

Project Notes:

Project Title:

Name: Adel Benchemam

Note Well: There are NO SHORT-cuts to reading journal articles and taking notes from them. Comprehension is paramount. You will most likely need to read it several times, so set aside enough time in your schedule.

Knowledge Gaps:	2
Literature Search Parameters:	3
Tags:	3
Article #1 Notes: Wildlife conservation using drones and artificial intelligence in Africa	5
Article #2 Notes: Social Media Addiction	7
Article #3 Notes: AI in medical field	9
Article #4 Notes: Running the 400 meter	12
Article #5 Notes: Artificial Intelligence based voice assistant	14
Article #6 Notes: AI voice assistant tools	16
Article #7 Notes: ARIA the bot	19
Article #8 Notes: Humanizing voice Assistants	21
Article #9 Notes: Design and Development of Intelligent Voice Personal Assistant using Python	23
Article #10 Notes: Artificial Intelligence Based A Communicative Virtual Voice Assistant Using Python & Visual Code Technology	25
Article #11 Notes: AI-based Desktop Voice Assistant	27
Article #12 Notes: Development of GUI for Text-to-Speech Recognition using Natural Language Processing	30
Article #13 Notes: Noise reduction algorithm for robust speech recognition using MLP neural network	33
Article #14 Notes: Indonesian Automatic Speech Recognition system using CMUSphinx toolkit and limited dataset	35
Article #15 Notes: Training of automatic speech recognition system on noised speech	38
Article #16 Notes: Robust Speech Recognition via Large-Scale Weak Supervision	41
Article #17 Notes: Desktop based Smart Voice Assistant using Python Language Integrated with Arduino	44
Article #18 Notes: Voice-Based Virtual-Controlled Intelligent Personal Assistants	47
Article #19 Notes: CommonVoice: A Massively-Multilingual Speech Corpus	49

Article #20 Notes: Large scale deep neural network acoustic modeling for YouTube video transcription52

Patent #1 Notes: Implementations for voice assistant on devices 54

Patent #2 Notes: Voice commands for transitioning between device states 57

Knowledge Gaps:

This list provides a brief overview of the major knowledge gaps for this project, how they were resolved and where to find the information.

Knowledge Gap	Resolved By	Information is located	Date resolved
What tools can be used for voice recognition?	Asking my old coworker and asking Mass Academy seniors	Email	09/23/2024
How can the personality of a voice assistant affect how people engage with it?	Article from science direct that tested out this question	Humanizing voice assistant: The impact of voice assistant personality on consumers' attitudes and behaviors - ScienceDirect Article 8 notes	10/06/2024
What assistants have been made with python?	Article named Artificially developed intelligent system using Python	Article 9 notes	10/06/2024
How can I interact with web elements?	Looking into tools, Selenium will be used	The source code is located on GitHub, the documentation is on the official selenium website	11/05/2024

Literature Search Parameters:

These searches were performed between (Start Date of reading) and XX/XX/2024.

List of keywords and databases used during this project.

Database/search engine	Keywords	Summary of search
Google Scholar	Voice Recognition	Took notes on it in article 5
Google Scholar	Automatic Speech Recognition	Found (mostly IEEE) articles and took notes on assistants that use ASR. Since ASR isn't what I'm working on, I won't be making my own NLP, however, it's good to know which models have high accuracy.
Bing	Python GUI control	PyAutoGUI: PyAutoGUI · PyPI Welcome to PyAutoGUI's documentation! — PyAutoGUI documentation Can be used to control mouseclicks and keyboard control WebBrowser can be used to control opening of tabs and windows.
Bing	Selenium web driver	https://selenium-python.readthedocs.io/getting-started.html#simple-usage This can allow me to interact with specific elements of the screen. For example, I can say "click button submit" and the program will recognize

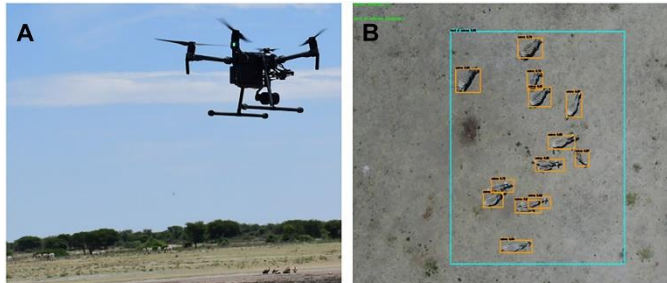
Tags:

Tag Name

#imageRecognition	#voiceRecognition
#running	#tools
#machinevision	#personality
#medical	#AutomaticSpeechRecognition
#prosthetics	# SpeechCorpus
#training	#MachineLearning
#acousticmodel	

Article #1 Notes: Wildlife conservation using drones and artificial intelligence in Africa

Article notes should be on separate sheets

Source Title	Wildlife conservation using drones and artificial intelligence in Africa
Source citation (APA Format)	Tinao, P., & Jamisola, R. S. (2023). Wildlife conservation using drones and artificial intelligence in Africa. <i>Science Robotics</i> , 8(85). https://doi.org/10.1126/scirobotics.adm7008
Original URL	https://www.science.org/doi/10.1126/scirobotics.adm7008
Source type	Journal publication
Keywords	Conservation, Artificial intelligence, drones
#Tags	#imageRecognition
Summary of key points + notes (include methodology)	Traditional ways of protecting wildlife include causing permanent scars, putting GPS trackers on the animals, or other methods that involve capturing animals. These threaten the wildlife's behavior and require too much invasive effort. Drone(backed with AI) surveillance provides a solution to this as a way to monitor these herds without harming them and knowing when to make decisions to help increase wildlife population.
Research Question/Problem/Need	How does Artificial Intelligence protect wildlife?
Important Figures	Fig.1 
VOCAB: (w/definition)	Deep learning- a type of machine learning based on artificial neural networks in which multiple layers of processing are used to extract progressively higher level features from data.(Oxford language)

	Neural Networks - a computer system modeled on the human brain and nervous system.(Oxford language)
Cited references to follow up on	Chalmers, C., Fergus P., Wich, S., & Montanez, A. C. (2019, October 16). <i>Conservation ai: Live stream analysis for the detection of endangered species using convolutional neural networks and drone technology</i> . arXiv. https://arxiv.org/abs/1910.07360
Follow up Questions	<ol style="list-style-type: none"> 1. In what ways can we apply the same concept of tracking on animals that can evade these drones such as those who live underground or fly? 2. What threats can we use AI to detect in order to protect these wildlife animals? 3. Would there be ways that allow humans to help these AI models learn and keep track of the animals more accurately?(like photography from the ground)

Article #2 Notes: Social Media Addiction

Article notes should be on separate sheets

Source Title	Are you addicted to social media?
Source citation (APA Format)	Lee Health. (2021, December 6). <i>Are you addicted to social media?</i> Lee Health. https://www.leehealth.org/health-and-wellness/healthy-news-blog/mental-health/are-you-addicted-to-social-media
Original URL	https://www.leehealth.org/health-and-wellness/healthy-news-blog/mental-health/are-you-addicted-to-social-media
Source type	Website
Keywords	Dopamine, addiction
#Tags	#dopamine
Summary of key points + notes (include methodology)	This article relates to an issue that pretty much everyone faces: social media dependency. It states that social media applications are highly stimulating and make users release a neurotransmitter called dopamine. Dopamine is a chemical messenger between neurons that makes someone feel good. It is released when eating food or shopping and is used by your brain as a reward when interacting with a social media platform. This article points out that most people have a habit of using social media. Although this may seem harmful, studies have shown that the use of social media shows dependance such as alcohol and drugs. Some even show symptoms similar to those of substance use. Fortunately, this article offers ways to get over the bad habit of doom scrolling. It encourages users to social media detox, limit the hours on their phone, and use a real alarm. The real issue happens when people use social media as a way to cope with stress. It can be a distraction from work, school, and physical health. Social media detoxing is the solution, as it focuses on conversation and reduces dependency on social media.
Research Question/Problem/Need	Why is it easy to become addicted to social media?
Important Figures	none
VOCAB: (w/definition)	Dopamine - sends chemical messages in your brain to let you know that something feels good(source)
Cited references to follow up on	None
Follow up Questions	<ol style="list-style-type: none"> 1. How does detoxing rewire you to be less likely to use social media? 2. What do social media companies do in order to keep the audience captivated?

- | | |
|--|--|
| | <p>3. What other methods, other than detoxing, can help escape the dopamine cycle?</p> |
|--|--|

Article #3 Notes: AI in medical field

Article notes should be on separate sheets

Source Title	Artificial intelligence meets medical robotics
Source citation (APA Format)	Yip, M., Salcudean, S. E., Goldberg, K., Althoefer, K., Menciassi, A., Opfermann, J. D., Krieger, A., Swaminathan, K., Walsh, C. J., Huang, H. H., & I-Chieh, I. (2023). Artificial intelligence meets medical robotics. <i>Science</i> , 381(6654), 141–146. https://doi.org/10.1126/science.adj3312
Original URL	https://www.science.org/doi/10.1126/science.adj3312
Source type	Journal publication
Keywords	Prosthetics, surgery, autonomy, semi-autonomy
#Tags	#machinevision #medical #prosthetics
Summary of key points + notes (include methodology)	<p>This science publication relates to one of my favorite classes and something I want to help work on in the future. The science publication shows how technology is currently changing lives in the medical field. Some machines allow doctors to perform surgeries remotely. Currently, robots are still working with human direction as we do not have the technology to make them fully autonomous. Currently, the highest level of autonomy these robots have is being able to create strategies which will later on be selected by humans for which is more effective. Technology is also being used in rehabilitation centers in order to be able to help people move certain parts of their bodies that may have been lost through accidents. Artificial intelligence has learned to use computer vision in order to be able to identify problems in the body. Artificial intelligence has also enabled the improvement of exoskeletons which can now facilitate rehabilitation and collect data for training. Computer vision can also predict your intention through your muscle movements and move a prosthetic accordingly, which will aid people with lost limbs. AI's introduction to the medical field can act as a supplement to the surgeons as they can accomplish tedious subtasks. One of these subtasks can be debridement which consists of removing foreign fragments from the body. This kind of task is prone to human error, but with the help of artificial intelligence we're able to spot these mistakes and remove them.</p>
Research Question/Problem/Need	How can Artificial Intelligence improve the medical field?

Important Figures

Smart medical robots

Various applications of artificial intelligence (AI), including machine learning, machine vision, and haptic control, have resulted in the development of robotic devices that can be used in all aspects of patient care, including diagnostics, surgical procedures, rehabilitation, and limb replacement. The use of robotics in medicine aims to ensure consistent, safe, and efficient treatment, as well as allowing data gathering for improvement and potentially increasing access to treatment in underserved communities and remote regions, and to those affected by natural disasters.

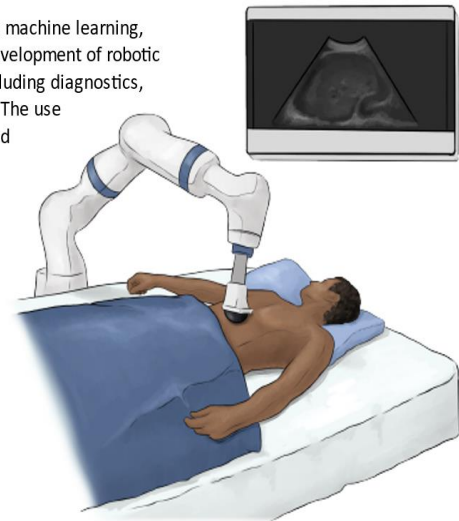
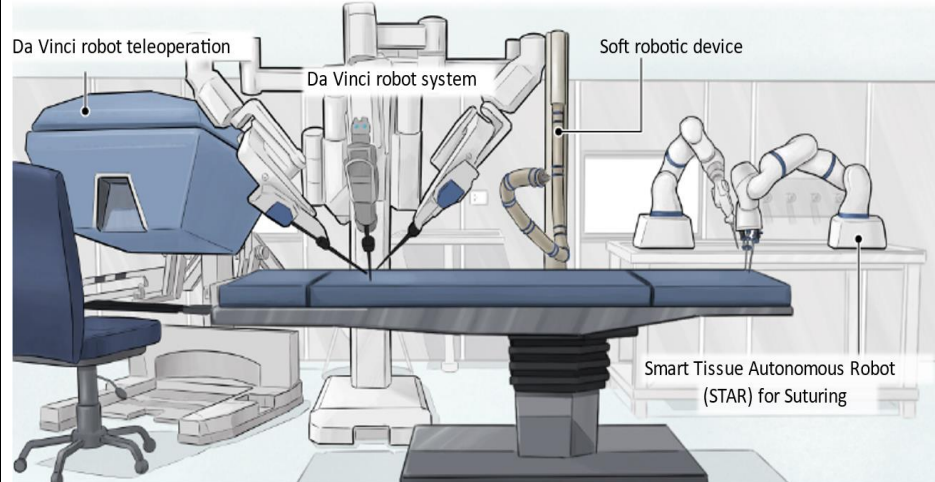


Image-guided robotics can aid diagnostics and delivery of interventions through analyzing medical images. Moreover, autonomous robots are being developed to carry out imaging, e.g., ultrasound imaging to access standard imaging planes consistently. Image guidance is also important for autonomous surgical planning and procedures such as endoscopy.

