International Journal of Innovative Research in Computer Science & Technology (IJIRCST)
ISSN: 2347-5552, Volume-9, Issue-3, May 2021
https://doi.org/10.21276/ijircst.2021.9.3.14
Article ID IRPD1156, Pages 77-82
www.ijircst.org

AI-Based Voice Assistance Using AWS

Lovekesh Kumar[1], Kanchan Yadav[2], Juhi Singh[3], and Khushboo Tripathi[4]

[1,2,3,4] Department of Computer Science and Engineering, Amity University, Gurgaon, Haryana, India

Correspondence should be addressed to Lovekesh Kumar; Kundu8007@gmail.com

ABSTRACT - Amazon Alexa is a voice-controlled pplication that is rapidly gaining popularity. In this paper user interactions with this technology, and focused on the types of tasks requested of Alexa, the variables that affect user behaviours with Alexa, and Alexa's alternatives. AI-based voice assistance using AWS is offering the users a way to acquire such competence. Particularly, we focus on developing skills for the Alexa assistant, as it is the most widespread. It's open a new world, a world where the user can talk to a machine as if it were a human and the machine will perform the work you request. Ideally, such conversations should be solely between the user and the voice assistance. For the hands-free feature that the user raved about and the other for speech recognition and understanding which is one key feature of the Echo.

KEYWORDS- Conversational Agents, Human information Behaviour, Human Information Interactions, Intelligent Personal Assistants, Voice-Controlled Agents

I. INTRODUCTION

AI-based voice assistance using AWS is a voice-controlled virtual assistant. It is inspired by the computer voice which ran on the star ship Enterprise i.e. Amazon Alexa which changing the way of interact with technology. A smart speaker that connects to the Amazon Alexa Voice Service. [1] The data about Alexa usage were collected via the online questionnaire and FQA (frequently questioned answers. It enables users to provide human information interaction. All through the use of voice-Z. Wake words activate Alexa to start recording. After that "Echo" sends those audio clips to the cloud where the request can be explicated. In addition to "Alexa" the Echo [11] devices also register the words "Echo", "Amazon" and "Computer". Most of the Alexa devices are consistently listening, once a wake word of choice is listen to by the device, you can get their attention by using wake words: Alexa, Echo, Amazon, etc. The default wake word is "Alexa", but you can change the default wake word within the Alexa app by going to:

- Menu
- Settings
- Device settings

You have to say the wake word followed by a command when you use Alexa on a smart speaker. An example of this would be, "Alexa, tell me about SAP (study abroad program)". Some older devices weren't completely hands-free they required to push a button to wake them up. Using Amazon Alexa on smartphones isn't quite as good as on smart speakers [2]. The smartphone app supports wake words, but only when the Alexa app is open and running. Additionally, the feature has to be turned on in the settings. You will have to open the console and tap on the Alexa button to use voice commands.

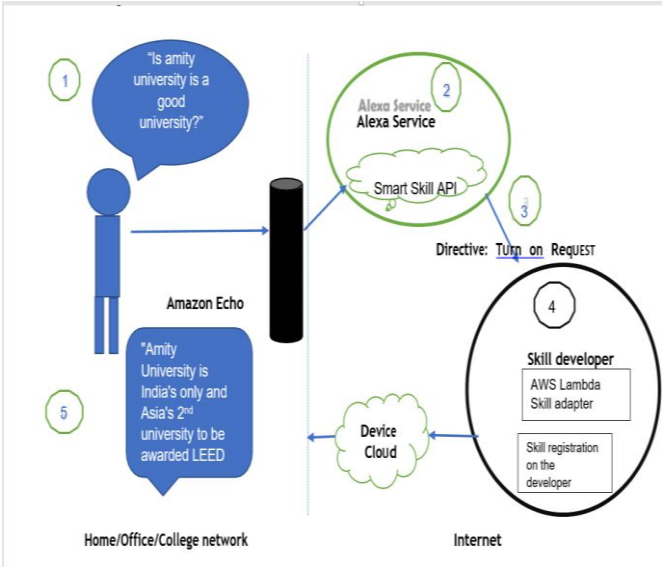

Fig. 1: Reserve order of Amazon's Alexa to receive and send information back to the device

The device hub sets the desired state of each device configured by the user.

