

# SpeakEasy

*Integrating context-sensitive visual and aural language acquisition into your daily life*

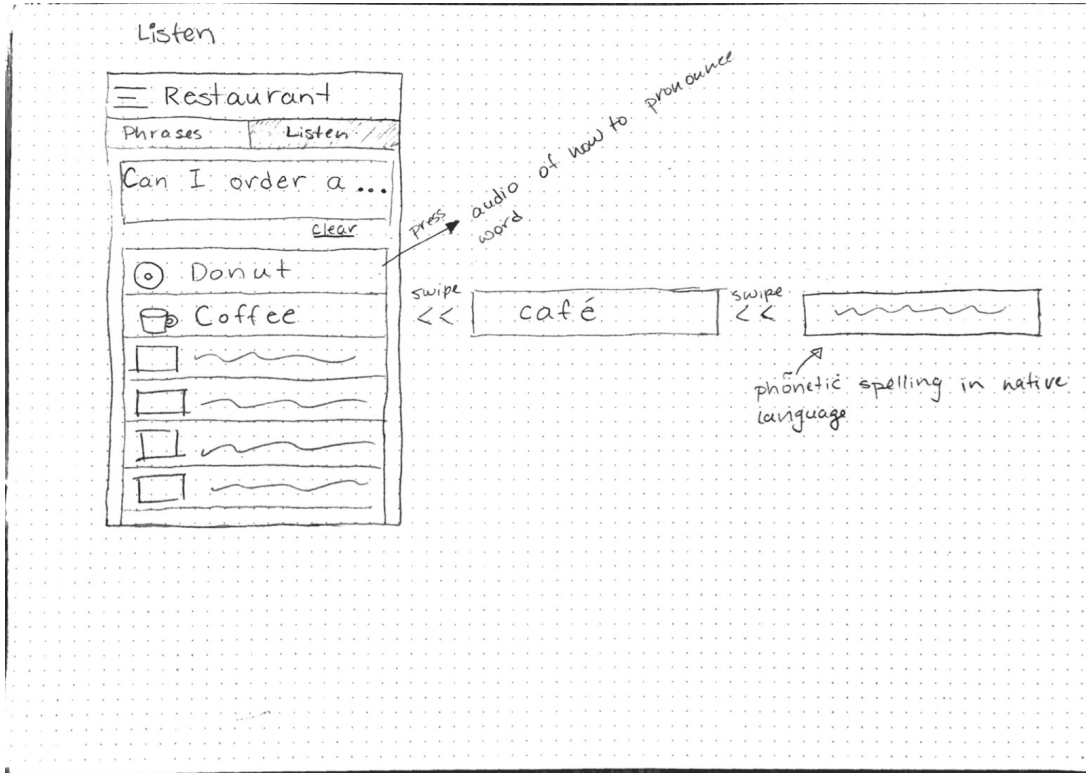
Gabriela Groth (Manager, Design), Eric Chew (Development), Tommy Truong (Documentation), Carlos Araujo (User testing)

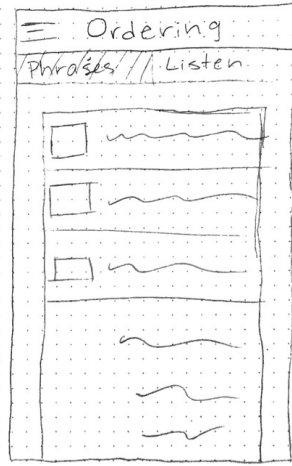
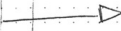
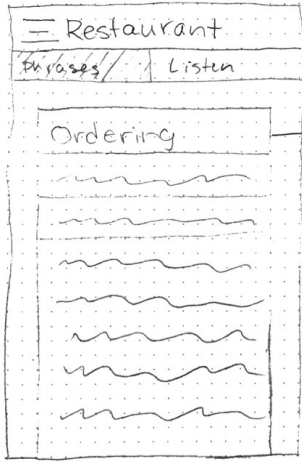
## **MISSION STATEMENT**

Learning a new language is often difficult and boring. The current solutions do not integrate well into a busy life. These solutions include computer programs like Rosetta Stone or group language classes that only allow learners to practice one or two hours a day. Furthermore, students using these solutions can only learn a language isolated from real-world situations. Our app, SpeakEasy, is designed to overcome these limitations. SpeakEasy is a mobile application which helps immerse users in another language by providing on-the-go language instruction specific to the users' immediate context and environment. SpeakEasy provides instant speech suggestions and corrections appropriate to a user's location, allowing them to learn and master a language even while performing other activities.

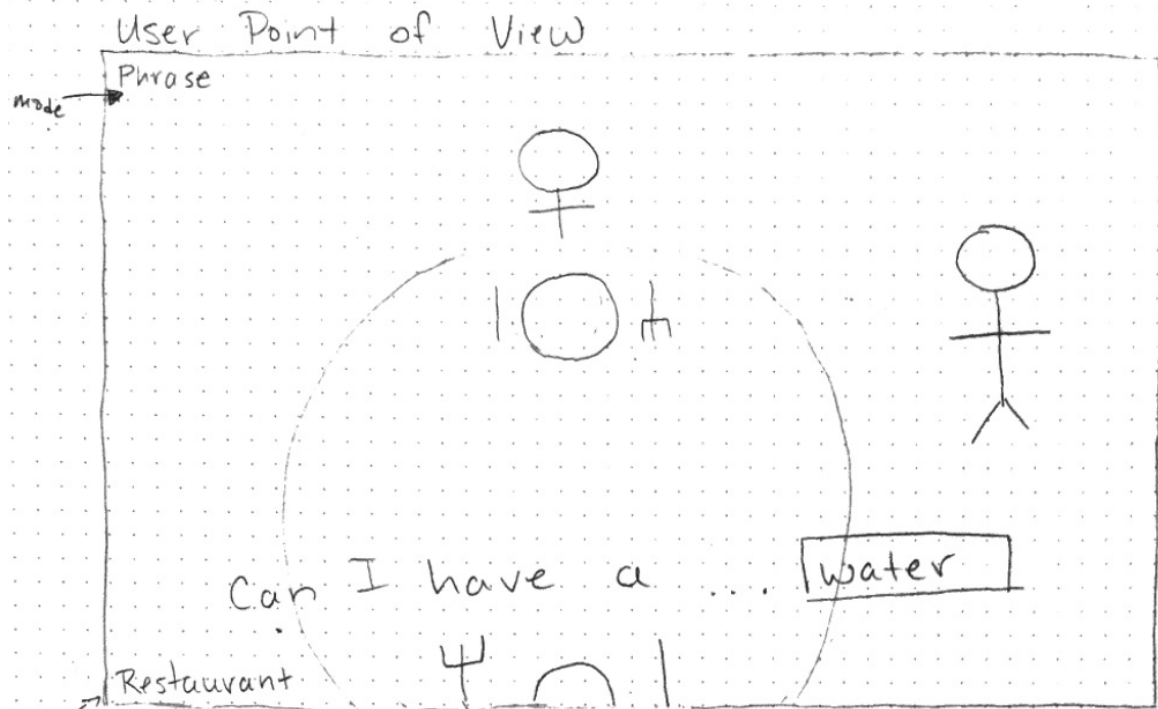
# UI SKETCHES

## DESIGN 1 - MOBILE PHONE APPLICATION INTERFACE

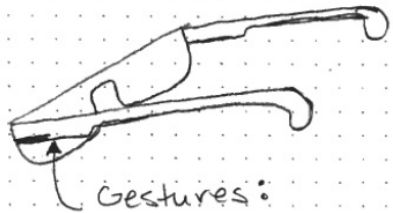




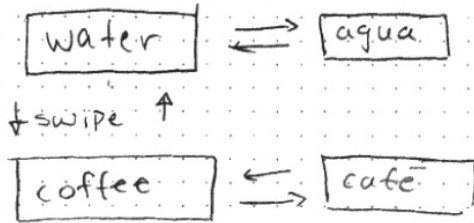
## DESIGN 2 - SMART EYEWEAR APPLICATION INTERFACE



