+ neuron

UX Architectural Guide

April 2017

TABLE OF CONTENTS

EXECUTIVE SUMMARY

OVERVIEW 3

MAP CREATION 4

MAP READING 5

WIREFRAMES

MAP CREATION 6

MAP READING 8

Overview

EXECUTIVE SUMMARY

The Waytuit app proposes using computer vision as an indoor navigation tool for visually impaired people. Computer vision is used to create 3D models of indoor environments and embed crowdsourced semantic annotation. It relies on sighted volunteers with smartphones to act as distributed physical crawlers to collect video footage of a building's interior. These videos are used to build a 3D point-cloud model of the space for localization and wayfinding. Once video has been collected, an online annotation interface allows remote crowd workers to add labels for important visual cues. These labels are then embedded in the underlying 3D model. Someone who is vision impaired can locate themselves to sub-meter accuracy simply by capturing a photo of their surroundings on a smartphone. Once their location is determined, all labels in their vicinity are easily retrieved and available to them relative to their current orientation.

CREATING THE UNDERLYING 3D MAP

1. User must enter their location (e.g. Stanford Shopping Center) before beginning recording video of an indoor environment.
2. User is informed of whether a 3D map exists for this location. If no 3D map exists for selected location, user is instructed to begin video recording at threshold of main entrance. If 3D map exists for selected location, user has option to begin video recording at one or multiple confirmed indoor locations using dead reckoning.
3. Once the user's starting location is known, they are able to begin recording video.
4. Using the camera of their smartphone or other capable device, the user captures their surroundings as they move through the indoor space. The environment is recorded as a series of still-frame images at ~4fps, rather than video, which is typically 30fps or higher.
5. The user may end the recording at any time.
6. Once recording is ended, the user will be asked to review the footage quality before submission.
7. Using the dead-reckoning point and the smartphone's build-in sensors (accelerometer, gyroscope, magnetometer, and altimeter), the recorded still-frame photos are stitched together to generate a 3D map.

ADDING LABELS TO THE UNDERLYING 3D MAP
(can be done from remote location)

1. The still-frame images are scrubbed for any text that is recognizable. This text is then converted to a label and embedded in the 3D map.
2. Additional labels are added by remote crowd workers via an online annotation interface. They will be tasked with clarifying any illegible text as well as labeling important, non-textual visual cues in video still-frame photos (e.g. drinking fountains, elevators, etc).

DETERMINING THE USER'S LOCATION

1. To access an indoor map, visually impaired users must "check-in" to a location for which a map has been generated. This is necessary to optimize performance, as it will reduce the queryable data to a much smaller subset.
2. To locate the user within the 3D map, they must capture their surroundings using the camera of their smartphone or other capable device. Surroundings are recorded as a series of still-frame photos in the same way the environment is originally mapped.
3. Using SIFT features from the captured photos, the user's location within the 3D model is determined.

INTERACTING WITH THE 3D MAP

Exploratory: Once location is determined, the user may retrieve all labels in their vicinity and display/verbalize them relative to their current direction. Using the smartphone's built-in sensors (and continuous photo capture, if necessary) real-time localization is performed to provide contextual information to the user as they move around the environment.

Destination Routing: Once location is determined, the user may also enter (via speech-to-text) a specific destination within the indoor environment (e.g. menswear department) and be routed here using a combination of metric-based directions (i.e. walk 20 steps ahead) and landmark-based directions (i.e. walk until you have passed the cashier on your left).

SUPPLEMENTARY AIDS

Text Recognition: In addition to the 3D point-cloud representation, the application will also allow visually impaired users to use the camera on their smartphone or other capable device to identify text and read it aloud (using Google Text Recognition API). This would be a particularly useful feature for reading product labels in stores.

