Voice Controlled Radio Mobile Application with Speech Recognition Using Alan Studio

Amaan Mithani, Dr. Pooja Bagane, Karan Samant, Jashn Anand, Gaurav Gandhi

Department of Computer Science, Symbiosis Institute of Technology, Affiliated to Symbiosis International University, Pune, poojabagane@gmail.com

Abstract

In this paper, a very captivating proposal is offered, a mobile-based service that is a union of the ground-breaking Alan Studio, Flutter & Dart. The projected idea of the Voice Controlled Mobile Application offers a naive method and comfort to the user. This application delivers all the mechanisms mandatory for a user to be able to use speech & voice as an intermediate to shuffle between radio stations of his/her choice and also the option to have a conversation with the application. A voice-controlled structure implanted in a mobile application. Music has many health benefits – improves mood, reduces stress, improves memory, improves cognition and many more. By the participation and establishment of technology, it is detected novel behaviours of listening to music are substituting longstanding methods. Thus, new ways of listening to music, such as radio are in demand now and this application gives the user the experience of hands-free radio, along with voice commands such as a conversation, weather and news too. Alan AI's responses are compared to Siri and Google Assistant and they showed staggering results.

Keywords. Alan Studio, Flutter; Radio, Voice Controlled Mobile App.

1. INTRODUCTION

Advancements in technology are being developed to make life easier, and voice control is one of the main machineries that is slowly used on more and more gadgets. One of the most in-demand skills is voice control. Voice control is becoming more common in apps and services. Voice control capabilities give the customer a very simple and hands-free experience, allowing them to access the facility without using the device by hand. Current routines do not let us remove time out of our hectic schedules to type/select every song, artist, or genre that we desire. We want things done quickly and without having to use our hands. Music has many health benefits – improves mood, reduces stress, improves memory, improves cognition and many more. By the participation and establishment of technology, it is detected novel behaviours of listening to music are substituting longstanding methods. There are no linguistic borders when it comes to a good melody, thus music is popular all around the world. When people are exercising, driving, doing domestic chores, or doing anything else that requires their hands to be active, they prefer to listen to music to make their activity more enjoyable. We created Bharat Play, an AI radio with an in-app voice assistant named Alan that makes listening to music much easier. With only one voice command, you can listen to your favourite radio station. The system's hands-free method goes a long way toward encouraging the user to cooperate frequently, as the user wishes to

practice speech facility than inputting instructions. One of the most significant rewards of the suggested application is that speech recognition is not restricted to mobiles, laptops, or desktops, but is also implemented in all types of devices with which handlers cooperate, such as smart tv's and smart watches.

2. LITERATURE SURVEY

As speech communications are much prevalent, a rising amount of research has aided to clarify many questions regarding whose voice is maximum effective, as well as the key obstacles and confines of voice assistants. Numerous researches have looked at popular assistants like Google Assistant Alexa, Siri have become interwoven into users' daily lives, as well as how their associations with these assistants have evolved as time passes. Voice interactions are particularly beneficial in situations where the user's visual and ligaments are busy, for example while driving/cooking/working out. Users regularly request their voice assistants to 'answer search or type queries, control music, set timers, and manage associated smart home devices', according to usage records and self-report statistics [2]. Fig 1. Shows how Google Home and Amazon Alexa compare in terms of command criteria. In both the charts, music has a very high ratio, 26.1% and 28.5% respectively. This shows that across all voice assistants regardless of the brand, people do trust their voice assistant to play songs.

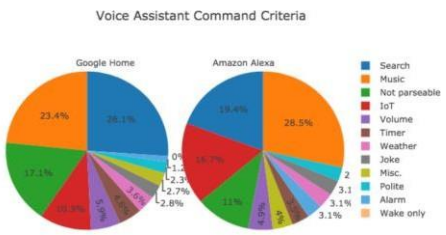


Fig 1. Voice assistant criteria Google Home vs Alexa

Similarly, project principles and finest practises for designing for speech are yet in their early phases, and voice assistants face many unanswered technical challenges. Voice interfaces, especially those deprived of a screen, frequently suffer from a lack: unlike graphical user interfaces, where the potentials for communication are noticeable to the user. Furthermore, speech recognition mistakes plague voice assistants, leading in erroneous transcriptions of what the user said. Liable on the customer's pronunciation or usage of specialised language and proper nouns, theseerrors can be particularly common.