Surgical robots

Various levels of autonomy can be used in robots that carry out surgical tasks.



Surgical robots can take various forms, including teleoperated devices that allow surgeons to carry out complex procedures without fear of tissue damage from hand tremor. Soft robotic devices are under development for minimally invasive surgical procedures, providing haptic feedback to surgeons as well as ensuring safe manipulation of and navigation through soft tissues. Semi-autonomous robots that can undertake surgical subtasks, such as suturing and debridement, are also under development. This could potentially lead to fully autonomous surgical robots.

Rehabilitation devices and advanced prosthetics

Wearable devices can improve recovery from injury and facilitate human-in-the-loop intervention.



Soft exosuits

Wearable robots in the form of soft exosuits can autonomously

Upper limb prosthetics

Recognizing user intent is important for upper limb prosthetics to

Lower limb prosthetics

Replacement of lower limbs requires prosthetics to adapt

VOCAB: (w/definition)	<p>Debridement - the removal of damaged tissues or foreign fragments from a wound.</p> <p>Machine Vision - the ability of a computer to see, in this case it can be used to sense when a user of a prosthetic intends to grasp something.</p>
Cited references to follow up on	<p>Burgner-Kahrs, J., Rucker, D. C., & H. Choset. (2015) Continuum Robots for Medical Applications: A Survey. <i>IEEE Transactions on Robotics</i>, 31(6), 1261-1280, https://doi.org/10.1109/TRO.2015.2489500.</p> <p>Fichtinger, G., Troccaz, J., Haidegger, T. (2022). Image-Guided Interventional Robotics: Lost in Translation?. <i>Proceedings of the IEEE</i>, 110 (7), 932-950. https://doi.org/10.1109/JPROC.2022.3166253.hal-03654928</p> <p>Schreiber, D., Yu, Z., Jiang, H., Henderson, T., Li, G., & Yu, J. (2022, May 23-27) CRANE: a 10 Degree-of-Freedom, Tele-surgical System for Dexterous Manipulation within Imaging Bores. <i>2022 International Conference on Robotics and Automation (ICRA)</i>, Philadelphia, PA, USA, pp. 5487-5494. IEEE. https://doi.org/10.1109/ICRA46639.2022.9811732.</p>
Follow up Questions	<ol style="list-style-type: none"> 1. How can AI be further implemented in order to help people medically? 2. Which tedious tasks can AI take up in order to help doctors, not just surgeons? 3. What steps does AI still have to go through in order to be fully autonomous?

Article #4 Notes: Running the 400 meter

Article notes should be on separate sheets

Source Title	Chemistry and Sport - Athletics: The 400m Event
Source citation (APA Format)	Royal Society of Chemistry. (n.d.). <i>Chemistry and Sport - Athletics: 400m</i> . Royal Society of Chemistry Education. https://edu.rsc.org/resources/chemistry-and-sport-athletics-400m/857.article
Original URL	https://edu.rsc.org/resources/chemistry-and-sport-athletics-400m/857.article
Source type	Online source
Keywords	Energy, aerobic, anaerobic
#Tags	#running
Summary of key points + notes (include methodology)	<p>This article pertains to my hobby of running. I run cross country and outdoor track, and I am interested to know how the body reacts to different kinds of exercises. I want to see the different effects of running at different distances and their intensities. My favorite event, also known as the most straining one, is the 400-meter dash. It was first introduced in ancient Olympic games and reintroduced in modern Olympic games in 1896. The article by RSC Education explains the energy sources that go into the 400-meter dash. Due to the event being too short to pace and recover yet too long to run at top speed, the race pushes runners to their limits. The article describes three energy sources that runners utilize during the 400 meters. The first is ATP-PC, which quickly creates short-term ATP, only lasting about 10 seconds. After this short span of time, anaerobic glycolysis takes place. It is slower at providing ATP, which in turn builds up lactic acid. Essentially, lactic acid is a product made when the body is generating energy in order to provide for the muscle when there is an insufficient amount of oxygen. I can say from experience that this makes the last stretch the most difficult, as the buildup makes the body and limbs feel heavy. Lastly, the body utilizes aerobic glycolysis: using oxygen to create ATP. This method is the most efficient and happens at the end of the race.</p> <p>I mainly want to focus on the anaerobic glycolysis phase, as it is the one that makes the 400 meters so difficult. It makes seconds feel extra slow as your limbs begin to fail on you. The 400 meters are grueling and make the last stretch feel like a battle between yourself as you force yourself to finish and push through the pain.</p>
Research Question/Problem/Need	How does the body use its energy sources while an athlete is running the 400 meter?
Important Figures	None
VOCAB: (w/definition)	ATP-PC - energy system provides immediate energy through the breakdown of

	<p>these stored high energy phosphates.</p> <p>Anaerobic glycolysis - produces more ATP per molecule of glucose but more slowly than ATP PC and it can be produced for a longer period of time</p> <p>Lactic Acid - a chemical your body produces when your cells break down carbohydrates for energy.(cleveland clinic)</p>
Cited references to follow up on	None
Follow up Questions	<ol style="list-style-type: none"> 1. How do external factors impact the body while running? 2. How would the energy sources change if the race was longer or shorter? 3. What are ways to limit lactic acid during the race? 4. How do the foods that athletes eat right before a race impact their body during a race?

Article #5 Notes: Artificial Intelligence based voice assistant

Article notes should be on separate sheets

Source Title	Artificial Intelligence-based Voice Assistant IEEE Conference Publication IEEE Xplore
Source citation (APA Format)	Subhash, S., Srivatsa, P. N., Siddesh, S., Ullas, A., & Santhosh, B. (2020, July 27-28). Artificial Intelligence-based Voice Assistant. <i>2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4)</i> , London, United Kingdom. pp. 593-596. IEEE. https://doi.org/0.1109/WorldS450073.2020.9210344
Original URL	https://ieeexplore.ieee.org/abstract/document/9210344
Source type	Conference Paper
Keywords	Voice Recognition, Text To Speech
#Tags	#voiceRecognition
Summary of key points + notes (include methodology)	<p>Goal: to make a simple assistant through voice user interface This is an idea I want to pursue and improve.</p> <p>This is a voice-based assistant that will listen to what the user has asked for and then follow through with whatever command it is given. These commands are created using Text to Speech technology that will turn all audio files into a text string. The program will then look through the string and identify any command that is given. Once it finds the command it will follow through with it. The program opens website links, opens apps, restarts the computer, calls for help, tells the weather, and tells the location.</p> <p>They met their goal of making a non-hardware intensive program and made it simple.</p> <p>I believe I can add more to this program as I want to add more interactivity to websites as the user can ask the program to scroll down. I also think I may use Pytorch for this project as it is a python framework used to train machine learning models. I may also use OpenAI's API in order to add more ways the program can assist you, such as answering more questions. I may also use Pytorch's facial recognition in order to track the users face and control the computer that way(example: wink = click)</p>
Research Question/Problem/Need	How can we make communication between humans and computers easier?
Important Figures	https://ieeexplore.ieee.org/mediastore/IEEE/content/media/9203790/9210258/9210344/RP-102-399-fig-1-source-large.gif

	This shows how people communicate with this program to make interactions with computers easier.
VOCAB: (w/definition)	ASR: Automatic speech recognition, the main principle behind the working of AI-based Voice Assistant. It records speech and cleans out background noise.
Cited references to follow up on	Chowdury, S. S., Talukdar, A., Mahmud. A., & Rahman, T. (2018, October 28-31). Domain specific Intelligent personal assistant with bilingual voice command processing. <i>TENCON 2018 - 2018 IEEE Region 10 Conference</i> , Jeju, Korea (South), 2018, pp. 0731-0734. IEEE. https://doi.org/10.1109/TENCON.2018.8650203 .
Follow up Questions	<ol style="list-style-type: none"> 1. What other functions can be implemented into this? 2. How can this application be made compatible with all computer/microphone set ups? 3. What issues can arise that can hinder the performance of this application, and how can we fix them?

Article #6 Notes: AI voice assistant tools

Article notes should be on separate sheets

Source Title	Smart AI Voice Assistant through Generative Text Transformer and NLP Implementation in Python IEEE Conference Publication IEEE Xplore
Source citation (APA Format)	Bajpai, D., Kiran, M. U., Reddy, B. H., & Natarajan, S. K. (2024, June 21-23). Smart AI Voice Assistant through Generative Text Transformer and NLP Implementation in Python. <i>2024 4th International Conference on Intelligent Technologies (CONIT)</i> , Bangalore, India, pp. 1-6. IEEE. https://doi.org/10.1109/CONIT61985.2024.10626557
Original URL	https://ieeexplore.ieee.org/abstract/document/10626557?casa_token=mP2ku9rgA9YAAAAA:8xL_gJVauJlaiBGK6gDbxS02BgbWcifgUS4Jvwf6R0ZGcWulRRcSyMcuV_2CtnyLvQEzSjq4sw
Source type	Conference Paper
Keywords	Productivity, Ethics, Privacy, Transforms, Transformers, Natural language processing, User experience, NLP, AI, GTT, Speech Recognition, Contextual Relevance, Unified Framework, Empowerment
#Tags	#voiceRecognition #tools
Summary of key points + notes (include methodology)	<p>This article was helpful in clarifying how voice recognition works.</p> <p>The main parts of it include:</p> <p>Automatic Speech Recognition: an algorithm that will transcribe speech into text form and overcomes challenges such as accents and pronunciation.</p> <p>Natural Language Processing: language models that allow semantic understanding that will make the computer comprehend the meaning of user input.</p> <p>Using Google TTS to transcribe text into string in order to make a smart AI email application [2]. A good voice recognition bot can be offline to not rely on the internet.</p> <p>A lot of these AI models seem to use Python, OpenAI's ChatGPT, and the use of open source libraries.</p> <p>Data protection, encrypt information in order to protect sensitive details such as bank accounts. I don't think I'll make a program that will save anything from the user so I don't have to worry about data safety.</p> <p>Virtual assistant goal: Improve user intractability and convenient assistance.</p> <p>New idea: allow the AI to perform math.</p> <p>GTTS (Google Text-to-Speech), Speech Recognition System[14], Selenium: searching capabilities [15], Wolfram API for math, Play Sound API allows program to interact with user, and Pyaudio[16].</p>
Research Question/ Problem/ Need	Objective: To dissect the technological underpinnings of Smart AI and evaluate its performance by analyzing ASR and NLP.

Important
Figures

$$y = f(x)$$

Y is the output

X is the input

F is the mathematical computation performed by the AI

Fig. 2.

Process of weather details from the internet (a) call for the weather by assistant (b) process of weather details from the internet

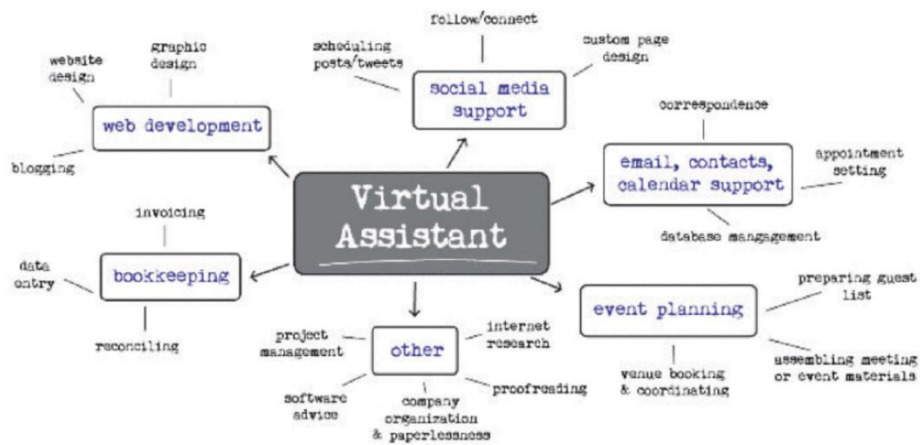
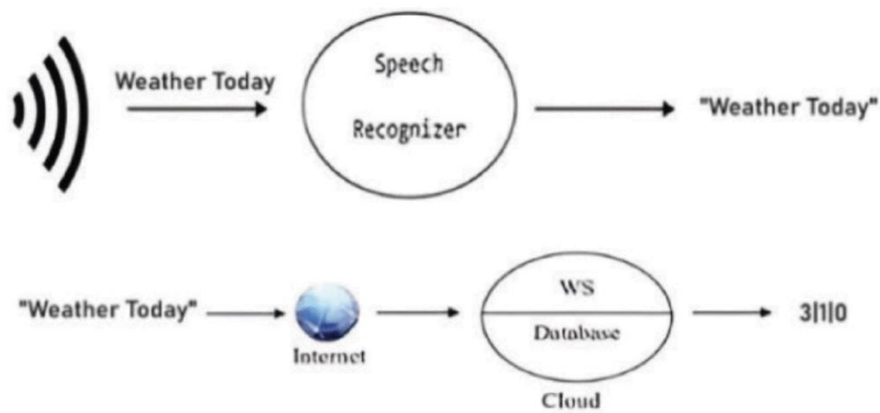


Fig. 3.

