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

Ivan Gondoprastowo

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

Abstract This project proposes and investigates a new approach to deliver a better result and user experience of intelligent personal assistant. This project is highly relevant to the design and development of applications and services for next generation mobile devices and wearable devices. Currently, most intelligent personal assistant implementations such as Apple Siri, Google Voice, and Samsung S Voice adopt a one-stage human-computer interaction approach. The approach made by these technologies analyzes users' commands and directly presents the results. This project investigates whether a multi-stage human-computer interaction approach would improve the quality and utility of the results. The study has shown that these improvements yield a better user experience and user satisfaction.

1. Introduction

This project addresses the current limitations of digital voice assistants such as Apple Siri [1] and Google Now [2]. These digital voice assistants combine speech recognition, 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. This results in specific system behavior when a verbal command is received; that is, the system will try to automatically understand the context and the semantics, predict what the user wants it to perform, and then perform that action, based on the single verbal command.

While the process seems natural, the best current implementations often perform poorly and return unpredictable or unwanted results or actions. A study conducted by Piper Jaffray in 2012 [3], a Minneapolis-based investment bank, graded the speech technology for resolving requests with the letter "D" (poor) for accuracy. This is far from ideal. There is an exponential growth in voice based input. Starner [4] predicts huge potential for voice based input method adoption for the next generation of wearable computers, such as Google Glass and smart watches. Ultimately, the growing adoption of intelligent voice assistants creates the need to solve the problem as soon as

possible, as users are starting to utilize the application to perform tasks while doing other activities, such as driving or jogging.

This project examines a different approach for input method application. Multi-stage human-computer interaction to refine verbal commands requires little user effort yet will greatly improve the quality and utility of the results. It can also avoid the necessity of users having to refine their commands by restarting the whole process, which is an intuitive reaction to poor search results. Furthermore, having the system ask for more details creates more natural human-computer interaction. Ultimately, better results and more natural interaction will create a better user experience.

This project is highly relevant to the design and development of applications and services for next generation mobile devices and wearable computers, such as Google Glass. The project will compare the user experience with the two approaches, single-stage verbal commands and multi-stage interaction for command refining. We will measure the user experience by having volunteer testers try both methods. The user experience subject study will include (but not limited to) several measurements of the user experience such as task completion time, system response time, quality of results, and overall user satisfaction. The result of this experiment can influence user experience design decision of next generation wearable devices applications and services.

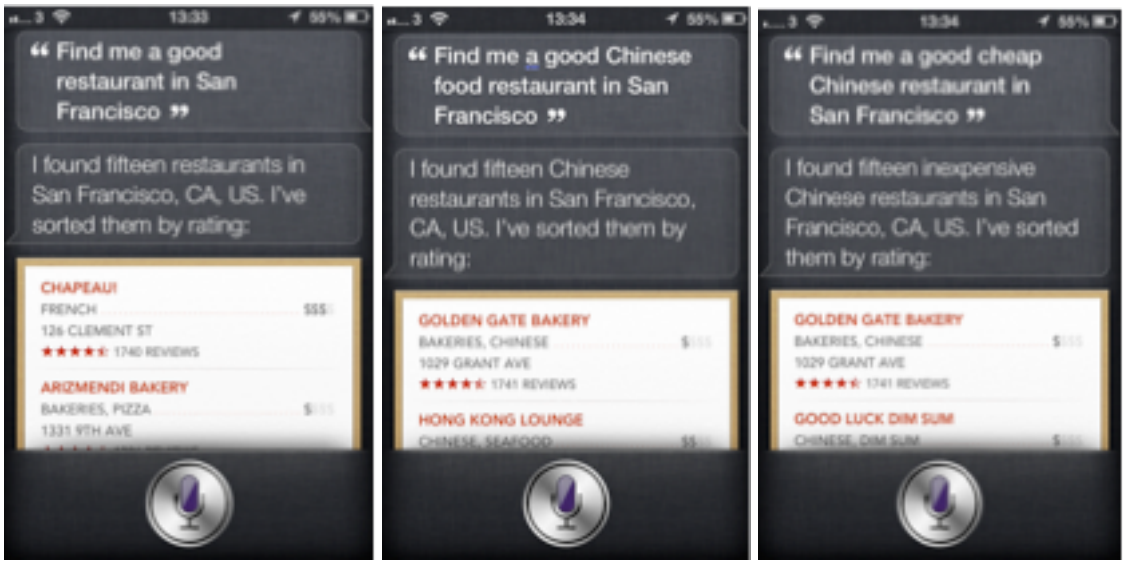

Figure 1. Siri as an implementation of an intelligent personal assistant

The rest of the report is going to be organized as follows. Section 2 describes the methodology we used to achieve the goal of the project. Section 3 discusses the experiment system design, followed by the implementation of the system in Section 4. Section 5 discusses the experimental study design. Section 6 presents the result and findings that we obtained from the experiments conducted. Section 7 analyzes and discusses the results and implications. Section 8 discusses the limitation of the system, experiments, and future plan of the project. Finally, Section 9 presents our conclusion.

2. Methodology

In order to measure and compare the user experience for current intelligent personal assistant implementation and the proposed approach of enabling multi-stage human-computer interaction, first we built our own implementation of intelligent personal assistant that incorporates both the current intelligent personal assistant interaction approach and the proposed multi-stage human-computer interaction approach. We restricted our implementation of the system to support only commands for finding places to eat or drink in 25 countries that are supported by Yelp! API [5]. This includes building the user interface that would be accessible through standard web browsers on mobile phone which incorporates speech recognizer [6] and text-to-speech [7] capabilities, building the REST style[8] web service that would connect the interface with the back-end part of the system, and building the system that would analyze commands using natural language processing tools such as machine-trained named entity recognition classifier and a set of logics. The system connects to external web services (i.e. Yelp! API) in order to retrieve the information. Moreover, in order to support the experiment, the system implements A/B testing [9] and with logging capabilities for easier result analysis. The implementation of the system will be discussed in more detail in the later section.

