Jarvis Emulator  
Concept of Operations  
COP 4331, Fall 2015

**Modification History**

|  |  |  |  |
| --- | --- | --- | --- |
| **Version** | **Date** | **Who** | **Comment** |
| v0.0 | 9/14/2015 | Robin Schiro | Created document |
| v1.0 | 9/15/15 | Robin Schiro | Updated details for ‘Current System’ |
| V2.0 | 9/15/2015 | Jimmy Lam | Added ‘Needs’ and ‘Impacts’ |
| V3.0 | 9/16/2015 | Jimmy Lam | Edited ‘Needs’ and ‘Impacts’ |
| V3.1 | 9/16/2016 | Manuel Gonzalez | Added ‘Users and Modes of Operation’ and ‘Analysis’ |

**Team Members:**

* Jimmy Lam
* Julian Rojas
* Manuel Gonzalez
* Robin Schiro

1. **Current System**
   1. Currently, most PC users cannot identify themselves and interact with their computers using a system of facial recognition. In an environment in which one computer is shared between multiple people, a user might have to enter a password to log on. After that, he must focus on his monitor to open desired applications and visit websites that interest him. These activities require the user to physically input information into the computer.
   2. Moreover, most desktop sessions do not allow multiple users to interact with the computer in the same setting. If the computer is “occupied” by someone, another person must wait his turn before taking advantage of the computer’s resources.
2. **The Proposed System: Needs**
   1. Firstly, our project needs the Windows operating system, as we will be coding a Windows desktop application. Our project requires a webcam for facial detection and recognition, as well as a microphone for users to perform voice commands. A language database is needed for ‘Jarvis’ to recognize the voice commands and construct sentences to respond to the user. We will need a front end GUI for the user to see when ‘Jarvis’ is speaking to improve the interaction level between the user and ‘Jarvis’. The GUI will also be used to allow users to create a profile and configure settings with ‘Jarvis’ (for example, whether ‘Jarvis’ will log in when the user enters the room). Finally, we will need web APIs for ‘Jarvis’ to report information that is relevant to the user (based on user input in the GUI and based on which user ‘Jarvis’ detects).
3. **The Proposed System: Users and Modes of Operation**

The system will handle only one type of user, no specific roles will be present. Each user will be authenticated through facial recognition, this will allow the application to be customized depending on which user is interacting with it. The system should be able to handle multiple users at once.

The system will perform a set of actions reacting to the user's behavior (leaving the room, speaking, not looking to the camera, moving, etc.), as well as executing commands detected through speech. Some actions may not be available for some users, depending on their profiles.

1. **The Proposed System: Operational Scenarios**
   1. Normal Scenarios:
      1. The user comes within the field of view of the camera; the application is able to detect a face. If the application has not previously been exposed to the users face, it will prompt them to complete a profile for it to be able to recognize them. Otherwise, it will greet them by their name.
      2. When a user is interacting with the JARVIS, it will be able to follow commands. For example, if the user starts a sentence with a keyword, it will take in a command that has already been established. The user says “JARVIS, open Microsoft Word” and the application will then answer “Opening Microsoft Word” as the new application loads.
      3. JARVIS will be able to detect multiple faces in the same frame, and it will be able to recognize them.
   2. Atypical Scenarios
      1. The most common one would be that the application would not be able to detect faces depending on the lighting or conditions of the room.
      2. Too many faces in the frame may cause the program to crash or confuse people’s profiles.
      3. Not understanding the commands given by the user or doing something completely different from what the user requested.
      4. A user that has previously interacted with the application comes into the field of view, and the application is not able to recognize him, therefore a second profile is created. Duplicate information leads to waste in computer’s memory.
2. **The Proposed System: Operational Features**
   1. MUST HAVE:
      1. The ability to recognize, distinguish and track people’s faces
      2. A GUI that allows the user to manually configure their own profiles, such as preferences, most used applications, most visited websites
      3. It must be able to perform speech recognition and construction
      4. The ability to report information from websites (such as weather)
      5. The ability to open and close applications, and log in or out
   2. WOULD LIKE TO HAVE:
      1. The ability to analyze mouth movement from users to be able to determine whom it is interacting with
      2. An animated face to talk to the user
      3. Recognition of emotions and gestures
3. **The Proposed System: Expected Impacts**
   1. **Enhanced Desktop Experience, Better Personalization, More Productivity**

When using the ‘Jarvis Emulator,’ users will have an enhanced desktop experience. Users will be able to open and close applications, and sign in and out of their computer simply through voice command and facial recognition. Not only that, ‘Jarvis’ will provide a personal experience with users as it will be able to detect who is talking and, if their faces are stored in ‘Jarvis’s’ memory, address that person by name. If the face is not stored in memory, the new user can easily create an account with ‘Jarvis’, where it will change its behavior based on who is using it. ‘Jarvis’ will also help to improve user productivity; instead of users searching for weather or news, users can tell ‘Jarvis’ which websites they want information from, and ‘Jarvis’ will automatically update the users with all the information relevant to the users, saving them time to focus on their other daily needs.

1. **The Proposed System: Analysis**

**Expected Improvements:**

* The system should seamlessly integrate facial recognition, speech recognition and artificial intelligence.
* The system should feel like it “knows” the user.
* The user should be able to use our system without changing any of his/her usual routine.

**Disadvantages:**

* The user may feel uncomfortable of having a camera constantly watching what he/she is doing.

**Limitations:**

* It requires camera and a microphone to function.
* Only some online information will be available through the application's speech construction due to the difference in web API formats
* Due to the limited amount of information known about the user, the system may not be able to accurately predict some of the user's needs.

**Risks:**

* The current algorithm for facial recognition has some problems detecting specific users. We may not be able to fix such errors during our short development period.
* Many of the algorithms and techniques that we will be using are a bit out of the scope for undergraduate studies, and could be difficult to implement.

**Alternatives and Tradeoffs:**

* All algorithms that cannot be enhanced will be discussed for further research.

Mainly, the Web APIs our system should support. Time constraints prevents us from having an extensive list of supported websites for the application to interpret and present to the user, so we will have to choose whether or not to implement specific websites. However, a system can be designed for add-ons and plug-ins that could increase the amount of websites supported.