1. Users say, "Is amity university is a good university."
2. Alexa Skill service receives the request and routes this intent to the Smart Skill API.
3. A directive is composed including the 'Turn-On-Request' name in the directive header and the appliance ID (located in the directive payload) corresponding to the friendly name.
4. The skill adapter hosted in AWS Lambda receives the directive. Included in the directive is an access token to determine the user's account making the request. A call is made to device cloud API to turn on the scene matching the appliance ID for the associated user.
5. Now, the device cloud receives a request from the skill adapter and communicates to a device hub or controller to turn on the scene preconfigured by the user.

In Fig. 1, the process starts with signal processing, which gives many chances to Alexa to make sense of the audio by cleaning the signal. If the wake word is detected, then a signal is sent in the speech recognition software in the cloud, which converts the audio clips into text format. In speech recognition output space is huge as it looks at all the words in the English language and the only technology which is capable of scaling in the cloud. To convert the audio into text, Alexa will analyse characteristics of the user's speech such as frequency and pitch to give you feature value. The sequence of the words will be determined by a decoder, then the decoder will determine that which is the most likely sequence of words from the given input features and the model, which split into two pieces. The first of these pieces is the prior, which gives you the most likely sequence based on a huge amount of existing text, without looking at the features, the other is the acoustic model which looking at pairings of audio and transcripts. These are integrated and zestful coding is applied, which has to happen in actual time.

II. LITERATURE SURVEY

The field of Amazon Alexa having speech recognition has seen some major advancements or innovations. Almost all the digital devices which are coming nowadays are coming with voice assistants which help to control the device with speech recognition only. A new set of techniques is being developed constantly to improve the performance of voice automated search. Google's Speech Recognition API program is used by the speech recognition module which is imported in python, this module is imported by using the command "import speech recognition as sr". This module is used to recognize the voice which is given as input by the user. [12]

STT is time-consuming because in this process firstly, the system has to listen to the user and different users have different systems where some are easy to understand while some are not easily audible. Total execution time depends on this step. Once the speech is converted to text executing commands and giving the results back to the user is not a time-consuming step. N M. A. Jawale (2019) et al. proposed in today's world, Python programming language has been used to evolve in many artificial intelligence applications. Programmers are broadly classified into three categories namely, novice users, knowledge intermittent, and expert one. This paper explores the use of voice recognition technology in the field of programming, specifically for the corresponding program with Python programming language.[9] In evaluation study, it found helpful for new Python programmers and come up with new learning inflect for programmers wherein beginner can experience nuisance untied program writing. Kishore Kumar R1 (2018) et al. presented to develop an economically effective and performance-wise efficient virtual assistant using python based on the concepts of Speech Recognition, Natural Language Processing and Artificial Intelligence.[8] People who are using it can give voice inputs and the device itself responds through voice commands by itself. The whole project is put in action through a python script which includes online Speech-Text conversion and Text Speech conversion codes written.[17] The device will casual manner so that the user has a friendly experience with the device and feels it like his or her own assistant.[16]

