Name: Brennon Farmer

Name of Project Advisor: X. Ma

Group Name: Myr R

Descriptive Title: Voice Recognition and Control Software for Smart Mirrors

Introduction

Smart mirrors allow a user to perform various tasks such as: view weather reports, play music, check the time, review a calendar, etc., while looking at the user's reflection. To allow users to navigate through the smart mirror's tasks (without touching the mirror), robust and responsive software must be applied. Voice control software is one of the common answers for this design problem. Many smart devices such as smart phones, computers, and televisions use voice control to navigate through apps, menus and options. This technical review will summarize some commercial voice control software used for smart mirrors and other smart devices, explain the underlying technology of the software, and provide the methods for implementing the technology.

Commercial Applications of Voice Control Software

Smart Mirror

Amazon's Alexa is the most common voice control software implemented in smart mirrors. The line of smart mirrors created by Seura, "Seura Mirrors +Alexa", are one brand of mirrors that use Amazon Alexa [1]. Integration of Amazon Alexa and Amazon Voice Service with any device is free and open source. Amazon Voice Service allows developers to create new voice phrases (custom skills) which cause Alexa to perform an action [2]. The custom skills can be created in Node.js, Java, Python, C#, or Go. Another voice control software is Google Assistant [3]. The Google Assistant SDK (Software Development Kit) is free and open source. Custom actions (user interface navigation tasks) can be developed. This allows new voice-controlled commands to be added to the Google Assistance interface [4].

Other Smart Devices

Dragon speech recognition is a voice control software. The software uses deep learning algorithms to perform learning on the acoustic models produced from the user. This allows the software to learn the speech patterns of the user. The software has up to 99% accuracy at $300 (Dragon Professional Individual) [5]. Another voice control software is Braina Pro. The software uses AI and voice recognition to allow the use of voice commands. It is also capable of converting speech to text. Like

Google Assistant and Amazon Alexa, custom voice commands can be created by the user. Unlike Dragon, no learning is used to recognize speech patterns, instead, the cognitive behaviors of the AI are responsible for recognizing speech patterns. The product costs $239 [6].

Underlying Technology of Voice Control Software

Preprocessing

An input signal is read from the microphone of a device [7]. Next, the signal is preprocessed to separate the voiced portion of the signal from the non-voiced or silent portion [8]. The short-time energy method is usually applied for preprocessing. The method compares the energy in the window of the inout signal to a threshold. If the energy of the window is above the threshold, the sample of the windows are kept; otherwise, the samples are dismissed. This process works since the energy of the voiced portion of the signal is higher than the silent/unvoiced portion [7].

Feature Extraction

Next, the resulting signal's feature is extracted. The most common feature extraction method is Mel Frequency Cepstral Coefficient method. The method requires framing the signal into short frames. Next, the spectral density of the power spectrum is to be calculated for each frame [7]. After, the Mel filter bank is applied to the power spectrums, Next, the energy in each filter is summed. Afterwards, the log of all filter bank energies is taken. Next, the Discrete Cosine Transform (DCT) is used on the log filter bank energies. If the DCT coefficients are not at least two or higher than 13, the coefficients are dismissed [8].

Speaker Modeling

The features extracted from the input signal are used to train a classifier. A trained classifier enables the classification of words spoken by the user [7]. One popular classifier is the Hidden Markov Model. The model uses machine learning and discrete states to estimate the word spoken by the user. By giving the model a sequence of MFCCs, the algorithm of the model can combine the sounds (from the MFCCs) to generate the possible words of the user. With states, the model can increase the likelihood of generating the correct sequence of letters, since the sequence of letters are various states [7].

Implementation of Voice Control Software on Smart Mirror

To implement voice control software, hardware is required. Speakers must be implemented to allow the voice control software to recognize voices. For some software, a certain Operating System or functionality must be installed on the system. For example, the Briana software requires Windows OS [6].

Another example is Amazon Alexa. Internet access is required to use Amazon Alexa and the custom voice commands [3]. The developer that is planning to use voice control must be wary of the hardware that is used. For small smart devices, the hardware will consist of a SoC (System on Chip). For example, a typical SoC is a Raspberry Pi. The maximum RAM that this SoC has is only 4 GB (Rasberry Pi 4) [9]. The developer must know how much RAM an Operating System requires, if a certain OS will be used to support voice control software. Also, the connectivity of the Wi Fi support must be known if Internet required software (Amazon Alexa) is to be applied. If the Wi FI connectivity is below par, the responsiveness of the smart device will decrease.

Reference List

[1] Seura.com. (2019). Séura Voice Mirrors. [online] Available at: https://www.seura.com/products/smart/alexa/hotel/ [Accessed 29 Sep. 2019].

[2] J. Vanian, "This Is Amazon's Latest Effort to Get Alexa Into More Devices", Fortune, 2019. [Online]. Available: https://fortune.com/2017/08/17/amazon-alexa-devices-software-kit/. [Accessed: 30- Sep- 2019].

[3] "Custom Conversational Actions | Actions on Google | Google Developers", Google Developers, 2019. [Online]. Available: https://developers.google.com/actions/conversational/overview. [Accessed: 30- Sep- 2019].

[4] "Overview | Google Assistant SDK | Google Developers", Google Developers, 2019. [Online]. Available: https://developers.google.com/assistant/sdk/overview. [Accessed: 30- Sep- 2019].

[5] *Nuance.com*, 2019. [Online]. Available: https://www.nuance.com/content/dam/nuance/en_us/collateral/dragon/feature-matrix/fm-dragon-naturally-speaking-en-us.pdf. [Accessed: 30- Sep- 2019]. "Braina - Artificial Intelligence Software for Windows", *Brainasoft.com*, 2019. [Online]. Available: https://www.brainasoft.com/braina/. [Accessed: 30- Sep- 2019].

[6] N. M., "Voice Recognition Technique: A Review", *International Journal for Research in Applied Science & Engineering Technology*, vol. 5, no., pp. 262-268, 2017. Available: https://www.ijraset.com/fileserve.php?FID=7636. [Accessed 30 September 2019].

[7] "Practical Cryptography", *Practicalcryptography.com*, 2019. [Online]. Available: http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/. [Accessed: 30- Sep- 2019].

[8] Raspberry Pi (Trading) Ltd., “Raspberry Pi 4 Model B,” Raspberry Pi 4 Model B Datasheet, Jun. 2019 [Revised June 2019].