Virtual Assistants in Emergency Response Systems

Sanchari Biswas
Computer and Information Science
Temple University
sanchari.biswas@temple.edu

Inzamamul Islam
Electrical Engineering
Temple University
inzamam@temple.edu

***Abstract*—*The* collaboration between people and machines has grown largely for real-time applications since the inclusion of artificial intelligence into 21st-century technology. However, the advent of virtual assistants is still slow-paced in the emergency response systems infrastructure. In this paper, we analyse three popular Virtual Assistants — Amazon Alexa, Google Assistant, and Apple Siri — against practical parameters - accent, speed of speech, volume, and noise introduction - to deduce which of the three have a better accuracy and also to get a picture of whether the Virtual Assistants are ready to be introduced into the Emergency Response Systems Infrastructure.**

I. Introduction:

Virtual assistants have become an indispensable part of many households. When compared to manual operation schemes, the evolution of Voice Assistant agents has brought substantial changes in system work flow and process control. System performance increase, faster response time, precise and smart system control, real-time monitoring, data rendering at a faster pace, and resource management are all common characteristics of such adjustments. Virtual assistants are being extensively used around us for tasks such as shopping, creating to-do lists, generating alarms and notifications, and controlling smart equipment [1].

Emergency response are critical systems that serve as means for emergency response teams to locate and move resources to emergency sites.It seems perfectly reasonable to assume that this area would profit hugely by the assistance of Virtual Assistants. Patients are being introduced to internet-connected gadgets in a variety of ways, and VA is now one of them. Tracking health information, whether from fetal monitors, electrocardiograms, temperature monitors, or blood glucose levels, is critical for some patients. Many of these tests necessitate further engagement with a medical expert. This allows VAs to communicate more useful data to patients, reducing the need for direct patient-physician engagement. Another area the VA may help is reminding people to take their prescriptions.

According to a new study, virtual digital assistants like Siri, Alexa, and Google Assistant have the potential to give users with reliable and relevant information during medical emergencies, but their present versions aren't quite up to the task. In an experiment, the four most popular virtual digital assistants (VDAs) were quizzed on first aid for a variety of medical emergencies. Even when the virtual assistant got the inquiry right, the replies were frequently inaccurate [2]. While it's commendable that the VAs advised persons experiencing heart attack symptoms to dial 911, it would be much better if they also provided guidance on what to do while waiting for an ambulance [2].

Keeping this challenges in mind, we tried to understand why VAs are still not up to the mark in performance in emergency situations. For that we tried to analysis its performance in various situations. If we can find how VA's are interpreting the command it is getting and among them how many of the words it can understand correctly, it might be helpful while doing research with VA in this field. We tried to come up with an comparison process by which we can distinguish between understanding level of different voice assistant.

There are different types of VA's available of different company. Each of them has different algorithm to understand human language. Among them, we worked with Alexa, Google assistant, and Siri due to our availability. Besides thses are the most common VA's out of all others.

In this paper, we stared with brief history Virtual Assistant to understand VA better. Then tried to define the emergency situation where we want to make use of virtual assistant. Then took some existing research paper as a reference, tried to understand how they compared VA's. Keeping those ideas in mind, we tried to develop an automatic process where we can show a better result of comparisons.

II. Background and related work:

Though virtual assistants have become popular very recently but the idea of virtual assistants is pretty old. Because speech is humans' primary and most natural mode of communication, efforts to develop systems that can interpret and respond to spoken language have a long history. The earliest effective systems that identify speech automatically date back to the early 1950s. Researchers at Bell Laboratories built Audrey, the first true speech recognition device, in 1952. The use of pattern recognition algorithms in the 1960s was a turning point in the development of automatic voice recognition [3]. The Speech Understand Research program (SUR) at Carnegie Mellon University was started in the 1970s by the US Department of Defense Advanced Research Project Agency with the goal of developing and researching speech recognition technology. With the introduction of Apple's Siri in 2011, spoken dialogue systems have progressed from simple question-answering systems to more complex virtual assistants (VAs) that learn from users' speech. Speech interaction based on VAs can thus be voice-based, i.e. with a single modality (e.g. Amazon Echo), or voice-enhanced, i.e. with a multi-modal interface (e.g. Google Assistant) (e.g. Google Assistant). In the private context, current VAs aid the user in carrying out tasks via speech, typically in simple information retrieval or service execution activities [4].

A. Use of VA in Emergency Critical system:

Critical situations are serious, unexpected, and often dangerous situations requiring immediate action. To handle critical situations we often rely on those system which can be trustworthy in any circumstances, usually defined as Critical system. So, Critical systems are highly reliable systems that should retain their reliability as they evolve without incurring prohibitive costs. On the other hand, Safety-critical systems (SCS) or life-critical systems are a sub-domain of critical system, those systems whose failure could result in loss of life, significant property damage, or damage to the environment. There are lots of scopes of SCS, among them Infrastructure (eg-Circuit breakers, Fire alarms, Life support systems, Telecommunications) Medicine (eg-Emergency response systems, Robotic surgery, Pacemaker devices, Healthcare information systems), Recreation (eg-Parachutes, Scuba equipments), Transport (eg-Railway signalling, Airbag systems, Aircrew life support systems, Launch vehicle safety systems).

A continually listening VA device shall be able to detect an emergency situation and then act on that. An critically endangered individual could be just a shout away in any room of the house, through VA's, Internet-connected car, or smartphone. Smart speakers can now alert you to impending severe weather conditions, call a friend for assistance, provide basic medical information, assist you in alerting someone when you've fallen, and detect burglars. That is the extent of their emergency response capabilities.

