

# HRI Project

Michał Ostyk-Narbutt, 1854051

Prof. Nardi, Artificial Intelligence 2A

January 21, 2019



**SAPIENZA**  
UNIVERSITÀ DI ROMA

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Pipeline</b>	<b>2</b>
<b>3</b>	<b>Tools</b>	<b>2</b>
<b>4</b>	<b>Implementation of SDS.</b>	<b>3</b>
4.1	Scenario . . . . .	3
4.2	Dialogue . . . . .	3
<b>5</b>	<b>Conclusions</b>	<b>4</b>

# 1 Introduction

The goal of this project was to implement a task-oriented Spoken Dialogue System (SDS) in a proposed scenario. This report presents an attempt of performing a Human Robot Interaction (HRI), between a bartending-bot and a human.

## 2 Pipeline

The bartending-bot, interacts with the human by asking them a series of questions. The architecture of a SDS problem is depicted on Figure 1. At first the user speaks or responds to a question, or to ask one. This speech is recorded and analyzed through a **Speech Understanding/Recognition** API. Prior to analysis in the Dialogue Manager, the words/sentences are tokenized. Next, the program attempts perform Language understanding (LU), through Domain and User Intent identification. Post identification, the Dialogue Manager, may attempt to find a policy through searching through a specific knowledge database. After the system output's a response, a speech generator/synthesizer will communicate it to the user in audio form.

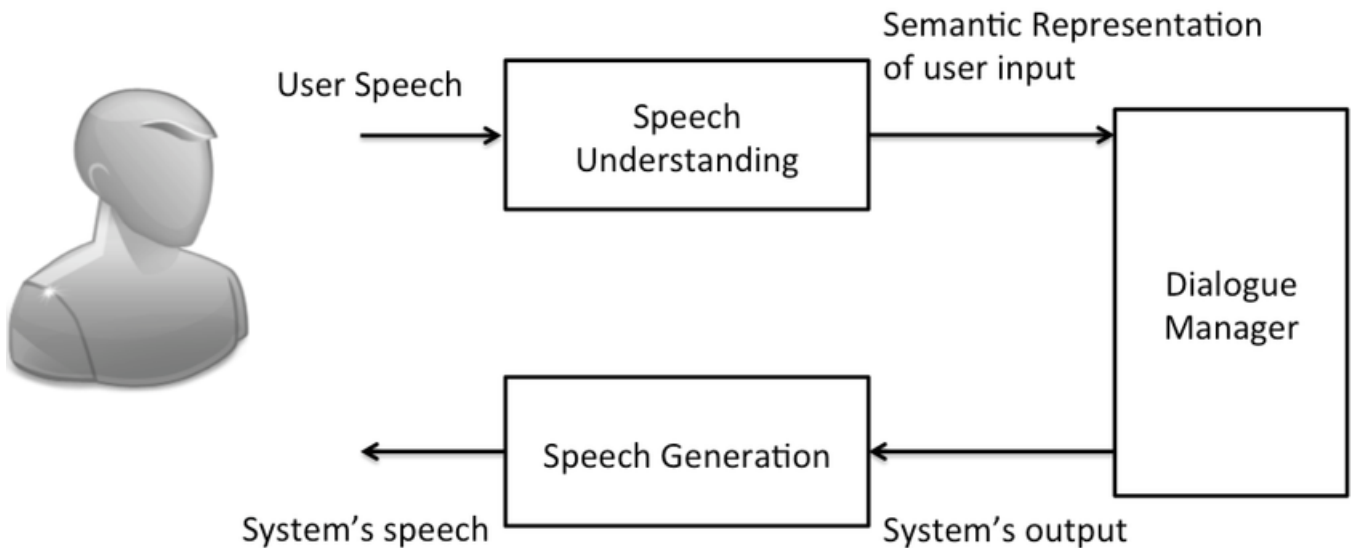


Figure 1: SDS architecture [1]

## 3 Tools

The implementation of this project was done in the multi-paradigm programming language Python; version 3.6.4. To perform natural language processing, the following helper tools have been utilized:

- Automatic Speech Recognition (ASR) – Google Speech Recognition [2],
- Dependency Parser – Spacy [3],
- Text-to-speech – SAY (built in with Mac OS).

All of the code associated with this project and its python library requirements can be found under this link here [4].

## 4 Implementation of SDS.

### 4.1 Scenario

You are at a bar on a Friday after a long work-week, you want to drink but are unsure of what exactly. Hence you ask the bartender what can he recommend you. The bartender is an expert and knows many drinks. He proposes the base alcohol of the drinks and then will ask you if you want a specific drink or if not, then he will recommend something really refreshing.

Every time you speak, the system generates a dependency graph. On this basis it tries to perform language understanding. Most of the time it looks for case specific nouns, adjectives as well as compounds. In the following section, an implementation of the scenario is presented.

