Tell Me How to Cook

Presented by: Itika Gupta, Sabita Acharya

Introduction

Introduction

Introduction

Problem Statement

- Existing systems are distractive. They restrict the user's ability to perform anything else while using the system.
- Youtube videos on cooking, and cook books are not interactive.

Someone Tell Me How to Cook!

Related Work

Authors: Romain Laroche, Jan Dziekan, Laurent Roussarie, and Piotr Baczyk.

Cooking Coach Spoken/Multimodal Dialogue Systems

Cooking Coach Spoken/Multimodal Dialogue Systems

Romain Laroche, Orange Labs, France, romain.laroche@orange.com
Jan Dziekan, Orange Labs, Poland, jan.dziekan@telekomunikacja.pl
Laurent Roussarie, Orange Labs, France, laurent.roussarie@orange.com
Piotr Baczyk, Orange Labs, Poland, piotr.baczyk@orange.com

Abstract

The use of a cookbook is quite awkward. Your working space is dirty, overloaded with kitchen utensils, your hands are probably busy, dirty or both and it is difficult for your eyes to keep track of the step you are working on. As a result, you keep going back and forth between the work surface and the cookbook, you lose time, track and may even forget a component as an unfortunate result of the mess.

Spoken dialogue systems are relevant for helping in a cooking task. Indeed, it can be performed without any device. This demonstration shows and describes Cooking Coach, two dialogue systems (one vocal and one multimodal) helping the user to search for a recipe, to check that she has all the ingredients and to prepare the recipe.

1 Spoken Dialogue System

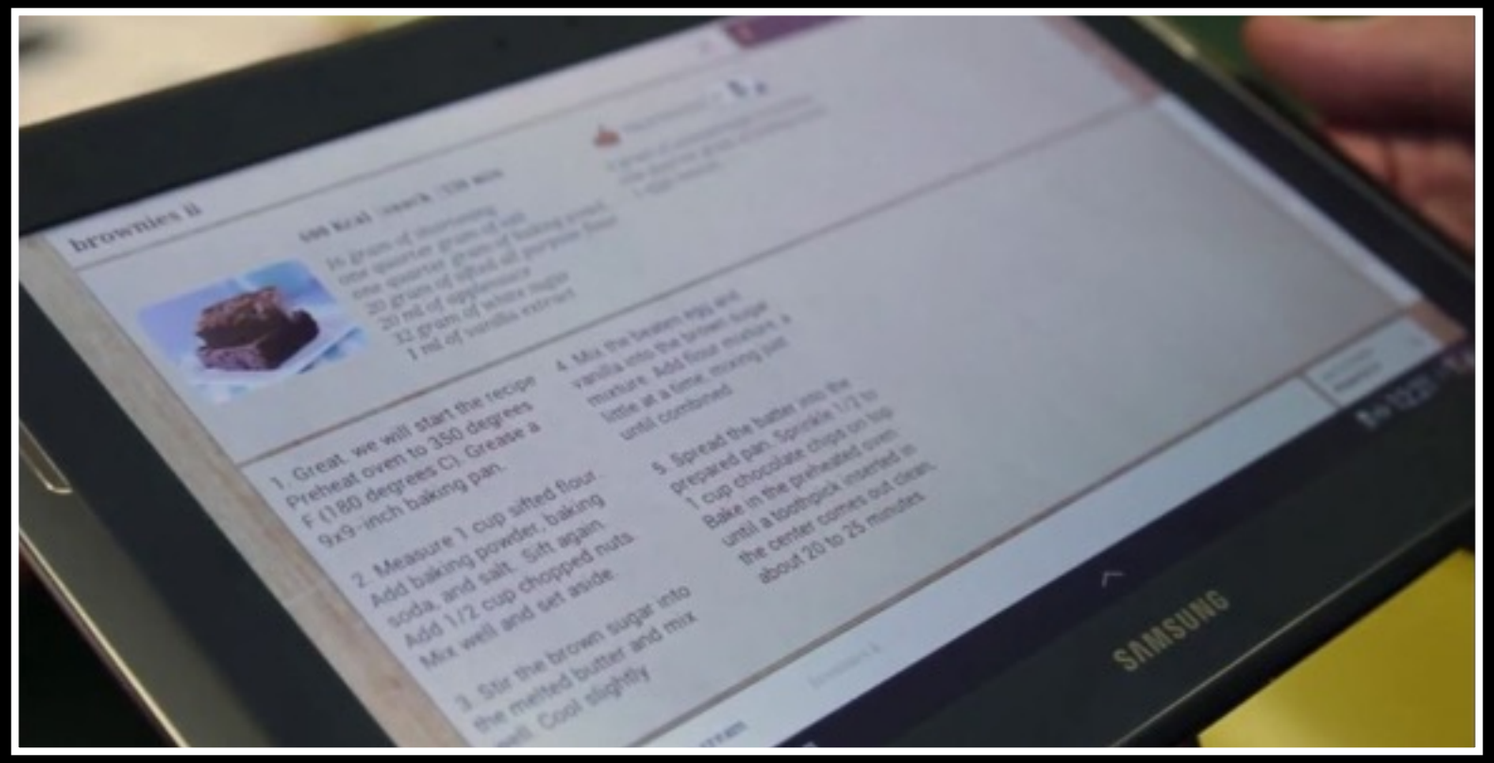
Section 1 describes the spoken version of Cooking Coach. See Section 2 for the multimodal version.

capture and restitution. The role of this client is to replace the Voice Platform that is commonly used for phone-based Interactive Response Services. We interface with the Dragon Mobile SDK, developed by Nuance, in order to get Voice Activity Detection, Automatic Speech Recognition and Text-To-Speech. On the other end, the dialogue application is deployed on an Application Server, which may have access to Information Systems if required.

A dialogue turn is typically processed as follows. Once Dragon Mobile has translated the audio signal into text, the Android client sends an HTTP request to the application server. An HTTP Server receives it and communicates its content to the Phase Engine's servlets. These latter asks the Semantic Analyser to interpret it and then processes the resulting request according to the dialogue logic. The servlets may load user libraries, which in turn, may have access to Information Systems. Once the servlets have planned the system's reaction, it calls a Java Server Page that generates the VoiceXML page to communicate to the client through the HTTP Server. The client parses the VoiceXML file, accesses the Nuance Server to synthesize the text (Text-To-Speech), plays the resulting sound to the user.

1.2 Dialogue Example

Related Work



Related Work

Authors: Joana Paulo Pardal and Nuno J. Mamede.

Starting to Cook a Coaching Dialogue System in the Olympus framework.