Following the development of intelligent personal assistant for the study, we designed and conducted a preliminary study to discover the appropriate level of depth and number of questions that maximizes both the user experience and the result. In order to support this, we modified the system to only support multi-stage human-computer interaction, and would ask series of questions based on the filters that are supported by the Yelp! API. The system would analyze user's command and find out what other filters could be applied to narrow down the result set. Users are recommended to stop at a point where they feel the results are satisfying. The system was then reverted to the version that incorporates A/B testing (i.e. switches randomly between current and the proposed approach of interaction). The result of the experiment was analyzed and incorporated to the implementation of the system for the main study. Detailed description and discussion of the preliminary study will be presented in Section 5.

After the analysis of preliminary study and incorporation of the findings to the system used for the experiment, we designed and conducted the main study. The main study incorporates tasks which resemble daily use cases of intelligent personal assistant. Subject plays a role of a young graduate living in San Francisco, California, U.S.A, and was presented with tasks related to finding places to eat in the U.S. The main study design will be discussed in more detail in Section 5. Finally, the result of the experiment was then analyzed together with the user activity log.

3. Intelligent Personal Assistant System Design

As mentioned in the earlier part of the report, in order to conduct a proper experiment, we need to analyze our proposed approach and isolate the part of the current implementation of intelligent personal assistant that can be used as the basis for comparison of the experiment. As the current implementation of intelligent personal assistant does not incorporate multi-stage human-computer interaction, we decided to build our own system for our experiment. This system would replicate

the essential parts of the system with simplified capability and incorporate the current human-computer interaction approach and the proposed multi-stage human-computer interaction approach.

Due to the complexity of the system, we divided the system into three parts, the user interface (front-end) part, the web service, and the back end:

3.1 User Interface

The user interface part of the system is responsible for retrieving command and presenting results to the user. This part of the system is accessible through Google Chrome [10] browser, which is available across all platforms including mobile. This part of the system provides both regular (typing) and voice input method. The user interface enables users to access the web service available and retrieve data from it.

The other responsibility of this front-end part of the system is to present the result and ask questions for refinement when multi-stage human-computer interaction approach is being applied. In order to create a better user experience, apart from presentation of result, system will also provide a verbal reply for every request.

3.2. Web Service

In order to connect the front-end and the back-end part of the system, we decided to build a REST web service. This web service acts as an interface for the back-end part of the system which provides accessibility to modules that are implemented in the back-end part of the system for the front-end.

The web service is responsible to handle requests and commands from the front-end part of the system and pass it to the back-end part of the system for processing. Upon completion of the process, the web service will forward the response from the back-end part of the system to the front-end part of the system for presentation of result or request for further refinement.

3.3. Back-end System

The back-end part of the system is responsible for most of the operations and computations needed upon the receipt of the command. It recognizes what the user intend to look for by utilizing set of logics and natural language processing tools to handle the natural language based command and turn it into set of actions and keywords to be passed on to the module that would connect to external web services. This module encode the important items and will then interact with the Yelp! API to retrieve the information needed to satisfy a command. The result from the Yelp! API will then be parsed and processed before being sent back to the web service as a response for a request.

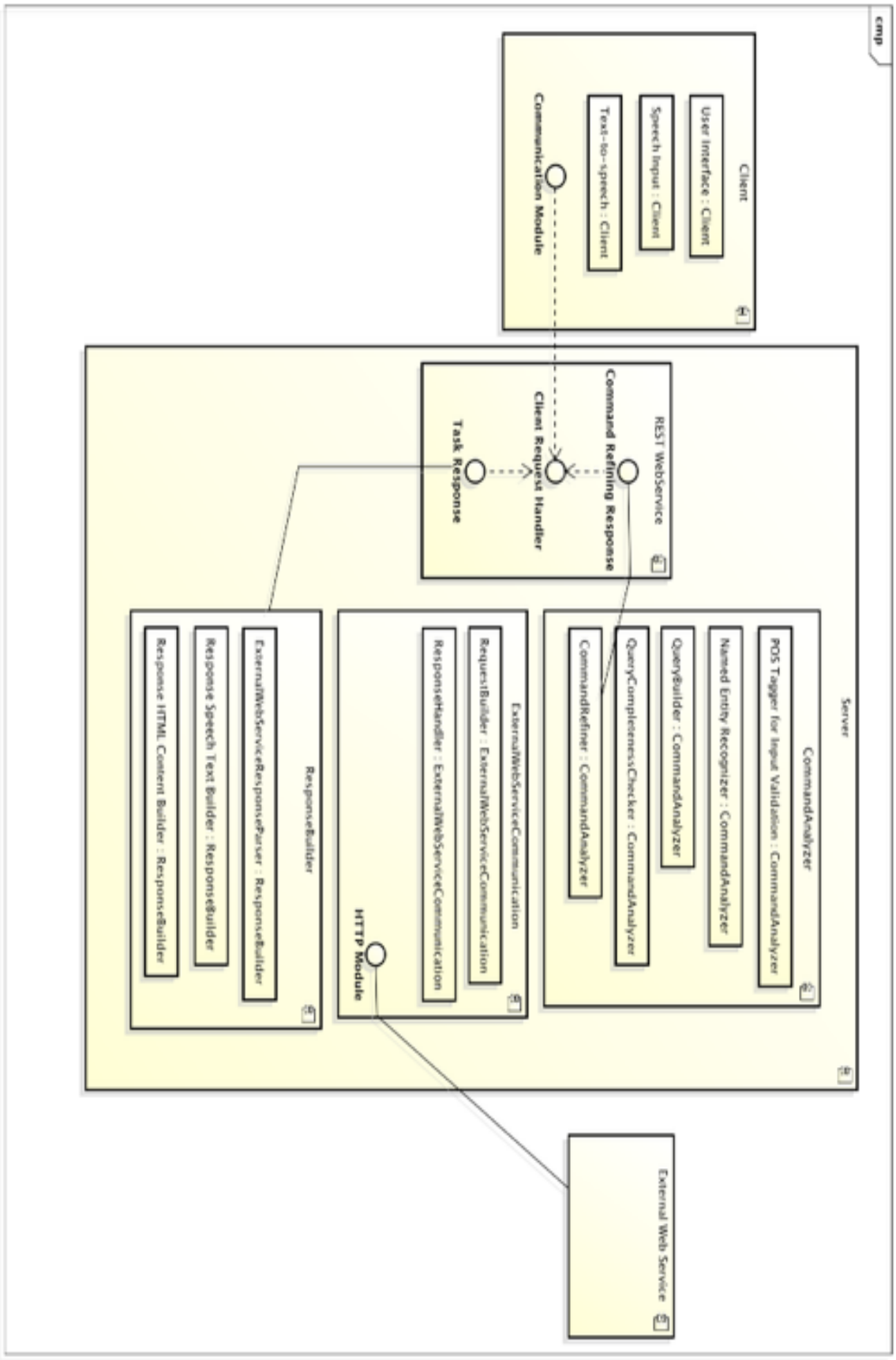

Figure 2. Component Diagram of the Overall System Implementation

Another important part of the system design is the user experience and usage flow, as well as the process flow. For the implementation of the system to be used for this experiment, we designed two

flows that will represent the current implementation of human interaction and the proposed multi-stage human-computer interaction.