Virtual Assistant (Requirements)

VOCAB: (w/definition)	<p>ASR: Transcribes sound and will overcome accents or pronunciations</p> <p>NLP: Semantic understanding, a machine learning technology that gives computers the ability to interpret, manipulate, and comprehend human language.(google)</p>
Cited references to follow up on	<p>Singh, S., Arora, D. K., Dar, I. N., Moghni, A., Kumar, S. & Kumar, A. (2022, February 23-25). <i>ARIA The Bot. 2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM)</i>, Gautam Buddha Nagar, India, pp. 167-174. IEEE. https://doi.org/10.1109/ICIPTM54933.2022.9753961</p> <p>Shoeb, M., Kolluru, V.R., Naga Venkat Sai, M., Mustafa Baig, M., Razia, S. Kumar, A., Mozar, S. (eds). (2022). Implementation of Artificial Intelligence Based Sustainable Smart Voice Assistance. 4th International Conference on Communication and Cyber Physical Engineering (ICCCE 2021). Singapore. <i>Lecture Notes in Electrical Engineering</i>, 828. Springer. https://doi.org/10.1007/978-981-16-7985-8_112</p> <p>Pandey, D., Maitrey, S., & Seth, D. (2022, August 13–14). Artificially developed intelligent system using Python. <i>In Proceedings of the AIP Conference</i>, 2597(1). 30002. Ghaziabad, India. AIP https://doi.org/10.1063/5.0116687</p>
Follow up Questions	<ol style="list-style-type: none"> 1. How can we make AI safe from recording sensitive information? 2. What opensource tools can be used to transcribe sounds? 3. How can I make an AI interact with websites, not just search features?

Article #7 Notes: ARIA the bot

Article notes should be on separate sheets

Source Title	ARIA The Bot
Source citation (APA Format)	Singh, S., Arora, D. K., Dar, I. N., Moghni, A., Kumar, S., & Kumar, A. (2022, February 23-25). ARIA The Bot. <i>2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM)</i> , Gautam Buddha Nagar, India, pp. 167-174. IEEE. https://doi.org/10.1109/ICIPTM54933.2022.9753961 .
Original URL	https://ieeexplore.ieee.org/abstract/document/9753961/keywords#keywords
Source type	Conference paper
Keywords	Bot (Internet), Text recognition, Operating systems, Encyclopedias, Motion pictures, Software, Libraries
#Tags	#voiceRecognition #LanguageModel
Summary of key points + notes (include methodology)	<p>This is a voice recognition system named ARIA which is inspired by JARVIS from Marvel movies. The goal of ARIA is to understand the Indian English accent and perform the tasks that it is asked to. Unlike another model, which was difficult to use as someone had to set up and carry a raspberry pi. It also is meant to deal with the issue of virtual assistants not having a wide range of abilities. The strengths of this bot compared to other robots is its ability to search things on the internet AND on the user's computer.</p> <p>The tools that were used to create this were: pyqt5, pyttsx3, speech Recognition, pywin32, and pyaudio.</p>
Research Question/Problem/Need	How can we make a voice assistant that can be utilized on and offline?
Important Figures	Fig 5.5 Opening of YouTube

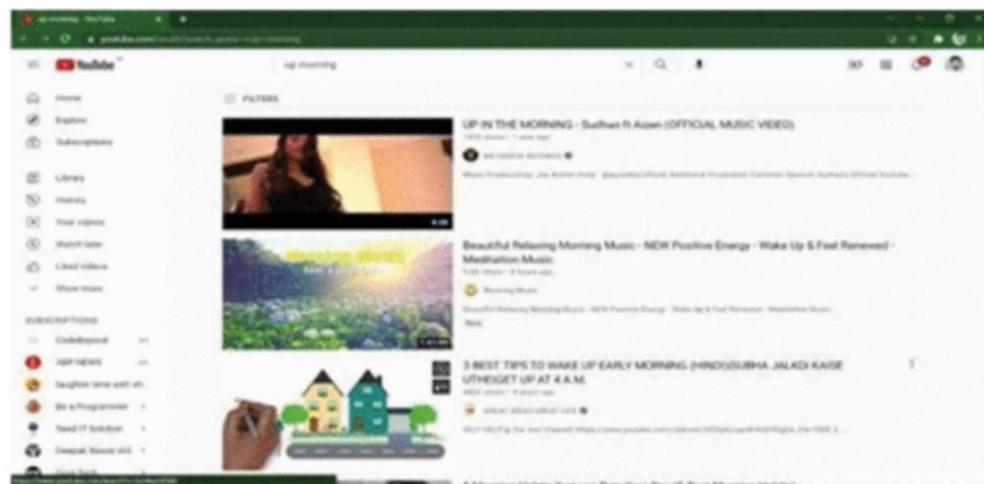
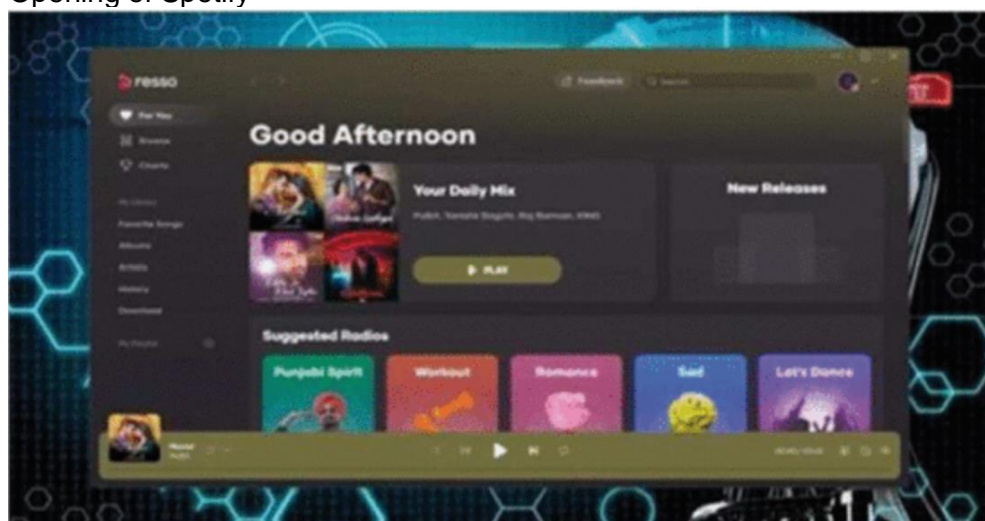


Fig 5.6
Opening of Spotify



VOCAB: (w/definition)

API - Application programming interface, a kind of software interface that offers a service or other piece of software(IBM).
GUI – General User Interface, what the user sees when they open an app.

Cited references to follow up on

Sinha, A. (2018). A review of Voice Based Personal Assistant. *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*

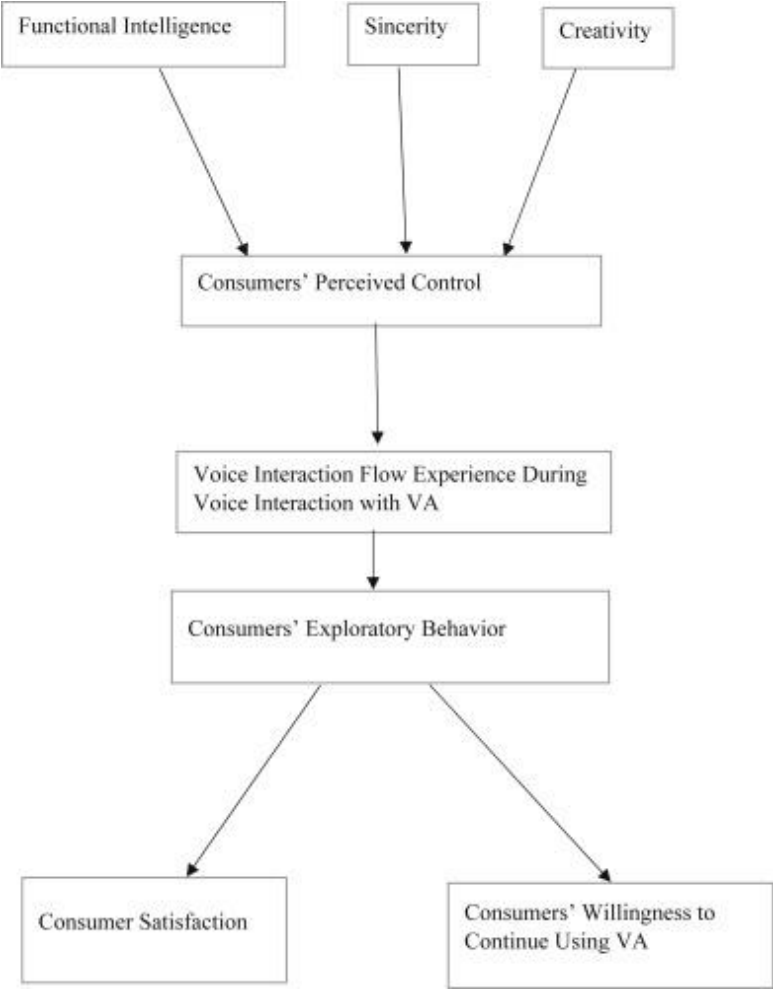
Follow up Questions

1. How can this program be modified to recognize other accents?
2. How can the accessibility when it comes to other apps be increased in this bot?
3. How can usability be increased? (making a user-friendly GUI?)

Article #8 Notes: Humanizing voice Assistants

Article notes should be on separate sheets

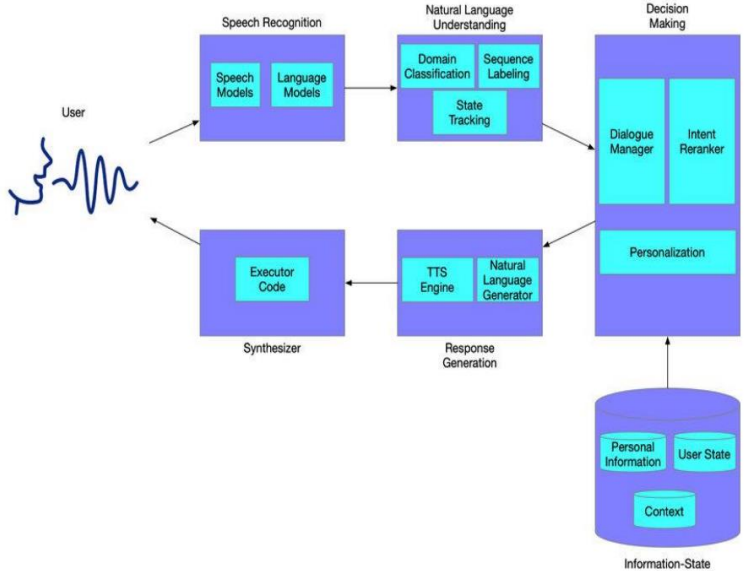
Source Title	Humanizing voice assistant: The impact of voice assistant personality on consumers' attitudes and behaviors
Source citation (APA Format)	Poushneh, A. (2021). Humanizing voice assistant: The impact of voice assistant personality on consumers' attitudes and behaviors. <i>Journal of Retailing and Consumer Services</i> , 58, 102283. https://doi.org/10.1016/j.jretconser.2020.102283
Original URL	https://www.sciencedirect.com/science/article/pii/S0969698920312911
Source type	Journal Article
Keywords	Voice assistant personality (VAP), Voice interaction flow experience, Control, Focused attention, Exploratory behavior, Satisfaction, Willingness to continue using voice assistant (VA)
#Tags	#Personality
Summary of key points + notes (include methodology)	<p>Voice Assistants aren't human; however, they can have a fake personality programmed into them. This personality increases engagement from users. This has been tested with 275 consumers (157 males, and 120 females) between 21–41-years old who all frequently use voice assistants. They interacted with voice assistants for about 7 minutes and ranked them based off of 7 personality traits.</p> <p>Functional intelligence: ability to complete tasks. Aesthetic Appeal: how appealing the UI for the VA is. Protective Quality: how environmentally aware, loving, and protective VA is Sincerity: honesty and agreeable information. Creativity: how thoughtful, creative, and brilliant VA is in providing information that is trendy, smooth, contemporary, and up-to-date. Sociability: How engaging the VA is. Emotional Intelligence: when jokes and empathy are displayed.</p> <p>Findings: These traits positively influence the users control and makes them more engaged and willing to use the assistant in the future.</p>
Research Question/Problem/ Need	<p>Purpose: explore personality traits associated with voice assistant mobile applications.</p> <ol style="list-style-type: none"> 1) what are the salient personality traits of voice assistant? 2) how do the personality traits drive consumers attitude and behaviors through voice interaction flow experience?

