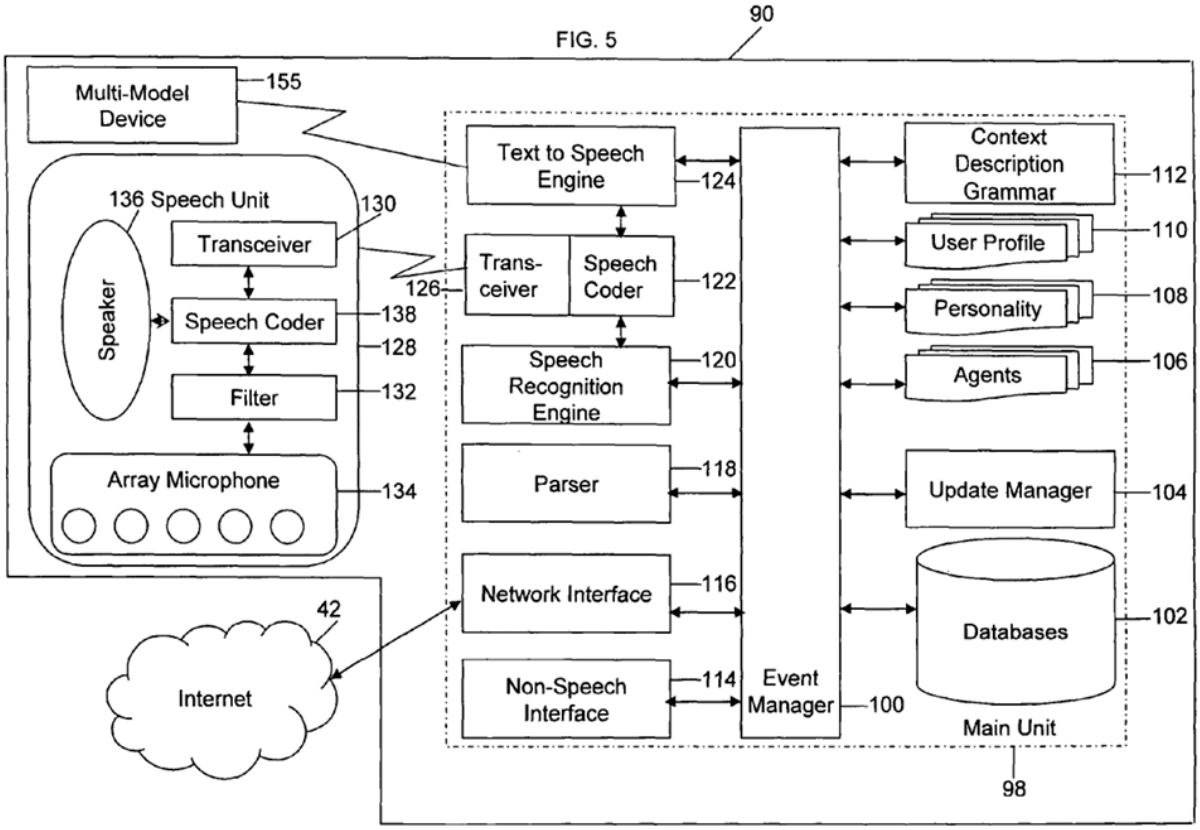
generating, based on the comparison, one or more rank scores for individual context entries of the plurality of context entries; and

identifying, based on the one or more rank scores, the one or more context entries from among the plurality of context entries; and

determine, based on the determined one or more words and the context information, the command or the request associated with the natural language utterance.

’957 Patent at claim 1.

50. Figure 5 of the ’957 Patent, reproduced below, shows a block diagram of an embodiment of the interactive natural language processing system.



'957 Patent, Fig. 5.

AMAZON'S USE OF THE PATENTED TECHNOLOGY

51. Amazon is a digital and retail conglomerate that provides *inter alia* internet cloud services, software products, digital products, and internet-enabled hardware in the United States and worldwide. In its 10-K filing for fiscal year 2022, Amazon reported consolidated total net sales of over $513 billion.

52. Among the products and services that Amazon makes, uses, sells, and/or offers to sell in the United States, and/or imports into the United States, are: the "Alexa Products"—Amazon's Alexa virtual assistant and offerings that include Alexa, including the Echo product line (such as Echo 1st Gen., Echo 2nd Gen., Echo Dot 1st Gen., Echo Dot 2nd Gen., Echo Dot 3rd Gen., Echo Dot Kids Edition, Echo Show 1st Gen., Echo Show 2nd Gen., Echo Show 5, Echo

Spot, Echo Plus 1st Gen., Echo Plus 2nd Gen., Echo Auto, and Echo Look), Amazon's Alexa apps, Music apps, and Shopping apps on a smartphone or other mobile device, Amazon's Alexa cloud, Alexa Voice Services, and Amazon.com website, and any other device, app, or instrumentality that includes, provides access to, or works with Alexa (such as Amazon Tap, Amazon Dash Wand, Echo Wall Clock—servers, network infrastructure, tablets, and internet of things ("IoT") devices (collectively, the "Accused Products"), which infringe the Asserted Patents as described in the counts below.

53. On information and belief, Amazon also provides third-party developers of automobiles, electronic hardware, and software with interfaces to the Alexa Products in order to encourage those developers to design, make, use, import into the United States, offer to sell, and sell products and services capable of being voice-controlled by the Accused Products.

FIRST COUNT
(Infringement of U.S Patent No. 7,693,720)

54. Dialect incorporates by reference the allegations set forth in Paragraphs 1–53 as though fully set forth herein.

55. The claims of the '720 Patent are valid and enforceable.

56. The claims of the '720 Patent are directed to patentable subject matter. Particularly, the '720 Patent is directed to a novel, tangible voice recognition system. The inventive, tangible claimed structures of the '720 Patent improve on the natural language processing of a natural language utterance by a user. The claimed inventions provide specific concrete solutions to the problem of speech recognition in existing systems.

57. In violation of 35 U.S.C. § 271(a) and without authority from Plaintiff, Amazon has directly infringed by making, using, offering for sale, selling, and/or importing into the Commonwealth of Virginia, this judicial district, and elsewhere in the United States products and

services that embody the invention disclosed and claimed in the ’720 Patent, including at least the Alexa Products, operating via tablets, smartphones, automobile infotainment systems, or other devices supporting in-vehicle Alexa functionality such as the Echo Auto, Alexa Auto SDK and Alexa Mobile Accessory Kit (collectively, the “Accused Automotive Products and Services”).

58. Each of the Accused Automotive Products and Services contains elements that are identical or equivalent to each claimed element of the patented invention pointed out by at least Claim 1 of the ’720 Patent.

59. Each of the Accused Automotive Products and Services comprises a mobile system responsive to a user generated natural language speech utterance.

60. For example, Amazon describes how the Alexa Auto SDK and Alexa Mobile Accessory Kit provide mobile solutions for hands-free interactions.[13]

Head Unit with Alexa Built-in

By building Alexa into your vehicle with the Alexa Auto SDK, your customers can interact with Alexa, hands-free, using the in-cabin microphone and speakers. You can also display digital content on the infotainment screen, like album artwork, audiobook covers, search results, and more.

Alexa Auto SDK »

[13] https://developer.amazon.com/en-US/alexa/alexa-auto/vehicles#experiences

Aftermarket Device with Alexa Built-in

Bring Alexa to the millions of vehicles already on the road by building Alexa into your automotive aftermarket device. The Alexa Mobile Accessory (AMA) Kit enables device makers to easily connect Bluetooth-enabled audio devices with Alexa using the Amazon Alexa App on Android and iOS devices.

Learn more about AMA »

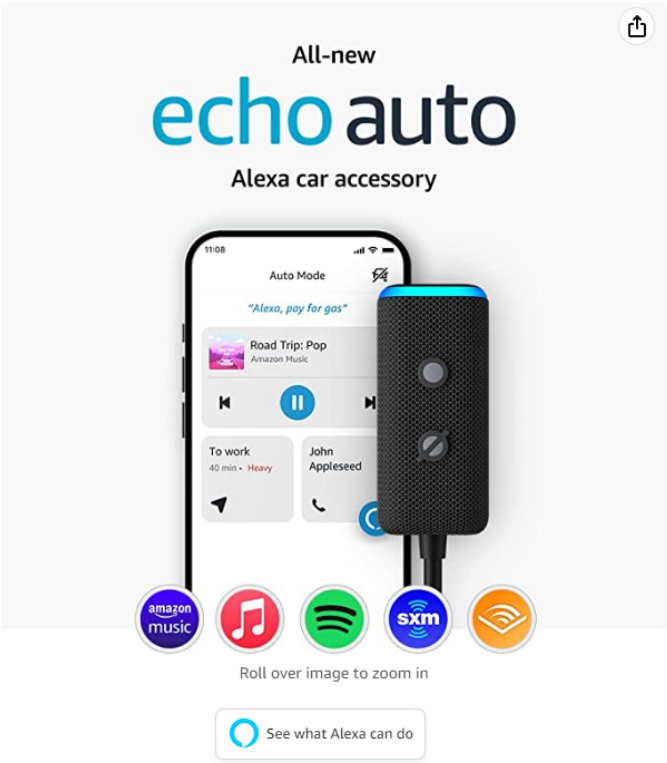
Amazon further describes the Alexa Auto SDK as enabling speech recognition, Automatic Speech Recognition and Natural Language Understanding.[14]

The Alexa Auto SDK enables core Alexa functionality, such as speech recognition and text-to-speech, and other capabilities such as streaming media, controlling smart home devices, notifications, weather reports, and tens of thousands of custom skills. It can be updated over-the-air and now includes the Local Voice Control extension, which ensures that customers get a reliable response from Alexa regardless of connectivity. On a normal commute, drivers lose connectivity or experience intermittent or transitional connectivity while driving in dense urban areas, through tunnels, in remote areas, or in parking garages. With the Local Voice Control extension to the Alexa Auto SDK, automakers embed a scaled-down version of Alexa's cloud-based Automatic Speech Recognition (ASR) and Natural Language Understanding (NLU) into the vehicle's infotainment system that increases resiliency of the cloud-based capabilities. This ensures that a customer can always ask, "Alexa, turn on the A/C," "Alexa, tune to 90.5 FM," or "Alexa, call Jackie," and get the appropriate response and action from Alexa.

Amazon further describes Echo Auto as enabling hands-free features.[15]

[14] https://developer.amazon.com/en-US/blogs/alexa/alexa-auto/2019/09/enabling-offline-access-to-alexa-in-vehicles-with-local-voice-control-extension-to-alexa-auto-sdk-v2-0

[15] https://developer.android.com/training/cars/apps/poi



Amazon Music, Apple Music, Spotify, and more, or listen live to radio stations. Use Follow Me Music to resume media playback.

- CALL AND MESSAGE WITH YOUR VOICE - Use your voice to make calls, reply to text messages, drop in on Alexa-enabled devices in your home or broadcast announcements.
- GET YOUR FAVORITE ENTERTAINMENT HANDS-FREE - Binge a hot new podcast, catch up on the news, or listen to best-selling Audible books.
- ALEXA CAN HELP WITH AMAZON ROADSIDE ASSISTANCE - Just say, "Alexa, call Roadside Assistance" and you'll be connected to an agent who can assess the situation and request pay-as-you-go services as needed.
- CONTROL YOUR SMART HOME FROM THE ROAD - Ask Alexa to set the thermostat, turn off the lights, check if your front door is locked, and more while you're away from home.
- DESIGNED TO PROTECT YOUR PRIVACY - Built with multiple layers of privacy controls, including a mic-off button and light indicator.

We want you to know

Echo Auto is designed to enable Alexa hands-free features, such as play music, make calls, answer text messages and more for cars that don't have a voice assistant built-in. Echo Auto connects to the Alexa app on your phone using your existing data plan and plays through your car's speakers via Bluetooth connection or Aux input. Carrier charges may apply.

Car and Phone compatibility - Check if your car is compatible and check if your phone is compatible. Note: Echo Auto is compatible with most cars with Bluetooth or Aux input.

Fast car charger - This product is Quick Charge 3.0 certified for fast charging.

Mounting your Echo Auto - The included adhesive car mount should not be placed on soft plastic, leather, or fabric surfaces. Choose a hard, durable surface away from the airflow.

61. Each of the Accused Automotive Products and Services comprises a speech unit connected to a computer device on a vehicle, wherein the speech unit receives a natural language speech utterance from a user and converts the received natural language speech utterance into an electronic signal.

62. For example, Amazon describes how Alexa can be built into a vehicle that comprises an in-cabin microphone and speakers connected to the automotive head unit.[16]

[16] https://developer.amazon.com/en-US/alexa/alexa-auto/vehicles#experiences

Head Unit with Alexa Built-in

By building Alexa into your vehicle with the Alexa Auto SDK, your customers can interact with Alexa, hands-free, using the in-cabin microphone and speakers. You can also display digital content on the infotainment screen, like album artwork, audiobook covers, search results, and more.

Alexa Auto SDK »

Amazon further describes how Alexa Mobile Accessory Kit manufacturers must support encoding for the received natural language utterance to convert the utterance into an electronic signal.[17]

Codec requirements for AMA

AMA supports the following codecs:

- Opus
- mSBC

The following table shows the recommended parameter settings for the supported codecs.

Note: An AMA enabled device must support at least one codec with the corresponding parameter settings.

AMA software requirements

Your Bluetooth-enabled accessory must support AMA transport and control protocols, including the following requirements:

- The accessory must encode the user's utterances with the encoding technologies described in the Codec Requirements for AMA. Then, the accessory sends the encoded utterances to the Alexa app on a phone by using the AMA Kit Protocol Specification.
- The accessory must implement functionality to receive Alexa responses from a mobile phone by using the AMA Kit Protocol Specification.
- The accessory must implement Over-the-Air (OTA) updates through the device companion app, and not through the Alexa app.
- The accessory must provide device information, such as serial number, name, supported transport, and device type to the AMA protocol stack.

[17] https://developer.amazon.com/en-US/docs/alexa/ama-kit/ama-kit-hw-security-reqs.html

Amazon further requires that the Alexa Mobile Accessory Kits include Bluetooth connections that can connect to a computing device on board the vehicle.[18]

Bluetooth and Bluetooth Low Energy (BLE) requirements for AMA

The Bluetooth implementation for a device must meet the following requirements to support AMA:

- Bluetooth v4.2 dual mode + LE compliance (with DLE support)
- Standard pairing, authentication, link key, and encryption operation
- BT Protocols support for SPP/A2DP/HFP /SDP/RFCOMM/SCO
- Accessories that implement mSBC must be MFi Certified

63. Each of the Accused Automotive Products and Services comprises a natural language speech processing system connected to the computer device on the vehicle, wherein the natural language speech processing system receives, processes, and responds to the electronic signal using data received from a plurality of domain agents.

64. For example, Amazon describes Alexa as recognizing speech and determining what the user wants before sending a request to invoke a skill that can fulfill the request.[19]

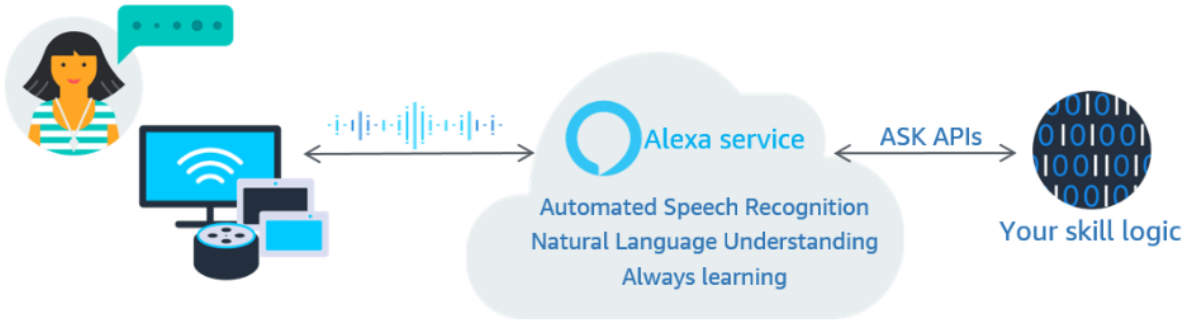
---

[18] *Id.*

[19] https://developer.amazon.com/en-US/docs/alexa/ask-overviews/what-is-the-alexa-skills-kit.html

A user accesses content in a skill by asking Alexa to *invoke* the skill. Alexa is always ready to invoke new skills. When a user says the wake word, "Alexa," and speaks to an Alexa-enabled device, the device streams the speech to the *Alexa service* in the cloud. Alexa recognizes the speech, determines what the user wants, and then sends a request to invoke the skill that can fulfill the request. The Alexa service handles the speech recognition and natural language processing. Your skill runs as a service on a cloud platform. Alexa communicates with your skill by using a request-response mechanism over the HTTPS interface. When a user invokes an Alexa skill, your skill receives a POST request containing a JSON body. The request body contains the parameters necessary for your skill to understand the request, perform its logic, and then generate a response.

The following diagram shows the voice-activated processing flow to invoke a skill with the Alexa service.



65. Each of the Accused Automotive Products and Services comprises a speech recognition engine that recognizes at least one of words or phrases from the electronic signal using at least the data received from the plurality of domain agents, wherein the data used by the speech recognition engine includes a plurality of dictionary and phrase entries that are dynamically updated based on at least a history of a current dialog and one or more prior dialogs associated with the user.

66. For example, Amazon describes how Alexa recognizes words or phrases using information from different Alexa intents and based on at least a history of the current dialog.[20]

[20] https://www.amazon.science/blog/how-alexa-is-learning-to-converse-more-naturally

Different Alexa intents (weather, search, directions) might assign the same data (San Francisco) to different slots (WeatherLocation, City, Town). Tracking references by carrying slots over from one intent to the next is crucial for maintaining a coherent conversation.

Amazon further describes Alexa tracking "previously provided information."[21]

Alexa Conversations is especially useful for use cases where the conversation can take a number of unanticipated paths as the user naturally talks to the skill, such as when a user chooses a movie, orders food, or makes a reservation. For example, when ordering a pizza, a user might do the following:

- Answer more than one question at once ("Medium, two toppings.")
- Ask questions and expect Alexa to track previously provided information ("How many people does that feed?")
- List values ("Pepperoni and green pepper.")
- Make a correction ("Make that a large.")

Further, Amazon describes how Alexa uses phrase slot types.[22]

[21] https://developer.amazon.com/en-US/docs/alexa/conversations/about-alexa-conversations.html
[22] https://developer.amazon.com/en-US/docs/alexa/custom-skills/slot-type-reference.html

Phrases

Phrase slot types allow for input from a user with fewer constraints on format and content. Make sure that your skill uses no more than one phrase slot per intent.

The following phrase slot type is supported.

Slot Type	Short Description	Examples	Supported Languages
`AMAZON.SearchQuery`	A search query like you might enter into a search engine, for example for searching a database.	Why is the sky blue What is the capital of Maryland Restaurants near me	• Arabic (SA) • English (AU) • English (CA) • English (IN) • English (UK) • English (US) • French (CA) • French (FR) • German (DE) • Hindi (IN) • Italian (IT) • Japanese (JP) • Portuguese (BR) • Spanish (ES) • Spanish (MX) • Spanish (US)

As another example, Amazon describes best practices for maintaining lists of key words and phrases.[23]

The most important rule is to design intents with a list of sample utterances that captures key words and phrases that can invoke the skill in a name- free manner. Keep your utterances specific to your skills use-cases and avoid targeting generic utterances. The Alexa ML/AI models are designed in part to route traffic to skills with a high correlation between skill functionality and target utterances.

[23] https://developer.amazon.com/en-US/docs/alexa/custom-skills/understand-name-free-interaction-for-custom-skills.html

Alexa Products further record previous interactions and maintain a history "to learn more about you as they listen."[24]

> Your interactions with Alexa are used to help make Alexa more accurate and useful to you over time and to build new Alexa experiences. But what does that actually mean?
>
> Our customers want Alexa to get smarter and more helpful to them every day. To do that, we use your Alexa requests to train our speech recognition and natural language understanding systems using machine learning. In more simple terms, we analyze data to teach Alexa to more accurately make decisions or predictions. For example, when you ask if it's raining outside, Alexa decides that you're asking about the weather —and when you say, "Alexa, play music," Alexa predicts what music you might like based on your previous requests.

67. Each of the Accused Automotive Products and Services comprises a parser that interprets the recognized words or phrases, wherein the parser uses at least the data received from the plurality of domain agents to interpret the recognized words or phrases, wherein the parser interprets the recognized words or phrases by determining a context for the natural language speech utterance; selecting at least one of the plurality of domain agents based on the determined context; and transforming the recognized words or phrases into at least one of a question or a command, wherein the at least one question or command is formulated in a grammar that the selected domain agent uses to process the formulated question or command.

68. For example, Amazon describes the Alexa Auto SDK as enabling speech recognition.[25]

---

[24] https://www.amazon.com/b?ie=UTF8&node=23608617011

[25] https://developer.amazon.com/en-US/blogs/alexa/alexa-auto/2019/09/enabling-offline-access-to-alexa-in-vehicles-with-local-voice-control-extension-to-alexa-auto-sdk-v2-0

The Alexa Auto SDK enables core Alexa functionality, such as speech recognition and text-to-speech, and other capabilities such as streaming media, controlling smart home devices, notifications, weather reports, and tens of thousands of custom skills. It can be updated over-the-air and now includes the Local Voice Control extension, which ensures that customers get a reliable response from Alexa regardless of connectivity. On a normal commute, drivers lose connectivity or experience intermittent or transitional connectivity while driving in dense urban areas, through tunnels, in remote areas, or in parking garages. With the Local Voice Control extension to the Alexa Auto SDK, automakers embed a scaled-down version of Alexa's cloud-based Automatic Speech Recognition (ASR) and Natural Language Understanding (NLU) into the vehicle's infotainment system that increases resiliency of the cloud-based capabilities. This ensures that a customer can always ask, "Alexa, turn on the A/C," "Alexa, tune to 90.5 FM," or "Alexa, call Jackie," and get the appropriate response and action from Alexa.

Amazon further describes how Alexa finds the most relevant skill for a given utterance..[26]

Alexa uses a two-step, scalable, and efficient neural shortlisting-reranking approach to find the most relevant skill for a given utterance. This post describes the first of those two steps, which relies on a neural model we call Shortlister. (I'll describe the second step in a follow-up post tomorrow.) Shortlister is a scalable and efficient architecture with a shared encoder, a personalized skill attention mechanism, and skill-specific classification networks. We outline this architecture in our paper "Efficient Large-Scale Neural Domain Classification with Personalized Attention", which we will present at the 56th Annual Meeting of the Association for Computational Linguistics (ACL 2018) in July.

Amazon further describes how Alexa formulates a question or command based on the grammar of different Alexa intents.[27]

[26] https://www.amazon.science/blog/the-scalable-neural-architecture-behind-alexas-ability-to-select-skills
[27] https://www.amazon.science/blog/how-alexa-is-learning-to-converse-more-naturally

Different Alexa intents (weather, search, directions) might assign the same data (San Francisco) to different slots (WeatherLocation, City, Town). Tracking references by carrying slots over from one intent to the next is crucial for maintaining a coherent conversation.

Amazon further describes how grammar has "been a tool in Alexa's NLU toolkit since well before the first Echo device shipped."[28]

[28] https://www.amazon.science/blog/tools-for-generating-synthetic-data-helped-bootstrap-alexas-new-language-releases

Rules of Grammar

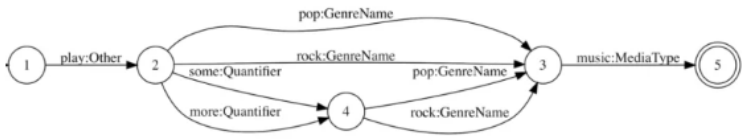
Grammars have been a tool in Alexa's NLU toolkit since well before the first Echo device shipped. A grammar is a set of rewrite rules for varying basic template sentences through word insertions, deletions, and substitutions.

Below is a very simple grammar, which models requests to play either pop or rock music, with or without the modifiers "more" and "some". Below the rules of the grammar is a diagram of a computational system (a finite-state transducer, or FST) that implements them.

```
pop = GenreName
rock = GenreName
music = MediaType
some = Quantifier
more = Quantifier

pop_rock = pop | rock
some_pop = (some | more)? pop_rock

INTENT Music.PlayMusicIntent = play some_pop music
```



A toy grammar, which can model requests to play pop or rock music, with or without the modifiers "some" or "more", and a diagram of the resulting finite-state transducer. The question mark indicates that the some_more variable is optional.

69. Each of the Accused Automotive Products and Services comprises an agent architecture that communicatively couples services of each of an agent manager, a system agent, the plurality of domain agents, and an agent library that includes one or more utilities that can be used by the system agent and the plurality of domain agents, wherein the selected domain agent uses the communicatively coupled services to create a response to the formulated question or command and format the response for presentation to the user.

70. For example, Amazon describes how the Alexa platform provides the infrastructure to support a variety of skills.[29]

The following diagram shows the voice-activated processing flow to invoke a skill with the Alexa service.

In addition to voice interaction, skills might include complementary visuals and touch interactions.

How does a user interact with a skill?

Every Alexa skill has a *voice interaction model* that defines the words and phrases users can say to make the skill do what they want. This model determines how users communicate with and control your skill. A voice user interface is similar to a graphical user interface in a traditional app. Instead of clicking buttons and selecting options from dialog boxes, users make their requests and respond to questions by voice. Often, the voice interaction is of a much shorter duration than interaction with an app. When a user asks questions and makes requests, Alexa employs the interaction model to interpret and translate the words into a specific request to the identified skill.

The following table compares a voice user interface skill with a graphical user interface app for making airline reservations.

Action	Voice user interface	Typical graphical user interface
Make a request	User says, "Alexa, I want to fly to Denver from Seattle."	User clicks on the app, and then selects the origination and destination airports. User scrolls through the list of airports to find **Seattle**, and then scrolls to find **Denver**.
Collect more information from the user	Alexa replies, "When would you like to travel?" and then waits for a reply.	App displays a calendar, and then waits for the user to select a date.
Provide needed information	User replies, "February first." The skill makes the reservation and waits for confirmation.	User opens the calendar, selects **February 1**, and then chooses **OK**. User clicks a button to complete the request, and then waits for confirmation.
Request complete	Alexa replies, "Your reservation from Seattle to Denver on Monday, February first is all set."	App displays the result of the request. The user closes the app.

[29] https://developer.amazon.com/en-US/docs/alexa/ask-overviews/what-is-the-alexa-skills-kit.html#:~:text=Skills%20are%20like%20apps%20for,to%20control%20cloud%2Dconnected%20devices

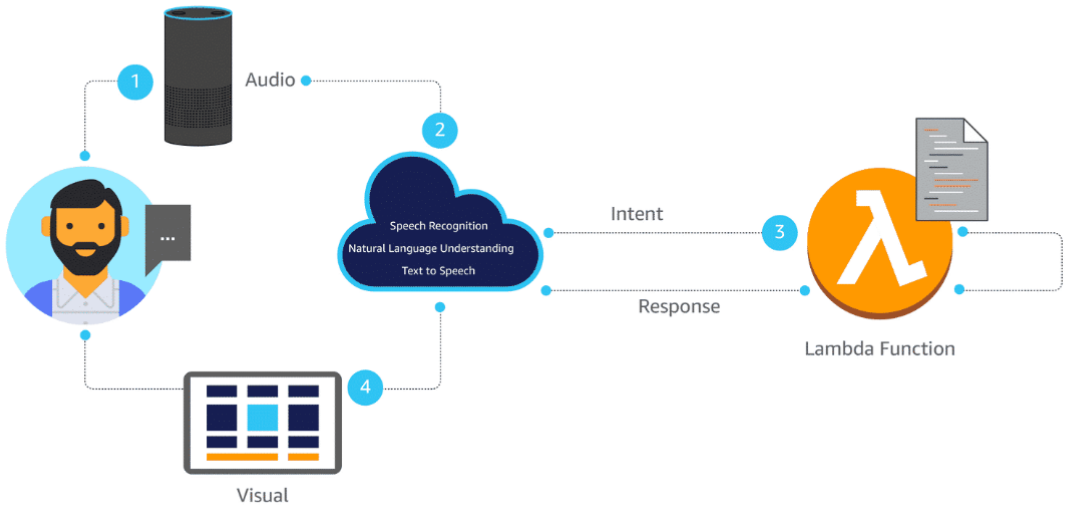
Further, Amazon describes how the Alexa infrastructure supports a user's interaction with an Alexa skill.[30]

How an Alexa skill works

The following simple workflow demonstrates how an Alexa skill works. In this example, the user invokes a simple Alexa skill called "Hello World."