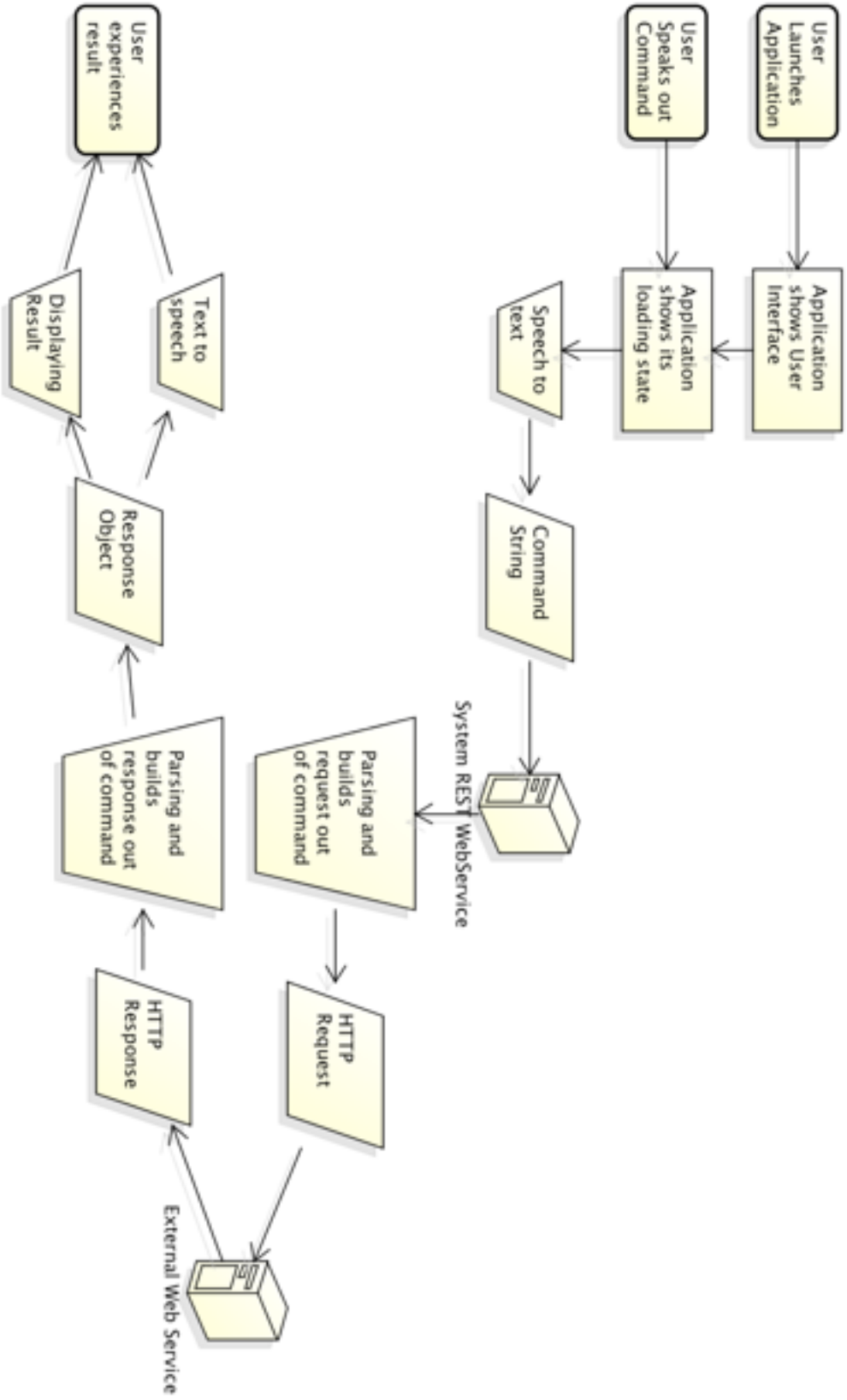

Figure 3. Usage Flow for Current Intelligent Assistant Human-computer Interaction Approach

