

## **CHAPTER II**

### **Review of Related Literature and Studies**

*The Review of Related Literature and Studies showcases previous studies and publications relevant to the project. This chapter gives light as to what motivated the proponent/s in pursuing the specific field of study.*

#### **I. Related Literature**

##### **Foreign**

According to Lee (2016), today's Intelligent Personal Assistants are infamously limited in their capabilities despite of all the technology advances. There are a lot of consumers, as well as tech writers alike that raise problems regarding Siri and Cortana's limitations. The first problem that the researcher mentioned about IPAs, was its purely processing power and capability. The second problem was about the integration of IPAs, in which the researcher mentioned Siri, Cortana and Facebook's M as to be limited because they are essentially closed books. Meaning they are all considered to be in closed-development, in fact that they do not open their technologies to the public and how it was developed accordingly. These IPAs can do an internet search and possibly answer a given question about a closest restaurant, but they can't also make a user's reservations in that restaurant and call an Uber base on the user's time of need. But given that an IPA will be a little more open, things can change quickly, because by making a more open IPA program and allowing it to interface with the other available APIs of other companies and services probably it can do everything for you. IPAs have the potential to bring about another paradigm shift in business to consumer communication. Instead of having to find and search through a company's website, IPAs allow you to just ask the question to your phone. The difference of an open IPA that can be integrated with other business, is that it can basically place an order, reverse an appointment, or even buy a product for you.

According to the website of Cloud Speech API, Google Speech API or Cloud Speech API, is an API that has the capability to integrate Google speech recognition technologies into the developer applications. It works as sending an audio and receive a transcribed text from the Cloud Speech API service. It is powered by Machine Learning that the most advanced deep learning neural network algorithms are applied to the user's audio for speech recognition with unparalleled accuracy. The API also recognizes 110 languages to support global user base. Streaming text results, returning partial recognition results as they become available with the recognized text appearing immediately while speaking, is one of those capability that the Speech API can do. Accuracy in noisy environments, is one of the problems of IPAs these days, because they tend to get the noise of the surrounding as well, while catching the users audio input, but with Google Speech API, the user doesn't need advanced signal processing or noise cancellation before sending audio to speech API because the service can handle noisy audio from different kinds of environments. It supports the Context-Aware Recognition, the it can be tailored to context by providing a separate set of word hints with each API call. And in any case, the speech API supports any device that can send a REST or gRPC request and it includes phones, PCs, tablets and IoT devices such as cars, TVs and speakers.

According to the very own website of Adapt made by Mycroft, it is well defined that Adapt Intent Parser is a framework that is flexible and extensible for intent definition and determination. It is intended to parse natural language text into structured intent that can be invoked programmatically. Adapt is a software library that is open to everyone that can be used for converting natural language into machine readable data structures. As the developers made this Intent parser to take in natural language and outputs a data structure that includes the intent, a match probability and a tagged list of entities. It is designed to run on devices with limited computing services. It can also be tagged as a rules-based artificial intelligence library. An example is, a user might want to create a natural language

interface that allows them to play a Pandora station. The user might say "Turn on Pandora", or "Play Pandora", or "Put on my Joan Jett Pandora station.".

The parser uses greedy and naïve implementation for intent determination. Given an utterance, it uses the Adapt parsing tools to come up with sorted collection of tagged parses. A valid parse result contains no overlapping tagged entities, and its confidence is the sum of the tagged entity confidences, which are weighted based on the percentage of the utterance (per character) that the entity match represents. This system makes heavy use of generators to enable greedy algorithms to short circuit large portions of computation. It has the following features, that makes it unique among the others. It is lightweight, because Adapt is written to run on embedded systems with limited resources such as IoT devices like remote control and hubs. It is portable, due to Adapt is written in Python, which is a language with near universal support and it is designed to run cross-platform and can be deployed in software environments that include Android, iOS, Windows and Linux. It is reliable, because it is small enough to run locally on your embedded device and it eliminates the need to parse intent using the cloud and surely that the applications will function even when the network is offline. Adapt is Rules-based, which means that any functional AI needs to make user of both rules based intent parsing and deep learning. And to add more spice to how great Adapt is, it is the only intent parser that is openly available to the public which makes it easy to deploy a rules-based system to complement deep learning approaches.