1. To launch the skill, the user says, "Alexa, open Hello World." A complete statement or request to Alexa by the user is called an utterance.
2. The Alexa-enabled device sends the utterance to the Alexa service in the cloud. The Alexa service processes the utterance through automatic speech recognition to convert the utterance to text, and then through natural language understanding to recognize the intent of the text.
3. Alexa sends a JavaScript Object Notation (JSON) request to handle the intent to the resource compute service that hosts your skill (usually an AWS Lambda function). The resource compute service acts as the back end and executes your skill code to handle the intent. In this case, the server returns, "Welcome to the Hello World skill."

The following diagram demonstrates what happens when a user interacts with an Alexa skill. It assumes that you are using AWS Lambda's serverless compute service to host your skill code.



71. Amazon has long known about the '720 Patent.

72. Amazon has been aware of the '720 Patent from multiple communications, presentations, and licensing negotiations between VoiceBox Technologies and Amazon between at least 2011 and 2017 and from VoiceBox Technologies personnel hired by Amazon.

73. Amazon knew or should have known that Amazon's actions infringe one or more of the claims of the '720 Patent because Amazon has the technical expertise to understand the

[30] https://developer.amazon.com/it-IT/docs/alexa/workshops/build-an-engaging-skill/why-build/index.html

scope and content of the ’720 Patent, because Amazon is a major provider of voice recognition products and services, and because Amazon knows the design, function, and operation of the Accused Automotive Products and Services, as well as the nature and extent of their use by others. At a minimum, Amazon has knowledge of the ’720 Patent at least as of the filing of this Complaint.

74. Further, on information and belief, Amazon has actively induced and/or contributed to infringement of at least Claim 1 of the ’720 Patent in violation of at least 35 U.S.C. § 271(b), (c), and (f).

75. Users of the Accused Automotive Products and Services directly infringe at least Claim 1 of the ’720 Patent when they use the Accused Automotive Products and Services in the ordinary, customary, and intended way.

76. On information and belief, Amazon’s inducements in violation of 35 U.S.C. § 271(b) include, without limitation and with specific intent to encourage infringement, knowingly inducing businesses and consumers to use the Accused Automotive Products and Services within the United States in the ordinary, customary, and intended way by, directly or through intermediaries, supplying the Accused Automotive Products and Services to businesses and consumers within the United States, and instructing and encouraging such businesses and consumers to use the Accused Automotive Products and Services in the ordinary, customary, and intended way by activities related to selling, marketing, advertising, promotion, support, and distribution of the Accused Automotive Products and Services, which Amazon knew infringes at least Claim 1 of the ’720 Patent, or, alternatively, was willfully blind to the infringement.

77. On information and belief, Amazon’s inducements in violation of 35 U.S.C. § 271(b) further include, without limitation and with specific intent to encourage the infringement, knowingly inducing Amazon’s customers to commit acts of infringement with respect to the

Accused Automotive Products and Services within the United States, by, directly or through intermediaries, instructing and encouraging such customers to import, make, use, sell, offer to sell, or otherwise commit acts of infringement with respect to the Accused Automotive Products and Services in the United States by activities related to selling, marketing, advertising, promotion, support, and distribution of the Accused Automotive Products and Services, which Amazon knew infringes at least Claim 1 of the ’720 Patent, or, alternatively, was willfully blind to the infringement.

78. On information and belief, in violation of 35 U.S.C. § 271(c), Amazon’s contributory infringement further includes offering to sell or selling within the United States, or importing into the United States, components of the patented invention of and/or a material or apparatus for use in practicing at least Claim 1 of the ’720 Patent, constituting a material part of the invention. On information and belief, Amazon knows and has known the same to be especially made or especially adapted for use in an infringement of the ’720 Patent, and such components are not a staple article or commodity of commerce suitable for substantial noninfringing use.

79. On information and belief, in violation of 35 U.S.C. § 271(f)(1), Amazon’s infringement further includes without authority supplying or causing to be supplied in or from the United States all or a substantial portion of the components of the patented invention of at least Claim 1 of the ’720 Patent, where such components are uncombined in whole or in part, in such manner as to actively induce the combination of such components outside of the United States in a manner that would infringe the patent if such combination occurred within the United States.

80. On information and belief, in violation of 35 U.S.C. § 271(f)(2), Amazon’s infringement further includes without authority supplying or causing to be supplied in or from the United States components of the patented invention of at least Claim 1 of the ’720 Patent that are

especially made or especially adapted for use in the invention and not staple articles or commodities of commerce suitable for substantial noninfringing use, where such components are uncombined in whole or in part, knowing that such components are so made or adapted and intending that such components will be combined outside of the United States in a manner that would infringe the patent if such combination occurred within the United States.

81. Amazon is not licensed or otherwise authorized to practice the claims of the ’720 Patent.

82. Thus, by its acts, Amazon has injured Dialect and is liable to Dialect for directly and/or indirectly infringing one or more claims of the ’720 Patent, whether literally or under the doctrine of equivalents, including without limitation Claim 1.

83. As a result of Amazon’s infringement of the ’720 Patent, Dialect has suffered monetary damages, and seeks recovery, in an amount to be proven at trial, adequate to compensate for Amazon’s infringement, but in no event less than a reasonable royalty with interest and costs.

84. On information and belief, in addition to Amazon’s knowledge of the ’720 Patent as set forth above both prior to and as a result of the filing of this Complaint, Amazon has had, and continues to have, the specific intent to infringe, through its deliberate and intentional infringement or, alternatively, through its willfully blind disregard of the ’720 Patent by knowing there was a high probability of infringement but taking deliberate actions to avoid confirming that infringement. The filing of this action has also made Amazon aware of the unjustifiably high risk that its actions constituted and continue to constitute infringement of the ’720 Patent. On information and belief, discovery will reveal additional facts and circumstances from which Amazon’s knowledge and intent to infringe (or willful indifference), both before and after the filing of this action, may be inferred.

85. Accordingly, Amazon's infringement of the '720 Patent has also been and continues to be deliberate, intentional, and willful, and this is therefore an exceptional case warranting an award of enhanced damages and attorneys' fees and costs pursuant to 35 U.S.C. §§ 284 and 285.

86. Amazon's infringement of Dialect's rights under the '720 Patent will continue to damage Dialect, causing irreparable harm for which there is no adequate remedy at law, unless enjoined by this Court.

SECOND COUNT
(Infringement of U.S Patent No. 8,015,006)

87. Dialect incorporates by reference the allegations set forth in Paragraphs 1–86 as though fully set forth herein.

88. The claims of the '006 Patent are valid and enforceable.

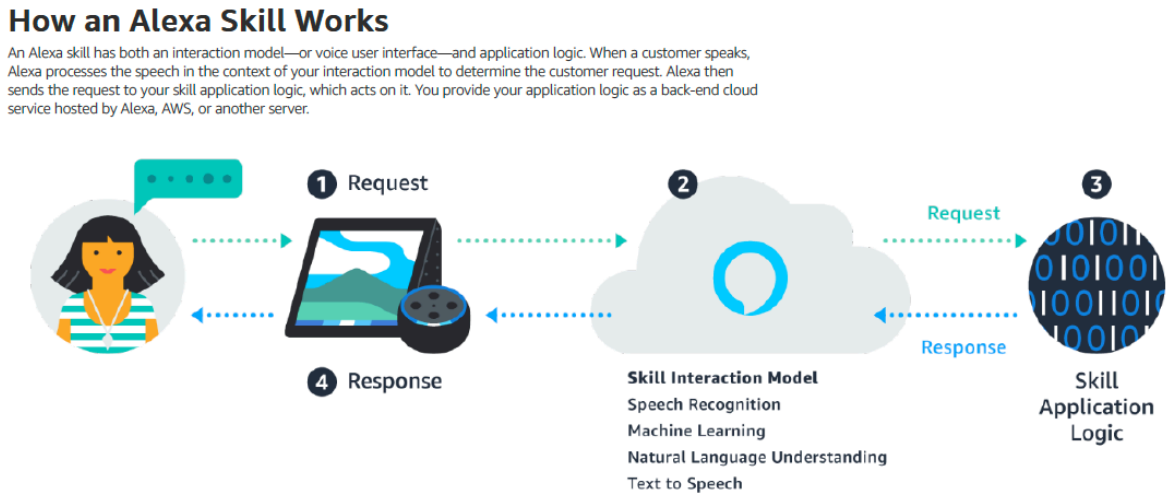
89. The claims of the '006 Patent are directed to patentable subject matter. Particularly, the '006 Patent is directed to a novel, tangible voice recognition system. The inventive, tangible claimed structures of the '006 Patent improve on the natural language processing of a natural language utterance by a user. The claimed inventions provide specific concrete solutions to the problem of speech recognition in existing systems.

90. In violation of 35 U.S.C. § 271(a) and without authority from Plaintiff, Amazon has directly infringed by making, using, offering for sale, selling, and/or importing into the Commonwealth of Virginia, this judicial district, and elsewhere in the United States products and services that embody the invention disclosed and claimed in the '006 Patent, including at least the Accused Products.

91. Each of the Accused Products contains elements that are identical or equivalent to each claimed element of the patented invention pointed out by at least Claim 5 of the '006 Patent.

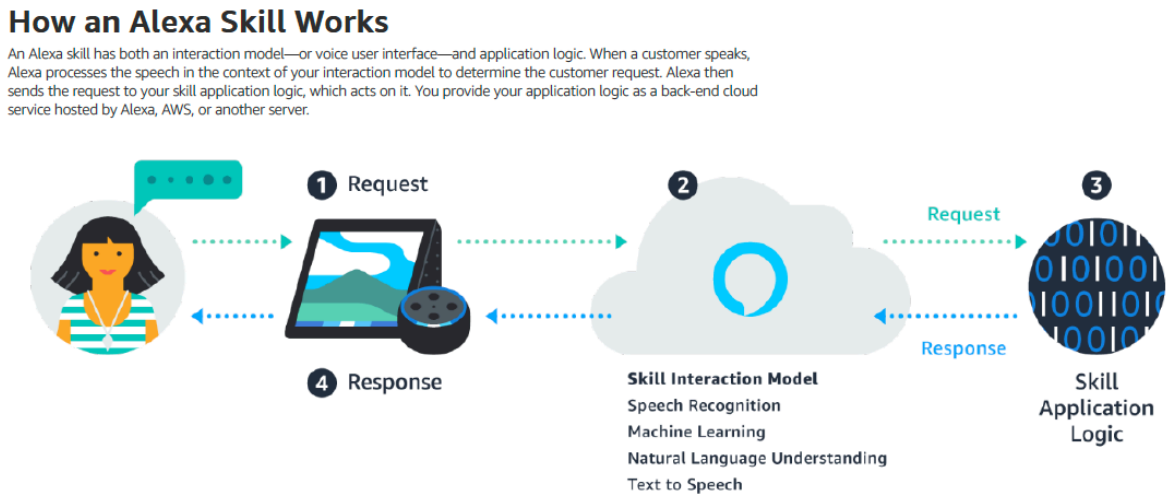
92. Each of the Accused Products comprises a method for processing natural language speech utterances with context-specific domain agents.

93. For example, Amazon describes Alexa as processing the speech or a customer.[31]



94. Each of the Accused Products comprises receiving, at a speech unit coupled to a processing device, a natural language speech utterance that contains a request.

95. For example, Amazon describes Alexa as receiving a user request at an Alexa device such as an Echo Dot or Echo Show.[32]



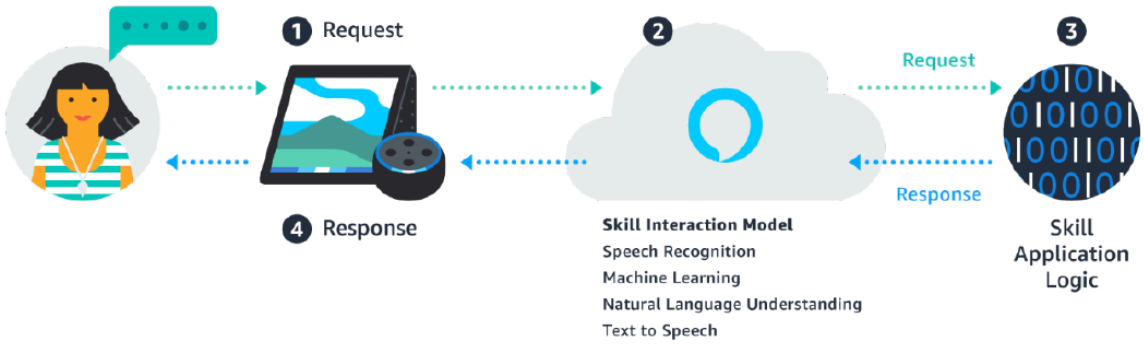
[31] https://developer.amazon.com/en-US/alexa/alexa-skills-kit/get-deeper/custom-skills
[32] *Id.*

96. Each of the Accused Products comprises recognizing, at a speech recognition engine coupled to the processing device, one or more words or phrases contained in the utterance using information in one or more dictionary and phrase tables.

97. For example, Amazon describes Alexa as processing the user request with speech recognition and natural language understanding.[33]

How an Alexa Skill Works

An Alexa skill has both an interaction model—or voice user interface—and application logic. When a customer speaks, Alexa processes the speech in the context of your interaction model to determine the customer request. Alexa then sends the request to your skill application logic, which acts on it. You provide your application logic as a back-end cloud service hosted by Alexa, AWS, or another server.



Amazon further describes Alexa as using a "wide range of sentences, phrases, and words that users are likely to say.[34]

Match a variety of utterances

Don't assume that users say the exact phrase that you anticipate for an intent. While the user might say, "Plan a trip," they also might say "Plan a vacation to Hawaii." To make sure your skill can respond to a variety of user utterances, provide a wide range of sentences, phrases, and words that users are likely to say.

[33] *Id.*
[34] https://developer.amazon.com/en-US/alexa/alexa-haus/voice-fundamentals

98. Each of the Accused Products comprises parsing, at a parser coupled to the processing device, information relating to the utterance to determine a meaning associated with the utterance and a context associated with the request contained in the utterance, wherein the parsed information includes the one or more recognized words or phrases.

99. For example, Amazon describes Alexa as using intents, slots, and sample utterances for a skill, where an intent represents an action that fulfills a request.[35]

Create Intents, Utterances, and Slots

Note: Sign in to the developer console to build or publish your skill.

To create your custom interaction model, you create the *intents*, *slots*, and *sample utterances* for your skill. An *intent* represents an action that fulfills a user's spoken request. Intents can optionally have arguments called *slots*. The *sample utterances* are set of likely spoken phrases mapped to the intents.

For details and code examples that show how you access intents and slots sent in user requests during skill runtime, see Handle Requests Sent by Alexa.

Amazon further describes Alexa intents assigning identified data to different slots based on the intent.[36]

[35] https://developer.amazon.com/en-US/docs/alexa/custom-skills/create-intents-utterances-and-slots.html
[36] https://www.amazon.science/blog/how-alexa-is-learning-to-converse-more-naturally

Different Alexa intents (weather, search, directions) might assign the same data (San Francisco) to different slots (WeatherLocation, City, Town). Tracking references by carrying slots over from one intent to the next is crucial for maintaining a coherent conversation.

Amazon's Alexa Machine Learning team further describes a system for referring back to slots in the context during a conversation.[37]

Abstract

In the slot-filling paradigm, where a user can refer back to slots in the context during a conversation, the goal of the contextual understanding system is to resolve the referring expressions to the appropriate slots in the context. In large-scale multi-domain systems, this presents two challenges - scaling to a very large and potentially unbounded set of slot values, and dealing with diverse schemas. We present a neural network architecture that addresses the slot value scalability challenge by reformulating the contextual interpretation as a decision to *carryover* a slot from a set of possible candidates. To deal with heterogenous schemas, we introduce a simple data-driven method for transforming the candidate slots. Our experiments show that our approach can scale to multiple domains and provides competitive results over a strong baseline.

Index Terms: spoken dialog, state tracking, anaphora resolution, contextual understanding.

[37] "Contextual Slot Carryover for Disparate Schemas," Naik (Amazon) et al. 2018.

100. Each of the Accused Products comprises formulating, at the parser, the request contained in the utterance in accordance with a grammar used by a domain agent associated with the determined context.

101. For example, Amazon describes Alexa as formulating requests contained in the user utterance in accordance with Alexa intents.[38]

Different Alexa intents (weather, search, directions) might assign the same data (San Francisco) to different slots (WeatherLocation, City, Town). Tracking references by carrying slots over from one intent to the next is crucial for maintaining a coherent conversation.

Amazon further describes how grammar has "been a tool in Alexa's NLU toolkit since well before the first Echo device shipped."[39]

[38] https://www.amazon.science/blog/how-alexa-is-learning-to-converse-more-naturally
[39] https://www.amazon.science/blog/tools-for-generating-synthetic-data-helped-bootstrap-alexas-new-language-releases

Rules of Grammar

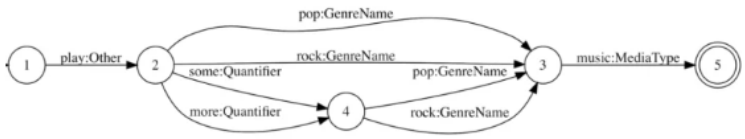
Grammars have been a tool in Alexa's NLU toolkit since well before the first Echo device shipped. A grammar is a set of rewrite rules for varying basic template sentences through word insertions, deletions, and substitutions.

Below is a very simple grammar, which models requests to play either pop or rock music, with or without the modifiers "more" and "some". Below the rules of the grammar is a diagram of a computational system (a finite-state transducer, or FST) that implements them.

```
pop = GenreName
rock = GenreName
music = MediaType
some = Quantifier
more = Quantifier

pop_rock = pop | rock
some_pop = (some | more)? pop_rock

INTENT Music.PlayMusicIntent = play some_pop music
```



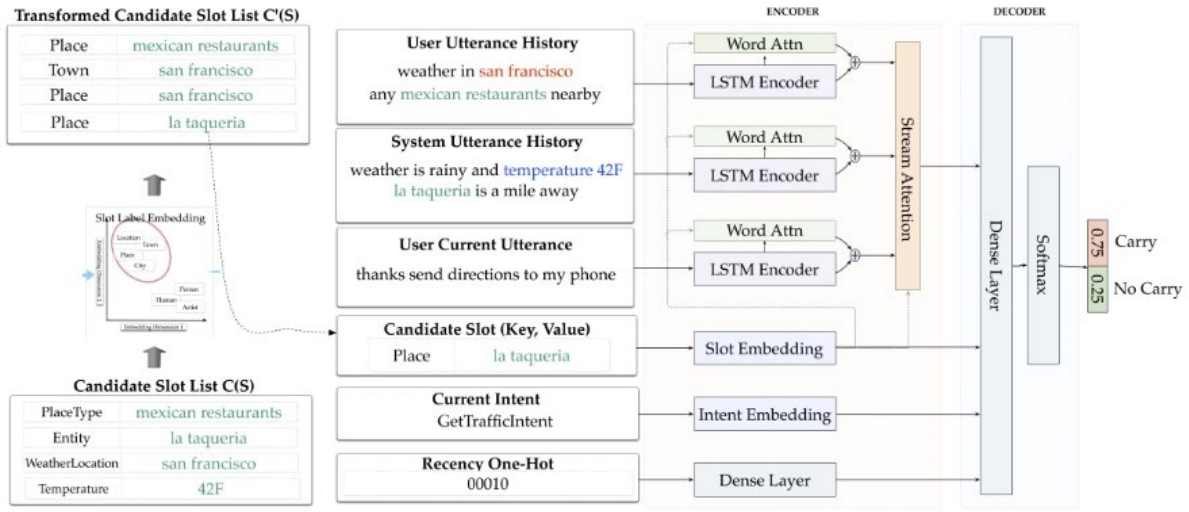
A toy grammar, which can model requests to play pop or rock music, with or without the modifiers "some" or "more", and a diagram of the resulting finite-state transducer. The question mark indicates that the some_more variable is optional.

102. Each of the Accused Products comprises determining one or more required values and one or more optional values associated with formulating the request in the grammar used by the domain agent.

103. For example, Amazon describes Alexa as supporting slots that are variables that a function acts upon.[40]

[40] https://www.amazon.science/blog/how-alexa-is-learning-to-converse-more-naturally

Today, Alexa analyzes the semantic content of utterances according to the categories domain, intent, and slot. "Domain" describes the type of application — or "skill" — that the utterance should invoke; for instance, *mapping* skills should answer questions about geographic distance. "Intent" describes the particular function the skill should execute, such as measuring driving distance. And "slots" are variables that the function acts upon, such as point of origin and destination.



Our system architecture

104. Each of the Accused Products comprises extracting one or more criteria and one or more parameters from one or more keywords contained in the one or more recognized words or phrases, wherein the parser extracts the one or more criteria and the one or more parameters using procedures sensitive to the determined context.

105. For example, Amazon describes Alexa as identifying and assigning keywords into different slots based on the intent.[41]

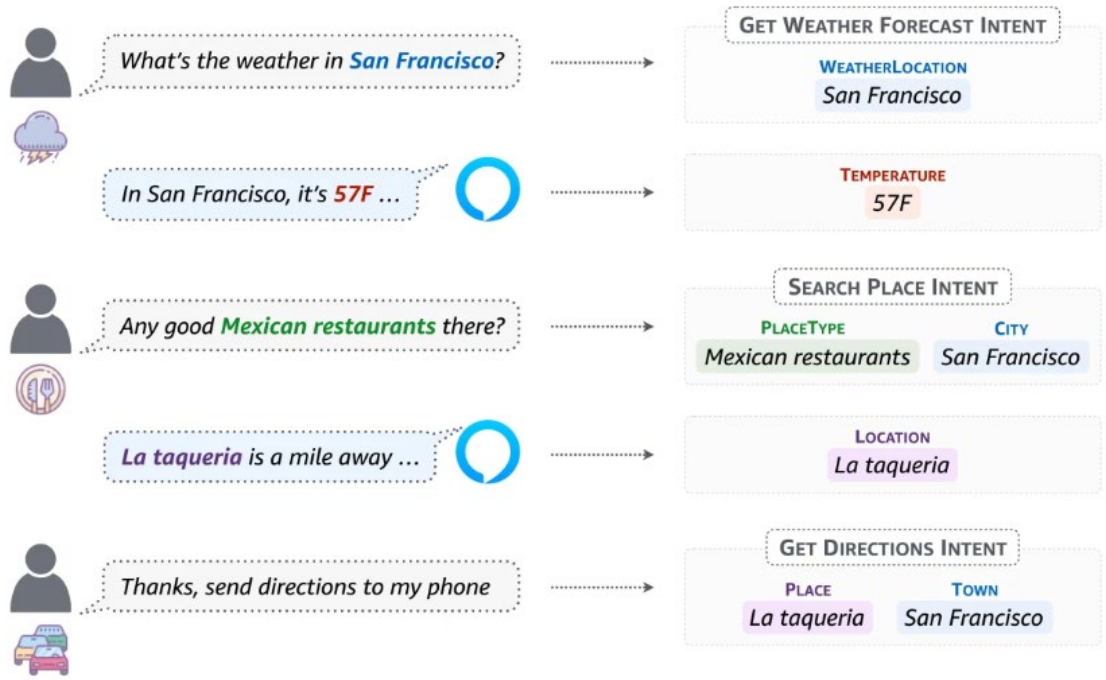
[41] https://www.amazon.science/blog/how-alexa-is-learning-to-converse-more-naturally

Different Alexa intents (weather, search, directions) might assign the same data (San Francisco) to different slots (WeatherLocation, City, Town). Tracking references by carrying slots over from one intent to the next is crucial for maintaining a coherent conversation.

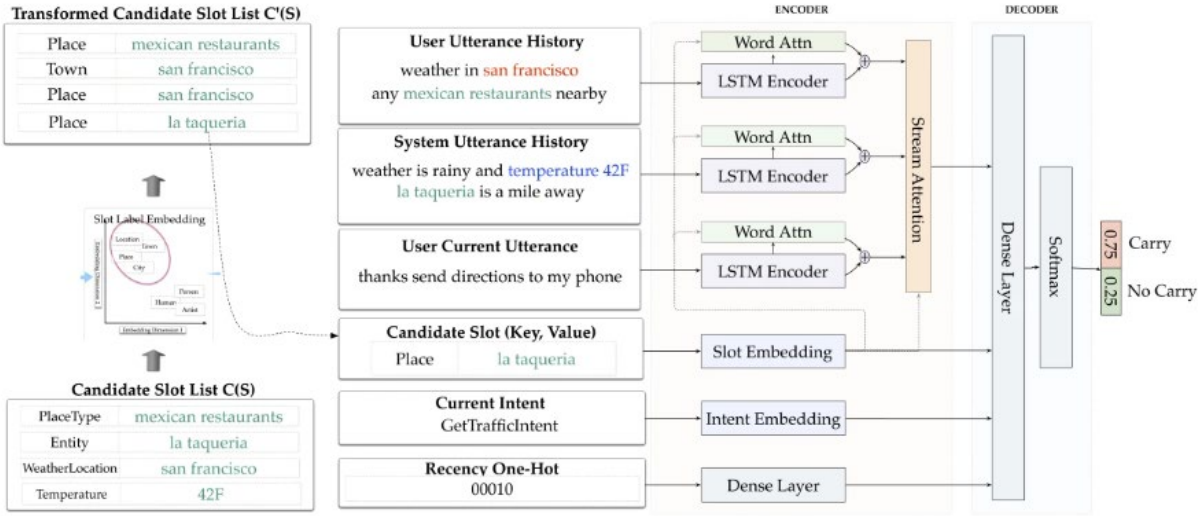
106. Each of the Accused Products comprises inferring one or more further criteria and one or more further parameters associated with the request using a dynamic set of prior probabilities or fuzzy possibilities.

107. For example, Amazon describes Alexa as inferring parameters for a request using the history of the current interaction.[42]

[42] https://www.amazon.science/blog/how-alexa-is-learning-to-converse-more-naturally



Different Alexa intents (weather, search, directions) might assign the same data (San Francisco) to different slots (WeatherLocation, City, Town). Tracking references by carrying slots over from one intent to the next is crucial for maintaining a coherent conversation.



Our system architecture

Amazon further states that their approach makes decisions "about slot values mentioned in context" and "the probability that any given carryover decision is the correct one."[43]

[43] *Id.*

> We make an important distinction between our approach and conventional dialogue state tracking, which maintains a probability distribution across all possible values that a given target slot can take on. Our system, by contrast, (i) cares only about slot values mentioned in context and (ii) makes independent decisions about the probability that any given carryover decision is the correct one. This helps us scale to the large number of slots across all Alexa skills.

108. Each of the Accused Products comprises transforming the one or more extracted criteria, the one or more extracted parameters, the one or more inferred criteria, and the one or more inferred parameters into one or more tokens having a format compatible with the grammar used by the domain agent, wherein the one or more tokens include all the required values and one or more of the optional values associated with formulating the request in the grammar used by the domain agent.

109. For example, Amazon describes Alexa as formulating requests based on the combination of data assigned to slots based on the intent.[44]

[44] https://www.amazon.science/blog/how-alexa-is-learning-to-converse-more-naturally

Different Alexa intents (weather, search, directions) might assign the same data (San Francisco) to different slots (WeatherLocation, City, Town). Tracking references by carrying slots over from one intent to the next is crucial for maintaining a coherent conversation.