Location



Gestures:

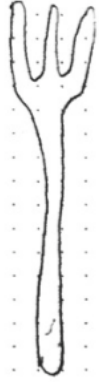


- <swipe ↑ or ↓> scrolls through auto complete suggestions
- <click> pronounce word
- <swipe ← or →> scroll through translations and pronunciation guide

mode →

Focus

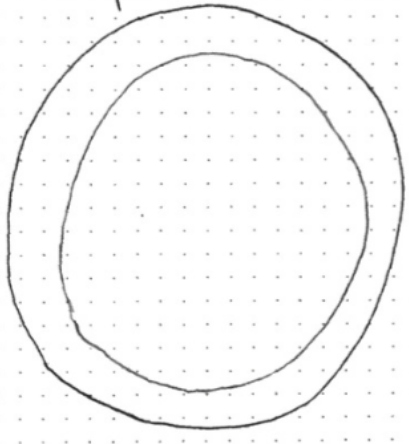
fork



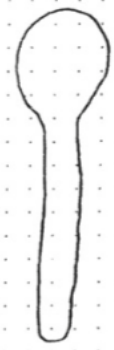
knife



plate



spoon

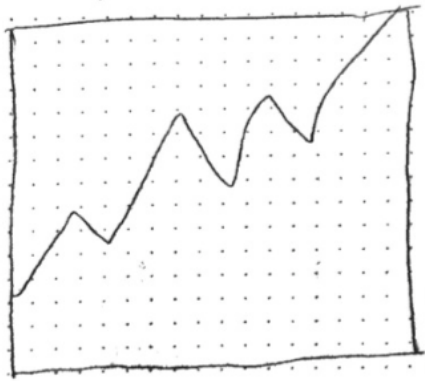


Statistics

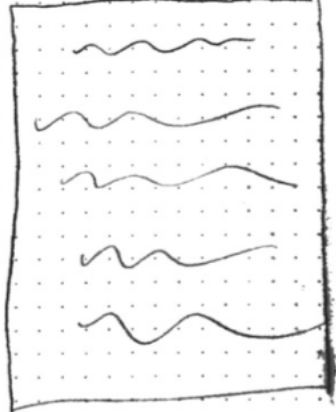
Level 6

Words spoken : 40

New words used : 5



Corrections



## SELECTED INTERFACE DESIGN

Our selected design interface was a combination of specific components from each design that we sketched. As an application that focuses mainly on the integration of context sensitive visual and aural language acquisition into one's daily life, we decided to center our interface around a smart eyewear platform. Using a smart eyewear platform allowed for easy integration into the users lives and an intuitive way to pick up visual and audio information from the users context.

Since our application supports the task of intuitive suggestion, translation, and pronunciation, we thought that smart eyewear would provide an advantage over our mobile phone alternative. Seeing that most of our users would be using this functionality when trying to communicate with others, we wanted the ability to display suggestions, look through translations, and see/hear pronunciation to be as unobtrusive as possible. Being able to project these to the user directly in their field of vision was a direct edge over the mobile phone option, which would have required users to momentarily divert their attention away from the conversation to read off of their device. The placement of speakers already next to the ear in most smart eyewear also allowed for simple audio delivery of pronunciation. When considering the mobile phone option, we realized that the current forms of audio delivery would have been rude or distracting, requiring users to wear headphones or take time out of the conversation to bring their device close to their ears.

The smart eyewear option also provided an effective way to provide contextually sensitive language support. The ability for smart eyewear to key into the users field of vision made it a better choice over the mobile phone for learning about the current context. Not only could the eyewear use location information to figure out the context, but it could also easily gather real time visual information. The mobile phone, on the other hand, could only obtain location information and static images from online databases, and would require unintuitive or time consuming methods for the user to gather real time visual information. The placement of smart eyewear on the face also meant that gathering audio data to use would require no external and intrusive placement of microphones that could distract from a conversation.

In order to provide a good feedback and progress, we decided to split the feedback mechanisms among the smart eyewear and the mobile phone interfaces. Immediate feedback mechanisms were integrated into the smart eyewear interface, since it would have to be non-interruptive and easily accessible during conversation. However, we realized that more latent forms of feedback, such as viewing progress and reviewing

mistakes, usually happened after conversation and therefore did not need to be as integrated into the users workflow. At the same time, displaying and navigating through all the information would have resulted in a cluttered and confusing interface for the small screen sizes and limited gestures of smart eyewear. Therefore, displaying this information on a mobile platform was the more natural choice.

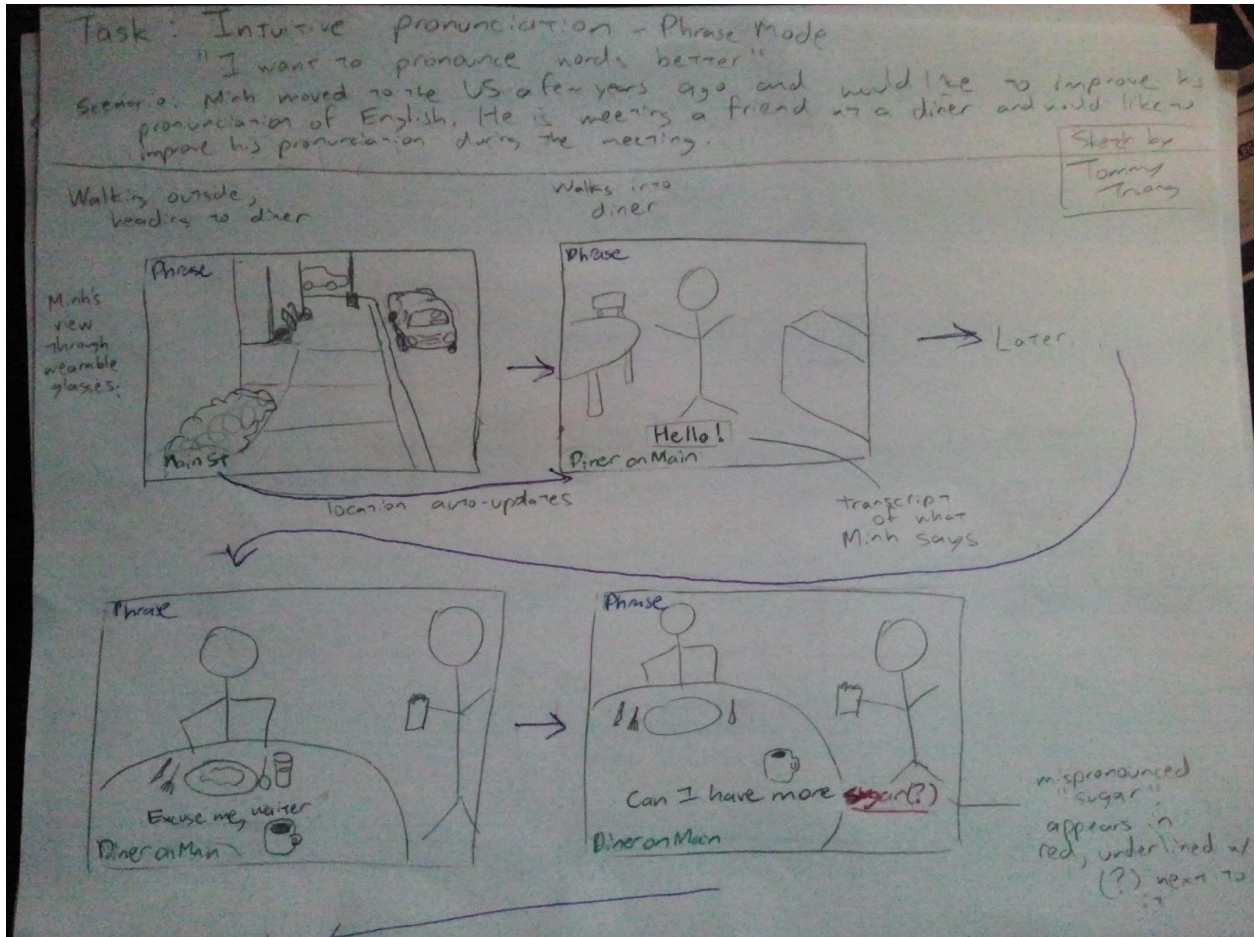
<b>FUNCTIONALITY</b>	<b>DESCRIPTION</b>
Focus Mode	Focus mode is an interface that allows users to focus on various objects around them while virtually displaying their names. The names displayed are location sensitive, using terminology relevant to the area the user is in.
Phrases Mode	Phrases mode is an interface that allows users to view context sensitive phrase suggestions. As the user speaks, phrase suggestions narrow down and act as autocomplete options to their current sentence. The phrases displayed are location sensitive, using terminology relevant to the area the user is in.
Translation	Users can swipe through translations of words and phrases in their native language.
Pronunciation	Users can swipe through phonetic translations of words and phrases to better understand pronunciation. They can also hear how words are pronounced. The accent of these pronunciations are location sensitive.
Immediate Pronunciation Correction	As users speak words in the foreign language, the words are highlighted in different colors to represent the accuracy of their pronunciation.
Statistics	Users can access statistics of their overall progress. These statistics include things such as the number of words learned and the number of words