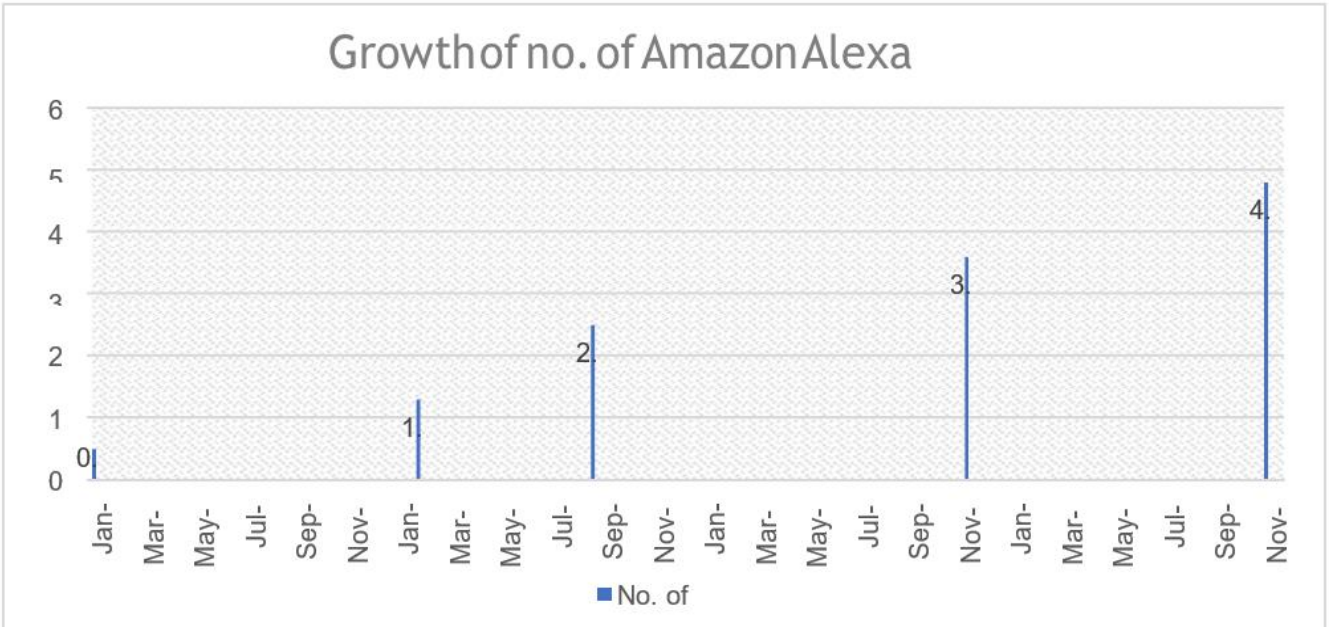

Fig. 2: Total number of Amazon Alexa skills from January 2016 to September 2019

In the above Fig. 2, Amazon's Alexa has gone from having 0.5 skills to over 4.8 skills as of September 2019. Introduced in November 2014, Alexa is a virtual assistant artificial intelligence technology developed by Amazon and used through Amazon's services. In today's era Virtual assistants like Amazon's Alexa are elevation becoming a customary place feature of many user electronics devices. As the worldwide adoption of digital voice-controlled agents continues to increase rapidly, so does the number of skills being performed by these devices. Alexa is the platform supported by the highest number of devices.

III. PROPOSED METHODOLOGY

The console of an Alexa developer permit developers to test and submit their skills for verification before they are made public to end-users. Once a skill is submitted for distribution, Amazon validates certain requirements.

A. Structure

Alexa Skill act as an action at the mention of a keyword set by the programmer. This is usually a greeting that has the assistant's name, like "Alexa". It uses natural language processing to decipher the command given to it, and then either answer your question or obey your command.
The proposed system will mainly consist of a mic, a speaker. At the backend, a database, voice recognition software, a voice sample, and the main Alexa skill will control and run all the components related to a system[7].

B. Working on the proposed model

The Alexa Skill responds to the commands given by the user. To accept the command, at first, the voice recognition tool of the system has to be awakened to accept and execute the request [4]. To awake the system, a user has to speak a wake word like "Location" or "Alexa" or by any other name that the programmer associates with the software. User needs to speak it's in the following ways i.e command/request/question just after he/she has said the wake-word. The wake-word initiates the voice recognition tool and all the words spoken after the wake-word gets accepted by the voice recognition- skill service.[14] This skill then analysis the input and bifurcate it as a request or question or command. After analysing and understanding the input, the software gathers the related information about the query from the internet. As soon as the skill gathers the necessary and relevant information, the system provides the information to the user.

Alexa Skill follows the structure outlined in **Fig. 3**., there is a speech communication between the user and the device. The device interacts with the skill (or application). Internally, two main elements, namely the skill interface and the skilled service communicate through JSON encoded messages.[5]