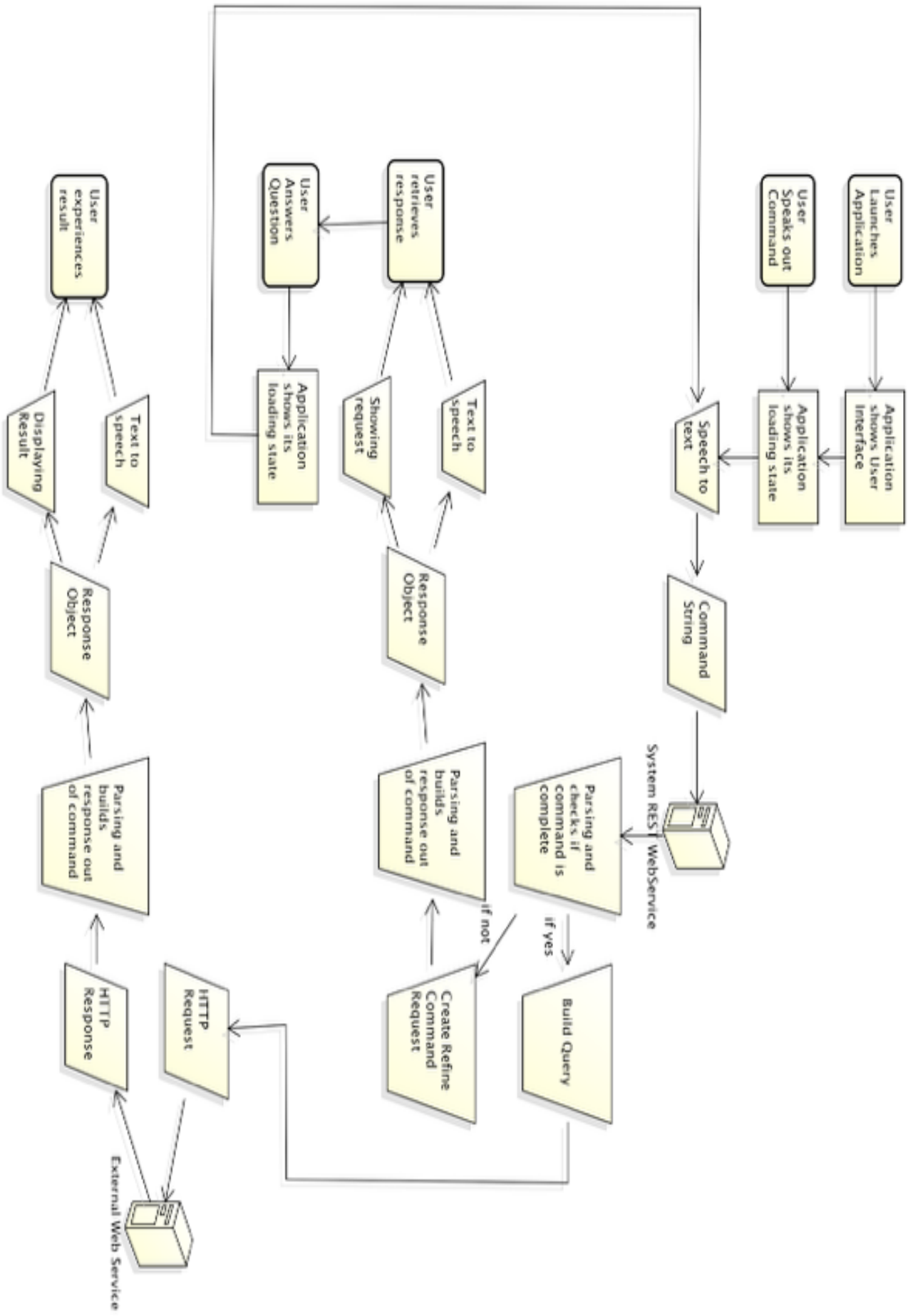

Figure 4. Usage Flow for Multi-stage Human-computer Interaction Approach

4. Intelligent Personal Assistant System Implementation

In this section, we present the implementation method used to develop the intelligent personal assistant system for our experiment. Identical to the previous section, we will divide this into three parts:

4.1. User Interface Implementation

First, we break down the features of the user interface of our intelligent personal assistant. The user interface is designed to be compatible with interaction and gestures on mobile interface, as currently, implementations of intelligent personal assistant are mostly adopted on mobile devices. For this reason, the user interface (front-end) part of the system is being implemented by utilising a mobile interface framework, jQuery Mobile [11]. This would allow us to quickly re-use the design of most user interface elements that are necessary for our system implementation.

In addition to that, to support the voice input method, which is the main command input method for the system, we utilize Google Chrome Speech Input API [12]. This speech recognizer is embedded to the user interface of the system by using JavaScript. Our implementation makes it possible to show the intermediate recognized words on the text box prior to the final recognition result.

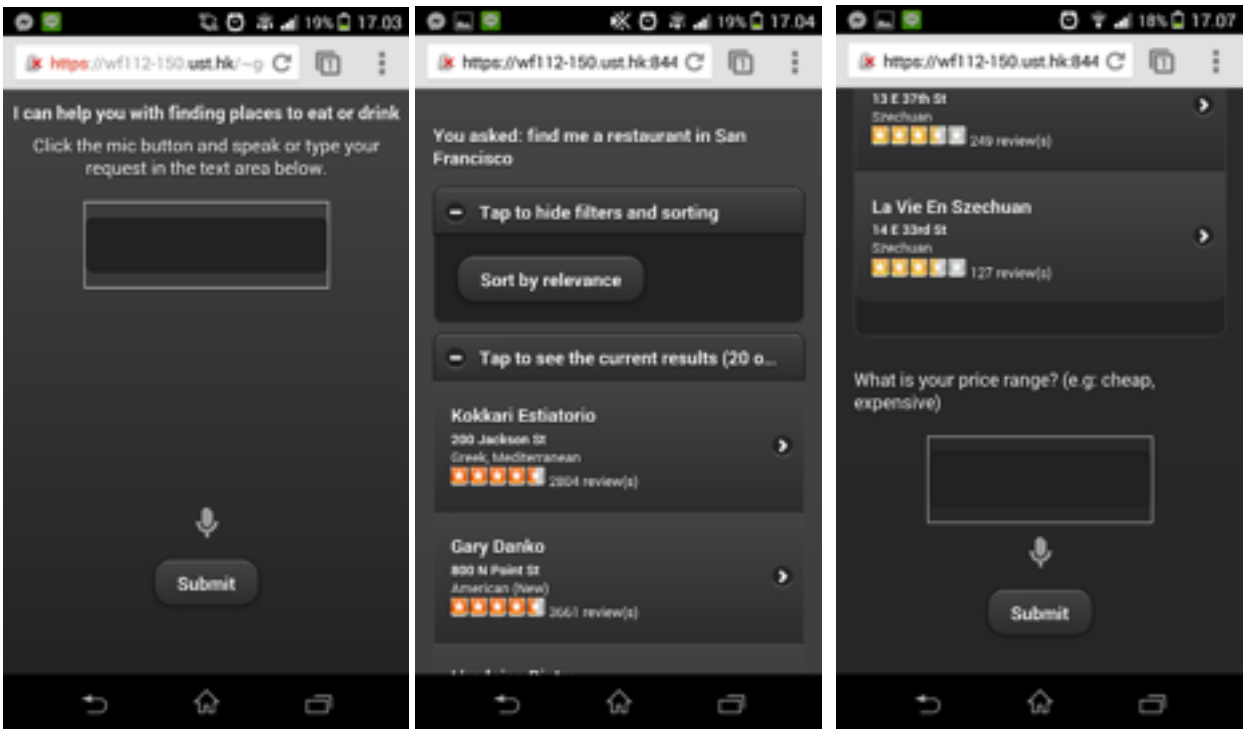

Figure 5. System User Interface

Another feature of our implementation of user interface is the voice feedback that is given to the user along with the presentation of the result on the interface. This is being done by utilizing Google Translate text-to-speech API. In order to incorporate this to our own implementation, we had to implement a text-to-speech gateway which would receive a string of words from our back-end implementation, and sign it with specific encrypted code to

imitate a request being sent through Google Translate service. This workaround allows us to use Google Translate text-to-speech feature that otherwise will not be accessible from outside Google Translate service.