B. Related Work:

Various research is going on the usability of VA in emergency critical situations. For that we also need to understand how they are performing to understand human voice. There might be a different scenario where sound quality might be compromised. As we want to use VA in critical emergency situation, we need to understand its comprehension level and where we might need to avoid using it.

Virtual assistants (VAs), such as Alexa, Google Assistant, and Siri, are Artificial Intelligence assistance devices that replicate human communication and perform web-based searches and other requests [5]. The VAs' inability to offer correct health information is hampered by their faulty comprehension of complicated medical language syntax [6]. The ability to detect unique words spoken by different persons is a vital basis of any medical contact, as like any relevant healthcare advice can only be given once the language has been properly understood by service people. The voice recordings from the 46 participants were used in the first publication. All of the participants were fluent in English, with 12 of them having foreign accents that were distinct from a "Canadian accent." Four of the 12 individuals with non-Canadian foreign accents were British, three were Spanish/Filipino, two were African, two were Eastern European, and one was Chinese. Then using two table (one is for the brand name and other one is for the generic name) audio command was played. Each speech recording was played back from a laptop using a Jabra Speak 410 speaker placed directly next to the VA devices' microphones during analysis. A 4th-generation Amazon Echo smart speaker was used to test Alexa; a Google Pixel 4a smartphone was used to test Google Assistant; and an iPhone 7 smartphone was used to test Siri. The latest software was installed on all hardware devices, and the device language was adjusted to English (Canada) [7]. This study found no significant interactions of participant accent in comprehension accuracy. Participants with a Canadian accent and those with a foreign accents saw the same impact. From this work, we garner the importance of difference in accents as a significant factor while considering the responsiveness of Virtual Assistants.

There has been existing work [8] that compared virtual assistants against cancer screening terminologies shows that there has been little research done to evaluate the quality of health data given by these devices . In May 2020, five investigators used their personal iPhones to perform the investigation in the San Francisco Bay Area. Four of the five investigators were native English speakers (2 males, 3 women). There are notable discrepancies across the four most popular voice assistants(Alexa, Siri, Google Assistent and Cortata) when it comes to responding to questions concerning cancer screening, and there is opportunity for improvement across the board. Their suggestion was to guarantee that assistants deliver reliable information, software developers should consider collaborating with health professionals, particularly guideline developers and evidence-based medicine practitioners [8].

We also considered existing work that investigates personalized voice characters for in-car speech interfaces [9]. The findings demonstrate that personalized assistants are well received when they are tailored to the preferences of the users. Assistants need a more serious tone for driving-related use cases than for amusement use cases [9]. Statement of Contribution the paper shared, findings from a real-world driving research (N=55) in which it evaluated the impact of personalized voice assistants on trust, user experience, acceptance, and workload to that of a non-personalized voice assistant. This article begins with collecting requirements for in-car assistant personality qualities in order to investigate customised voice assistants for driving scenarios. Then we created a set of assistants, which we tested in a real-world driving scenario [9]. The authors of this paper invited 19 people, 12 of them were men, ranging in age from 19 to 53 years old (M=35, SD=11). Then they came up with six situations, three of which were related to driving and three of which were tied to entertainment. Participants had the opportunity to converse with each of the eight speech aides. The responses of the assistants were pre-recorded by a voice actress for each helper and matched their position in the model. They had participants complete out a MeCue and an Acceptance Scale questionnaire after each visit [10]. Finally, they held a semi-structured interview with the participants. The results shows that the correct matching of assistant characters to the user's personality is a crucial prerequisite for positive effects of penalization. It reflects on how we can improve in-car voice interaction through personalising [9].

None of these papers have interrogated the Virtual Assistants against parameters such as volume, speed of speech, or noise tolerance. But, in our scenario, emergency response situations, each of the above plays an important factor alongside accents. Volume of the commands played to the VAs are influenced by how far or near the subject is from the device. In probable cases, the subject might be incapacitated and hence incapable of moving closer to or further away from the device. The speed of speech is also influenced by stressful or other such situations [11]. Similarly, external noises, eg. police siren, commotion, vehicular sounds, etc, are common inclusions to the original audio in emergency situations. We will go into deeper detail later in the III section discussing each of these parameters.

III. Methodology:

Our work deals primarily with the three Virtual Assistants: Amazon Alexa, Google Assistant, and Apple Siri. We analyse them based on how accurately they transcribe voice commands. Our analysis is based on the parameters: accent, speed of speech, volume, and noise tolerance. Figure 1 shows the block diagram of our entire workflow.

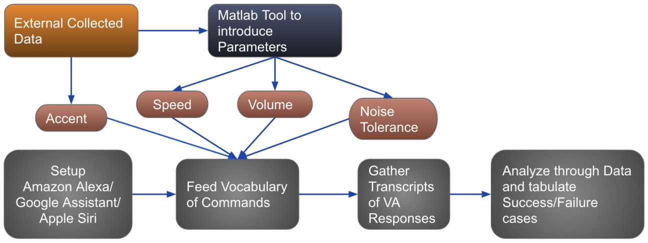

External Collected Data
Matlab Tool to introduce Parameters
Accent
Speed
Volume
Noise Tolerance
Setup Amazon Alexa/ Google Assistant/ Apple Siri
Feed Vocabulary of Commands
Gather Transcripts of VA Responses
Analyze through Data and tabulate Success/Failure cases

Figure 1: *Block diagram of the workflow of data collection, analysis, and tabulation of the Virtual Assistants*

We begin our analysis methodology from the setup step. In this step, we zero-ize and initialize our three VAs - Amazon Alexa, Google Assistant and Apple Siri. This ensures setting a common ground and involves removing any voice recognition or personalization profiles, adjusting volumes and sensitivity to an optimum base level, and placing all the three VAs in nearly identical locations. After the setup is completed, our understanding involves getting the information regarding accuracy of transcripts back from the VAs. For this reason, we need to have the VAs listen to our audio commands and recognize them. This brings into picture our next few functional blocks.