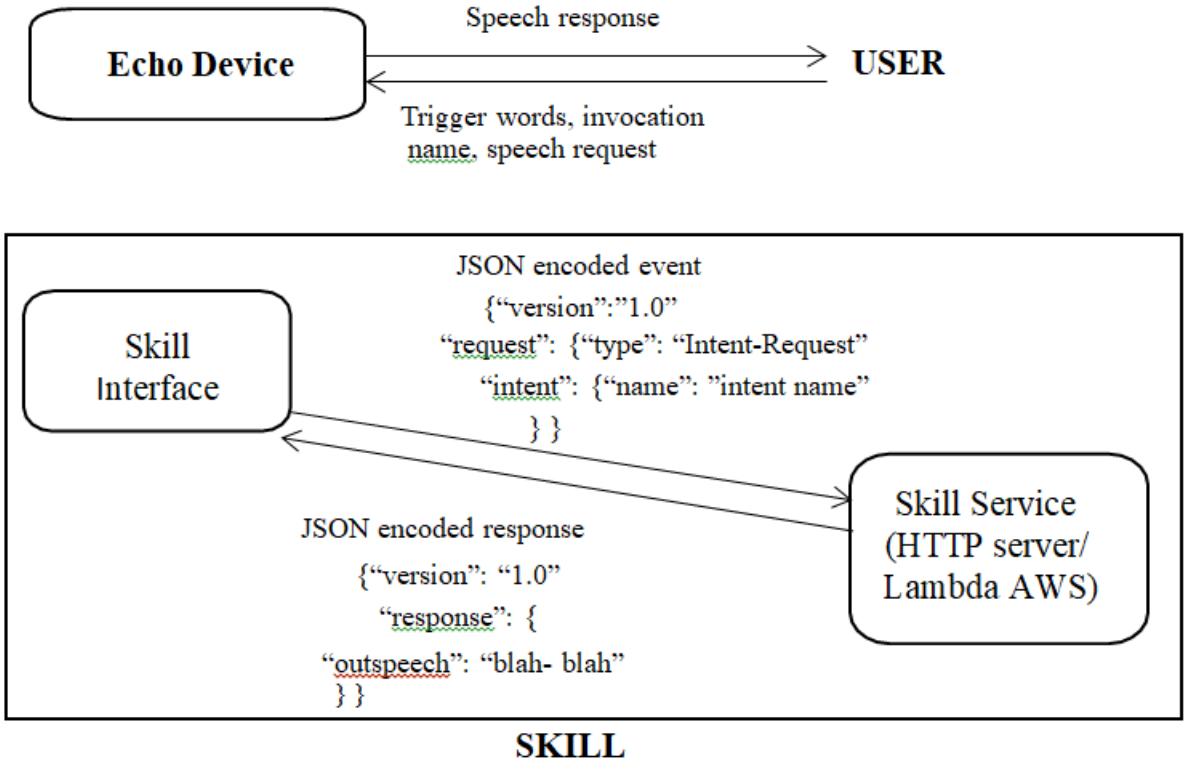

Fig. 3: Structure of an Alexa skill form, skill interface and the skilled service communicate through JSON ncoded

IV. WORKING PRINCIPLE

The proposed system connects to the Internet by using any wi-fi network. It's always on and listening for the wake word to wake the device up. Once it hears an audio clip, the device gathers the voice commands that heed after the wake-word and conveys a signal to a natural voice recognition service in the cloud called The Voice Service, which interprets them and sends back the appropriate response.[15] Most Alexa-operated devices are always listening, and you can get their recognition by using a wake word of choice: Alexa, Echo, Amazon, or Computer. [3] The default wake word is "Alexa", but you can change the wake word within the Alexa app, selecting your device from the list, and tapping the Wake Word option. To operate Alexa on a smart speaker, utter the wake word heed by a command. An example of this would be, "Is Amity University is a good university". Some primeval devices like the Amazon Tap requisite you to push a button to wake them up, so they weren't thoroughly hands- free.[10] The above command has 3 main parts:

- Wake word
- Invocation name
- Utterance

When users visualize 'Alexa' which awakens the device. The wake word poses the Alexa into the listening mode and is primed to take commands from the user. All the usage skills must have an invocation name to start them. The invocation name is the keyword used to trigger a distinct "skill". Users can integrate invocation names with an action, command, or question. The utterance decides what the user wants Alexa to perform. Utterances are locution the users will use when requesting Alexa. Alexa recognizes the user's objective from the given utterance and acknowledges accordingly. After this Alexa authorizes devices to convey the user's commands to a cloud-based service called Alexa Voice Service (AVS). All the complex operations such as Automatic Speech Recognition (ASR) and Natural Language Understanding (NLU) accomplish by the Alexa Voice Service which is the brain of the Alexa-enabled device. Alexa Voice Service processes the retaliation and recognizes the user's objective, if needed then it makes the web service request to a third-party server[13].

