Sonal Raj - CS110487

Shashi Kant - CS110462

Spock: Advanced Human Computer Interaction System

BTP Project Illustration Keynote

Objective

What goals will Spock achieve ?

Objectives

- Obtain a viable knowledge of the working of existing algorithms and methods of interaction, and build up a concept system to improve these techniques.
- Code a Human Computer Interaction Engine capable to techniques like motion and gesture recognition, speech synthesis (biometric authentications, iris control and many other features could be included in future versions and are beyond the scope of this project.).
- Implement an Application Programming Interface (API) to extend the functionality of this engine to multitude of applications running on the system to implement it in their own way.

Motivation

Why on earth did we want to do this?

Motivations

- Definitely **not** marks or Compulsion ! ☺
- Inspired by brilliant pieces of work like "The Sixth Sense" by Pranav Mistry
- Vision to overcome the limitations of existing systems like Microsoft Kinect and Nintendo.
- Sci-fi movies like Star Trek or the Jarvis interactive computer systems in the Iron Man movie series

Motivations

Pranav Mistry's work at MIT on Sixth Sense can be checked out at: http://www.pranavmistry.com/projects/sixthsense/

Iron Man's J.A.R.V.I.S computer system can be referenced at : http://marvel-movies.wikia.com/wiki/J.A.R.V.I.S.

Mission Statement

"This Project's mission is to develop a technological product which will change the way you interacted with your computers using life-like sensing methods."

Evolution

Natural User Interface (NUI)

Existing Work Survey

What all led to this?

Projects

- “Sixth Sense”
  - Pranav Mistry
  - MIT Labs

Projects

“Sixth Sense”

Pranav Mistry

MIT Labs

Projects

“Sixth Sense”

Pranav Mistry

MIT Labs

CMUSphinx toolkit is a leading speech recognition toolkit with various tools used to build speech applications.

CMUSphinx
Open Source Toolkit For Speech Recognition
Project by Carnegie Mellon University

Kinect is a line of motion sensing input devices by Microsoft for Xbox 360 and Xbox One video game consoles and Windows PCs.

SIRI is a voice assistant for iOS. Uses natural speech synthesis tools.

hello

“Sphinx”
Carnegie Mellon University

“Microsoft Kinect”
MS R&D

“SIRI for IOS”
Apple Inc.

Papers and Publications

1. **Real-time hand gesture recognition using range cameras,** *Hervé Lahamy and Derek Litchi,* Department of Geomatics Engineering, University of Calgary, NW, Calgary, Alberta
2. **Real-Time Human Pose Recognition in Parts from Single Depth Images,** *Jamie Shotton Andrew Fitzgibbon,* Microsoft Research Cambridge & Xbox Incubation.
3. **Minimum variance modulation filter for robust speech recognition**. *Yu-Hsiang Bosco Chiu and Richard M Stern,* Carnegie Mellon University, Pittsburgh, USA.
4. **Some recent research work at LIUM based on the use of CMU Sphinx**, *Yannick Estève ET. Al.* LIUM, University of Le Mans, France

hello

Accomplished Work

What have we tried to do?