3. METHODOLOGY

3.1 Alan Studio

3.2 Unique terminology and workflow

The Alan Platform enables application developers to define dialog scripts for their in-app voice assistant. These dialog scripts enable the Alan Platform to build a Domain Language Model for the application and later use the non-verbal context to understand user language very accurately.[1]

ii. Voice conversations are dynamic

Today, most conversational solutions require you to design a static, template-driven dialog flow. This type of design expects users to stick to the step-by-step structure, which can result unnatural, frustrating, and unsuccessful interactions. Most such conversations are rigid and robotic, and far from the dynamic flow of human interactions. Fig 8 gives a representation of these types of voice conversations.

Alan provides a means to load entities dynamically and customize them during the dialog session. The conversational flow, therefore, can be updated during runtime, on the fly, without the need to rebuild the AI conversational model. All this is feasible because the Alan Platform uses a hybrid model for dialog management. The hybrid approach builds voice dialogs through a combination of machine learning models tuned on your application data and procedural programming in JavaScript that has access to the Dynamic Intent and Entities. shows the different approaches taken for dialog management.

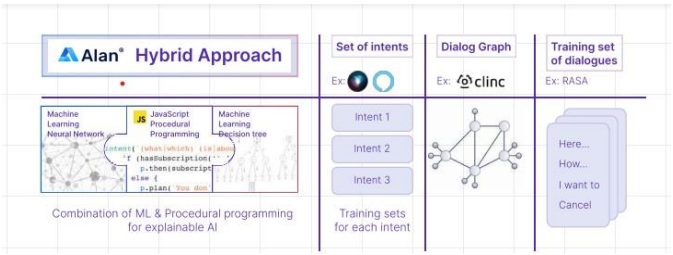


Fig 3 Hybrid Approach for Dialog Management to enable human-like conversations.

Source : Alan Whitepaper

iii. Quickly iterate

When the in-app voice assistant goes live, Alan collects conversational data and provides you with advanced voice analytics tools to gain real-time insights into users' interactions in each context. This will help you better understand your users' needs and deliver a more accurate experience. With all the key metrics and conversational data at their disposal, app developers will augment the voice conversation scripts to improve the user experience and expand the cohort that is exposed to the voice assistant.

3.2 Flutter and Dart

Flutter is a cross-platform UI toolkit that permits for code reclaim across operating systems for example iOS and Android while also consenting applications to interact straight with fundamental platform services [5]. The goal is for developers to be able to produce high-recital programmes that feel normal across platforms, embracing changes where they are while allotting as much code as feasible.

We have used Flutter and Dart to create the front end- the audio player and integrate the Alan Package from Alan Studio. This was all coded in Android Studio.

4. DESIGN IMPLEMENTATION

The aforementioned technologies come together to provide the user with a flawless software that allows him to listen to his favourite genre of music while browsing through various radio stations. The app is quite accurate in its operation, providing the user with songs coming from the maximum dependable and reliable radio stations.

Fig 4 shows the flowchart of the application when the user says the 'Hey Alan!' wake word.

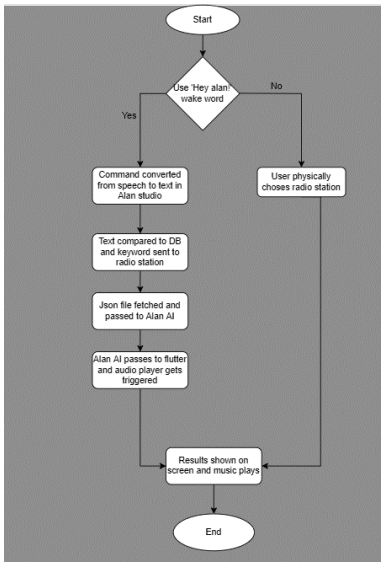


Fig 4 Flowchart of the App

5. EXPERIMENTAL RESULTS

Here we can see how Alan compares to Siri and Google Assistant. The first parameter we tested was the response time shown in Table 1. Here we can see we have tested 8 instances between the three voice assistants and by no surprise Siri has topped the ratings with the lowest average response time, Google Assistant coming second and Alan Voice coming last. But a thing to note is that in the system we have developed, music is the main motive and we can see that in Siri and Google Assistant the 'Stop Music' function does not work but in Alan voice it does. Moreover, there is no way for Siri and Google Assistant to play a particular radio station but Alan can, as we see in the Accuracy Table Shown in Table 2.

Figure 5 shows Table 1's response time comparison in the form of a bar chart visualised by colour coding the different assistants.

In playing a specific genre we can clearly see that Alan took far less time than Siri, which is a great insight to the working of our system and the wake word 'Hey Alan' performed really well when compared to the popular voice assistants.