Starting to Cook a Coaching Dialogue System in the Olympus framework

Joana Paulo Pardal and Nuno J. Mamede

Abstract This paper presents the COOKCOACH, a system that guides the user while cooking a recipe during a spoken dialogue coaching session. We describe our experience creating the first version of our coaching dialogue system that helps the user cook a selected dish. We use the CMU dialogue systems framework OLYMPUS. The main challenge we faced was the change of paradigm: instead of the system being driven by the user, the user is guided by the system. The result is a system capable of dictating generic tasks to the user. The methodology allows us to create different systems in diverse domains for similar guidance tasks that support the user while

Related Work

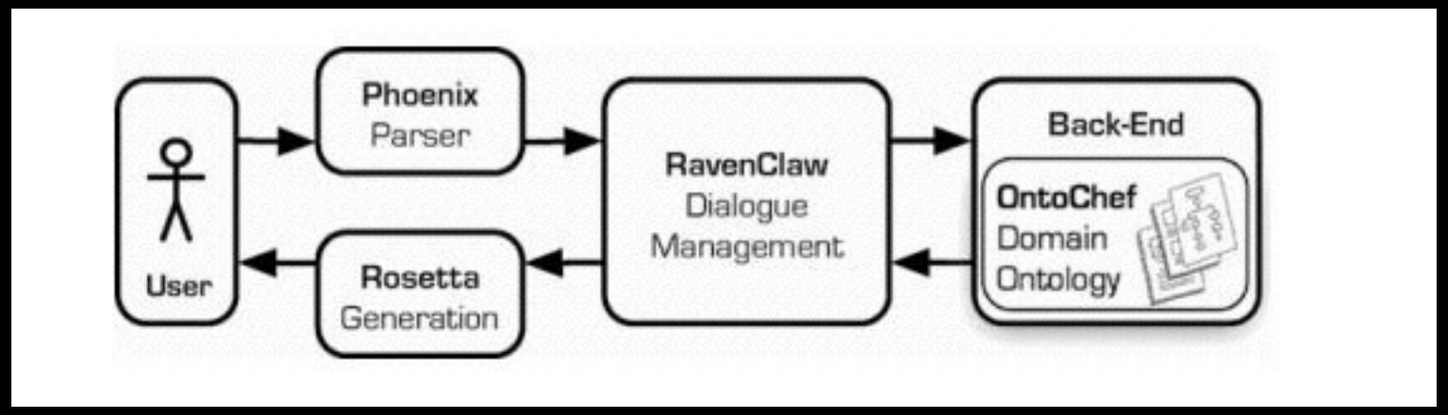


Fig 1. CookCoach Olympus-based architecture

U: I want to cook a chocolate cake.
S: OK! We will need this ingredients.
U: Next
S: 4 oz of sweet dark chocolate.
U: Next
S: 5 eggs
U: Can you repeat?
S: 5 eggs
U: Next
S: 1/2 cup of butter
U: How much is that in liters?
S: 125 mL of butter
`(...)`

S: Separate egg whites and egg yolks.
U: Next step
S: Whisk the eggs whites.
S: Use an electric mixer.
S: Whisk until forming stiff peaks.
U: How? `(go down)`
S: Put eggs whites in a container.
U: How?
S: Let me show you with a brief movie...
`(plays video)`
U: OK! Got it! Next step `(go up)`
S: Add sugar as needed.
`(...)`

Fig 2. Sample conversation from CookCoach

-Coo
They
-Task
initia
-The
differ
utens
meas

Related work

Author: R. M. Wasigner.

Dialog based user interfaces featuring a home cooking assistant.

Dialog based user interfaces featuring a home cooking assistant

Author: R. M. Wasinger
School of Electrical and Information Engineering
The University of Sydney
email: rainer.wasinger@wasinet.com
Supervisor: Prof. Hansen Yee

Abstract: Natural language dialog components have improved significantly since they were first discussed more than 20 years ago. This paper leverages the current position of the maturing speech industry and explores effective user interface design for speech applications. In particular, discussion focuses on the creation of a home cooking assistant based on a multi-modal design. Issues relating to the implementation of dynamically changing multiple-context rule grammars are also discussed. So too are methods of interacting with an underlying database.

Keywords

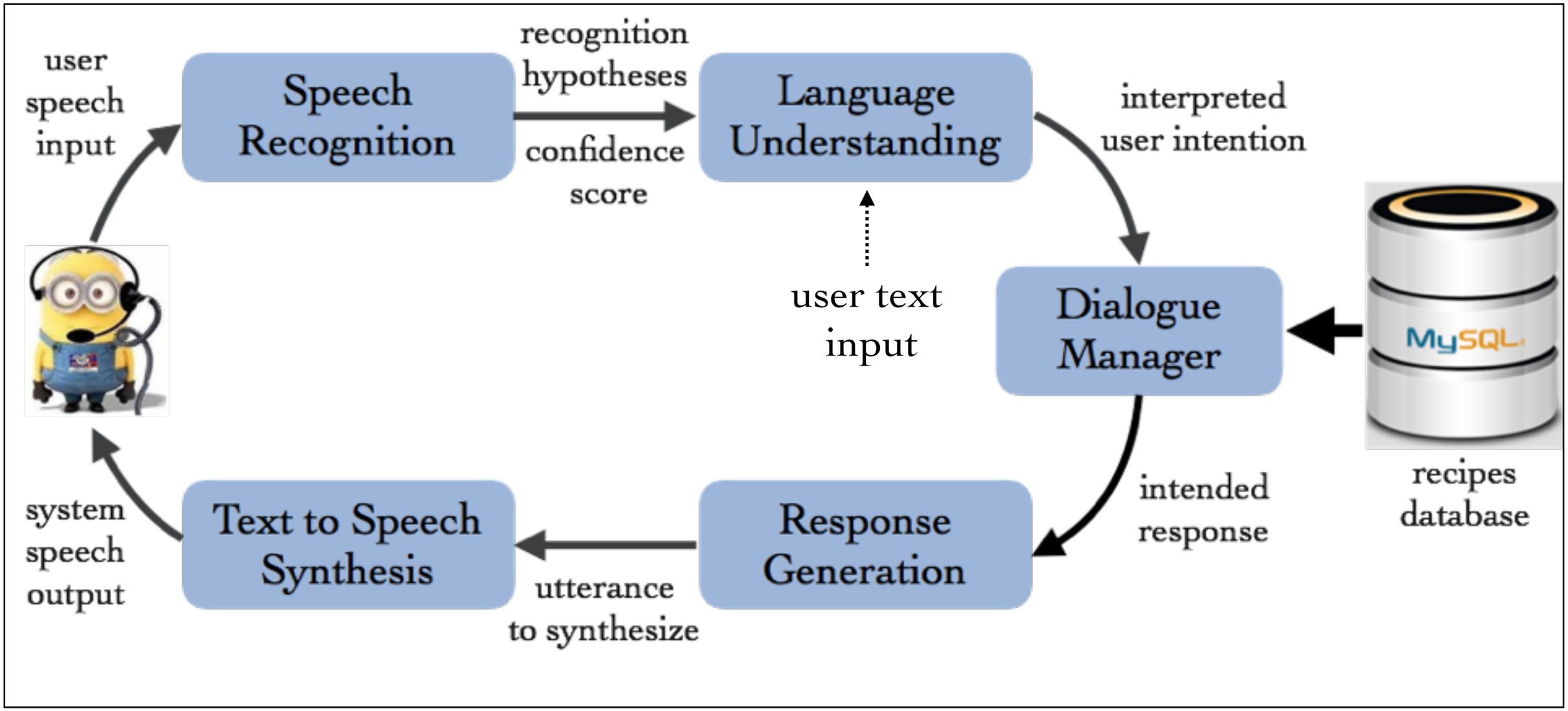
Dialog based user interfaces; multi-modality; multiple-context and dynamically changing rule grammars; home cooking assistant.