<p>Important Figures</p>	 <pre> graph TD FI[Functional Intelligence] --> CPC[Consumers' Perceived Control] S[Sincerity] --> CPC C[Creativity] --> CPC CPC --> VIFED[Voice Interaction Flow Experience During Voice Interaction with VA] VIFED --> CE[Consumers' Exploratory Behavior] CE --> CS[Consumer Satisfaction] CE --> CW[Consumers' Willingness to Continue Using VA] </pre> <p>Figure 1</p>
<p>VOCAB: (w/definition)</p>	<p>Flow experience: immersion while using the product.</p>
<p>Cited references to follow up on</p>	<p>Araujo, T. (2018). Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. <i>Computers in Human Behavior</i>, 85(85), 183–189. https://doi.org/10.1016/j.chb.2018.03.051</p>
<p>Follow up Questions</p>	<ol style="list-style-type: none"> 1. How does the voice of the AI affect how people engage with it? 2. How does this effect change with different age groups? 3. How does the rate of speech affect how people immerse themselves?

Article #9 Notes: Design and Development of Intelligent Voice Personal Assistant using Python

Article notes should be on separate sheets


Source Title	Design and Development of Intelligent Voice Personal Assistant using Python
Source citation (APA Format)	Appalaraju, V., Rajesh, V., Saikumar, K., Sabitha , P., & Kiran, K. R.(2021, March 9). Design and Development of Intelligent Voice Personal Assistant using Python. <i>2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)</i> , Greater Noida, India. pp. 1650-1654. IEEE. https://doi.org/10.1109/ICAC3N53548.2021.9725753 .
Original URL	Design and Development of Intelligent Voice Personal Assistant using Python IEEE Conference Publication IEEE Xplore
Source type	Conference paper
Keywords	Text recognition, Speech recognition, Writing, Motion pictures, Feature extraction, Task analysis, Smart devices, intelligent voice personal assistant, speech recognition, artificial intelligence, virtual voice assistant
#Tags	#voiceRecognition
Summary of key points + notes (include methodology)	<p>The python speech to text library was used in order to process what a user has said.</p> <p>Tools that were used: Py Audio, Py jokes, play sound, IMDb, information from Wikipedia using the Wikipedia API, and PyWhatKit is a Python library for exchanging Whatsapp messages at a set time.</p> <p>Accepts prerecorded and live recorded speech. I might implement a slower system that will record a user and then take a while to responses but with better accuracy as Assembly AI doesn't have a lot of live recording features.</p>
Research Question/Problem/ Need	Goal: create a personal desktop assistant that can conduct tasks using the user's voice as a command.

<p>Important Figures</p>	 <p>Fig. 1. Components of the projected voice assistants</p>
<p>VOCAB: (w/definition)</p>	<p>Support Vector Machine - a type of supervised learning algorithm that can be used for classification or regression tasks.</p>
<p>Cited references to follow up on</p>	<p>Pandey, A., Vashist, V., Tiwari, P., Sikka, S., & Makkar, P. (2020). Smart voice-based virtual personal assistants with artificial intelligence. <i>Artificial Computational Research Society</i>, 1(3).</p>
<p>Follow up Questions</p>	<ol style="list-style-type: none"> 1. How can the NLP be improved? 2. What tools can I use to change the voice from the standard robot voice in order to build more engagement with users? 3. How can I ensure getting relevant and up to date information that won't be at risk of being misleading?(Wikipedia isn't the best source)

Article #10 Notes: Artificial Intelligence Based A Communicative Virtual Voice Assistant Using Python & Visual Code Technology

Article notes should be on separate sheets

Source Title	Artificial Intelligence Based A Communicative Virtual Voice Assistant Using Python & Visual Code Technology
Source citation (APA Format)	Jain, R., Sharma, V., Mangilal, Kardam, R., & Rani, M. (2021). Artificial Intelligence Based A Communicative Virtual Voice Assistant Using Python & Visual Code Technology. <i>World Journal of Research and Review (WJRR)</i> , 13(5), 23-25. https://www.wjrr.org/download_data/WJRR1305017.pdf
Original URL	Artificial Intelligence Based A Communicative Virtual Voice Assistant Using Python & Visual Code Technology (wjrr.org)
Source type	Journal article
Keywords	Desktop Assistant, Python, Machine Learning, Text to Speech, Speech to Text, Language Processing, Voice Recognition, Artificial Intelligence, Internet Of Things (IOT), Virtual Assistant.
#Tags	#voiceRecognition
Summary of key points + notes (include methodology)	<p>Speech recognition is everywhere, they make using devices very easy for children, blind people, and people with other disabilities. The tools that were used were speech_recognition library, OS to work with files, https://docs.python.org/3/library/webbrowser.html web browser to open links. I think I may use this for my project in order to be able to navigate websites.</p> <p>They've used datetime which is a library I am already using. They are also using Pyttsx3 in order to turn text to audio. I am also using this but after reading article 8, I think I might change it to ElevenLabs in order to have a more human voice.</p> <p>They also use SMTPLIB in order to send and handle emails. I think I may add this to my project goals.</p> <p>Just like the other assistants, this AI turns recordings to text and analyzes the commands. It also asks for clarification when the user is not clear.</p>
Research Question/Problem/ Need	Making a voice assistant will make navigation of a device easier for those with disabilities or who are too young to use a device.

<p>Important Figures</p>	 <pre> graph TD START([START]) --> Record[Record the voice using Microsoft speech software] Record --> GetMFC[Get Mel-Frequency Cepstrum Coefficient stored into reference template] GetMFC --> Measure[Get measuring similarity between training and testing input voice signal] Measure --> Receive[Receive external voice command (Speaker)] Receive --> Match{Match with reference template} Match --> Send[Send the signal to activate decision command] Send --> STOP([STOP]) Match --> Receive </pre> <p>Figure 7</p>
<p>VOCAB: (w/definition)</p>	<p>No new vocab</p>
<p>Cited references to follow up on</p>	<p>P. Nguyen, G. Heigold, & G. Zweig. (2010). Speech Recognition With Flat Direct Models. <i>IEEE Journal of Selected Topics in Signal Processing</i>, 4(6), 994-1006. https://doi.org/10.1109/JSTSP.2010.2080812.</p>
<p>Follow up Questions</p>	<ol style="list-style-type: none"> 1. How can the voice recognition of this AI be improved? 2. How can the information this AI gets be more accurate? 3. How can WebBrowser be used in order to navigate websites through voice control?

Article #11 Notes: AI-based Desktop Voice Assistant

Article notes should be on separate sheets

Source Title	AI-based Desktop Voice Assistant
Source citation (APA Format)	Kunekar, P., Deshmukh, A., Gajalwad, S., Bichare, A., Gunjal, K., & Hingade, S. (2023, January 20-21). AI-based Desktop Voice Assistant. <i>2023 5th Biennial International Conference on Nascent Technologies in Engineering (ICNTE)</i> , Navi Mumbai, India. pp. 1-4. IEEE. https://doi.org/10.1109/ICNTE56631.2023.10146699 .
Original URL	https://ieeexplore.ieee.org/document/10146699
Source type	Conference proceeding
Keywords	Technological innovation, Portable computers, Virtual assistants, Process control, Speech recognition, Natural language processing, Task analysis, Artificial intelligence, Natural language processing, Text to Speech, Voice assistant
#Tags	#voiceRecognition
Summary of key points + notes (include methodology)	<p>virtual assistant that can translate Indian language Marathi. Many IT companies use voice assistants to work with Google assistant, Cortana, and Amazon Alexa.</p> <p>Voice recording gets turned into a machine readable format and then use pyaudio to visualize the result of what was received.</p> <p>User input is taken through google speech recognition system. Can write and read files. It can act like a chatbot. It returns news and weather. It can also modify the News that a user wants to listen to.</p> <p>Goal: help user to perform various tasks with voice commands only. Overall: Made a desktop application that uses google's speech recognition API. This can be used to retrieve information from IoT and communicate with nearby devices such as an Alexa.</p>
Research Question/Problem/ Need	Engineering goal: to make a desktop assistant that utilizes natural language processing in order to assist users with retrieving information and communicate with other virtual assistants in the room.

Important Figures

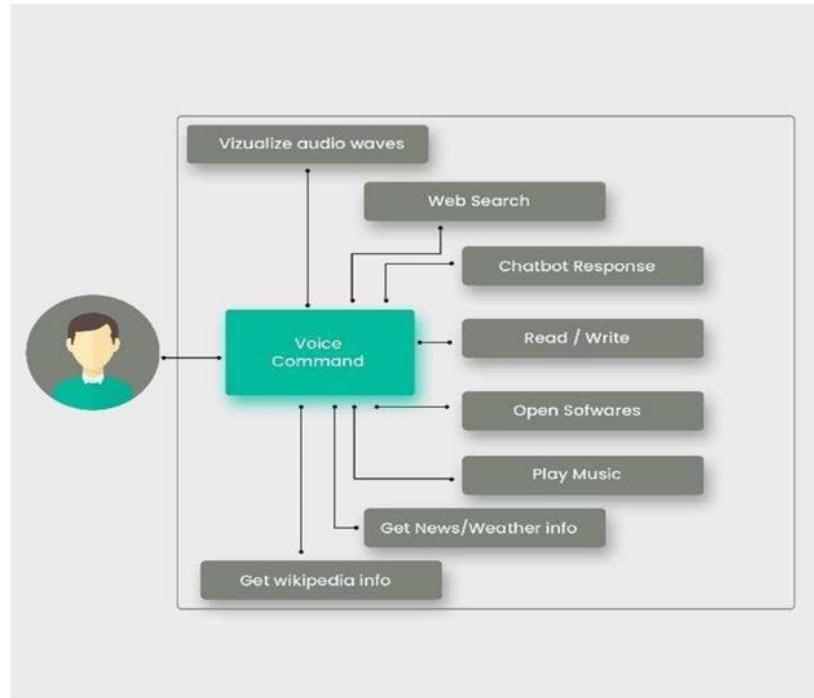


Fig 2. Use case diagram of Voice Assistant



Fig 4. Screenshot of Input module

VOCAB: (w/definition)

Internet of Things - The Internet of Things (IoT) refers to a network of physical devices, vehicles, appliances, and other objects embedded with sensors, software, and network connectivity, enabling them to collect and exchange data.

Audio Visualization: the wavy effect that is emitted by assistants to make it seem like audio is visual.

Cited references to follow up on	Singh, N., Yagyasen, D., Singh, S. V., Kumar, G., & Agrawal, H. (2021). Voice assistant using Python. <i>International Journal of Innovative Research in Technology</i> , 8(2), 1-5
Follow up Questions	<ol style="list-style-type: none"> 1. What makes a User Interface appealing to users? 2. How can I fix weaknesses and challenges this AI faces, like false information on the internet? 3. How can I make my program interact with other voice assistants?