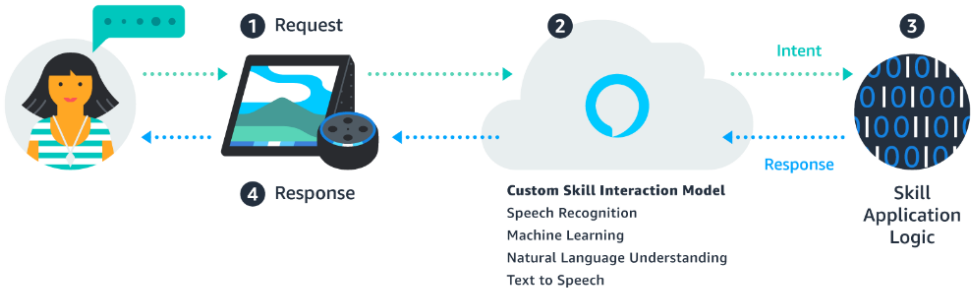
110. Each of the Accused Products comprises processing the formulated request with the domain agent associated with the determined context to generate a response to the utterance.

111. For example, Amazon describes Alexa as processing a skill interaction model and skill application logic to produce a response.[45]

[45] https://developer.amazon.com/en-US/alexa/alexa-skills-kit/get-deeper/custom-skills

How Custom Skills Work

When you build a custom skill, you define your own interaction model, or voice use interface, rather than relying on one provided by Alexa. Your custom interaction model specifies how customers' spoken input maps to the requests, or intents, that your skill can handle. When a customer speaks to your skill, Alexa uses your interaction model to determine the underlying customer intent, and sends it to your skill application logic.

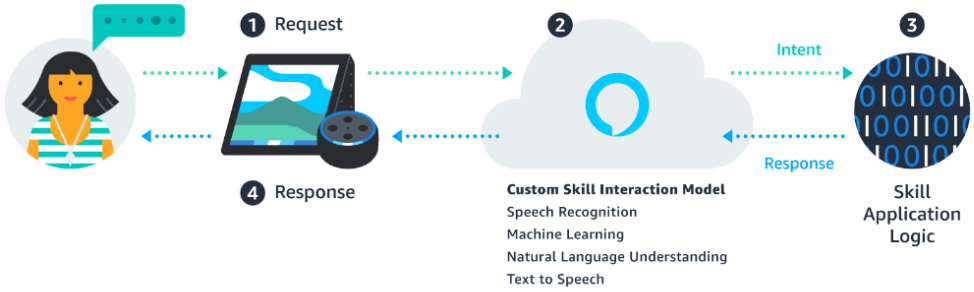


112. Each of the Accused Products comprises presenting the generated response to the utterance via the speech unit.

113. For example, Amazon describes Alexa as providing text to speech services to generate the response to the user's request.[46]

How Custom Skills Work

When you build a custom skill, you define your own interaction model, or voice use interface, rather than relying on one provided by Alexa. Your custom interaction model specifies how customers' spoken input maps to the requests, or intents, that your skill can handle. When a customer speaks to your skill, Alexa uses your interaction model to determine the underlying customer intent, and sends it to your skill application logic.



114. Amazon has long known about the '006 Patent.

115. Amazon has been aware of the '006 Patent from multiple communications, presentations, and licensing negotiations between VoiceBox Technologies and Amazon between at least 2011 and 2017 and from VoiceBox Technologies personnel hired by Amazon.

---

[46] *Id.*

116. Amazon knew or should have known that Amazon's actions infringe one or more of the claims of the '006 Patent because Amazon has the technical expertise to understand the scope and content of the '006 Patent, because Amazon is a major provider of voice recognition products and services, and because Amazon knows the design, function, and operation of the Accused Products, as well as the nature and extent of their use by others. At a minimum, Amazon has knowledge of the '006 Patent at least as of the filing of this Complaint.

117. Further, on information and belief, Amazon has actively induced and/or contributed to infringement of at least Claim 5 of the '006 Patent in violation of at least 35 U.S.C. § 271(b) and (c).

118. Users of the Accused Products directly infringe at least Claim 5 of the '006 Patent when they use the Accused Products in the ordinary, customary, and intended way.

119. On information and belief, Amazon's inducements in violation of 35 U.S.C. § 271(b) include, without limitation and with specific intent to encourage infringement, knowingly inducing businesses and consumers to use the Accused Products within the United States in the ordinary, customary, and intended way by, directly or through intermediaries, supplying the Accused Products to businesses and consumers within the United States, and instructing and encouraging such businesses and consumers to use the Accused Products in the ordinary, customary, and intended way by activities related to selling, marketing, advertising, promotion, support, and distribution of the Accused Products, which Amazon knew infringes at least Claim 5 of the '006 Patent, or, alternatively, was willfully blind to the infringement.

120. On information and belief, Amazon's inducements in violation of 35 U.S.C. § 271(b) further include, without limitation and with specific intent to encourage the infringement, knowingly inducing Amazon's customers to commit acts of infringement with respect to the

Accused Products within the United States, by, directly or through intermediaries, instructing and encouraging such customers to import, make, use, sell, offer to sell, or otherwise commit acts of infringement with respect to the Accused Products in the United States by activities related to selling, marketing, advertising, promotion, support, and distribution of the Accused Products, which Amazon knew infringes at least Claim 5 of the ’006 Patent, or, alternatively, was willfully blind to the infringement.

121. On information and belief, in violation of 35 U.S.C. § 271(c), Amazon’s contributory infringement further includes offering to sell or selling within the United States, or importing into the United States, components of the patented invention of and/or a material or apparatus for use in practicing at least Claim 5 of the ’006 Patent, constituting a material part of the invention. On information and belief, Amazon knows and has known the same to be especially made or especially adapted for use in an infringement of the ’006 Patent, and such components are not a staple article or commodity of commerce suitable for substantial noninfringing use.

122. Amazon is not licensed or otherwise authorized to practice the claims of the ’006 Patent.

123. Thus, by its acts, Amazon has injured Dialect and is liable to Dialect for directly and/or indirectly infringing one or more claims of the ’006 Patent, whether literally or under the doctrine of equivalents, including without limitation Claim 5.

124. As a result of Amazon’s infringement of the ’006 Patent, Dialect has suffered monetary damages, and seeks recovery, in an amount to be proven at trial, adequate to compensate for Amazon’s infringement, but in no event less than a reasonable royalty with interest and costs.

125. On information and belief, in addition to Amazon’s knowledge of the ’006 Patent as set forth above both prior to and as a result of the filing of this Complaint, Amazon has had, and

continues to have, the specific intent to infringe, through its deliberate and intentional infringement or, alternatively, through its willfully blind disregard of the ’006 Patent by knowing there was a high probability of infringement but taking deliberate actions to avoid confirming that infringement. The filing of this action has also made Amazon aware of the unjustifiably high risk that its actions constituted and continue to constitute infringement of the ’006 Patent. On information and belief, discovery will reveal additional facts and circumstances from which Amazon’s knowledge and intent to infringe (or willful indifference), both before and after the filing of this action, may be inferred.

126. Accordingly, Amazon’s infringement of the ’006 Patent has also been and continues to be deliberate, intentional, and willful, and this is therefore an exceptional case warranting an award of enhanced damages and attorneys’ fees and costs pursuant to 35 U.S.C. §§ 284 and 285.

127. Amazon’s infringement of Dialect’s rights under the ’006 Patent will continue to damage Dialect, causing irreparable harm for which there is no adequate remedy at law, unless enjoined by this Court.

THIRD COUNT
(Infringement of U.S Patent No. 8,140,327)

128. Dialect incorporates by reference the allegations set forth in Paragraphs 1–127 of the Complaint as though fully set forth herein.

129. The claims of the ’327 Patent are valid and enforceable.

130. The claims of the ’327 Patent are directed to patentable subject matter. Particularly, the ’327 Patent is directed to a novel, tangible voice recognition system. The inventive, tangible claimed structures of the ’327 Patent improve on the natural language processing of a natural

language utterance by a user. The claimed inventions provide specific concrete solutions to the problem of speech processing in existing systems.

131. In violation of 35 U.S.C. § 271(a) and without authority from Plaintiff, Amazon has directly infringed by making, using, offering for sale, selling, and/or importing into the Commonwealth of Virginia, this judicial district, and elsewhere in the United States products and services that embody the invention disclosed and claimed in the ’327 Patent, including at least the Accused Products.

132. Each of the Accused Products contains elements that are identical or equivalent to each claimed element of the patented invention pointed out by at least Claim 14 of the ’327 Patent.

133. Each of the ’327 Patent Accused Products comprises a system for filtering and eliminating noise from natural language speech utterances.

134. For example, Amazon describes Alexa devices as using the same technology as the Amazon Alexa Premium Voice Far-Field Development Kit.[47]

[47] https://developer.amazon.com/blogs/alexa/post/80facfd2-1176-4c4f-94ac-4c5c781011ca/amazon-alexa-premium-far-field-voice-development-kit#:~:text=The%20Amazon%20Alexa%20Premium%20Far%2DField%20Voice%20Development%20Kit%20uses,playing%20music%20at%20loud%20volumes

PRODUCT BRIEF

Amazon Alexa Premium Voice Far-Feild Development Kit

Engineered for Performance

The Amazon Alexa Premium Voice Far-Field Development Kit is designed to help commercial device manufacturers enable high quality far-field voice experiences in their products. This development kit gives manufacturers access to the audio technology that powers the Amazon Echo family of devices. It supports both tabletop 360˚ far-field voice activation applications, as well as applications that require voice-activation from one direction, such as devices with a display or placed near a wall. Both solutions incorporate Amazon's proprietary software and algorithm technology for "Alexa" wake word recognition, beam forming, noise reduction, and acoustic echo cancellation, and provides accurate far-field voice recognition in noisy environments and from long distances.

Turnkey Kit for Evaluation, Prototyping, and Engineering Development

This development kit simplifies the task of designing high-quality audio front ends for voice processing systems, allowing OEMs to focus on core product functionality. Included in the kit are two microphone array boards, a digital signal processor board which contains the audio processing hardware to optimize the quality of the voice signal, and a Raspberry Pi 3© single-board computer with the AVS Device SDK to handle interactions with the AVS cloud. Simply register the device, add a set of powered speakers, and download and install the software.

Features and Benefits

Premium Far-Field Performance

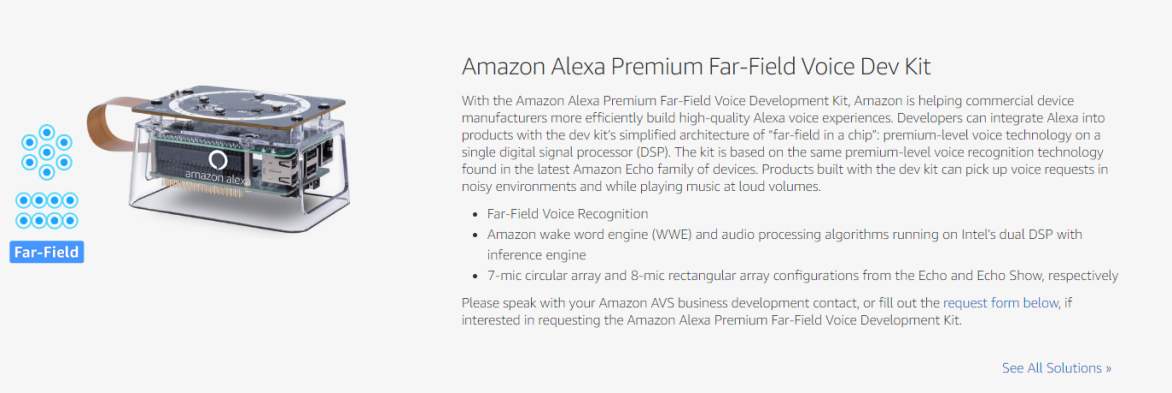
- 7- and 8- microphone arrays optimized for premium far-field audio performance in devices with 360˚ and 180˚ use cases
- Amazon audio processing and wake word trigger algorithms provide excellent voice recognition performance, even in noisy environmental conditions

Easy to Integrate

- Modular "far-field in a chip" architecture incorporates the audio processing algorithms and wake word trigger engine on the digital signal processor, to simplify integration with the device processor and software system stack
- Raspberry Pi 3© single-board computer can be replaced with production ready SoC for system prototyping and product development
- AVS Device SDK simplifies development of production-ready AVS client software

Amazon further describes this technology as allowing devices to pick up voice requests in noisy environments and while playing music at loud volumes.[48]

[48] https://developer.amazon.com/en-US/alexa/solution-providers/dev-kits/amazon-premium-voice



135. Each of the '327 Patent Accused Products comprises a microphone array configured to add one or more nulls to a beam pattern steered to point in a direction associated with a user speaking a natural language utterance to capture an input speech signal corresponding to the natural language utterance, wherein the one or more nulls notch out point or limited area noise sources from the input speech signal.

136. Amazon describes the technology of Alexa devices as using a microphone array that supports beam forming.[49]

[49] https://developer.amazon.com/blogs/alexa/post/80facfd2-1176-4c4f-94ac-4c5c781011ca/amazon-alexa-premium-far-field-voice-development-kit#:~:text=The%20Amazon%20Alexa%20Premium%20Far%2DField%20Voice%20Development%20Kit%20uses,playing%20music%20at%20loud%20volumes

PRODUCT BRIEF

Amazon Alexa Premium Voice Far-Feild Development Kit

Engineered for Performance

The Amazon Alexa Premium Voice Far-Field Development Kit is designed to help commercial device manufacturers enable high quality far-field voice experiences in their products. This development kit gives manufacturers access to the audio technology that powers the Amazon Echo family of devices. It supports both tabletop 360˚ far-field voice activation applications, as well as applications that require voice-activation from one direction, such as devices with a display or placed near a wall. Both solutions incorporate Amazon's proprietary software and algorithm technology for "Alexa" wake word recognition, beam forming, noise reduction, and acoustic echo cancellation, and provides accurate far-field voice recognition in noisy environments and from long distances.

Turnkey Kit for Evaluation, Prototyping, and Engineering Development

This development kit simplifies the task of designing high-quality audio front ends for voice processing systems, allowing OEMs to focus on core product functionality. Included in the kit are two microphone array boards, a digital signal processor board which contains the audio processing hardware to optimize the quality of the voice signal, and a Raspberry Pi 3© single-board computer with the AVS Device SDK to handle interactions with the AVS cloud. Simply register the device, add a set of powered speakers, and download and install the software.

Features and Benefits

Premium Far-Field Performance

- 7- and 8- microphone arrays optimized for premium far-field audio performance in devices with 360˚ and 180˚ use cases
- Amazon audio processing and wake word trigger algorithms provide excellent voice recognition performance, even in noisy environmental conditions

Easy to Integrate

- Modular "far-field in a chip" architecture incorporates the audio processing algorithms and wake word trigger engine on the digital signal processor, to simplify integration with the device processor and software system stack
- Raspberry Pi 3© single-board computer can be replaced with production ready SoC for system prototyping and product development
- AVS Device SDK simplifies development of production-ready AVS client software

Amazon's applied scientist in the Alexa Speech group specifically discussed the beam former for dealing with the noise robustness of the system.[50]

Your group has three papers at ICASSP this year. What are they about?

Two are about multichannel acoustic modeling. Right now, we are using seven microphones, and we beamform the signal to do acoustic modeling. We optimize the beam former based on speech quality, but we optimize the acoustic model based on word error rate.

With our new work, we are incorporating beamforming into the acoustic model, so we are jointly optimizing both the beam former and the acoustic model, and both of them are optimized toward reducing word error rate.

The other paper is about improving the conventional beamforming system. The challenge we are dealing with is the noise robustness of the system. Under clean conditions, ASR is already performing very well, but we still have problems dealing with noisy conditions. So the main idea is to train a teacher model on the clean data, then artificially add noise — music noise, audio playback noise, household noise.

Then we use the noisy data to train the student model, paired with the teacher's predictions from the clean data. We are hoping to create a student that performs similarly to the teacher but in the noisy domain.

Another example is in beamforming. You can do traditional beamforming where you combine the microphones with filters that affect the amplitude and phase to focus on sound from a specific direction. But if that's all you do, then you still have to figure out what direction to point it in. So we train it on the sound of interest. This is the speech that we want to listen to, so when it combines the microphones together, it can combine them in a way that optimizes for the specific direction of the speech and the overall recognition end to end.

Amazon further describes in its documentation of Alexa "Hardware Configurations" how beamforming results "in an increase in SNR and a reduction in reverberation in the audio signal."[51]

Beamforming

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 SNR and a reduction in reverberation in the audio signal from the desired direction that improves the accuracy of speech recognition systems, especially for far-field. For ASR, always use linear processing based versions of beamforming.

Amazon has further published papers about the SIR Beam Selector for Amazon Echo Devices.[52]

SIR Beam Selector for Amazon Echo Devices Audio Front-End

Xianxian Zhang, Trausti Kristjansson, Philip Hilmes

Amazon Inc., Sunnyvale, California, USA
{xianxi, traustik, philmes}@amazon.com

Amazon further describes how "Amazon scientists recommend car manufacturers use a signal-processing technique called beamforming that steers microphone arrays toward a voice signal."[53]

In design meetings, Amazon scientists collaborate with car manufacturers on techniques that allow the car's built-in system to focus on the person speaking to Alexa and cancel out acoustic echo and background noise.

For example, Amazon scientists recommend car manufacturers use a signal-processing technique called beamforming that steers microphone arrays toward a voice signal coming from a particular source, such as the driver, while suppressing audio interference from all other directions.

137. Each of the '327 Patent Accused Products comprises an adaptive filter coupled to the microphone array, wherein the adaptive filter is configured to: receive the input speech signal corresponding to the natural language utterance from the microphone array and compare

environmental noise to the input speech signal to set one or more parameters associated with the adaptive filter.

138. Amazon's applied scientist in the Alexa Speech group specifically discussed using an adaptive filter for echo cancellation changes based on what it's hearing.[54]

How do you combine signal processing and machine learning?

Acoustic echoes are handled really well by traditional signal-processing techniques, but we may use machine learning to contrive a control mechanism to figure out how to adapt that signal processing to a specific situation. So if you have an Echo device that's playing music, we use acoustic-echo cancellation filters that have learned what is playing, and then, listening to the microphones, it uses that reference to directly cancel the audio. But its performance is very dependent on how it's adapting the filter to cancel the audio.

So we use machine learning algorithms to understand what's going on in the environment and then adapt our canceling the sound. The adaptive filter for echo cancellation changes based on what it's hearing, but you can tell it to change or you can tell it, no, don't change. So we listen to understand, for instance, somebody's trying to talk to Alexa. Let's freeze the filter so that it doesn't change and try to cancel what the person is saying.

Amazon further describes the AEC algorithm that "estimates the acoustic echo path and the acoustic echo between the loudspeaker and microphone components," which "is then subtracted form the microphone signal."[55]

In speech recognition systems, the term "acoustic echo" refers to the signal that a loudspeaker plays and that a microphone captures from the loudspeaker. The acoustic echo is a source of interference for the ASR engine because the microphone simultaneously captures the echo and the user utterance. The goal of AEC is to remove the acoustic echo component from the microphone signal so that the ASR engine accurately understands the user utterance. The AEC algorithm adaptively estimates the acoustic echo path and the acoustic echo between the loudspeaker and microphone components. The estimated acoustic echo is then subtracted from the microphone signal to obtain a near echo-free microphone signal. An AEC-processed microphone signal should be free from acoustic echo.

139. Each of the '327 Patent Accused Products comprises an adaptive filter configured to use band shaping and notch filtering to remove narrow-band noise from the input speech signal received from the microphone array according to the one or more parameters.

---

[51] https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/audio-hardware-configurations.html
[52] https://assets.amazon.science/da/c2/71f5f9fa49f585a4616e49d52749/sir-beam-selector-for-amazon-echo-devices-audio-front-end.pdf
[53] https://www.amazon.science/news-and-features/the-science-behind-alexa-in-vehicles
[54] https://www.amazon.jobs/fr/landing_pages/icassp
[55] https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/audio-hardware-configurations.html

140. Amazon's applied scientist in the Alexa Speech group specifically discussed using an adaptive filter for echo cancellation changes.[56]

How do you combine signal processing and machine learning?

Acoustic echoes are handled really well by traditional signal-processing techniques, but we may use machine learning to contrive a control mechanism to figure out how to adapt that signal processing to a specific situation. So if you have an Echo device that's playing music, we use acoustic-echo cancellation filters that have learned what is playing, and then, listening to the microphones, it uses that reference to directly cancel the audio. But its performance is very dependent on how it's adapting the filter to cancel the audio.

So we use machine learning algorithms to understand what's going on in the environment and then adapt our canceling the sound. The adaptive filter for echo cancellation changes based on what it's hearing, but you can tell it to change or you can tell it, no, don't change. So we listen to understand, for instance, somebody's trying to talk to Alexa. Let's freeze the filter so that it doesn't change and try to cancel what the person is saying.

141. Each of the '327 Patent Accused Products comprises an adaptive filter configured to suppress cross-talk and environmentally caused echoes in the input speech signal received from the microphone array using adaptive echo cancellation.

142. Amazon's applied scientist in the Alexa Speech group specifically discussed using an adaptive filter for echo cancellation changes.[57]

How do you combine signal processing and machine learning?

Acoustic echoes are handled really well by traditional signal-processing techniques, but we may use machine learning to contrive a control mechanism to figure out how to adapt that signal processing to a specific situation. So if you have an Echo device that's playing music, we use acoustic-echo cancellation filters that have learned what is playing, and then, listening to the microphones, it uses that reference to directly cancel the audio. But its performance is very dependent on how it's adapting the filter to cancel the audio.

So we use machine learning algorithms to understand what's going on in the environment and then adapt our canceling the sound. The adaptive filter for echo cancellation changes based on what it's hearing, but you can tell it to change or you can tell it, no, don't change. So we listen to understand, for instance, somebody's trying to talk to Alexa. Let's freeze the filter so that it doesn't change and try to cancel what the person is saying.

[56] https://www.amazon.jobs/fr/landing_pages/icassp
[57] https://www.amazon.jobs/fr/landing_pages/icassp

Amazon further describes in its documentation of Alexa "Hardware Configurations" how "Acoustic Echo Cancellation" subtracts acoustic echo "from the microphone signal to obtain a near echo-free microphone signal."[58]

Acoustic Echo Cancellation (AEC)

In speech recognition systems, the term "acoustic echo" refers to the signal that a loudspeaker plays and that a microphone captures from the loudspeaker. The acoustic echo is a source of interference for the ASR engine because the microphone simultaneously captures the echo and the user utterance. The goal of AEC is to remove the acoustic echo component from the microphone signal so that the ASR engine accurately understands the user utterance. The AEC algorithm adaptively estimates the acoustic echo path and the acoustic echo between the loudspeaker and microphone components. The estimated acoustic echo is then subtracted from the microphone signal to obtain a near echo-free microphone signal. An AEC-processed microphone signal should be free from acoustic echo.

Amazon further describes how it recommends a "technique called acoustic echo cancellation" to car manufacturers for integration of Alexa.[59]

Another recommended technique called acoustic echo cancellation blocks out the noise from the car's speakers so that music played by the speaker does not interfere with the customer requests. Manufacturers can also deploy various noise reduction algorithms, which learn to separate a speaker's voice from other background noise.

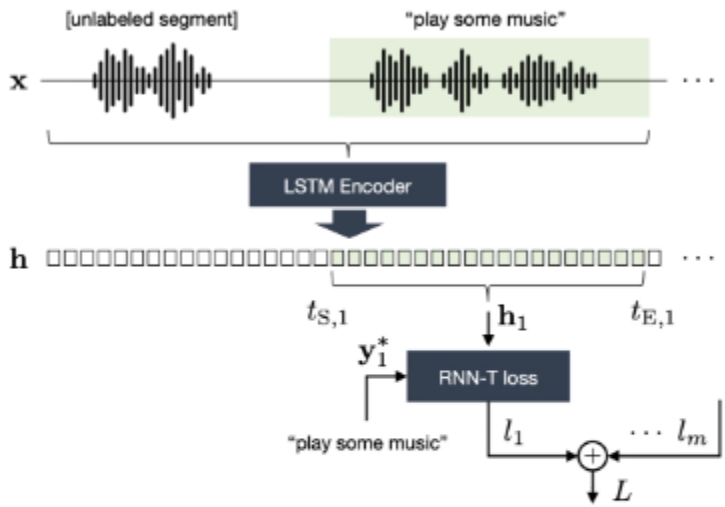
Inside the vehicle lab, Amazon scientists use software and acoustic tools to simulate road noise, music, and other acoustic elements as they test and validate Alexa integration with the car manufacturer's built-in systems.

143. Each of '327 Patent Accused Products comprises a speech coder arranged between the adaptive filter and a speech recognition engine, wherein the speech coder is configured to receive the input speech signal passed through the adaptive filter and use adaptive lossy audio compression to remove momentary gaps from the input speech signal and variable rate sampling

[58] https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/audio-hardware-configurations.html
[59] https://www.amazon.science/news-and-features/the-science-behind-alexa-in-vehicles

to compress and digitize the input speech signal, wherein the speech coder optimizes the adaptive lossy audio compression and the variable rate sampling to only preserve components in the input speech signal that will be input to the speech recognition engine.

144. For example, Amazon describes Alexa encoding segments of the audio stream.[60]



During the training of a context-aware ASR model, a long-short-term-memory (LSTM) encoder encodes both unlabeled and labeled segments of the audio stream, so the model can use the entire input audio to improve ASR accuracy. (From "Improving RNN-T ASR accuracy using context audio")

Furthermore, Amazon describes the compression techniques used by Alexa.[61]

On top of these innovations, however, we still had to develop some new compression techniques to get the RNN-T to run efficiently on-device. A neural network consists of simple processing nodes each of which is connected to several others. The connections between nodes have associated *weights*, which determine how much one node's output contributes to the computation performed by the next node.

[60] https://www.amazon.science/blog/on-device-speech-processing-makes-alexa-faster-lower-bandwidth
[61] *Id.*

145. Each of ’327 Patent Accused Products comprises a transceiver configured to communicate the digitized input speech signal from a buffer in the speech coder to the speech recognition engine at a rate that depends on available bandwidth associated with a communication link that connects the transceiver and the speech recognition engine.

146. For example, Amazon describes Alexa sending audio streams to the cloud in frames for off-device processing.[62]

> 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.

Amazon further describes how Alexa uses Opus to send captured audio to the Alexa Voice Service (AVS).[63]

Codec specifications

Send all captured audio to AVS in either PCM or Opus, and adhere to the following specifications:

Specification	PCM	Opus
Number of channels	Single channel (mono)	Single channel (mono)
Sample size	16-bit linear PCM (LPCM)	16 bit
Sample rate	16 kHz	16 kHz
Bitrate	256 Kpbs	32 Kpbs or 64 Kpbs, hard constant bitrate
Byte order	Little endian	Little endian
Frame size	10 ms	20 ms
DATA frame size	320 bytes	80 bytes (32 Kpbs) or 160 bytes (64 Kpbs)
Complexity	N/A	= 4

[62] https://www.amazon.science/blog/on-device-speech-processing-makes-alexa-faster-lower-bandwidth
[63] https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/speechrecognizer.html

Furthermore, the Opus codec is "highly tuned to allow it to quickly optimize between optimal fidelity for the available bandwidth."[64]

> Opus is highly tuned to allow it to quickly optimize between optimal fidelity for the available bandwidth. It also has a variable bitrate (VBR) mode that can range from 6 kbit/s to 510 kbit/s to help minimize bandwidth while maintaining consistent quality for applications that require that.