Interface (GUI). Speech is transient, invisible and asymmetric. It is transient because once it has been spoken it is gone. Graphics displayed to a screen on the other hand are persistent. It is invisible because you cannot see speech. Finally,

Related work

Fig 3. GUI for home cooking assistant

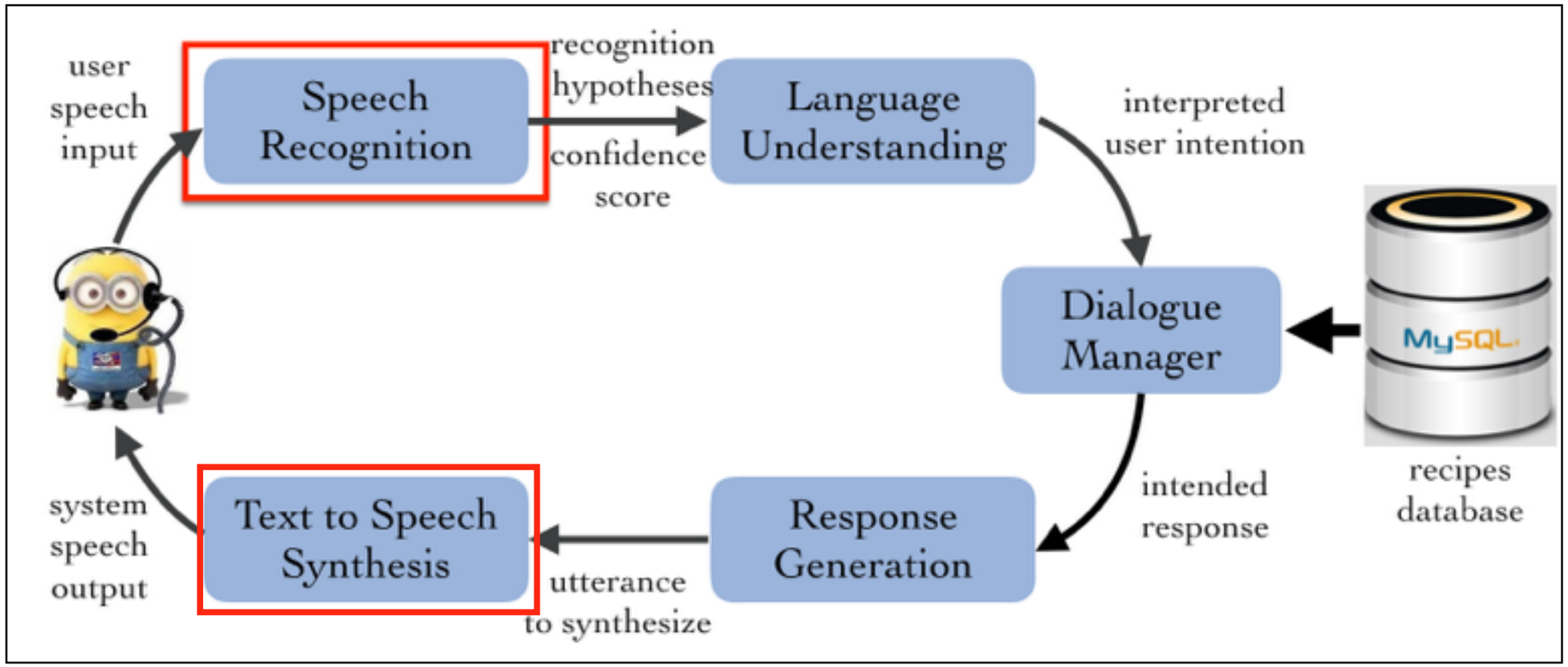
Architecture



Architecture

Speech Recognition and Text-To-Speech Synthesis

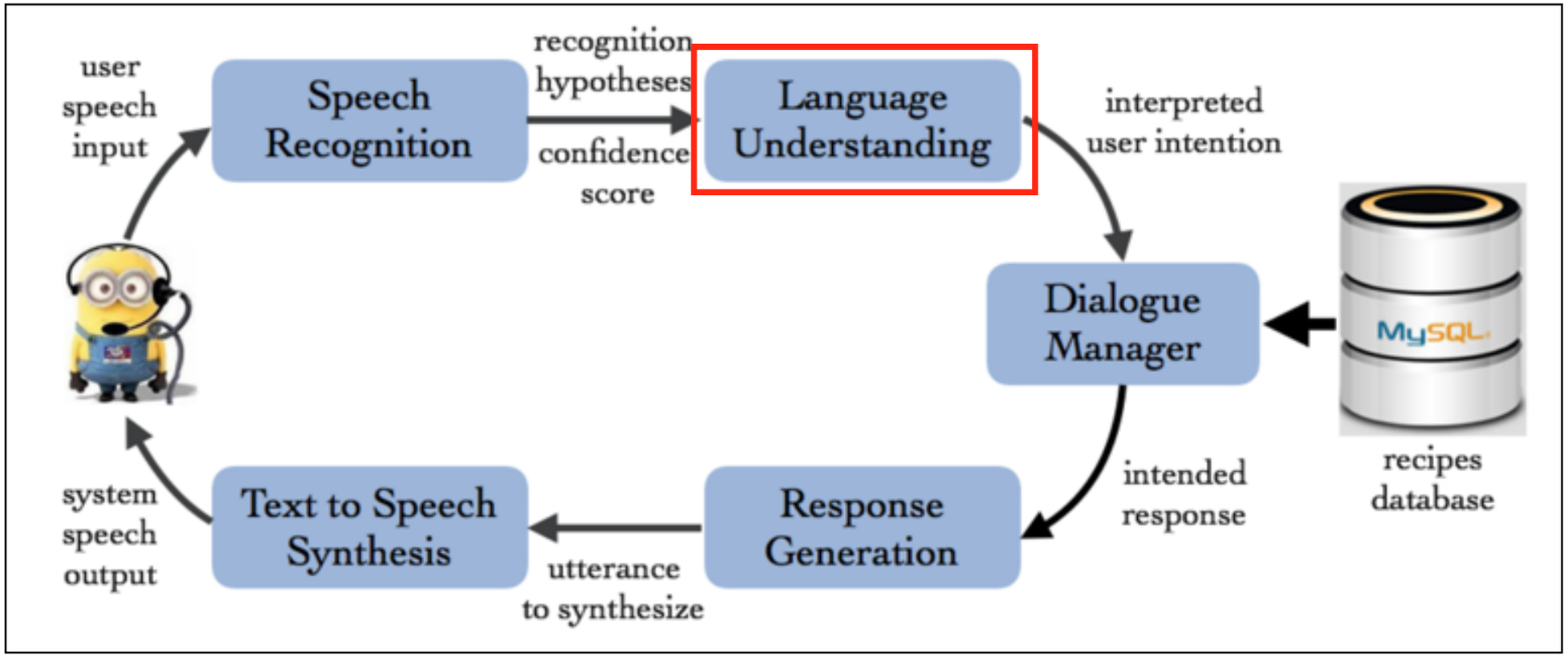
- J.A.R.V.I.S Speech API.
- Google Speech Engine.



