Journal of Engineering Sciences
Vol 12, Issue 7, July/2021
ISSN NO:0377-9254

Quiz through Voice User Interface

Kotesawara Rao A. [1], Sai Jahnavi B.[1], Neelima D.[1], Sarvani K.[1]
[1] Andhra Loyola Institute of Engineering and Technology, ITI Rd, Beside Ramesh Hospital, Jayaprakash Nagar address, 520008, Vijayawada, India
*Corresponding email:neelimadevadi111@gmail.com

Abstract. Speech Recognition is a topic in computer science and Artificial Intelligence with the goal of understanding and comprehending "what" was spoken. Voice User Interfaces (VUIs) are effective and intuitive to improve user interaction with various types of information-technology-based systems. VUIs are increasingly becoming suitable for use in various practical applications, e.g., voice-operated smart phones or smart speakers such as Amazon Echo or Google Home. There are many applications which conduct quiz through Voice User Interface, but in those quiz may or may not be conducted through Voice User Interface. In this system a Voice User Interface is created for specifically for the purpose of conduction of quiz. This system takes voice command from the user as input where speech recognition, speech synthesis are performed and produce result in the form of voice i.e., voice to voice interaction for quiz conduction. Furthermore, the user can interact with the system in addition to attempting quiz. The main goal of Quiz through Voice User Interface is enabling hand free mode. This system deals with conduction of quiz through speech recognition using Natural Language Processing by using javascript programming. The result of this system is that the quiz is always conducted through a Voice User Interface unlike other available applications.

Keywords: voice user interface, natural language processing, speech recognition, speech synthesis, voice command.

1 Introduction

A voice-user interface (VUI) makes spoken human interaction with computers possible, using speech recognition to understand spoken commands and answer questions, and typically text to speech to play a reply. A voice command device (VCD) is a device controlled with a voice user interface.

Voice user interfaces have been added to automobiles, home automation systems, computer operating systems, home appliances like washing machines and microwave ovens, and television remote controls. They are the primary way of interacting with virtual assistants on smartphones and smart speakers. Older automated attendants (which route phone calls to the correct extension) and interactive voice response systems (which conduct more complicated transactions over the phone) can respond to the pressing of keypad buttons via DTMF tones, but those with a full voice user interface allow callers to speak requests and responses without having to press any buttons.

Newer VCDs are speaker-independent, so they can respond to multiple voices, regardless of accent or dialectal influences. They are also capable of responding to several commands at once, separating vocal messages, and providing appropriate feedback, accurately imitating a natural conversation.

Voice user interfaces (VUIs) allow people to use voice input to control computers and devices. VUIs have advantages over typing, such as faster input speed, hands-free use, and intuitive use. Thanks to recent improvements in the accuracy of speech recognition and synthesis through machine learning approaches, VUIs are rapidly becoming suitable for various practical purposes in devices such as voice-operated smartphones and smart speakers like Amazon Echo or Google Home.

Each voice assistant has its own weaknesses, strengths and quirks. The feature of conduction of quiz through voice user interface should be included in VCDs such as Amazon Echo, Google Home. But at the same time the feature of conduction of quiz through voice user interface may or may not be included in voice assistants in mobile phones and computer operating system. The result for voice command of quiz in such assistant is irrelevant from the desired result.

2 Literature Review

Implementation of Voice User Interfaces to Enhance Users' Activities on Moodle

Authors: Toshihiro Kita, Chikako Nagaoka, Naoshi Hiraoka and Martin Dougiamas

Voice user interfaces (VUIs) are effective and intuitive to improve user interaction. If VUIs, which enable hands-free and intuitive use, become available to a learning management system (LMS) such as Moodle, learning activities on the LMS can be made easier and help motivate the users. In addition, if it is possible to search LMS help documentation by speech via a VUI and listen to the search results, the work efficiency of LMS content creators may be improved. In this research, we have developed a voice app for learners to attempt quizzes on Moodle sites, and a voice app for users to search MoodleDocs (Moodle online documents).