Furthermore, the Internet Engineering Task Force (IETF) RFC 6716—Definition of the Opus Audio Codec—specifies that an Opus encoder "should select which coding mode to use at run-time depending on the conditions."[65]

```
For a normal encoder where both the SILK and the CELT modules are
included, an optimal encoder should select which coding mode to use
at run-time depending on the conditions.  In the reference
implementation, the frame size is selected by the application, but
the other configuration parameters (number of channels, bandwidth,
mode) are automatically selected (unless explicitly overridden by the
application) depending on the following:

o  Requested bitrate

o  Input sampling rate

o  Type of signal (speech vs. music)

o  Frame size in use
```

147. Amazon has long known about the '327 Patent.

148. Amazon has been aware of the '327 Patent from multiple communications, presentations, and licensing negotiations between VoiceBox Technologies and Amazon between at least 2011 and 2017 and from VoiceBox Technologies personnel hired by Amazon.

149. Amazon knew or should have known that Amazon's actions infringe one or more of the claims of the '327 Patent because Amazon has the technical expertise to understand the

[64] https://www.wowza.com/blog/opus-codec-the-audio-format-explained
[65] https://datatracker.ietf.org/doc/html/rfc6716

scope and content of the ’327 Patent, because Amazon is a major provider of voice recognition products and services, and because Amazon knows the design, function, and operation of the Accused Products, as well as the nature and extent of their use by others. At a minimum, Amazon has knowledge of the ’327 Patent at least as of the filing of this Complaint.

150. Further, on information and belief, Amazon has actively induced and/or contributed to infringement of at least Claim 14 of the ’327 Patent in violation of at least 35 U.S.C. § 271(b), (c), and (f).

151. Users of the Accused Products directly infringe at least Claim 14 of the ’327 Patent when they use the Accused Products in the ordinary, customary, and intended way.

152. On information and belief, Amazon’s inducements in violation of 35 U.S.C. § 271(b) include, without limitation and with specific intent to encourage infringement, knowingly inducing businesses and consumers to use the Accused Products within the United States in the ordinary, customary, and intended way by, directly or through intermediaries, supplying the Accused Products to businesses and consumers within the United States, and instructing and encouraging such businesses and consumers to use the Accused Products in the ordinary, customary, and intended way by activities related to selling, marketing, advertising, promotion, support, and distribution of the Accused Products, which Amazon knew infringes at least Claim 14 of the ’327 Patent, or, alternatively, was willfully blind to the infringement.

153. On information and belief, Amazon’s inducements in violation of 35 U.S.C. § 271(b) further include, without limitation and with specific intent to encourage the infringement, knowingly inducing Amazon’s customers to commit acts of infringement with respect to the Accused Products within the United States, by, directly or through intermediaries, instructing and encouraging such customers to import, make, use, sell, offer to sell, or otherwise commit acts of

infringement with respect to the Accused Products in the United States by activities related to selling, marketing, advertising, promotion, support, and distribution of the Accused Products, which Amazon knew infringes at least Claim 14 of the ’327 Patent, or, alternatively, was willfully blind to the infringement.

154. On information and belief, in violation of 35 U.S.C. § 271(c), Amazon’s contributory infringement further includes offering to sell or selling within the United States, or importing into the United States, components of the patented invention of and/or a material or apparatus for use in practicing at least Claim 14 of the ’327 Patent, constituting a material part of the invention. On information and belief, Amazon knows and has known the same to be especially made or especially adapted for use in an infringement of the ’327 Patent, and such components are not a staple article or commodity of commerce suitable for substantial noninfringing use.

155. On information and belief, in violation of 35 U.S.C. § 271(f)(1), Amazon’s infringement further includes without authority supplying or causing to be supplied in or from the United States all or a substantial portion of the components of the patented invention of at least Claim 14 of the ’327 Patent, where such components are uncombined in whole or in part, in such manner as to actively induce the combination of such components outside of the United States in a manner that would infringe the patent if such combination occurred within the United States.

156. On information and belief, in violation of 35 U.S.C. § 271(f)(2), Amazon’s infringement further includes without authority supplying or causing to be supplied in or from the United States components of the patented invention of at least Claim 14 of the ’327 Patent that are especially made or especially adapted for use in the invention and not staple articles or commodities of commerce suitable for substantial noninfringing use, where such components are uncombined in whole or in part, knowing that such components are so made or adapted and

intending that such components will be combined outside of the United States in a manner that would infringe the patent if such combination occurred within the United States.

157. Amazon is not licensed or otherwise authorized to practice the claims of the ’327 Patent.

158. Thus, by its acts, Amazon has injured Dialect and is liable to Dialect for directly and/or indirectly infringing one or more claims of the ’327 Patent, whether literally or under the doctrine of equivalents, including without limitation Claim 14.

159. As a result of Amazon’s infringement of the ’327 Patent, Dialect has suffered monetary damages, and seeks recovery, in an amount to be proven at trial, adequate to compensate for Amazon’s infringement, but in no event less than a reasonable royalty with interest and costs.

160. On information and belief, in addition to Amazon’s knowledge of the ’327 Patent as set forth above both prior to and as a result of the filing of this Complaint, Amazon has had, and continues to have, the specific intent to infringe, through its deliberate and intentional infringement or, alternatively, through its willfully blind disregard of the ’327 Patent by knowing there was a high probability of infringement but taking deliberate actions to avoid confirming that infringement. The filing of this action has also made Amazon aware of the unjustifiably high risk that its actions constituted and continue to constitute infringement of the ’327 Patent. On information and belief, discovery will reveal additional facts and circumstances from which Amazon’s knowledge and intent to infringe (or willful indifference), both before and after the filing of this action, may be inferred.

161. Accordingly, Amazon’s infringement of the ’327 Patent has also been and continues to be deliberate, intentional, and willful, and this is therefore an exceptional case

warranting an award of enhanced damages and attorneys' fees and costs pursuant to 35 U.S.C. §§ 284 and 285.

162. Amazon's infringement of Dialect's rights under the '327 Patent will continue to damage Dialect, causing irreparable harm for which there is no adequate remedy at law, unless enjoined by this Court.

FOURTH COUNT
(Infringement of U.S Patent No. 8,195,468)

163. Dialect incorporates by reference the allegations set forth in Paragraphs 1–162 of the Complaint as though fully set forth herein.

164. The claims of the '468 Patent are valid and enforceable.

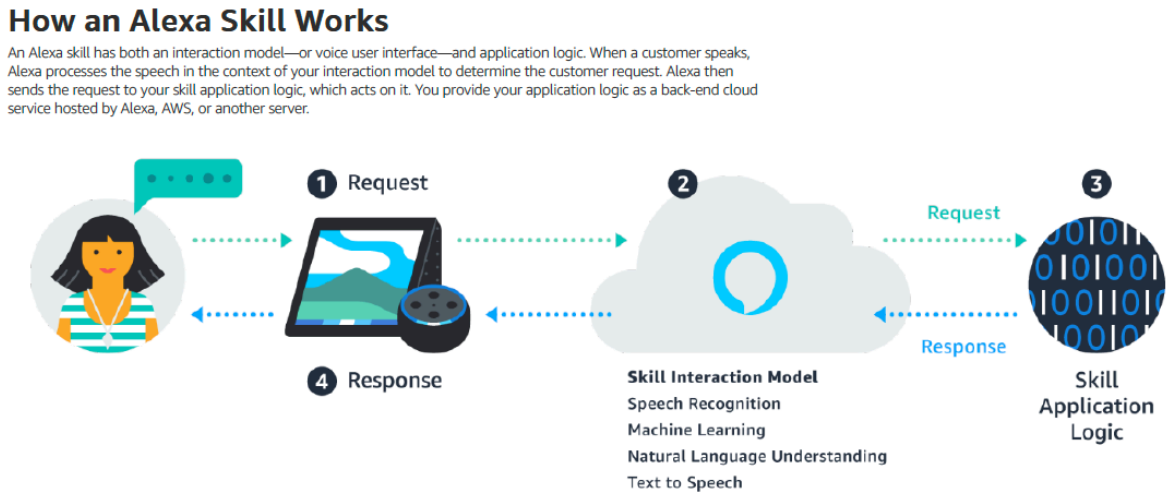
165. The claims of the '468 Patent are directed to patentable subject matter. Particularly, the '468 Patent is directed to a novel, tangible voice recognition system. The inventive, tangible claimed structures of the '468 Patent improve on the natural language recognition of a natural language utterance by a user. The claimed inventions provide specific concrete solutions to the problem of speech processing in existing systems.

166. In violation of 35 U.S.C. § 271(a) and without authority from Plaintiff, Amazon has directly infringed by making, using, offering for sale, selling, and/or importing into the Commonwealth of Virginia, this judicial district, and elsewhere in the United States products and services that embody the invention disclosed and claimed in the '468 Patent, including at least the Alexa Products, operating via tablets, smartphones, or other devices supporting non-speech inputs such as the Echo Show (collectively, the "Accused Multi-modal Products and Services")..

167. Each of the Accused Multi-modal Products and Services contains elements that are identical or equivalent to each claimed element of the patented invention pointed out by at least Claim 19 of the '468 Patent.

168. Each of the Accused Multi-modal Products and Services comprises a method for processing multi-modal natural language inputs.

169. For example, Amazon describes Alexa as processing natural language requests from a customer via multi-modal input systems such as the Echo Show.[66]



Amazon further describes the Alexa Multimodal Response Builder.[67]

[66] https://developer.amazon.com/en-US/alexa/alexa-skills-kit/start
[67] https://developer.amazon.com/en-US/docs/alexa/alexa-presentation-language/apl-authoring-tool.html#use-mmrb

Use the Multimodal Response Builder

The Multimodal Response Builder provides a guided experience to create a visual response in three steps:

1. Select from a set of templates designed to look good across a broad range of devices.
2. Customize the response by specifying the content to display within the template
3. Preview the response both in the developer console and on a device.

After you finish customizing the response, the Multimodal Response Builder generates a code example you can copy into a request handler in your skill to display the response. Rebuild the interaction model for your skill and then test your skill on a device or with the developer console simulator.

You can create a new document in the Multimodal Response Builder to get started, and then edit the document in the full authoring tool if you want to do more complex customizations.

To open the Multimodal Response Builder

1. In the developer console, open the skill for which you want to create this document.
2. In the left-hand navigation, click **Multimodal Responses**.
3. Click **Create with Response Builder**.
4. Step through the pages to complete the three steps
   - Select a template. The set of templates includes responsive templates as well as other visual designs.
   - Customize
   - Preview and test

Note: Populating the primary text or title fields of a template with long strings might create a poor visual experience. Make sure you test your response across different viewport and make sure all of your content displays as you expect.

Amazon further describes the Alexa Multi-Modal API for supporting multi-modal inputs.[68]

[68] https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/avs-apl-overview.html

Alexa Multi-Modal API Overview

Note: Learn how to improve your skills with APL with Build visually rich experiences using APL at the Alexa Learning Lab.

The Alexa Voice Service (AVS) provides a collection of APIs for Alexa Built-in devices with visual displays, such as smart TVs, which allow those devices to receive and present Alexa visual responses. Some of these APIs help device makers to publish the visual characteristics and functionality of their visual displays to enable Alexa Skills and services to respond with appropriate visual content for a device display.

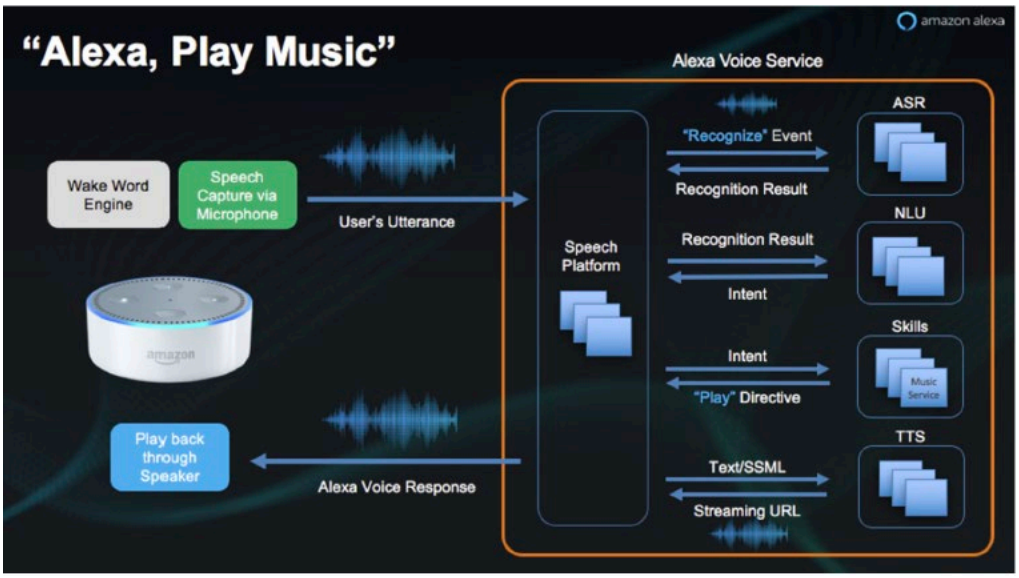
To add visual Alexa experiences to your device, you have several options for different levels of implementation:

- **Full APL integration** – To add a visual Alexa experience to your device, integrate the APL Core Library and implement an APL View Host in your code. See the APL Core Library repository and documentation to learn how to implement these two requirements.
- **Receive and render APL on a device** – To receive and render APL documents on your device screen, implement the Alexa.Presentation.APL namespace.
- **Publish device characteristics to Alexa** – To publish information about the visual characteristics of your device and its display for Alexa Skill developers, you can implement Alexa.Display, Alexa.Display.Window, and Alexa.InteractionMode.
- **AVS Device SDK** – Simplify your APL implementation by integrating the AVS Device SDK, which includes a sample app for voice-only devices and an IPC Server sample app for smart screen devices: AVS Device SDK repository.

170. Each of the Accused Multi-modal Products and Services comprises receiving a multi-modal natural language input at a conversational voice user interface, the multi-modal input including a natural language utterance and a non-speech input provided by a user, wherein a transcription module coupled to the conversational voice user interface transcribes the non-speech input to create a non-speech-based transcription.

171. For example, Amazon describes how the Alexa Voice Service processes a natural language request.[69]

[69] https://developer.amazon.com/fr/blogs/alexa/post/b015c5bd-2b01-4f7f-923e-dd34c998f3d0/avs-tech-series-the-basics-of-amazon-alexa-developer-tools-and-services



Amazon further discloses that Alexa converts the sound to text and that text is processed to understand what the user means.[70]

How Alexa Works

When I'm in my living room and ask Alexa about the weather, I trigger a complex system. First, the on-device chip detects the wake word (Alexa). Once detected, the microphones record what I'm saying and stream the sound for analysis in the cloud. At a high level, there are two phases to analyze the sound of my voice. First, Alexa converts the sound to text. This is known as Automatic Speech Recognition (ASR). Once the text is known, the second phase is to understand what I mean. This is Natural Language Understanding (NLU). The output of NLU is an Intent (what does the customer want) and associated parameters. In this example ("Alexa, what's the weather today ?"), the intent might be " `GetWeatherForecast` " and the parameter can be my postcode, inferred from my profile.

This whole process uses Artificial Intelligence heavily to transform the sound of my voice to phonemes, phonemes to words, words to phrases, phrases to intents. Based on the NLU output, Alexa routes the intent to a service to fulfill it. The service might be internal to Alexa or external, like one of the skills activated on my Alexa account. The fulfillment service processes the intent and returns a response as a JSON document. The document contains the text of the response Alexa must say.

Amazon further describes how Alexa Presentation Language processes multi-modal inputs including speech and non-speech elements.[71]

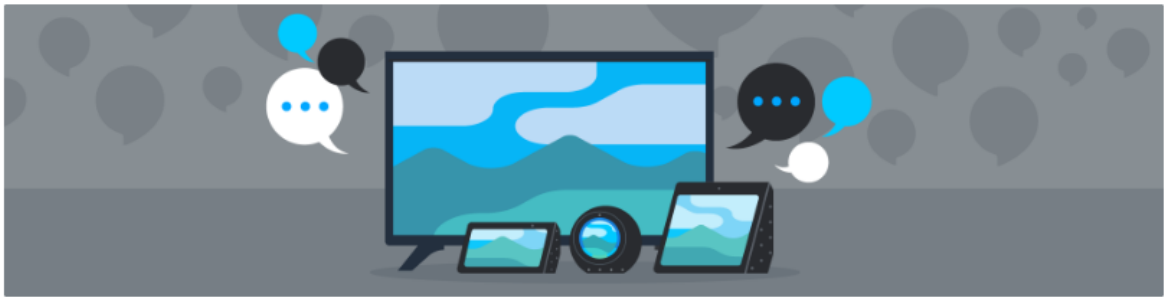
[70] https://aws.amazon.com/blogs/aws/majority-of-alexa-now-running-on-faster-more-cost-effective-amazon-ec2-inf1-instances/

[71] https://developer.amazon.com/en-US/blogs/alexa/alexa-skills-kit/2020/07/new-alexa-presentation-language-1-4

Build Interactive, Multimodal Experiences with Alexa Presentation Language (APL) 1.4

Arunjeet Singh Jul 22, 2020 Share:

Multimodal Intermediate Alexa Live



We are excited to announce the next version of the Alexa Presentation Language (APL) that enables you to easily build interactive visual experiences with new capabilities and improved tooling. APL 1.4 lets you add editable text boxes, drag and drop UI controls, and back navigation so customers can return to previous screens. You can utilize new Alexa responsive components and templates to quickly add visuals to your skills to enhance the voice experience, and live preview your APL documents in the authoring tool. Also, we added a major update to the Alexa Skills Kit (ASK) toolkit for Visual Studio Code (VS Code) that adds APL rendering and local debugging. Learn more about APL 1.4 in our technical documentation.

New APL Features

APL 1.4 supports user gestures and new components you can incorporate into your multimodal skills.

- New Gestures: Take advantage of new drag and drop UI controls such as single finger drag, swipe to delete, and long press gestures.
- Editable Text Boxes: Add new editable text boxes that allow customers to input a text response via touch or TV remote.
- Grid Layout Support: Use the GridSequence component to add a list of text and images to a fixed grid layout that scrolls in a single direction, vertically or horizontally. You can take advantage of the pre-built AlexaGridList responsive template that uses GridSequence.
- Back Navigation: Add back navigation so customers can easily return to previous screens in your skill with touch. Back navigation requires requesting the backstack extension from the APL extension framework.

Amazon further describes how Alexa captures and responds to key presses by the user.[72]

Keyboard handlers

APL provides support for keyboard events. You can define keyboard handlers to capture and respond to key presses. For details, see Keyboard Events and Handlers.

Amazon provides sample code of keyboard event handlers.[73]

handleKeyDown

An array of keyboard event handlers to run when the user presses a key on the keyboard or when a key auto-repeats. The `keyDown` event is generated whenever possible, not just for text entry. For example, pressing the "shift" key should generate a `keyDown` event.

The event generated has the form:

```
"event": {
  "source": {
    "type": "COMPONENT_TYPE", // The type of the component (e.g., "Pager", "TouchWrapper")
    "handler": "KeyDown",
    ...                       // Component source properties
  },
  "keyboard": {
    "altKey":   Boolean
    "code":     String,
    "ctrlKey":  Boolean,
    "key":      String,
    "metaKey":  Boolean,
    "repeat":   Boolean,
    "shiftKey": Boolean
  }
}
```

The `event.source.type` property contains the name of the component, such as "`TouchWrapper`" or "`ScrollView`". Refer to Event source for a description of `event.source` properties.

For more details on the `keyboard` property, see Keyboard Event Handlers.

The `handleKeyDown` event handler runs in normal mode.

172. Each of the Accused Multi-modal Products and Services comprises identifying the user that provided the multi-modal input.

173. For example, Amazon describes Alexa Voice ID, which recognizes users when they speak.[74]

[72] https://developer.amazon.com/en-US/docs/alexa/alexa-presentation-language/apl-for-screen-devices.html
[73] https://developer.amazon.com/en-US/docs/alexa/alexa-presentation-language/apl-actionable-component.html
[74] https://www.amazon.com/gp/help/customer/display.html?nodeId=GYCXKY2AB2QWZT2X

Digital Services and Device Support › Alexa Settings › Alexa Profile Settings ›

What Is Alexa Voice ID?

Alexa voice ID helps Alexa recognize you when you speak and provide a personalized experience.

You can create an Alexa voice ID for a personalized experience when Alexa recognizes your voice. With Alexa voice ID, Alexa can call you by name and provide enhanced personalization. To set up Alexa voice ID, go to Create an Alexa Voice ID.

Amazon further describes how Voice ID is enabled for Alexa skills.[75]

Request Recognized Speaker Contact Information for Use in Your Skill

Note: Sign in to the developer console to build or publish your skill.

If your skill implements personalization for recognized speakers, and the Alexa user is a recognized speaker with a voice ID and **Personalize skills** enabled in the Alexa app, your skill can request the user's permission to access the contact information associated with the user's voice ID. The contact information includes the name — either the full name or given name (first name) — and mobile number.

After receiving consent from the user, your skill can use the contact information to enable helpful experiences like the following:

- Personalize responses to the user, such as *"Welcome back, <first name>!"*
- Make a booking or appointment with your service on behalf of the user
- Send detailed or follow-up information to the user by text

You access this contact information by using the Person Profile API.

Note

For sample code, see the Person Profile Demo in GitHub.

For details about personalization, see Add Personalization to Your Skill.

If you want to enable the user to verbally consent to share their contact information with your skill, see Use Voice-Forward Consent in Your Alexa Skill.

[75] https://developer.amazon.com/en-US/docs/alexa/custom-skills/request-recognized-speaker-contact-information.html

174. Each of the Accused Multi-modal Products and Services comprises creating a speech-based transcription of the natural language utterance using a speech recognition engine and a semantic knowledge-based model, wherein the semantic knowledge-based model includes a personalized cognitive model derived from one or more prior interactions between the identified user and the conversational voice user interface, a general cognitive model derived from one or more prior interactions between a plurality of users and the conversational voice user interface, and an environmental model derived from an environment of the identified user and the conversational voice user interface.

175. For example, Amazon describes how Alexa converts the sound into text.[76]

> **How Alexa Works**
>
> When I'm in my living room and ask Alexa about the weather, I trigger a complex system. First, the on-device chip detects the wake word (Alexa). Once detected, the microphones record what I'm saying and stream the sound for analysis in the cloud. At a high level, there are two phases to analyze the sound of my voice. First, Alexa converts the sound to text. This is known as Automatic Speech Recognition (ASR). Once the text is known, the second phase is to understand what I mean. This is Natural Language Understanding (NLU). The output of NLU is an Intent (what does the customer want) and associated parameters. In this example ("Alexa, what's the weather today ?"), the intent might be "`GetWeatherForecast`" and the parameter can be my postcode, inferred from my profile.
>
> This whole process uses Artificial Intelligence heavily to transform the sound of my voice to phonemes, phonemes to words, words to phrases, phrases to intents. Based on the NLU output, Alexa routes the intent to a service to fulfill it. The service might be internal to Alexa or external, like one of the skills activated on my Alexa account. The fulfillment service processes the intent and returns a response as a JSON document. The document contains the text of the response Alexa must say.

Amazon further describes how Voice ID allows Alexa to personalize responses based on prior interactions with the identified user.[77]

---

[76] https://aws.amazon.com/blogs/aws/majority-of-alexa-now-running-on-faster-more-cost-effective-amazon-ec2-inf1-instances/

[77] https://www.amazon.com/gp/help/customer/display.html?nodeId=GYCXKY2AB2QWZT2X

Digital Services and Device Support › Alexa Settings › Alexa Profile Settings ›

What Is Alexa Voice ID?

Alexa voice ID helps Alexa recognize you when you speak and provide a personalized experience.

You can create an Alexa voice ID for a personalized experience when Alexa recognizes your voice. With Alexa voice ID, Alexa can call you by name and provide enhanced personalization. To set up Alexa voice ID, go to Create an Alexa Voice ID.

Amazon further describes Alexa's model as including training from a large generalized corpus.[78]

Our overarching strategy for Alexa's AI has been to combine machine learning (ML) — in particular, deep learning — with the large-scale data and computational resources available through AWS. But these performance improvements are the result of research on a variety of specific topics that extend deep learning, including

- semi-supervised learning, or using a combination of unlabeled and labeled data to improve the ML system;
- active learning, or the learning strategy where the ML system selects more-informative samples to receive manual labels;
- large-scale distributed training, or parallelizing ML-based model training for efficient learning on a large corpus; and
- context-aware modeling, or using a wide variety of information — including the type of device where a request originates, skills the customer uses or has enabled, and past requests — to improve accuracy.

Amazon further describes how the Intent Request History REST API can "get the aggregated and anonymized transcriptions of user speech data" from "customer interactions over the past 30 days."[79]

[78] https://www.amazon.science/blog/alexa-at-five-looking-back-looking-forward
[79] https://developer.amazon.com/en-US/docs/alexa/smapi/intent-request-history.html

Use the Intent Request History REST API to get the aggregated and anonymized transcriptions of user speech data and intent request details for a skill. The API returns data from customer interactions over the past 30 days, updated each day. You can use the data to learn how users interact with your skill to identify improvements to your interaction model.

This API supports text search and sort on the results. Data is available for skills that have at least 10 unique users per locale in a day. Alexa returns data for those locales that meet the threshold only.

Amazon further describes Alexa Hunches using information from the user's environment in supporting skill.[80]

4. More context-aware and proactive

Today, through an optional feature called Hunches, Alexa can learn how you interact with your smart home and suggest actions when she senses that devices such as lights, locks, switches, and plugs are not in the states that you prefer. We are currently expanding the notion of Hunches to include another Alexa feature called Routines. If you set your alarm for 6:00 a.m. every day, for example, and on waking, you immediately ask for the weather, Alexa will suggest creating a Routine that sets the weekday alarm to 6:00 and plays the weather report as soon as the alarm goes off.

Earlier this year, we launched Alexa Guard, a feature that you can activate when you leave the house. If your Echo device detects the sound of a smoke alarm, a carbon monoxide alarm, or glass breaking, Alexa Guard sends you an alert. Guard's acoustic-event-detection model uses multitask learning, which reduces the amount of labeled data needed for training and makes the model more compact.

This fall, we will begin previewing an extended version of Alexa Guard that recognizes additional sounds associated with activity, such as footsteps, talking, coughing, or doors closing. Customers can also create Routines that include Guard — activating Guard automatically during work hours, for instance.

176. On information and belief, each of the Accused Multi-modal Products and Services comprises merging the speech-based transcription and the non-speech-based transcription to create a merged transcription.

[80] https://www.amazon.science/blog/alexa-at-five-looking-back-looking-forward

177. Each of the Accused Multi-modal Products and Services comprises identifying one or more entries in a context stack matching information contained in the merged transcription.

178. For example, Amazon describes how Alexa compares text combinations to a list of sample utterances that can invoke a skill in a name-free manner.[81]

[81] https://developer.amazon.com/en-US/docs/alexa/custom-skills/understand-name-free-interaction-for-custom-skills.html

The most important rule is to design intents with a list of sample utterances that captures key words and phrases that can invoke the skill in a name- free manner. Keep your utterances specific to your skills use-cases and avoid targeting generic utterances. The Alexa ML/AI models are designed in part to route traffic to skills with a high correlation between skill functionality and target utterances.

Guidelines

To help ensure that Alexa directs to your skill only those customer requests, questions, or statements that can be supported by the skill, build sample utterances that capture your skill's functionality. Provide sample utterances that you predict customers can use to interact with the skill naturally. For a better user experience, provide a large variety of sample utterances written in different forms. For example, name-free utterances for a horoscope skill might include:

- *"What is the horoscope for gemini?"*
- *"Today's horoscope"*
- *"I want to know my horoscope"*
- *"What does my star sign say?"*

Amazon further describes how HypRank in Alexa performs "intent-slot semantic analysis for a skill" such that user inputs are compared to "a list of hypotheses."[82]

[82] https://www.amazon.science/blog/hyprank-how-alexa-determines-what-skill-can-best-meet-a-customers-need

HypRank comprises two components:

1. Hypothesis representation for each skill; and 2. A bidirectional long short-term memory (bi-LSTM) model for re-ranking a list of hypotheses.