Architecture

Language Understanding

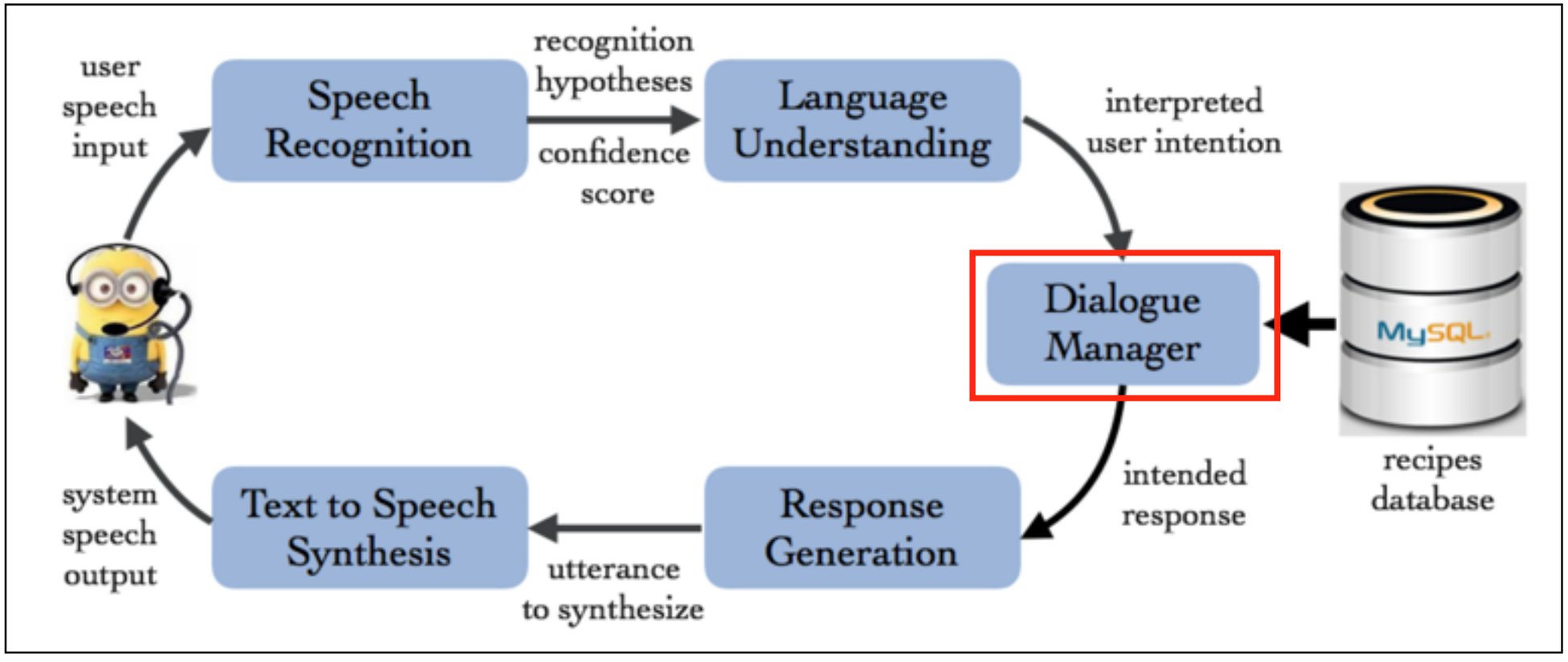
- Dialogue acts recognized: Answer, Statement, and Acknowledgement (Affirmation or Negation).



Architecture

Dialogue Manager

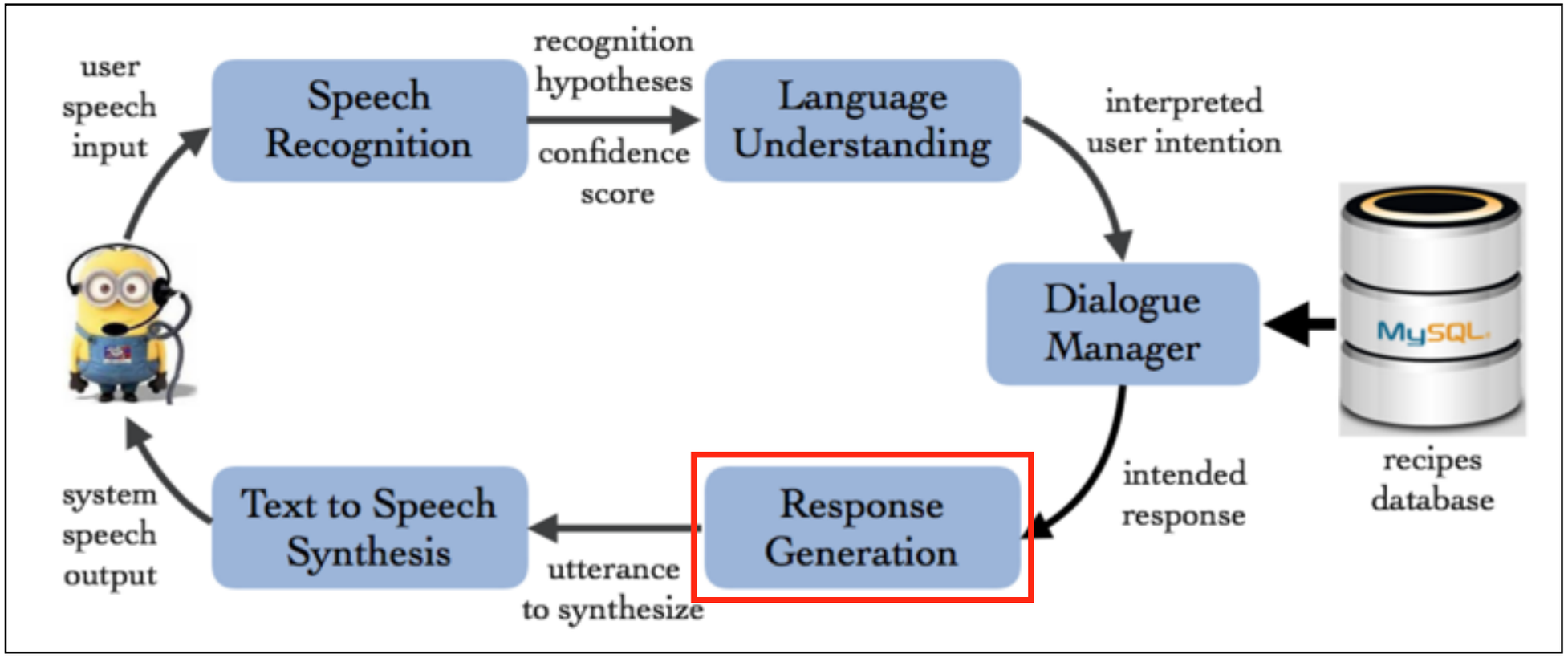
- Dialogue Control: system dialogue acts such as Statements, Questions, Info-Request.
- Dialogue Context: keeps track of the current recipe name, the procedure, and the ingredients.



Architecture

Response Generation

- Using templates obtained from YouTube videos.



Domain knowledge

We say that this is a
"Baking with Julia". I
bagels.

An interesting point:
Google speech reco

Domain knowledge

Laura: Okay let's work on our dough. First thing I am going to do is dissolve my one packet of yeast in one quarter cup of nice warm water. If it's too hot, it slows the yeast. I am getting a pinch of sugar which is yeast's favorite food.

Julia: Good.

Laura: Okay now I am going to let that nice and happy over there. Now I need my mixing bowl and here I have two cups of again, warm water. Anytime you put liquid into it, it needs to be about the same temperature.

Julia: Yes.

Laura: I am going to 3 table spoons of vegetable shortening which helps to give my bagels the texture.

Julia: Oh yes they were softy chewy.

Laura: Okay I am going to put into this 1 table spoon of salt and I need 1 table spoon of sugar and color and flavor.

Dataset

- Recipes are obtained from The USENET Cookbook.
- Currently we have 5 recipes in our toy corpus.

USENET contains
ranging from main
salad and dessert

Recipe Example

AFRICAN VEGETABLE STEW

Recipe Name

A spicy vegetable stew

Short Description

I got this recipe from my mom, who got it from a friend. It's very easy to make.

Long Description

Ingredients

Ingredients

(Serves 4)

Servings

- 1 very large onion, chopped
- 1 bunch Swiss chard
- 1 can garbanzo beans (known also as chick peas, ceci, etc.)
- 1/2 cup raisins
- 1/2 cup uncooked rice
- several fresh tomatoes (or 1 large can of canned tomatoes)
- 1 clove garlic (or more to taste)
- 2 yams
- salt and pepper to taste
- tabasco sauce to taste

Procedure

Procedure

1. Fry onion, garlic, and white stems of chard until barely limp. Add chopped greens and fry a bit.
2. Either peel the yams or scrub them well with a vegetable brush, then slice them into thick slices. Add garbanzos, raisins, yams, tomatoes, salt, and pepper. Cook a couple of minutes.
3. Make a well in the center of the mixture in the pot. Put the rice in the well and pat it down until it's wet. Cover and cook until rice is done!about 25 minutes.
4. Add Tabasco sauce to taste.