In order to play the commands to the VAs with minimal manual interaction and bias, we proceed to the external collection of data. This involves collecting our data from an external audio archive [12], and manipulating this data per our requirement and our analysis factors. This is performed in the 'MATLAB Tool to Introduce Parameters' functional block. We see from the pipelines that this block incorporates our speed, volume, and noise factors within the audio from the archive with accent already as a factor. These modified audio files are then played, using the same MATLAB tool, in a continuous loop with appropriate pauses to account for the VA's response time. This response, in the case of our analysis, is immaterial and we only concentrate on the transcripts of the audio the VAs listen to.

Moving on our flow diagram, the next functional block we encounter is the 'Gather Transcript of VA Responses' block. Most of the Virtual Assistants have the transcripts saved for all the audio commands that have been played to them. These transcripts are, in most cases, sorted by latest time. Our plan of action involves acquiring a copy of these transcripts and matching them to the audio commands played to them to analyse and tabulate their accuracy. Each of the Virtual Assistants have a different interface and, hence, a different set of steps to acquire the transcripts. We provide more detail on these steps later in Section III-E.

Following the transcript gathering, we reach our final step in the analysis, where we analyse and tabulate the data collected. We perform this step in order to extract meaningful numbers, percentages, and comparison graphs from the collected vast raw data. We use a number of tools in this step as explained later in Section IV, but most of our operations are performed on Numbers Version 11.2.

In the upcoming sections, we provide a further detailed explanation of each step.

A. Setup Virtual Assistants:

We start our workflow with the setup of the three Virtual Assistants.

1) Amazon Alexa: The Alexa device we used is an Echo Dot (3rd generation). For Alexa, we removed the voice recognition profile, because we wished to play audios for it which were not in our voices and wanted Alexa to be able to recognize them. After that, we simply rebooted the device and once it successfully came up, we proceeded to our data generation for VAs step.

2) Google Assistant: We used a Google Mini device (2nd generation) for command interpretation. We did not need to remove any voice recognition, because Google is more generic to reply for any command it gets. After that, proceeded to our data generation for VAs step and collected the script.

3) Apple Siri: We used Siri from Iphone. For that before every command, we needed to concatenate our own voice “Hey Siri” with our rest command. Because Siri does not reply for other voice commands. However, that will not affect our research, because anyone who is going to get the assistant, he/she will going to use his/her own devices.

B. External Collected Data:

Finding medical data sets is difficult due to the fact that medical data is very well protected by privacy regulations all around the world. Aside from that, we wanted to prevent any emergency instruction that may prompt a call to 911. That's why we start with a very generalized data set. Our main focus to understand which VA can understand better in different scenario, so that we can trust it better.

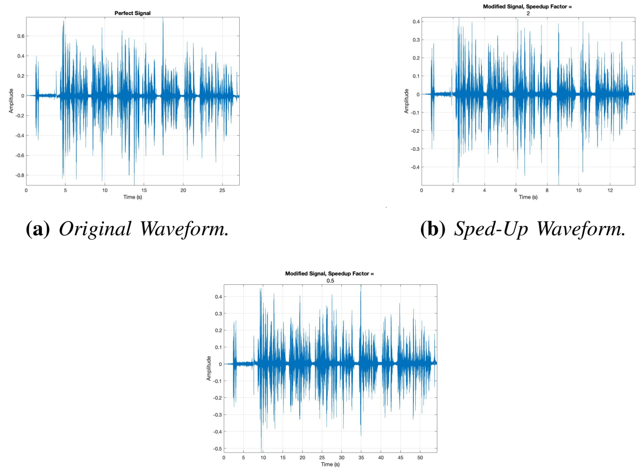
We used the exact same voice command from a website[12], where four participants of four different accents were chosen.

For the selection of the sentences, we tried to consider to collect sentence which are more relevant to real word command. For example: one sentence we choose "it was a sign from the gods to foretell war or heavy rain", it has some word "War" and "heavy rain". How VA's are understanding this words, give us an idea how VA's response will be in an disastrous situations. The other sentence we used “The letter implied that the animal could be suffering from a rare form of foot and mouth disease”, which has some similar words that we could relate with disease. Then we used voice chopper to collect our desired command part from that part. Then using matlab[13] simulation, variation in data has been generated.

C. Data Generation for VAs:

We use MATLAB [13] to modulate the input audio files, according to our parameters: speed of speech, volume, and noise tolerance. We then play them back to our three Virtual Assistants and wait for their response.

1) MATLAB [13] Code:: Our MATLAB [13] code starts with the input of the audio files. We then use the audioread [13] function to read in the audio file as a sampled data part and a sampling rate part. Next, we create a function that grabs the audio files one after another and processes them per each of our parameters and plots as well as plays back the output using the playblocking [13] function. This function blocks the program execution until our device output device (in this case, our laptop speaker) has finished playing the entire audio. Only then, it moves on to the next part of the program. If we do not use this function, we would get overlapping audio, where one play starts before the previous audio has finished playing. After playback of the audio, we use the pause [13] function, which takes in a numeral in seconds as its argument. This function pauses the execution for a certain time period. We use this function to account for the time the Virtual Assistant will take to respond to our command. Alongside playing the audio file, we also plot it with the time on the X-axis and the amplitude on the Y-axis.