According to the article, Mycroft Releases Key AI Component as Open Source (2016), Mycroft released this important piece of code as open source. To be more specific, it is the Adapt Intent Parser, which transforms what the user speaks into something that can be interpreted by an application on a device, such as phone and desktop. Mycroft released the Adapt Intent parser as an open source so it can be used by anyone in other projects and they can also help to make it better. Another reason stated on this article, as to why this release is very important, is its being a lightweight solution that will allow developers to integrate it within devices with limited hardware power. It is also mentioned in the article

that, it is given that users can talk to iPhone or Android device and there are a few speech engines out there that can do the same job but the main difference of Adapt from them is that those engines cannot be used by anyone. (Stahie, 2016)

According to the article from Network World, Raspberry Pi-based home AI project open-sources key components (2016), Mycroft.ai has opened an important part of that stack to developers everywhere. It is called Adapt Intent Parser, which is an intent parser library that is designed to be a powerful tool for converting human speech into machine data and it smoothly translates a verbal command into a usable set of instructions for a program. It will process simple voice commands, that can put media services like YouTube, Netflix, Pandora and many others at a user's fingertips together with smart home technology like SmartThings or Philips Hue. Adapt was launched in Kickstarter in August 2015, and reached its funding goal of \$99,00 during the original period. (Gold, 2016)

Based on the Website of Mimic by Mycroft and VocaliD, Mimic is a fast-lightweight text-to-speech engine based on Carnegie Mellon University's FLITE software. It takes the text in and reads it out loud to be able to create a high-quality voice. The Mimic's low-latency, small resource footprint and good quality voice set it apart from other open source text-to-speech projects. This is powerful tool that also help solve other problems. Mycroft partnered with VocaliD to help Dr. Rupal Patel and her team to bring realistic voices to people with speech disorders. Its technology allows the users to create customized voices that really represent the people who use it. Mimic works on Linux, Android & Windows and we are working on iOS support. This text-to-speech engine offers several voices that can use different speech modelling techniques such as diphone, clustergen and hts. It can differ a lot on size, naturalness and intelligibility.

According to the article, Mycroft AI Partners With VocaliD On Mimic Text To Speech Engine, Mycroft AI partnered with VocaliD to develop a high quality text-to-speech software for the open source community. The new TTS engine called Mimic, is based on Flite. In which Flite is an open source software that is developed at Carnegie Mellon University. VocaliD will be contributing their broad expertise in speech science to

the project, while Mycroft AI's software team works to improve the performance and portability of the TTS engine. VocaliD's technology allows them to create synthetic voices that are unique as fingerprints. This will allow their clients to speak with a personalized voice that fits their gender, age and background, in which something that until now has been impossible but with the use of Mimic engine, personalized voice service will enable clients to sound more natural than ever. Working together VocaliD and Mycroft are engaging the imagination of the open source community and collaborating on the world's fastest and most natural sounding text to speech engine.

According to the Microsoft Patterns and Practices about Message bus (2017), communication between applications usually creates dependencies between the applications. The sender must communicate with the receivers and the receiver should recognize the messages from all the given senders. Usually, the applications of an integration solution have different interfaces, but with changing the interfaces of proprietary applications is difficult. It is not feasible to change the interface for all the applications of the integration solution even one of the application's interface will be changed. There are some integration solutions that consists of a fixed set of applications. With an integration solution that has low extensibility and modifiability requirements typically does not need to accommodate new applications. To the provided problems presented, there is a solution to be used where all applications will be connected through a logical component known as Message Bus. It specializes in transporting messages between applications. The message bus contains three key elements such as, a set of agreed-upon message schemas, a set of common command messages and a shared infrastructure for sending bus messages to recipients. On the idea of Message Bus, the bus does not preserve message ordering. There is an internal optimization, routing, buffering or even the underlying transport mechanism might affect how the messages travel to the receiving applications. Then, the order in which messages reach each receiver is nondeterministic. To be able to preserve the order of messages, it requires additional logic. This logic could also be provided by the bus and the logic could therefore be transparent for the participating

applications. In which, an application that sends messages through the bus must prepare the messages so that the messages comply with the type of messages the bus expects. As a resulting context, there are benefits that was mentioned in using a Messages Bus. First is the improved modifiability, where each application has a single connection to the message bus instead of multiple dedicated connections to each of the other applications. Second is the reduced application complexity, where the messages bus will transport messages between senders and receivers. Third is, improved performance, in which there are no intermediaries between the communication applications. Last is the improved scalability, where adding applications has a constant low cost. The Figure 7.0 from the study shows the integration solution that uses a message bus.