On the user interface level, user activity logging is being done by using JavaScript, specifically on every action/gestures performed by the users. This allows us to detect whether users performed voice input or text input, on which part of the interface did user pay more attention to, on which element did user tap, and record the time for each actions, especially, scrolling.

4.2. Web Service Implementation

As described in the previous section, the web service acts as a bridge between the front-end and back-end functionality of the system. This is being made available by implementing representational state transfer (REST) architectural style. Implementation wise, this is simply another way of setting up a Java based web server. However, this allows clients to access the server through an interface that is defined slightly differently from basic Java servlet-based web server.

In implementing REST web service on Java web server, in our case, Tomcat 7.0, we use Jersey, a reference implementation of JAX-RS, which is a Java API for RESTful Web Services [13]. This allows us to use annotations defined in JSR311, making it easy for us to build a REST based web services by using Java programming language.

The steps taken upon receipt of response through a specified URL are as follows. First, we will determine whether this is a new request or continuing request (i.e. refinement for previous command). We will then represent the command as an object and recognizes the command to extract important attributes of it. Following that, if the request is a new request, we will randomly choose between using the flow for current approach or multi-stage human-computer interaction approach. System will then proceed with respective procedures and pass the response back to the front-end part of the system. In the occasion where a continuing request is received, system will try to analyze user's response and refine the important attributes of the command based on the previously detected attributes.

4.3. Back-end Implementation

The back-end system is implemented using Java. It interacts with the front-end through the REST web service. The back-end functionality of the system handles most of the operations that need to be done in order to process a command. Firstly, the incoming command will be analyzed and important attributes will be extracted. This extraction is being done by using a named entity recognizer classifier [14] developed by Stanford Natural Language Processing Group and a set of logic that will be applied to extract and formulate important attributes of the command. Following that, this set of important attributes is formulated to a form of query and signed with a unique access token provided by the Yelp! API. Upon passing of the request, a response will be retrieved from the API. A module is implemented to parse this response into a compatible object for further processing. Lastly, our response builder module

will build a response that contains the result of the API access. This response will be passed to the REST web service to be sent back to the calling front-end part.

Another notable feature of the back-end implementation is the logging module. For each operation done by the system, a record of its input and output will be recorded. This includes recording all the commands retrieved, result of analysis of important attributes of the command, modules called, implementation selected (usage flow chosen), and formulated query for external services access as well as its response. The user activity log from the user interface will also be recorded. In order to implement this module, we utilized a Java-based logging library, log4j [15].

The following figure describes how the service implemented.

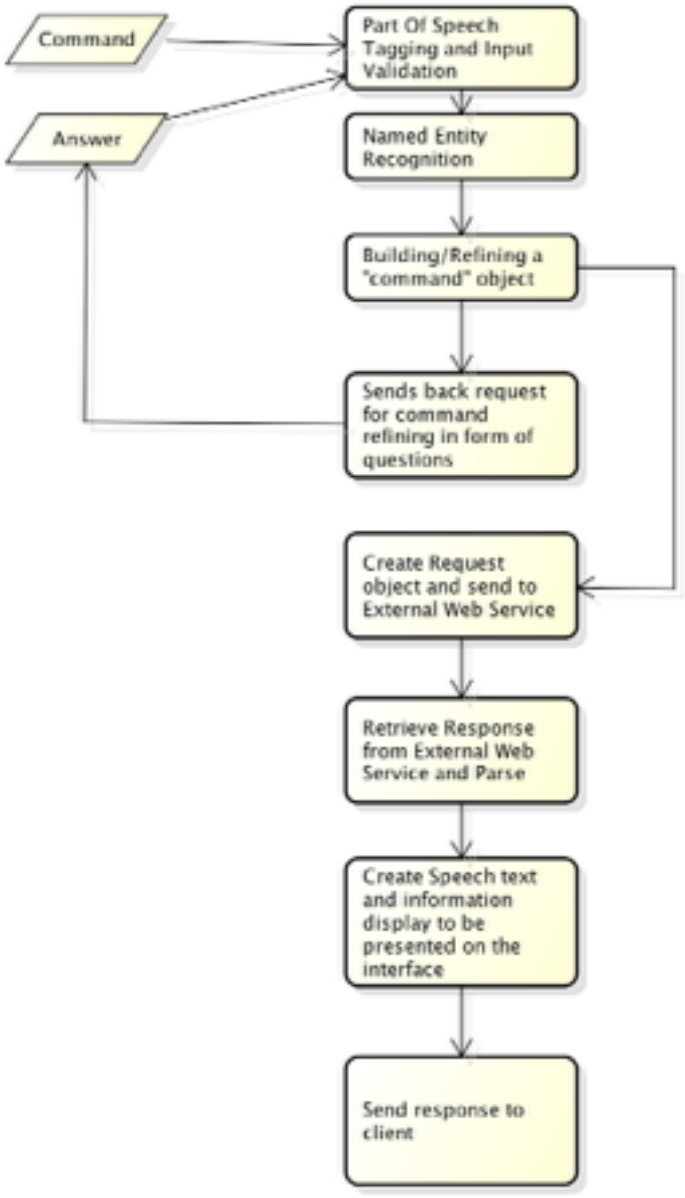

Figure 6. System Implementation Flow Diagram

5. Experimental Study Design

It is important to note that effectively, what we are trying to study is the effect of having multi-stage human-computer interaction as compared to the current single-stage human-computer interaction that is being applied in most intelligent personal system implementation. Translating this to design of our experiment, our basis for comparison will be user satisfaction, task completion, user experience, and user perception on the performance of the system when subject is following the current single-stage human-computer interaction approach flow.