Article #12 Notes: Development of GUI for Text-to-Speech Recognition using Natural Language Processing

Article notes should be on separate sheets

Source Title	Development of GUI for Text-to-Speech Recognition using Natural Language Processing
Source citation (APA Format)	Mukherjee, P., Santra, S., Bhowmick, S., Paul, A., Chatterjee, P., & Deyasi, A. (2018, May 04-05). Development of GUI for text-to-speech recognition using natural language processing. <i>2018 2nd International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech)</i> , Kolkata, India. pp. 1-4. https://doi.org/10.1109/IEMENTECH.2018.8465238
Original URL	Development of GUI for Text-to-Speech Recognition using Natural Language Processing IEEE Conference Publication IEEE Xplore
Source type	Conference Proceeding
Keywords	Speech recognition, synthesizers, speech synthesis, natural language processing, computer applications, linguistics, text-to-speech, natural language processing, speech synthesizer, speech recognition, signal transformation.
#Tags	#VoiceRecognition #GUI
Summary of key points + notes (include methodology)	<p>TTS takes text as input and analyzes using a TTS engine and sends back an audio as output.</p> <p>TTS system worked upon Natural Language Generator.</p> <p>Optical character recognition process.</p> <p>It converts the phonetic and prosodic information into a wave form based upon approximation formula.</p> <p>A synthesizer is developed to work with TTS conversion and then save a file as an MP3.</p> <p>Synthesized speech is a collection of small pieces of recorded speech which are stored in a knowledge base.</p> <p>The front interfaces engage with phonetic conversion to each unit, and divides and marks to form a speech tree or pattern tree using the speech unit which configures the tune and rhythm through phrases, clauses, and sentences. This process of transcriptions is known as text-to-phoneme (TTP) or grapheme-to-phoneme (GTP) conversion.</p> <p>Symbolic representation of TTS.</p> <p>Speech synthesis is done in many ways: Concatenative Synthesis (Unit-selection Synthesis, Diphone Synthesis, and Domain Specific Synthesis), Formant Synthesis, Articulatory Synthesis, HMM-based Synthesis, sinewave Synthesis, etc.</p>

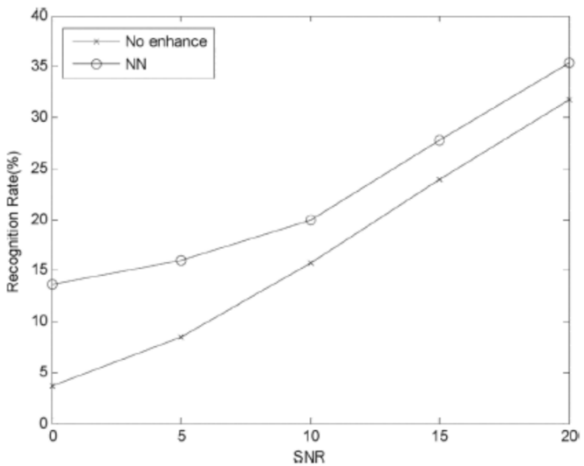
	<p>Design:</p> <p>System is developed using C#, and dot net 3.5</p> <p>TTS Gramaty converts tts either by typing the text into the text field provided or by copying from an external doc in the local machine and then pasting it in the text field provided in the application. Then it begins the reading process.</p> <p>Full screen has a input field which will take in text and a button can allow the user to treat it like an audiobook.</p> <p>There is a volume controller to control the audio output volume, a Speech Speed Controller, to control the speech rate. Speak now, Pause, Resume and stop buttons with their respective purpose.</p>
<p>Research Question/Problem/ Need</p>	<p>What are the components of speech synthesis?</p>
<p>Important Figures</p>	<div data-bbox="613 793 1101 1612"> </div> <div data-bbox="571 1621 683 1661"> <p>Figure 1</p> </div> <div data-bbox="571 1684 686 1724"> <p>Figure 2</p> </div>

	<pre> graph TD subgraph PART_OF_NLP [PART OF NLP] P[PHONEMES + PROSODY] --> SLG[Segment List Generation] end subgraph PART_OF_SPEECH_PROCESSING [PART OF SPEECH PROCESSING] SC[Speech corpus] --> SSA[Selective Segmentat] PSD[Parametric Segment Database] --> SSA SSA --> SSD[Speech Segment Database] SSA --> SA[Speech Analysis] SSD --> SA SA --> E[Equalization] E --> SCOD[Speech Coding] end subgraph PART_OF_SOUND_PROCESSING [PART OF SOUND PROCESSING] SSDS[Synthetic Segment Database] --> SU[Speech Uncodeing] SU --> PM[Prosody Matching] SLG --> PM PM --> SC[Segment Concatenation] SC --> SS[Signal Synthesis] SS --> S[Speech] end SCOD --> DSP[DSP Units] SLG --> DSP DSP --> PM DSP --> SC DSP --> SS DSP --> S </pre>
VOCAB: (w/definition)	Digital Signal Processing - the manipulation of signals after they have been converted into a digital format
Cited references to follow up on	Thu, C. S. T., & Zin, T. (2014). Implementation of text to speech conversion. <i>International Journal of Engineering Research & Technology</i> , 3(3), 911–915. https://www.ijert.org/implementation-of-text-to-speech-conversion
Follow up Questions	<ol style="list-style-type: none"> 1. What parts of my program should involve TTS? 2. Does TTS voice affect the way people engage in the app? 3. What challenges does TTS often face?

Article #13 Notes: Noise reduction algorithm for robust speech recognition using MLP neural network

Article notes should be on separate sheets

Source Title	Noise reduction algorithm for robust speech recognition using MLP neural network
Source citation (APA Format)	Ghaemmaghami, M. P., Razzazi, F., Sameti, H., Dabbaghchian, S. & BabaAli, B. (2009, November 28-29). Noise reduction algorithm for robust speech recognition using MLP neural network. <i>2009 Asia-Pacific Conference on Computational Intelligence and Industrial Applications (PACIIA)</i> , Wuhan, China, pp. 377-380. IEEE. https://doi.org/10.1109/PACIIA.2009.5406411 .
Original URL	Noise reduction algorithm for robust speech recognition using MLP neural network IEEE Conference Publication IEEE Xplore
Source type	Conference proceeding
Keywords	Noise reduction, Noise robustness, Speech recognition, Neural networks, Working environment noise, Speech enhancement, Feature extraction, Application software, Speech processing, Databases, MLP neural network, log spectral, robust speech recognition
#Tags	#MachineLearning
Summary of key points + notes (include methodology)	<p>MLP reduces the difference between noisy and clean speech. Used this application through different environments without needing to retrain data because the only time data needs to be trained is in the preprocessing stage with a small portion of noisy data which is created by artificially adding types of noises from databases.</p> <p>Speech enhancement - reduces the distortion caused by ambient noise Feature extraction - the features representing the speech signal are designed in order to be less sensitive to noise conditions. Model Compensation - to determine the influence of noise on the distributions of speech features and to modify the models used in the recognition to take into account the influence of the noise.</p> <p>multistream and missing features have been proposed for dealing with the mismatch problem. These techniques Less weight to noisy parts of the speech signal using signal to noise ratio and views differences between frequency bands.</p> <p>Used weight initialization to speed up the training of the neural networks Artificially added noises from databases, compared the energy ratio of the clean speech signal including silence periods and the added noise</p>

	<p>within each sentence.</p> <p>Described a nonlinear noise reduction algorithm motivated the MMSE criterion.</p> <p>Experimental results show that these methods improve recognition accuracy in different cases.</p> <p>Designed to apply to feature extraction so that it can be used in ASR systems</p>
Research Question/Problem/ Need	Creating a speech recognition system that reduces background noise.
Important Figures	 <p>Figure 2</p>
VOCAB: (w/definition)	<p>Multi layer perceptron - a type of artificial neural network with multiple layers of neurons.</p> <p>Signal to Noise Ratio - a measure of how clear a signal is compared to the background noise.</p>
Cited references to follow up on	<p>Rabiner, L., & Juang, B. H. (1993). <i>Fundamentals of speech recognition</i>. Prentice Hall.</p> <p>Yu, D., Deng, L., Droppo, J., Wu, J., Gong, Y., & Acero, A. (2008). Robust Speech Recognition Using a Cepstral Minimum-Mean-Square-Error-Motivated Noise Suppressor. <i>IEEE Transactions on Audio, Speech, and Language Processing</i>, 16(5), 1061-1070. https://doi.org/10.1109/TASL.2008.921761</p>
Follow up Questions	<ol style="list-style-type: none"> 1. How can this be improved? 2. How does an MLP work? 3. Are there any other algorithms that can improve noise reduction and then be applied to ASR?

Article #14 Notes: Indonesian Automatic Speech Recognition system using CMUSphinx toolkit and limited dataset

Article notes should be on separate sheets

Source Title	Indonesian Automatic Speech Recognition system using CMUSphinx toolkit and limited dataset
Source citation (APA Format)	Prakoso, H., Ferdiana, R., & Hartanto, R. (2016, November 26-30). Indonesian Automatic Speech Recognition system using CMUSphinx toolkit and limited dataset. <i>2016 International Symposium on Electronics and Smart Devices (ISESD)</i> , Bandung, Indonesia, pp. 283-286. IEEE. https://doi.org/10.1109/ISESD.2016.7886734 .
Original URL	Indonesian Automatic Speech Recognition system using CMUSphinx toolkit and limited dataset IEEE Conference Publication IEEE Xplore
Source type	Conference Paper
Keywords	Hidden Markov models, Acoustics, Training, Automatic speech recognition, Speech, Signal to noise ratio, Indonesian Automatic Speech recognition, CMUSphinx, acoustic model, SNR
#Tags	#ASR #AcousticModel
Summary of key points + notes (include methodology)	<p>Built an automatic speech recognition model for an Indonesian language using CMUSphinx(which is a Hidden Markov Model based ASR tool) with a limited data set.</p> <p>Investigated the implementation in different noise environments and got the maximum accuracy of 80% in a 27.764 dB environment.</p> <p>The earliest ASR systems are Dragon and Carnegie Mellon University's Harpy. However, the harpy system only recognizes 1011 words vocabulary. Accuracy greater than 85% can only be obtained when the task is restricted in some way.</p> <p>To create an ASR Model you need a: acoustic model, language model, and lexicon. Creating an acoustic model is costly.</p> <p>Provided its own dictionary, audio dataset, and language model that was used to train the acoustic model with CMUSphinx toolkit.</p> <p>Arabic ASR has an accuracy of 96.67% with a small vocabulary. Most widely used algorithm in modern ASR is the Hidden Markov Model to build an acoustic model.</p> <p>Back end component named unit matching system, responsible for recognizing the observed feature of the speech signal by the concatenate information from the acoustic model, language model,</p>

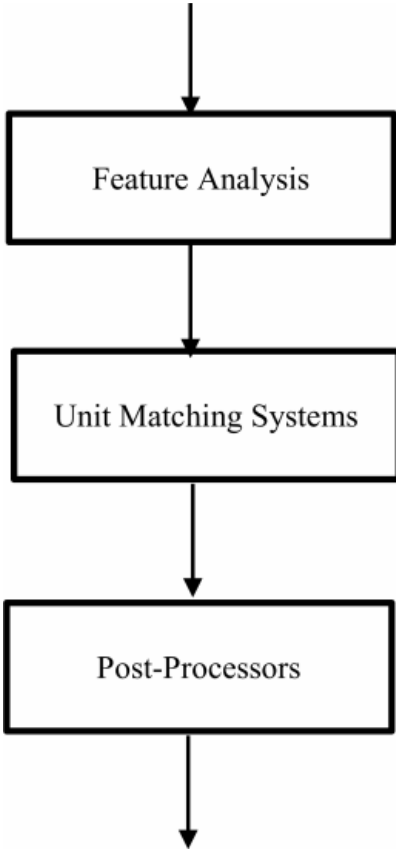
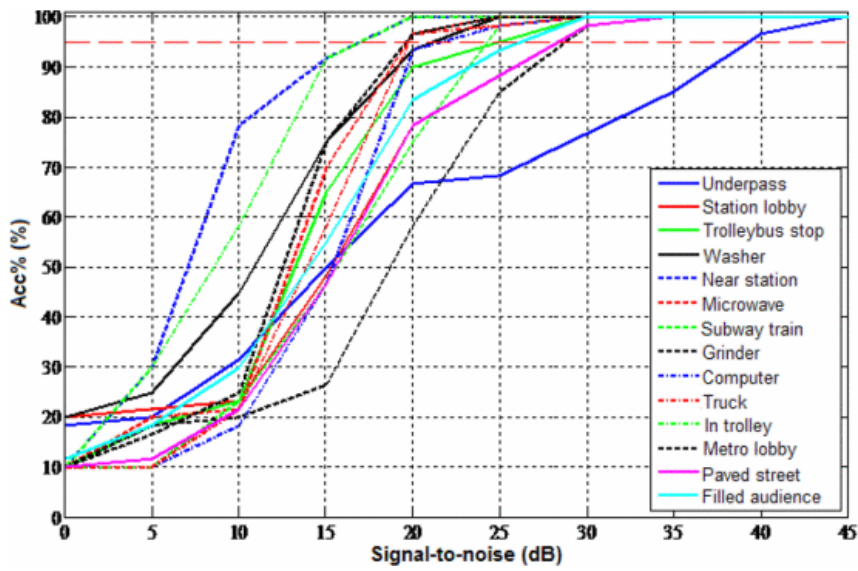
	<p>and the lexicon.</p> <p>Language model is a collection of probabilistic data assigned to a stream of words.</p> <p>Unigram or Ngram models can be used for estimating the probability of words sequence. Unigram models are used to find information. Meanwhile, Ngram models are for approximating long sentences and sequences which aren't seen during data training.</p> <p>Lexicon is the vocab: words and expressions</p> <p>Acoustic model creation:</p> <p>dictionary files with Indonesian digit vocab</p> <p>language model that contains ngram resulted from a toolkit named Imtool</p> <p>video</p> <p>tested by training and testing using CMUSphinx to measure accuracy</p> <p>tested different accuracies in different environments.</p> <p>Measures using sentence error ratio, after testing with a sample of 100 words, 14 were wrong, the accuracy is 86%.</p> <p>ASR system has an accuracy of 80%, 55%, 50%, 25%, 11% and 3% with SNR value are 27.764 dB;22.798 dB;12.865 dB;7.265 dB;5.651 dB and 1.122 dB respectively</p>
Research Question/Problem/ Need	To make an ASR model to recognize an indonesian language.
Important Figures	 <pre> graph TD A[] --> B[Feature Analysis] B --> C[Unit Matching Systems] C --> D[Post-Processors] D --> E[] style A fill:none,stroke:none style E fill:none,stroke:none </pre> <p>The diagram illustrates the ASR system components in a vertical flow. It starts with an input arrow pointing to a box labeled 'Feature Analysis'. An arrow then points down to a box labeled 'Unit Matching Systems'. Another arrow points down to a box labeled 'Post-Processors'. Finally, an arrow points down from the 'Post-Processors' box, indicating the output of the system.</p>

	Figure 1: An automatic speech recognition system
VOCAB: (w/definition)	<p>Hidden Markov Model - A statistical model called a Hidden Markov Model (HMM) is used to describe systems with changing unobservable states over time. Hidden Markov Model in Machine learning - GeeksforGeeks</p> <p>Acoustic model- a type of machine-learning model that is used in speech recognition systems. The model is trained to recognize the acoustic properties of human speech and generate a corresponding text transcription. Acoustic models are typically trained with large datasets of spoken words, phrases, and sentences. These datasets are used to create a “map” of the sounds of human speech. What Is A Acoustic Model In Speech Recognition - Try Speak Free!</p>
Cited references to follow up on	<p>Lowerre, B. (1990, May 1). <i>The Harpy speech understanding system: Readings in speech recognition</i>. Association for Computer Machinery. https://dl.acm.org/doi/abs/10.5555/108235.108277.</p>
Follow up Questions	<ol style="list-style-type: none"> 1. How can an Acoustic model be made? 2. Why is an HMM the most used model? 3. How can I make a language model multilingual?