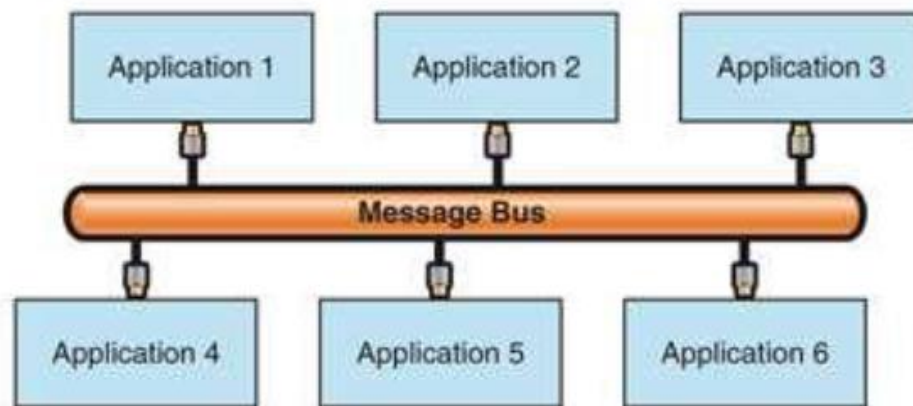


Figure 7.0 Applications communicating through a message bus

## **II. Related Studies**

### **Foreign**

#### **Analysis and Comparison of Intelligent Personal Assistants**

In the study made by Bahceci (2016), it aims to analyze and compare the current major implementations of IPAs in order to determine which implementation is the most developed as of today and is greatly contributing to the sustainable future of Artificial Intelligence. The scope of this study is restricted to the three different companies, namely, Apple Inc., Google Inc., and Microsoft Corporation.

The researcher mentioned that, “Intelligent Personal Assistants (IPA) and their current software implementations exists in a range of applications, usually integrated in the Operating System (OS) from different developers and organizations, such as in personal computers, mobile computers and in Internet of Things (IOT). “(Bahceci, 2016, p. 3)

The first IPA that was analyzed in this study was Apple’s Siri. It was introduced in the year 2011, and as early as 2016 it supports 17 natural human languages. It can control the following native applications coming with the operating systems: reminders, weather, stocks, messaging, email, calendar, contacts, notes, music, clocks, web browser, Wolfram Alpha and Apple Maps. The researcher mentioned that Apple created the Siri SDK together with the release of iOS 10. It was named, SiriKit, mainly to give the developers the ability to implement Siri usage in their own apps but in order to support SiriKit, application must support one of these following domains: VoIP calling, Messaging, Payments, Photo, Workouts, Ride booking, and CarPlay and Restaurants reservations which requires additional support from Apple.

Second IPA that was analyzed in this study was Google Assistant, previously names as Google Now. It was released on July 9, 2012. The researcher mentioned that little is known about Google Assistant’s system that makes them as a closed book. It was not known, what language its implemented in and what dependencies it has. Google’s IPA is

given that it is under closed development and there no such SDK or tools existent for accessing the IPA's features.

The last IPA that was analyzed and discussed was, Microsoft's Cortana. It is an IPA that is can access and change reminders, recognize natural voice, and answer questions using information from the search engine, Bing. It is the only personal assistant under discussion that is cross platform and presented an SDK for third party developers before the release of SiriKit. Microsoft has opened their SDK for Cortana for third party developers and gives much freedom for developers to integrate Cortana into their own applications, and even supports new actions. However, to do these stuffs, developers need to register first their actions, without any cost and will most likely be reviewed by a developer working for the company of Microsoft and Cortana Team.

### **Recommender Systems with Personality**

In the study conducted by Azaria & Hong (2016), it was stated that an intelligent agent must be personal, in which it does know its user's preferences and can recommend relevant content, it is a dynamic learner, instructable, supportive and affable. An IPA should be personal in terms of each should be personalized that it knows the user's preferences and can recommend relevant content, doing it in a privacy-sensitive manner. It should be a dynamic learner, in which it is learning and getting knowledge about the world and about the user. The IPA should be supportive, in terms of providing a long-term assistance, and as well as initiating actions in order to pursue user's goals and tasks. Affable, in which it offers interaction methods that appears natural to the user. Lastly, instructable, that will allow the user to customize its agent according to her personal needs.