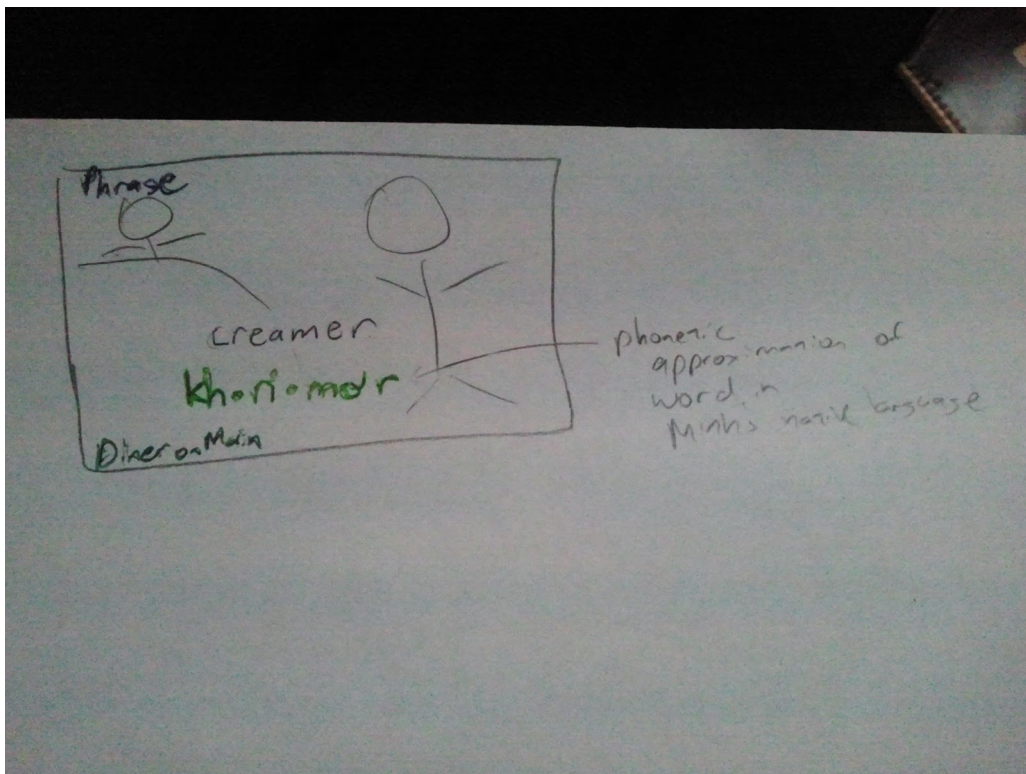
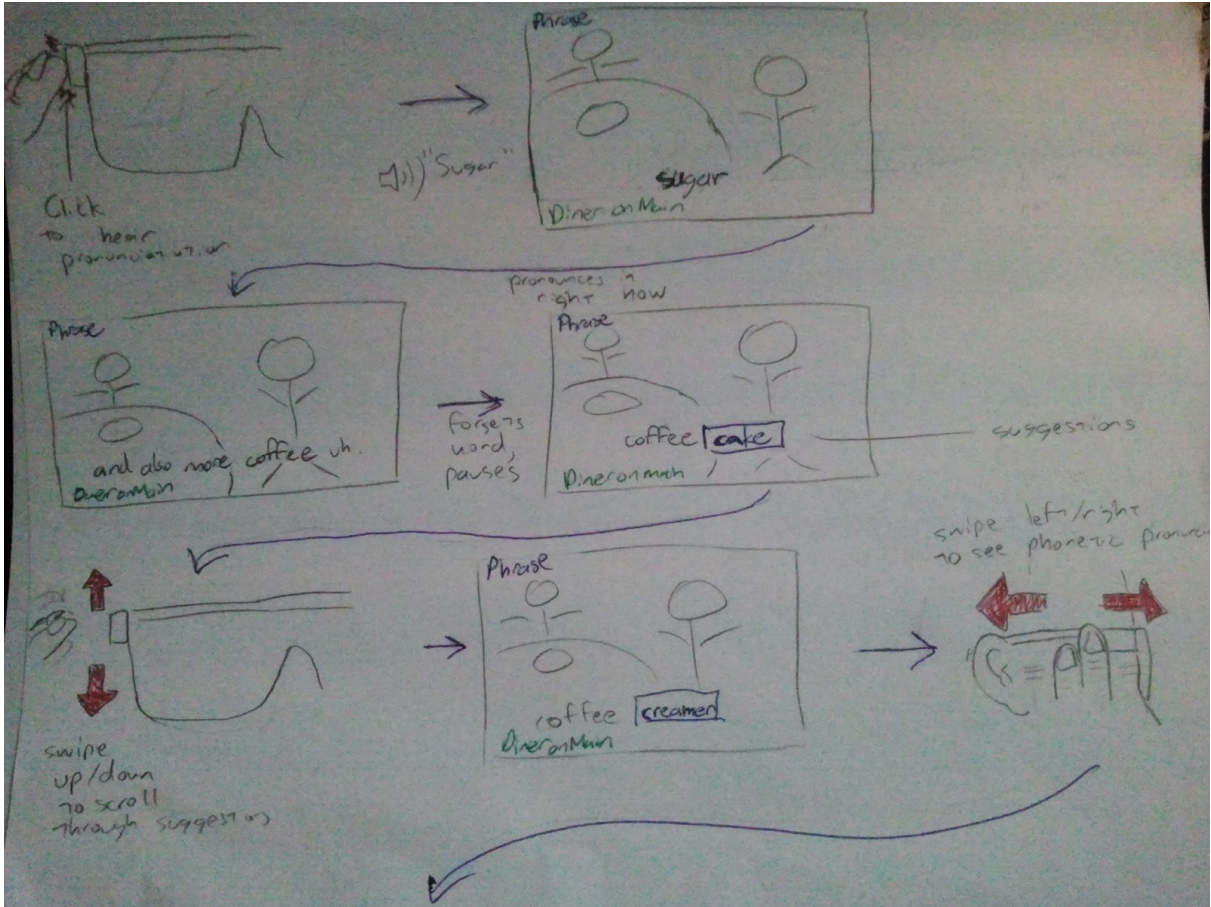
	missed.
Word Practice	Users can practice the pronunciation and meaning of words that they missed.