Article #15 Notes: Training of automatic speech recognition system on noised speech

Article notes should be on separate sheets

Source Title	Training of automatic speech recognition system on noised speech
Source citation (APA Format)	Prodeus, A., & Kukharicheva, K. (2016, October 18-20). Training of automatic speech recognition system on noised speech. <i>2016 4th International Conference on Methods and Systems of Navigation and Motion Control (MSNMC)</i> , Kiev, Ukraine, pp. 283-286. IEEE. https://doi.org/10.1109/MSNMC.2016.7783147
Original URL	https://ieeexplore.ieee.org/document/7783147
Source type	Conference Paper
Keywords	Speech, Training, Signal to noise ratio, Noise measurement, Automatic speech recognition, Navigation, automatic speech recognition, speech recognition accuracy, training technique, clean speech, noised speech.
#Tags	#Training
Summary of key points + notes (include methodology)	<p>2 techniques of training on noised speech are compared with training clean speech.</p> <p>They were compared based on accuracy with the usage of 14 kinds of noise.</p> <p>These were noises of household appliances and computers, street and transport, teaching rooms, and lobbies.</p> <p>Training on noised speech allows reaching the 95% recognition accuracy for minimal SNR 10 dB. Training on clean speech allows reaching the same accuracy for minimal SNR ratio 20 dB.</p> <p>F-35 was the first US fighter aircraft with an automatic recognition system able to hear pilots' spoken commands in order to operate the aircraft's weapons.</p> <p>The first approach: ASR is trained on clean speech, the second approach has ASR on noisy speech</p> <p>Fully matched training method is very effective for SNR = 5dB Speech recognition accuracy = 75% whereas accuracy for clean speech is 25%</p> <p>Multi-style learning makes clean speech recognition better.</p> <p>$s(t)=k \cdot x(t)+n(t), k=100.05^{(SNR0-SNR)}$, $x(t)$ is clear speech signal, $n(t)$ is noise, SNR is signal-to-noise ratio for</p>

	<p>saved clear speech signal. SNR0 value was varied in the range 0–45 dB</p> <p>Used Russian number names in order to test these, and noises such as household appliances, computers, and transport vehicles.</p> <p>95% accuracy for street noise is only achieved when the SNR is over 28 dB. meanwhile with the FMT technique. SNR 5-15 dB will produce the same accuracy result but will decrease later on.</p> <p>SMT technique allows to reach the 95% recognition accuracy for SNR, \geq 10 dB. FMT can do the same, but is much more demanding to the volume of ASR system memory. When training on clean speech, 95% recognition accuracy was reached only for five of the fourteen kinds of noise interference for SNRr \geq 20 dB.</p>
Research Question/Problem/ Need	Comparing different techniques of training ASR models and then seeing their results.
Important Figures	 <p>Fig. 1. Acc% for ASR system trained on clean speech.</p>

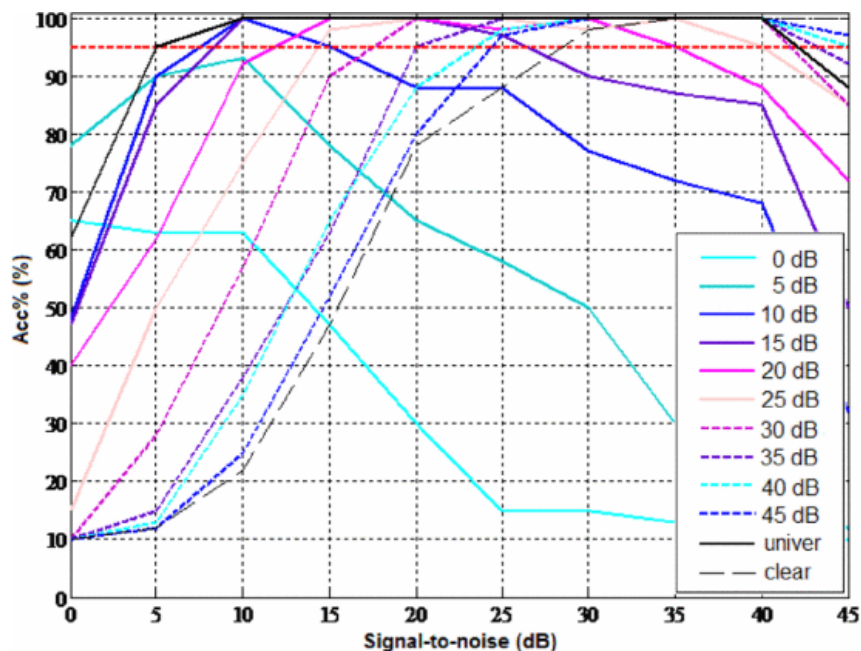


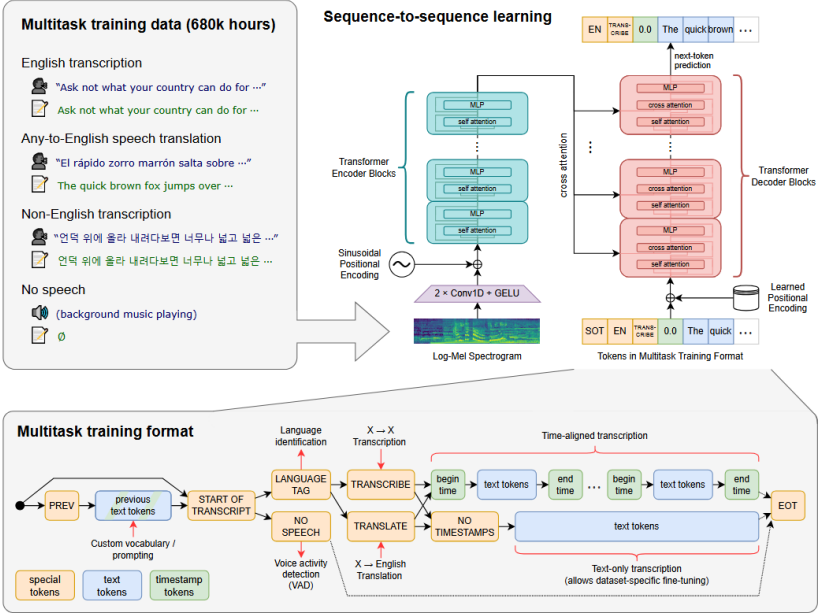
Fig. 2.
Acc% for ASR system trained by FMT technique.

VOCAB: (w/definition)	Spectrum matched training - produce training with varying SNR for noise that will affect the ASR system during recognition. fully matched training - ASR system is trained on speech with the same SNR and noise spectrum for which ASR system will be tested.
Cited references to follow up on	J. Rajnoha. (2009). Multi-Condition Training for Unknown Environment Adaptation in Robust ASR Under Real Conditions. <i>Acta Polytechnica</i> , 49(2-3), 3-7. https://doi.org/10.14311/1105
Follow up Questions	<ol style="list-style-type: none"> 1. What is the reason that different background noises cause a different accuracy level? 2. Which of these methods (or better methods) is commonly found in ASR systems today? 3. Why do different levels of sound under signal to noise ratios create different accuracies?

Article #16 Notes: Robust Speech Recognition via Large-Scale Weak Supervision

Article notes should be on separate sheets

Source Title	Robust Speech Recognition via Large-Scale Weak Supervision
Source citation (APA Format)	Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2022, December 6). <i>Robust speech recognition via large-scale weak supervision</i> . arXiv. https://arxiv.org/abs/2212.04356
Original URL	https://arxiv.org/abs/2212.04356
Source type	Online source
Keywords	Audio and Speech Processing, Computation and Language, Machine Learning, Sound
#Tags	#AutomaticSpeechRecognition
Summary of key points + notes (include methodology)	<p>Created a speech processing system that is trained on 680000 hours. It comes close to levels of human recognition.</p> <p>A model can be accurate in some metrics but not in others. The goal of the system is to be able to work out of the box without fine-tuning.</p> <p>The dataset covers 96 languages through 117000 hours of data. It also has X-> en translation data which includes 125000 hours.</p> <p>This model has the goal to be able to be robust in diverse conditions such as accents or background noises.</p> <p>Diversity for voices helps the model. However, diversity in transcription doesn't help.</p> <p>Since this dataset is large-scale and has weak supervision. This means that a lot of the data can be fuzzy.</p> <p>Used GPT-2 vocabulary. Encoder-decoder transformer as this can scale to higher amounts.</p> <p>One of the problems with this model is that it can have noisy data as it is provided by the internet.</p> <p>It is used for multiple jobs such as translation and voice activity detection.</p> <p>It reduces the reliance on fine tuned datasets and works well in conditions with accents and background noise.</p> <p>However, there are performance gaps for different languages as the data behind those languages is not equal.</p> <p>Doesn't use Word Error Rate Metric to test the accuracy of the models because it penalizes all differences of the output compared to the input. This means that even a large block of text that would be considered correct by humans can have a lot of WER errors.</p>

Research Question/Problem/ Need	<p>To create a robust speech recognition system that can be applied to multiple functions.</p>
Important Figures	 <p>Multitask training data (680k hours)</p> <ul style="list-style-type: none"> English transcription <ul style="list-style-type: none"> "Ask not what your country can do for ..." Ask not what your country can do for ... Any-to-English speech translation <ul style="list-style-type: none"> "El rápido zorro marrón salta sobre ..." The quick brown fox jumps over ... Non-English transcription <ul style="list-style-type: none"> "언덕 위에 올라 내려다보면 너무나 넓고 넓은 ..." 언덕 위에 올라 내려다보면 너무나 넓고 넓은 ... No speech <ul style="list-style-type: none"> (background music playing) ∅ <p>Sequence-to-sequence learning</p> <p>The diagram shows a Transformer Encoder (blue) and a Transformer Decoder (red). The encoder processes a Log-Mel Spectrogram (2 × Conv1D + GELU) and Sinusoidal Positional Encoding. The decoder generates tokens in the Multitask Training Format, including SOT, EN, TRANS, CRASE, 0.0, The, quick, brown, and next-token prediction. Cross-attention is shown between the encoder and decoder.</p> <p>Multitask training format</p> <p>The diagram shows the flow from previous text tokens to START OF TRANSCRIPT, then to LANGUAGE TAG, TRANSCRIBE, TRANSLATE, and finally to text tokens. It includes special tokens (special, text, timestamp) and a custom vocabulary / prompting mechanism. The format also includes a time-aligned transcription section with begin time, text tokens, end time, and EOT.</p>
VOCAB: (w/definition)	<p>Transformers- a kind of deep learning model that is efficient as it trains data in parallel. This makes them require less time to train. https://www.ibm.com/topics/transformer-model</p> <p>Word Error Rate – a kind of metric that would measure how correct the output is relative to the input.</p> $WER = (\text{num inserted} + \text{num deleted} + \text{num substituted}) / \text{num words in the reference}$ <p>https://medium.com/nlplanet/two-minutes-nlp-intro-to-word-error-rate-wer-for-speech-to-text-fc17a98003ea</p> <p>Mel Spectrogram - A time frequency representation of sound https://www.fon.hum.uva.nl/praat/manual/MelSpectrogram.html</p> <p>Fine tuning – retraining a model to fit the case that you are facing better. https://www.ibm.com/topics/fine-tuning</p>

Cited references to follow up on	Ardila, R., Branson, M., Davis, K., Henretty, M., Kohler, M., Meyer, J., Morais, R., Saunders, L., Tyers, F. M., & Weber, G. (2019). Common voice: A massively-multilingual speech corpus. <i>arXiv</i> . https://doi.org/10.48550/arXiv.1912.06670
Follow up Questions	<ol style="list-style-type: none">1. What advantages do Transformers bring compared to RNN's and CNN's?2. How will this performance gap between different languages be fixed?3. In what situations can transformers fail or struggle with?

