

Multi-stage Human-computer Interaction for Command Refining on an Intelligent Personal Assistant

Project Supervisor: Prof. David P. Rossiter
Ivan Gondoprastowo

Department of Computer Science and Engineering, The Hong Kong University of Science and Technology

Introduction



Figure 1. Siri's responses to different commands

Your intelligent personal assistant **is not** that intelligent

1. It combines speech recognition, advanced artificial intelligence, machine learning and natural language processing techniques to try to **understand** verbal commands given by the users and then perform specific tasks according to those commands

2. It tries to automatically understand the context and the semantics, **predict** what the user wants it to perform, and then perform that action, based on a **single verbal command**

A study conducted by Piper Jaffray in 2012, a Minneapolis-based investment bank, graded the technology for resolving requests with the letter "D" (poor) for accuracy

So, what if:

1. It does not have force itself to try to understand the command if it really does not understand

2. It does not have to perform an analysis on your historical actions or understand the situation you are in, in order to perform your command

3. It does not have to try to predict things that maybe you do not even know yet

4. It can help you to think about what you want

in other words, what if they can ask you questions that are relevant to your initial command?

"Find me a restaurant"

What kind of restaurant do you want?

"hmm..Chinese restaurant"

Where do you want it to be located?

"near me"

I found fifteen Chinese restaurants fairly close to your location

Objective

1. To investigate the multi-stage human-computer interaction for command refining, a new approach to deliver better results and user experience on an intelligent personal assistant
2. To compare the proposed multi-stage interaction to the single-stage interaction widely adopted on implementations of Intelligent Personal Assistant such as Siri and Google Voice

Methodology

In order to achieve the objective of the project, we broke down our work into three steps:

1. Building an Intelligent Personal Assistant that incorporates both single-stage approach and multi-stage human-computer interaction implementation
2. Design and conduct a preliminary study to support the multi-stage interaction design choice
3. Design and conduct the main study to compare between the single-stage approach and the multi-stage interaction

System Design and Implementation

In order to eliminate factors that might be resulted from different implementations of intelligent personal assistant system, and to allow us to compare between the commonly adopted and the proposed multi-stage approach, we chose to build our own intelligent personal assistant that helps users to find places to eat or drink, which incorporates both approaches.