(a) *Original Waveform.*

(b) *Sped-Up Waveform.*

(c) *Sped-Down Waveform.*

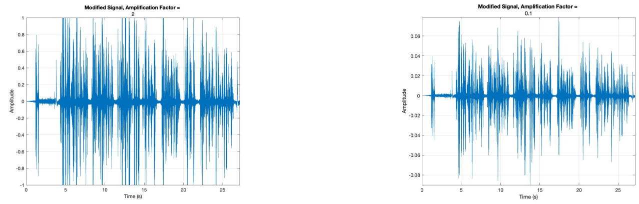
Figure 2: *Graphs according to speed of output.*

Figure 2a shows the amplitude vs time plot for an original unaltered audio waveform with a South-Western accent. We do not alter any aspect of this waveform and play it back to the external speaker as-is. This indicates that the data we collected in Section III-B is being plotted and played back here, without any change in amplitude or frequency, i.e, without being sped up, slowed down, increased or decreased volume, or any added noise.

Figure 2b shows the amplitude vs time plot for the sped-up waveform. What it means in a real life scenario is, people often talk faster or slower when faced with certain situations [11]. Here, we use the stretchAudio [13] function to increase or decrease the frequency of the waveform and then plot and play it back. In figure 2b, we speed the audio up twice for visualization reasons but in our original setup, we have increased the speed parameters in the range of 1.1 to 1.6 times, with an 0.1 increment. This is to determine the speed level at which each of the three Virtual Assistants stop understanding. We start at 0.1 because despite this being almost unnoticeable, we needed to start at a base level, which was faster than normal but yet understandable to all three Virtual Assistants. We then, with our gradual increments, determine the threshold for each of the three VAs.

For the sped down part, we start at the speed parameter of 0.1 and increment it with 0.1 increments up to 0.9. In figure 2c, we show the amplitude vs time plot of an audio waveform sped down to the factor of 0.5. Here again, although 0.1 sounds like an unrealistically low volume, we use it as a starting point where no VA can properly recognize and transcribe a command, and then with gradual increments deduce the threshold of each VA.

This modification of speed (frequency) makes the audio that is being played back to be faster (or slower), i.e, lesser (or greater) time required to pronounce words, including the pauses in between. While this is not a very realistic speech speed-up (or slow-down) simulation, this gives us a rough closeness as to how the VAs handle the situation, because our range of increment (or decrement) of speed of speech is in a low, realistic range.



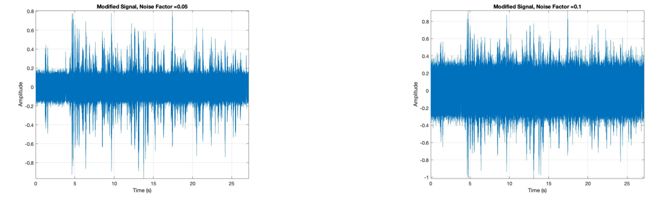
(a) *Amplified Up Waveform.*

(b) *Amplified Down Waveform.*

Figure 3: *Graphs according to amplitude of output.*

Figure 3a shows the amplitude vs time plot for the amplified waveform. In this case, we multiply the input signal, by an amplitude factor, and then plot and output our resulting waveform. We have started here with a range of 0.1 and gradually incremented in steps of 0.2 up to 0.9, and then again from 1.2 up to 2. We do this for the same reason as of the speed up, to start from a baseline when all the three VAs can either perfectly recognize the commands or all the three VAs totally fail to acknowledge the command. Our short steps then serve as a confirmation that we get closest in threshold to when and how well each of the VAs switch in their performance. In figure 3a, the amplitude factor that we have multiplied with is 2. In figure 3b, the amplitude factor is 0.1.

This modification of volume (amplitude) makes the audio that is being played back to be louder (or lower) in volume. This, in real life situation, takes into consideration the fact that wounded individuals might have lower volume of speaking or might be far from or unable to get closer to the VA device. Similarly, in an enclosed space, such as a vehicle, the volume might be louder than in normal settings.



(a) *Noisy (5% added) Waveform.*

(b) *Noisy (10% added) Waveform.*

Figure 4: *Graphs according to noise incorporation in output.*

Figure 4a and figure 4b show the amplitude vs time plot for the noisy wave forms with 5 and 10 percentages respectively of white Gaussian noises introduced. Noises are the most obvious parameters to consider, especially in an emergency situation. Noises are any audio input to the VA, other than the desired command audio. There might be a variety of overlapping noise sources to the original, eg. a police siren, people's conversations, and all other kinds. In our experiment setup, we start with a noise of 0.2 percent. We then increase it in steps of 0.2 up to 6 percent. We do this to start at a low noise value, where the command is recognizable accurately by all the three VAs, and we can deduce very closely the threshold where each of the VAs cannot recognize with accuracy anymore. Our noise has a frequency of between 725 and 1600 Hz, which is the frequency range of a Rumbler siren [14]. This provides us with a realistic measure of noise interference. We use Gaussian white noise to serve as a generalization of all the kinds of noises there can be, but we can also replace this with specific sound files to see the output for specific category of noises.