It was also mentioned in this study that, current personal agents today are rigid, limited to predefined commands and not extendable by the end users. An example was given, "If Google Now finds a flight reservation in a user's email it will recommend the user to leave for the airport when it is time to do so. However, this capability is predefined (handcrafted) and the user cannot extend this functionality by requesting the agent to notify



her if she should leave for an important meeting located out of town.” (Azaria & Hong, 2016, p. 3)

### **“What can I help you with?”: Infrequent users’ experiences of Intelligent Personal Assistants**

According to the study of Cowan (2016), Intelligent Personal Assistants (IPAs) are widely available on devices such as smartphones. However, most people do not use them regularly. This paper is conducted with the use of focus groups where 20 participants, most of whom were infrequent users, discussed their experiences using an Intelligent Personal Assistants on a smart phone. The researchers focused their recruitment of participants on Siri users specifically. Given that Siri is the most commonly used IPA with 71% of participants in other previous studies, citing that they used Siri. The analyzation of the study focused on users’ views, everyday practices and barriers to use which resulted to six core problems such as, issues with supporting hands-free interaction, problems with performance with regards to user accent and speech recognition more widely, problems around integration with third party apps, platforms and systems, social embarrassment being a barrier to using mobile IPAs in public, the human-like nature of IPAs and issues of trust, data privacy, transparency and ownership. If these barriers cannot be addressed, developers may be best served in accepting that IPAs will remain a relatively niche application and focus on maximizing the experience of power users. The researchers believed that enhanced user experiences and widespread adoption is an achievable target.

Issues with speech recognition and strategies used when interacting with Siri was mentioned on the results of focus group findings. The participants of the study talked about the noticeable system improvements, but equally referred to examples of Siri misrecognizing what they have said. It was presented that several examples of how Siri still does not accurately pick up what they say especially when they are in noisy places. Also, many participants noted that there are issues of Siri not dealing well with accents that are local or non-native. One of the participants mentioned referring to Siri not accurately

picking up what they said, *“I think it does struggle with some of our idioms and our accents. It just doesn't understand. So, there's a few of the tasks that I had to repeat myself several times, or had to phrase things differently because it wasn't picking it up.”* (Cowan, 2017, p. 5) Given on the same context, some of the participants mentioned of having to think carefully about their words and phrasing before they will talk to Siri as pausing and single word editing were problematic for them. One of the participants mentioned, *“Yeah, when I was trying to edit it, it was just one word that it got wrong. I said "patterns" and it said like "patters," and so I was just trying to fix one word. I was asking my brother to send me a list of patterns, and I couldn't get Siri to just edit one word. I had to say the whole message again and I was like, this is not really good.”* (Cowan, 2017, p. 5)

Another issue mentioned on this study was about the Integration with application, platforms and systems. Participants across all focus groups discussed issues with respect to Siri's interoperability with third party apps, platforms and devices and expressed a strong preference that Siri becomes more customizable and better integrated. The lack of integration and ability to customize of participants' regular app usage habits was a key barrier in their eyes. It was found out that limiting the access of IPAs to only proprietary services and applications was seen as a significant barrier to the users.

### **About User Preference and Willingness to Pay for a Secure and Privacy Protective Ubiquitous Personal Assistant**

In this study, it aimed to contribute to the existing research by assessing the users' preferences and willingness to pay for a highly secure and privacy stringent UPA. The researchers described the concept of Intelligent Personal Assistant as a computer system that interacts with a user via voice recognition, can gather and evaluate contextual information and has the ability to provide assistance.

It was mentioned in this study that existing intelligent Assistants like Siri, Cortana, Google Now, Sirius and Alexa are sophisticated and excellent at what they do, but they

remain specialized and very limited in their support capabilities. (Wilson, Zibuschka & Hinz, 2017)

### **End-user Interactions with Intelligent and Autonomous Systems**

In the study made by Stumpf, et al. (2012), which focuses on the approaches and challenges to explore making these personalized systems transparent, controllable and ultimately trustworthy to end users. It aims to help in building connections among different researchers, as well as industrial practitioners using real-world problems to facilitate the exchange of approaches and ideas about how to make better support on end users.

It was mentioned in this study that, many consumers and business specialists often interact with systems nowadays in a form of “human in the loop” learning, but interacting with these even well-designed systems is limited and often uninformative for the end users due to internal complexity and the black box nature of most intelligent systems.