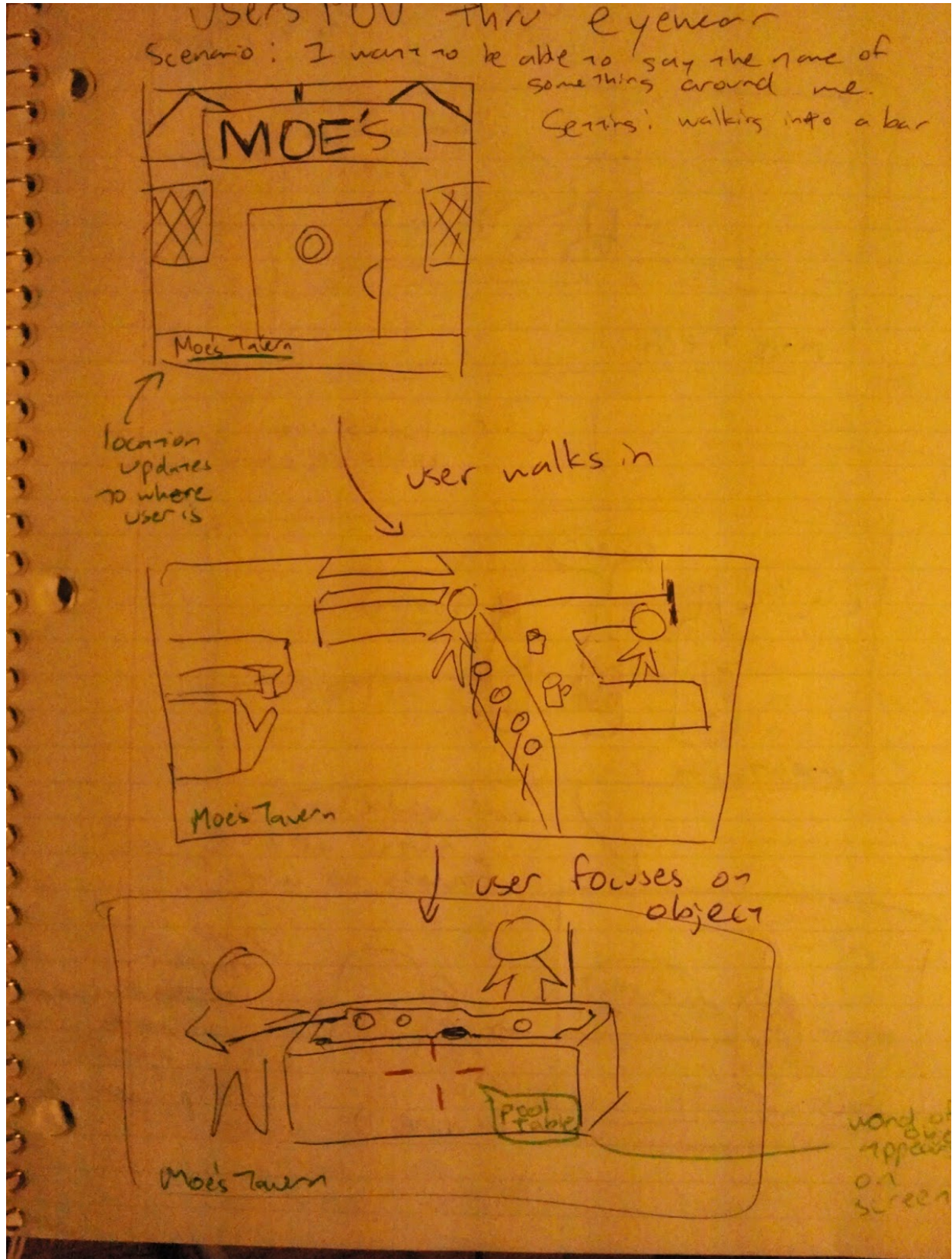
# UI STORYBOARDS

## Scenario for Task 1: Intuitive Pronunciation





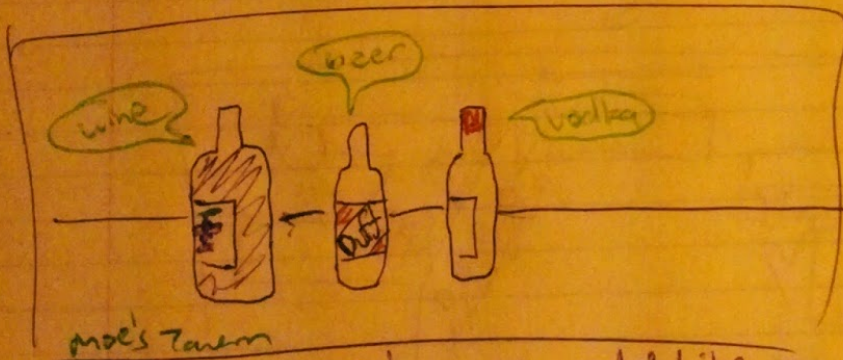
# Scenario for Task 2: Contextually Sensitive Language Support



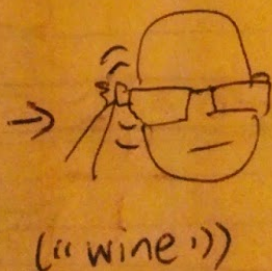
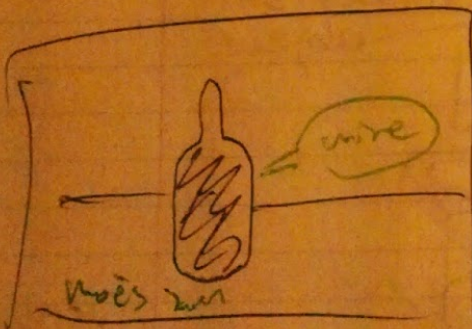
User turns around to order drink