For each skill in the k-best list, we form a hypothesis based on additional semantic and contextual signals. For example, we perform intent-slot semantic analysis for a skill. If a user says "play Michael Jackson," the Pop Music skill might infer the intent PlayMusic, while the Classic Music skill might infer the intent PlayTune. But the confidence scores that the skills assign their inferences could be useful for skill re-ranking. The hypothesis generator is constantly being updated, re-weighting signals and accommodating new functionality and changes in usage patterns.

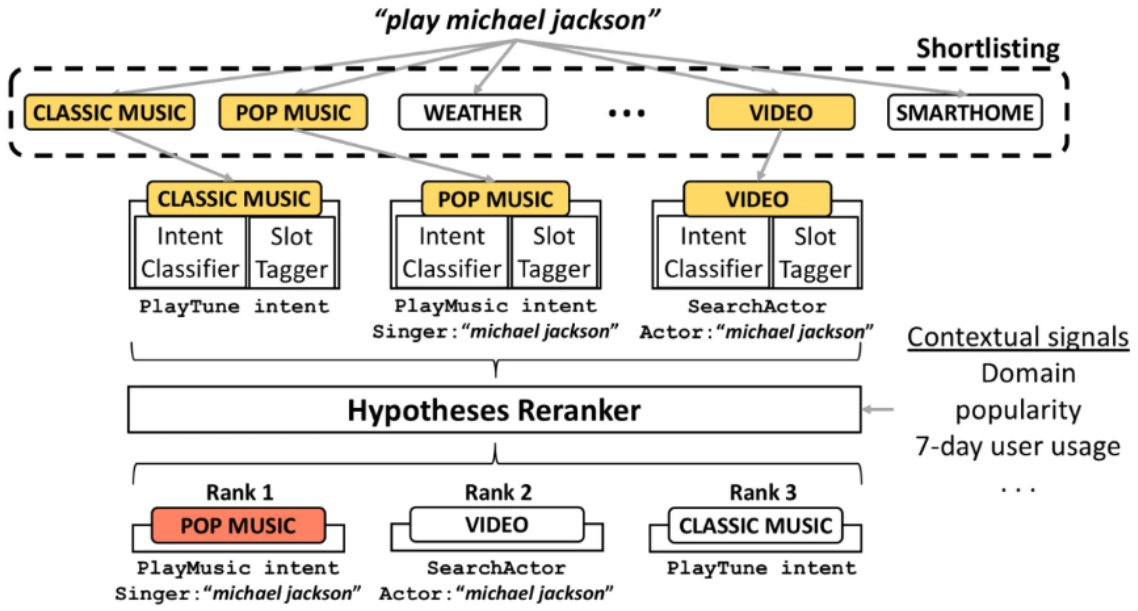
HypRank is unique because of its list-wise ranking approach using a bi-LSTM layer. LSTM models are common in natural-language processing because they factor in the order in which data are received: if you're trying to understand the sixth word in an utterance, it helps to know what the previous five were. Bidirectional LSTM models consider data sequences both forward and backward.

By leveraging the bi-LSTM layer, HypRank can evaluate an entire list of skill hypotheses before providing a re-ranking score for each hypothesis. This is distinct from point-wise approaches that look at each hypothesis in isolation or pair-wise approaches that look at pairs of hypotheses in a series of tournament-like competitions.

179. Each of the Accused Multi-modal Products and Services comprises determining a most likely context for the multi-modal input based on the identified entries.

180. For example, Amazon describes how the Hypotheses Reranker in Alexa uses "rich contextual signals" in order "to select the most pertinent skills."[83]

[83] https://www.amazon.science/blog/hyprank-how-alexa-determines-what-skill-can-best-meet-a-customers-need



A high-level flow of the two-step shortlisting-reranking approach

The Challenge

The problem here is essentially a domain classification problem over the k-best candidate skills returned by the shortlisting system, which we call Shortlister. The goal of Shortlister is to achieve high recall — to identify as many pertinent skills as possible — with maximum efficiency. On the other hand, the goal of the reranking network, HypRank, is to use rich contextual signals to achieve high precision — to select the most pertinent skills. Designing HypRank comes with its own challenges:

181. Each of the Accused Multi-modal Products and Services comprises identifying a domain agent associated with the most likely context for the multi-modal input.

182. For example, Amazon describes Alexa as using a two-step approach to find the most relevant skill for a given utterance.[84]

[84] https://www.amazon.science/blog/the-scalable-neural-architecture-behind-alexas-ability-to-select-skills

Alexa uses a two-step, scalable, and efficient neural shortlisting-reranking approach to find the most relevant skill for a given utterance. This post describes the first of those two steps, which relies on a neural model we call Shortlister. (I'll describe the second step in a follow-up post tomorrow.) Shortlister is a scalable and efficient architecture with a shared encoder, a personalized skill attention mechanism, and skill-specific classification networks. We outline this architecture in our paper "Efficient Large-Scale Neural Domain Classification with Personalized Attention", which we will present at the 56th Annual Meeting of the Association for Computational Linguistics (ACL 2018) in July.

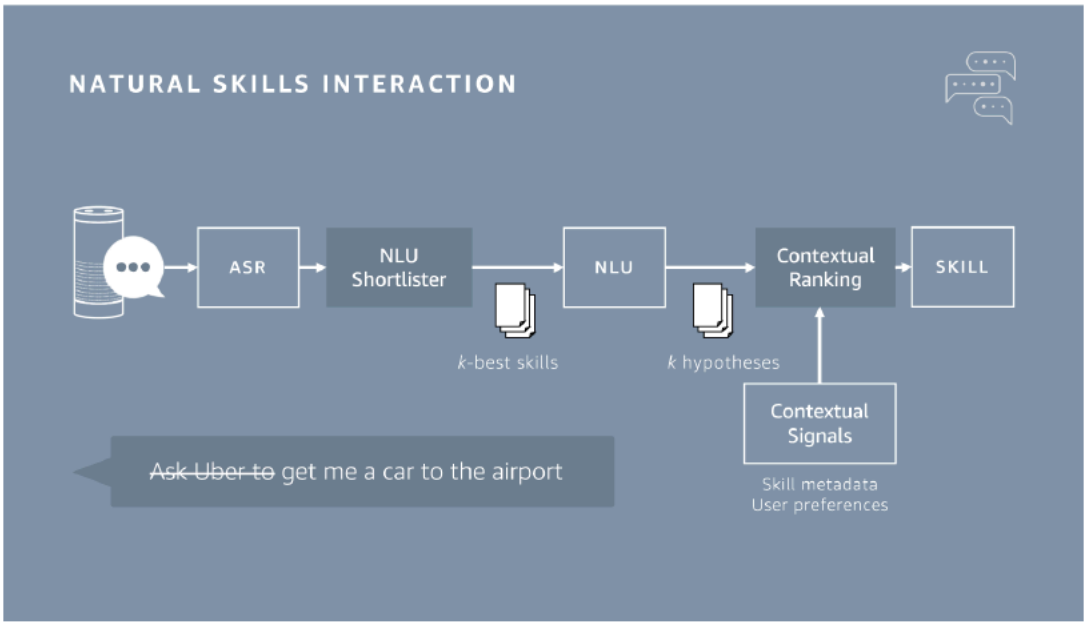
The shared encoder network is hierarchical: Its lower layers are character-based and orthography sensitive and learn to represent each word in terms of character structure or shape; its middle layers are word-based, and with the outputs from the lower layers, they learn to represent an entire utterance. The skill attention mechanism is a separate network that is personalized per user. It computes a summary vector that describes which skills are enabled in a given user's profile and how relevant they are to the utterance representation. Both the utterance representation vector and the personalized skill-summary vector feed into a battery of skill-specific classification networks, one network for each skill.

Amazon further describes how Alexa surfaces the most relevant skills.[85]

[85] https://developer.amazon.com/fr/blogs/alexa/post/0fecdb38-97c9-48ac-953b-23814a469cfc/skill-discovery

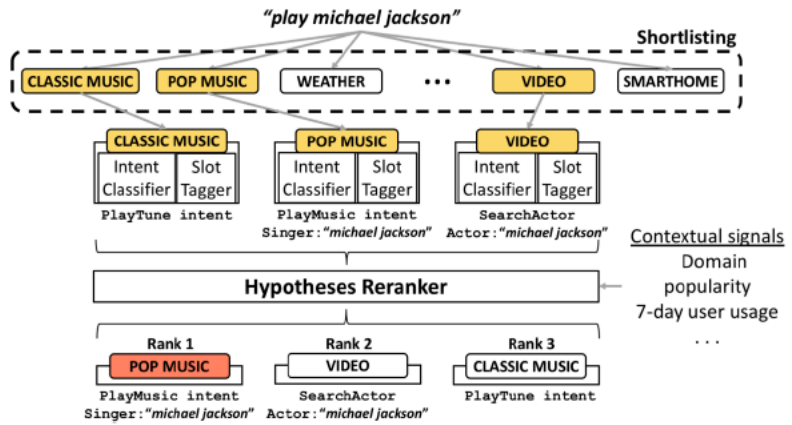
Surfacing the Most Relevant Skills via Shortlister and HypRank

It's important that customers are able to easily reengage with their skills – even if they describe what they want instead of calling the skill by name. Let's look again at the phrase "Alexa, remove an ink stain" where we have an utterance, "remove a {stain type} stain," and a slot value, stain type of "ink," that can be mapped to a skill's interaction model. To make this work, two techniques are employed to deliver the right experience. The first is a neural model architecture we call Shortlister, which solves a domain classification problem to find the most statistically significant (k-best) matches of candidate skills (details in a recent blog and research paper). The goal of Shortlister is to efficiently identify as many relevant skills as possible – it's high volume. The second is a hypothesis reranking network, we call HypRank. It uses contextual signals to select the most relevant skills – it's high precision (details in a recent blog and research paper). When combined, these two help classify which skill can provide the most likely correct response from an invocation name-free utterance. Beyond helping people reengage with the skills they've manually or automatically enabled, the Shortlister and HypRank surface specific and relevant skills to people that have never used them.



Amazon further provides examples of how the hypothesis reranker operates.[86]

[86] https://www.amazon.science/blog/hyprank-how-alexa-determines-what-skill-can-best-meet-a-customers-need



A high-level flow of the two-step shortlisting-reranking approach

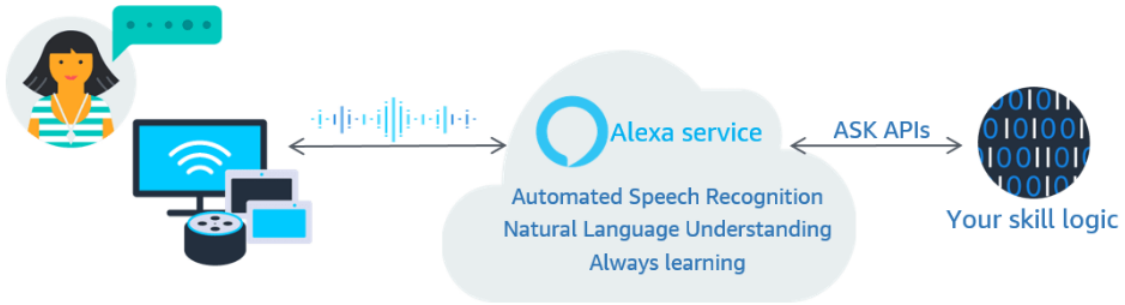
The Challenge

The problem here is essentially a domain classification problem over the k-best candidate skills returned by the shortlisting system, which we call Shortlister. The goal of Shortlister is to achieve high recall — to identify as many pertinent skills as possible — with maximum efficiency. On the other hand, the goal of the reranking network, HypRank, is to use rich contextual signals to achieve high precision — to select the most pertinent skills. Designing HypRank comes with its own challenges:

183. Each of the Accused Multi-modal Products and Services comprises communicating a request to the identified domain agent.

184. For example, Amazon describes how Alexa communicates a request to an Alexa skill.[87]

The following diagram shows the voice-activated processing flow to invoke a skill with the Alexa service.



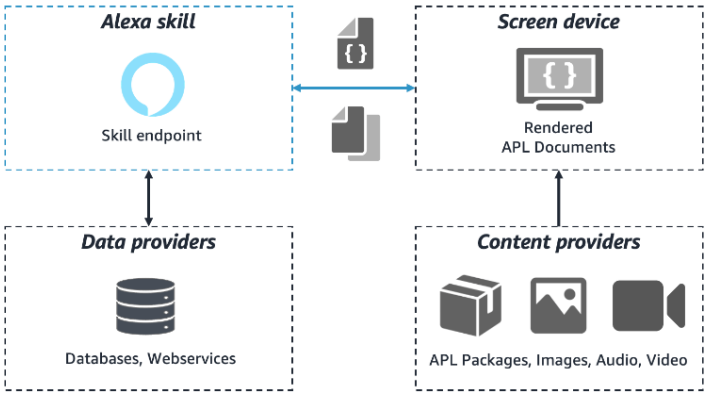
In addition to voice interaction, skills might include complementary visuals and touch interactions.

[87] https://developer.amazon.com/en-US/docs/alexa/ask-overviews/what-is-the-alexa-skills-kit.html

185. Each of the Accused Multi-modal Products and Services comprises generating a response to the user from content provided by the identified domain agent as a result of processing the request.

186. For example, Amazon describes how an Alexa skill processes a multi-modal response on a screen device.[88]

To plan your APL experiences, you need to understand the four technical *actors* that play a role in rendering APL *documents* on Alexa-enabled devices with screens. The following diagram shows a typical APL architecture at a high level.



APL architecture with four technical actors.

- **Alexa skill** — A skill endpoint that initiates an APL experience by sending out an APL *document* template in response to a regular skill request such as an `IntentRequest`.
- **Data providers** — APL templates often come with data that should be rendered with it so that the skill often attaches a JSON data source to the template that is sent to a screen device. This data typically doesn't reside within a skill, but resides in external data stores, such as a database that the skill reads from and writes to.
- **Screen device** — When the skill has sent an APL template with an attached data source, an Alexa-enabled device with a screen, such as an Echo Show or Fire TV, is responsible for inflating and rendering the template on screen.
- **Content providers** — APL templates that are sent by a skill often contain URL references to external content and media files; such as APL *packages*, images, audio and video files; that are stored publicly on the internet. An Alexa-enabled device with a screen resolves these references by downloading or streaming source files from external content providers, such as *content delivery networks* (CDN).

187. Amazon has long known about the ’468 Patent.

188. Amazon has been aware of the ’468 Patent from multiple communications, presentations, and licensing negotiations between VoiceBox Technologies and Amazon between at least 2011 and 2017 and from VoiceBox Technologies personnel hired by Amazon.

[88] https://developer.amazon.com/en-US/docs/alexa/alexa-presentation-language/apl-bp-understand-apl-architecture.html

189. Amazon knew or should have known that Amazon's actions infringe one or more of the claims of the '468 Patent because Amazon has the technical expertise to understand the scope and content of the '468 Patent, because Amazon is a major provider of voice recognition products and services, and because Amazon knows the design, function, and operation of the Accused Multi-modal Products and Services, as well as the nature and extent of their use by others. At a minimum, Amazon has knowledge of the '468 Patent at least as of the filing of this Complaint.

190. Further, on information and belief, Amazon has actively induced and/or contributed to infringement of at least Claim 19 of the '468 Patent in violation of at least 35 U.S.C. § 271(b) and (c).

191. Users of the Accused Multi-modal Products and Services directly infringe at least Claim 19 of the '468 Patent when they use the Accused Multi-modal Products and Services in the ordinary, customary, and intended way.

192. On information and belief, Amazon's inducements in violation of 35 U.S.C. § 271(b) include, without limitation and with specific intent to encourage infringement, knowingly inducing businesses and consumers to use the Accused Multi-modal Products and Services within the United States in the ordinary, customary, and intended way by, directly or through intermediaries, supplying the Accused Multi-modal Products and Services to businesses and consumers within the United States, and instructing and encouraging such businesses and consumers to use the Accused Multi-modal Products and Services in the ordinary, customary, and intended way by activities related to selling, marketing, advertising, promotion, support, and distribution of the Accused Multi-modal Products and Services, which Amazon knew infringes at least Claim 19 of the '468 Patent, or, alternatively, was willfully blind to the infringement.

193. On information and belief, Amazon's inducements in violation of 35 U.S.C. § 271(b) further include, without limitation and with specific intent to encourage the infringement, knowingly inducing Amazon's customers to commit acts of infringement with respect to the Accused Multi-modal Products and Services within the United States, by, directly or through intermediaries, instructing and encouraging such customers to import, make, use, sell, offer to sell, or otherwise commit acts of infringement with respect to the Accused Multi-modal Products and Services in the United States by activities related to selling, marketing, advertising, promotion, support, and distribution of the Accused Multi-modal Products and Services, which Amazon knew infringes at least Claim 19 of the '468 Patent, or, alternatively, was willfully blind to the infringement.

194. On information and belief, in violation of 35 U.S.C. § 271(c), Amazon's contributory infringement further includes offering to sell or selling within the United States, or importing into the United States, components of the patented invention of and/or a material or apparatus for use in practicing at least Claim 19 of the '468 Patent, constituting a material part of the invention. On information and belief, Amazon knows and has known the same to be especially made or especially adapted for use in an infringement of the '468 Patent, and such components are not a staple article or commodity of commerce suitable for substantial noninfringing use.

195. Amazon is not licensed or otherwise authorized to practice the claims of the '468 Patent.

196. Thus, by its acts, Amazon has injured Dialect and is liable to Dialect for directly and/or indirectly infringing one or more claims of the '468 Patent, whether literally or under the doctrine of equivalents, including without limitation Claim 19.

197. As a result of Amazon's infringement of the '468 Patent, Dialect has suffered monetary damages, and seeks recovery, in an amount to be proven at trial, adequate to compensate for Amazon's infringement, but in no event less than a reasonable royalty with interest and costs.

198. On information and belief, in addition to Amazon's knowledge of the '468 Patent as set forth above both prior to and as a result of the filing of this Complaint, Amazon has had, and continues to have, the specific intent to infringe, through its deliberate and intentional infringement or, alternatively, through its willfully blind disregard of the '468 Patent by knowing there was a high probability of infringement but taking deliberate actions to avoid confirming that infringement. The filing of this action has also made Amazon aware of the unjustifiably high risk that its actions constituted and continue to constitute infringement of the '468 Patent. On information and belief, discovery will reveal additional facts and circumstances from which Amazon's knowledge and intent to infringe (or willful indifference), both before and after the filing of this action, may be inferred.

199. Accordingly, Amazon's infringement of the '468 Patent has also been and continues to be deliberate, intentional, and willful, and this is therefore an exceptional case warranting an award of enhanced damages and attorneys' fees and costs pursuant to 35 U.S.C. §§ 284 and 285.

200. Amazon's infringement of Dialect's rights under the '468 Patent will continue to damage Dialect, causing irreparable harm for which there is no adequate remedy at law, unless enjoined by this Court.

FIFTH COUNT
(Infringement of U.S Patent No. 9,031,845)

201. Dialect incorporates by reference the allegations set forth in Paragraphs 1–200 as though fully set forth herein.

202. The claims of the ’845 Patent are valid and enforceable.

203. The claims of the ’845 Patent are directed to patentable subject matter. Particularly, the ’845 Patent is directed to a novel, tangible voice recognition system. The inventive, tangible claimed structures of the ’845 Patent improve on the natural language processing of a natural language utterance by a user. The claimed inventions provide specific concrete solutions to the problem of speech recognition in existing systems.

204. In violation of 35 U.S.C. § 271(a) and without authority from Plaintiff, Amazon has directly infringed by making, using, offering for sale, selling, and/or importing into the Commonwealth of Virginia, this judicial district, and elsewhere in the United States products and services that embody the invention disclosed and claimed in the ’845 Patent, including at least the Accused Automotive Products and Services.

205. Each of the Accused Automotive Products and Services contains elements that are identical or equivalent to each claimed element of the patented invention pointed out by at least Claim 1 of the ’845 Patent.

206. Each of the Accused Automotive Products and Services comprises a mobile system for processing natural language utterances.

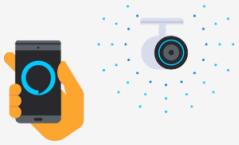
207. For example, Amazon describes how vehicles integrating the Alexa Auto SDK and Alexa Mobile Accessory Kit provide mobile solutions for hands-free interactions.[89]

[89] https://developer.amazon.com/en-US/alexa/alexa-auto/vehicles#experiences

Head Unit with Alexa Built-in

By building Alexa into your vehicle with the Alexa Auto SDK, your customers can interact with Alexa, hands-free, using the in-cabin microphone and speakers. You can also display digital content on the infotainment screen, like album artwork, audiobook covers, search results, and more.

Alexa Auto SDK »



Aftermarket Device with Alexa Built-in

Bring Alexa to the millions of vehicles already on the road by building Alexa into your automotive aftermarket device. The Alexa Mobile Accessory (AMA) Kit enables device makers to easily connect Bluetooth-enabled audio devices with Alexa using the Amazon Alexa App on Android and iOS devices.

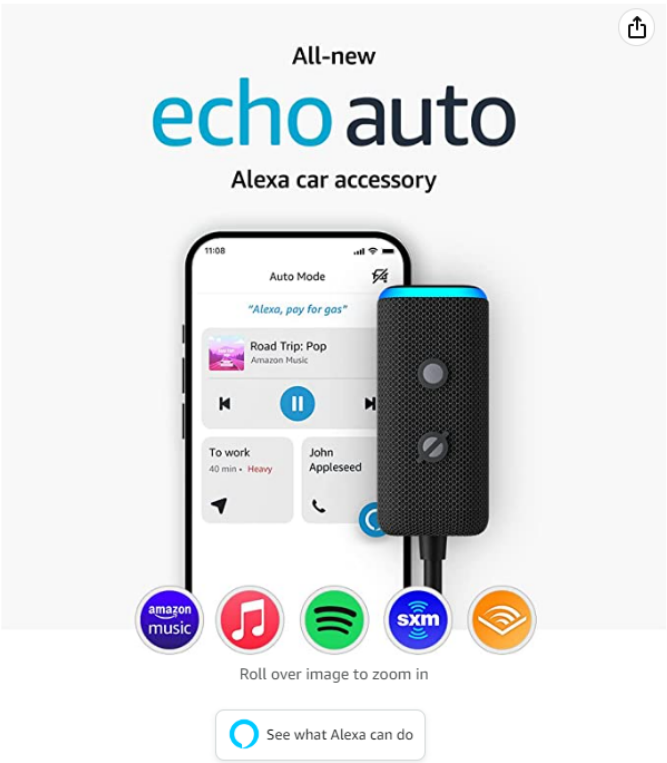
Learn more about AMA »

Amazon further describes the Alexa Auto SDK as enabling speech recognition, Automatic Speech Recognition and Natural Language Understanding.[90]

[90] https://developer.amazon.com/en-US/blogs/alexa/alexa-auto/2019/09/enabling-offline-access-to-alexa-in-vehicles-with-local-voice-control-extension-to-alexa-auto-sdk-v2-0

The Alexa Auto SDK enables core Alexa functionality, such as speech recognition and text-to-speech, and other capabilities such as streaming media, controlling smart home devices, notifications, weather reports, and tens of thousands of custom skills. It can be updated over-the-air and now includes the Local Voice Control extension, which ensures that customers get a reliable response from Alexa regardless of connectivity. On a normal commute, drivers lose connectivity or experience intermittent or transitional connectivity while driving in dense urban areas, through tunnels, in remote areas, or in parking garages. With the Local Voice Control extension to the Alexa Auto SDK, automakers embed a scaled-down version of Alexa's cloud-based Automatic Speech Recognition (ASR) and Natural Language Understanding (NLU) into the vehicle's infotainment system that increases resiliency of the cloud-based capabilities. This ensures that a customer can always ask, "Alexa, turn on the A/C," "Alexa, tune to 90.5 FM," or "Alexa, call Jackie," and get the appropriate response and action from Alexa.

Amazon further describes Echo Auto as enabling hands-free features.[91]



Amazon Music, Apple Music, Spotify, and more, or listen live to radio stations. Use Follow Me Music to resume media playback.

- CALL AND MESSAGE WITH YOUR VOICE - Use your voice to make calls, reply to text messages, drop in on Alexa-enabled devices in your home or broadcast announcements.
- GET YOUR FAVORITE ENTERTAINMENT HANDS-FREE - Binge a hot new podcast, catch up on the news, or listen to best-selling Audible books.
- ALEXA CAN HELP WITH AMAZON ROADSIDE ASSISTANCE - Just say, "Alexa, call Roadside Assistance" and you'll be connected to an agent who can assess the situation and request pay-as-you-go services as needed.
- CONTROL YOUR SMART HOME FROM THE ROAD - Ask Alexa to set the thermostat, turn off the lights, check if your front door is locked, and more while you're away from home.
- DESIGNED TO PROTECT YOUR PRIVACY - Built with multiple layers of privacy controls, including a mic-off button and light indicator.

We want you to know

Echo Auto is designed to enable Alexa hands-free features, such as play music, make calls, answer text messages and more for cars that don't have a voice assistant built-in. Echo Auto connects to the Alexa app on your phone using your existing data plan and plays through your car's speakers via Bluetooth connection or Aux input. Carrier charges may apply.

Car and Phone compatibility - Check if your car is compatible and check if your phone is compatible. Note: Echo Auto is compatible with most cars with Bluetooth or Aux input.

Fast car charger - This product is Quick Charge 3.0 certified for fast charging.

Mounting your Echo Auto - The included adhesive car mount should not be placed on soft plastic, leather, or fabric surfaces. Choose a hard, durable surface away from the airflow.

208. Each of the Accused Automotive Products and Services comprises one or more physical processors at a vehicle that are programmed to execute one or more computer program instructions which, when executed, cause the one or more physical processors to: receive a natural language utterance associated with a user.

[91] https://developer.android.com/training/cars/apps/poi

209. For example, Amazon describes the use of the Alexa Auto SDK to provide a "scaled-down version of Alexa's cloud based Automatic Speech Recognition (ASR) and Natural Language Understanding (NLU) into the vehicle's infotainment system."[92]

> The Alexa Auto SDK enables core Alexa functionality, such as speech recognition and text-to-speech, and other capabilities such as streaming media, controlling smart home devices, notifications, weather reports, and tens of thousands of custom skills. It can be updated over-the-air and now includes the Local Voice Control extension, which ensures that customers get a reliable response from Alexa regardless of connectivity. On a normal commute, drivers lose connectivity or experience intermittent or transitional connectivity while driving in dense urban areas, through tunnels, in remote areas, or in parking garages. With the Local Voice Control extension to the Alexa Auto SDK, automakers embed a scaled-down version of Alexa's cloud-based Automatic Speech Recognition (ASR) and Natural Language Understanding (NLU) into the vehicle's infotainment system that increases resiliency of the cloud-based capabilities. This ensures that a customer can always ask, "Alexa, turn on the A/C," "Alexa, tune to 90.5 FM," or "Alexa, call Jackie," and get the appropriate response and action from Alexa.

210. Each of the Accused Automotive Products and Services comprises performing speech recognition on the natural language utterance.

211. For example, Amazon describes Alexa Auto SDK as supporting Automatic Speech Recognition (ASR) and Natural Language Understanding (NLU).[93]

212. Each of the Accused Automotive Products and Services comprises parsing and interpreting the speech recognized natural language utterance.

213. For example, Amazon describes how in "vehicles with Alexa" that the system can interpret natural language questions.[94]

> "Alexa, where's the nearest coffee shop?"
>
> In vehicles with Alexa, drivers can pose questions like that — while keeping their eyes on the road and hands on the wheel. Amazon's cloud-based voice assistant technology works with the car's navigation system to figure out where the nearest coffee shop is and guide drivers to it.