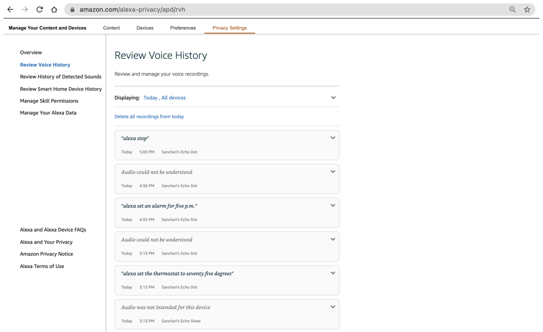
D. Feed Commands to VA:

After the initial setup of our VAs (Amazon Alexa, Google Assistant, and Apple Siri), we play our MATLAB [13] program to each of the three for all the audio commands. The pause [13] function enables the VA to complete its response, but we do not take note of this response, as we are only concerned with the transcript of our command as understood by the VA, word by word. Once we have finished playing all the audio files, modified by all of our parameters, it is time to collect the transcripts, which is described in detail in the next section.

E. Gather Transcripts of VA Responses:

Each of our VAs have a distinct way of collecting the transcripts of the commands played to them over time. We collected these transcripts corresponding to all the commands we played and then proceeded to our next part of action, result analysis. We discuss next, in brief, the collection procedure of the transcripts of our VAs.

1) Amaxon Alexa: To access the command history of an Alexa device, we navigate to the Content and Devices section of the Amazon profile for that device. Under Privacy Settings, we select Alexa Privacy. We find our command history under the Review Voice History section. This section looks like figure 5.



amazon.com/alexa-privacy/apd/rvh

Manage Your Content and Devices | Content | Devices | Preferences | Privacy Settings

Overview
Review Voice History
Review History of Detected Sounds
Review Smart Home Device History
Manage Skill Permissions
Manage Your Alexa Data

Review Voice History

Review and manage your voice recordings.

Displaying: Today , All devices

Delete all recordings from today

"alexa stop"
Today 5:00 PM Sanchari's Echo Dot

Audio could not be understood
Today 4:36 PM Sanchari's Echo Dot

"alexa set an alarm for five p.m."
Today 4:35 PM Sanchari's Echo Dot

Audio could not be understood
Today 3:13 PM Sanchari's Echo Dot

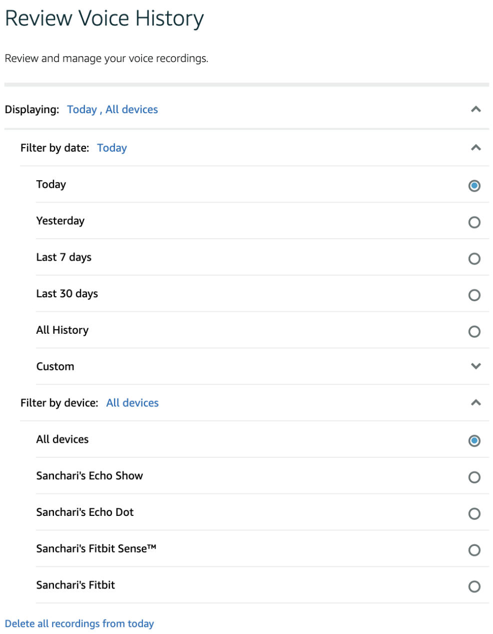
"alexa set the thermostat to seventy five degrees"
Today 3:13 PM Sanchari's Echo Dot

Audio was not intended for this device
Today 3:13 PM Sanchari's Echo Show

Alexa and Alexa Device FAQs
Alexa and Your Privacy
Amazon Privacy Notice
Alexa Terms of Use

Figure 5: *Review Voice History section of Amazon Alexa*

This voice history can be filtered based on the device the commands are played on and the time frame as shown in figure 6.



Review Voice History

Review and manage your voice recordings.

Displaying: Today , All devices

Filter by date: Today

Today ●
Yesterday ○
Last 7 days ○
Last 30 days ○
All History ○
Custom

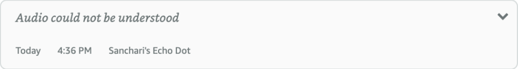
Filter by device: All devices

All devices ●
Sanchari's Echo Show ○
Sanchari's Echo Dot ○
Sanchari's Fitbit Sense™ ○
Sanchari's Fitbit ○

Delete all recordings from today

Figure 6: *Voice History filters for Amazon Alexa*

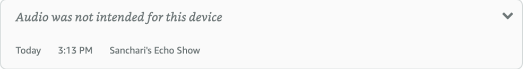
In special cases, we get two kind of responses from Amazon Alexa, other than correct or incorrect transcription. These are 'Audio could not be understood' (as shown in figure 7) and 'Audio was not intended for this device' (as shown in figure 8).



Audio could not be understood

Today 4:36 PM Sanchari's Echo Dot

Figure 7: *'Audio could not be understood' for Amazon Alexa*

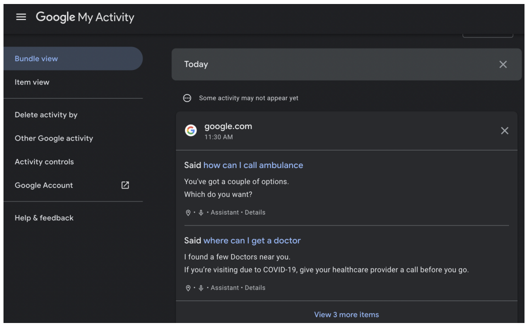


Audio was not intended for this device

Today 3:13 PM Sanchari's Echo Show

Figure 8: *'Audio was not intended for this device' for Amazon Alexa*

2) Google Assistant: To access the command history of an Google mini device, we navigate to the "Google My Activity" then choose "Bundle View" tab. We find our command history. This section looks like-



Today

Some activity may not appear yet

google.com
11:30 AM

Said how can I call ambulance

You've got a couple of options.
Which do you want?