### **Short Introduction to Text-to-Speech Synthesis**

This paper discusses that a Text-to-Speech (TTS) synthesizer is a computer-based system that should be able to read any text aloud, whether it was directly introduced in the computer by an operator or scanned and submitted to an Optical Character Recognition (OCR) system. Let us try to be clear. There is a fundamental difference between systems we are about to discuss here and any other talking machine in the sense that we are interested in the automatic production of new sentences. This definition still needs some refinements. Systems that simply concatenate isolated words or parts of sentences, denoted as Voice Response Systems, are only applicable when a limited vocabulary is required, and when the sentences to be pronounced respect a very restricted structure, as is the case for the announcement of arrivals in the train stations for instance. In the context of TTS synthesis, it is impossible to record and store all the words of the language. It is thus more suitable to

define Text-to-Speech as the automatic production of speech, through a grapheme-to-phoneme transcription of the sentences to utter.

### **Flite: A Small Fast Run-Time Synthesis Engine**

According to this study of Black & Lenzo (2011), it states and defines Flite as small, fast run-time synthesis library suitable for embedded systems and servers, Flite is designed as an alternative run-time synthesis platform for Festival in applications where speed and size are important. Voices built using the FestVox process may be compiled into efficient representations that can be linked against Flite to produce complete text-to-speech synthesizers. The Flite library is much faster and much smaller than equivalent Festival system. This paper describes the motivations and the basic structure of the library, and gives figures of its sizes and speed. Some intended enhancements are also discussed.

### **Message Bus and Distributed Technology**

According to the study of Albrecht, Hook & Bushouse, applications have become increasingly large and monolithic during the recent years despite of successful efforts to structure software internally at the class library level. Facilities are needed to allow components from different data systems, which may be very different internally, to be combined to create heterogenous applications. With the core technologies needed to achieve flexibility are the Message Bus, distributed objects and applications framework. In this paper, the concepts of the message bus and distributed objects were introduced. It was mentioned that the size of the programs and their relatively high level, user oriented interface makes them inflexible to use and construct new applications. The high degree of the integration characteristic of programming at the class library level makes it hard to construct heterogenous applications that use modules from different system. Addressed these problems, there are technologies that are being developed to overcome this. It is the Distributed Objects, that allows major software modules to be represented as objects which can be used either stand-alone or as components of distributed applications. To connect it all together, message bus is the answer. Message Bus provides flexible services and methods

for distributed objects to communicate with another and share data. The concept of Message bus is illustrated in Fig 8.0 presenting that message bus is a facility used to bind, at runtime, distributed objects (program components) to construct applications.

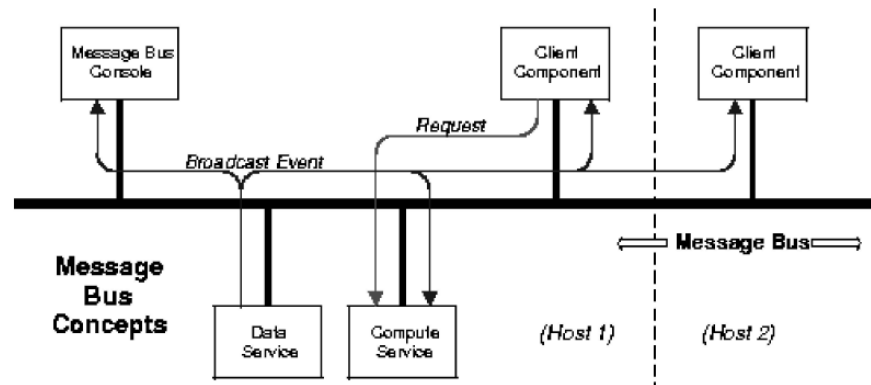


Fig 8.0 Message Bus Concept

### **Bridging the Communication Gap: Conceptualizing Issues of Learnability in Using Intelligent Personal Assistants**

In this paper made by Chen & Wang (2017), mentioned that Intelligent Personal Assistant became a part of people's daily life. It helps the user to access information in a timely manner thru dialog interfaces but current designs may not fulfill users' expectation because of people's limited understanding of the capabilities and constraints of intelligent personal assistant.

The researchers mentioned that there are situations where IPA misunderstood the users' speech input, that leads to giving feedback that is not correct. If IPAs processed users' intention inaccurately, then it would result to the user being frustrated because of the failure. And in order to meet users' expectation, the researchers noted that communication between humans and IPAs shouldn't be an exception, and common ground is still needed to ensure satisfactory experience out of the interaction.

Another problem mentioned by the researchers in this paper, is that IPAs give only general feedback without even providing further information when an error occurs or

when the system detects it. IPAs often lack clues, in which users' need to learn how to communicate with IPAs. These feedback serves as flexible scaffolds rather than coercive instructions, that aims to preserve perceived controllability of users.