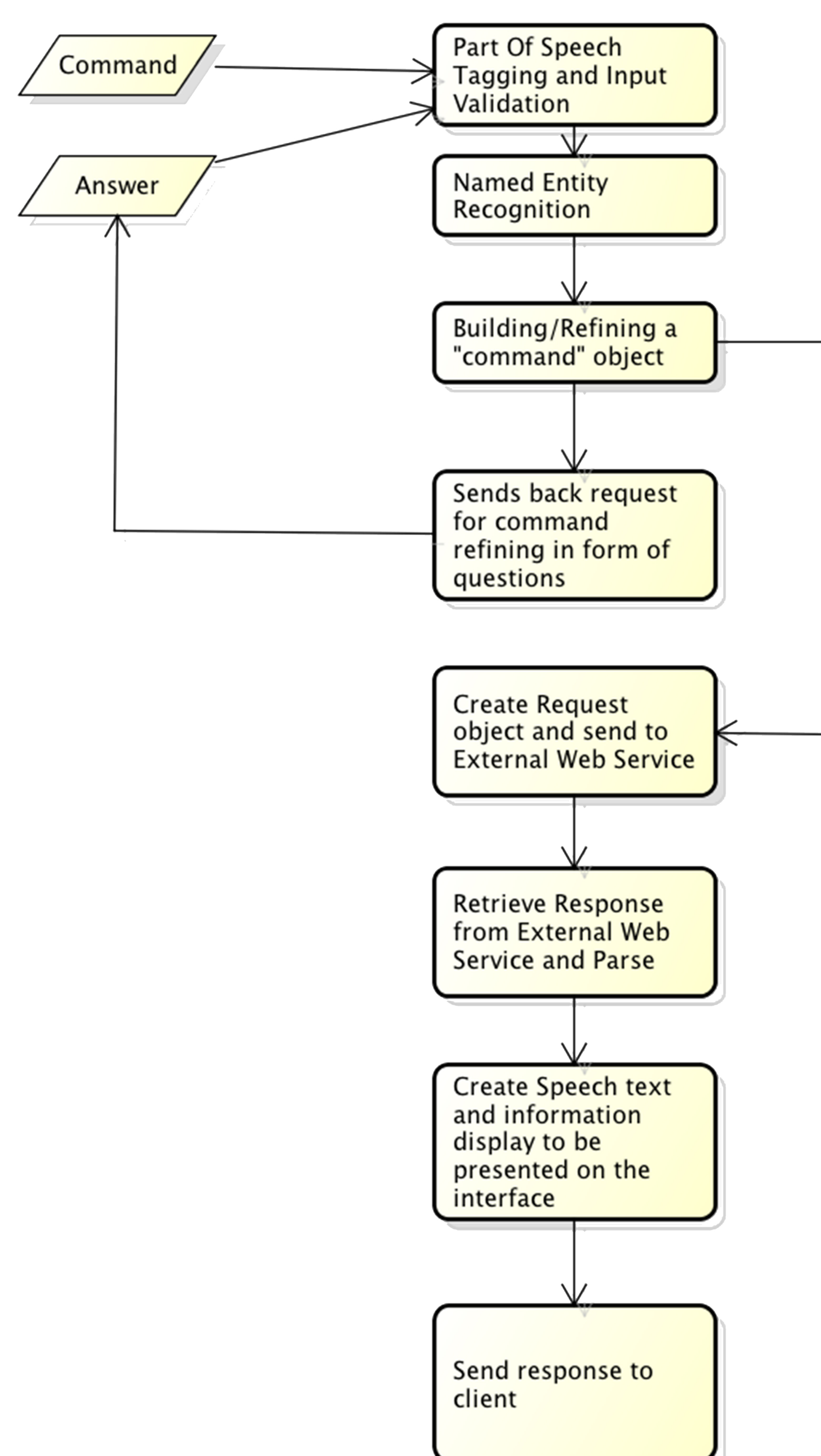


Figure 2. Process Flow Chart

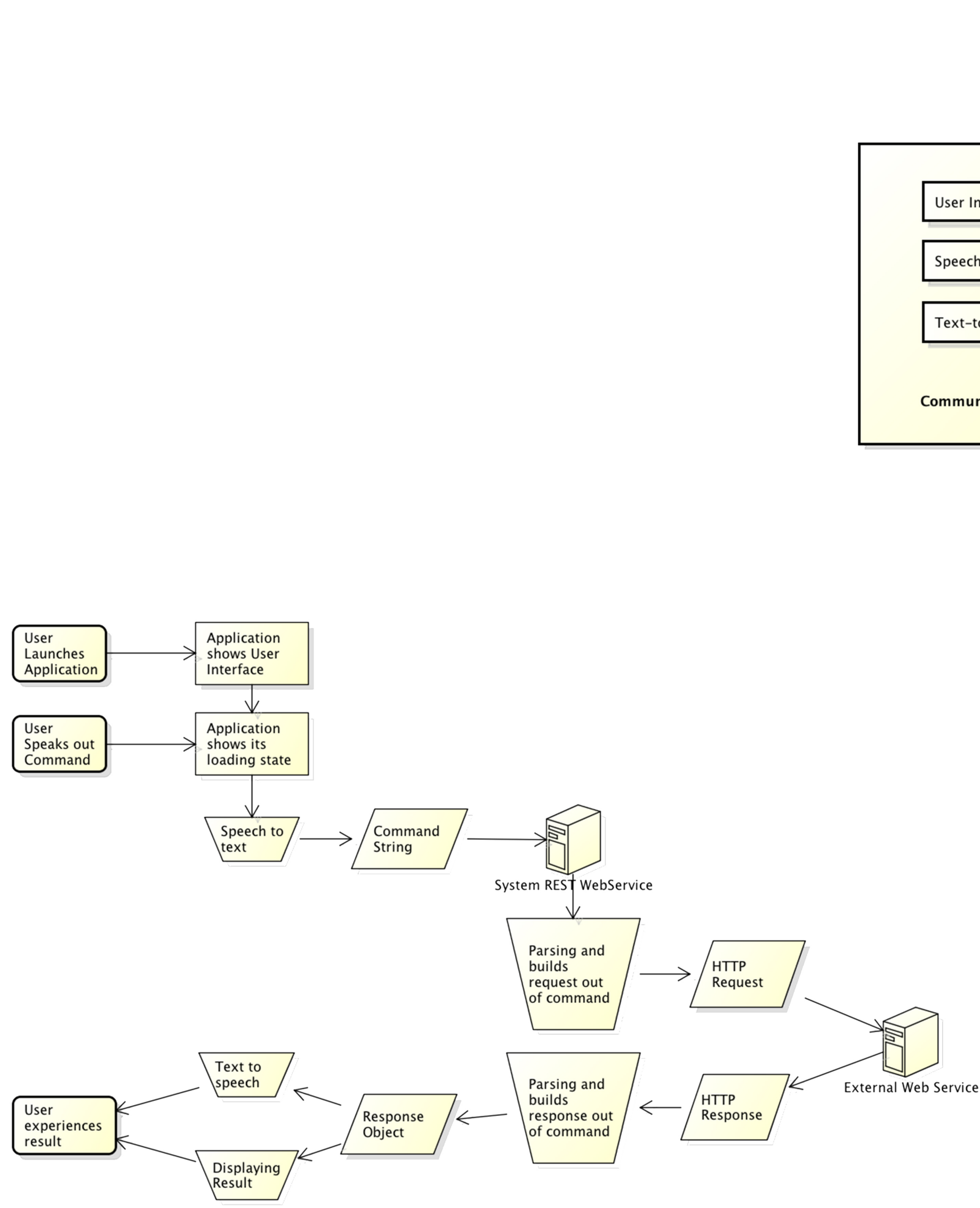


Figure 4. Single-stage Interaction Flow

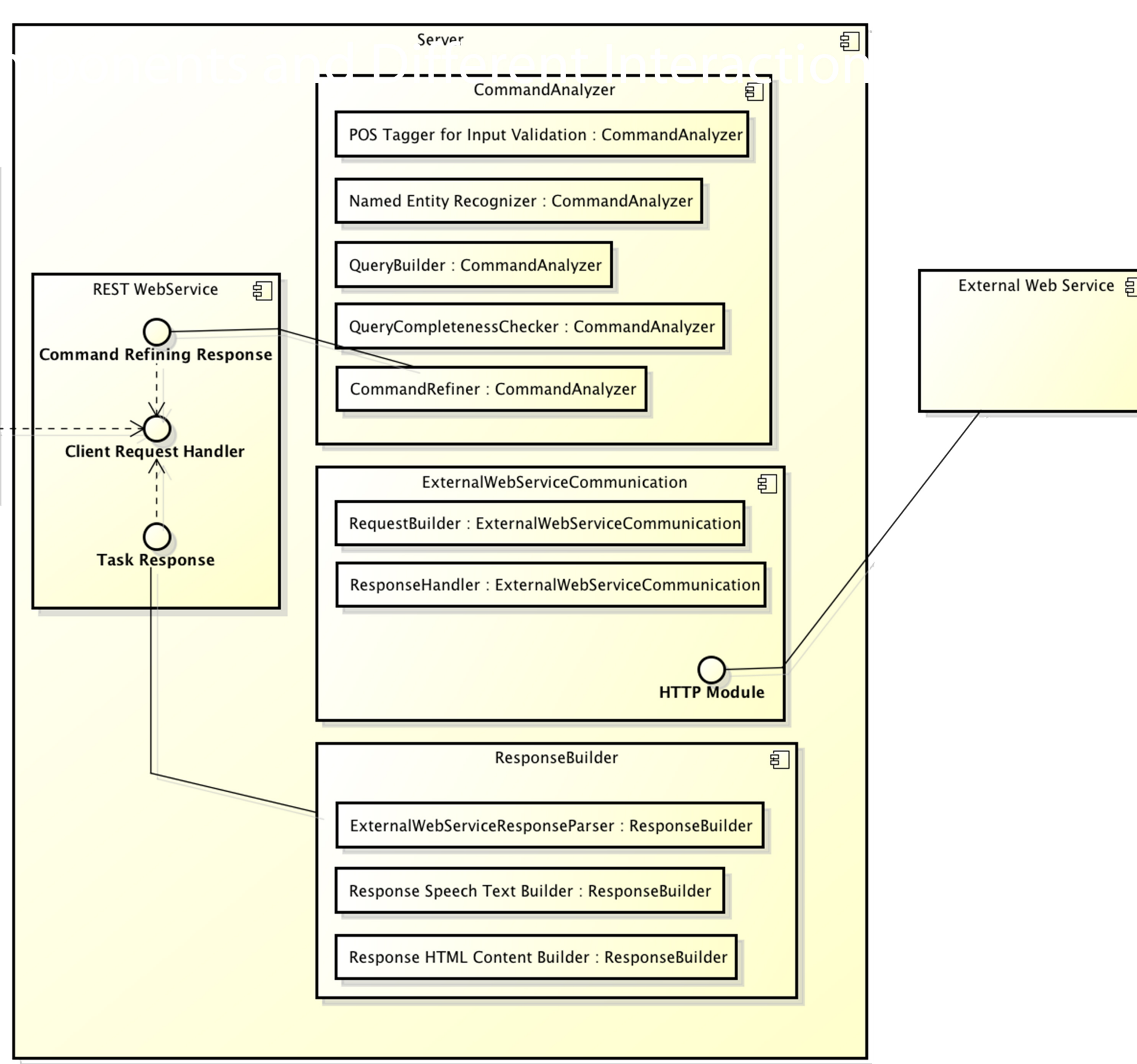


Figure 3. System Component Diagram

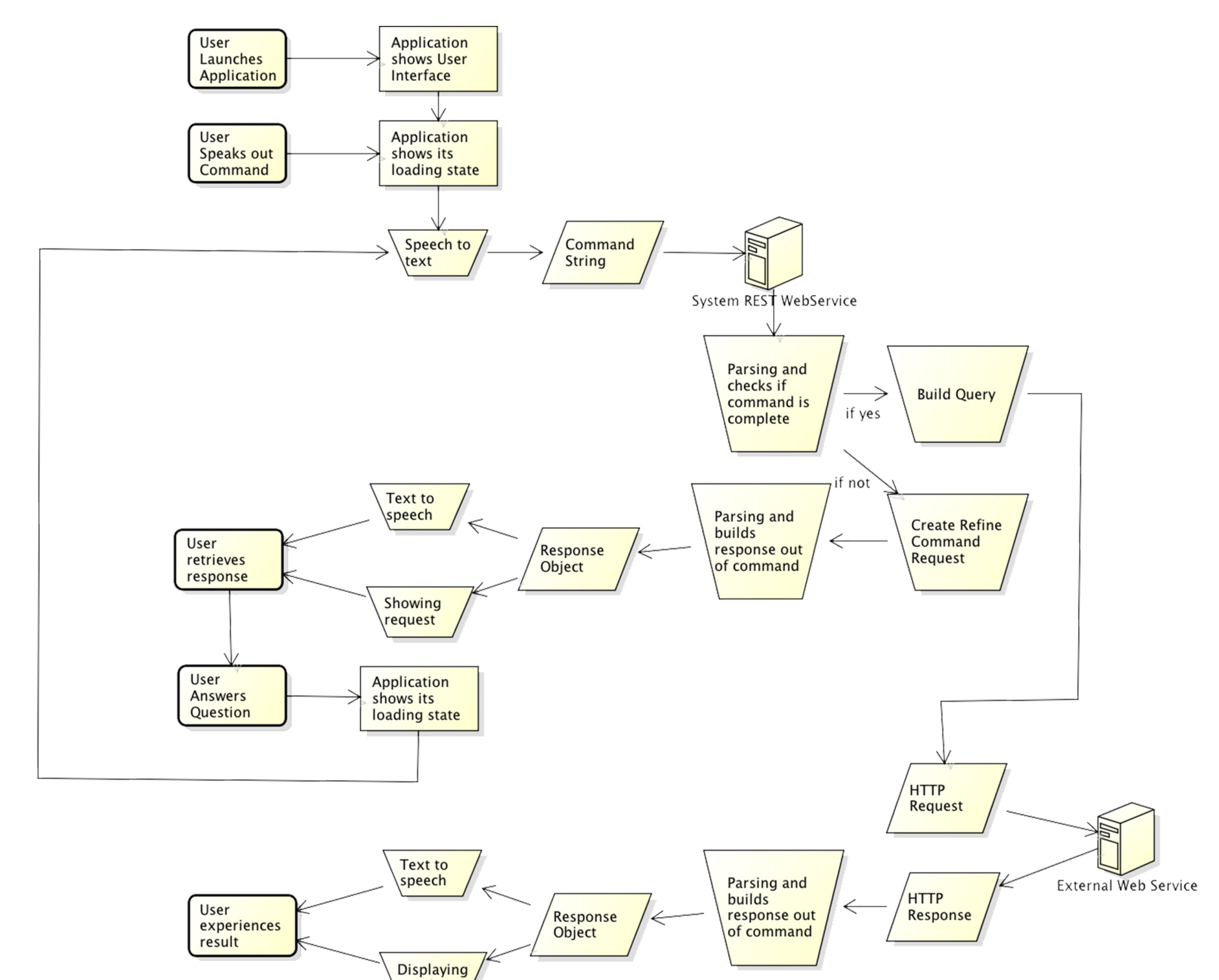