Notes

Additional Notes

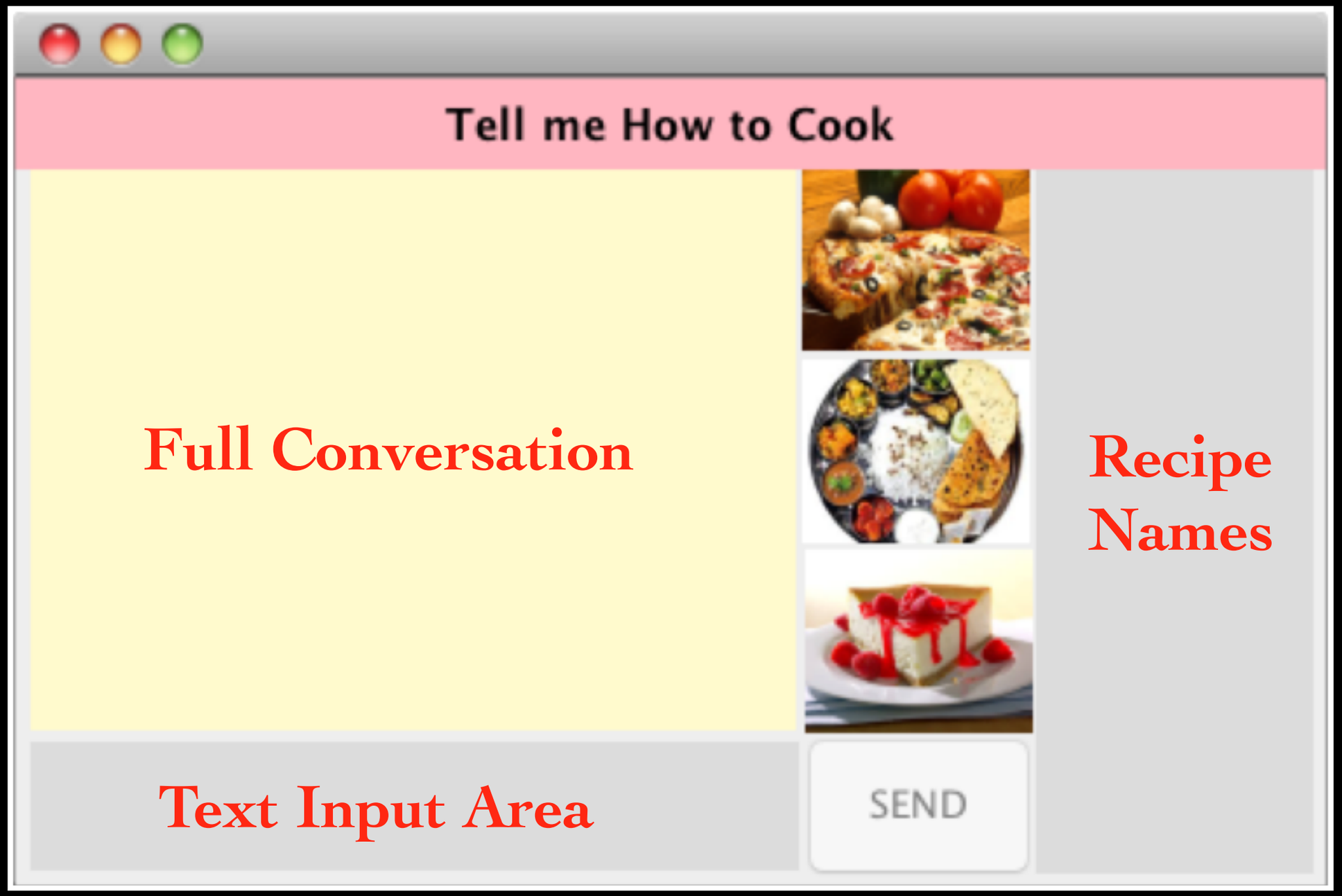
I find that the flavors of the vegetables don't stand out unless you put in a lot of tabasco. But then, I've got a cast-iron tongue. You should add enough to make the stew seem spicy to you.

Rating

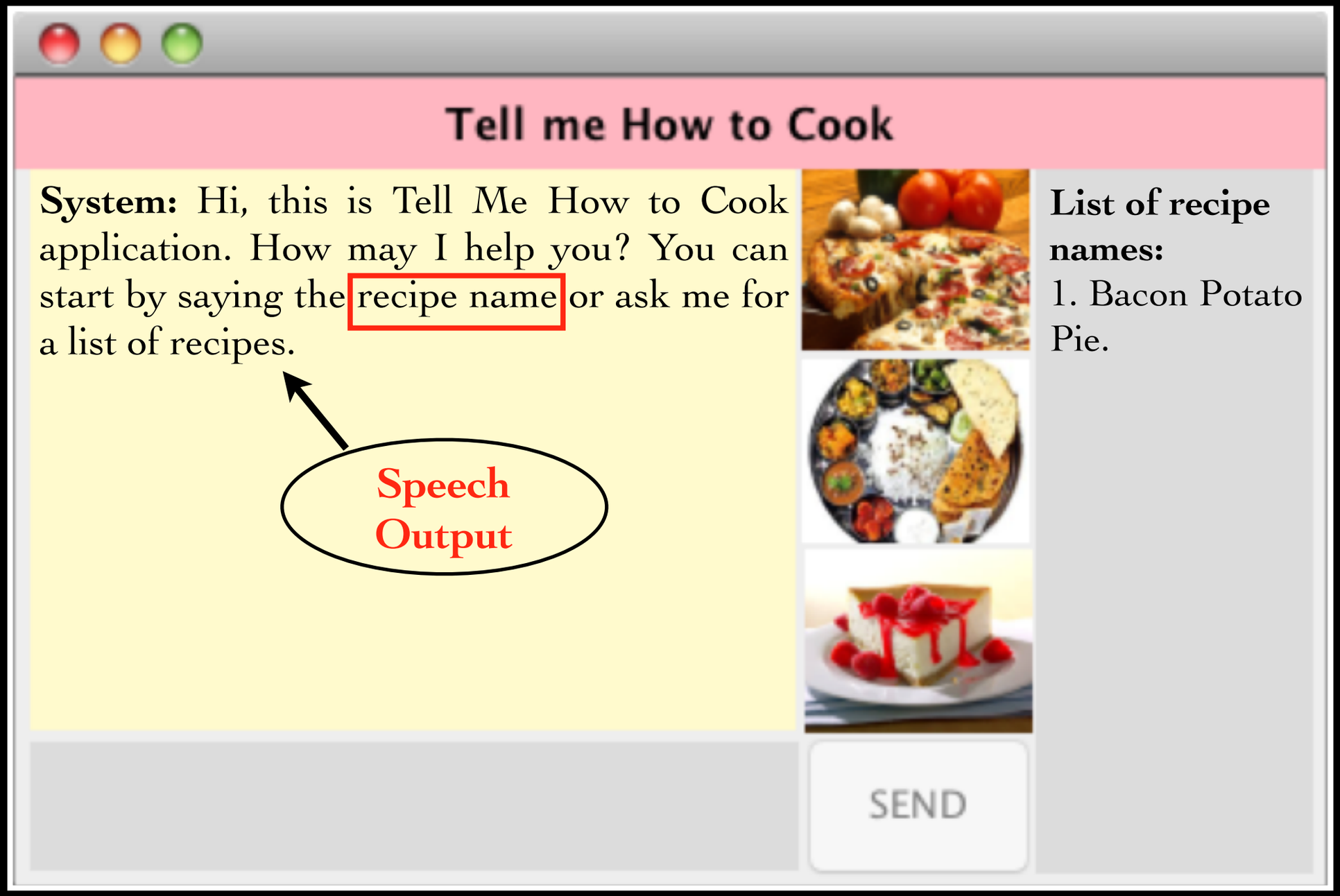
Difficulty: easy.
Time: 15 minutes preparation, 30 minutes cooking.
Precision: no need to measure.