### 4.2 Dialogue

Bartender: Long day huh? What do you want to drink?

Bartender: I can make you a drink with either

Bartender: vodka

Bartender: rum

Bartender: gin

Bartender: brandy

Bartender: tequila

Bartender: whiskey

Bartender: wine

Bartender: or liqueur

Listening... :

Done, now processing...

You: I want a drink with whiskey

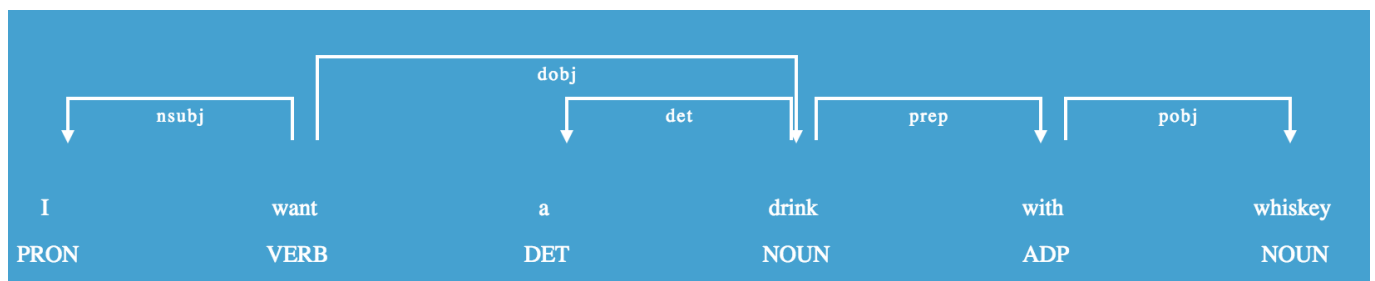


Figure 2: Spoken sentence dependency graph - "I want a drink with whiskey".

Bartender: Great choice! We have the finest selection of whiskey

Bartender: Do you have any specific drink with whiskey in mind?

Listening... :

Done, now processing...

You: no I don't

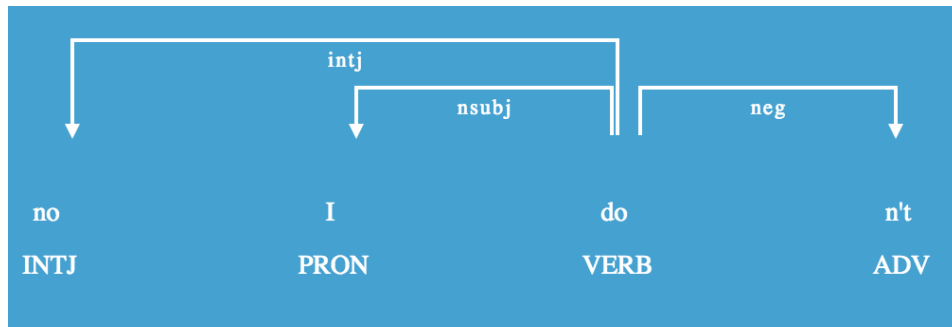


Figure 3: Spoken sentence dependency graph - "No I don't".

Bartender: Let me offer you something  
 Bartender: Okay, How about I make you a Cherries Jubilee  
 Listening... :  
 Done, now processing...  
 You: ok great

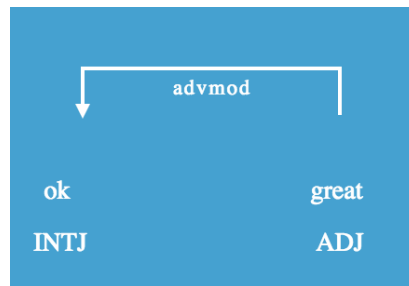


Figure 4: Spoken sentence dependency graph - "ok great".

Bartender: great, starting to prepare it right now  
 Bartender: Here you go, a nice cold Cherries Jubilee

## 5 Conclusions

Overall, the bartender is not-stupid, having an immense knowledge of drinks and can make them quite fast. However, when dealing with something as specific as the names of drinks, the bartender struggles. This is due to the fact that numerous drinks have been named after Cities and countries implying Proper nouns, as well as some are combinations i.e. 'Gin and Tonic'. Also when names include adjectives i.e. 'Mr. Big', 'Canadian Club Old Fashioned'. To conclude, Google Speech recognition and spacey are superb tools to start of the journey with learning HRI and NLP, but to provide a proper intelligent system, requires a lot of time, large databases and a lot of trial and error.

## References

- [1] Jean-Baptiste, Emilie. (2016). Statistical Task Modeling of Activities of Daily Living for Rehabilitation.
- [2] <https://pypi.org/project/SpeechRecognition/>
- [3] <https://spacy.io/>
- [4] <https://github.com/Ostyk/bartending-bot>