### **Understanding User Satisfaction with Intelligent Assistants**

In the study made by Kiseleva, et al. (2016), the researchers described a user study that was designed to measure user satisfaction over a range of typical scenarios of use: controlling a device, web search, and structured search dialog. The researchers used this data to study how user satisfaction varied with different usage scenarios and what signals can be used for modeling satisfaction in different scenarios. They found that the notion of satisfaction varies across different scenarios and show that, in some scenarios (e.g. making a phone call), task completion is very important while for others (e.g. planning a night out), the amount of effort spent is key. The researchers also studied about how the nature and complexity of the task at hand affect user satisfaction, and found that preserving the conversation context is essential and that overall task-level satisfaction cannot be reduced to query-level satisfaction alone.

It was mentioned in this study that users are more satisfied with the simple task and less satisfied with more complex task. It only shows that the complexity of context in structured search dialogs, when viewed independently the quality of the results is comparable for each step of the interaction, and the high levels of satisfaction for the simple task confirm that the quality is high, yet the satisfaction levels go down considerably when tasks are of increasing complexity and suggests that intelligent assistant loses context of a conversation, requiring more effort and interaction to restart the dialog and get back on the track.

## **Perspectives for Evaluating Conversational AI**

According to the study made by Jadeja and Varia (2016), defining success in the case of conversational AI is difficult and need to use some appropriate metrics for the evaluation of conversational AI. The researchers proposed four different perspectives namely user experience, information retrieval, linguistic and artificial intelligence for the evaluation and provide background details of conversational AI systems containing desirable characteristics of personal assistants, differences between chatbot and an AI based personal assistant. Personalization importance and how it can be achieved is explained in detail. The researchers also highlight the current challenges in the development of an ideal conversational AI (personal assistant) along with guidelines for achieving personalized experience for users.

It was stated in this study that Personal Assistants could initially recognize only certain commands and words but the mammoth task of identifying natural language completely is an important. The provision of vocabulary to the personal assistants, practical language recognition based on local slangs and accents is the challenges faced initially. The challenge we have to face is the understanding of context and reference based on history and previous data apart from natural language.

## **Multimodal Interaction and Believability: How can we design and evaluate the next generation of IPA?**

In the study conducted by Li and Lee (2017), believability of Intelligent personal assistants (IPA) has proven to be an important building block of successful human-agent interaction. Yet, only a handful of studies have focused on proposing and validating possible approaches to enhance such believability. The researchers hypothesize that IPAs that are capable of handling multimodal interaction, such as facial expression and hand gestures, can appear more believable to human users. The study illustrated some reasons why such interaction can improve believability, in turn enhancing user interpersonal

rapport with agent. It also discusses design of a study to evaluate believability in human-agent interactions.

In this study, it was discussed that IPAs can interact with a user through only verbal channel, commonly referred to as Voice User Interface (VUI). The drawback of using only VUIs in IPAs are obvious in three aspects: (1) Conversational Content: IPAs are incapable of conveying emotion, varying their social response, and reacting to users emotional state; (2) Performance: Responses from IPAs demonstrate poor variety of social behavior, preventing users to engage in a long-time interaction with them; (3) Function: IPAs are not good at comprehending users intent and can only carry out limited conversation spanning one or a few turns. Such conversations mostly take the form where question from a user triggers a response from the agent, and known facts from prior interactions are disregarded in subsequent IPA responses. These shortcomings of VUI demonstrate that verbal interactional one is insufficient for successful human- to-agent interaction. IPAs often lack necessary input to determine users current emotional state and need.

### **Factors Affecting User's Information Requests**

According to the study of Arguello, et al., conversational search interfaces have two important characteristics. First is they can accept voice requests from users and second they aim to provide users with more human-like interactions. The researchers study the effects of two factors (medium and target) on: (RQ1) participant's perceptions about their information requests, (RQ2) the different characteristics of their information requests, and (RQ3) participant's strategies when requesting information appropriate for a novice or expert. The result shows that both factors had a strong effect on participants' requests (including retrieval performance), and that target had a stronger effect on participant's perceptions and their choice of strategy in requesting novice-or expert-appropriate information. The researchers also analyze information requests collected using crowdsourcing and address three research questions.