• • Assistant • Details

Said where can I get a doctor

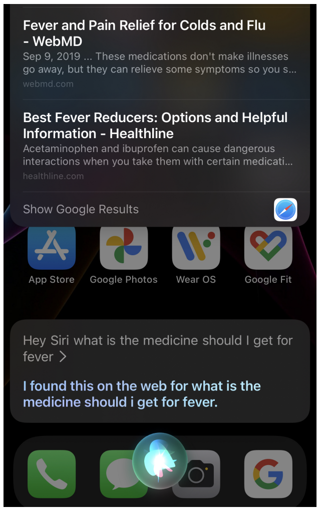
I found a few Doctors near you.
If you're visiting due to COVID-19, give your healthcare provider a call before you go.

• • Assistant • Details

View 3 more items

Figure 9: *Google Assistant Data Collection Screen*

3) Apple Siri: For Siri, to access the command history we used iPhone device. When siri listen something, it might not understand but if we turn on the setting "always show speech" (from Setting> SiriSearch> Siri Responses> Always Show Speech) is shows the transcript in screen. This section looks like-



Fever and Pain Relief for Colds and Flu - WebMD
Sep 9, 2019 ... These medications don't make illnesses go away, but they can relieve some symptoms so you s...
webmd.com

Best Fever Reducers: Options and Helpful Information - Healthline
Acetaminophen and ibuprofen can cause dangerous interactions when you take them with certain medicati...
healthline.com

Show Google Results

Hey Siri what is the medicine should I get for fever >

I found this on the web for what is the medicine should i get for fever.

Figure 10: *Siri Data Collection Screen*

IV. Evaluation:

A. Segregation, Analysis, and Tabulation

We played four commands in four different accents: American, Australian, Canadian, and British. Our other parameters being speed, volume, and noise tolerance. Our output numbers consist of three factors: 1) number of words in the original played command, 2) number of words in the transcript of the VA, and 3) number of correct words in the transcript of the VA in comparison to the words in the original played command. For each parameter, we take the averages of the output numbers over the four commands and the number of times played, which is the same for each case. We measure the correctness or, as we have called it, the accuracy percentage of each VA on any parameter as the percentage of the ratio of the number of words transcribed correctly to the number of words in the original input command played.

$$Accuracy\% = \frac{No.ofwordstranscribedcorrectly}{No.ofwordsininputcommand}X100\%$$

Our first analysis involves comparing the original, unaltered wave forms classified by various accents. We calculate the averages and the accuracy percentage and get the table I.

Accents (Original wave form)	Amazon Alexa	Google Assistant	Apple Siri
American	100%	100%	100%
Australian	86%	100%	100%
Canadian	62%	100%	67.5%
British	72%	100%	87.5%

Table I: *Accuracy Percentage based on the accents of the original wave form*

From this table I, we see that all of the VAs perform exceptionally when exposed to American accent. Both Google Assistant and Apple Siri perform great for Australian accent, but Amazon Alexa has a lower accuracy at 86%. In the case of Canadian accent, both Amazon Alexa and Apple Siri have a accuracy percentage between 60 and 70 %, Google Assistant performs great here as well with a 100% accuracy. For British accent, Alexa has an accuracy % of 72, Siri of 87.5 and Assistant still at 100.

Our second analysis involves comparing the altered wave forms classified according to their parameters for the three Virtual Assistants. Table II tabulates our findings for the VA Amazon Alexa.

We observe that Alexa does pretty well under perfect, loud, and somewhat noisy conditions. However, in slow, low (as 0.1 as amplitude factor) or more than 5% added noise conditions, the performance of Alexa is considerably reduced. We also notice that for low volumes, Alexa often returns 'Audio was not intended for this device', and for very high speeds, Alexa often returns 'Audio could not be understood'.

Table III tabulates our findings for the VA Apple Siri.

We observe that Siri does pretty well in most of the cases except when the volume is very low (as in 0.1 as amplitude factor) and when the noise is pretty high (greater than 5%). In these few cases, the performance of Siri is greatly decreased.

Table IV tabulates our findings for the VA Google Assistant.

Alexa	Words Transcribed Total	Words Transcribed Correctly
Perfect	82%	60%
Fast 1.5	54%	50%
Fast 2	23%	8%
Slow 0.5	18%	12%
Slow 0.75	56%	48%
Loud 1.5	82%	60%
Loud 2	82%	60%
Low 0.1	12%	8%
Low 0.5	70%	63%
Low 0.9	82%	60%
Noisy 0.01	74%	58%
Noisy 0.03	56%	36%
Noisy 0.05	32%	18%
Noisy 0.07	14%	5%
Noisy 0.09	6%	2%

Table II: *Accuracy Percentage based on the modification parameters on the original wave form for Amazon Alexa*

Siri	Words Transcribed Total	Words Transcribed Correctly
Perfect	87.5%	63.5%
Fast 1.5	87.5%	63.5%
Fast 2	60%	55%
Slow 0.5	82.5%	77.5%
Slow 0.75	87.5%	81%
Loud 1.5	75%	63.5%
Loud 2	87.5%	75%
Low 0.1	20%	16%
Low 0.5	73.5%	55.5%
Low 0.9	100%	87.5%
Noisy 0.01	87.5%	75%
Noisy 0.03	60%	75%
Noisy 0.05	50%	42.5%
Noisy 0.07	33%	20%
Noisy 0.09	16%	5%