---

[92] https://developer.amazon.com/en-US/blogs/alexa/alexa-auto/2019/09/enabling-offline-access-to-alexa-in-vehicles-with-local-voice-control-extension-to-alexa-auto-sdk-v2-0t

[93] *Id.*

[94] https://www.amazon.science/news-and-features/the-science-behind-alexa-in-vehiclest

Amazon states that its intention is to take the home Alexa experience "and extend that experience to cars."[95] Amazon further describes Alexa as matching the transcribed text of the user's speech to the user's intent.[96]

How Alexa Works

When I'm in my living room and ask Alexa about the weather, I trigger a complex system. First, the on-device chip detects the wake word (Alexa). Once detected, the microphones record what I'm saying and stream the sound for analysis in the cloud. At a high level, there are two phases to analyze the sound of my voice. First, Alexa converts the sound to text. This is known as Automatic Speech Recognition (ASR). Once the text is known, the second phase is to understand what I mean. This is Natural Language Understanding (NLU). The output of NLU is an Intent (what does the customer want) and associated parameters. In this example ("Alexa, what's the weather today ?"), the intent might be " `GetWeatherForecast` " and the parameter can be my postcode, inferred from my profile.

This whole process uses Artificial Intelligence heavily to transform the sound of my voice to phonemes, phonemes to words, words to phrases, phrases to intents. Based on the NLU output, Alexa routes the intent to a service to fulfill it. The service might be internal to Alexa or external, like one of the skills activated on my Alexa account. The fulfillment service processes the intent and returns a response as a JSON document. The document contains the text of the response Alexa must say.

214. Each of the Accused Automotive Products and Services comprises determining a domain and a context that are associated with the parsed and interpreted natural language utterance.

215. For example, Amazon describes Alexa as using a two-step approach to find the most relevant skill for a given utterance.[97]

---

[95] *Id.*

[96] https://aws.amazon.com/blogs/aws/majority-of-alexa-now-running-on-faster-more-cost-effective-amazon-ec2-inf1-instances/

[97] https://www.amazon.science/blog/the-scalable-neural-architecture-behind-alexas-ability-to-select-skills

Alexa uses a two-step, scalable, and efficient neural shortlisting-reranking approach to find the most relevant skill for a given utterance. This post describes the first of those two steps, which relies on a neural model we call Shortlister. (I'll describe the second step in a follow-up post tomorrow.) Shortlister is a scalable and efficient architecture with a shared encoder, a personalized skill attention mechanism, and skill-specific classification networks. We outline this architecture in our paper "Efficient Large-Scale Neural Domain Classification with Personalized Attention", which we will present at the 56th Annual Meeting of the Association for Computational Linguistics (ACL 2018) in July.

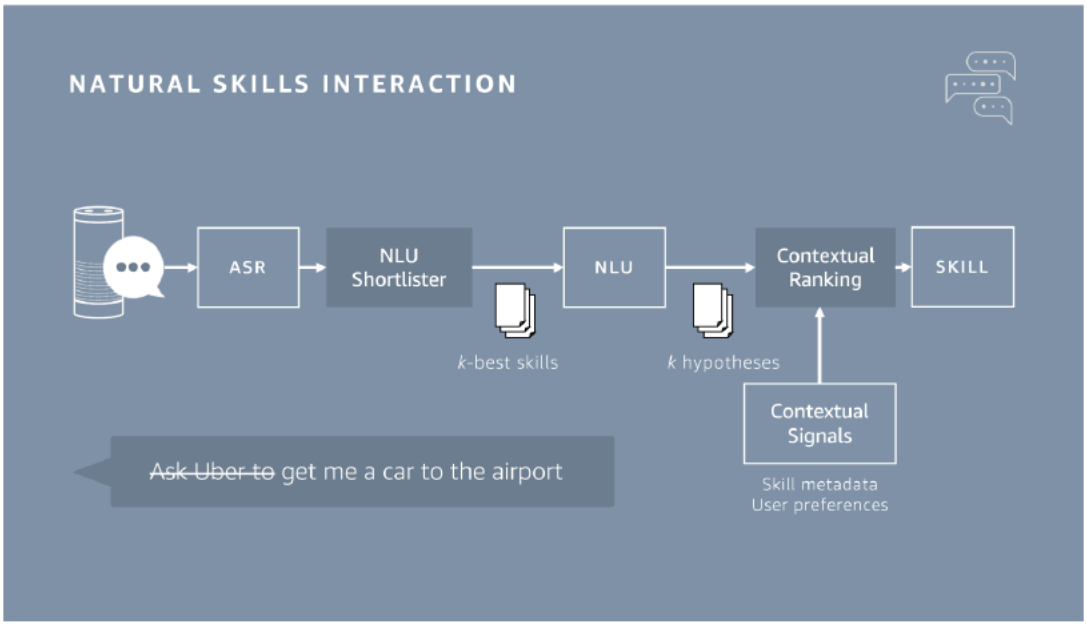
The shared encoder network is hierarchical: Its lower layers are character-based and orthography sensitive and learn to represent each word in terms of character structure or shape; its middle layers are word-based, and with the outputs from the lower layers, they learn to represent an entire utterance. The skill attention mechanism is a separate network that is personalized per user. It computes a summary vector that describes which skills are enabled in a given user's profile and how relevant they are to the utterance representation. Both the utterance representation vector and the personalized skill-summary vector feed into a battery of skill-specific classification networks, one network for each skill.

Amazon further describes how Alexa surfaces the most relevant skills.[98]

[98] https://developer.amazon.com/fr/blogs/alexa/post/0fecdb38-97c9-48ac-953b-23814a469cfc/skill-discovery

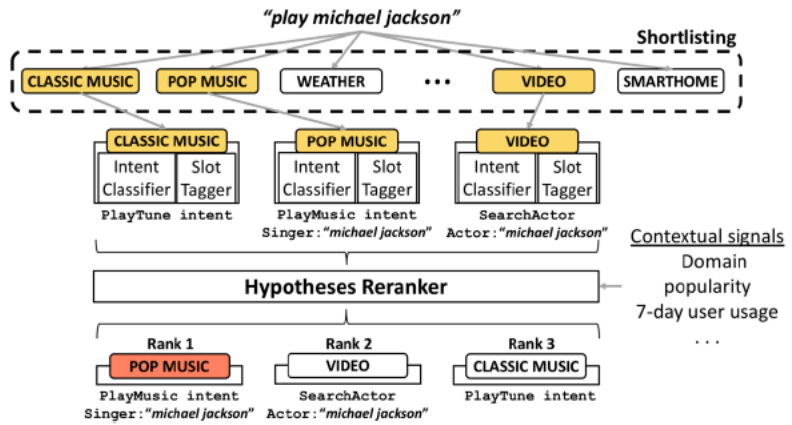
Surfacing the Most Relevant Skills via Shortlister and HypRank

It's important that customers are able to easily reengage with their skills – even if they describe what they want instead of calling the skill by name. Let's look again at the phrase "Alexa, remove an ink stain" where we have an utterance, "remove a {stain type} stain," and a slot value, stain type of "ink," that can be mapped to a skill's interaction model. To make this work, two techniques are employed to deliver the right experience. The first is a neural model architecture we call Shortlister, which solves a domain classification problem to find the most statistically significant (k-best) matches of candidate skills (details in a recent blog and research paper). The goal of Shortlister is to efficiently identify as many relevant skills as possible – it's high volume. The second is a hypothesis reranking network, we call HypRank. It uses contextual signals to select the most relevant skills – it's high precision (details in a recent blog and research paper). When combined, these two help classify which skill can provide the most likely correct response from an invocation name-free utterance. Beyond helping people reengage with the skills they've manually or automatically enabled, the Shortlister and HypRank surface specific and relevant skills to people that have never used them.



Amazon further provides examples of how the hypothesis reranker operates.[99]

[99] https://www.amazon.science/blog/hyprank-how-alexa-determines-what-skill-can-best-meet-a-customers-need



A high-level flow of the two-step shortlisting-reranking approach

The Challenge

The problem here is essentially a domain classification problem over the k-best candidate skills returned by the shortlisting system, which we call Shortlister. The goal of Shortlister is to achieve high recall — to identify as many pertinent skills as possible — with maximum efficiency. On the other hand, the goal of the reranking network, HypRank, is to use rich contextual signals to achieve high precision — to select the most pertinent skills. Designing HypRank comes with its own challenges:

216. Each of the Accused Automotive Products and Services comprises formulating a command or query based on the domain and the context.

217. For example, Amazon describes Alexa as processing the user's command to tune to 90.5 FM.[100]

[100] https://developer.amazon.com/en-US/blogs/alexa/alexa-auto/2019/09/enabling-offline-access-to-alexa-in-vehicles-with-local-voice-control-extension-to-alexa-auto-sdk-v2-0t

218. Each of the Accused Automotive Products and Services comprises determining whether the command or query is to be executed on-board or off-board the vehicle.

219. For example, Alexa describes the Local Voice Control extension to the Alexa Auto SDK as providing for commands or queries to be executed on-board the vehicle instead of on the Alexa cloud hardware.[101]

> The Alexa Auto SDK enables core Alexa functionality, such as speech recognition and text-to-speech, and other capabilities such as streaming media, controlling smart home devices, notifications, weather reports, and tens of thousands of custom skills. It can be updated over-the-air and now includes the Local Voice Control extension, which ensures that customers get a reliable response from Alexa regardless of connectivity. On a normal commute, drivers lose connectivity or experience intermittent or transitional connectivity while driving in dense urban areas, through tunnels, in remote areas, or in parking garages. With the Local Voice Control extension to the Alexa Auto SDK, automakers embed a scaled-down version of Alexa's cloud-based Automatic Speech Recognition (ASR) and Natural Language Understanding (NLU) into the vehicle's infotainment system that increases resiliency of the cloud-based capabilities. This ensures that a customer can always ask, "Alexa, turn on the A/C," "Alexa, tune to 90.5 FM," or "Alexa, call Jackie," and get the appropriate response and action from Alexa.
>
> "The vehicle environment is unique and requires online and offline voice capabilities to be truly successful," says Matt Arcaro, Research Manager, Next-Generation Automotive Research at IDC. "This dual-mode support not only reduces the impact of variable network conditions, but will give vehicle users the confidence to increasingly think of voice as an effective and efficient means for vehicle interaction and customization."
>
> The following features are now available through the Alexa Auto SDK:
>
> - **Music and Radio**: control media playback, stream music, audiobooks, and podcasts from leading providers, tune radio, change local media sources, turn up the volume, and adjust equalizer.
> - **Calling and Messaging**: dial by name and number, control call in progress, make Alexa announcements, use Alexa Drop In, join meetings, and send and receive Alexa messages.
> - **Search and Navigation**: navigate to favorite destinations, search for places by name, brand, category, address, or intersection, get details such as hours of operation, phone number, and ratings, determine route info like traffic and ETA, and cancel navigation in progress.
> - **Car Control**: turn on air conditioning, set desired cabin temperature, control temperature by zone, set fan mode and speed, defrost front and rear windshield, and control interior cabin lighting.

220. Each of the Accused Automotive Products and Services comprises executing the command or query at the vehicle in response to a determination that the command or query is to be executed on-board the vehicle.

221. For example, Amazon describes commands that can be executed at the vehicle when the command is to be executed on-board.[102]

---

[101] *Id.*
[102] https://developer.android.com/training/cars/media

The Alexa Auto SDK enables core Alexa functionality, such as speech recognition and text-to-speech, and other capabilities such as streaming media, controlling smart home devices, notifications, weather reports, and tens of thousands of custom skills. It can be updated over-the-air and now includes the Local Voice Control extension, which ensures that customers get a reliable response from Alexa regardless of connectivity. On a normal commute, drivers lose connectivity or experience intermittent or transitional connectivity while driving in dense urban areas, through tunnels, in remote areas, or in parking garages. With the Local Voice Control extension to the Alexa Auto SDK, automakers embed a scaled-down version of Alexa's cloud-based Automatic Speech Recognition (ASR) and Natural Language Understanding (NLU) into the vehicle's infotainment system that increases resiliency of the cloud-based capabilities. This ensures that a customer can always ask, "Alexa, turn on the A/C," "Alexa, tune to 90.5 FM," or "Alexa, call Jackie," and get the appropriate response and action from Alexa.

"The vehicle environment is unique and requires online and offline voice capabilities to be truly successful," says Matt Arcaro, Research Manager, Next-Generation Automotive Research at IDC. "This dual-mode support not only reduces the impact of variable network conditions, but will give vehicle users the confidence to increasingly think of voice as an effective and efficient means for vehicle interaction and customization."

The following features are now available through the Alexa Auto SDK:

- **Music and Radio**: control media playback, stream music, audiobooks, and podcasts from leading providers, tune radio, change local media sources, turn up the volume, and adjust equalizer.
- **Calling and Messaging**: dial by name and number, control call in progress, make Alexa announcements, use Alexa Drop In, join meetings, and send and receive Alexa messages.
- **Search and Navigation**: navigate to favorite destinations, search for places by name, brand, category, address, or intersection, get details such as hours of operation, phone number, and ratings, determine route info like traffic and ETA, and cancel navigation in progress.
- **Car Control**: turn on air conditioning, set desired cabin temperature, control temperature by zone, set fan mode and speed, defrost front and rear windshield, and control interior cabin lighting.

222. Each of the Accused Automotive Products and Services comprises invoking a device that communicates wirelessly over a wide area network to process the command or query such that the command or query is executed off-board the vehicle in response to a determination that the command or query is to be executed off-board the vehicle.

223. For example, Amazon describes the use of online functionality for accessing Alexa in the car.[103]

Today, we announced version 2.0 of the Alexa Auto SDK, which now offers a full-suite of offline functionality, in addition to online functionality, for accessing Alexa in the car, including: music and radio, search and navigation, calling and messaging, and car control. Automakers and their suppliers can use the Alexa Auto SDK to integrate Alexa into their vehicles, helping keep customers entertained and productive—even when connectivity is intermittent or not available.

224. Amazon has long known about the '845 Patent.

[103] *Id.*

225. Amazon has been aware of the ’845 Patent from multiple communications, presentations, and licensing negotiations between VoiceBox Technologies and Amazon between at least 2011 and 2017 and from VoiceBox Technologies personnel hired by Amazon.

226. Amazon knew or should have known that Amazon’s actions infringe one or more of the claims of the ’845 Patent because Amazon has the technical expertise to understand the scope and content of the ’845 Patent, because Amazon is a major provider of voice recognition products and services, and because Amazon knows the design, function, and operation of the Accused Automotive Products and Services, as well as the nature and extent of their use by others. At a minimum, Amazon has knowledge of the ’845 Patent at least as of the filing of this Complaint.

227. Further, on information and belief, Amazon has actively induced and/or contributed to infringement of at least Claim 1 of the ’845 Patent in violation of at least 35 U.S.C. § 271(b), (c), and (f).

228. Users of the Accused Automotive Products and Services directly infringe at least Claim 1 of the ’845 Patent when they use the Accused Automotive Products and Services in the ordinary, customary, and intended way.

229. On information and belief, Amazon’s inducements in violation of 35 U.S.C. § 271(b) include, without limitation and with specific intent to encourage infringement, knowingly inducing businesses and consumers to use the Accused Automotive Products and Services within the United States in the ordinary, customary, and intended way by, directly or through intermediaries, supplying the Accused Automotive Products and Services to businesses and consumers within the United States, and instructing and encouraging such businesses and consumers to use the Accused Automotive Products and Services in the ordinary, customary, and intended way by activities related to selling, marketing, advertising, promotion, support, and

distribution of the Accused Automotive Products and Services, which Amazon knew infringes at least Claim 1 of the ’845 Patent, or, alternatively, was willfully blind to the infringement.

230. On information and belief, Amazon’s inducements in violation of 35 U.S.C. § 271(b) further include, without limitation and with specific intent to encourage the infringement, knowingly inducing Amazon’s customers to commit acts of infringement with respect to the Accused Automotive Products and Services within the United States, by, directly or through intermediaries, instructing and encouraging such customers to import, make, use, sell, offer to sell, or otherwise commit acts of infringement with respect to the Accused Automotive Products and Services in the United States by activities related to selling, marketing, advertising, promotion, support, and distribution of the Accused Automotive Products and Services, which Amazon knew infringes at least Claim 1 of the ’845 Patent, or, alternatively, was willfully blind to the infringement.

231. On information and belief, in violation of 35 U.S.C. § 271(c), Amazon’s contributory infringement further includes offering to sell or selling within the United States, or importing into the United States, components of the patented invention of and/or a material or apparatus for use in practicing at least Claim 1 of the ’845 Patent, constituting a material part of the invention. On information and belief, Amazon knows and has known the same to be especially made or especially adapted for use in an infringement of the ’845 Patent, and such components are not a staple article or commodity of commerce suitable for substantial noninfringing use.

232. On information and belief, in violation of 35 U.S.C. § 271(f)(1), Amazon’s infringement further includes without authority supplying or causing to be supplied in or from the United States all or a substantial portion of the components of the patented invention of at least Claim 1 of the ’845 Patent, where such components are uncombined in whole or in part, in such

manner as to actively induce the combination of such components outside of the United States in a manner that would infringe the patent if such combination occurred within the United States.

233. On information and belief, in violation of 35 U.S.C. § 271(f)(2), Amazon's infringement further includes without authority supplying or causing to be supplied in or from the United States components of the patented invention of at least Claim 1 of the '845 Patent that are especially made or especially adapted for use in the invention and not staple articles or commodities of commerce suitable for substantial noninfringing use, where such components are uncombined in whole or in part, knowing that such components are so made or adapted and intending that such components will be combined outside of the United States in a manner that would infringe the patent if such combination occurred within the United States.

234. Amazon is not licensed or otherwise authorized to practice the claims of the '845 Patent.

235. Thus, by its acts, Amazon has injured Dialect and is liable to Dialect for directly and/or indirectly infringing one or more claims of the '845 Patent, whether literally or under the doctrine of equivalents, including without limitation Claim 1.

236. As a result of Amazon's infringement of the '845 Patent, Dialect has suffered monetary damages, and seeks recovery, in an amount to be proven at trial, adequate to compensate for Amazon's infringement, but in no event less than a reasonable royalty with interest and costs.

237. On information and belief, in addition to Amazon's knowledge of the '845 Patent as set forth above both prior to and as a result of the filing of this Complaint, Amazon has had, and continues to have, the specific intent to infringe, through its deliberate and intentional infringement or, alternatively, through its willfully blind disregard of the '845 Patent by knowing there was a high probability of infringement but taking deliberate actions to avoid confirming that

infringement. The filing of this action has also made Amazon aware of the unjustifiably high risk that its actions constituted and continue to constitute infringement of the ’845 Patent. On information and belief, discovery will reveal additional facts and circumstances from which Amazon’s knowledge and intent to infringe (or willful indifference), both before and after the filing of this action, may be inferred.

238. Accordingly, Amazon’s infringement of the ’845 Patent has also been and continues to be deliberate, intentional, and willful, and this is therefore an exceptional case warranting an award of enhanced damages and attorneys’ fees and costs pursuant to 35 U.S.C. §§ 284 and 285.

239. Amazon’s infringement of Dialect’s rights under the ’845 Patent will continue to damage Dialect, causing irreparable harm for which there is no adequate remedy at law, unless enjoined by this Court.

SIXTH COUNT
(Infringement of U.S Patent No. 9,263,039)

240. Dialect incorporates by reference the allegations set forth in Paragraphs 1–239 as though fully set forth herein.

241. The claims of the ’039 Patent are valid and enforceable.

242. The claims of the ’039 Patent are directed to patentable subject matter. Particularly, the ’039 Patent is directed to a novel, tangible voice recognition system. The inventive, tangible claimed structures of the ’039 Patent improve on the natural language processing of a natural language utterance by a user. The claimed inventions provide specific concrete solutions to the problem of speech recognition in existing systems.

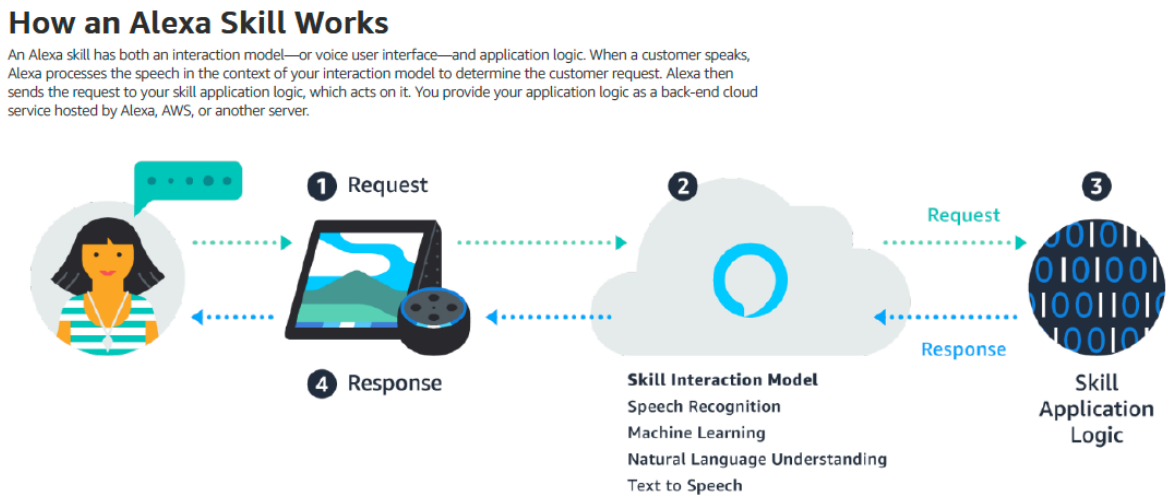
243. In violation of 35 U.S.C. § 271(a) and without authority from Plaintiff, Amazon has directly infringed by making, using, offering for sale, selling, and/or importing into the

Commonwealth of Virginia, this judicial district, and elsewhere in the United States products and services that embody the invention disclosed and claimed in the ’039 Patent, including at least the Alexa Products, operating via tablets, smartphones, or other devices supporting non-speech inputs such as the Echo Show (collectively, the “Accused Multi-modal Products and Services”).

244. Each of the Accused Multi-modal Products and Services contains elements that are identical or equivalent to each claimed element of the patented invention pointed out by at least Claim 13 of the ’039 Patent.

245. Each of the Accused Multi-modal Products and Services comprises a method of processing speech and non-speech communications.

246. For example, Amazon describes Alexa as performing speech recognition on user inputs.[104]



Amazon further describes Alexa as processing non-speech inputs.[105]

[104] https://developer.amazon.com/en-US/alexa/alexa-skills-kit/get-deeper/custom-skills
[105] https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/avs-apl-overview.html

Alexa Multi-Modal API Overview

Note: Learn how to improve your skills with APL with Build visually rich experiences using APL at the Alexa Learning Lab.

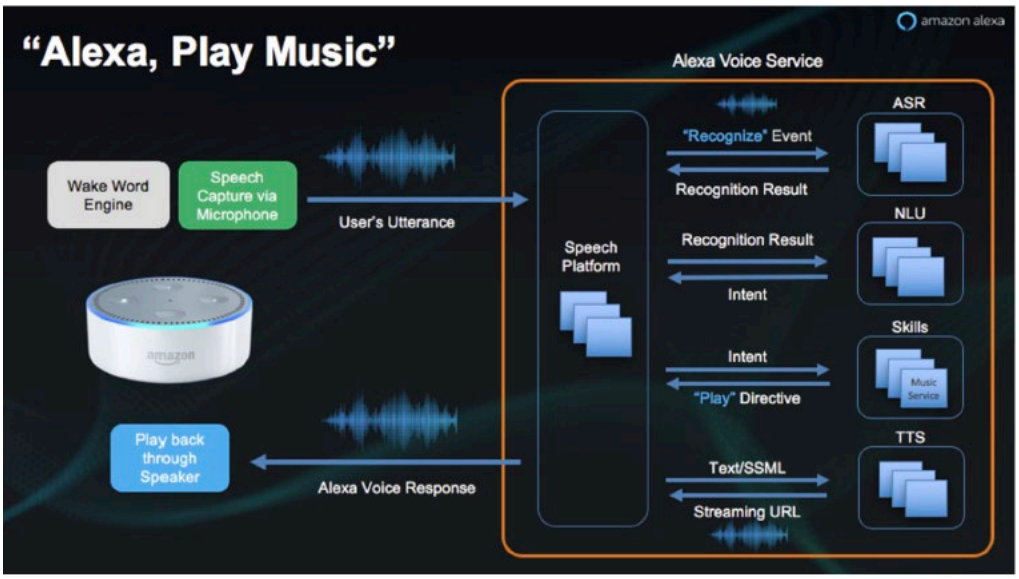
The Alexa Voice Service (AVS) provides a collection of APIs for Alexa Built-in devices with visual displays, such as smart TVs, which allow those devices to receive and present Alexa visual responses. Some of these APIs help device makers to publish the visual characteristics and functionality of their visual displays to enable Alexa Skills and services to respond with appropriate visual content for a device display.

To add visual Alexa experiences to your device, you have several options for different levels of implementation:

- **Full APL integration** – To add a visual Alexa experience to your device, integrate the APL Core Library and implement an APL View Host in your code. See the APL Core Library repository and documentation to learn how to implement these two requirements.
- **Receive and render APL on a device** – To receive and render APL documents on your device screen, implement the Alexa.Presentation.APL namespace.
- **Publish device characteristics to Alexa** – To publish information about the visual characteristics of your device and its display for Alexa Skill developers, you can implement Alexa.Display, Alexa.Display.Window, and Alexa.InteractionMode.
- **AVS Device SDK** – Simplify your APL implementation by integrating the AVS Device SDK, which includes a sample app for voice-only devices and an IPC Server sample app for smart screen devices: AVS Device SDK repository.

247. Each of the Accused Multi-modal Products and Services comprises receiving the speech and non-speech communications.

248. For example, Amazon discloses that Alexa receives speech inputs.[106]



[106] https://developer.amazon.com/fr/blogs/alexa/post/b015c5bd-2b01-4f7f-923e-dd34c998f3d0/avs-tech-series-the-basics-of-amazon-alexa-developer-tools-and-services

Amazon further discloses Alexa processing the sound input into text and then parsing it to understand what the user means.[107]

How Alexa Works

When I'm in my living room and ask Alexa about the weather, I trigger a complex system. First, the on-device chip detects the wake word (Alexa). Once detected, the microphones record what I'm saying and stream the sound for analysis in the cloud. At a high level, there are two phases to analyze the sound of my voice. First, Alexa converts the sound to text. This is known as Automatic Speech Recognition (ASR). Once the text is known, the second phase is to understand what I mean. This is Natural Language Understanding (NLU). The output of NLU is an Intent (what does the customer want) and associated parameters. In this example ("Alexa, what's the weather today ?"), the intent might be "`GetWeatherForecast`" and the parameter can be my postcode, inferred from my profile.

This whole process uses Artificial Intelligence heavily to transform the sound of my voice to phonemes, phonemes to words, words to phrases, phrases to intents. Based on the NLU output, Alexa routes the intent to a service to fulfill it. The service might be internal to Alexa or external, like one of the skills activated on my Alexa account. The fulfillment service processes the intent and returns a response as a JSON document. The document contains the text of the response Alexa must say.

Amazon further describes Alexa's ability to receive non-speech inputs.[108]

[107] https://aws.amazon.com/blogs/aws/majority-of-alexa-now-running-on-faster-more-cost-effective-amazon-ec2-inf1-instances/

[108] https://developer.amazon.com/en-US/docs/alexa/alexa-presentation-language/apl-authoring-tool.html#use-mmrb

Use the Multimodal Response Builder

The Multimodal Response Builder provides a guided experience to create a visual response in three steps:

1. Select from a set of templates designed to look good across a broad range of devices.
2. Customize the response by specifying the content to display within the template
3. Preview the response both in the developer console and on a device.

After you finish customizing the response, the Multimodal Response Builder generates a code example you can copy into a request handler in your skill to display the response. Rebuild the interaction model for your skill and then test your skill on a device or with the developer console simulator.

