

**SAYA FREE SPEECH**  
**RECOGNITION MINI PROJECT**

**AVIAD OTMAZGIN**  
**AMIR BARON**

## **INTRODUCTION**

Our mini project handles with the speech recognition part on saya.

Currently, saya can recognize only a small vocabulary of approximately 40 words.

Our mini project's target is to allow saya to do "free speech recognition", meaning that she would be able to recognize any word that was spoken to her.

## **GENERAL COMPONENTS**

Our project contains three main parts: the JAVA SPEECH API, CLOUDGARDEN and DRAGON NATURALLY SPEAKING software.

## **JAVA SPEECH API (JSAPI)**

The Java Speech API defines a standard, easy-to-use, cross-platform software interface to state-of-the-art speech technology. Two core speech technologies are supported through the Java Speech API: *speech recognition* and *speech synthesis*. Speech recognition provides computers with the ability to listen to spoken language and to determine what has been said. In other words, it processes audio input containing speech by converting it to text.

The Java Speech API was developed through an open development process. With the active involvement of leading speech technology companies, with input from application developers and with months of public review and comment, the specification has achieved a high degree of technical excellence. As a specification for a rapidly evolving technology, Sun will support and enhance the Java Speech API to maintain its leading capabilities.

The Java Speech API is an extension to the Java platform. Extensions are packages of classes written in the Java programming language (and any associated native code) that application developers can use to extend the functionality of the core part of the Java platform.