Artificial Intelligence-based Voice Assistant

Authors: Subhas S, Prajwal N Srivasta,Siddesh S,Ullas A,santhosh B

Voice control is a major growing feature that change the way people can live. The voice assistant is commonly

being used in smartphones and laptops. AI-based Voice assistants are the operating systems that can recognize human voice and respond via integrated voices. This voice assistant will gather the audio from the microphone and then convert that into text, later it is sent through GTTS (Google text to speech). GTTS engine will convert text into audio file in English language, then that audio is played using play sound package of python programming Language.

A robot quizmaster that can localize, separate, and recognize simultaneous utterances for a fastest-voice-first quiz game

Authors: Izaya Nishimuta, Naoki Hirayama, Kazuyoshi Yoshii, Katsutoshi Itoyama and Hiroshi G. Okuno

This paper presents an interactive humanoid robot that can moderate a multi-player fastest-voice-first-type quiz game by leveraging state-of-the-art robot audition techniques such as sound source localization and separation and speech recognition. In this game, a player who says "Yes" first gets a right to answer a question, and players are allowed to barge in a questionary utterance of the quizmaster. The robot needs to identify which player says "Yes" first, even if multiple players respond at almost exactly the same time, and must judge the correctness of the answer given by the player. To enable natural human-robot interaction, we believe that the robot should use its own microphones (i.e., ears) embedded in the head, rather than having pin microphones attached to individual players. In this paper we use a robot audition system called HARK for separating the mixture of audio signals recorded by the ears into multiple source signals (i.e., almost the simultaneous utterances of "Yes" and the questionary utterance) and estimating the direction of each source. To judge the correctness of an answer, we use a speech recognizer called Julius. Experimental results showed that our robot can correctly identify which player spoke first when the players' utterances differed by 60 msec.

3 Research Methodology

3.1 Natural Language Processing

Natural language processing (NLP) refers to the branch of computer science—and more specifically, the branch of artificial intelligence or AI—concerned with giving computers the ability to understand text and spoken words in much the same way human beings can.

NLP combines computational linguistics—rule-based modeling of human language—with statistical, machine learning, and deep learning models. Together, these technologies enable computers to process human language in the form of text or voice data and to 'understand' its full meaning, complete with the speaker or writer's intent and sentiment.

NLP drives computer programs that translate text from one language to another, respond to spoken commands, and summarize large volumes of text rapidly—even in real time. There's a good chance you've interacted with NLP in the form of voice-operated GPS systems, digital assistants, speech-to-text dictation software, customer service chatbots, and other consumer conveniences. But NLP also plays a growing role in enterprise solutions that help streamline business operations, increase employee productivity, and simplify mission-critical business processes.

3.2 Hidden Markov Algorithm

Hidden Markov model (HMM) is the base of a set of successful techniques for acoustic modeling in speech recognition systems. The main reasons for this success are due to this model's analytic ability in the speech phenomenon and its accuracy in practical speech recognition systems. Another major specification of HMM is its convergent and reliable parameter training procedure. Spoken utterances are represented as a non-stationary sequence of feature vectors. Therefore, to evaluate a speech sequence statistically, it is required to segment the speech sequence into stationary states.

In a typical HMM, the speech signal is divided into 10-millisecond fragments. The power spectrum of each fragment, which is essentially a plot of the signal's power as a function of frequency, is mapped to a vector of real numbers known as cepstral coefficients. The dimension of this vector is usually small—sometimes as low as 10, although more accurate systems may have dimension 32 or more. The final output of the HMM is a sequence of these vectors.

3.3 Speech Recognition

Speech recognition, also known as automatic speech recognition (ASR), computer speech recognition, or speech-to-text, is a capability which enables a program to process human speech into a written format. While it's commonly confused with voice recognition, speech recognition focuses on the translation of speech from a verbal format to a text one whereas voice recognition just seeks to identify an individual user's voice.

