CHAPTER I

**THE STUDY AND ITS PROBLEMS**

**A. BACKGROUND OF THE STUDY**

The artificial intelligence, voice recognition, natural language processing and learning machines space has been growing rapidly. The idea of having a personal assistant which connects us to the web and the ever growing Internet of Things is becoming ever more commonplace.

An intelligent personal assistant (or simply IPA) is a software agent that can perform tasks or services for an individual. These tasks or services are based on user input, location awareness, and the ability to access information from a variety of online sources (such as weather or traffic conditions, news, stock prices, user schedules, retail prices, etc.). Examples of such an agent are Apple's Siri, Google's Google Now, Amazon Alexa, Microsoft's Cortana, Braina (application developed by Brainasoft for Microsoft Windows), Samsung's S Voice, LG G3's Voice Mate, BlackBerry's Assistant, SILVIA, HTC's Hidi, IBM's Watson (computer), Facebook's M (app) and One Voice Technologies (IVAN).

According to venture capitalist Chi-Hua Chien of Kleiner Perkins Caufield & Byers, examples of tasks that may be performed by a smart personal agent-type of Intelligent Automated Assistant include schedule management (e.g., sending an alert to a dinner date that a user is running late due to traffic conditions, update schedules for both parties, and change the restaurant reservation time) and personal health management (e.g., monitoring caloric intake, heart rate and exercise regimen, then making recommendations for healthy choices).

Intelligent personal assistant technology is enabled by the combination of mobile devices, application programming interfaces (APIs), and the proliferation of mobile apps. However, intelligent automated assistants are designed to perform specific, one-time tasks specified by user voice instructions, while smart personal agents perform ongoing tasks (e.g., schedule management) autonomously.

Simply put, artificial intelligence is a sub-field of computer science. Its goal is to enable the development of computers that are able to do things normally done by people -- in particular, things associated with people acting intelligently.

Stanford researcher [John McCarthy](http://en.wikipedia.org/wiki/John_McCarthy_%28computer_scientist%29) coined the term in 1956 during what is now called [The Dartmouth Conference](http://www.livinginternet.com/i/ii_ai.htm), where the core mission of the AI field was defined. However, the five decades since the inception of AI have brought only very slow progress, and early optimism concerning the attainment of human-level intelligence has given way to an appreciation of the profound difficulty of the problem. Artificial Intelligence are designed to perceive human mind or cognitive functions. It is designed to think as something stated. It response as if you are communicating with someone. Today, artificial intelligence is widely used in medical diagnosis, robotics, games and etc. It is commonly used because Artificial Intelligence is very useful in their respective fields. Mainstream thinking in psychology regards human intelligence not as a single ability or cognitive process but rather as an array of separate components. Research in AI has focused chiefly on the following components of intelligence: learning, reasoning, problem-solving, perception, and language-understanding. But analysing use data and providing recommendations like that is relatively easy, and it doesn’t quite give the feel of a “personal assistant.” The real task – and the real difficulty – for IPAs is recording, understanding, and effectively responding to human speech.

While we take them for granted, our hearing and language systems are remarkable feats of computation

While understanding language is relatively easy for most four-year-olds, decoding human speech is a remarkably complex task. In fact, decoding a verbal request requires more than 100 times as much processing power as responding to a textual search request. This is because while our ears have been honed by millions of years of evolution to pick up and decode human speech, IPAs only have built-in microphones, and these microphones do little to delineate human speech from surrounding noise.

Artificial Intelligence plays an important role in today’s technology. Researchers and Developers believe that the appropriateness of Artificial Intelligence used to a certain level may offer a chance for man to communicate without any restrictions and make computing an innovative and futuristic success.

**B. ALGORITHMS**

**1.) Hidden Markov Model Statistical Algorithm for Speech Recognition**

A hidden Markov model (HMM) is a statistical Markov model in which the system being modelled is assumed to be a Markov process with unobserved (hidden) states. An HMM can be presented as the simplest dynamic Bayesian network. The mathematics behind the HMM were developed by L. E. Baum and co-workers. It is closely related to an earlier work on the optimal nonlinear filtering problem by Ruslan L. Stratonovich, who was the first to describe the forward-backward procedure.

In simpler Markov models (like a Markov chain), the state is directly visible to the observer, and therefore the state transition probabilities are the only parameters. In a hidden Markov model, the state is not directly visible, but the output, dependent on the state, is visible. Each state has a probability distribution over the possible output tokens. Therefore, the sequence of tokens generated by an HMM gives some information about the sequence of states. The adjective 'hidden' refers to the state sequence through which the model passes, not to the parameters of the model; the model is still referred to as a 'hidden' Markov model even if these parameters are known exactly.