Difficulty
Time

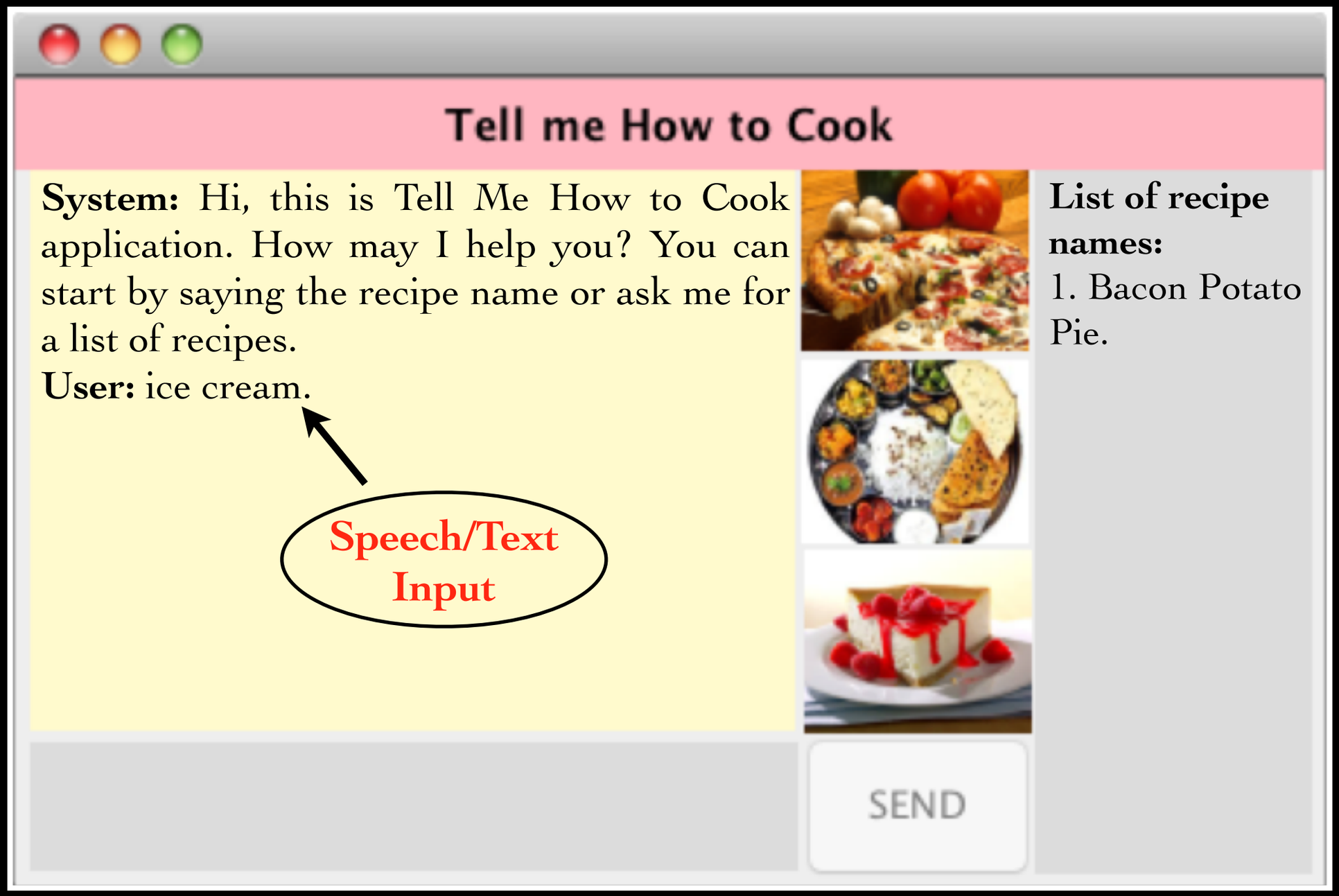
GUI for our System



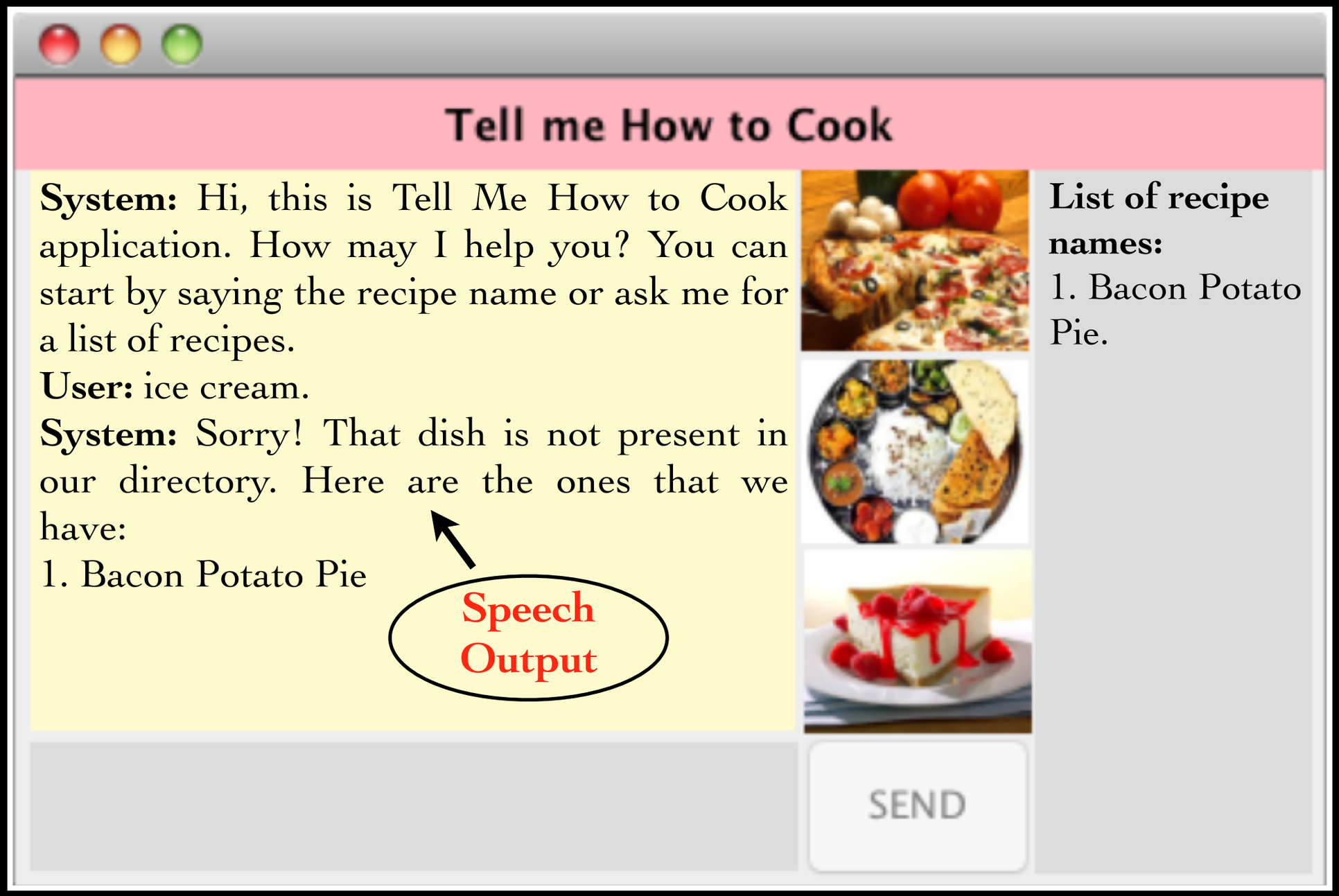
Dialogue Flow (1)