In order to compare between the two approaches, we used A/B testing method and have the subjects use our implementation to accomplish several tasks of similar difficulty and complexity level, and have them compare the four metrics described above between each task and the task immediately preceding it.

The main study comprises of a case with 6 tasks that need to be completed. These tasks are related to finding places to eat or drink at some given location. Subject is given a role of recent graduate who are working in a city in the U.S. In the experiment, subject commutes to work, school event, and visits his/her friend in another location in the U.S. Upon completion of each task, subject is requested to answer some questions with regard to the four metrics described above.

As mentioned in the previous sections, in order to discover the proper level of depth to mitigate the effect of having inappropriate amount of filter which could affect the user experience when using the system, we conducted a preliminary study to find out how many questions and what are the questions that are relevant to the user experience. We developed a study that comprises of four cases having one task each. The task would require subject to find a place to eat or drink given a situation / condition. This includes time pressure, economic condition pressure, and different dining partner(s). For each task, subject is requested to move on to the next task when they feel the results presented is satisfying, when they hit the final result set or when they feel annoyed. For each task, subjects are also required to fill in the number and type of questions that will give them optimal user experience and result.

The following sections will present, analyze and discuss the result of the experiment and the interesting findings from the experiment.

6. Experiment Results and Findings

In this section we will present the result of both preliminary and main study for this project. First, we present the information about the preliminary study and the result and findings of the preliminary study. The preliminary study was conducted to find out the appropriate level of depth for filtering and the number of questions that users believe will give the optimal user experience and result. The available filters are location, type of restaurant, price, distance, and whether or not to show only businesses with Yelp! deals.

Preliminary Study Result

The subjects are students with age ranging from 16-22 years old. Subjects comprised of Hong Kong nationality (29%), Mainland Chinese (28%), and other countries (43%). The first task simulates a condition where subject is required to find a place to eat, subject can choose any location for dinner,

without time pressure or economical condition restriction. For this task, 85% of the subjects successfully accomplished the task, giving the system 7.14 out of 10 for its performance. 57% of subjects stopped working on the task due to satisfying results, 29% of subjects got annoyed by the series of questions, and the rest (14%) stopped because they have reached the final result page (i.e. no more filter can be applied). On average, subjects feel that the number of questions that would provide them the best experience is 3 and the number of questions that is necessary to give a satisfying results is 2. Subjects believe that location, type of restaurants and price would be the appropriate filters with priority according to the order.

The second task requires subject to find a place to eat near a specified location, under time pressure, but no economical restriction. Simulating the time pressure, subject gets to try only once. For this task, 85% of the subjects successfully accomplished the task, giving the system 8 out of 10 for its performance. 72% of subjects stopped working on the task due to satisfying results, 14% of subjects got annoyed by the series of questions, and the rest (14%) stopped because they have reached the final result page. On average, subjects feel that the number of questions that would provide them the best experience is 2 and the number of questions that is necessary to give a satisfying results is 2. Subjects believe that location, distance and type of restaurants or price would be the appropriate filters with priority according to the order.

The third task requests subject to find a place to eat without any specified condition, but under time pressure and economic condition (i.e. limited budget). For this task, 85% of the subjects successfully accomplished the task, giving the system 6.85 out of 10 for its performance. 43% of subjects stopped working on the task due to satisfying results, 29% of subjects got annoyed by the series of questions, and the rest (28%) stopped because they have reached the final result page. On average, subjects feel that the number of questions that would provide them the best experience is 3 and the number of questions that is necessary to give a satisfying results is 3. Subjects believe that price, type of restaurant and distance would be the appropriate filters with priority according to the order.

The last task of the preliminary study is designed to discover the appropriate filters and number of questions by eliminating the extraneous variable that might come from different ways of formulating the initial command. For instance, some subjects might ask “Where can I find cheap restaurants in San Francisco”, while some others might say “Find me a cheap restaurant”. The two examples show the different in initial depth of the command, with the former has higher depth as it incorporates the location while the latter does not. For this task, subjects are required to find a place to eat by saying “Find me a restaurant”. The task yields 85% accomplishment rate with averaged overall performance of 7.71 out of 10, with 85% of termination resulted of user satisfaction of the result. On average, subjects feel that the number of questions that would provide them the best experience is 3 and the number of questions that is necessary to give a satisfying results is 3. Subjects believe that location, type of restaurants and price would be the appropriate filters with priority according to the order.

At the end of the experiment, subjects are being asked about the performance of the system and number of questions to give them optimal user experience and results. The following table summarizes the finding from the first study.

	Average across all tasks	**End-of-experiment Average**
Performance	7.43 / 10	7.57 / 10
No. of questions yielding best result	2.64	3.42
No. of questions yielding best user experience	3.32	2.86
Filters	Location, Type, Price	Location, Type, Price

From the preliminary study, it can be concluded that the number of questions that would optimize the result and user experience is 3. As shown on the table above, the filters that are important and useful are location, type, and price respectively. This is applied to the system by setting the maximum number of questions to 3 with filters available for specifying location, type, and price.

The preliminary study also yields some interesting results that eventually was incorporated to the design of the system for the main study. Firstly, it was found that subjects prefer that the intermediate results are shown even when the system is still asking for further filtering, as 57% of subjects stated that it is better to have it shown rather than collapsed as implemented in the preliminary study. Secondly, it was discovered that subjects tend to accept a result set when they see businesses that they have been to, as 57% of subjects found a restaurant that they are familiar with in the result set and all of them accept the result set. Another interesting discovery is the fact that subjects generally want to be able to see the business page or get more detailed information about the business while still having the result page available. This is not possible in most implementation of the intelligent personal assistant available in the market. For this, 57% of subjects pressed the link, got redirected to the page, and asked if they can go back to the result page.

Main Study Result

The main study compares the user satisfaction of the result, user experience, task accomplishment rate and perceived system performance when the different flows are being applied. The first flow which will be the basis for comparison is the current single-stage human-computer interaction approach implemented on most available intelligent personal assistant. The second flow is the proposed multi-stage human-computer interaction that allows for command refining. Note that the base flow and the proposed flow are being applied randomly and users do not get to know which flow they are currently in.