You can create a new document in the Multimodal Response Builder to get started, and then edit the document in the full authoring tool if you want to do more complex customizations.

To open the Multimodal Response Builder

1. In the developer console, open the skill for which you want to create this document.
2. In the left-hand navigation, click **Multimodal Responses**.
3. Click **Create with Response Builder**.
4. Step through the pages to complete the three steps
   - Select a template. The set of templates includes responsive templates as well as other visual designs.
   - Customize
   - Preview and test

Note: Populating the primary text or title fields of a template with long strings might create a poor visual experience. Make sure you test your response across different viewport and make sure all of your content displays as you expect.

Amazon further describes the Alexa Multi-Modal API.[109]

[109] https://developer.amazon.com/en-US/docs/alexa/alexa-voice-service/avs-apl-overview.html

Alexa Multi-Modal API Overview

Note: Learn how to improve your skills with APL with Build visually rich experiences using APL at the Alexa Learning Lab.

The Alexa Voice Service (AVS) provides a collection of APIs for Alexa Built-in devices with visual displays, such as smart TVs, which allow those devices to receive and present Alexa visual responses. Some of these APIs help device makers to publish the visual characteristics and functionality of their visual displays to enable Alexa Skills and services to respond with appropriate visual content for a device display.

To add visual Alexa experiences to your device, you have several options for different levels of implementation:

- **Full APL integration** – To add a visual Alexa experience to your device, integrate the APL Core Library and implement an APL View Host in your code. See the APL Core Library repository and documentation to learn how to implement these two requirements.
- **Receive and render APL on a device** – To receive and render APL documents on your device screen, implement the Alexa.Presentation.APL namespace.
- **Publish device characteristics to Alexa** – To publish information about the visual characteristics of your device and its display for Alexa Skill developers, you can implement Alexa.Display, Alexa.Display.Window, and Alexa.InteractionMode.
- **AVS Device SDK** – Simplify your APL implementation by integrating the AVS Device SDK, which includes a sample app for voice-only devices and an IPC Server sample app for smart screen devices: AVS Device SDK repository.

249. Each of the Accused Multi-modal Products and Services comprises transcribing the speech and non-speech communications to create a speech-based textual message and a non-speech-based textual message.

250. For example, Amazon describes Alexa as transcribes the speech.[110]

How Alexa Works

When I'm in my living room and ask Alexa about the weather, I trigger a complex system. First, the on-device chip detects the wake word (Alexa). Once detected, the microphones record what I'm saying and stream the sound for analysis in the cloud. At a high level, there are two phases to analyze the sound of my voice. First, Alexa converts the sound to text. This is known as Automatic Speech Recognition (ASR). Once the text is known, the second phase is to understand what I mean. This is Natural Language Understanding (NLU). The output of NLU is an Intent (what does the customer want) and associated parameters. In this example ("Alexa, what's the weather today ?"), the intent might be " `GetWeatherForecast` " and the parameter can be my postcode, inferred from my profile.

This whole process uses Artificial Intelligence heavily to transform the sound of my voice to phonemes, phonemes to words, words to phrases, phrases to intents. Based on the NLU output, Alexa routes the intent to a service to fulfill it. The service might be internal to Alexa or external, like one of the skills activated on my Alexa account. The fulfillment service processes the intent and returns a response as a JSON document. The document contains the text of the response Alexa must say.

Amazon further describes how to capture and respond to key presses by the user.[111]

---

[110] https://aws.amazon.com/blogs/aws/majority-of-alexa-now-running-on-faster-more-cost-effective-amazon-ec2-inf1-instances/

[111] https://developer.amazon.com/en-US/docs/alexa/alexa-presentation-language/apl-for-screen-devices.html

Keyboard handlers

APL provides support for keyboard events. You can define keyboard handlers to capture and respond to key presses. For details, see Keyboard Events and Handlers.

251. On information and belief, each of the Accused Multi-modal Products and Services comprises merging the speech-based textual message and the non-speech-based textual message to generate a query.

252. For example, Amazon describes Alexa processing the transcript to determine what the user means.[112]

How Alexa Works

When I'm in my living room and ask Alexa about the weather, I trigger a complex system. First, the on-device chip detects the wake word (Alexa). Once detected, the microphones record what I'm saying and stream the sound for analysis in the cloud. At a high level, there are two phases to analyze the sound of my voice. First, Alexa converts the sound to text. This is known as Automatic Speech Recognition (ASR). Once the text is known, the second phase is to understand what I mean. This is Natural Language Understanding (NLU). The output of NLU is an Intent (what does the customer want) and associated parameters. In this example ("Alexa, what's the weather today ?"), the intent might be " `GetWeatherForecast` " and the parameter can be my postcode, inferred from my profile.

This whole process uses Artificial Intelligence heavily to transform the sound of my voice to phonemes, phonemes to words, words to phrases, phrases to intents. Based on the NLU output, Alexa routes the intent to a service to fulfill it. The service might be internal to Alexa or external, like one of the skills activated on my Alexa account. The fulfillment service processes the intent and returns a response as a JSON document. The document contains the text of the response Alexa must say.

253. Each of the Accused Multi-modal Products and Services comprises searching the query for text combinations.

254. For example, Amazon describes Alexa as listening for recognized text combinations to select an appropriate skill.[113]

---

[112] https://aws.amazon.com/blogs/aws/majority-of-alexa-now-running-on-faster-more-cost-effective-amazon-ec2-inf1-instances/

[113] https://developer.amazon.com/en-US/docs/alexa/custom-skills/understand-name-free-interaction-for-custom-skills.html

When Alexa receives a request from a customer without a skill name, such as "Alexa, play relaxing sounds with crickets," Alexa looks for skills that might fulfill the request. Alexa determines the best choice among eligible skills and hands the request to the skill. To provide a signal for Alexa to consider when routing name free requests and enable customers to launch your skill and intents without knowing or having to remember the skill's name, consider adding Name Free Interactions (NFI) container to the skill.

About Skill and Intent Launch phrases

Skill launch Phrases are an optional new way to teach Alexa how an end-customer might invoke your skill as a modal launch, along with the standard invocation pattern of "Alexa, open ". For example, "Alexa, Open Tasty Recipes Skill" you might have something like "can you give me a tasty recipe".

Intents and Intent Launch Phrases: The nameFreeInteraction container allows you to identify which intents are natural ingress points (or invocation points) for your skill. Where end users might currently need to say "Alexa, open Tasty Recipes Skill and give me a recipe for meatballs", you can indicate to Alexa that your getRecipe intent is a natural ingress point, and provide common natural phrases that would be used to deep-link into this intent such as simply, "give me a tasty recipe for meatballs". The phrases provided in the nfi container give you the mechanism to indicate which sample utterances/use cases are of the highest importance so that Alexa can weight these utterances among the rest. Note: slots are not supported for skill or intent launch phrases within the nameFreeInteraction container, however the sample utterances you provide in your interaction model that contain slots will still be used for training data.

The most important rule is to design intents with a list of sample utterances that captures key words and phrases that can invoke the skill in a name- free manner. Keep your utterances specific to your skills use-cases and avoid targeting generic utterances. The Alexa ML/AI models are designed in part to route traffic to skills with a high correlation between skill functionality and target utterances.

Guidelines

To help ensure that Alexa directs to your skill only those customer requests, questions, or statements that can be supported by the skill, build sample utterances that capture your skill's functionality. Provide sample utterances that you predict customers can use to interact with the skill naturally. For a better user experience, provide a large variety of sample utterances written in different forms. For example, name-free utterances for a horoscope skill might include:

- *"What is the horoscope for gemini?"*
- *"Today's horoscope"*
- *"I want to know my horoscope"*
- *"What does my star sign say?"*

255. Each of the Accused Multi-modal Products and Services comprises comparing the text combinations to entries in a context description grammar.

256. For example, Amazon describes Alexa as comparing text combinations to a list of sample utterances that can invoke a skill in a name-free manner.[114]

The most important rule is to design intents with a list of sample utterances that captures key words and phrases that can invoke the skill in a name- free manner. Keep your utterances specific to your skills use-cases and avoid targeting generic utterances. The Alexa ML/AI models are designed in part to route traffic to skills with a high correlation between skill functionality and target utterances.

Guidelines

To help ensure that Alexa directs to your skill only those customer requests, questions, or statements that can be supported by the skill, build sample utterances that capture your skill's functionality. Provide sample utterances that you predict customers can use to interact with the skill naturally. For a better user experience, provide a large variety of sample utterances written in different forms. For example, name-free utterances for a horoscope skill might include:

- *"What is the horoscope for gemini?"*
- *"Today's horoscope"*
- *"I want to know my horoscope"*
- *"What does my star sign say?"*

Amazon further describes how grammar has "been a tool in Alexa's NLU toolkit since well before the first Echo device shipped."[115]

[114] https://developer.amazon.com/en-US/docs/alexa/custom-skills/understand-name-free-interaction-for-custom-skills.html

[115] https://www.amazon.science/blog/tools-for-generating-synthetic-data-helped-bootstrap-alexas-new-language-releases

Rules of Grammar

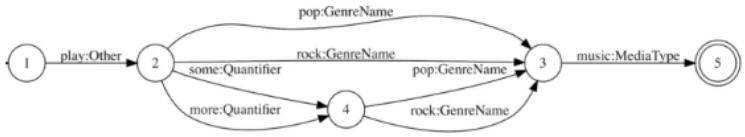
Grammars have been a tool in Alexa's NLU toolkit since well before the first Echo device shipped. A grammar is a set of rewrite rules for varying basic template sentences through word insertions, deletions, and substitutions.

Below is a very simple grammar, which models requests to play either pop or rock music, with or without the modifiers "more" and "some". Below the rules of the grammar is a diagram of a computational system (a finite-state transducer, or FST) that implements them.

```
pop = GenreName
rock = GenreName
music = MediaType
some = Quantifier
more = Quantifier

pop_rock = pop | rock
some_pop = (some | more)? pop_rock

INTENT Music.PlayMusicIntent = play some_pop music
```



A toy grammar, which can model requests to play pop or rock music, with or without the modifiers "some" or "more", and a diagram of the resulting finite-state transducer. The question mark indicates that the some_more variable is optional.

257. Each of the Accused Multi-modal Products and Services comprises accessing a plurality of domain agents that are associated with the context description grammar.

258. For example Amazon describes Alexa invoking custom skills.[116]

[116] https://developer.amazon.com/en-US/docs/alexa/custom-skills/understanding-how-users-invoke-custom-skills.html

About invoking custom skills

Users can invoke Alexa custom skills in two ways:

- Use a phrase supported by the Alexa service in combination with the invocation name for a custom skill to request information, ask a question, or tell Alexa to do something ("Alexa, *ask* Daily Horoscopes *for* the horoscope for Gemini"). See the following sections:
  - Invoke a skill with a specific request (intent)
  - Invoke a skill with no specific request (no intent)
  - Make sure that the sample utterances support the invocation phrases
  - Other words and phrases users can include in their utterances
- Indirectly invoke an Alexa custom skill through a name-free interaction. In this case, the user asks Alexa to perform a task, without naming the skill that should fulfill the request. ("Alexa, what is my horoscope?"). See the following section: Indirectly invoke a skill with name-free interaction.

Amazon describes a variety of methods for invoking skills.[117]

invocation

The act of beginning an interaction with a particular Alexa ability. For example, if a customer wants to wake Alexa to use the Horoscope skill, "Alexa, ask Horoscope for today's reading."
Alexa then follows up after the invocation and asks, "What horoscope sign would you like?"

Types of invocations include: full intent invocation, partial intent invocation, and no intent invocation.

full intent invocation

A customer's request that contains all information Alexa needs to make the request actionable. For example, "Alexa, ask History Buff what happened on June 3rd."

partial intent invocation

A customer's request that contains the customer's intent, but is missing a required slot. For example, "Alexa, ask Horoscope for today's reading." Here, the required zodiac sign is missing, and Alexa needs to obtain that information from the customer.

no intent invocation

A customer's request with no intent or slot information. For example, "Alexa, open History Buff."

259. Each of the Accused Multi-modal Products and Services comprises generating a relevance score based on results from comparing the text combinations to entries in the context description grammar.

260. For example, Amazon describes Alexa's two-step neural shortlisting-reranking approach to find the most relevant skill for a given utterance.[118]

---

[117] https://developer.amazon.com/en-US/docs/alexa/ask-overviews/alexa-skills-kit-glossary.html
[118] https://www.amazon.science/blog/the-scalable-neural-architecture-behind-alexas-ability-to-select-skills

Alexa uses a two-step, scalable, and efficient neural shortlisting-reranking approach to find the most relevant skill for a given utterance. This post describes the first of those two steps, which relies on a neural model we call Shortlister. (I'll describe the second step in a follow-up post tomorrow.) Shortlister is a scalable and efficient architecture with a shared encoder, a personalized skill attention mechanism, and skill-specific classification networks. We outline this architecture in our paper "Efficient Large-Scale Neural Domain Classification with Personalized Attention", which we will present at the 56th Annual Meeting of the Association for Computational Linguistics (ACL 2018) in July.

The shared encoder network is hierarchical: Its lower layers are character-based and orthography sensitive and learn to represent each word in terms of character structure or shape; its middle layers are word-based, and with the outputs from the lower layers, they learn to represent an entire utterance. The skill attention mechanism is a separate network that is personalized per user. It computes a summary vector that describes which skills are enabled in a given user's profile and how relevant they are to the utterance representation. Both the utterance representation vector and the personalized skill-summary vector feed into a battery of skill-specific classification networks, one network for each skill.

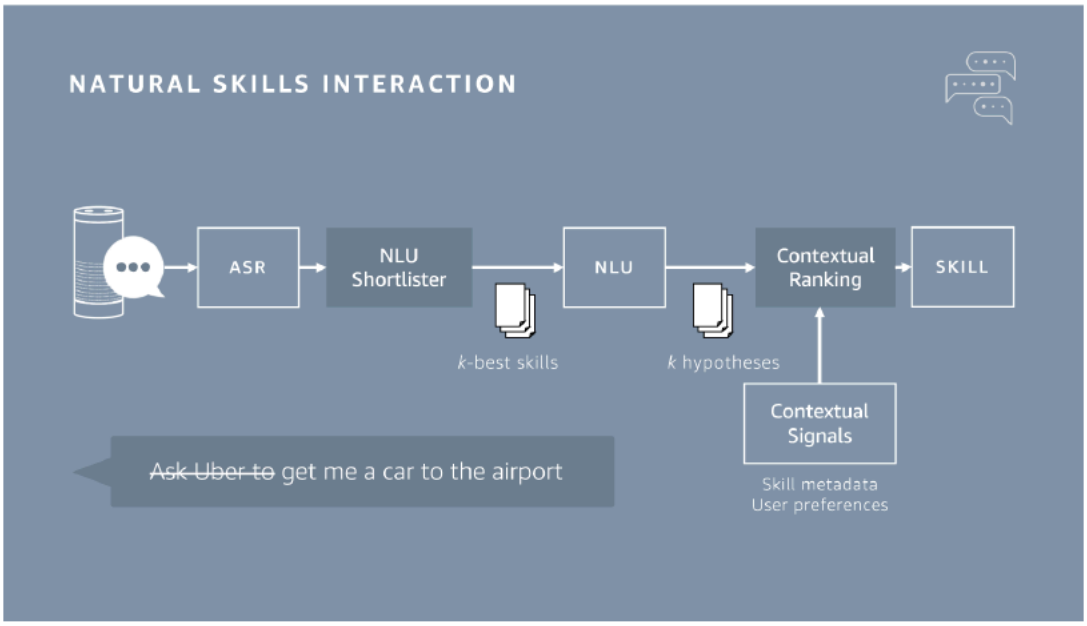
261. Each of the Accused Multi-modal Products and Services comprises selecting one or more domain agents based on results from the relevance score.

262. For example, Amazon describes how Alexa uses Shortlister and HypRank to select the most relevant skills.[119]

[119] https://developer.amazon.com/fr/blogs/alexa/post/0fecdb38-97c9-48ac-953b-23814a469cfc/skill-discovery

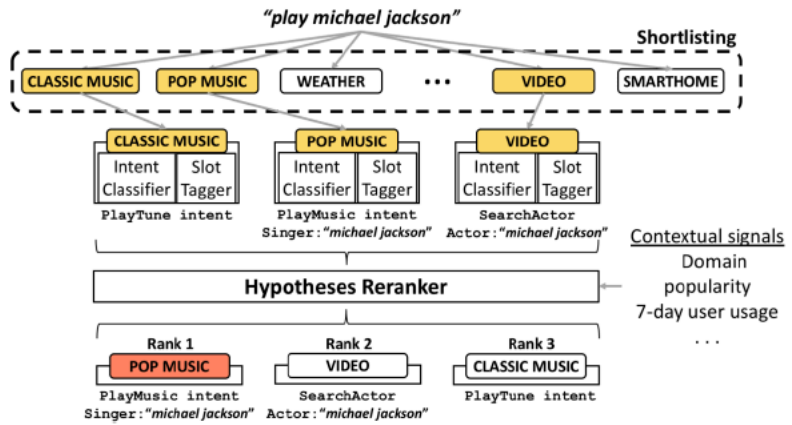
Surfacing the Most Relevant Skills via Shortlister and HypRank

It's important that customers are able to easily reengage with their skills – even if they describe what they want instead of calling the skill by name. Let's look again at the phrase "Alexa, remove an ink stain" where we have an utterance, "remove a {stain type} stain," and a slot value, stain type of "ink," that can be mapped to a skill's interaction model. To make this work, two techniques are employed to deliver the right experience. The first is a neural model architecture we call Shortlister, which solves a domain classification problem to find the most statistically significant (k-best) matches of candidate skills (details in a recent blog and research paper). The goal of Shortlister is to efficiently identify as many relevant skills as possible – it's high volume. The second is a hypothesis reranking network, we call HypRank. It uses contextual signals to select the most relevant skills – it's high precision (details in a recent blog and research paper). When combined, these two help classify which skill can provide the most likely correct response from an invocation name-free utterance. Beyond helping people reengage with the skills they've manually or automatically enabled, the Shortlister and HypRank surface specific and relevant skills to people that have never used them.



Amazon has further provided a specific example of the operation of the Hypotheses Reranker.[120]

[120] https://www.amazon.science/blog/hyprank-how-alexa-determines-what-skill-can-best-meet-a-customers-need



A high-level flow of the two-step shortlisting-reranking approach

The Challenge

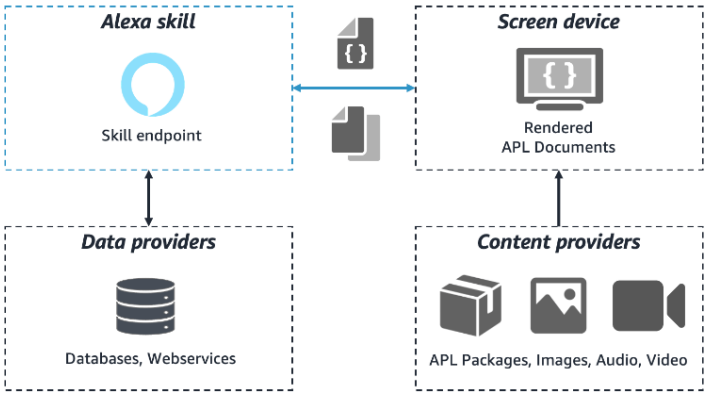
The problem here is essentially a domain classification problem over the k-best candidate skills returned by the shortlisting system, which we call Shortlister. The goal of Shortlister is to achieve high recall — to identify as many pertinent skills as possible — with maximum efficiency. On the other hand, the goal of the reranking network, HypRank, is to use rich contextual signals to achieve high precision — to select the most pertinent skills. Designing HypRank comes with its own challenges:

263. Each of the Accused Multi-modal Products and Services comprises obtaining content that is gathered by the selected domain agents.

264. For example, Amazon describes how Alexa collects information from content providers by an Alexa skill.[121]

[121] https://developer.amazon.com/en-US/docs/alexa/alexa-presentation-language/apl-bp-understand-apl-architecture.html

To plan your APL experiences, you need to understand the four technical *actors* that play a role in rendering APL *documents* on Alexa-enabled devices with screens. The following diagram shows a typical APL architecture at a high level.



APL architecture with four technical actors.

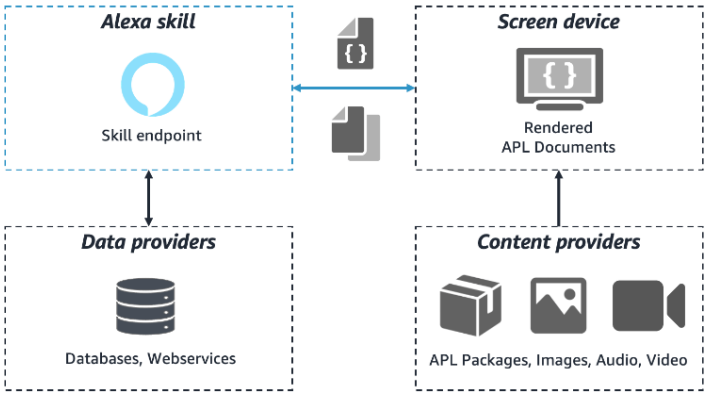
- **Alexa skill** — A skill endpoint that initiates an APL experience by sending out an APL *document* template in response to a regular skill request such as an `IntentRequest`.
- **Data providers** — APL templates often come with data that should be rendered with it so that the skill often attaches a JSON data source to the template that is sent to a screen device. This data typically doesn't reside within a skill, but resides in external data stores, such as a database that the skill reads from and writes to.
- **Screen device** — When the skill has sent an APL template with an attached data source, an Alexa-enabled device with a screen, such as an Echo Show or Fire TV, is responsible for inflating and rendering the template on screen.
- **Content providers** — APL templates that are sent by a skill often contain URL references to external content and media files; such as APL *packages*, images, audio and video files; that are stored publicly on the internet. An Alexa-enabled device with a screen resolves these references by downloading or streaming source files from external content providers, such as *content delivery networks* (CDN).

265. Each of the Accused Multi-modal Products and Services comprises generating a response from the content, wherein the content is arranged in a selected order based on results from the relevance score.

266. For example, Amazon describes how Alexa arranges the content from the content providers based on the APL template to be displayed on the screen device.[122]

[122] https://developer.amazon.com/en-US/docs/alexa/alexa-presentation-language/apl-bp-understand-apl-architecture.html

To plan your APL experiences, you need to understand the four technical *actors* that play a role in rendering APL *documents* on Alexa-enabled devices with screens. The following diagram shows a typical APL architecture at a high level.



APL architecture with four technical actors.

- **Alexa skill** — A skill endpoint that initiates an APL experience by sending out an APL *document* template in response to a regular skill request such as an `IntentRequest`.
- **Data providers** — APL templates often come with data that should be rendered with it so that the skill often attaches a JSON data source to the template that is sent to a screen device. This data typically doesn't reside within a skill, but resides in external data stores, such as a database that the skill reads from and writes to.
- **Screen device** — When the skill has sent an APL template with an attached data source, an Alexa-enabled device with a screen, such as an Echo Show or Fire TV, is responsible for inflating and rendering the template on screen.
- **Content providers** — APL templates that are sent by a skill often contain URL references to external content and media files; such as APL *packages*, images, audio and video files; that are stored publicly on the internet. An Alexa-enabled device with a screen resolves these references by downloading or streaming source files from external content providers, such as *content delivery networks* (CDN).

Amazon further describes how Alexa Shopping Actions API can add a shopping experience to a skill and how the API can further provide recommended products to a user in response to a request.[123]

[123] https://developer.amazon.com/en-IN/docs/alexa/alexa-shopping/alexa-shopping-actions-for-alexa-skills-api-reference.html#recommend-action

Request

To enable the customer to hear recommendations for specific products, add a `Connections.StartConnection` directive to your response.

Request example

```
{
    "type": "Connections.StartConnection",
    "uri": "connection://AMAZON.RecommendShoppingProducts/1",
    "input": {
        "products": [{
                "asin": "ASIN1"
            },
            {
                "asin": "ASIN2",
            },
            {
                "asin": "ASIN3"
            }
        ],
        "attribution": {
            "associateId": "some.associate.id.1",
            "trackingId": "some.tracking.id.1"
        }
    },
    "token": "RecommendProducts",
    "onCompletion": "SEND_ERRORS_ONLY"
}
```

267. Amazon has long known about the ’039 Patent.

268. Amazon has been aware of the ’039 Patent from multiple communications, presentations, and licensing negotiations between VoiceBox Technologies and Amazon between at least 2011 and 2017 and from VoiceBox Technologies personnel hired by Amazon.

269. Amazon knew or should have known that Amazon’s actions infringe one or more of the claims of the ’039 Patent because Amazon has the technical expertise to understand the scope and content of the ’039 Patent, because Amazon is a major provider of voice recognition products and services, and because Amazon knows the design, function, and operation of the Accused Multi-modal Products and Services, as well as the nature and extent of their use by others. At a minimum, Amazon has knowledge of the ’039 Patent at least as of the filing of this Complaint.

270. Further, on information and belief, Amazon has actively induced and/or contributed to infringement of at least Claim 13 of the ’039 Patent in violation of at least 35 U.S.C. § 271(b), and (c).

271. Users of the Accused Multi-modal Products and Services directly infringe at least Claim 13 of the ’039 Patent when they use the Accused Multi-modal Products and Services in the ordinary, customary, and intended way.

272. On information and belief, Amazon’s inducements in violation of 35 U.S.C. § 271(b) include, without limitation and with specific intent to encourage infringement, knowingly inducing businesses and consumers to use the Accused Multi-modal Products and Services within the United States in the ordinary, customary, and intended way by, directly or through intermediaries, supplying the Accused Multi-modal Products and Services to businesses and consumers within the United States, and instructing and encouraging such businesses and consumers to use the Accused Multi-modal Products and Services in the ordinary, customary, and intended way by activities related to selling, marketing, advertising, promotion, support, and distribution of the Accused Multi-modal Products and Services, which Amazon knew infringes at least Claim 13 of the ’039 Patent, or, alternatively, was willfully blind to the infringement.

273. On information and belief, Amazon’s inducements in violation of 35 U.S.C. § 271(b) further include, without limitation and with specific intent to encourage the infringement, knowingly inducing Amazon’s customers to commit acts of infringement with respect to the Accused Multi-modal Products and Services within the United States, by, directly or through intermediaries, instructing and encouraging such customers to import, make, use, sell, offer to sell, or otherwise commit acts of infringement with respect to the Accused Multi-modal Products and Services in the United States by activities related to selling, marketing, advertising, promotion,

support, and distribution of the Accused Multi-modal Products and Services, which Amazon knew infringes at least Claim 13 of the ’039 Patent, or, alternatively, was willfully blind to the infringement.

274. On information and belief, in violation of 35 U.S.C. § 271(c), Amazon’s contributory infringement further includes offering to sell or selling within the United States, or importing into the United States, components of the patented invention of and/or a material or apparatus for use in practicing at least Claim 13 of the ’039 Patent, constituting a material part of the invention. On information and belief, Amazon knows and has known the same to be especially made or especially adapted for use in an infringement of the ’039 Patent, and such components are not a staple article or commodity of commerce suitable for substantial noninfringing use.

275. Amazon is not licensed or otherwise authorized to practice the claims of the ’039 Patent.

276. Thus, by its acts, Amazon has injured Dialect and is liable to Dialect for directly and/or indirectly infringing one or more claims of the ’039 Patent, whether literally or under the doctrine of equivalents, including without limitation Claim 13.

277. As a result of Amazon’s infringement of the ’039 Patent, Dialect has suffered monetary damages, and seeks recovery, in an amount to be proven at trial, adequate to compensate for Amazon’s infringement, but in no event less than a reasonable royalty with interest and costs.