User looks at available choices

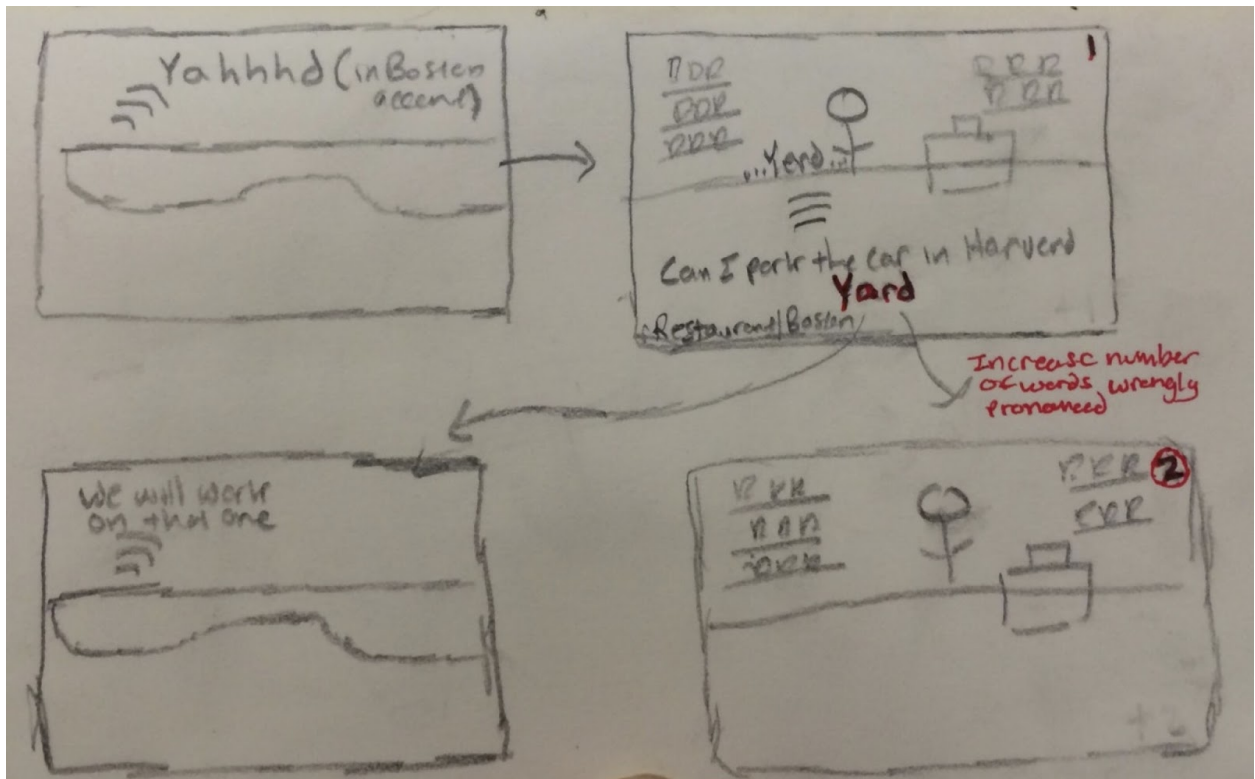
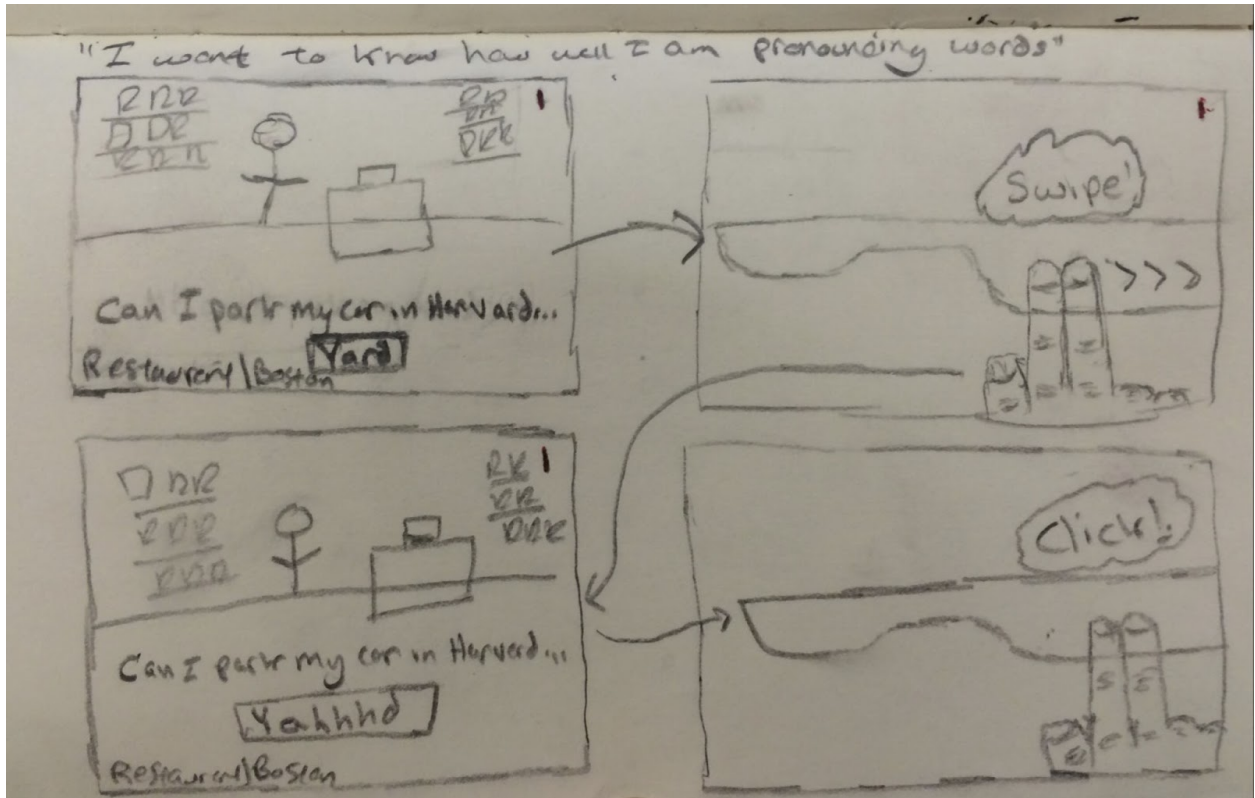


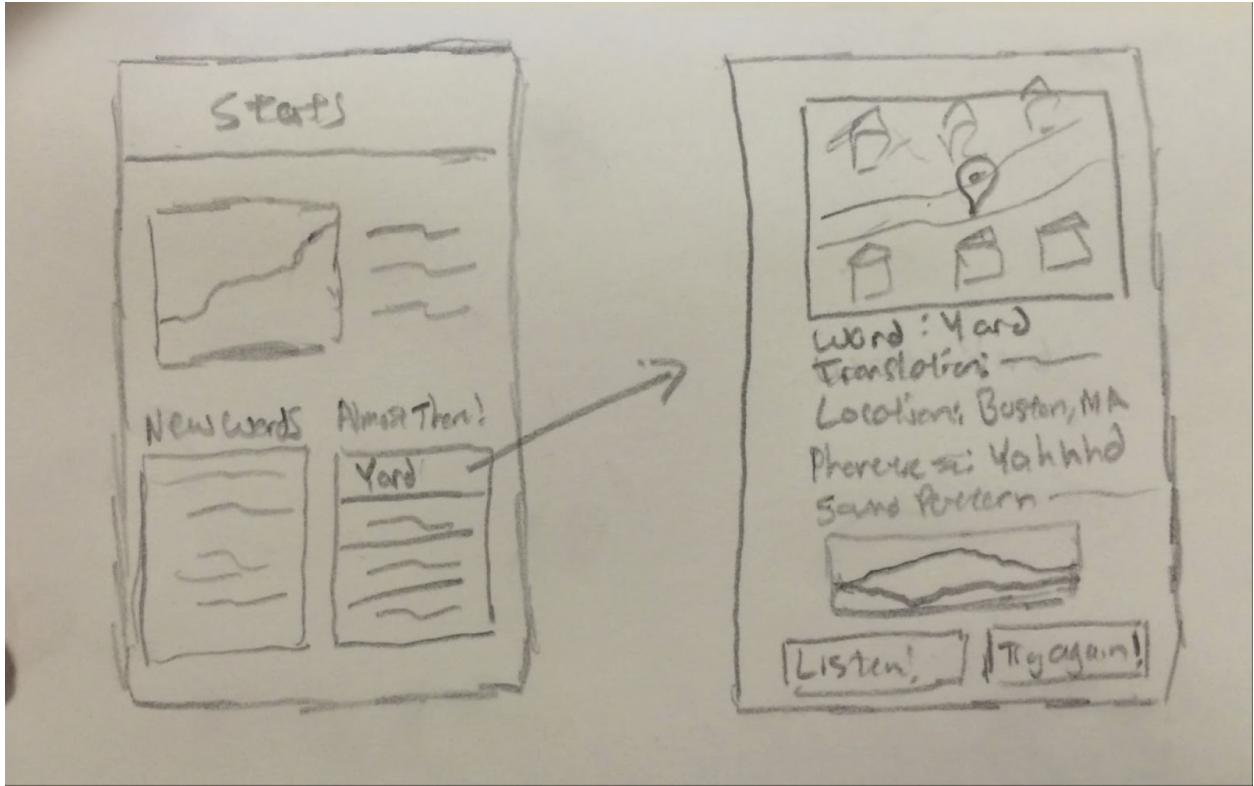
User would like a glass of wine, focuses on wine