Tech to be Used

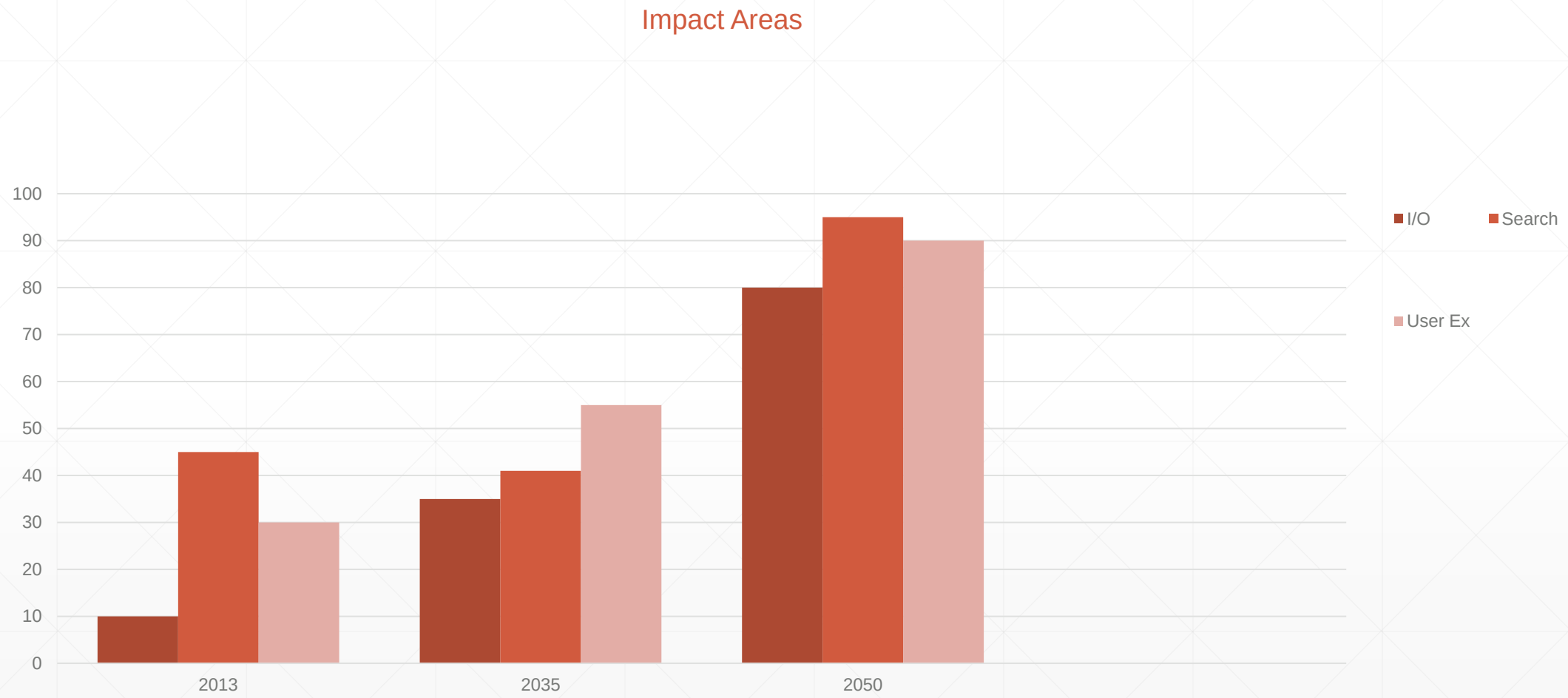
- **Languages**
  1. C/C++
  2. Python
  3. Bash
  4. JavaScript

- **Frameworks**
  - OpenCV
  - CMUSphinx
  - Emscripten

Impact Areas

- Lively Computing experience to “muggle” computer users.
- Robust and Capable devices.
- Advanced Search
- Security feed analysis
- And many more . . .

Impact prediction



Future Work and Applications

- Boon to physically challenged disabled users. The blind can now command a computer with voice and gesture with accurate dictation tools within Spock to control a system, rather than typing on Braille keyboards.
- Used in Embedded devices for robot and appliance control
- take search to the next level. You can not only search for your keywords within a video or audio file, but with powerful system architecture, you can do it in near real-time
- In combination of suitable machine Learning and AI Techniques, it can be used in Defense and Military operations for monitoring cross-border conversations or surveillance data in an automated and much more efficient way compared to manned analysis.

Progress so far.

Where have we reached?

Progress

- **Technologies learnt and configured**
- Studied documentations of the OpenCV Framework and the Sphinx Speech toolkit from CMU
- Fiddled around with configurations quite a bit. Thereby, their configurations in Linux as well as Windows was carried out and learnt successfully.
- Interface implementations of the above frameworks.

Progress

- **Coding Work:** Noise Reduction and image synthesis: [**Algorithm**] This was done by
- subtracting the RGB value of the pixels of the previous frame from the RGB values of the pixels of the current frame.
- Then this image is converted to octachrome (8 colors only - red, blue, green, cyan, magenta, yellow, white, black). This makes most pixels neutral or grey.
- This is followed by the greying of those pixels not surrounded by 20 non-grey pixels, in the function crosshair (IplImage*img1, IplImage* img2). The non-grey pixels that remain represent proper motion and noise is eliminated. A database is provided with the code which contains a set of points for each gesture.

Progress

- **Coding Work:** Speech synthesis:
- Split the waveform by on utterances by silences
- Algorithm to match all possible combination of words with the audio models used in speech recognition acoustic model phonetic dictionary Language Model.
- We have developed a system command acoustic model for commanding a Linux system which adapts commands to recognized speech
- Other concepts used in speech recognition are : Lattice, N-best lists, word confusion networks, speech database, text databases are currently being implemented.