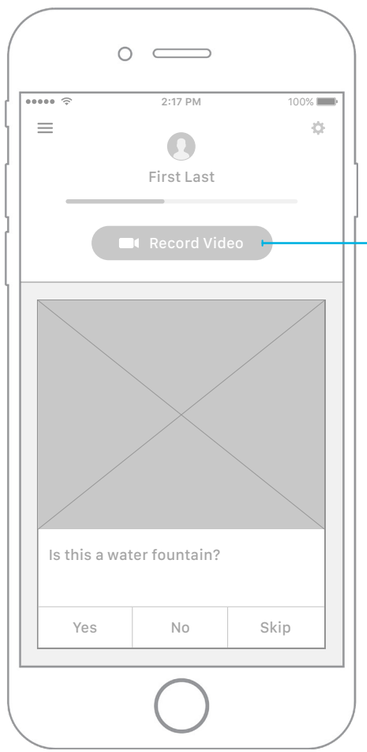
FUTURE CAPABILITIES

Outdoor Locating using Google Street View

1. Map Creation

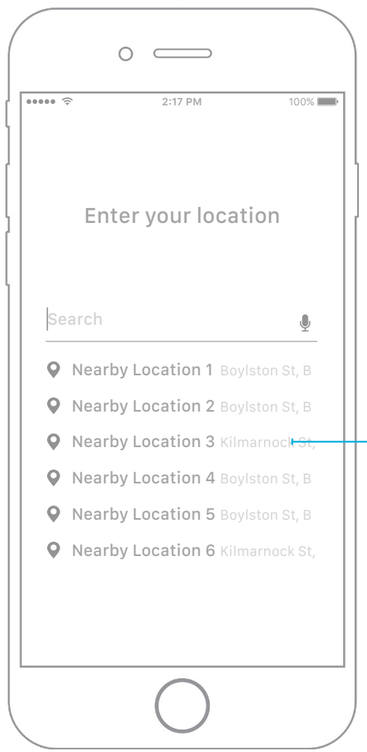
WIREFRAMES

1.A SIGHTED USER HOMEPAGE



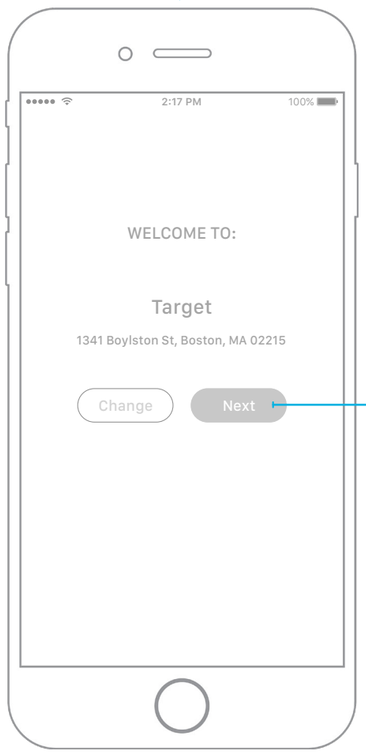
A sighted user has to option to record a new video of a building's interior (must be on-location) or can contribute to the community database by helping to label key frames of pre-recorded videos. In this case, the user choosed to record a video

1.B LOCATION ENTRY

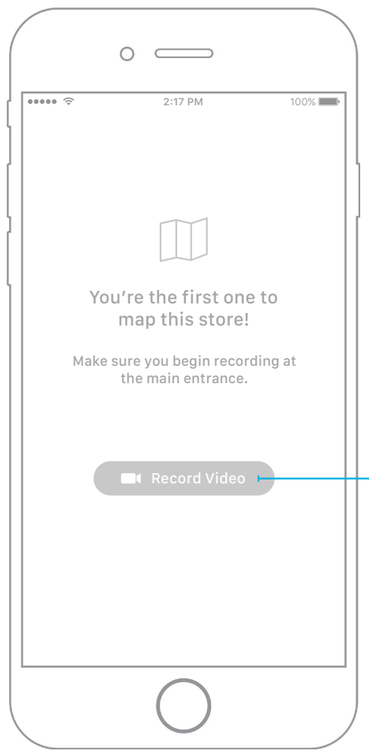


User must enter their location before beginning recording video of an indoor environment.