Table III: *Accuracy Percentage based on the modification parameters on the original wave form for Apple Siri*

Assistant	Words Transcribed Total	Words Transcribed Correctly
Perfect	100%	100%
Fast 1.5	55.5%	33.5%
Fast 2	20.5%	10.5%
Slow 0.5	100%	87.5%
Slow 0.75	100%	87.5%
Loud 1.5	100%	100%
Loud 2	100%	100%
Low 0.1	50%	12%
Low 0.5	100%	87.5%
Low 0.9	100%	100%
Noisy 0.01	100%	87.5%
Noisy 0.03	62%	38.5%
Noisy 0.05	50%	28%
Noisy 0.07	50%	28%
Noisy 0.09	25%	5%

Table IV: *Accuracy Percentage based on the modification parameters on the original wave form for Google Assistant*

We observe that Google Assistant does pretty well in most of the cases except when faster speech (almost 2 times the speed of original) and extreme noisy conditions (more than 7% added noise). In these few cases, the performance of Google Assistant is low but overall it has a very good performance.

We next plot the above performances in a graphical form for us to be able to draw concrete conclusions based on this surfeit of data.

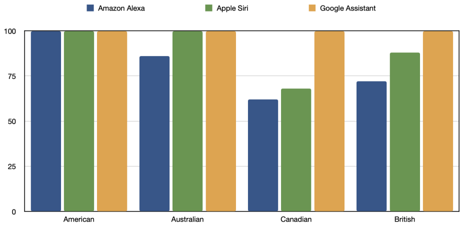


Figure 11: *Graph plotted percentage of words properly transcribed, by accent*

Figure 11 depicts the performances of the three Virtual Assistants - Amazon Alexa, Apple Siri, and Google Assistant - against the four accent parameters - American, Australian, Canadian, and British accents for the original audio wave form. We see hence that for American accents, all the three VAs perform accurately. In the case of Australian accent, the performance of Amazon Alexa is reduced on average to 86% whereas the other two VAs still have an accurate performance. For Canadian accent, Google Assistant continues to give an accurate performance, whereas the performance of Apple Siri and Amazon Alexa are reduces to 68% and 62% respectively. Finally, we have the British accent, for which Google Assistant continues to have an accurate performance, but the performances of Apple Siri and Amazon Alexa are reduced to 88% and 72% respectively. From this analysis, for these four accents considered in our analysis, we can conclude that Google Assistant provides the best performance among the three VAs.

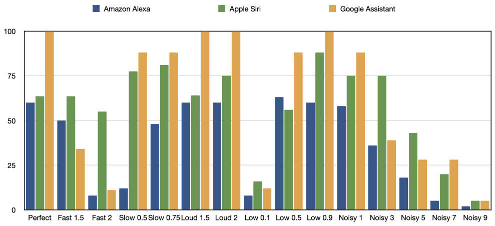


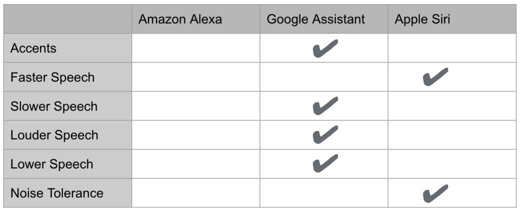
Figure 12: *Graph plotted percentage of words properly transcribed, by other external factors*

Figure 12 depicts the performances of the three Virtual Assistants - Amazon Alexa, Apple Siri, and Google Assistant - against the three external parameters - speed, volume, and noise - used on modification to the original wave forms. The results are averaged over the four commands, the four accents, and the total number of times played to eliminate the presence of biasing factors, if any.

From the graph, we can see that for faster speech and noisy inputs, Apple Siri performs better compared to Google Assistant and Amazon Alexa. For the other external factors, slower speech, louder volumes, and lower volumes, Google Assistant has a better accuracy than Apple Siri and Amazon Alexa.

B. Findings

We analyze and tabulate the vast amount of data that we have gathered from the transcripts of the Virtual Assistants, based on four English commands in four different accents - American, Australian, Canadian, and British, and based on three external parameters - speed of speech, volume, and noise introduced. This provides us with a meaningful and easy to understand set of information to work with. Thus, at this point, we are able to draw some conclusions regarding the performance and accuracy of the three Virtual Assistants in the context of our examination.



	Amazon Alexa	Google Assistant	Apple Siri
Accents		✓	
Faster Speech			✓
Slower Speech		✓	
Louder Speech		✓	
Lower Speech		✓	
Noise Tolerance			✓

Figure 13: *Table depicting the conclusive accuracy of the three Virtual Assistants*

The figure 13 has been constructed from the data obtained in the previous section. The tabulation of this data, as we can observe from figures 11 and 12, lets us undergo a comparative study of the three Virtual Assistants. For example, if we take a look at our figure 11 with the accuracy of the VAs according to the accent of the audios fed, we see that Google Assistant has a hundred percent accuracy for all the four accents. This is why we check the Google Assistant column for the Accents row in table 13. For the other rows and columns, we take a look at the figure 12. From here, we can see that Apple Siri performs better than the other two VAs for the factors of Faster Speech and Noise Tolerance, Hence, we check the Apple Siri column corresponding to those two rows in our table 13. For all the other factors, Google Assistant has a comparatively better performance than Amazon Alexa and Apple Siri. So, we check the rows of those parameters and the column for Google Assistant in our table 13. This provides us with the finding that based on our external factors, out of these three Virtual Assistants, Google Assistant has better performance in more of the instances and hence has a better overall comparative accuracy for our analysis.