In this study, participants reported less difficulty and greater satisfaction and confidence in the human versus search engine condition (i.e., when they had greater expectations about the system's ability to respond to requests). Second, while the interaction effect was not significant, participants reported being the least confident and satisfied in the voice and search engine (V-S) condition. In terms of RQ2, the results suggest three main trends. First, consistent with previous comparisons between voice and textual queries the spoken requests were longer, had more complex grammatical structure (resembling natural language), and yielded worse retrieval results when issued unmodified to a commercial web search API. Second, the same trend was true for requests intended for a human versus search engine. Finally, both factors had a strong additive effect—information requests were the longest, most complex, and yielded the worst retrieval results in the voice and human condition (V-H). In terms of RQ3, the results suggest four main trends. First, our qualitative analysis found six strategies adopted by participants where requesting information for a novice or expert. Second, target had a greater effect than medium in influencing participants to use certain strategies. When requesting information from a human versus a search engine, participants were more likely to mention the complexity of the information desired, more likely to describe the purpose of the information (not significant), more likely to describe undesired information (not significant), and more likely to use a combination of strategies. Third, all strategies hurt performance when the request was submitted unmodified to a commercial web search API. Finally, the strategies associated with the worst retrieval performance (i.e., 'purpose', 'complexity', and 'not want') were more common in the human versus search engine condition.

### **Comparative Analysis of Mobile Phone Application Solutions Accessibility for Informing Visually Impaired Persons in traffic Environment**

In the study made by Perisa, et al. (2016), which focuses on comparative analysis of operating systems(OS) and the accessibility of mobile phones. The research shows what

kind of mobile phone, depending on OS, is accessible for visually impaired persons and point to the problems which still exist in available solutions. The purpose is to give necessary information to visually impaired on suitable mobile phone solutions and simple choice of OS for simpler usage and informing in traffic network.

It was stated in the study that virtual assistants have issues because of language limitations and speech misunderstanding. A quantitative analysis of using Amazon Echo, Apple Siri and Google services was also conducted. It was stated that there is a slightly 80% of the respondents were very satisfied on how virtual assistants recognize the voice of the user and how accurate the feedback information was.

### **Instructable Intelligent Personal Agent**

In the study conducted by Azaria, et. al (2016), humans often learn from natural language instruction. As users become increasingly accustomed to interacting with mobile devices using speech, their interest in instructing these devices in natural language is likely to grow. The researchers introduced Learning by Instruction Agent (LIA), which is an intelligent personal agent that users can teach to perform new action sequences to achieve new commands, using solely natural language interaction. LIA uses a CCG semantic parser to ground the semantics of each command in terms of primitive executable procedures defining sensors and effectors of the agent.

In this study, it was stated that the success rate of the parsing and execution, out of a total of 12,654 commands, 15.4% (1,952) were parsed to an unknown command (which in most cases is a parse failure), and an additional 5.4% (679 commands) resulted in an execution error (e.g. trying to change an immutable field or setting an email to a non-email value). When considering only those who completed the tasks, this error rate drops to 9.9% (659 out of 6,649) of unknown commands, and 4.9% of execution errors. This improvement is expected, since those who completed all tasks interacted longer with LIA and thus the researchers more likely to provide commands that LIA understood. Furthermore, the subjects who provided commands that LIA could not execute were more

likely to drop out early. In addition to unknown commands and unsuccessful executions, there were also undesired executions in which LIA executed an action that differed from the subject's intent. This problem can be caused either by a parse error (this happened several times when a subject wanted to set a to b's value, but said set a as b, and, if both a and b were mutable fields, b was set to a's value, due to the more common meaning of the token "as") or human error (e.g. "set recipient to inbox email's recipient", where the subject really meant to use the sender of the email in the inbox).

### **Managing Uncertainty in Time Expressions for Virtual Assistants**

In the study made Rong, et. al (2017), that in communications and planning, people often express uncertainty about time using imprecise temporal expressions (ITEs). Unfortunately, modern virtual assistants often lack system support to capture the intents behind these expressions. This can result in unnatural interactions and undesirable interruptions (e.g., having a work reminder delivered at 12pm when out at lunch, because the user said "this afternoon"). In this study it explores the existing practices, expectations, and preferences surrounding the use of ITEs. The researchers used mixed methods approach which includes surveys, interviews, and an analysis of a large corpus of written communications. The researchers found that that people frequently use a diverse set of ITEs in both communication and planning. These uses reflect a variety of motivations, such as conveying uncertainty or task priority. In addition, the researchers found that people have a variety of expectations about time input and management when interacting with virtual assistants.