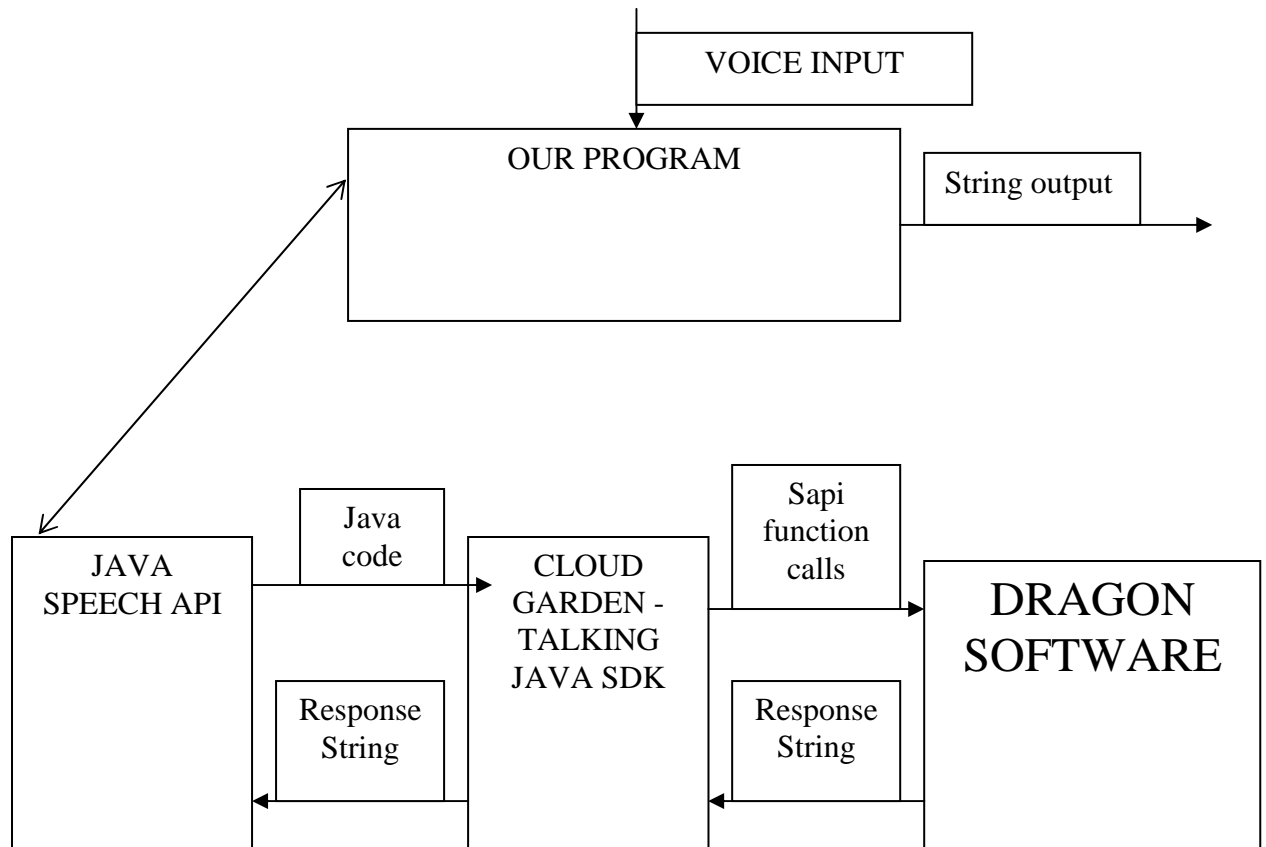
### **CLOUDGARDEN**

CloudGarden has produced a full implementation of [Sun's Java Speech API](#) for Windows platforms, allowing a large range of SAPI4, SAPI5 and DRAGON NATURALLY SPEECH compliant Text-To-Speech and Speech-Recognition engines (in many different languages) to be programmed using the standard Java Speech API.

### **DRAGON NATURALLY SPEAKING**

Dragon naturally speaking is speech recognition software. It contains speech recognition engine based on SAPI4 of Microsoft, supported by cloud garden. In other words, cloud garden uses the DRAGON software to implement the JSAPI.

## PROJECT OVERVIEW (schematics)



## **INTRODUCTION TO SPEECH RECOGNITION**

Speech recognition is the process of converting spoken language to written text or some similar form.

The major steps of a typical speech recognizer are:

- *Grammar design*: recognition grammars define the words that may be spoken by a user and the patterns in which they may be spoken. A grammar must be created and activated for a recognizer to know what it should listen for in incoming audio. Grammars are described below in more detail.
- *Signal processing*: analyze the spectrum (frequency) characteristics of the incoming audio.
- *Phoneme recognition*: compare the spectrum patterns to the patterns of the phonemes of the language being recognized.
- *Word recognition*: compare the sequence of likely phonemes against the words and patterns of words specified by the active grammars.
- *Result generation*: provide the application with information about the words the recognizer has detected in the incoming audio. The result information is always provided once recognition of a single utterance (often a sentence) is complete, but may also be provided during the recognition process. The result always indicates the recognizer's best guess of what a user said, but may also indicate alternative guesses.

There are two ways to work with the engine of speech recognition: the first one, is using the rule grammar technique and the other is to use the dictation grammar technique as described below:

### **Dictation Grammar VS Rule Grammar**

#### **RULE GRAMMAR:**

In a rule-based speech recognition system, an application provides the recognizer with rules that define what the user is expected to say. These rules constrain the recognition process. Careful design of the rules, combined with careful user interface design, will produce rules that allow users reasonable freedom of expression while still limiting the range of things that may be said so that the recognition process is as fast and accurate as possible.

#### **DICTATION GRAMMAR:**

Dictation grammars impose fewer restrictions on what can be said, making them closer to providing the ideal of free-form speech input. The cost of this greater freedom is that they require more substantial computing resources, require higher quality audio input and tend to make more errors.

A dictation grammar is typically larger and more complex than rule-based grammars. Dictation grammars are typically developed by statistical training on large collections of written text.

## **PROJECT DESCRIPTION AND PROGRAMMING ISSUES**

Our main goal was to create free speech recognition software that will replace the current non-free speech recognition software on saya's use.