278. On information and belief, in addition to Amazon’s knowledge of the ’039 Patent as set forth above both prior to and as a result of the filing of this Complaint, Amazon has had, and continues to have, the specific intent to infringe, through its deliberate and intentional infringement or, alternatively, through its willfully blind disregard of the ’039 Patent by knowing there was a high probability of infringement but taking deliberate actions to avoid confirming that

infringement. The filing of this action has also made Amazon aware of the unjustifiably high risk that its actions constituted and continue to constitute infringement of the ’039 Patent. On information and belief, discovery will reveal additional facts and circumstances from which Amazon’s knowledge and intent to infringe (or willful indifference), both before and after the filing of this action, may be inferred.

279. Accordingly, Amazon’s infringement of the ’039 Patent has also been and continues to be deliberate, intentional, and willful, and this is therefore an exceptional case warranting an award of enhanced damages and attorneys’ fees and costs pursuant to 35 U.S.C. §§ 284 and 285.

280. Amazon’s infringement of Dialect’s rights under the ’039 Patent will continue to damage Dialect, causing irreparable harm for which there is no adequate remedy at law, unless enjoined by this Court.

SEVENTH COUNT
(Infringement of U.S Patent No. 9,495,957)

281. Dialect incorporates by reference the allegations set forth in Paragraphs 1–280 of the Complaint as though fully set forth herein.

282. The claims of the ’957 Patent are valid and enforceable.

283. The claims of the ’957 Patent are directed to patentable subject matter. Particularly, the ’957 Patent is directed to a novel, tangible voice recognition system. The inventive, tangible claimed structures of the ’957 Patent improve on the natural language recognition of a natural language utterance by a user. The claimed inventions provide specific concrete solutions to the problem of speech processing in existing systems.

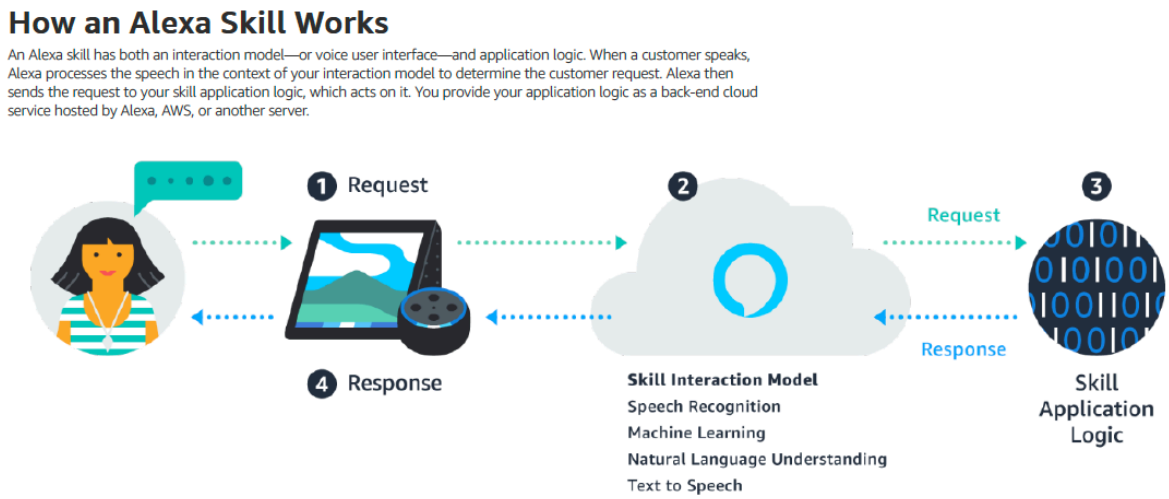
284. In violation of 35 U.S.C. § 271(a) and without authority from Plaintiff, Amazon has directly infringed by making, using, offering for sale, selling, and/or importing into the

Commonwealth of Virginia, this judicial district, and elsewhere in the United States products and services that embody the invention disclosed and claimed in the ’039 Patent, including at least the Accused Products.

285. Each of the Accused Products contains elements that are identical or equivalent to each claimed element of the patented invention pointed out by at least Claim 1 of the ’957 Patent.

286. Each of the Accused Products comprises a system for processing a natural language utterance, the system including one or more processors executing one or more computer program modules.

287. For example, Amazon’s Alexa devices comprise one or more processors executing one or more computer program modules. Furthermore, Amazon describes Alexa as processing natural language utterances. [124]



288. Each of the Accused Products comprises generating a context stack comprising context information that corresponds to a plurality of prior utterances, wherein the context stack includes a plurality of context entries.

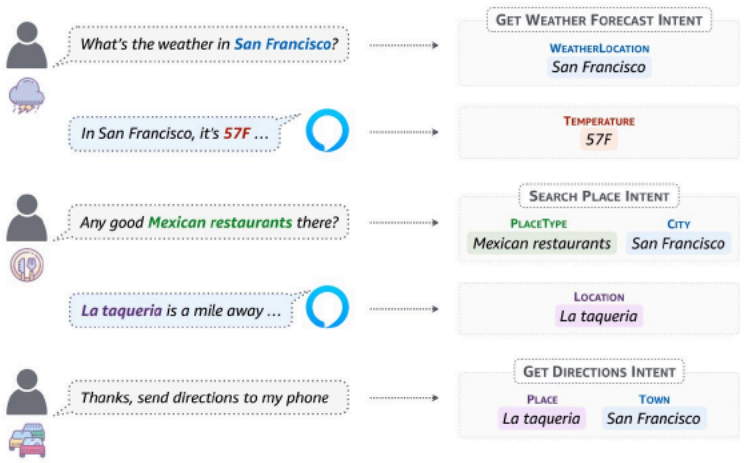
[124] https://developer.amazon.com/en-US/alexa/alexa-skills-kit/get-deeper/custom-skills

289. For example, Amazon describes the use of a shortlisting-reranking approach used by Alexa.[125]

Alexa uses a two-step, scalable, and efficient neural shortlisting-reranking approach to find the most relevant skill for a given utterance. This post describes the first of those two steps, which relies on a neural model we call Shortlister. (I'll describe the second step in a follow-up post tomorrow.) Shortlister is a scalable and efficient architecture with a shared encoder, a personalized skill attention mechanism, and skill-specific classification networks. We outline this architecture in our paper "Efficient Large-Scale Neural Domain Classification with Personalized Attention", which we will present at the 56th Annual Meeting of the Association for Computational Linguistics (ACL 2018) in July.

The shared encoder network is hierarchical: Its lower layers are character-based and orthography sensitive and learn to represent each word in terms of character structure or shape; its middle layers are word-based, and with the outputs from the lower layers, they learn to represent an entire utterance. The skill attention mechanism is a separate network that is personalized per user. It computes a summary vector that describes which skills are enabled in a given user's profile and how relevant they are to the utterance representation. Both the utterance representation vector and the personalized skill-summary vector feed into a battery of skill-specific classification networks, one network for each skill.

Amazon further provides examples of how Alexa uses records of past interactions to interpret a natural language request.[126]



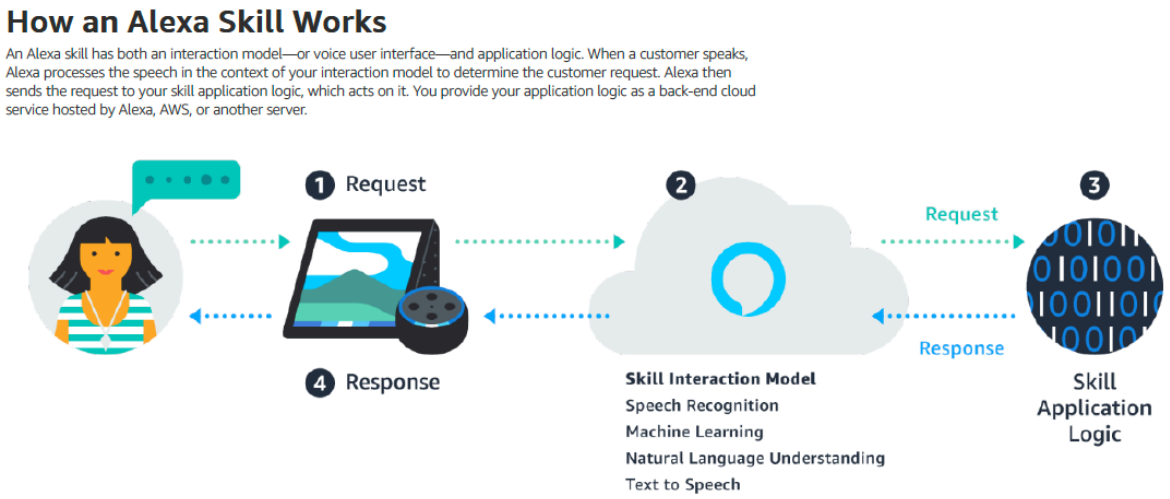
Different Alexa intents (weather, search, directions) might assign the same data (San Francisco) to different slots (WeatherLocation, City, Town). Tracking references by carrying slots over from one intent to the next is crucial for maintaining a coherent conversation.

[125] https://www.amazon.science/blog/the-scalable-neural-architecture-behind-alexas-ability-to-select-skills

[126] https://www.amazon.science/blog/how-alexa-is-learning-to-converse-more-naturally

290. Each of the Accused Products comprises receiving the natural language utterance, wherein the natural language utterance is associated with a command or is associated with a request.

291. For example, Amazon describes Alexa as receiving the natural language utterance, which includes a request.[127]



292. Each of the Accused Products comprises determining one or more words of the natural language utterance by performing speech recognition on the natural language utterance.

293. For example, Amazon discloses that Alexa transcribes the speech to text before processing the text to understand what the user means.[128]

[127] https://developer.amazon.com/en-US/alexa/alexa-skills-kit/get-deeper/custom-skills
[128] https://aws.amazon.com/blogs/aws/majority-of-alexa-now-running-on-faster-more-cost-effective-amazon-ec2-inf1-instances/

How Alexa Works

When I'm in my living room and ask Alexa about the weather, I trigger a complex system. First, the on-device chip detects the wake word (Alexa). Once detected, the microphones record what I'm saying and stream the sound for analysis in the cloud. At a high level, there are two phases to analyze the sound of my voice. First, Alexa converts the sound to text. This is known as Automatic Speech Recognition (ASR). Once the text is known, the second phase is to understand what I mean. This is Natural Language Understanding (NLU). The output of NLU is an Intent (what does the customer want) and associated parameters. In this example ("Alexa, what's the weather today ?"), the intent might be " `GetWeatherForecast` " and the parameter can be my postcode, inferred from my profile.

This whole process uses Artificial Intelligence heavily to transform the sound of my voice to phonemes, phonemes to words, words to phrases, phrases to intents. Based on the NLU output, Alexa routes the intent to a service to fulfill it. The service might be internal to Alexa or external, like one of the skills activated on my Alexa account. The fulfillment service processes the intent and returns a response as a JSON document. The document contains the text of the response Alexa must say.

294. Each of the Accused Products comprises identifying, from among the plurality of context entries, one or more context entries that correspond to the one or more words, wherein the context information includes the one or more context entries, wherein identifying the one or more context entries comprises: comparing the plurality of context entries to the one or more words.

295. For example, Amazon describes Alexa identifying one or more words based on the Alexa intent to assign the data to different slots.[129]

Different Alexa intents (weather, search, directions) might assign the same data (San Francisco) to different slots (WeatherLocation, City, Town). Tracking references by carrying slots over from one intent to the next is crucial for maintaining a coherent conversation.

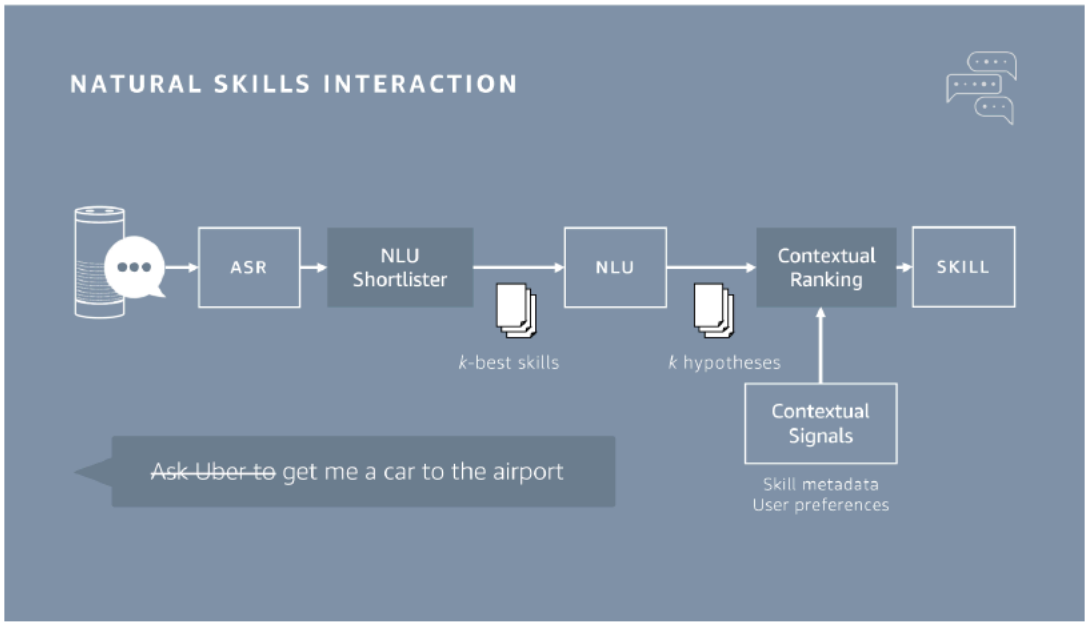
296. Each of Accused Products comprises generating, based on the comparison, one or more rank scores for individual context entries of the plurality of context entries.

[129] https://www.amazon.science/blog/how-alexa-is-learning-to-converse-more-naturally

297. For example, Amazon describes how Alexa uses Shortlister and HypRank to surface the most relevant skills.[130]

Surfacing the Most Relevant Skills via Shortlister and HypRank

It's important that customers are able to easily reengage with their skills – even if they describe what they want instead of calling the skill by name. Let's look again at the phrase "Alexa, remove an ink stain" where we have an utterance, "remove a {stain type} stain," and a slot value, stain type of "ink," that can be mapped to a skill's interaction model. To make this work, two techniques are employed to deliver the right experience. The first is a neural model architecture we call Shortlister, which solves a domain classification problem to find the most statistically significant (k-best) matches of candidate skills (details in a recent blog and research paper). The goal of Shortlister is to efficiently identify as many relevant skills as possible – it's high volume. The second is a hypothesis reranking network, we call HypRank. It uses contextual signals to select the most relevant skills – it's high precision (details in a recent blog and research paper). When combined, these two help classify which skill can provide the most likely correct response from an invocation name-free utterance. Beyond helping people reengage with the skills they've manually or automatically enabled, the Shortlister and HypRank surface specific and relevant skills to people that have never used them.



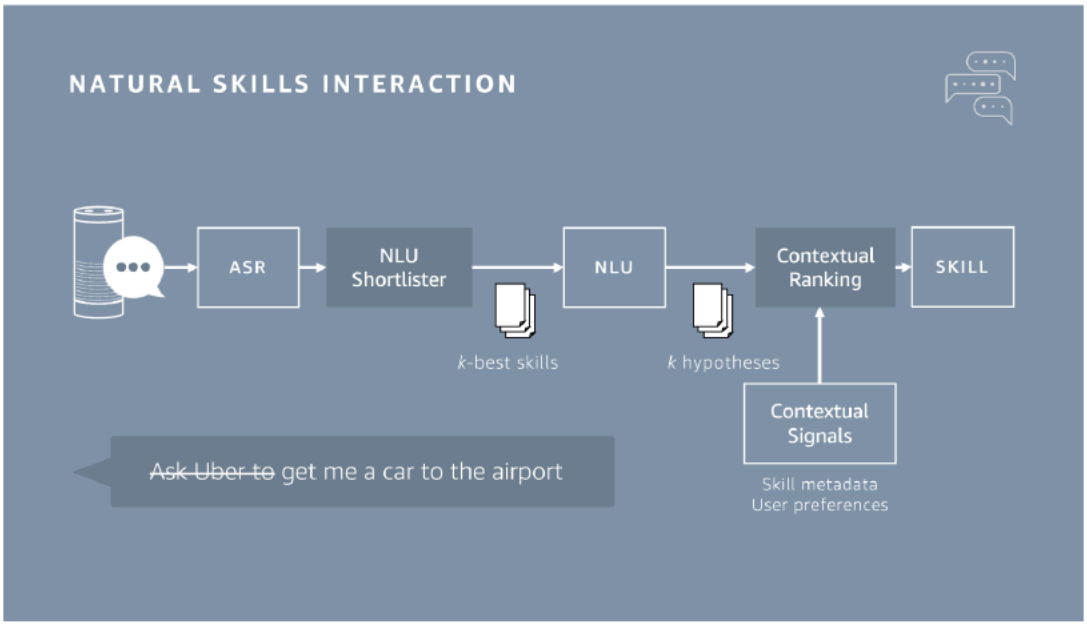
298. Each of the '957 Patent Accused Products comprises identifying, based on the one or more rank scores, the one or more context entries from among the plurality of context entries.

299. For example, Amazon describes how Alexa uses Shortlister and HypRank to identify the most relevant skills.[131]

[130] https://developer.amazon.com/fr/blogs/alexa/post/0fecdb38-97c9-48ac-953b-23814a469cfc/skill-discovery
[131] *Id.*

Surfacing the Most Relevant Skills via Shortlister and HypRank

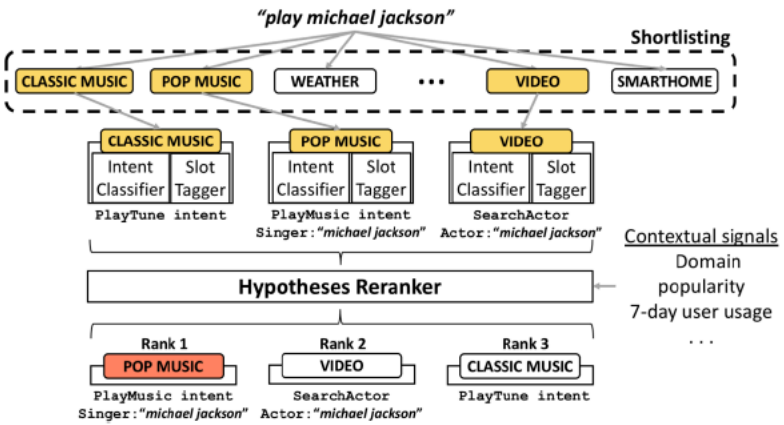
It's important that customers are able to easily reengage with their skills – even if they describe what they want instead of calling the skill by name. Let's look again at the phrase "Alexa, remove an ink stain" where we have an utterance, "remove a {stain type} stain," and a slot value, stain type of "ink," that can be mapped to a skill's interaction model. To make this work, two techniques are employed to deliver the right experience. The first is a neural model architecture we call Shortlister, which solves a domain classification problem to find the most statistically significant (k-best) matches of candidate skills (details in a recent blog and research paper). The goal of Shortlister is to efficiently identify as many relevant skills as possible – it's high volume. The second is a hypothesis reranking network, we call HypRank. It uses contextual signals to select the most relevant skills – it's high precision (details in a recent blog and research paper). When combined, these two help classify which skill can provide the most likely correct response from an invocation name-free utterance. Beyond helping people reengage with the skills they've manually or automatically enabled, the Shortlister and HypRank surface specific and relevant skills to people that have never used them.



300. Each of the Accused Products comprises determining, based on the determined one or more words and the context information, the command or the request associated with the natural language utterance.

301. For example, Amazon describes how Alexa uses the Hypotheses Reranker to determine the command in a specific example.[132]

[132] https://www.amazon.science/blog/hyprank-how-alexa-determines-what-skill-can-best-meet-a-customers-need



A high-level flow of the two-step shortlisting-reranking approach

The Challenge

The problem here is essentially a domain classification problem over the k-best candidate skills returned by the shortlisting system, which we call Shortlister. The goal of Shortlister is to achieve high recall — to identify as many pertinent skills as possible — with maximum efficiency. On the other hand, the goal of the reranking network, HypRank, is to use rich contextual signals to achieve high precision — to select the most pertinent skills. Designing HypRank comes with its own challenges:

302. Amazon has long known about the '957 Patent.

303. Amazon has been aware of the '957 Patent from multiple communications, presentations, and licensing negotiations between VoiceBox Technologies and Amazon between at least 2011 and 2017 and from VoiceBox Technologies personnel hired by Amazon.

304. Amazon knew or should have known that Amazon's actions infringe one or more of the claims of the '957 Patent because Amazon has the technical expertise to understand the scope and content of the '957 Patent, because Amazon is a major provider of voice recognition products and services, and because Amazon knows the design, function, and operation of the Accused Products, as well as the nature and extent of their use by others. At a minimum, Amazon has knowledge of the '957 Patent at least as of the filing of this Complaint.

305. Further, on information and belief, Amazon has actively induced and/or contributed to infringement of at least Claim 1 of the ’957 Patent in violation of at least 35 U.S.C. § 271(b), (c), and (f).

306. Users of the Accused Products and Services directly infringe at least Claim 1 of the ’957 Patent when they use the Accused Products in the ordinary, customary, and intended way.

307. On information and belief, Amazon’s inducements in violation of 35 U.S.C. § 271(b) include, without limitation and with specific intent to encourage infringement, knowingly inducing businesses and consumers to use the Accused Products within the United States in the ordinary, customary, and intended way by, directly or through intermediaries, supplying the Accused Products to businesses and consumers within the United States, and instructing and encouraging such businesses and consumers to use the Accused Products in the ordinary, customary, and intended way by activities related to selling, marketing, advertising, promotion, support, and distribution of the Accused Products, which Amazon knew infringes at least Claim 1 of the ’957 Patent, or, alternatively, was willfully blind to the infringement.

308. On information and belief, Amazon’s inducements in violation of 35 U.S.C. § 271(b) further include, without limitation and with specific intent to encourage the infringement, knowingly inducing Amazon’s customers to commit acts of infringement with respect to the Accused Products within the United States, by, directly or through intermediaries, instructing and encouraging such customers to import, make, use, sell, offer to sell, or otherwise commit acts of infringement with respect to the Accused Products in the United States by activities related to selling, marketing, advertising, promotion, support, and distribution of the Accused Products, which Amazon knew infringes at least Claim 1 of the ’957 Patent, or, alternatively, was willfully blind to the infringement.

309. On information and belief, in violation of 35 U.S.C. § 271(c), Amazon's contributory infringement further includes offering to sell or selling within the United States, or importing into the United States, components of the patented invention of and/or a material or apparatus for use in practicing at least Claim 1 of the '957 Patent, constituting a material part of the invention. On information and belief, Amazon knows and has known the same to be especially made or especially adapted for use in an infringement of the '957 Patent, and such components are not a staple article or commodity of commerce suitable for substantial noninfringing use.

310. On information and belief, in violation of 35 U.S.C. § 271(f)(1), Amazon's infringement further includes without authority supplying or causing to be supplied in or from the United States all or a substantial portion of the components of the patented invention of at least Claim 1 of the '957 Patent, where such components are uncombined in whole or in part, in such manner as to actively induce the combination of such components outside of the United States in a manner that would infringe the patent if such combination occurred within the United States.

311. On information and belief, in violation of 35 U.S.C. § 271(f)(2), Amazon's infringement further includes without authority supplying or causing to be supplied in or from the United States components of the patented invention of at least Claim 1 of the '957 Patent that are especially made or especially adapted for use in the invention and not staple articles or commodities of commerce suitable for substantial noninfringing use, where such components are uncombined in whole or in part, knowing that such components are so made or adapted and intending that such components will be combined outside of the United States in a manner that would infringe the patent if such combination occurred within the United States.

312. Amazon is not licensed or otherwise authorized to practice the claims of the '957 Patent.

313. Thus, by its acts, Amazon has injured Dialect and is liable to Dialect for directly and/or indirectly infringing one or more claims of the ’957 Patent, whether literally or under the doctrine of equivalents, including without limitation Claim 1.

314. As a result of Amazon’s infringement of the ’957 Patent, Dialect has suffered monetary damages, and seeks recovery, in an amount to be proven at trial, adequate to compensate for Amazon’s infringement, but in no event less than a reasonable royalty with interest and costs.

315. On information and belief, in addition to Amazon’s knowledge of the ’957 Patent as set forth above both prior to and as a result of the filing of this Complaint, Amazon has had, and continues to have, the specific intent to infringe, through its deliberate and intentional infringement or, alternatively, through its willfully blind disregard of the ’957 Patent by knowing there was a high probability of infringement but taking deliberate actions to avoid confirming that infringement. The filing of this action has also made Amazon aware of the unjustifiably high risk that its actions constituted and continue to constitute infringement of the ’957 Patent. On information and belief, discovery will reveal additional facts and circumstances from which Amazon’s knowledge and intent to infringe (or willful indifference), both before and after the filing of this action, may be inferred.

316. Accordingly, Amazon’s infringement of the ’957 Patent has also been and continues to be deliberate, intentional, and willful, and this is therefore an exceptional case warranting an award of enhanced damages and attorneys’ fees and costs pursuant to 35 U.S.C. §§ 284 and 285.

317. Amazon’s infringement of Dialect’s rights under the ’957 Patent will continue to damage Dialect, causing irreparable harm for which there is no adequate remedy at law, unless enjoined by this Court.

NOTICE

318. Plaintiff has complied with the notice requirement of 35 U.S.C. § 287 and has not and does not currently distribute, sell, offer for sale, or make products embodying the Asserted Patents.

PRAYER FOR RELIEF

WHEREFORE, Plaintiff prays for judgment and seeks relief from Defendants as follows:

a. For judgment that Amazon has infringed and continues to infringe the claims of the ’720, ’006, ’327, ’468, ’845, ’039, and ’957 Patents;

b. For a permanent injunction against Amazon and its respective officers, directors, agents, servants, affiliates, employees, divisions, branches, subsidiaries, parents, and all other acting in active concert therewith from infringement of the ’720, ’006, ’327, ’468,’845, ’039, and ’957 Patents;

c. For an accounting of all damages sustained by Plaintiff as a result of Amazon’s acts of infringement;

d. In the event Amazon is not permanently enjoined, for a mandatory future royalty payable on each and every future sale by Amazon of a product or service that is found to infringe one or more of the Asserted Patents and on all future products and services which are not colorably different from products and services found to infringe;

e. For a judgment and order finding that Amazon’s infringement is willful and/or egregious and awarding to Plaintiff enhanced damages pursuant to 35 U.S.C. § 284;

f. For a judgment and order requiring Amazon to pay Plaintiff's damages, costs, expenses, and pre- and post-judgment interest for its infringement of the '720, '006, '327, '468,'845, '039, and '957 Patents as provided under 35 U.S.C. § 284;

g. For a judgment and order finding that this is an exceptional case within the meaning of 35 U.S.C. § 285 and awarding to Plaintiff its reasonable attorneys' fees; and

h. For such other and further relief in law and in equity as the Court may deem just and proper.

DEMAND FOR JURY TRIAL

Pursuant to Rule 38(b) of the Federal Rules of Civil Procedure, Plaintiff demands a trial by jury in this action for all issues triable by a jury.

Dated: May 1, 2023

Respectfully Submitted,
/s/ Walter D. Kelley, Jr.

Walter D. Kelley, Jr. (VA Bar No. 21622)
Tara Zurawski (VA Bar No. 73602)
HAUSFELD LLP
888 16th Street N.W., Suite 300
Washington, DC 20006
Tel: (202) 540-7200
Fax: (202) 540-7201
wkelley@hausfeld.com
tzurawski@hausfeld.com

Garland Stephens
Robert Magee
Richard Koehl
Jeff Risher
BLUE PEAK LAW GROUP LLP
3139 West Holcombe Blvd.
PMB 8160
Houston, TX 77025
Tel: (281) 972-3036
garland@bluepeak.law
robert@bluepeak.law
richard@bluepeak.law
jeff@bluepeak.law
Telephone: 281-972-3036