1.C LOCATION CONFIRMATION



1.D EXISTING MAP
No Map Exists



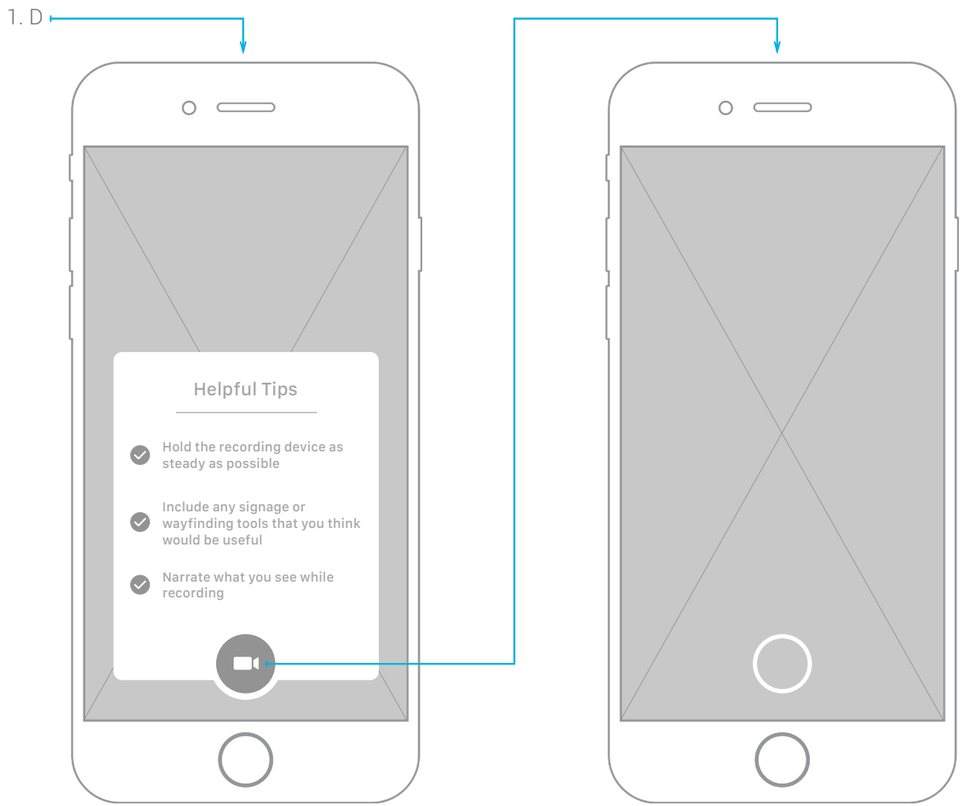
If 3D map exists for selected location, user has option to begin video recording at one or multiple confirmed indoor locations using dead reckoning.

If no 3D map exists for selected location, user is instructed to begin video recording at threshold of main entrance.

1. Map Creation cont'd

1.E **RECORD VIDEO**
Helpful Tips (Unchecked)

1.F **RECORD VIDEO**
Helpful Tips (Checked)

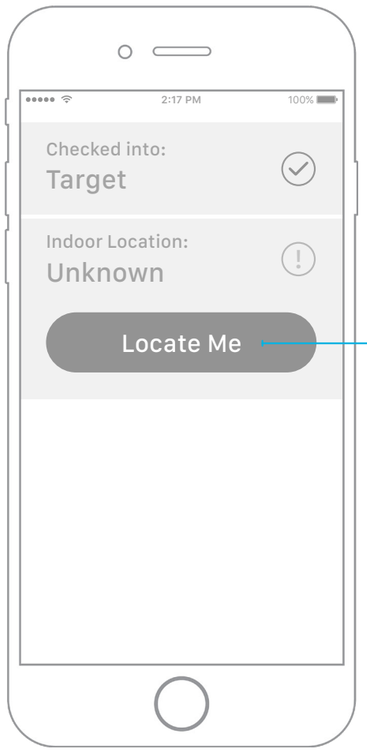


Using the camera of their smartphone, the user captures their surroundings as they move through the indoor space. The environment is recorded as a series of still-frame images at ~4fps.

Using the dead-reckoning point and the smartphone's build-in sensors (accelerometer, gyroscope, magnetometer, and altimeter), the recorded still-frame photos are stitched together to generate a 3D map.