Hidden Markov models are especially known for their application in temporal pattern recognition such as speech, handwriting, gesture recognition, part-of-speech tagging, musical score following, partial discharges and bioinformatics.

A hidden Markov model can be considered a generalization of a mixture model where the hidden variables (or latent variables), which control the mixture component to be selected for each observation, are related through a Markov process rather than independent of each other. Recently, hidden Markov models have been generalized to pairwise Markov models and triplet Markov models which allow consideration of more complex data structures and the modelling of nonstationary data.

**2.) Stanford Pattern Based Information Extraction and Diagnostics**

Bootstrapped systems have been commonly used to learn entities (Riloff, 1996; Collins and Singer, 1999). SPIED-Learn is based on the system described in Gupta and Manning (2014), which builds upon the previous bootstrapped pattern learning work and proposed an improved measure to score patterns. It learns entities for given classes from unlabelled text by bootstrapping from seed dictionaries. Patterns are learned using labelled entities, and entities are learned based on the extractions of learned patterns. The process is iteratively performed until no more patterns or entities can be learned. The following steps give a short summary of the iterative learning of entities belonging to a class DT:

1. Data labelling: The text is labelled using the class dictionaries, starting with the seed dictionaries in the first iteration. A phrase matching a dictionary phrase is labelled with the dictionary’s class.

2. Pattern generation: Patterns are generated using the context around the positively labelled entities to create candidate patterns for DT.

3. Pattern learning: Candidate patterns are scored using a pattern scoring measure and the top ones are added to the list of learned patterns for DT. The maximum number of patterns learned is given as an input to the system by the developer.

4. Entity learning: Learned patterns for the class are applied to the text to extract candidate entities. An entity scorer ranks the candidate entities and adds the top entities to DT’s dictionary. The maximum number of entities learned is given as an input to the system by the developer.

5. Repeat steps 1-4 for a given number of iterations.

**3.) Flite Text to Speech Synthesis Algorithm**

Flite is a small, fast run-time synthesis library suitable for embedded systems and servers. Flite is designed as an alternative for Festival in applications where speed and size are important. Voices are 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 algorithm is much faster and much

smaller than the equivalent Festival system. Flite is the core library. For synthesis, this library requires three further three parts to make a complete synthesizer. language model: providing phoneset, tokenization rules, text analysis, prosodic structures etc. lexicon: a pronunciation model including a lexicon and letter to sound rules for out of vocabulary words. Voice- A voice depends upon the primitives provided by the language model. The first two of these can be shared across voices of the same language. Each of these subsection is compiled into

separate libraries. Each voice definition in Fest Vox include specific tokenization rules and prosody rules. However, we can provide the basic tools. For basic diphone voices for known languages and simple generic limited domain voices built using the FestVox build model. Each word in the lexicon is converted to a list of letters (plus part of speech) and held in a sorted table, and each entry has an index into a list of phones (with lexical stress marked on vowels).

**C. STATEMENT OF PROBLEMS**

1.) No Intelligent Personal Assistant Development Platform exists that caters the fundamental human - computer interaction processing that is openly available for the public.

2.) APIs and Services may not communicate very well when developed in an isolated environment without setting a proper foundation where all of them can work together seamlessly, thus limiting the resources and databases accessed by assistants and restricting the quality of IPA.

3.) Proprietary IPAs does not let users gain direct access to their own data and since most of your personal data are stored in remote server, should attackers gain control of this, they could use that information to gain access to your personal accounts.

4.) Developing your own IPA can be a very expensive job, since most of the API and technologies are proprietary and are “closed books” owned by the giant players in the field.

5.) Most proprietary IPAs are limited and are scoped only depending on the owner company’s discretion, and gives users 0% to fully customize and make IPAs more “Personal”.

6.) IPAs provided by a specific company forces users to use only a specific kind, type or brand of systems, libraries and dependencies, thus limiting room for further developments and enhancements.

**D. OBJECTIVES OF THE STUDY**

As user demand scales for intelligent personal assistants (IPAs) such as Apple’s Siri, Google’s Google Now, and Microsoft’s Cortana, we are approaching the computational limits of current datacenter architectures. It is an open question how future server architectures should evolve to enable this emerging class of applications, and the lack of an Open Source IPA Development Platform workload is an obstacle in addressing this question.

