

ESTi

A digital voice assistant that prepares ready-to-review home construction scope of work reports.

Interaction Specification
Angela Yung + Tiffanie Horne

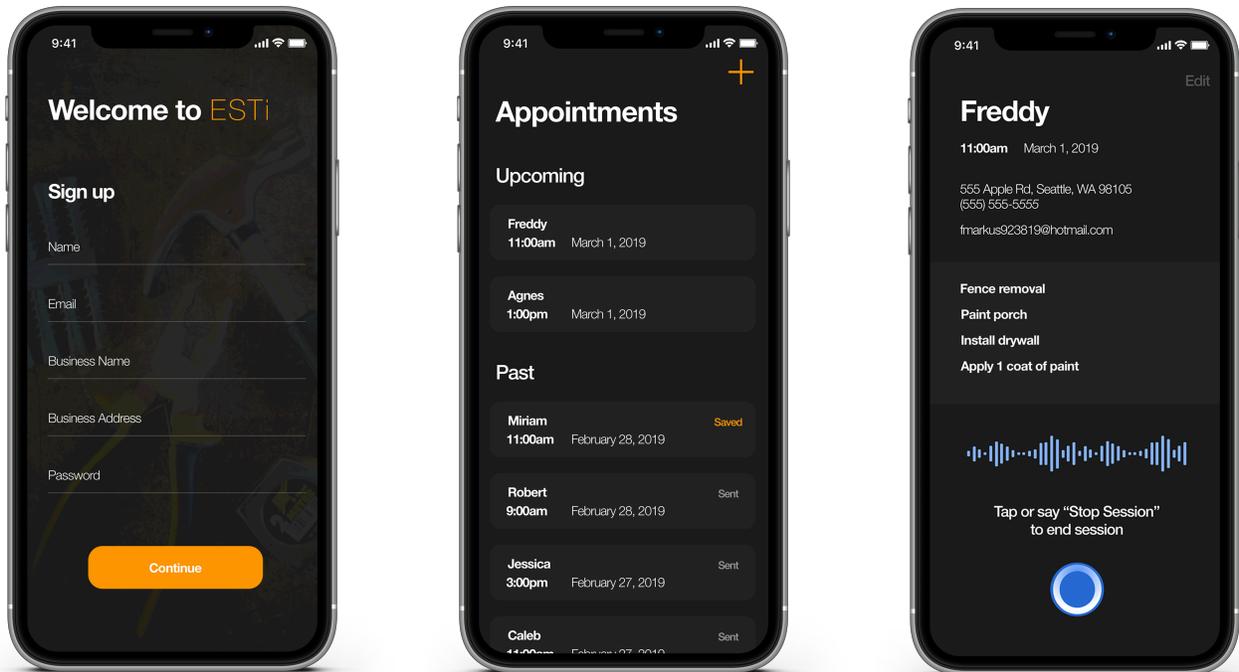


Figure 1. ESTi voice interface

OVERALL CONCEPT

ESTi is a smartphone-based voice assistant that converts speech into a ready-to-view, concise scope of work report (SOW). A SOW is a vital component of every successful home improvement construction project and the first step in producing a quote for a client. When a home improvement contractor meets a client to discuss the SOW, ESTi takes action through the use of Natural Language Processing (NLP)[1]. ESTi uses varying web services to request, aggregate and validate data from commercial database sources and interoperate with the smart-phone based device to populate SOW summary.

ESTi addresses the ongoing need of cost estimating to be made faster and more efficiently for professionals working in the residential construction industry by addressing the first step: summarizing the scope of work. ESTi efficiently summarizes the scope of work allowing the general contractor to more quickly generate an estimate of the costs associated with building and overseeing the project. Current practices prevent automaticity during each stage of cost estimating, but especially when establishing the scope of work. Some contractors have well-established procedures, most do not [2]. This need for appropriate software solutions is where ESTi steps in.

ESTi target's a user base of general contractors; namely, home design, improvement, and remodeling professionals who use a smart mobile phone and need a hands-free transcription tool. As an example, consider a contractor providing an estimate of all the cost associated with replacing a leaky roof. Imagine this contractor desires to secure the job. Contractors often meet the homeowner at their home to conduct on-site inspections. They use their cell phones to take pictures and make video notes while some jot notes on a pre-generated costs estimate form. Current methods to devising the scope-of-work require clarifying, listing, and summarizing the clients requirements. Organizing photos, sorting through video footage and reviewing Current practices are problematic because they are administrative intensive, time-consuming, and lack efficiency.

USER JOURNEY

Micah is an independent contractor who specializes in home improvement.

Business is booming which is great but Micah is now having trouble keeping up with pr quotes to her clients. Her turnaround time is about a week because of the time it takes review the videos she takes during appointments and gather estimates on parts and

Enter: ESTi, a contractor's companion. Micah is able to input all her appointments onto ahead of time. When the appointment starts, she goes to her app and taps on the client appointment. She says "start session" and begins her normal activities.

ESTi listens in the background to the appointment and generates estimates based on tl information that Micah and the client talked about. At the end of the appointment, Mica simply says "end ESTi" and the app will begin compiling an estimates summary. Micah have a chance to go through the summary to confirm or change any of the items. She c either save this review for later or complete it on site and send the estimates to the clier away.