Figure 5. Multi-stage Interaction Flow

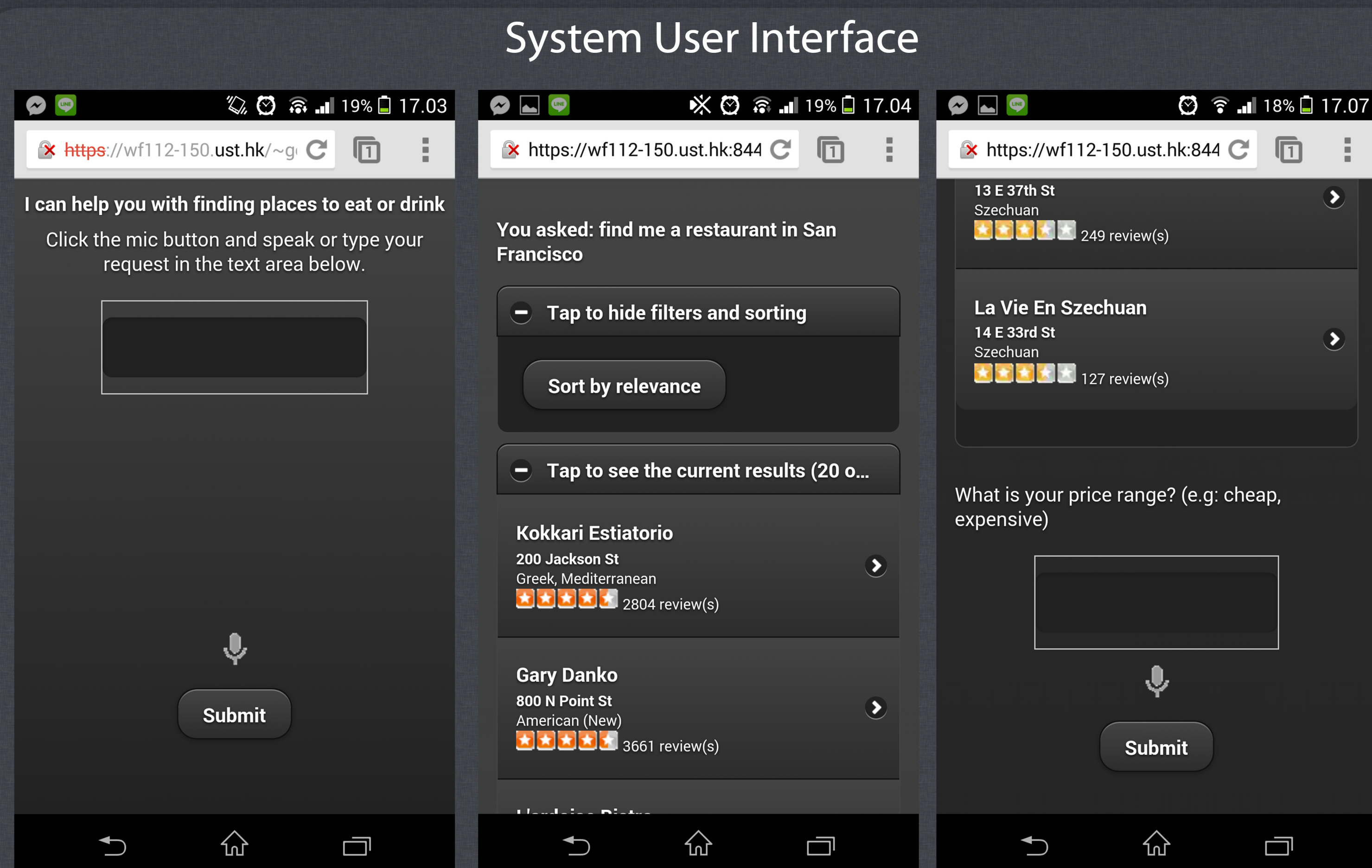


Figure 6. System User Interface

Preliminary Study

To achieve the first objective, we conducted a preliminary study that tried to examine the proper depth of filtering and the relevant filters. The study requires subjects to use our implemented voice assistant to solve for tasks that are set to simulate different economic conditions, different ranges for location, and different time window. Subjects were using a version of the system that is modified to fit the purpose of this study. The system supported only the multi-stage interaction implementation, and for each task, users are required to answer some questions related to their experience and satisfaction.