V. RESULTS

The speech interface was evaluated on the overall time, efficiency of speech to text, and the system was evaluated based on user response. User response is categorized as poor. satisfactory and good. Response Time and Performance:

Table 1: Result Table

Request type	Result
Approx Time	**7-8 sec**
No. of Queries	**80**
No of Hits	**72**
No of Failure	**8**
Percentage	**90%**

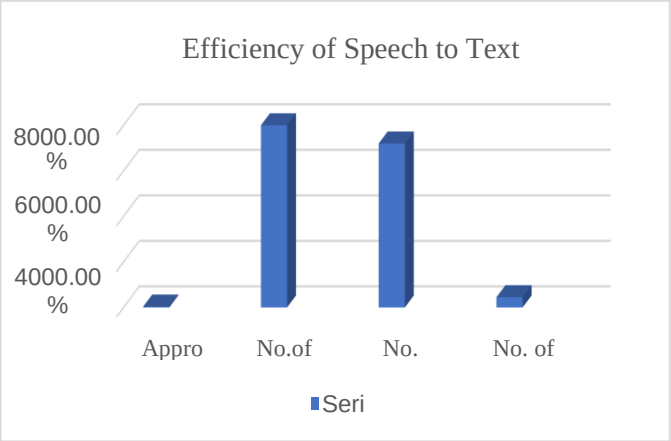

Fig. 5: The time and efficiency of speech to text conversion is dependent on the quality of the mic and internet speed

Using Table 1 and Fig. 5 Query execution time: Total time is taken to execute a query(end-to-end): 19 seconds. For the Alexa system, Echo devices are the input/output devices that are in the cloud. Audio is sent from an Echo device to Alexa in the cloud only when the wake word is detected or the action button is pressed.

VI. FUTURE WORKS

As speech recognition accuracy goes from, say, 95% to 99%, all of us in the room will go from barely using it today to using it all the time. Voice Assistants using AWS are overwhelmingly focused on natural language interfaces. The assistants that speak our language and communicate like a person have come to define the product class. This focus on natural language interface has distorted the distribution of assistants to distinct computing devices.

- Natural language understanding; speech synthesis.
- Contextual; associative communication

VII. CONCLUSION

The significant players in voice assiduity are contending to make voice a part of everyday life. First and foremost, they do this by massively discounting their smart

speakers and voice hardware so that it is affordable. However, they are also implementing voice wherever and whenever technologically possible. At Alexa Skill, we focus exclusively on delivering the richest and most meaningful analytics tools for our users. Our obsession is to empower users through compelling and actionable insights that drive measurable results for their study. You can compute on getting the wrest insights you entail without having to navigate through irrelevant content to get the information you need right now. This is because, at Alexa, we believe strongly in substance over style.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

[1] Lopatovska and H. Oropeza, "User interactions with 'Alexa' in public academic space," Proc. Assoc. Inf. Sci. Technol., vol. 55, no. 1, pp. 309–318, Jan. 2018.

[2] K. Pollmann, C. Ruff, K. Vetter, and G. Zimmermann, "Robot vs. voice assistant: Is playing with pepper more fun than playing with alexa?," in ACM/IEEE International Conference on Human-Robot Interaction, 2020, pp. 395–397.

[3] Lopatovska, "Personality dimensions of intelligent personal assistants," in CHIIR 2020 - Proceedings of the 2020 Conference on Human Information Interaction and Retrieval, 2020, pp. 333–337.

[4] Kei Hashimoto, Junichi Yamagishi, William Byrne, Simon King, Keiichi Tokuda, "An analysis of machine translation and speech synthesis in speech-to-speech translation system" proceedings of 5108978-1-4577-0539-7/11/$26.00 ©2011 IEEE.