3.4 Speech Synthesis

Speech synthesis is artificial simulation of human speech with by a computer or other device. The counterpart of the voice recognition, speech synthesis is mostly used for translating text information into audio information and in applications such as voice-enabled services and mobile applications. Apart from this, it is also used in assistive technology for helping vision-impaired individuals in reading text content.

3.5 Model development

With Web Speech API, which provide SpeechRecognition interface and SpeechSynthesis interface, a Voice User Interface is developed for quiz using javascript programming language. Recognition class in SpeechRecognition is where all the processing takes place. The primary purpose of a Recognition instance is, of course, to recognize speech. Each instance comes with a variety of settings and functionality for recognizing speech from an audio source. Speak class in SpeechSynthesis is where all the processing of text to speech occurs.

The different methods that are available in SpeechRecognition and SpeechSynthesis are used for the conversion of speech to text and text to speech. Generally for this an action is performed for the conversions but in this process model we included a processor (voice assistant) which automatically performs the action required for conversion from speech to text and text to speech. Due to this along with attempting a quiz with voice user interface, interaction with the voice assistant is possible to some extent.

4 Results

The result of this system is quiz through voice user interface (Fig. 1).

Figure 1 – Quiz through voice user interface

Along with the conduction of quiz, interaction with the voice assistant can also be done (Fig. 2).

Figure 2 – Interaction with voice user interface

In this system quiz can be attempted in two ways. The first way is through the user interface that is implemented int the system. The second way is through the interaction of the system. In both ways quiz can be attempted through voice commands.

5 Conclusions

In this research, we developed a system for the conduction of quiz through voice user interface. Along with it, a voice assistant is also developed through which the user can interact with the system. While implementing system with a voice user interface consideration of its advantages while being aware of its limitations are to be considered.

The future scope of this system is include more dynamic features for the interaction with the system.

References

1. Toshihiro Kita, Chikako Nagaoka, Naoshi Hiraoka and Martin Dougiamas. (2019). Implementation of Voice User Interfaces to Enhance Users' Activities on Moodle. *4th International Conference on Information Technology (InCIT)*. Bangkok, Thailand, doi: 10.1109/INCIT.2019.8912086.
2. S Subhash, Prajwal N Srivatsa, S Siddesh, A Ullas, B Santhosh. (2020). Artificial Intelligence-based Voice Assistant. *Fourth World Conference on Smart Trends in Systems, Security and Sustainability(WorldS4)*. IEEE. London, UK, doi: 10.1109/WorldS450073.2020.9210344.
3. Hyo-Jung Oh, Chung-Hee Lee, Yi-Gyu Hwang, Myung-Gil Jang, Jeon Gue Park and Yun Kun Lee. (2007). A Case study of Edutainment Robot: Applying Voice Question Answer to Intelligent Robot. *16th IEEE International Symposium on Robot and Human Interactive Communication*. Jeju, Korea (South), doi : 10.1109/ROMAN.2007.4415119.
4. Tae-Kook Kim. (2020). Short Research on Voice Control System Based on Artificial Intelligence Assistant. *International Conference on Electronics, Information, and Communication (ICEIC)*. Barcelona, Spain, doi: 10.1109/ICEIC49074.2020.9051160.
5. Jon Sticklen, Mark Urban-Lurain and Daina Briedis. (2008). Work in progress - effective engagement of millennial students using web-based voice-over slides and screen demos to augment traditional class delivery. *38th Annual Frontiers in Education Conference*. Saratoga Springs, NY, USA, doi: 10.1109/FIE.2008.4720397.