	Response time (ms)		
Parameters	Alan	Siri	Google Assistant
Wake word	2160	1720	**1092**
Play a song	6530	**6020**	6150
Stop music	**3000**	--	--
Play next	4450	**2650**	2920
Conversation(name)	3420	3210	**1530**
calculator	4160	3530	**2960**
Play genre	6860	9000	**6330**
Play radio city FM	**6700**	--	--

Table 1. Response time comparisons

	Accuracy (1 or 0)		
Parameters	Alan	Siri	Google Assistant
Wake word	1	1	1
Play a song	1	1	1
Stop music	1	0	0
Play next	1	1	1
conversation	1	1	1
calculator	1	1	1
Play genre	1	1	1
Play radio city FM	1	0	0
Total	**8/8**	6/8	6/8

Table 2. Accuracy Score Comparison

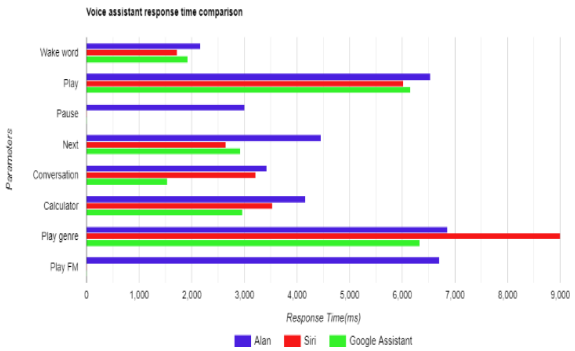


Fig 5 Voice assistants' response time comparison

We can see that Siri and Google both had 1 false positive and 1 true negative each, False positive because it played a random song when asked to play a paritcular FM and True negative because it did not even respond to the 'Stop' function when asked to pause the song.

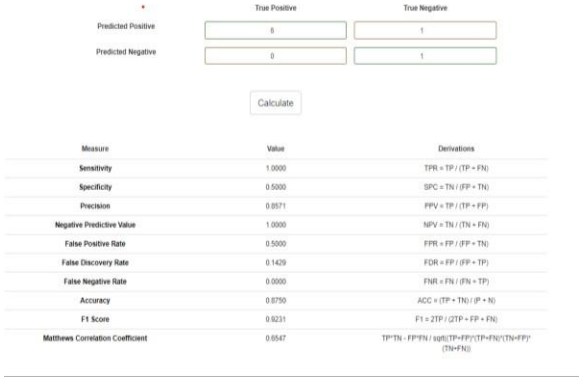
However Alan didn't have any faults and succesfully executed with and accuracy score of 1.00 in these tests.

	True Positive	True Negative
Predicted Positive	8	0
Predicted Negative	0	0

Calculate

Measure	Value	Derivations
Sensitivity	1.0000	TPR = TP / (TP + FN)
Specificity		SPC = TN / (FP + TN)
Precision	1.0000	PPV = TP / (TP + FP)
Negative Predictive Value		NPV = TN / (TN + FN)
False Positive Rate		FPR = FP / (FP + TN)
False Discovery Rate	0.0000	FDR = FP / (FP + TP)
False Negative Rate	0.0000	FNR = FN / (FN + TP)
Accuracy	1.0000	ACC = (TP + TN) / (P + N)
F1 Score	1.0000	F1 = 2TP / (2TP + FP + FN)
Matthews Correlation Coefficient		TP*TN - FP*FN / sqrt((TP+FP)*(TP+FN)*(TN+FP)*(TN+FN))

Fig 6 Confusion Matrix for Alan Voice



	True Positive	True Negative
Predicted Positive	8	1
Predicted Negative	0	1

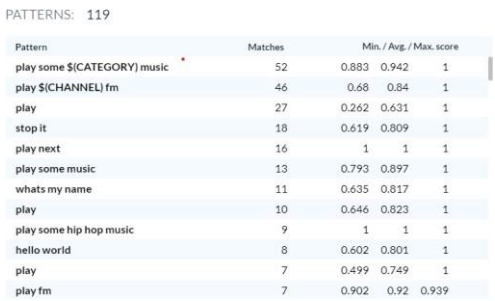
Calculate

Measure	Value	Derivations
Sensitivity	1.0000	TPR = TP / (TP + FN)
Specificity	0.5000	SPC = TN / (FP + TN)
Precision	0.8571	PPV = TP / (TP + FP)
Negative Predictive Value	1.0000	NPV = TN / (TN + FN)
False Positive Rate	0.5000	FPR = FP / (FP + TN)
False Discovery Rate	0.1429	FDR = FP / (FP + TP)
False Negative Rate	0.0000	FNR = FN / (FN + TP)
Accuracy	0.8750	ACC = (TP + TN) / (P + N)
F1 Score	0.9231	F1 = 2TP / (2TP + FP + FN)
Matthews Correlation Coefficient	0.6547	TP*TN - FP*FN / sqrt((TP+FP)*(TP+FN)*(TN+FP)*(TN+FN))