Dialogue Flow (2)

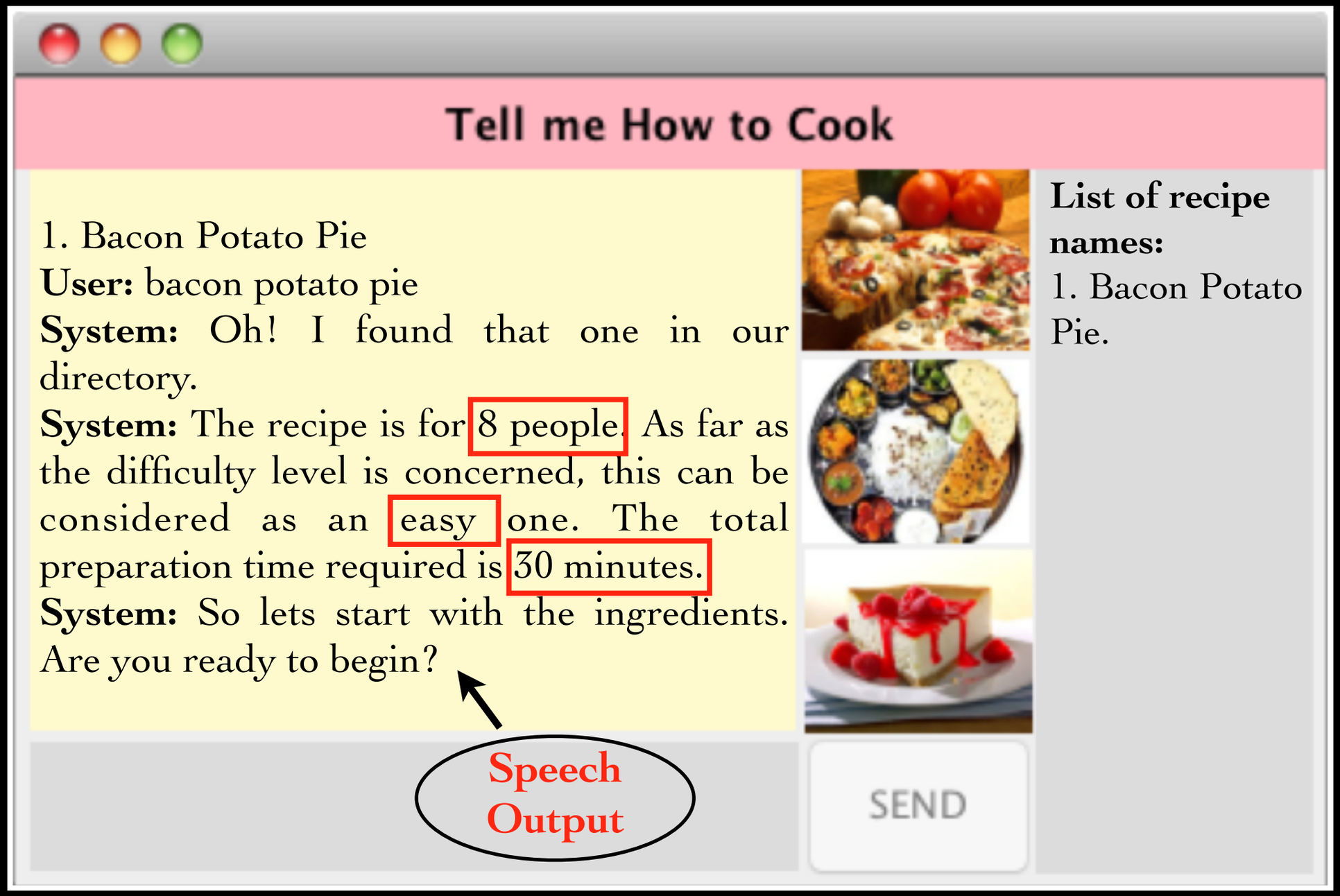


Dialogue Flow (3)



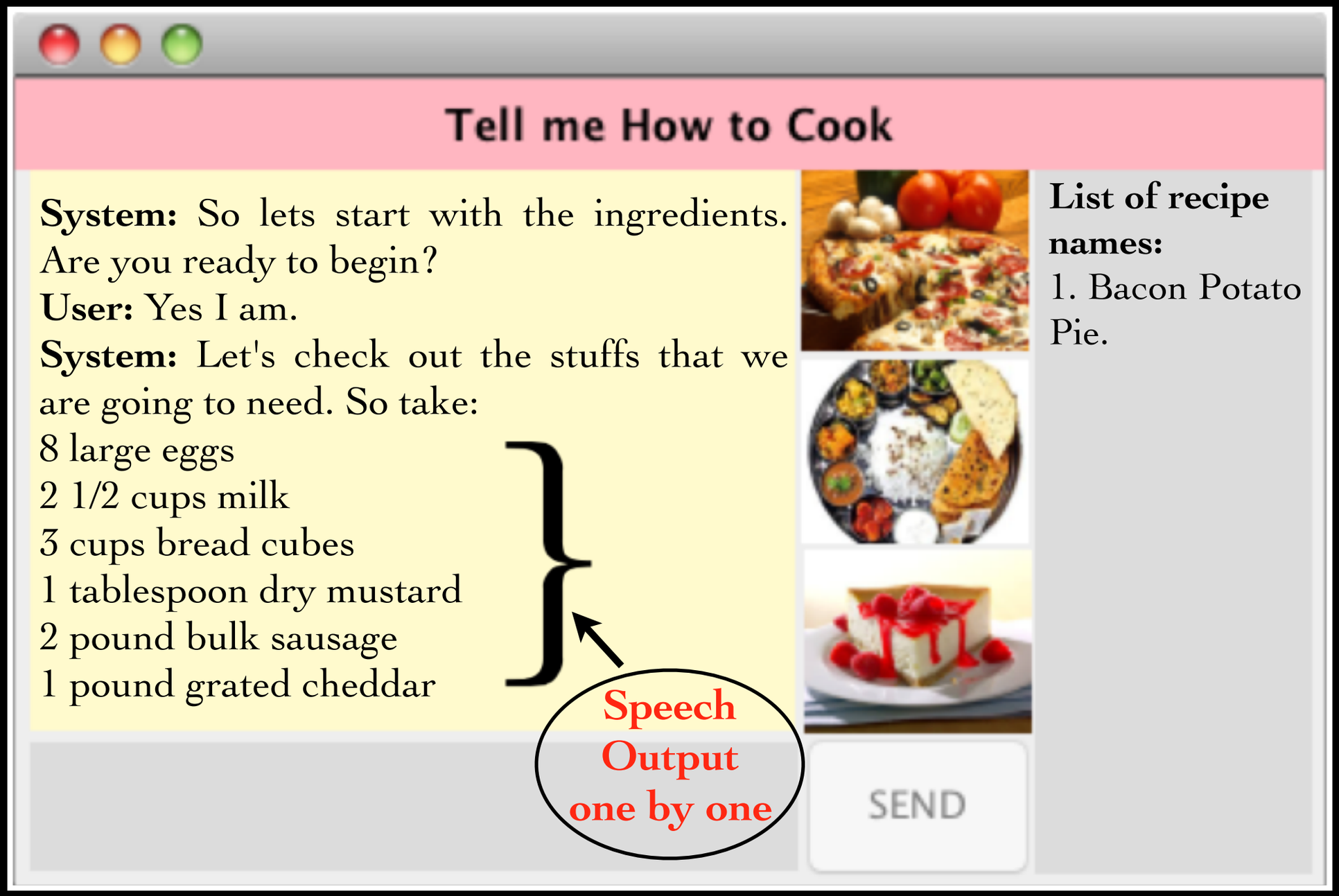
Dialogue Flow (3)

Dialogue Flow (4)



Dialogue Flow (5)

Dialogue Flow (6)



Next Steps

- Customize servings according to user's desire.
- Incorporate general and ingredient specific notes when available.
- Allow user to select dishes based on its difficulty level.
- Allow selection to be made based on dish preparation time
- Perform evaluation of the system.