The subjects are students with age ranging from 17-27 years old. Subjects represent different nationalities such as Hong Kong (30%), Indonesia (30%), and other countries, including Czech Republic, Swiss, and India. Subject pool is comprised of 60% male and 40% female, with 50-50 proportion for their exposure to intelligent personal assistant, and 30% of the pool have visited the U.S.

We define task accomplishment as whether or not subject is able to find the specific kind of place to eat at a particular location that complies with several filters/specification. With regard to the task accomplishment, it is discovered that base implementation (single-stage human-computer interaction) yields 39.39% task accomplishment rate. This is outperformed by the multi-stage human-computer interaction proposed which yields 81.48% task accomplishment rate (42.09% increase in task accomplishment, 106% improvement compared to the base implementation)

The user satisfaction is defined as the degree of satisfaction of the result set presented the users get by using the system. The study shows that 57% of subjects feel that they get a better satisfaction for the result set presented when they are facing the multi-stage human-computer interaction immediately preceding a single-stage interaction approach. For this 36% feels that it gives them a worse satisfaction rate and the rest (7%) stating indifference between both approach.

The user experience is defined as the degree of good experience the users get from using our system. The study discovers that 64% of subjects feel that they enjoy a better user experience for overall flow of the process when they are facing the multi-stage human-computer interaction immediately preceding a single-stage interaction approach. For this 29% feels that it gives them worse experience and the rest (7%) stating indifference between both approach.

The perceived system performance measures how well does the system perform according to the users. The study yields that 42% of subjects feel that the performance of the system improves when they are facing the multi-stage human-computer interaction immediately preceding a single-stage interaction approach. For this 29% feels that it gives them worse experience and the rest (29%) stating indifference between both approach.

The following section discusses the result of the main study as well as some interesting findings from the preliminary study and how it can be applied to current implementation of intelligent personal assistant.

7. Discussion

In this section, we discuss some of the important results obtained from both the preliminary study and the main study part of this project. Moreover, we also discuss some interesting findings from both study and its implication for the design of the intelligent personal assistant human-computer interaction. Finally, we will provide some suggestions that could be useful for future improvement of intelligent personal assistant user experience design.

The preliminary study was conducted to find out the proper depth of filtering and the number of questions that would yield the optimal user experience and satisfaction of retrieved results. While this can be considered as an insignificant part of the main study itself, the preliminary study provides an isolated experiment that helps us get some information to support our main study. While the main study focuses on comparing the user experience created by multi-stage human-computer interaction to the single-stage interaction used by most implementation of personal assistant available, it is crucial for us to understand how many questions that would yield the optimal user satisfaction of the experience in order to capture the optimal satisfaction yield by having multi-stage human-computer interaction approach. Failure to do so might yield unwanted

and invalid result that is resulted from unoptimized implementation of the multi-stage human-computer interaction.

The preliminary study gives us a guidance for the design of multi-stage human-computer interaction. Moreover, this result yielded is relevant for the design of current intelligent personal assistant human-computer interaction. The result implies four important and useful suggestion for improvement. Firstly, the multi-stage human-computer interaction is preferred when the questions are relevant. Secondly, the intermediate result should be shown, as this gives user a choice whether to proceed with further refining or to stop at a particular filter. Thirdly, providing users with capability to look for more information about the restaurant while still giving them an option to go back to the result page might yield a better user experience.

For instance, current implementation of Siri does not allow users to go back to the result page after choosing to see more information about a restaurant. Lastly, relevancy and ordering of filter options to the situations should also be considered. For example, during working hours, it might be better to have less number of questions and have filters for sorting by distance option asked first.

The main study was conducted to compare the single-stage human-computer interaction on an intelligent personal assistant, the approach that is being used on most available intelligent personal assistant system, which also serves as our basis for comparison, to the proposed multi-stage human-computer interaction on an intelligent personal assistant to provide users with capability of refining their initial command with available filter options. We predict that this approach will give users better user experience and satisfaction of result.

The study result shows that the proposed approach significantly increase the task accomplishment rate. The approach also yields better user experience, satisfaction of result, and user perception of system performance compared to single-stage human-computer interaction. This implies that the proposed approach can be a good option when designing for intelligent personal assistant user experience. Having the personal assistant asks you questions for command refinement helps users accomplish the task. Moreover, the approach also creates a better user experience if the number of questions and the questions asked are relevant. While further refinement might be necessary to bring an even better user experience and satisfaction of the result when multi-stage human-computer interaction is adopted, the result helps shed the light to a new direction or user experience design option.

8. Limitation and Future Plan

This section presents several limitations that the project has and discuss the future plan for the project.

With regard to the limitations of the project, there are few important items to be discussed. Firstly, the unavailability of data of businesses in Hong Kong that is open for public access forced us to use Yelp! API. Yelp! API provides a very robust developer services for finding places to eat. However, as mentioned above, only 30% of the subject pool has been to the U.S. Therefore, subjects might not know much about places to eat in the U.S. It also makes it harder for subjects to compare between results, as some subjects are not familiar with any places in the result set.

Secondly, the subject pool is not diverse enough in terms of their age and occupation, as all of them are students, with age ranges from 16-27. This lack of variety of samples might affect the external validity of the experiment. It is possible that people from different age or occupation groups have different preference with regard to human-computer interaction design on an intelligent personal assistant.

Thirdly, the system implemented is rather simple compared to the implementation of Siri. This is reasonable as Siri is an artificial intelligence based system that would learn as it is being used by people. Our implementation does not have such capability. Moreover, compared to the language processing technology of the system, Siri outperforms our implementation of system used for the experiment. The effect of having this difference in performance is mitigated by implementing both approaches on our intelligent personal assistant instead of using Siri implementation as the basis for comparison.

The fourth limitation is with regard to the device being used for the experiment. Due to budget limitation, we had to conduct the study on a macbook air machine instead of an android mobile phone. This could bring up an internal validity issue as the experience of using the system includes the device being used for accomplishing the task. Most intelligent personal assistant implementations are available on mobile devices. Our implementation is also designed to fit the mobile devices experience, but the experiment is conducted on a laptop machine.