Progress

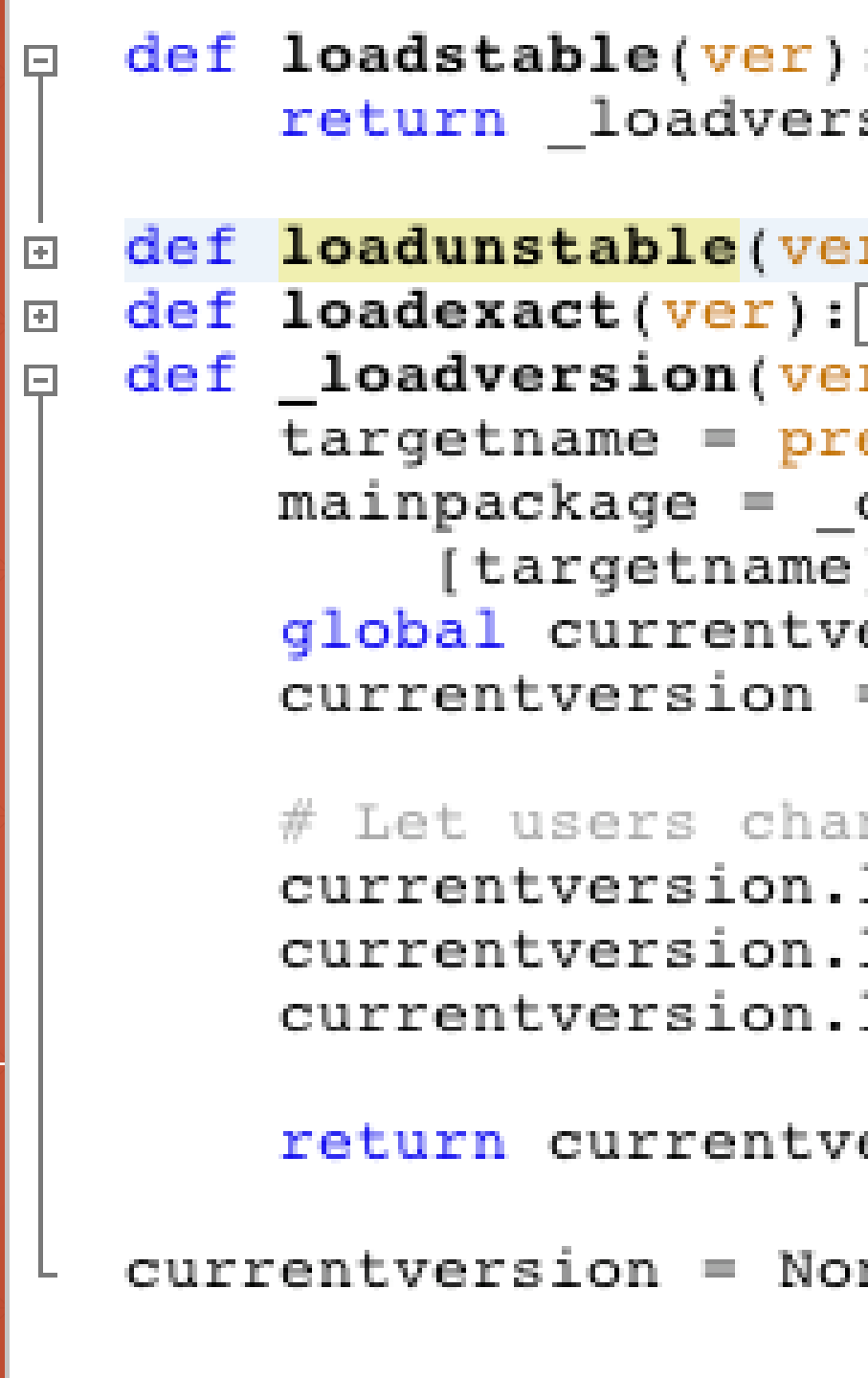
- **Coding Work:** Speech synthesis:
- Finally we developed an integration module to be included in any UNIX like system (meeting the system requirements criteria) so that

Progress

- **Coding Work:** API Generation
- We have used node.JS to integrate OpenCV modules into a web framework, so that it can be utilized easily on live web applications to perform operations from Spock.
- We have used Emscripten to create a JavaScript version of the PocketSphinx framework to harness the power of speech to control modern day web applications.

Code Glimpses

Lets take a peek at the real stuff ☺



Project Timeline

What and by when?

Timeline

Date	Work completed
December 2013	Study and Algorithmic Design and Implementation
January 2014	System Integration Modules prepared
February 2014	Unit Testing and deployments
March 2014	Basic API Specifications prepared
April-May 2014	Documentation Complete and report submission.

References

Who helped?

References

Projects

1. Sixth Sense, Pranav Mistry, MIT Labs
2. Microsoft Kinect, Microsoft Research and XBox
3. CMUSphinx, Carnegie Mellon university
4. SIRI and Google Voice Search, Apple and Google, Inc.

Papers and Publications.

1. **Real-time hand gesture recognition using range cameras,** *Hervé Lahamy and Derek Litchi,* Department of Geomatics Engineering, University of Calgary, NW, Calgary, Alberta
2. **Real-Time Human Pose Recognition in Parts from Single Depth Images,** *Jamie Shotton Andrew Fitzgibbon,* Microsoft Research Cambridge & Xbox Incubation.
3. **Minimum variance modulation filter for robust speech recognition**. *Yu-Hsiang Bosco Chiu and Richard M Stern,* Carnegie Mellon University, Pittsburgh, USA.
4. **Some recent research work at LIUM based on the use of CMU Sphinx**, *Yannick Estève ET. Al.* LIUM, University of Le Mans, France

Monitor the Progress of Spock and view the code at:
https://github.com/sonal-raj/Spock

Thank You

For your patience ☺