Main Study

To compare the satisfaction and user experience between the single-stage and multi-stage approach, we conducted the study by requiring subjects to use our implementation of voice assistant. This assistant implements A/B testing method; it randomly switched between single-stage and multi-stage interaction. The study comprised of a case of recent graduate living in San Francisco, with 6 tasks of finding places to eat in different places for different occasions. The tasks are designed to fit regular use cases of intelligent personal assistants. Upon completion of each task, subjects are required to indicate their satisfaction of result, user experience, perceived system performance, and compare each aspect to the immediately preceding.

Results and Findings

Preliminary Study

Number of Participants (n): 7	Average	End-of-task
System Performance	7.43 / 10	7.57 / 10
No. of questions that yield best results	2.64	3.42
No. of questions that yield best experience	3.32	2.86
Filters	1. Location 2. Type 3. Price	1. Location 2. Type 3. Price

Implications:

1. Multi-stage implementation should ask at most 3 questions
2. Filters relevant are Location, Type, and Price
3. Intermediate result should be displayed

Number of participants (n) = 10

Subject Statistics:

Age: 17-27 yrs old

Gender:
Male (60%),
Female (40%)

Exposure to voice assistant:
Yes (50%), No (50%)

English Fluency:
Intermediate (30%),
Fluent (60%),
Native (10%)

Result Highlights:

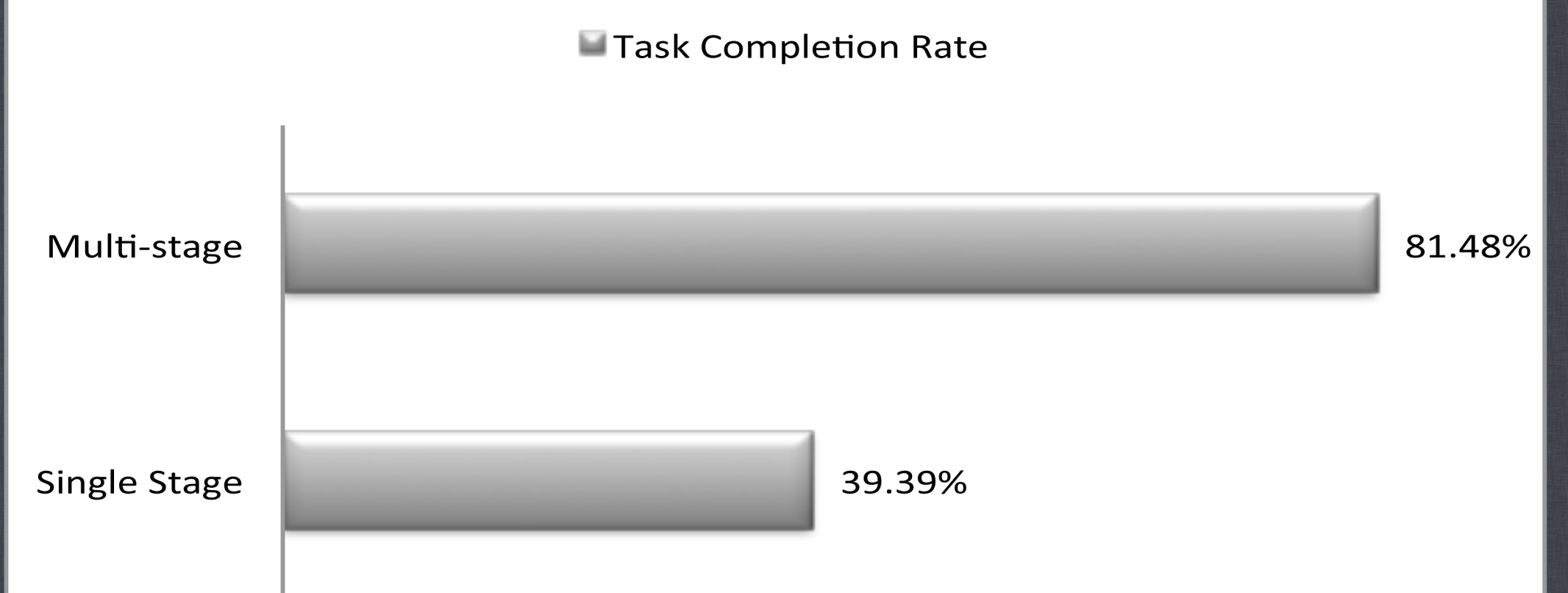
Improvement in task completion rate*: **106%**