[5] How to build a Hello World Alexa Skill. https://tutorials.botsfloor.com/how-to-build-hello-worldalexa-skill-bcea0d01ee8f

[6] MIT App Inventor, https://appinventor.mit.edu/

[7] Hirschberg, Julia, and Christopher D.(2015). Manning. "Advances in Natural Language Processing." Science 349 (6245): 261–266. doi:10.1126/science.aaa8685

[8] K. Srihari, V. Sakthivel, G. V. K. Reddy, S. Subhasree, P. Sankavi, and E. Udayakumar, Implementation of Alexa-Based Intelligent Voice Response System for Smart Campus, vol. 626. 2020.

[9] H. Phatnani, Mr. J. Patra and Ankit Sharma' "CHATBOT ASSISTING: SIRI" Proceedings of BITCON-2015 Innovations For National Development National Conference on Research and Development in Computer Science and Applications, E-ISSN2249–8974

[10] Roussev, V. Barreto, A. and Ahmed, I. (2016). API-Based Forensic Acquisition of Cloud Drives. Advances in Digital Forensics XII, pp. 213-235.

[11] 1redDrop. Amazon Echo sold 11 Million deviced by Dec 1, 2016: Morgan Stanley. http://1reddrop.com/2017/01/23/amazonecho-sold-11-million-devices-by-december-1-says- morgan-stanley/ [Accessed Jan. 2017].

[12] Gatner. Gartner Says Worldwide Spending on VPA-Enabled Wireless Speakers Will Top

[13] $2 Billion by 2020. https://www. gartner.com/newsroom/id/3464317 [Accessed Jan. 2017]

[14] Deepak Shende, RiaUmahiya, Monika Raghorte, AishwaryaBhisikar, AnupBhange, "AI Based Voice Assistant Using Python", Journal of Emerging Technologies and Innovative Research (JETIR), February 2019, Volume 6, Issue 2

[15] M. A. Jawale, A. B. Pawar, D. N. Kyatanavar, "Smart Python Coding through Voice Recognition", International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-10, August 2019.

[16] Hale, J. (2013). Amazon cloud drive forensic analysis. Digital Investigation, 10(3), pp. 259–265.

[17] "Automatic question generation for decision tree based state tying," Proc. of ICASSP, 1998

[18] "Speech recognition with flat direct models," IEEE Journal of Selected Topics in Signal Processing, 2010.

[19] Kei Hashimoto, Junichi Yamagishi, William Byrne, Simon King, Keiichi Tokuda, "An analysis of machine translation and speech synthesis in speech-to-speech translation system" proceedings of 5108978-1-4577-0539-7/11/$26.00 ©2011 IEEE.

ABOUT THE AUTHORS

Lovekesh Kumar has a Bachelor's of Technology in Computer Science and Engineering from Amity University. He worked with Python Programming Language and like many other Course in his Course of Education. He has two projects and four certifications in various areas of computer studies. His fields of interest are Python related Technologies (like Python-with ML, Django etc.)

Kanchan Yadav has a Bachelor's of Technology in Computer Science and Engineering from Amity University. She worked with Python Programming Language and like many other Course in his Course of Education. She has two projects and five certifications in various areas of computer studies. Her fields of interest are Python related Technologies (like Python-with ML, Django etc.)

Ms. Juhi Singh is currently an assistant professor at Amity University Haryana, India. She is research scholar at BMU, Rohtak, India. She has published more than 20 research articles in the reputed journals and presented her research in various conferences, and attended several national and international conferences and workshops. Her research interests include in Artificial Intelligence, Information security, Image processing and computational Intelligence.

Dr. Khushboo Tripathi has received her Ph.D. degree in Computer science from university of Allahabad, Allahabad. She has completed her M.Tech in Computer Science & Engineering from KNIT Sultanpur. She has done MCA from UPRTOU Allahabad, M.Sc and B.Sc. from University of Allahabad, Prayagraj, She has been in teaching and research for over 13 years and has got recognition in academic. She has good connect with industry and IT people. Her area of interest is Wireless Ad Hoc Networks, particularly, MANET and SENSOR networks, Advanced Networking, Security, Data Interpretation, Machine Learning and IoT. She has supervised

many

M. Tech., MCA and B. Tech students at Post graduation and graduation level for thesis and projects. She has published various papers in International and National reputed journals and conferences in India and Abroad. She is the senior member and reviewer of many professional organizations. She is reviewer of many reputed journals.