User taps ~~eyes~~ eyeglasses to ~~hear~~ hear how to pronounce

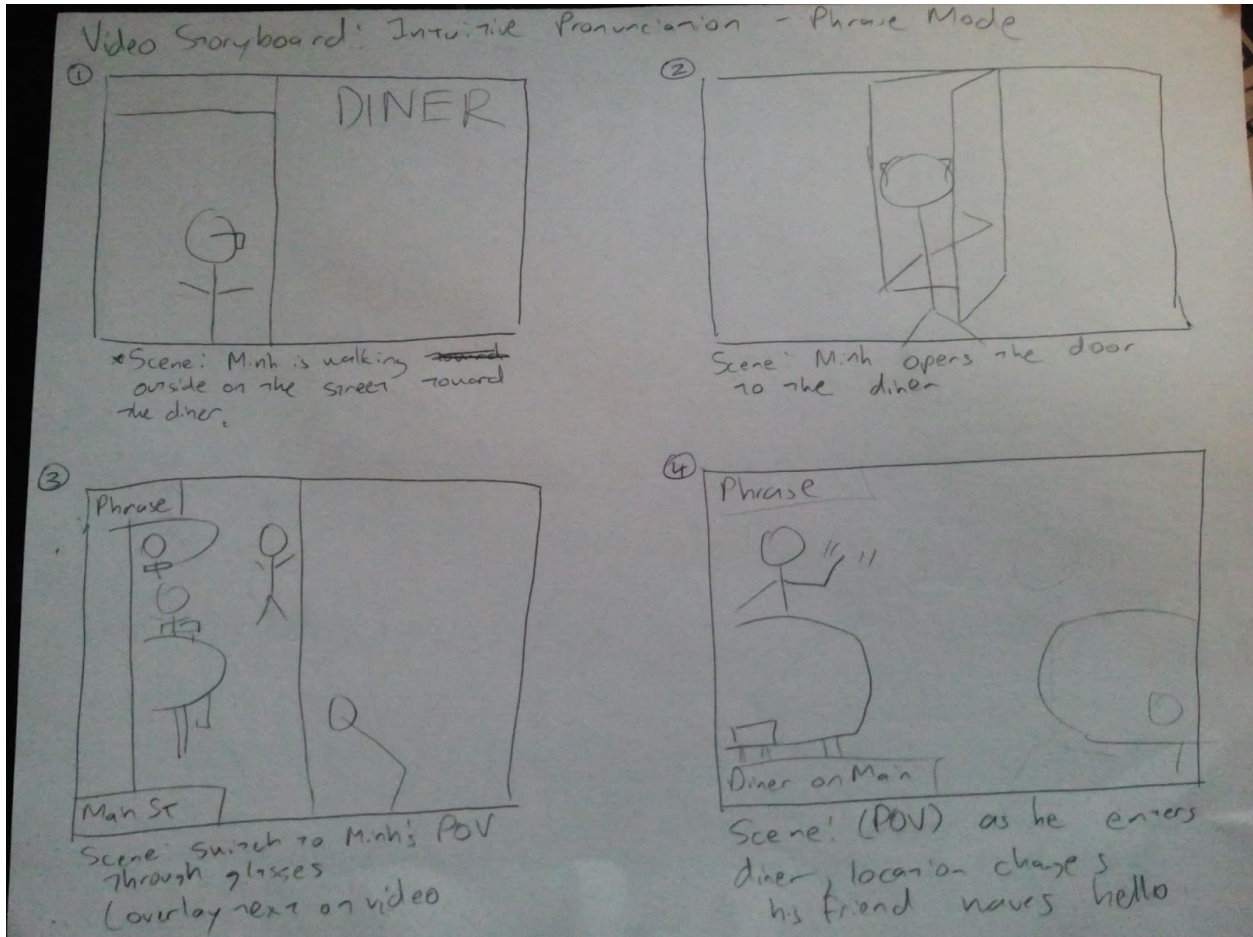
Scenario for Task 3: Active Language Learning Integration





# VIDEO PLANNING STORYBOARDS

## Scenario 1: Intuitive Pronunciation

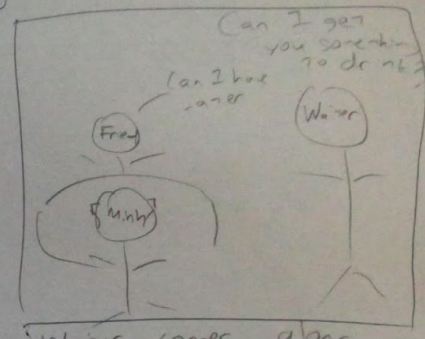


5



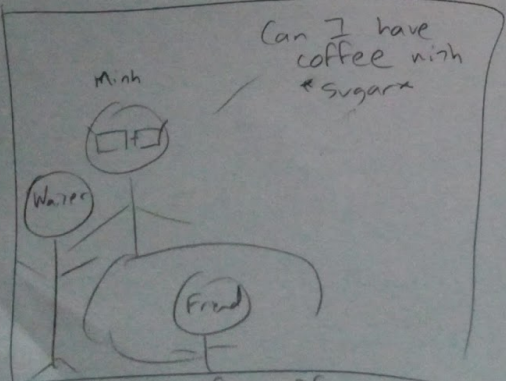
Switch back to third person POV, Minh & friend talk (and eat?)

6



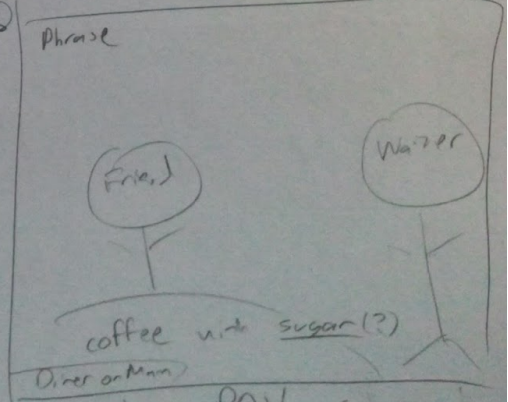
Waiter comes along, asks what they want to drink

7



Minh asks for coffee with sugar, but mispronounces sugar

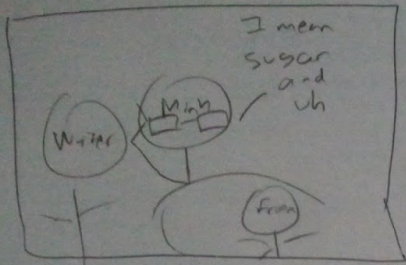
8



Phrase  
 Diner or Man  
 Switch to POV, sugar appears in red

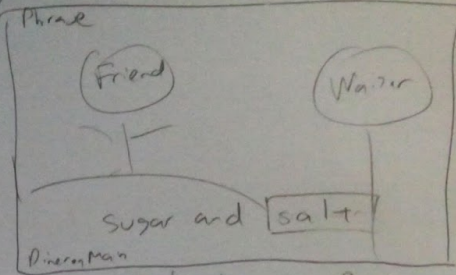


9



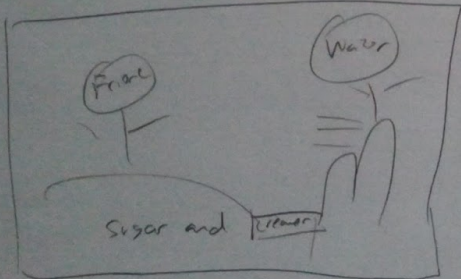
Mink taps on glasses, pronounces it correctly now, but forgets the word for "creamer"

10



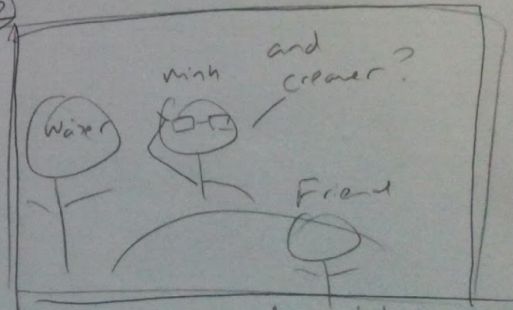
Switch back to POV, now has word suggestions