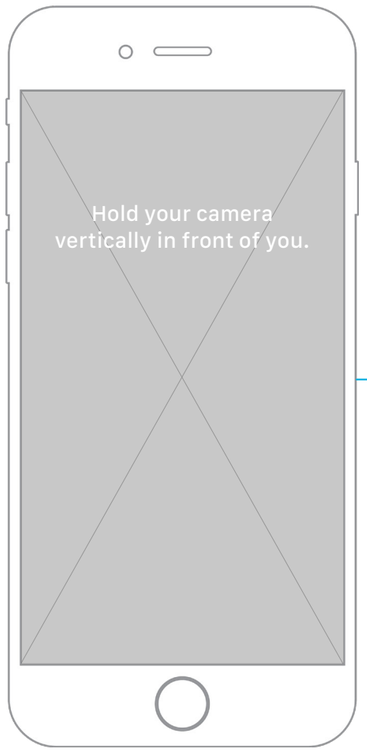
2. Map Reading / Determining the User's Location

2.A **BLIND USER HOMEPAGE**
Indoor Location: Unknown



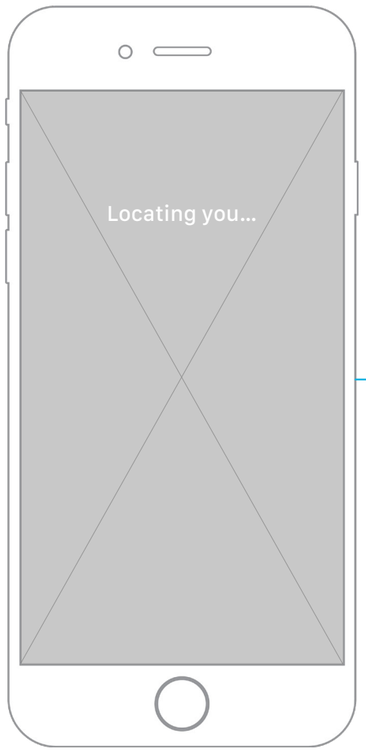
To access an indoor map, visually impaired users must be “checked-in” to a location for which a map has been generated and must also be located within the indoor environment.

2.B **LOCATE ME**
Step 1



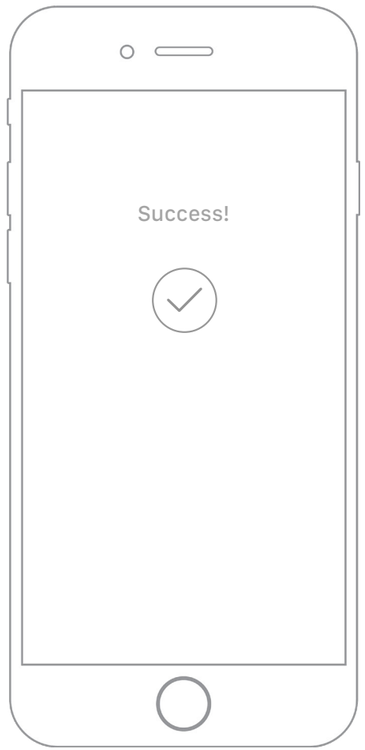
To locate the user within the indoor environment, they must capture their surroundings in the form of multiple still frame photos using the camera of their smartphone.

2.C **LOCATE ME**
Step 2



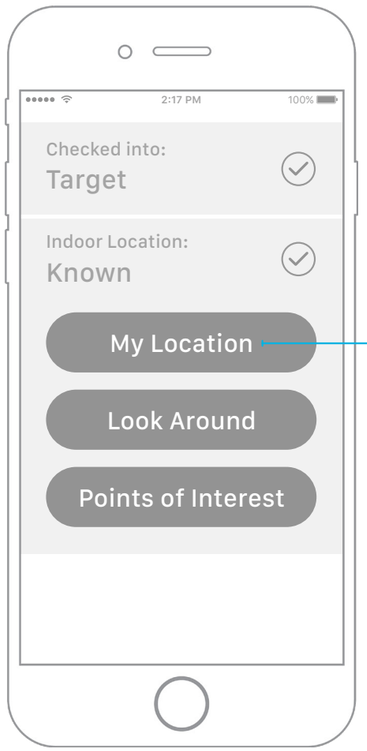
Using SIFT features from the captured photos, the user’s location within the 3D model is determined.