Fig 7 Confusion Matrix for Google Assistant and Siri

Alan studio analytics

As a user interacts with a voice assistant in the app, Alan AI captures voice data and analyses key conversational metrics.



PATTERNS: 119

Pattern	Matches	Min. / Avg. / Max. score		
play some $(CATEGORY) music	52	0.883	0.942	1
play $(CHANNEL) fm	46	0.68	0.84	1
play	27	0.262	0.631	1
stop it	18	0.619	0.809	1
play next	16	1	1	1
play some music	13	0.793	0.897	1
whats my name	11	0.635	0.817	1
play	10	0.646	0.823	1
play some hip hop music	9	1	1	1
hello world	8	0.602	0.801	1
play	7	0.499	0.749	1
play fm	7	0.902	0.92	0.939

Fig 8 Patterns

6. FUTURE WORK

No one expected speech recognition to become such a heavy power for forthcoming technological breakthroughs when it first emerged in the early 2010s with the debut of what is now one of the furthermost well-known devices, Siri. 'One-sixth of the population in the United States now possesses a smart speaker, according to estimates. This is just a small sample of the technology's potential and the huge expanse of possibility it offers.'

Predictions for future: Personalized Experiences, Voice Push Notifications, Search Behavior Will Shift, Inbuilt Security Features for Users, Voice Cloning and SmartDisplays.

7. CONCLUSION & DISCUSSION

The suggested technology allows a diverse group of users to listen to their favourite radio station no matter where they are. Because music is such an important part of everyone's lives, it allows users to enjoy their time whether travelling, exercising, or even with their hands busy. It also enables physically challenged persons to benefit from the most recent technological advances while also allowing them to relax by listening to their favourite type of music.

8. REFERENCES

[1] Chaprana, Aaditya & Kumar, Ranjeet & Saini, Ajay & Kumar, Akash. (2021). Voice Controlled News Web Application with Speech Recognition using Alan Studio. International Journal of Computer Applications. 183. 31-37. 10.5120/ijca2021921297.

[2] Terzopoulos, George & Satratzemi, Maya. (2020). Voice Assistants and Smart Speakers in Everyday Life and in Education. Informatics in Education. 19. 473-490. 10.15388/infedu.2020.21.

[3] Tolmeijer, Suzanne & Zierau, Naim & Janson, Andreas & Wahdatehagh, Jalil & Leimeister, Jan Marco & Bernstein, Abraham. (2021). Female by Default? – Exploring the Effect of Voice Assistant Gender and Pitch on Trait and Trust Attribution. 1-7. 10.1145/3411763.3451623.

[4] Bolton T, Dargahi T, Belguith S, Al-Rakhami MS, Sodhro AH. On the Security and Privacy Challenges of Virtual Assistants. Sensors (Basel). 2021 Mar 26;21(7):2312. doi: 10.3390/s21072312. PMID: 33810212; PMCID: PMC8036736.

[5] Napoli, Marco. (2019). Introducing Flutter and Getting Started. 10.1002/9781119550860.ch1.

[6] Sen, Sonali & Chakrabarty, Shamik & Toshniwal, Raghav & Bhaumik, Ankita. (2015). Design of an Intelligent Voice Controlled Home Automation System. International Journal of Computer Applications. 121. 39-42. 10.5120/21619-4904.

[7] Tulshan, Amrita & Dhage, Sudhir. (2019). Survey on Virtual Assistant: Google Assistant, Siri, Cortana, Alexa: 4th International Symposium SIRS 2018, Bangalore, India, September 19–22, 2018, Revised Selected Papers. 10.1007/978-981-13-5758-9_17.

[8] Assefi, Mehdi & Liu, Guangchi & Wittie, Mike & Izurieta, Clemente. (2015). Experimental Evaluation of Apple Siri and Google Speech Recognition.

[9] Ammari, Tawfiq & Kaye, Jofish & Tsai, Janice & Bentley, Frank. (2019). Music, Search, and IoT: How People (Really) Use Voice Assistants. ACM Transactions on Computer-Human Interaction. 26. 1-28. 10.1145/3311956.