Tell me How to Cook (After)

Comparison with Related Work

	Cooking Coach	CookCoach	Home cooking	Tell me how to cook
Platform/ framework	VoiceXML	Olympus	Java	**Java**
ASR/TTS	Nuance speech engines	Future work	IBM speech engines	**Google speech services**
Mode of Interaction	Speech, Gestures and Touch	Text based	Speech, Touch and Visual	**Speech based (text input)**
Type of Dialogue	System initiative	System initiative	User initiative	**System initiative**
Data Storage	Database	Domain ontology	Database	**Database**

Future Work

- Selection of dishes based on the available ingredients.
- Mixed initiative dialogue system.
- Giving information about ingredients that can be substituted.
- More natural responses.

Limitations of Our Approach

- Sample dialogues that were collected might not be representative of an actual speech only dialogue system.
- We are using rule-based approach. Language generations techniques might help in introducing variations in responses.
- Does not take into handle those cases where the user's dialogue act differs from what the system expects at that instance.
- Lots of acknowledgement might frustrate the user.

Advantages of Our Approach

- Easy to use, less distractive, and time saving.
- Gives recipe details as well as provides companionship through conversation.
- Can be used as a separate application or as a part of smart kitchen.
- Takes care of error conditions by frequently asking for confirmation.
- Allows user to specify the amount of servings.
- Database customizability.

database customisability: The user can store the recipes they want to make, unllike a book , which from the user's perspective contains both good and bad recipes

Takeaways from the Project

Human answers in vari

There are seve

- Developing a dialogue system is NOT an easy task.
- Some good dialogue system toolkits exist but they lack proper documentation.
- Current speech recognition systems are not perfect.
- Learnt to appreciate the existing dialogue systems and recognized improvements that are yet to be made.
- Choice of having system vs mixed initiative relies heavily on the application.

References

[1] Wasinger, Rainer M. "Dialog based user interfaces featuring a home cooking assistant." University of Sydney, Australia (2001).

[2] Pardal, Joana Paulo, and Nuno J. Mamede. "Starting to Cook a Coaching Dialogue System in the Olympus framework." In Proceedings of the Paralinguistic Information and its Integration in Spoken Dialogue Systems Workshop, pp. 255-267. Springer New York, 2011.

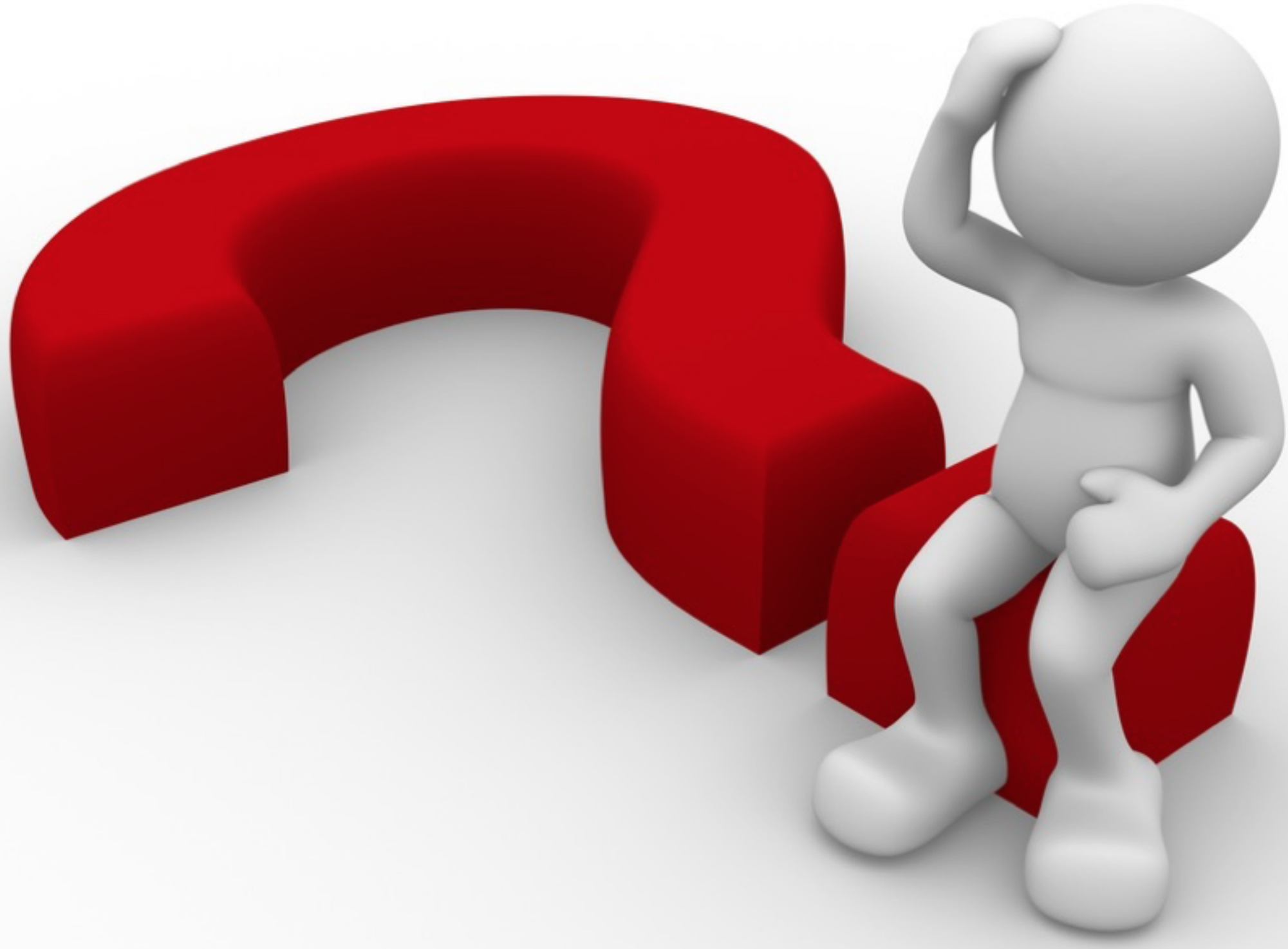
[3] Laroche, R., Dziekan, J., Roussarie, L., Baczyk, P. "Cooking Coach Spoken/ Multimodal Dialogue Systems. Conference Paper. August 2013. Conference: Cooking with Computers

[4] Schäfer, Ulrich, et al. "Ingredients and Recipe for a Robust Mobile Speech-Enabled Cooking Assistant for German." KI 2013: Advances in Artificial Intelligence. Springer Berlin Heidelberg, 2013. 212-223.

[5] Main Dishes. Accessed March 18, 2016. http://www.lysator.liu.se/etexts/recept/us/

[6] Croissants with Esther McManus https://www.youtube.com/watch?v=xps55MQ2Vgo

[7] Trophy Cupcakes on Martha Stewart
https://www.youtube.com/watch?v=k_877FkaGLk&nohtml5=False

[8] Cooking Matters: http://cookingmatters.org/

THANK
YOU