In order to achieve this goal, we decided to use dictation grammar. Moreover, at start we used the SAPI5 engine of Microsoft but the accuracy was very low.

After searching for alternatives, we found new software called DRAGON NATURALLY SPEAKING (described above). The use of the software's engine allowed us to create more powerful accurate free speech recognition software.

In addition, the software allows restriction of the size of the vocabulary, which improves massively the accuracy of the recognition.

# **GENERAL MANUAL**

## **Installation:**

- Install the DRAGON software.
- Download and install CLOUDGARDEN.
- Create a new user in DRAGON in the following way:
  1. In the DRAGON toolbar choose NaturallySpeaking → manage users.
  2. click "browse" and select the folder for saving user files.
  3. Click "new". In the "name" section click the vocabulary name and make sure that "skip initial training of this user" marked.
    - \* you can create a new empty vocabulary by clicking "advanced" → "vocabulary size" and then choose empty.IMPORTANT: in order to use the software features your vocabulary must include the next words: "finish", "change vocabulary" and all of the other vocabularies names available.
  4. Click next and follow the instructions.

## **Managing the vocabulary:**

- If you wish to add/remove words from your vocabulary, you can do that by the following way:

Open the installation folder on the hard drive, and choose "voctool.exe" and follow the further instructions.

You can choose to include words from a specific text file on your vocabulary through this tool.

**Activating the java free speech recognition software:**

First, you must make sure that the DRAGON software is closed. Now you can activate the program and start talking.

- If you want to change the vocabulary on real-time speak: "change vocabulary" and wait for reaction. Then, say the name of the vocabulary you want to change to and wait for response. You will get a message that the vocabulary was changed to the specified vocabulary.
- If you wish to terminate the program, say the exit command : "finish".

## **FUTURE DEVELOPMENT IN SPEECH RECOGNITION**

Today, using the free speech recognition engine DRAGON NATURALLY SPEAKING is very successful for limited vocabulary. The shorter the vocabulary is, the most accurate the speech recognition is.

Large vocabulary (general) speech recognition still isn't perfect. You still have to speak a little slower, and corrections are necessary. But the computer is pretty good at recognizing context, and letting you correct it and can even learn your language use patterns using your e-mail and document archive.

At some point in the future, speech recognition may become speech understanding. The statistical models that allow computers to decide what a person just said may someday allow them to grasp the meaning behind the words. Although it is a huge leap in terms of computational power and software sophistication, some researchers argue that speech recognition development offers the most direct line from the computers of today to true artificial intelligence. We can talk to our computers today. In 25 years, they may very well talk back.

## **DOCUMENTATION**

<http://www.cs.bgu.ac.il/~amirbaro>

## ABOUT

This project was made under the course:

**202-1-4011 - "topics on operating systems"**

Department: computer science

University: Ben Gurion University of the Negev

Made under the supervision of:

- Prof. Shlomi Dolev – [dolev@cs.bgu.ac.il](mailto:dolev@cs.bgu.ac.il)
- Mr. Michael Orlov – [orlovm@cs.bgu.ac.il](mailto:orlovm@cs.bgu.ac.il)

Presented by:

- Aviad Otmazgin - [otmazgin@cs.bgu.ac.il](mailto:otmazgin@cs.bgu.ac.il)
- Amir Baron - [amirbaro@cs.bgu.ac.il](mailto:amirbaro@cs.bgu.ac.il)

## **REFERENCE**

- **CloudGarden** - implementation of Sun's Java Speech API for Windows platforms.

(<http://www.cloudgarden.com>)

- **Michael Orlov's site** - information on Saya's software and hardware.

(<http://www.cs.bgu.ac.il/~orlov/teaching/saya>)

- **Dragon Naturally Speaking** - free speech recognition software.

(<http://www.nuance.com/naturallyspeaking>)

- **Java speech API** – information about the java speech interface.

(<http://java.sun.com/products/java-media/speech/>)