Speech-Based Chipotle Ordering System

Ja Jan Hsu
Leilei Liu
Leon Lai
Siddharth Coontoor
Turki Alenezi

OBJECTIVE

The objective of this project is to design, implement, and test an application that interacts with users via speech only.

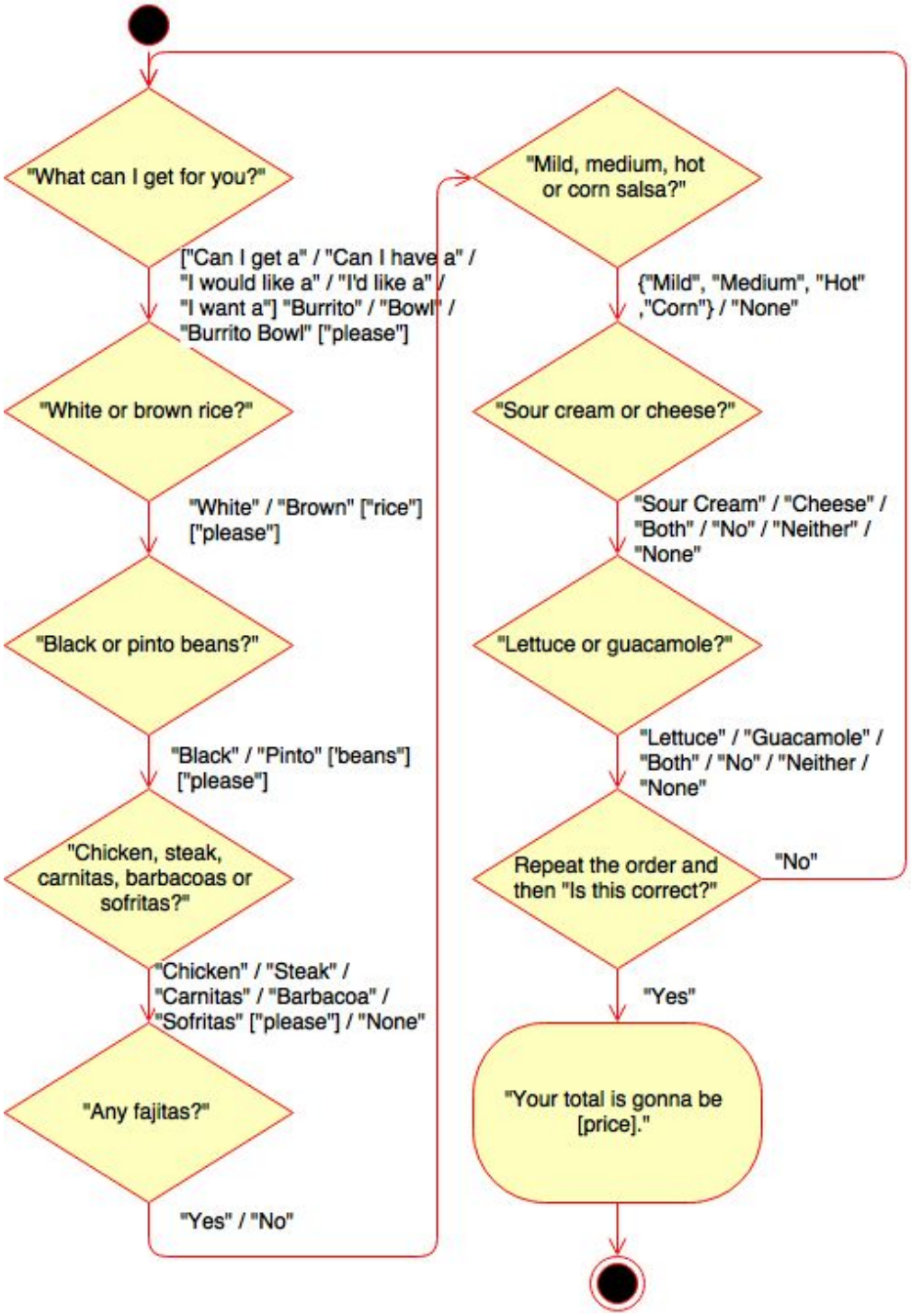
MOTIVATION

The Chipotle ordering process is structured and repetitive, suggesting that we can use an application to perform this ordering behavior. A speech ordering system can automate it and reduce the cost of employee training.

REQUIREMENTS AND DESIGN

This section includes an activity diagram to illustrate the system's behavior, design design rationale that addresses visibility concerns, the task analysis, and implementation details.

Activity Diagram



Operation Description

We followed the typical Chipotle ordering process in developing our speech-based application that allows customers to order food at any Chipotle branch.