It was stated in this study that several situations arose in which the language was complex or ambiguous. For example, a specific participant recorded the memo "Follow up with (a friend) about dinner plans either Saturday or Sunday." Upon seeing the item was scheduled for display on the following Saturday, the participant remarked that the interpretation was incorrect, then explained: "It's not like I was telling it to mark it for Saturday or Sunday, I was telling it I need to follow up with her about Saturday or Sun-

day.” Likewise, upon inputting “Buy tickets for the concert next month,” the other participant was disappointed the system misinterpreted “next month” as the time when tickets should be purchased, rather than the time at which the concert was occurring.

### **“Do Animals Have Accents?”: Talking with Agents in Multi-Party Conversation**

According to the study made by Porcheron, et al. (2017), the researchers explored the use of conversational agents, or so-called intelligent personal assistants (IPAs), in multi-party conversation amongst a group of friends while they are socializing in a café. The researchers examined the use of these IPAs in a mundane and common-place setting and employ an ethnomethodological perspective to draw out the character of the IPA-use in conversation. Additionally, the researchers highlight a number of nuanced practicalities of their use in multi-party settings. By providing a depiction of the nature and methodical practice of their use, we are able to contribute our findings to the design of IPAs.

It was stated in this study that in order to understand the failure to complete a query, members repeat or in some cases refine their query but very few times they do abandon the query. Their findings revealed that IPAs have difficulty to understand synonyms and homonyms in talk. IPAs presently provide limited functionality for conversational repair; if a device has not understood a phrase, it should ask people “could you ask your query using different words?” or, perhaps when a word is not recognized, “could you spell that?”, alleviating some of the identified problems.

### **“Like Having a Really bad PA”: The Gulf between User Expectation and Experience of Conversational Agents**

In the study made by Luger and Sullen (2016), the study showed the findings of interviews with 14 users of conversational agents (CAs) in an effort to understand the current interactional factors affecting everyday use. The researchers found that user expectations dramatically out of step with the operation of the systems, particularly in terms of known machine intelligence, system capability and goals. The researchers

consider the implications of these findings for the design of future systems using Norman's gulfs of execution and evaluation.

It was mentioned in the study that the operation of the CA systems failed to bridge the gap between user expectation and system operation. The study showed that users had poor mental models of how their CA worked and that these were reinforced through a lack of meaningful feedback regarding system capability and intelligence. Equally, where playful aspects and trigger responses were programmed into the systems, these served to act as engagement mechanisms, whilst concurrently setting unrealistic expectations that framed the ongoing user experience.

### **Prediction of Prospective User Engagement with Intelligent Assistants**

According to the study made by Sano, et al. (2016), intelligent assistants on mobile devices, such as Siri, have recently gained considerable attention as novel applications of dialogue technologies. A tremendous amount of real users of intelligent assistants provide us with an opportunity to explore a novel task of predicting whether users will continually use their intelligent assistants in the future. The researchers developed prediction models of prospective user engagement by using large-scale user logs obtained from a commercial intelligent assistant. Experiments demonstrated that our models can predict prospective user engagement reasonably well, and outperforms a strong baseline that makes prediction based past utterance frequency.

In this study, it was the features of the response frequency was described. It was described that the number of lengthy responses (more than 50 characters long) and the number of error messages decrease user engagement. Because longer responses require a longer reading time, they are prone to irritate users and consequently decrease user engagement and voice assist returns error messages (Sorry, I don't know.) when it fails to find appropriate responses to the user's utterances.

## **Employing Apple's Siri to practice pronunciation: A preliminary study on Arabic speakers**

According to the study made by Molden (2015), which reports on a preliminary study on Arabic Speakers using Siri to practice pronunciation and aims to uncover Siri's positive affordances for language learning in relation to its corrective feedback and interpretation of speech. The study reports on a small-scale study of how two Arabic learners of English as a second language used Siri to practice English pronunciation. The researcher discussed Siri's positive and negative affordances in relation to its accessibility; its base in intelligibility, its context dependency, its feedback, its validity, and its negotiation for meaning.

It was mentioned in this study that Siri provides feedback based on intelligibility, but the quality of this feedback may be up for question; and, even though certain communication breakdowns seem to foster changes in pronunciation, these breakdowns may also originate from other factors than pronunciation deficits. It is important to keep in mind, however, that Siri was never intended to be used as a language learning tool; it was its need for comprehensible input that united the developers' and language learners' purpose.