2.D **LOCATE ME**
Location Determined



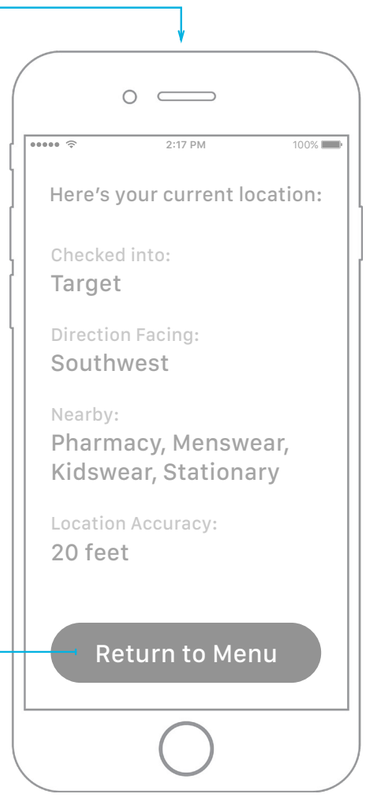
3. Map Reading / My Location

3.A **BLIND USER HOMEPAGE**
Indoor Location: Known



Once location is determined, the user has the ability to interest with the 3D map.

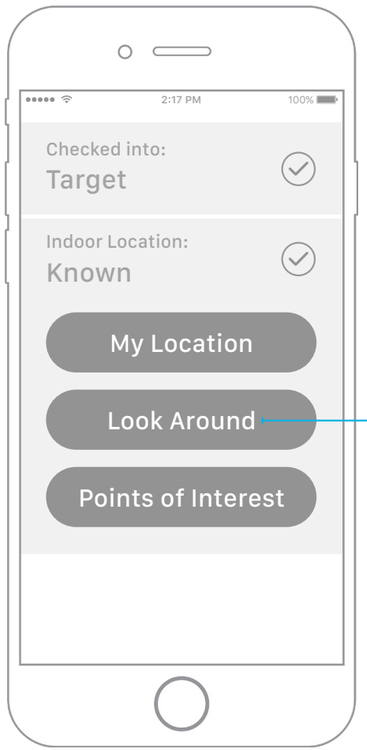
3.B **MY LOCATION**



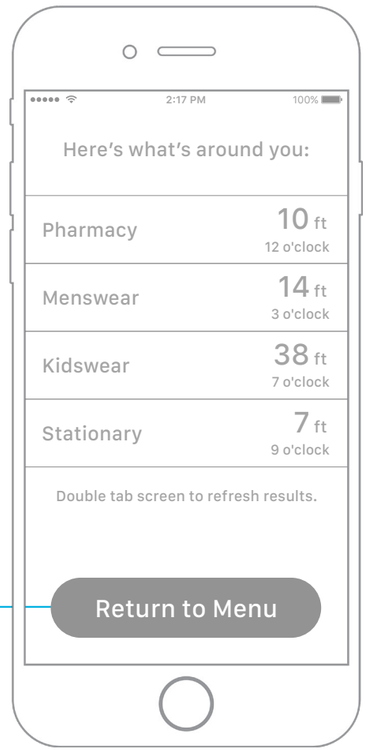
Information about the user's current location is displayed and read aloud using text-to-speech.

4. Map Reading / Look Around

4.A **BLIND USER HOMEPAGE**
Indoor Location: Known



4.B **LOOK AROUND**



4. A

Information about the user's surroundings is read aloud to them. 12 o'clock position is contextual to the user's current orientation at the time when the query is made.