Article #17 Notes: Desktop based Smart Voice Assistant using Python Language Integrated with Arduino

Article notes should be on separate sheets

Source Title	Desktop based Smart Voice Assistant using Python Language Integrated with Arduino
Source citation (APA Format)	Akash, S., Jayaram, N., & Jesudoss, A. (2022, May 25-27). Desktop based Smart Voice Assistant using Python Language Integrated with Arduino. <i>2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS)</i> , Madurai, India, pp. 374-379. IEEE. https://doi.org/10.1109/ICICCS53718.2022.9788267
Original URL	https://ieeexplore.ieee.org/document/9788267
Source type	Conference paper
Keywords	Pandemics, Virtual assistants, COVID-19, Market research, Internet of Things, Speech recognition, Artificial intelligence, User interfaces, Python, Natural language processing, Personal voice assistants, Voice Assistant, Speech to Text, Text to Speech, Internet of Things, Python
#Tags	#VoiceRecognition
Summary of key points + notes (include methodology)	<p>assistants help with entertainment and automation they add convenience and allow for multitasking goes through previous works and mentions their negatives such as lacking IoT and lack of complex training "Hello NOVA" is the wake-up word uses speech recognition module needs to: search information tell news relative to the location jokes to build personality opening files and folders in order to assist in navigation tell temperature or weather Use Internet of things in order to communicate with nearby devices</p> <p>Limitations ASR isn't perfect and background noise could interfere security concerns could be used by 3rd person that can make it retrieve information</p> <p>why python is used: simple language that supports object oriented programming growing popularity with ML and language processing plays any video</p>

	<p>random facts plays a game can turn off systems</p> <p>What can it do api keys to get news information from a news API uses OS in order to navigate the computer uses speech recognition module in order to convert audio. There is a threshold for which it recognizes speech Pyttsx3 can respond to users future steps is to make it more useful</p>
Research Question/Problem/ Need	<p>To create a voice assistant that uses IoT to interact with other devices. This is important as it decreases touch during pandemics</p>
Important Figures	
VOCAB: (w/definition)	<p>Internet of Things: a network of physical devices, vehicles, appliances, and other objects that use software and network connectivity to collect and exchange data (https://www.ibm.com/topics/internet-of-things)</p> <p>Object Oriented Programming: a programming paradigm that uses classes and objects to model real world entities and their relationships (https://www.geeksforgeeks.org/introduction-of-object-oriented-programming/)</p>

Cited references to follow up on	Moorthy, A. E., & Vu, K. Yamamoto, S. (eds). (2014). Voice Activated Personal Assistant: Acceptability of Use in the Public Space. <i>Human Interface and the Management of Information. Information and Knowledge in Applications and Services(HIMI 2014)</i> , pp. 8552, 324-334. Springer, Cham. https://doi.org/10.1007/978-3-319-07863-2_32
Follow up Questions	<ol style="list-style-type: none"> 1. How can this be improved? 2. Why does this use Arduino? Can't a software method be implemented? 3. What security issues can come from having an application with access to user files

Article #18 Notes: Voice-Based Virtual-Controlled Intelligent Personal Assistants

Article notes should be on separate sheets

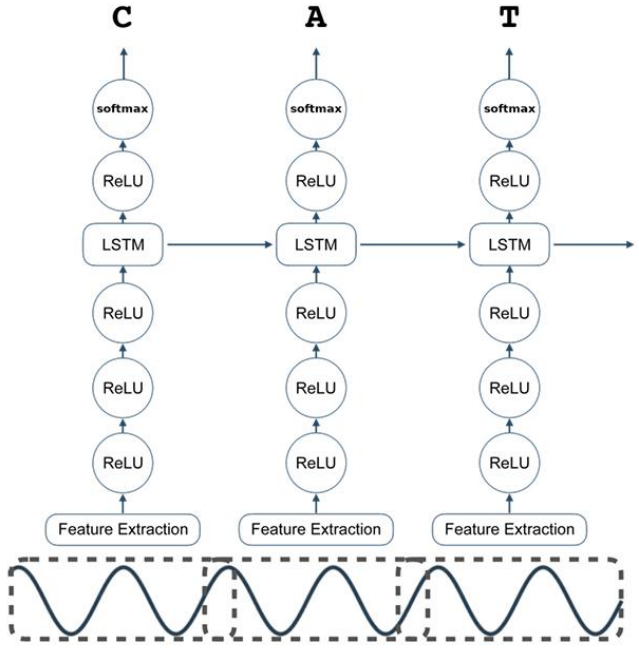
Source Title	Voice-Based Virtual-Controlled Intelligent Personal Assistants
Source citation (APA Format)	Yadav, S. P., Gupta, A., Nascimento, C. D. S., Albuquerque, V. H. C., Naruka, M. S., & Chauhan, S. S. (2023, April 20-21). Voice-based virtual-controlled intelligent personal assistants. <i>2023 International Conference on Computational Intelligence, Communication Technology and Networking (CICTN)</i> , Ghaziabad, India, pp. 563–568. IEEE. https://doi.org/10.1109/CICTN57981.2023.10141447
Original URL	https://ieeexplore.ieee.org/document/10141447
Source type	Conference paper
Keywords	Knowledge engineering, Vocabulary, Virtual assistants, Wearable computers, User-generated content, Speech recognition, Natural language processing, Virtual Assistant, Artificial Intelligence, Smart Home Devices, Personal Assistant, Voice-Based Virtual Assistant.
#Tags	#VoiceRecognition #Assistant
Summary of key points + notes (include methodology)	<p>This conference paper utilizes automatic speech recognition in the same way as other articles do.</p> <p>However, this one uses firebase to store information.</p> <p>It also utilizes news and weather API's in order to get relevant news</p> <p>Uses a wake up word</p> <p>IoT architecture in order to communicate with smart devices</p> <p>Facial recognition in order to make sure that the user is the owner of the device</p> <p>Uses wikipedia api in order to search queries submitted by users</p> <p>This was mainly a voice assistant meant for business use as this article mentions needing safe ways to store information about the user so they won't get sued.</p>
Research Question/Problem/ Need	Need to make an informative assistant that will have facial recognition.

<p>Important Figures</p>	<pre> graph TD Start([Start]) --> WakeUp{Is wake up?} WakeUp -- No --> Start WakeUp -- Yes --> FaceRec[Face Recognition] FaceRec --> ValidUser{Is Valid User?} ValidUser -- No --> Start ValidUser -- Yes --> UserSpeech[User Speech Input] UserSpeech --> SpeechRec[Speech Recognition] SpeechRec --> SkillsExec{Skills to execute?} SkillsExec -- No --> CreateNeg[Create Negative response] SkillsExec -- Yes --> CreatePos[Create Positive response] CreatePos --> ExecuteCmd[Execute Commands] CreateNeg --> BRAIN[BRAIN Console Output & Speech] ExecuteCmd --> BRAIN BRAIN --> End([End]) </pre>
<p>VOCAB: (w/definition)</p>	<p>None</p>
<p>Cited references to follow up on</p>	<p>Hoy, M. B. (2018). Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants. <i>Medical Reference Services Quarterly</i>, 37(1), 81–88. https://doi.org/10.1080/02763869.2018.1404391</p>
<p>Follow up Questions</p>	<ol style="list-style-type: none"> 1. How does IoT work? 2. Why is NLP being used in this project? 3. Which speech recognizer would be best for this program?

Article #19 Notes: CommonVoice: A Massively-Multilingual Speech Corpus

Article notes should be on separate sheets

Source Title	CommonVoice: A Massively-Multilingual Speech Corpus
Source citation (APA Format)	Ardila, R., Branson, M., Davis, K., Henretty, M., Kohler, M., Meyer, J., Morais, R., Saunders, L., Tyers, F. M., & Weber, G. (2020, March 5). <i>Common voice: A massively-multilingual speech corpus</i> . arXiv. https://arxiv.org/abs/1912.06670
Original URL	https://arxiv.org/pdf/1912.06670
Source type	Online source (ArXiv preprint)
Keywords	spoken corpus, Automatic Speech Recognition, low-resource languages
#Tags	#SpeechCorpus
Summary of key points + notes (include methodology)	<p>Mozilla's initiative named Common Voice is a living speech corpus which uses human participants of various linguistic backgrounds in order to contribute to automatic speech recognition research. It uses human voices and uses a voting system in order to determine if data is accurate.</p> <p>The files are released as MPEG-3 files with a 48kHz sampling rate. MPEG-3 is the most universally supported audio format as opposed to lossless compression</p> <p>I assume MPEG means it loses some data in order to save storage for the company</p> <p>Information that is stored about these voices are their accents, the client, their gender, and what was said.</p> <p>https://commonvoice.mozilla.org/en</p> <p>This is the website of the initiative and currently 32000 hours have been recorded but only 21000 have been validated.</p> <p>It also seems to have expanded to over a hundred languages. This means that it has potential to let models have translations through speech to text.</p> <p>I've noticed that the ASR model uses Long Short Term Memory(LSTM) which is great when fixing the errors of recurrent neural networks(forgotten context, and taking a while to process), however, this architecture might be outdated as most automatic speech recognition models use transformers which work very well when it comes to recognizing context, this makes it well suited in the cases of translation as words can affect what other words mean.</p>

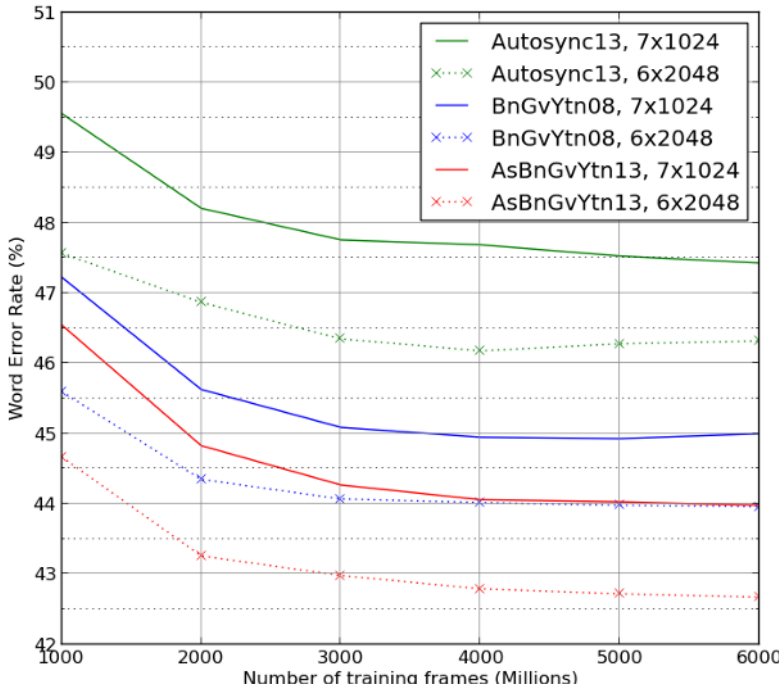
Research Question/Problem/ Need	<p>To create a speech corpus that includes many languages in order to improve automatic speech recognition.</p>
Important Figures	 <p>Figure 4: Architecture of Mozilla's DeepSpeech Automatic Speech Recognition model. A six-layer unidirectional CTC model, with one LSTM layer.</p>
VOCAB: (w/definition)	<p>Long Short Term Memory (LSTM) - architectures are capable of learning long-term dependencies in sequential data, which makes them well-suited for tasks such as language translation, speech recognition, and time series forecasting. https://www.geeksforgeeks.org/deep-learning-introduction-to-long-short-term-memory/</p> <p>Connectionist Temporal Classification(CTC) - a technique that is used with encoder-only transformer models for automatic speech recognition (https://huggingface.co/learn/audio-course/chapter3/ctc)</p> <p>MPEG-3 - a common audio format for consumer audio streaming and storage, and the standard for the transfer and playback of music on most digital audio players https://www.webopedia.com/definitions/mp3/</p>
Cited references to follow up on	<p>None as they usually dive into topics beyond the scope of this project.</p>
Follow up Questions	<ol style="list-style-type: none"> 1. How can you mitigate skewing of the data if many people falsely categorize an audio file?