The application starts by asking the customer "what can I get for you?" A valid answer to this question that the system recognizes must match its grammar rule, the "firststep" rule. The system does nothing and keeps listening until a valid answer is heard. Once the customer responds with a valid answer, the system stores the customer's choice and proceeds to the next question. Again, after asking the customer, the system waits for a valid answer matching the next rule in the grammar. After the customer provides grammatically correct answers to all questions, the system repeats the entire order to the customer and waits for confirmation from customer. In the case of confirmation, the system calculates and provides the customer with the total price of the order; otherwise, it starts the order over. Next, we illustrate how the system works in terms of recognition and justifies the design of each step.

A Chipotle employee typically first asks for the menu item that will be the basis of the order. Such items include the burrito and the burrito bowl, which is often call just the "bowl". Therefore, the first question is "what can I get for you?"Here, the system recognizes anything that optionally starts with phrases, such as "I would like a" and "can I get a", followed by a required item, which can be "bowl", "burrito", or "burrito bowl", optionally followed by "please". Here are some valid answers that match the grammar rule for the first question:

- I would like a bowl
- I'd like a burrito
- Can I get a bowl please
- burrito bowl please
- burrito

Typically, the first ingredient to be asked about is rice. White and brown rice are the two types available. Therefore, the second question is "white or brown rice?" The system recognizes answers that start with "white" or "brown" followed optionally by two possible items which are "rice" and "please". Here are some valid answers that match the grammar rule for the second question:

- brown
- white rice
- white rice please
- brown please

The choice of beans comes next in the routine ordering process, and the choices are black and pinto beans. Therefore, the third question is "black or pinto beans?" The system recognizes answers that starts with "black" or "pinto" followed optionally by two possible items which are "beans" and "please". Here are some valid answers that match the grammar rule for the third question:

- pinto
- black beans
- black beans please
- pinto please

Afterwards, to let the customer choose from the five meat choices or tofu, the system asks the fourth question, "what kind of meat? We have chicken, steak, barbacoa, carnitas and sofritas." The system recognizes answers that start with "chicken", "steak", "beef", "barbacoa", "carnitas", or " sofritas" followed optionally by "please". Here are some valid answers that match the grammar rule for the fourth question:

- chicken
- steak
- carnitas please

Conventionally, Chipotle employees then ask whether the customer wants fajitas in their orders as a yes/no question. The fifth question is thus "any fajitas?" The system recognizes answers that start with "yes", "yeah", or "no" followed optionally by "please" and "thanks". Here are some valid answers that match the grammar rule for the fifth question:

- yes please
- no thanks
- yes
- yes please thanks

Subsequently, to ask for the type of salsa, the sixth question is thus "what kind of salsa? mild, medium, hot, or corn?" The system recognizes answers that consist of the sauce options in any order separated optionally by "and", or it recognizes the word "none". Here are some valid answers that match the grammar rule for the sixth question:

- hot
- mild and medium
- mild, medium, and corn
- corn, hot, medium and mild
- none

The next step is typically to ask for any sour cream or cheese. Thus, the seventh question is "any sour cream or cheese?" The system recognizes answers that start with either or both options in any order followed optionally by "please". It can also recognize "no", "no thanks", and "neither". Here are some valid answers that match the grammar rule for the seventh question:

- sour cream please
- cheese
- cheese and sour cream
- both
- neither
- no
- no thanks

As the last step in the ingredient choosing process, the rest of the available ingredients to add are asked about. The eighth question is thus "any lettuce or guacamole?" The system recognizes answers that start with either or both options in any order followed optionally by "please". It also can recognize "no", "no thanks", and "neither". Here are some valid answers that match the grammar rule for the eighth question:

- lettuce please
- guacamole
- guacamole and lettuce
- both
- neither
- no

- no thanks

At this stage all the menu item and ingredient choices have been acquired. In a speech-only application no visual feedback is provided and thus visibility, in the context of design principles, is absent if the exact typical Chipotle ordering process is copied. Thus, before finalizing the ordering process, a confirmation message including all the menu item and ingredient choices is repeated as the system's response to the customer's inputs. The confirmation message repeats whatever the customer ordered and asks if that order is correct, for example "you ordered a bowl with brown rice, black beans, chicken, medium salsa and corn, sour cream, cheese, lettuce, is that right?" The system recognizes either "yes" or "no". If the customer says yes, the system calculates and provides the total price to the customer. If the customer's answer is no, the system apologizes and goes back to the first question and starts over.

Visibility Concerns

Users of the system includes first-time Chipotle patrons and regulars. For the first-timers, usability is of utmost concern. For regulars who are familiar with the menu items and ordering process, it is crucial to minimize the total duration of the ordering process. The concern of usability is already addressed by our choice of adopting the conventional ordering questions, which we trust to have formed out of usability. A design goal of this system is thus minimal average duration.

Speech-based systems must provide visibility via audible feedback based on what the user has spoken. Two feedback schemes were considered: a confirmation of the user's answer after each question and a confirmation of all of the user's answers at once towards the end of the ordering process.