System Performance: 7.7 / 10

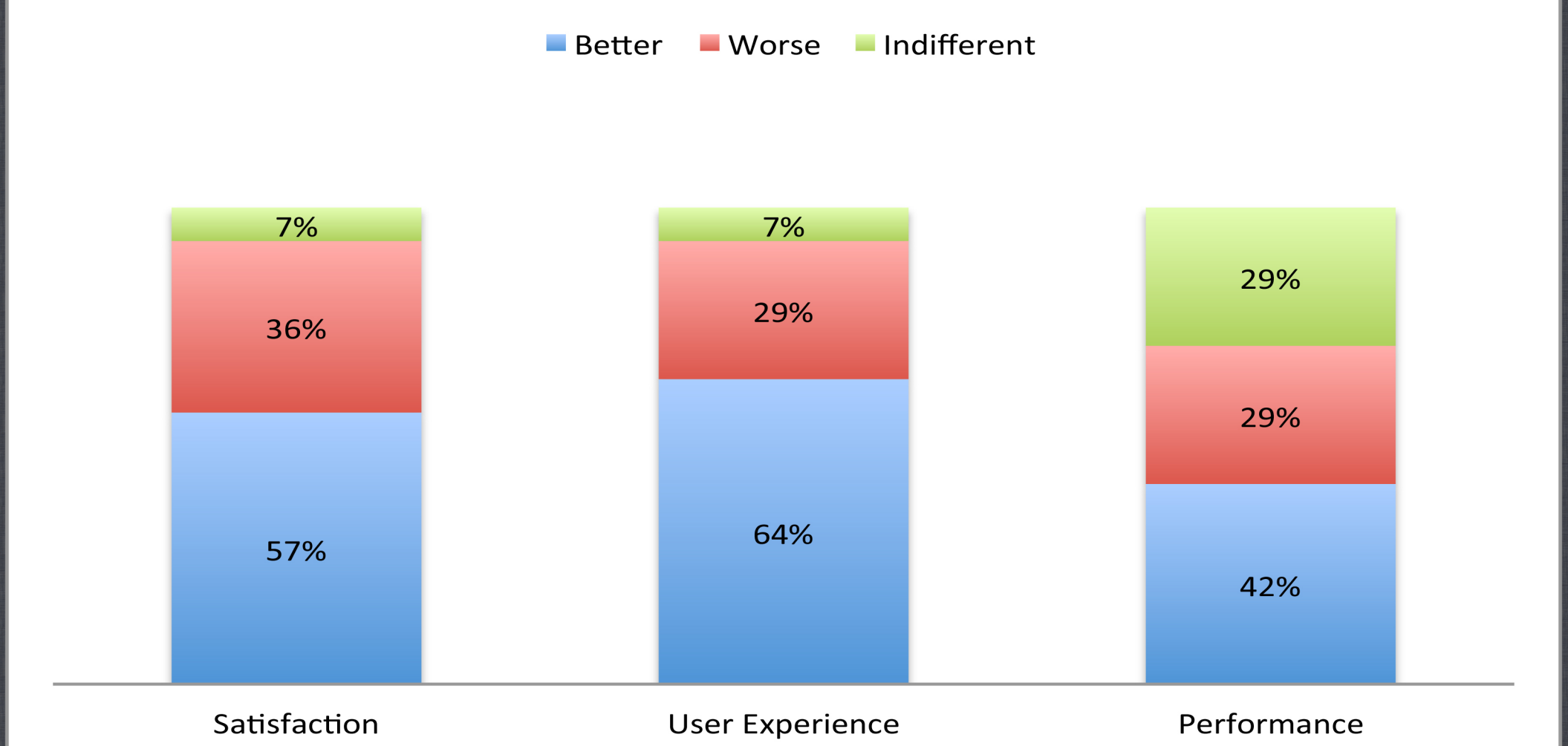
*Task Completion rate is defined as whether or not subject was able to find the specific kind of place to eat or drink at a particular location that complies with several filters/specification