11



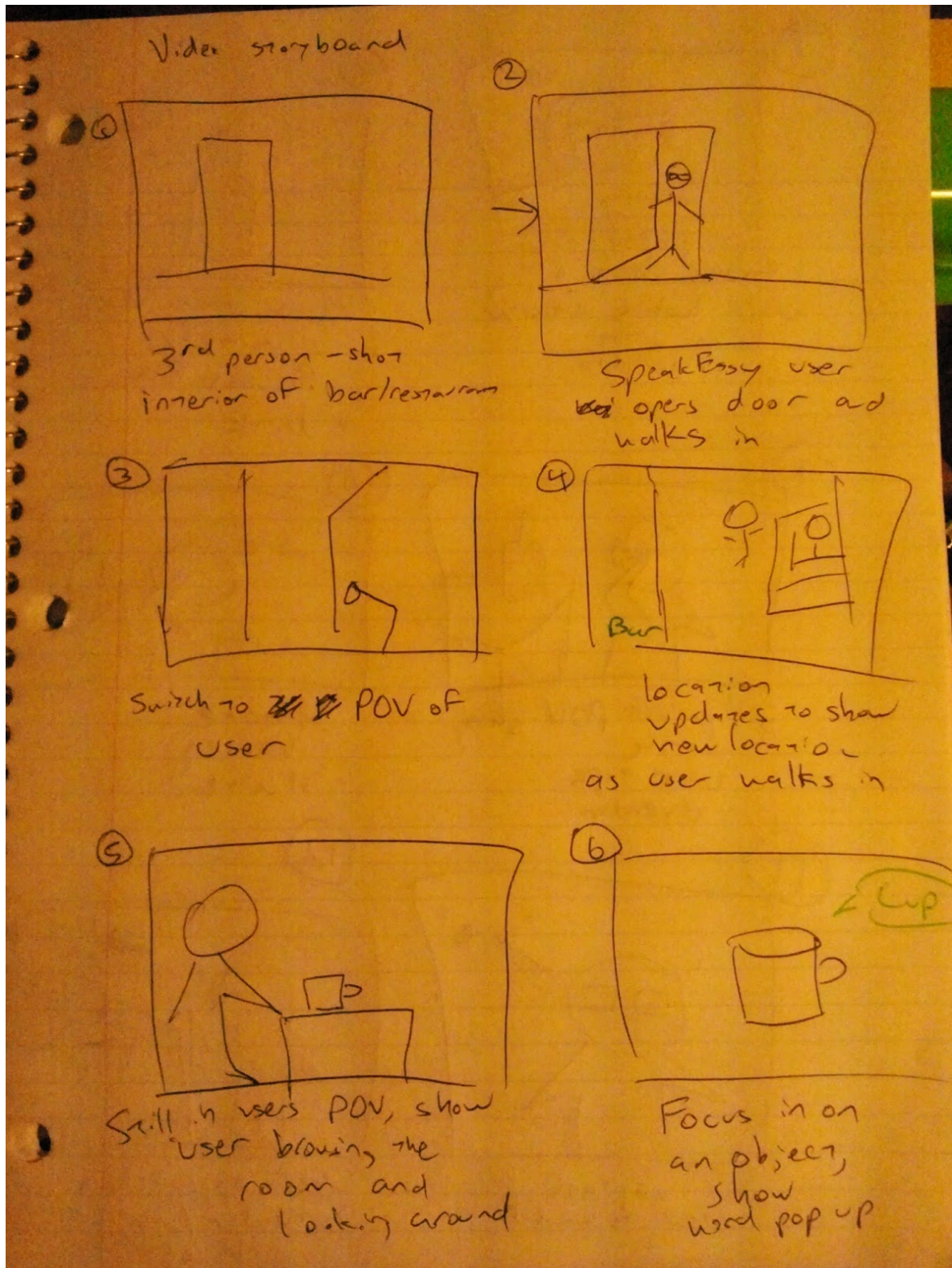
Still in Mink's POV, see finger swiping through options on side of video until it reaches creamer

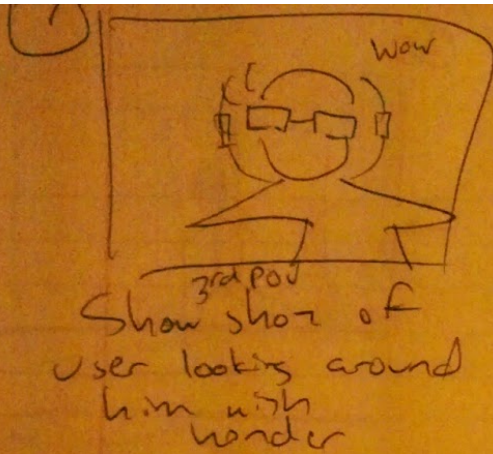
12



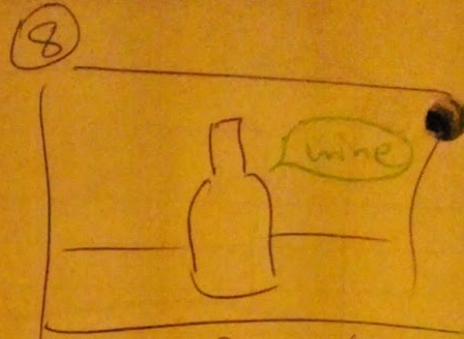
Mink remembers the word he wants to say and says it after tapping glasses

## Scenario 2: Contextually Sensitive Language Support





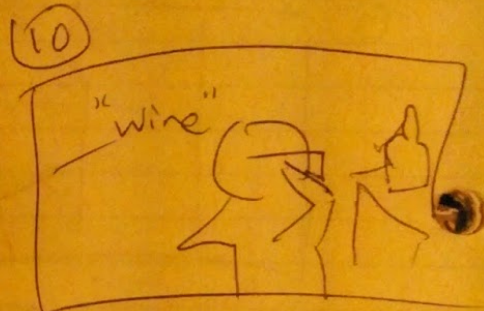
Show shot of user looking around him with wonder



Show from users view as he focuses on a drink



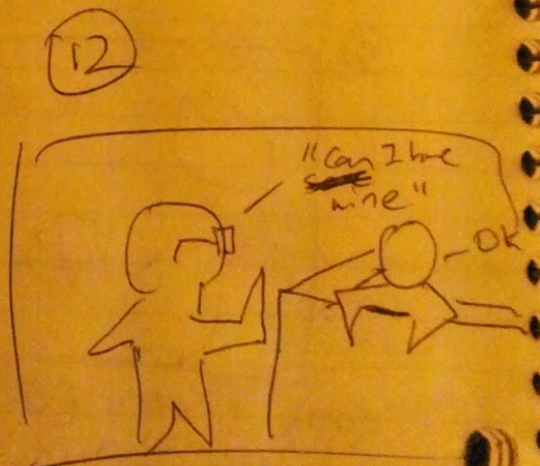
3rd person POV shot, the user taps eyewear



Someone voices "wine"