6. Xian-Jun Xia., Zhen-Hua Ling., Chen-Yu Yang., Li-Rong Dai. (2018). Improved unit selection speech synthesis method utilizing subjective evaluation results on synthetic speech. *2012 8th International Symposium on Chinese Spoken Language Processing*. Hong Kong, China, doi: 10.1109/ISCSLP.2012.6423524.
7. Aswin Shanmugam Subramanian, Chao Weng, Meng Yu, Shi-Xiong Zhang; Yong Xu, Shinji Watanabe, Dong Yu. (2020). Far-Field Location Guided Target Speech Extraction Using End-to-End Speech Recognition Objectives. *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP*). Barcelona, Spain, doi:10.1109/ICASSP40776.2020.9053692.
8. Kita,T., Nagaoka,C., Hiraoka,N., Suzuki,K., and Dougiamas, M. (2018). A Discussion on Effective Implementation and Prototyping of Voice User Interfaces for Learning Activities on Moodle. *10th International Conference on Computer Supported Education*. SciTePress, Volume 2: CSEDU. Funchal, Madeira, Portugal, doi:10.5220/0006782603980404
9. Andrei Stoica., Ken Q. Pu., Heidar Davoudi. (2020). NLP Relational Queries and Its Application. *2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI).* Las Vegas, NV, USA, doi:10.1109/IRI49571.2020.00064
10. Kei Hashimoto., Junichi Yamagishi., William Byrne., Simon King., Keiichi Tokuda. (2011). An analysis of machine translation and speech synthesis in speech-to-speech translation system. *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).* Prague, Czech Republic, doi:10.1109/ICASSP.2011.5947506.
11. Heng Lu., Zhen-Hua Ling., Li-Rong Da., Ren-Hua Wang. (2011). Building HMM based unit-selection speech synthesis system using synthetic speech naturalness evaluation score. *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).* Prague, Czech Republic, doi:10.1109/ICASSP.2011.5947567
12. A. Acero. (2000). An overview of text-to-speech synthesis. 2000 IEEE Workshop on Speech Coding. Proceedings. Meeting the Challenges of the New Millennium (Cat. No.00EX421). Delavan, WI, USA, doi:10.1109/SCFT.2000.878372.
13. A. Stolcke., E. Shriberg. (2001). Markovian combination of language and prosodic models for better speech understanding and recognition. *IEEE Workshop on Automatic Speech Recognition and Understanding, 2001. ASRU '01*. Madonna di Campiglio, Italy, doi:10.1109/ASRU.2001.1034615
14. Sharefah A., Al-Ghamdi., Joharah Khabti., Hend S. Al-Khalifa. (2017). Exploring NLP web APIs for building Arabic systems. *2017 Twelfth International Conference on Digital Information Management (ICDIM).* Fukuoka, Japan, doi: 10.1109/ICDIM.2017.8244649.
15. AP Gopi (2021), Secure Communication in Internet of Things Based on Packet Analysis, Machine Intelligence and Soft Computing, 2021.
16. A Naresh, PG Arepalli (2021), Traffic Analysis Using IoT for Improving Secured Communication, Innovations in the Industrial Internet of Things (IIoT), 2021.
17. Bharathi C R, (2017),"Identity Based Cryptography for Mobile ad hoc Networks", Journal of Theoretical and Applied Information Technology, Vol.95, Issue.5, pp.1173-1181.
18. Paul Jasmin Rani., Jason Bakthakumar., B. Praveen Kumaar., U. Praveen Kumaar., Santhosh Kumar. (2017). Voice controlled home automation system using Natural Language Processing (NLP) and Internet of Things (IoT). *2017 Third International Conference on Science Technology Engineering & Management (ICONSTEM).* Chennai, India, doi: 10.1109/ICONSTEM.2017.8261311.
19. Yasir Ali Solangi, Zulfiqar Ali Solangi, Samreen Aarain, Amna Abro, Ghulam Ali Mallah, Asadullah Shah. (2018). Review on Natural Language Processing (NLP) and Its Toolkits for Opinion Mining and Sentiment Analysis. *IEEE 5th International Conference on Engineering Technologies and Applied Sciences (ICETAS).* Bangkok, Thailand, doi: 10.1109/ICETAS.2018.8629198.