Main Study

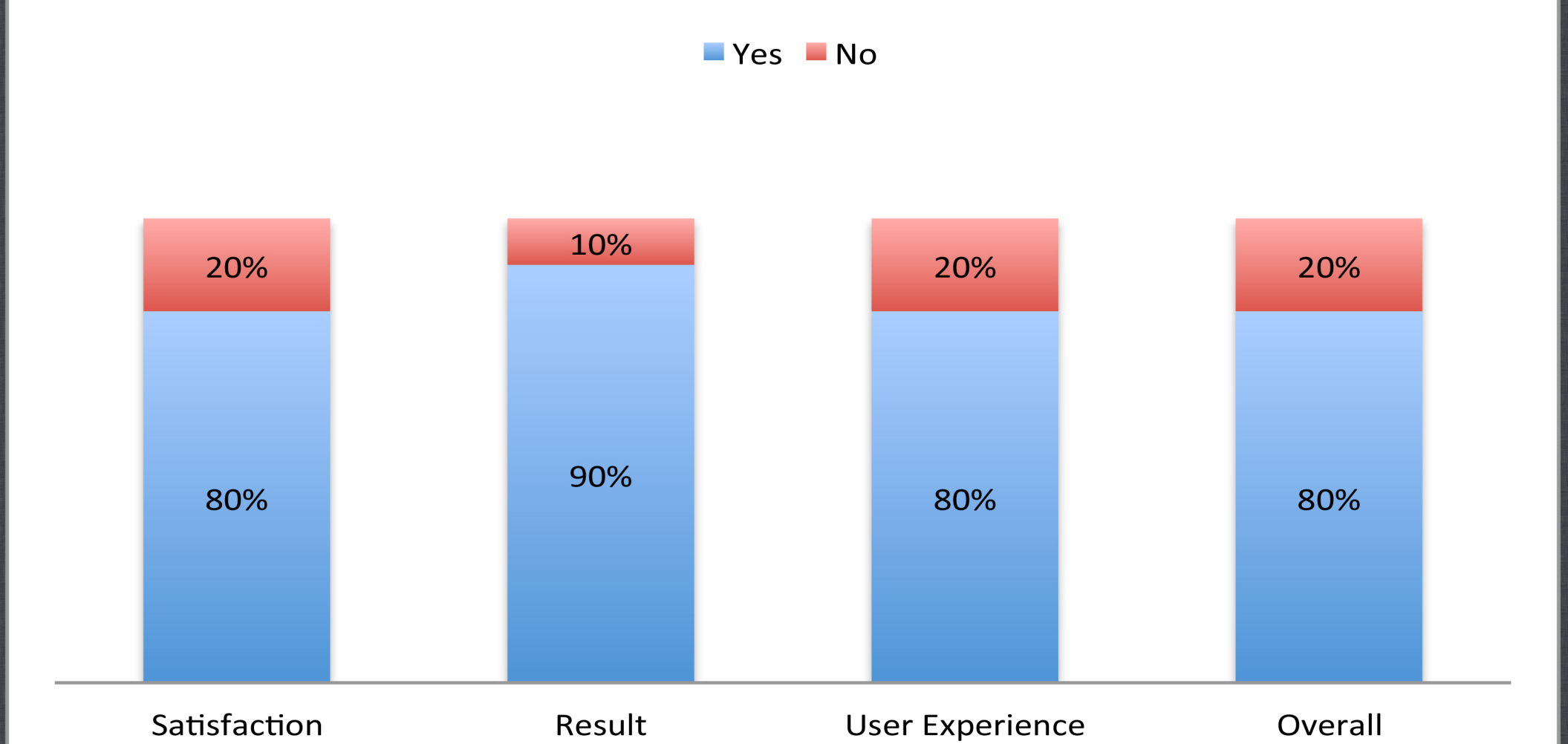
Task Completion Rate



Subject Perception of Multi-stage Interaction



Would Multi-stage interaction help improve these aspects?



Limitations and Future Works

1. Unavailability of data for businesses in Hong Kong. This can be improved by partnering with companies who provide data of businesses
2. Sample size is not big, lacking diversity and representativeness. This can be improved by adding more subjects to the pool and increase the diversity by bringing subjects from different background
4. System is not as capable as current implementation of voice assistants. This can be improved by refining the system and utilizing more computing power to take advantage of latest technologies in natural language processing, artificial intelligence and machine learning.
4. Lack of mobile devices to be used for experiment. This can be done by acquiring different kind of mobile devices (phone and tablets) to be used in the experiment

Conclusion

1. Multi-stage approach yielded significant improvement in task completion rate (106% improvement in task completion rate)
2. Multi-stage approach is perceived to give a better satisfaction (57%) and user experience (64%)
3. Multi-stage approach should be supported by relevant questions and proper depth of filtering, this depends on the nature of task