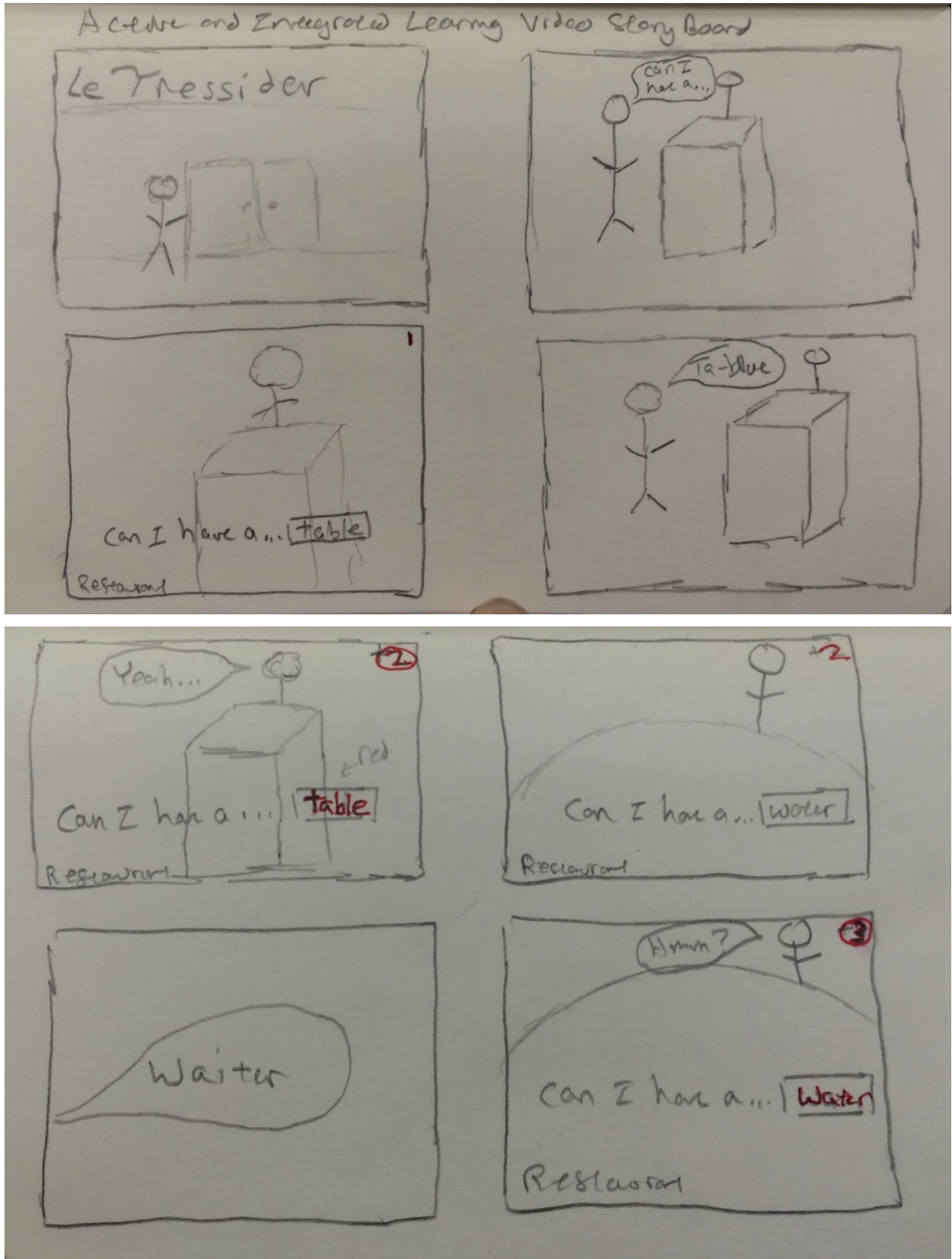


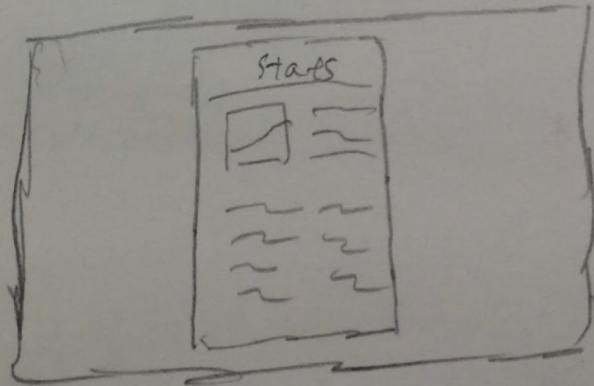
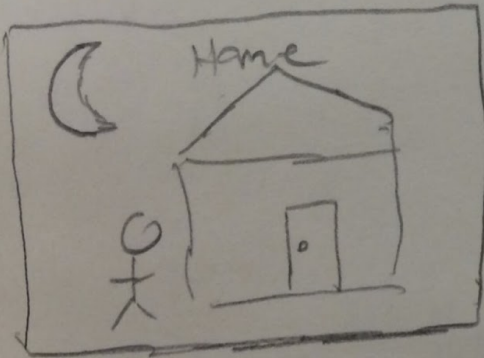
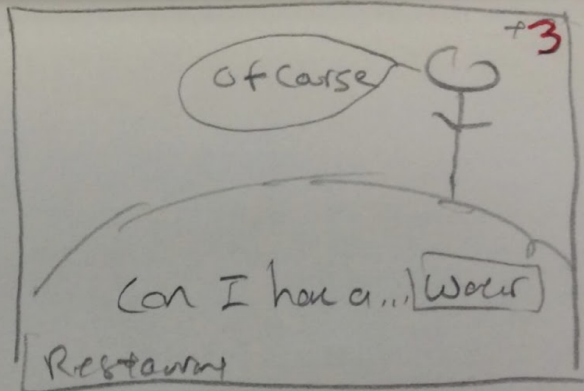
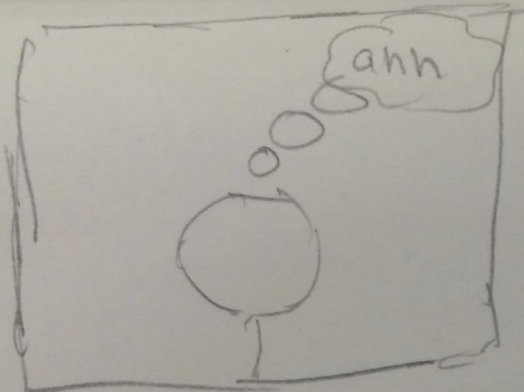
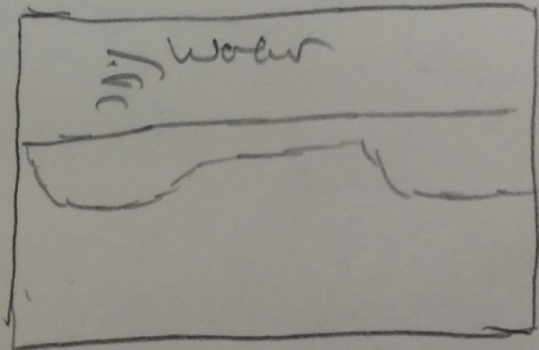
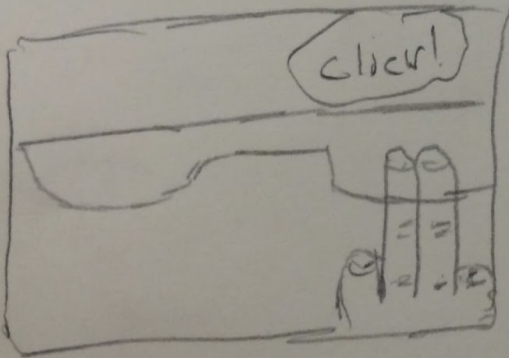
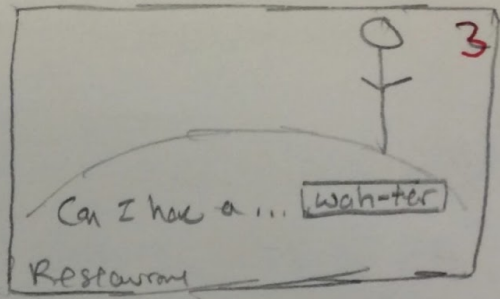
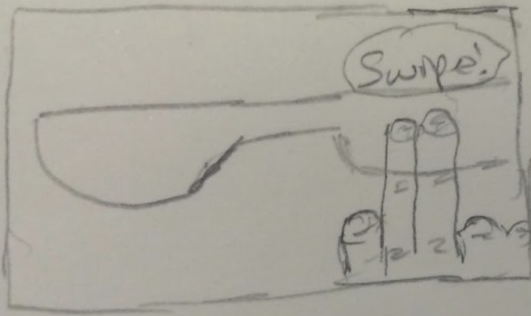
User repeats word to himself



User approaches bartender / waiter and orders

Scenario 3: Active Language Learning Integration





## CONCEPT VIDEO DESCRIPTION

### 1. What was difficult?

Editing the video was the most difficult part of the process. It took significant time to cut down the clips and use transitions to make the story work. In order to show our concept of having integrated language learning, we overlaid text in our video to simulate the smart eyewear part of our solution. It was especially difficult to line up our smart eyewear interface with the video.

### 2. What worked well?

Coming up with the video concept was the easiest part. It took longer to create the actual script for our video. Shooting the video went smoothly because we came up with detailed scenes before beginning to shoot.

We thought our video effectively demonstrated the different tasks that the main character of the video could do with it: learning phrases specific to his current context and location, learning to pronounce words correctly, and integrating language learning into his daily activities. Instead of explicitly labeling the three tasks in the video, we thought it was more natural to show a continuous story about the main character and the different situations he could navigate through with the app.

### 3. How long did it take for each phase of design prep, shooting, editing?

We spent an hour and a half preparing to take the video. We came up with scenes based on our video storyboards and figured out the locations we were going to use. Then we wrote an outline of our script for each scene.

We spent roughly two hours shooting the video. We used three locations and shot over two days. It took roughly four and a half hours to edit the video.

## VIDEO LINK

<https://www.dropbox.com/s/yo2g2c4g9q4fs8h/CS%20147%20Concept%20Video%20v2.mp4?dl=0>