V. Conclusion:

Our approach, while being a good indicator of the accuracy of the Virtual Assistants in terms of the external factors - accent, speed of speech, volume, and noise tolerance, have some drawbacks in respect to a realistic emergency response scenario. Most of these drawbacks evolve from the fact that human beings are not machines, and their response to any situation is unpredictable. So, firstly, the commands we play out to the VAs, we have no way of knowing for sure if an injured individual will always be in a state of mind to form a complete and accurate VA acceptable command. If he/she is unable to do such, most of the VAs might ask again for a command, which is not an ideal thing to do in such a situation. Secondly, we understand that emergency situations of various kinds cause different mental states in an individual which result in a faster or slower speech [11]. This faster or slower speech does not map exactly to the kind of speed up or down we are performing in our experiment. This is due to the fact that such a realistic speech is not uniformly sped up or down. Most often, it is the pauses that are sped up, whereas the time frame of pronunciation of each word stays the same. We have not, hence, recreated this scenario to the utmost accuracy in our experiment. Thirdly, we have only covered four of the accents, when there are an immense number of other accents widely spread. We, also, have not considered other prevalent languages like Spanish, Mandarin, or Hindi. Finally, our set of factors is limited due to the reason that we use an audio archive with various accents and a MATLAB Tool that we created to generate our modified factors. On the other hand, there can be a lot of other factors, eg. age, dialect, proficiency with Virtual Assistants, environment, that can give a more accurate analysis of Virtual Assistants in an emergency response situation.

Emergency response systems, being the means for emergency response teams to locate and move resources to emergency sites, have a very fine threshold for error when it comes to accuracy. Concluding only based on our analysis and associated factors, we can say , gauging by how our three Virtual Assistants perform, based on accents, speed of speech, low volumes, and noisy environments, they are not yet fit to be implemented in these scenarios and are in need of case-specific modification to their transcribing algorithms before this step can be realistically taken.

References

[1] H. Chung and S. Lee, "Intelligent virtual assistant knows your life," *arXiv preprint arXiv:1803.00466*, 2018.

[2] C. Picard, K. E. Smith, K. Picard, and M. J. Douma, "Can alexa, cortana, google assistant and siri save your life? a mixed-methods analysis of virtual digital assistants and their responses to first aid and basic life support queries," *BMJ Innovations*, 2020, ISSN: 2055-8074. DOI: 10.1136/bmjinnov-2018-000326. eprint: https://innovations.bmj.com/content/early/2020/01/07/bmjinnov-2018-000326.full.pdf. [Online]. Available: https://innovations.bmj.com/content/early/2020/01/07/bmjinnov-2018-000326.

[3] C. Rzepka, "Examining the use of voice assistants: A value-focused thinking approach," 2019.

[4] K. Collins and C. Metz, "Alexa vs. siri vs. google: Which can carry on a conversation best?" *The New York Times*, 2018.

[5] M. B. Hoy, "Alexa, siri, cortana, and more: An introduction to voice assistants," *Medical reference services quarterly*, vol. 37, no. 1, pp. 81–88, 2018.

[6] A. Palanica, A. Thommandram, A. Lee, M. Li, and Y. Fossat, "Do you understand the words that are comin outta my mouth? voice assistant comprehension of medication names," *NPJ digital medicine*, vol. 2, no. 1, pp. 1–6, 2019.

[7] A. Palanica and Y. Fossat, "Medication name comprehension of intelligent virtual assistants: A comparison of amazon alexa, google assistant, and apple siri between 2019 and 2021," *Frontiers in Digital Health*, vol. 3, p. 48, 2021.

[8] G. Hong, A. Folcarelli, J. Less, C. Wang, N. Erbasi, and S. Lin, "Voice assistants and cancer screening: A comparison of alexa, siri, google assistant, and cortana," *The Annals of Family Medicine*, vol. 19, no. 5, pp. 447–449, 2021.

[9] M. Braun, A. Mainz, R. Chadowitz, B. Pfleging, and F. Alt, "At your service: Designing voice assistant personalities to improve automotive user interfaces," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–11.

[10] J. D. Van Der Laan, A. Heino, and D. De Waard, "A simple procedure for the assessment of acceptance of advanced transport telematics," *Transportation Research Part C: Emerging Technologies*, vol. 5, no. 1, pp. 1–10, 1997.

[11] P. Howell, "Effect of speaking environment on speech production and perception," eng, *Journal of the human-environment system*, vol. 11, no. 1, pp. 51–57, Nov. 2008, PMC3024543[pmcid], ISSN: 1345-1324. DOI: 10.1618/jhes.11.51. [Online]. Available: https://doi.org/10.1618/jhes.11.51.

[12] *Citing - tool for generating a website's BibTex using the URL? - TeX - LaTeX - stack exchange*, https://www.alt-usage-english.org/audio$_{a}rchive.html$. [Online]. Available: https://www.alt-usage-english.org/audio_archive.html.

[13] MATLAB, *version 9.11.0 (R2021b)*. Natick, Massachusetts: The MathWorks Inc., 2021.

[14] F. Angione, C. Novak, C. Imeson, A. Lehman, B. Merwin, T. Pagliarella, N. Samardzic, P. D'Angela, and H. Ule, "Study of a low frequency emergency siren in comparison to traditional siren technology," in *Proceedings of Meetings on Acoustics 172ASA*, Acoustical Society of America, vol. 29, 2016, p. 030 008.