[10] Hoy, Matthew. (2018). Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants. Medical Reference Services Quarterly. 37. 81-88. 10.1080/02763869.2018.1404391.

[11] Julia Cambre, Alex C Williams, Afsaneh Razi, Ian Bicking, Abraham Wallin, Janice Tsai, Chinmay Kulkarni, and Jofish Kaye. 2021. Firefox Voice: An Open and Extensible Voice Assistant Built Upon the Web. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 250, 1–18.

[12] David F. McCall, Leslie M. Logue, Francis J. Zelina, Matthew V. Sendak, Julie R. Hinson, Ward L. Sanders, Steve Belinski, Brian E. Holtz. Voice controlled surgical suite. Google Patents US6591239B1, 1999-Present.

[13] Ian W. Freed, William Folwell Barton, Rohit Prasad. Preventing false wake word detections with a voicecontrolled device. Google Patents US9368105B1, 2014.

[14] Stephen S. Burns, Mickey W. Kowitz. Voice controlled wireless communication device system. Google Patents US7957975B2, 2016-Present.

[15] K. Christian, B. Kules, B. Shneiderman and A. Youssef, “A Comparison of Voice Controlled and Mouse Controlled Web Browsing,” Proceedings of the Fourth International ACM Conference on Assistive Technologies, no. ACM, pp. 72-79, 2000.

[16] Ammari, Tawfiq & Kaye, Jofish & Tsai, Janice & Bentley, Frank. (2019). Music, Search, and IoT: How People (Really) Use Voice Assistants. ACM Transactions on Computer-Human Interaction. 26. 1-28. 10.1145/3311956.

[17] Hoy, Matthew. (2018). Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants. Medical Reference Services Quarterly. 37. 81-88. 10.1080/02763869.2018.1404391.

[18] Julia Cambre, Alex C Williams, Afsaneh Razi, Ian Bicking, Abraham Wallin, Janice Tsai, Chinmay Kulkarni, and Jofish Kaye. 2021. Firefox Voice: An Open and Extensible Voice Assistant Built Upon the Web. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 250, 1–18.

[19] David F. McCall, Leslie M. Logue, Francis J. Zelina, Matthew V. Sendak, Julie R. Hinson, Ward L. Sanders, Steve Belinski, Brian E. Holtz. Voice controlled surgical suite. Google Patents US6591239B1, 1999-Present.

[20] Ian W. Freed, William Folwell Barton, Rohit Prasad. Preventing false wake word detections with a voicecontrolled device. Google Patents US9368105B1, 2014.

[21] Stephen S. Burns, Mickey W. Kowitz. Voice controlled wireless communication device system. Google Patents US7957975B2, 2016-Present.

[22] K. Christian, B. Kules, B. Shneiderman and A. Youssef, “A Comparison of Voice Controlled and Mouse Controlled Web Browsing,” Proceedings of the Fourth International ACM Conference on Assistive Technologies, no. ACM, pp. 72-79, 2000.

Biographies

.

Amaan Mithani. Currently a final year B.Tech in Computer Science student at Symbiosis Institute of Technology, (SIT) affiliated to Symbiosis International (Deemed University), Pune, India. His research is primarily in the field of machine leaning and deep learning.

Pooja Bagane, Research Scholar in Visvesaraya Technological University, Belgaum, Karnataka- India and also working as an Assistant Professor at Department of Computer Science, Symbiosis Institute of Technology, (SIT) affiliated to Symbiosis International (Deemed University), Pune, India. She has completed M.Tech. in CSE with Gold Medal. She was an academic topper throughout the years. She has published manuscripts in reputed journals. Her research, primarily in security, evolutionary algorithms, artificial intelligence & Machine Learning, and genetic algorithms

Karan Samant. Currently a final year B.Tech in Computer Science student at Symbiosis Institute of Technology, (SIT) affiliated to Symbiosis International (Deemed University), Pune, India. His research is primarily in the field of machine leaning and deep learning.

Gaurav Gandhi. Currently a final year B.Tech in Computer Science student at Symbiosis Institute of Technology, (SIT) affiliated to Symbiosis International (Deemed University), Pune, India. His research is primarily in the field of machine leaning and deep learning.

Jashn Anand. Currently a final year B.Tech in Computer Science student at Symbiosis Institute of Technology, (SIT) affiliated to Symbiosis International (Deemed University), Pune, India. His research is primarily in the field of machine leaning and deep learning