	<ol style="list-style-type: none">2. How can a performance gap between languages be reduced?3. Is Common Voice meant for recognition of short sentences or long speeches.
--	--

Article #20 Notes: Large scale deep neural network acoustic modeling for YouTube video transcription

Article notes should be on separate sheets

Source Title	Large scale deep neural network acoustic modeling with semi-supervised training data for YouTube video transcription
Source citation (APA Format)	Liao, H., McDermott E., & Senior, A. (2013, December 8-12). Large scale deep neural network acoustic modeling with semi-supervised training data for YouTube video transcription. 2013 IEEE Workshop on Automatic Speech Recognition and Understanding, Olomouc, Czech Republic, pp. 368-373. IEEE. https://doi.org/10.1109/ASRU.2013.6707758
Original URL	https://ieeexplore.ieee.org/document/6707758
Source type	Conference paper
Keywords	Training, YouTube, Acoustics, Approximation methods, Hidden Markov models, Context, Data models, Large vocabulary speech recognition, deep neural networks, deep learning, audio indexing
#Tags	#AcousticModel
Summary of key points + notes (include methodology)	<p>YT vids are hard to transcribe using standard Gaussian Mixture Model (GMM) have word error rates of over 50%</p> <p>Youtube videos aren't homogenous which can make it challenging for a model to transcribe what is stated</p> <p>A lot of a user's uploaded caption can be inaccurate which is why this filters have been put in order to remove inaccurate information</p> <p>Used Islands of confidence method to force align a audio with the transcript and rakes out incorrect word. This filtering gives higher quality data to the acoustic model</p> <p>Deep learning Neural Networks used for ASR training and experimented with many sizes of hidden layers.</p> <p>Using larger state inventories made accuracy better but was expensive</p> <p>tested a new model that had a word error rate of 40.9% instead need to use larger language models and use more supervised data training.</p>
Research Question/Problem/ Need	Engineering need: need to create an acoustic model that is able to minimize Word Error Rate in youtube videos.

Important Figures	 <table><caption>Approximate Word Error Rate (%) data from the graph</caption><thead><tr><th>Number of training frames (Millions)</th><th>Autosync13, 7x1024</th><th>Autosync13, 6x2048</th><th>BnGvYtn08, 7x1024</th><th>BnGvYtn08, 6x2048</th><th>AsBnGvYtn13, 7x1024</th><th>AsBnGvYtn13, 6x2048</th></tr></thead><tbody><tr><td>1000</td><td>49.5</td><td>47.5</td><td>47.2</td><td>45.5</td><td>46.5</td><td>44.5</td></tr><tr><td>2000</td><td>48.2</td><td>46.8</td><td>45.5</td><td>44.5</td><td>44.8</td><td>43.2</td></tr><tr><td>3000</td><td>47.8</td><td>46.3</td><td>45.1</td><td>44.1</td><td>44.3</td><td>43.0</td></tr><tr><td>4000</td><td>47.6</td><td>46.1</td><td>45.0</td><td>44.0</td><td>44.1</td><td>42.8</td></tr><tr><td>5000</td><td>47.5</td><td>46.2</td><td>45.0</td><td>44.0</td><td>44.1</td><td>42.7</td></tr><tr><td>6000</td><td>47.4</td><td>46.2</td><td>45.0</td><td>44.0</td><td>44.1</td><td>42.5</td></tr></tbody></table>	Number of training frames (Millions)	Autosync13, 7x1024	Autosync13, 6x2048	BnGvYtn08, 7x1024	BnGvYtn08, 6x2048	AsBnGvYtn13, 7x1024	AsBnGvYtn13, 6x2048	1000	49.5	47.5	47.2	45.5	46.5	44.5	2000	48.2	46.8	45.5	44.5	44.8	43.2	3000	47.8	46.3	45.1	44.1	44.3	43.0	4000	47.6	46.1	45.0	44.0	44.1	42.8	5000	47.5	46.2	45.0	44.0	44.1	42.7	6000	47.4	46.2	45.0	44.0	44.1	42.5
Number of training frames (Millions)	Autosync13, 7x1024	Autosync13, 6x2048	BnGvYtn08, 7x1024	BnGvYtn08, 6x2048	AsBnGvYtn13, 7x1024	AsBnGvYtn13, 6x2048																																												
1000	49.5	47.5	47.2	45.5	46.5	44.5																																												
2000	48.2	46.8	45.5	44.5	44.8	43.2																																												
3000	47.8	46.3	45.1	44.1	44.3	43.0																																												
4000	47.6	46.1	45.0	44.0	44.1	42.8																																												
5000	47.5	46.2	45.0	44.0	44.1	42.7																																												
6000	47.4	46.2	45.0	44.0	44.1	42.5																																												
VOCAB: (w/definition)	<p>Deep Learning - a subset of machine learning that uses multilayered neural network to simulate the human brain (https://www.ibm.com/topics/deep-learning)</p> <p>Gaussian Mixture Model (GMM) - a robust statistical tool in machine learning and data science, GMMs excel in estimating density and clustering data. (https://medium.com/@juanc.olamendy/understanding-gaussian-mixture-models-a-comprehensive-guide-df30af59ced7)</p>																																																	
Cited references to follow up on	<p>Wessel, F., Schlüter, R., Macherey, K., & Ney, H. (2001) Confidence measures for large vocabulary continuous speech recognition. <i>IEEE Transactions on Speech and Audio Processing</i>, 9(3), 288-298. https://doi.org/10.1109/89.906002</p>																																																	
Follow up Questions	<ol style="list-style-type: none">1. Why is it difficult for GMMs to identify words with high accuracy?2. What other models can be investigated to make a better acoustic model?3. Why are Hidden Markov Models useful?																																																	

Patent #1 Notes: Implementations for voice assistant on devices

Article notes should be on separate sheets

Source Title	Implementations for voice assistant on devices
Source citation (APA Format)	Mixer, K., & Shah, R. (2024). <i>Implementations for voice assistant on devices</i> (U.S. Patent No. 11,922,941B2). U.S. Patent and Trademark Office. https://patents.google.com/patent/US11922941B2/en?q=(Voice+assistant)&oq=Voice+assistant
Original URL	US11922941B2 - Implementations for voice assistant on devices - Google Patents
Source type	Patent
Keywords	Voice Recognition, Assistant, Language Processing
#Tags	#voiceRecognition
Summary of key points + notes (include methodology)	<p>Voice assistants take in input, send it to a remote system and then process the verbal input to respond to it. This recognizes how voice assistants are activated which is often done using push to talk.</p> <p>APIs give interfaces to hardware and other software like operating system, other applications that facilitate voice assistant functionality. In my case I am using Open AI's API for ChatGPT to help answer normal questions.</p> <p>I may also use Whisper as recommended by a classmate since it is accurate with transcription.</p> <p>Memory includes high-speed random access memory, such as DRAM, SRAM, DDR RAM, or other random access solid state memory devices I do not know how I will store any memory as I doubt that I will use it.</p>
Research Question/Problem/ Need	How should voice assistants be formatted?

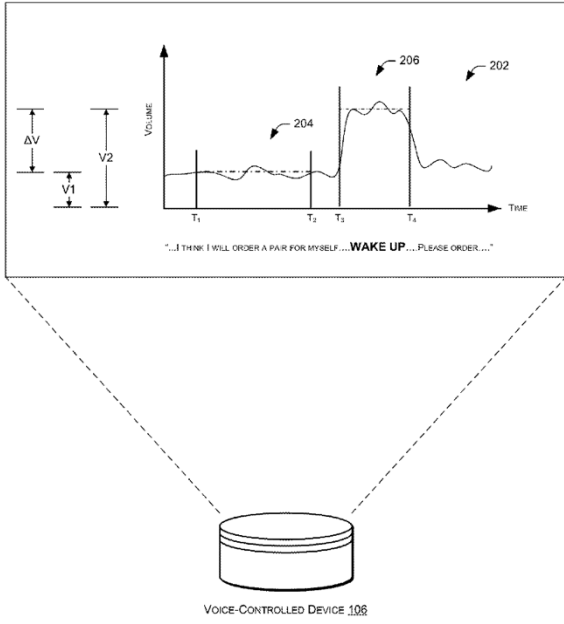
<p>Important Figures</p>	<p>U.S. Patent Mar. 5, 2024 Sheet 5 of 5 US 11,922,941 B2</p> <p>500</p> <p>At an electronic device comprising an audio input system, one or more processors, and memory storing one or more programs for execution by the one or more processors:</p> <pre> graph TD 502[Receive a verbal input at the device] --> 504[Process the verbal input] 504 --> 506[Transmit a request to a remote system, the request including information determined based on the verbal input] 506 --> 508[Receive a response to the request, where the response is generated by the remote system in accordance with the information based on the verbal input] 508 --> 510[Perform an operation in accordance with the response] 510 --> 512[Where one or more of the receiving, processing, transmitting, receiving and performing are performed by one or more voice processing modules of a voice assistant library executing on the electronic device, the voice processing modules providing a plurality of voice processing operations that are accessible to one or more application programs and/or operating software executing or executable on the electronic device] </pre> <p>Figure 5</p>
<p>VOCAB: (w/definition)</p>	<p>Random Access Memory (RAM) - a high-speed, short-term storage solution that gives applications, games, and the operating system itself, quick access to important information.</p> <p>Static RAM(SRAM) - a type of memory chip which is faster and requires less power than dynamic memory</p> <p>Dynamic Random Access Memory (DRAM) - a type of semiconductor memory that is typically used for the data or program code needed by a computer processor to function.</p> <p>DDR - a type of computer memory technology commonly used in personal computers and servers. It allows for faster transfer of data between the computer's memory and the processor</p>
<p>Cited references to follow up on</p>	<p>Loring, K., & Patel, P. (2001). <i>Network universal spoken language vocabulary</i> (U.S. Patent No. 6,195,641B1). U.S. Patent and Trademark Office.</p> <p>https://patents.google.com/patent/US6195641B1/en?q=(Voice+assistant)&oq=Voice+assistant</p>
<p>Follow up Questions</p>	<p>1. What concerns about voice assistants are there?</p>

	<ol style="list-style-type: none">2. What are the specific parts of the Automatic Speech Recogniton system?3. How can I implement an AI that will ignore unnecessary background noise?
--	---

Patent #2 Notes: Voice commands for transitioning between device states

Article notes should be on separate sheets

Source Title	Voice commands for transitioning between device states.
Source citation (APA Format)	Barton, F. W. (2015). <i>Voice commands for transitioning between device states</i> (U.S. Patent No. 9,047,857B1). U.S. Patent and Trademark Office. https://patentimages.storage.googleapis.com/6d/9f/b5/6207989ec4793f/US9047857.pdf
Original URL	https://patents.google.com/patent/US9047857B1
Source type	Patent
Keywords	Voice Recognition, Assistant, Language Processing, Automatic Speech Recognition
#Tags	#voiceRecognition
Summary of key points + notes (include methodology)	<p>This patent discusses how homes now rely on smart devices. It outlines how an assistant should transition between states. It should recognize predefined transition words that will enable it to switch between states. One of the states is constant recording where the device listens in on a wake-up word.</p> <p>The device listens not only for the wake-up word but also the loudness of the word in order to understand the intention. This patent claims that volume analysis should be implemented to better understand intention as sometimes a term can be misinterpreted when used in a normal conversation.</p> <p>The device will constantly be processing the sounds around it and compares the command to the volume of any other words that have been stated to know if the user wanted to use the assistant.</p> <p>If the wake up command was the same volume as all the other noises in the room, then the assistant will be less likely to respond.</p>
Research Question/Problem/ Need	How should assistants transition between states?

<p>Important Figures</p>	 <p style="text-align: center;">Fig. 2</p>
<p>VOCAB: (w/definition)</p>	<p>Hidden Markov model- Hidden Markov Models are close relatives of Markov Chains, but their hidden states make them a unique tool to use when you're interested in determining the probability of a sequence of random variables. (https://towardsdatascience.com/hidden-markov-models-explained-with-a-real-life-example-and-python-code-2df2a7956d65)</p> <p>In this case the HMM is being applied to the probability that the wake-up word was used on purpose in the environmental context.</p> <p>Gaussian mixture model - A Gaussian mixture of three normal distributions. Gaussian mixture models are a probabilistic model for representing normally distributed subpopulations within an overall population. Mixture models in general don't require knowing which subpopulation a data point belongs to, allowing the model to learn the subpopulations automatically. (brilliant.org)</p>
<p>Cited references to follow up on</p>	<p>Hansen, C. H., Shepherd, D. L., & Moncur, R. B. (1997). <i>User independent, real-time speech recognition system and method</i> (U.S. Patent No. 5,640,490A). U.S. Patent and Trademark Office. https://patents.google.com/patent/US5640490A/en</p>
<p>Follow up Questions</p>	<ol style="list-style-type: none"> 1. How can tone be implemented as well so that if a person is in a loud environment and accidentally yells the wake-up word they will not trigger the assistant?

	<ol style="list-style-type: none">2. What roadblocks can come from implementing the volume change?3. What factors lead to a loss of accuracy when it comes to ASR?
--	---