To determine which feedback scheme would result in the shortest average ordering process, we assumed that all question-and-answer stages (including confirmations, which are simply question-and-answer sequences) have the same length, u. If a process has probability r of repeating itself from the beginning after a complete run, the following equation determines the number of runs of the process in question:

$$\sum_{k=1}^{\infty} k\, r^{k-1} = \frac{1}{(1-r)^2}$$

Let *r* be equal to the speech recognition error rate, then having a confirmation after each stage makes the ordering process ((1+1)*8)u/(1-r)^2 long, while having a grand confirmation at the end makes the process (1+1*8)u/(1-r)^2 long, neglecting the price message. It is clear that the former scheme takes almost twice as long given any error rate. Therefore, we chose to have a grand order validation question at the end that repeats the entire order.

Task Analysis

Task	**Initiating signal or event**	**Action**	**Feedback**	**Potential errors**	**Error recovery**
Specifying base menu item	Voice prompt from the system.	User says something including "Bowl", "Burrito" or "Burrito bowl"	System asks the next question.	System does not recognize responses. System recognizes responses incorrectly.	System prompts error message and ask the question again and give customer options for the question. When the entire order has been given, system repeats the customer's order and asks for confirmation.
Choosing the type of rice	Voice prompt from the system with the available options.	User says "White", "Brown", "Half and Half" or "None"	System asks the next question.	System does not recognize responses. System recognizes responses incorrectly.	System prompts error message and ask the question again. When the entire order has been given, system repeats the customer's order and asks for confirmation.
Choosing the type of beans	Voice prompt from the system with the available options.	User says "Black", "Pinto" or "None"	System asks the next question.	System does not recognize responses. System recognizes	System prompts error message and ask the question again. When the entire order has been given, system

				responses incorrectly.	repeats the customer's order and asks for confirmation.
Choosing the type of meat/tofu	Voice prompt from the system with the available options.	User says "Chicken", "Steak", "Carnitas", "Barbacoa", "Sofritas" or "None"	System asks the next question.	System does not recognize responses. System recognizes responses incorrectly.	System prompts error message and ask the question again. When the entire order has been given, system repeats the customer's order and asks for confirmation.
Specifying the inclusion of fajitas	Voice prompt from the system with the available options.	User says "Yes" or "No"	System asks the next question.	System does not recognize response. System recognizes responses incorrectly.	System prompts error message and ask the question again. When the entire order has been given, system repeats the customer's order and asks for confirmation.
Choosing the type of salsa	Voice prompt from the system with the available options.	User says "Mild", "Medium", "Hot", or "Corn"	System asks the next question.	System does not recognize responses. System recognizes responses incorrectly.	System prompts error message and ask the question again. When the entire order has been given, system repeats the customer's order and asks for confirmation.
Specifying the inclusion of sour cream and/or cheese	Voice prompt from the system with the available options.	User says "Sour cream", "Cheese", "Both", or "None"	System asks the next question.	System does not recognize responses. System recognizes responses incorrectly.	System prompts error message and ask the question again. When the entire order has been given, system repeats the customer's order and asks for confirmation.
Specifying the inclusions of	Voice prompt from the system with the available options.	User says "Lettuce",	System asks the next question.	System does not	System prompts error message and

lettuce and/or guacamole		"Guacamole", or "None"		recognize responses. System recognizes responses incorrectly.	ask the question again. When the entire order has been given, system repeats the customer's order and asks for confirmation.
Confirming the order	Voice prompt from the system with the available options.	User says "Yes" or "No"	System reports the total price.	System does not recognize responses. System recognizes responses incorrectly.	System prompts error message and ask the question again. When the entire order has been given, system repeats the customer's order and asks for confirmation.

Implementation Details

- The application runs on the Microsoft Windows operating system, version "8" or later, and uses the Microsoft Speech API.
- The application was developed in C# using Microsoft Visual Studio.
- The speech recognition grammar utilizes Speech Recognition Grammar Specification Version 1.0 and is represented by an XML document.

FINDINGS

- We found that checking the speech recognition confidence level decreases the false positive rate; 90% was a reasonable confidence level threshold.

SUMMARY

- The ordering system follows the real ordering process in Chipotle.
- We provide hints and error handlings for users so they can better understand how to interact with the application.
- We provide various requesting phrases for users to answer a question. For example: can I have a _?; Can I get a _?; I want a _ ; I'd like a _; please; thanks.

- A grand confirmation dialogue instead of step-by-step confirmations maximizes ordering process efficiency.

LIMITATIONS

- In making design choices, the grand confirmation was assumed to take the same duration as other questions, when it was in reality considerably longer. It may in reality be more efficient to have step-by-step confirmations.
- The grammar does not allow all possible Chipotle orders.

REFERENCES

- Speech Recognition Grammar Specification Version 1.0. http://w3.org/TR/speech-grammar