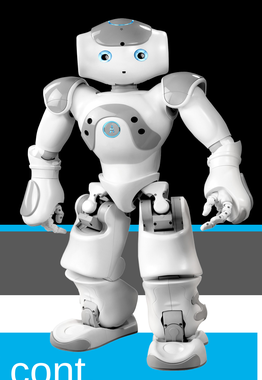
Speech based Control of Articulated Robots

Rahul Gurjar(12532), Rohun Tripati(11619)
Supervised By : Prof. Amitabh Mukherjee



INTRODUCTION

- Robots have entered every sphere of our life. Speech based control of Robots plays a vital role towards its advancement as it allows non-robotics-experts to interact with robots.
- Google Speech Processing takes as input audio file and predicts the text spoken.
- Stanford NLP parser generates parse trees from input text used for Semantic Role Labeling and the semantics roles are assigned by Illinois SRL - “The Curator”
- The FrameNet corpus is a English lexical database based on the concept of Semantic Frames. A Frame is a conceptual structure that comprises a situation along with all the elements in that situation.
- The Nao is a programmable robot provided by Aldebaran Robotics. The Nao simulator uses WeRobot and Choreographe.

RELATED WORK AND MOTIVATION

- Research in robotics systems has produced frameworks for robot middleware such as ROS, as developed by Quigley et al., which has been used in several domains of modern robotics research.
- The RoboFrameNet framework describes a system that uses natural language to command robot action through the intermediary of semantic frames. This module was implemented in ROS-fuerte but the package is retired.
- Automatic semantic role labeling was pioneered by Gildea and Jurafsky in 2002, and semantic role labeling was treated as a tagging problem on each constituent in a parse tree, solved using an argument identifier and an argument classifier.
- We intend to implement an end-to-end system, from speech recognition to action implementation and in the process extend Nao's action set.

METHODOLGY

Module 1 : Google ASR

- HTTP Post request with audio input is sent to the Google web speech API in flac format, which returns transcription with a confidence value for the same.

```
['{"result":[{"alternative":[{"transcript":"robot pick the ball","confidence":0.
85152686}],"final":true}],"result_index":0}']
```

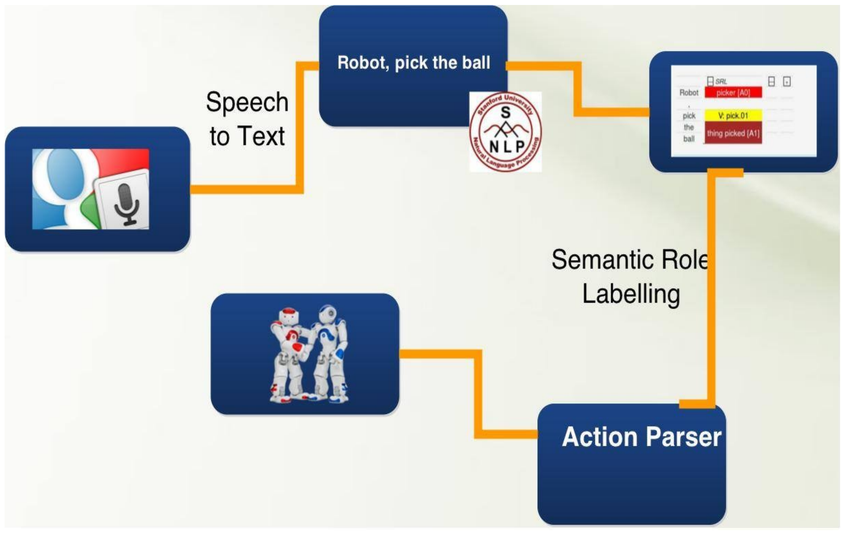
Module 2 : Stanford NLP Parser

- The output obtained from the Speech API is run through the Parser for generating the parse tree. The NLP Parser is based on a Probabilistic Context Free Grammar.

METHODOLOGY cont.

- Probabilistic parsers use knowledge of language gained from hand-parsed sentences to try to produce the most likely analysis of new sentences

```
(ROOT
  (S
    (VP (VB Pick)
      (VP
        (ADVP (RB up)
          (NP (DT the) (NN spoon))
          (PP (IN from)
            (NP (DT the) (NN table))))
```



Modules 3 & 4 : Semantic Role labeling

- Building an SRL is traditionally done in a two-stage architecture consisting of an argument identifier and an argument classifiers. The FrameNet corpus is the database of semantic frames used for SRL.
- The Curator extended the SRL building process by adding a pruning module before the argument identifier. Pruning filters out simple constituents unlikely to be arguments.
- The inference module runs after the classification stage and incorporates global information and enforces constraints

srl View

```
Predicate  pick [sense: 01]  [predicate: pick]
           <A0> Argument  Robot
           <A1> Argument  the ball
```

- Module 4 : The output from the SRL is used to create the corresponding action file for the Nao.

METHODOLOGY cont.

Module 5 : Robot Control

- We attempted to use PR2 robot in ROS for actions and gazebo for simulation. However, the packages for PR2 in ROS-Indigo were outdated and no results were obtained.
- We moved to Aldebaran Nao as our target robot. Choregraphe and Webots were used for simulation and programming.
- We implemented tasks like moving to predefined location, turning in place and picking a ball and prepared voice commands for them respectively.

EXPERIMENTS AND RESULTS

- Sample commands such as “Robot pick the ball” and “Robot move to origin” executed successfully.
- Error was generated by different modules during some executions:
  1. Google ASR – Some sample outputs failed such as “Nao pick the ball” which ASR interpreted as “Now pick the ball”.
  2. Semantic Role Labeling – Commands like “Pick the ball” failed as parser couldn't interpret robot as subject of the statements.
  3. Robot – Execution of commands which require robot to bend could result in failure.

CONCLUSION AND FUTURE WORK

- We have prepared an end to end system comprising five modules for Speech Based Control of the Nao Robot.
- Each module presents its own level of error, which propagates through the system affecting modules following it.
- To extend this project in the future, we need to expand the action set of the Nao robot.

REFERENCES

- Brian J Thomas and Odest Chadwicke Jenkins. RoboFrameNet: Verb-centric semantics for actions in robot middleware. ICRA2012
- FrameNet : https://framenet.icsi.berkeley.edu/fndrupal/
- Google Voice Recognition
- Stanford NL Parser : http://nlp.stanford.edu/software/lex-parser.shtml
- Nao Aldebaran : http://doc.aldebaran.com/2-1/software/webots/webots_index.html
- Utilized Nao Project : http://www.cogsys.wiai.uni-bamberg.de/teaching/ws1213/pj_b/project_report.pdf