Lastly, the number of subjects that we have conducted the experiments with is not big due to the time constrain. This possibly will affect the averaged result of the studies as the result might be prone to residual errors.

For further improvement on the project, we can aim to eliminate the limitations described above. For the first limitation, there are two approaches of tackling the problem. The first option is to conduct the experiment in the U.S. The second option is to get businesses data from Hong Kong based restaurant review companies. We predict that it is possible to partner up with Hong Kong based restaurant review providers such as OpenRice, or businesses and location data providers such as Kites. By slightly modifying the module of the system that interacts with the external service, we can set the system to send queries to their interface and parse the responses retrieved.

For the second limitation, further experiment should be conducted with subjects having different occupations representing different age groups. It is important to make the subject pool as diverse as possible to eliminate the external validity issue that might arise from the experiment.

Addressing the third limitation, while it is not possible to fully replicate Siri capability, improvement on the system used for the experiment might be able to eliminate the problem. It is also preferable to incorporate artificial learning capability and scalable computational resources to speed up the process of accomplishing the task.

Dealing with the fourth limitation, it is preferable to conduct the experiment on the mobile device. This can be done by acquiring an android based mobile device for further experiments.

Lastly, the limitation of number of subjects can be solved by more participants for the study. This solution can be combined with the one which addresses the second limitation. In other words, it is

favorable to have bigger subject pool with more diverse background and better distribution for each attribute of the subject.

9. Conclusion

The project proposes and investigates a novel approach to deliver a better result and user experience of intelligent personal assistant by applying multi-stage human-computer interaction that allows for command refining on an intelligent personal assistant. The project compares between the single-stage human-computer interaction which is adopted by most available implementation of an intelligent personal assistant to the proposed multi-stage human-computer interaction approach.

In order to conduct the experiment, a simple implementation of intelligent personal assistant was developed. The system incorporates A/B testing capability that would randomly choose between applying single-stage human-computer interaction or multi-stage human-computer interaction proposed when receiving a new command. This implementation utilizes a set of logic and natural language processing tools to extract the important attributes of the command, which eventually will be formulated to a specified format to be sent as a request to the external service provider, which in our case, is Yelp! API. Apart from that, the system also incorporates text-to-speech capability and voice input API provided by Google Chrome. The system user interface was built by using jQuery Mobile, to get the design optimized for mobile. All parts of the system are equipped with logging modules to record the user activity and information about processes being done for data analysis.

Throughout the project, we conducted two user experience studies. The preliminary study aims to to find out the proper depth of filtering and the number of questions that would yield the optimal user experience and satisfaction of retrieved results. The main study compares the user satisfaction of the result, user experience, task accomplishment rate and perceived system performance yielded by applying either the single-stage human-computer interaction approach or the proposed multi-stage human-computer interaction approach. The preliminary study yielded 3 as the number of questions to be asked to yield the optimal user experience and satisfaction of the results, with location, type of restaurants, and price being the filters favored by subjects. The main study discovers 42.09% improvement in task accomplishment rate when the multi-stage human-computer interaction is applied. The study also finds that the proposed approach yields better user experience, satisfaction of result, and perceived performance of the system.

In the latter part of the report, we presented the limitation of the project as well as the plan for further work that can be done. It can be concluded that the proposed multi-stage human-computer interaction promises a better overall user experience and satisfaction to the user. It also helps users accomplish their task in a more natural way. Incorporation of the proposed approach could definitely be a choice that intelligent personal assistant user experience designers could consider to improve the user experience. Ultimately, further exploration needs to be done for understanding the advantages of using the proposed approach for voice-input based applications, especially the ones implemented on mobile and wearable devices.

References

[1] Wikipedia. (2013). *Siri.* [Online]. Available: http://en.wikipedia.org/wiki/Siri.

[2] Wikipedia. (2013). *Google Now.* [Online]. Available: http://en.wikipedia.org/wiki/Google_Now.

[3] Bloomberg. (2012). *Apple's Siri Gets Below-Average Grade From Piper Jaffray.* [Online]. Available: http://www.bloomberg.com/news/2012-06-29/apple-s-siri-gets-below-average-grade-from-piper-jaffray.html.

[4] T. Starner. "The Role of Speech Input in Wearable Computing". *IEEE Pervasive Computing*, vol. 1(3), pp. 89-93, July 2002.

[5] Yelp. (2013). *Yelp.* [Online]. Available: http://www.yelp.com.

[6] S. Sampath and B. Bringert. (2010). *Speech Input API Specification* [Online]. Available: http://lists.w3.org/Archives/Public/public-xg-htmlspeech/2011Feb/att-0020/api-draft.html.

[7] G. Shires and H. Wennborg. (2010). *Web Speech API Specification* [Online]. Available: https://dvcs.w3.org/hg/speech-api/raw-file/tip/speechapi.html.

[8] R. T. Fielding. "Architectural styles and the design of network-based software architectures." *PhD Thesis*, University of California, Irvine, 2000.

[9] R. Kohavi, R. M. Henne, and D. Sommereld. "Practical Guide to Controlled Experiments on The Web: Listen to Your customers not to the Hippo," In *KDD '07: Proceedings of the 13th ACM International Conference on Knowledge Discovery and Data Mining,* New York, NY, USA, 2007, pp. 959-967.

[10] Google. (2013). *Chrome Browser - Google.* [Online]. Available: https://www.google.com/intl/en/chrome/browser/

[11] The jQuery Foundation. (2014). *About jQuery Mobile.* [Online]. Available: http://jquerymobile.com/about/

[12] Google. (2013). *Web Speech API Demonstration.* [Online]. Available: https://www.google.com/intl/en/chrome/demos/speech.html

[13] Wikipedia. (2013). *Representational state transfer.* [Online]. Available: http://en.wikipedia.org/wiki/Representational_state_transfer

[14] K. Toutanova and C. D. Manning. "Enriching the Knowledge Sources Used in a Maximum Entropy Part-of-Speech Tagger," *In Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora (EMNLP/VLC-2000),* 2000*,* pp. 63-70.

[15] Apache Software Foundation. (2012). *Apache Logging Services.* [Online]. Available: https://logging.apache.org/