5. Map Reading / Points of Interest

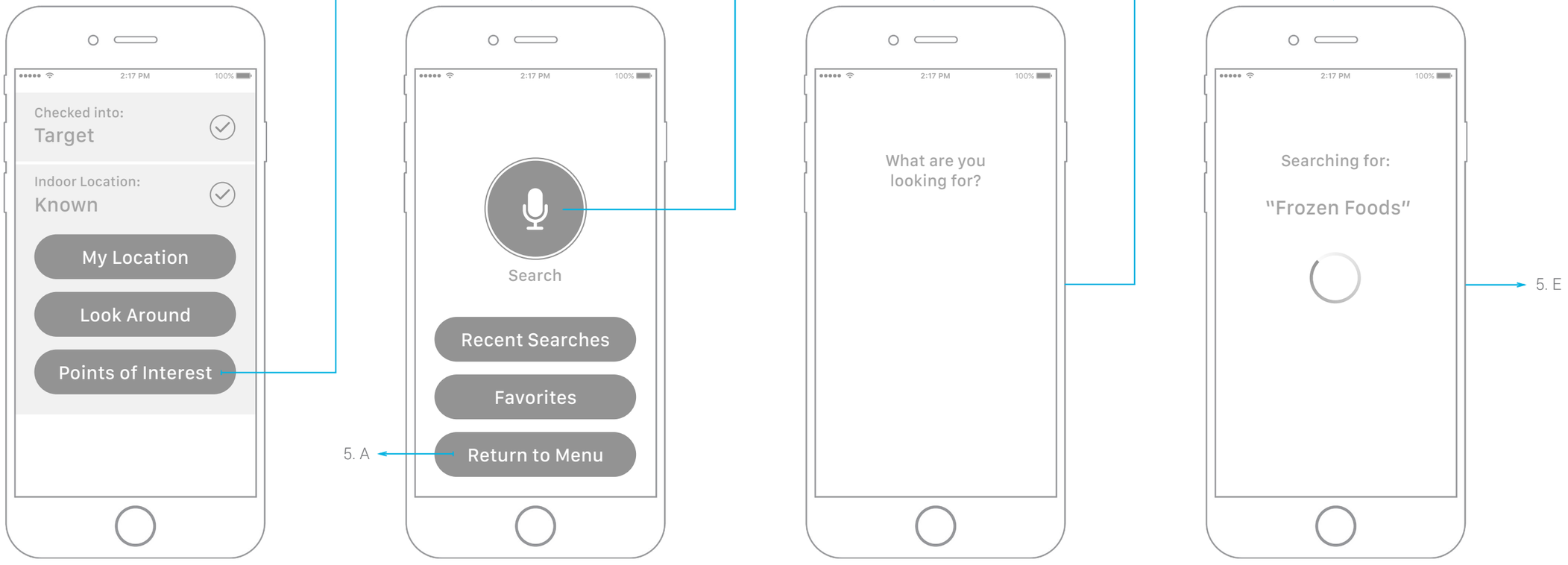
WIREFRAMES

5.A BLIND USER HOMEPAGE
Indoor Location: Known

5.B POINTS OF INTEREST
Menu

5.C POI SEARCH
Listening

5.D POI SEARCH
Searching

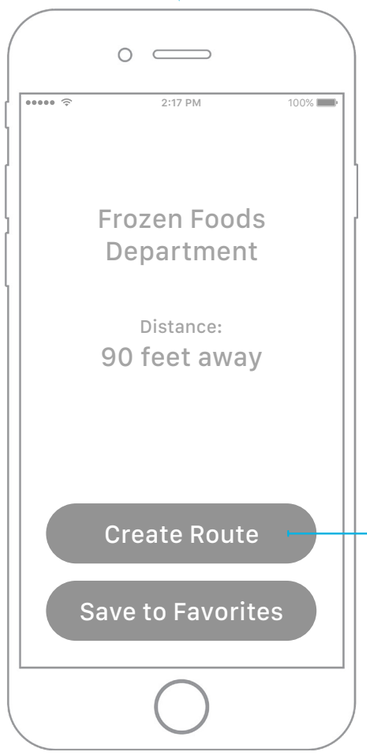


A user may search for an indoor POI either by keyword query using speech-to-text or retrieve a POI that was recently searched or saved to favorites.

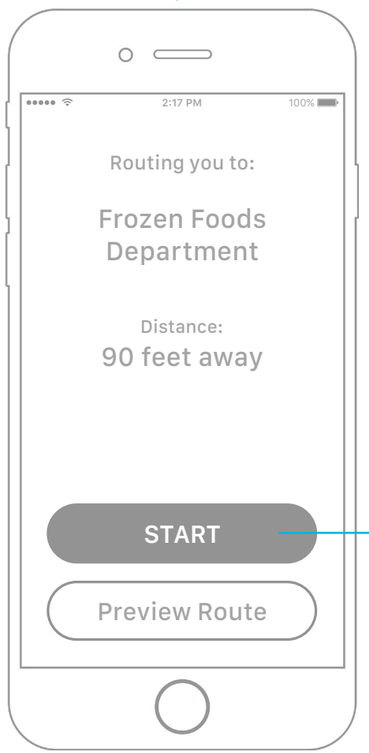
5. Map Reading / Points of Interest cont'd

5. D

5.E POI SEARCH RESULTS



5.F POI ROUTE CREATION



5.G POI ROUTING
In Progress

The route auto-advances using the smartphone's build-in sensors (accelerometer, gyroscope, magnetometer, and altimeter).