In this paper, we present the design of FORG, an open end-to-end IPA Development Platform used to build, code and extend a user’s own customized IPA with features such as Speech to Text Processing, Natural Language Processing, Text to Speech, and API and Web Services Compatibility.

We aim to give developers and enthusiast alike an algorithm or platform that will enable them to build, customize and redistribute their own Intelligent Personal Assistant.

**E. IMPORTANCE OF THE STUDY**

Intelligent Personal Assistants is the future of computing and further development of technologies in this area will surely be the trend for the future. It was one of the strongest fields of development during 2015 and it will also be in 2016 and will certainly be one of the sectors generating most profits for years or decades to come.

1. Intelligent machines can replace human beings in many areas of work. Robots can do certain laborious tasks. Painstaking activities, which have long been carried out by humans can be taken over by the robots. Owing to the intelligence programmed in them, the machines can shoulder greater responsibilities and can be programmed to manage themselves.
2. Smartphones are a great example of the application of artificial intelligence. In utilities like predicting what a user is going to type and correcting human errors in spelling, machine intelligence is at work.
3. Artificial intelligence can be utilized in carrying out repetitive and time-consuming tasks efficiently.
4. The greatest advantage of artificial intelligence is that machines do not require sleep or breaks, and are able to function without stopping. They can continuously perform the same task without getting bored or tired. When employed to carry out dangerous tasks, the risk to human health and safety is reduced.

**F. SIGNIFICANCE OF THE STUDY**

This study seeks to benefit the following groups of people:

1. To the ***computer and smartphone users***, to make computing a more interesting and worthwhile experience through the use and utilization of an A.I. Agent for automation and to ease them of the burden of executing certain computing tasks.
2. To the ***programmers and developers,*** this study will encourage them to partake in the improvement of Artificial Intelligence and to advance and develop applications and system that utilizes this technology.
3. To ***business owners and planners,*** A.I. is the current trend of today’s technology. It exists now in almost every device. This study can help them to take their business ideas to the next level through the use of A.I. agents to cater customer’s needs.
4. Lastly, to the ***future researchers,*** that they will further this study through investing more time in finding out the effectiveness and relevance and to what extent can A.I. help in today’s computing technologies.

**G. SCOPE AND LIMITATIONS**

This study focuses on the development of a Development Platform for building Intelligent Personal Assistants based from three core algorithms in the field of human – computer interaction. With this algorithm or platform, IPAs can be extended to include Machine Learning, Computer, Tasks, and Home Automation, Problem Solving and Queries. Software, APIs and SDKs that will be used in the development of this project are either open-source or licensed technologies. Additionally, functions and features of this project only covers topic Artificial Intelligence, Computational Knowledge, Machine Learning and Cognitive Perceptions in the field of Computer Programming and Development and not topics directly related to Medicine, Engineering and other subject areas not mentioned or related above.

This study will not cover or in any way will involve proprietary technologies and features of similar Intelligent Personal Assistants such as Microsoft ‘s Cortana, Apple’s Siri, IBM’s Watson, Amazon’s Alexa and Google’s Google Now, etc. Any similarity among these systems within this study is not intended and may be incidental.

**H. DEFINITION OF TERMS**

**Acoustic Model** is a set of words or sentences a Speech to Text Engine can understand.

**Actuator is a** Module or Code that executes a specific Intent.

**Application Program Interface (A.P.I)** is a set of routines, protocols, and tools for building software applications.

**Artificial Intelligence (A.I.)** is the simulation of human intelligence processes by machines.

**Assistant** is the term used for the Artificial Intelligence Entity, can be interchanged to Agent.

**Belief** is the knowledge base of the Agent about the world.

**Belief-Desire-Intention (B.D.I.)** is a software model developed for programming intelligent agents.

**Desire** is the current goal of the Agent in relation to the Belief.

**Environment** means the real world where the Agent takes input from.

**Intelligent Personal Assistant (IPA)** is a software agent that can perform tasks or services for an individual.

**Intention** is the method of action or the Agent will execute based from the Belief and Desire.

**Internet of Things** is a proposed development of the internet in which everyday objects have network connectivity, allowing them to send and receive data.

**Natural Language Processing (N.L.P.)** is a field of computer science, artificial intelligence, and computational linguistics concerned with the interactions between computers and human.

**Raspberry Pi** a brand of a programmable single computer board with ARM Architectures.

**Speech to Text (S.T.T.)** is a type of software engine that transcribes spoken speech into text.

**Text to Speech (T.T.S.)** is a type of software engine that speaks text in a computer voice.