EXPERIENCE DESIGN DETAILS

ESTi's capabilities are located within an app on your phone. The app uses a voice inter listen to your session and generate your estimates summary.

Setup: User Flow

Welcome to **ESTi**

Sign up

Name

Email

Business Name

Business Address

Password

Continue

Users will complete simple form to sign up and set up their account.

Let's get started!

I'll need to learn about your voice. Please record "start session" 5 times.

Continue

Users will need to record themselves on the app saying "start session"

Let's get started!

Now please record "end session" 5 times.

Continue

Likewise they will need to record themselves saying "end session"

Create an Appointment

Name

Email

Address

Appointment Date (Please Select) ▼

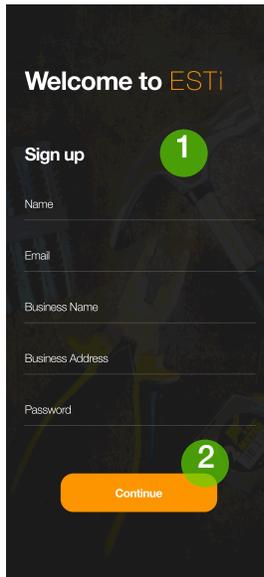
Appointment Time (Please Select) ▼

Save

Cancel

Once they have completed the voice setup, users can input their appointments which will show up on their homescreen after they are saved.

Setup: Annotated Screens



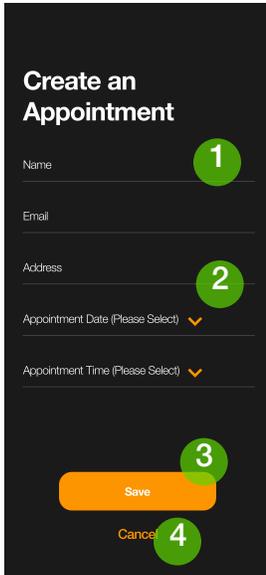
1 Users select a text line and are able to type in their details

2 Continue moves them to the next screen

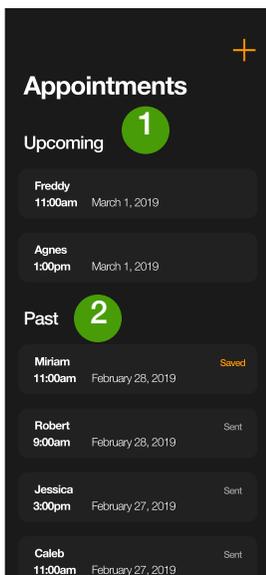


1 As the user speaks, a blue sound wave appears providing feedback to the user

2 Continue moves them to the next screen

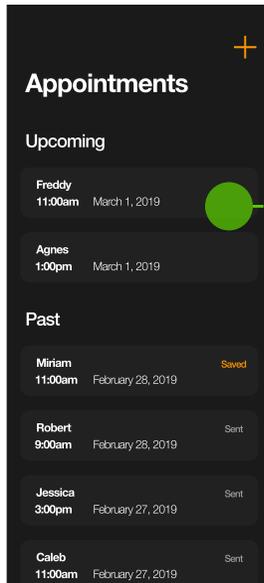


- 1 Users select a text line and are able to type in their details
- 2 To input date and time, users will click on the down arrow providing a drop down menu of dates/times
- 3 Users can save their appointment and will be redirected to their homepage where the new appointment will appear
- 4 Cancel will return them to the homescreen without saving the new appointment.

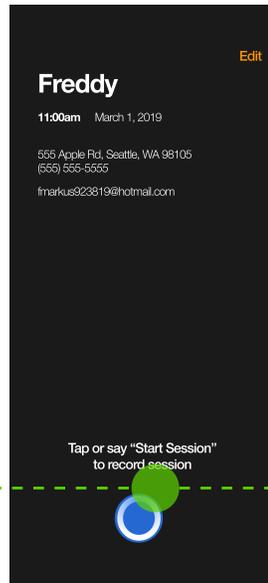


- 1 The home screen provides a list of the user's appointments, with the upcoming appointments appearing at the top in chronological order
- 2 Past appointments appearing below it in reverse chronological order.

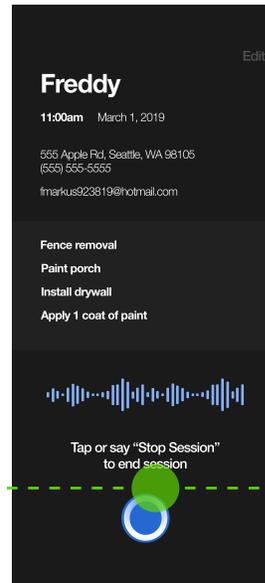
Main Screens: User Flow



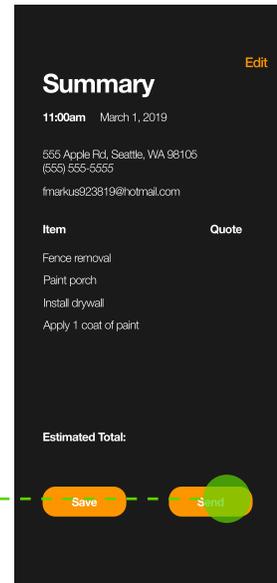
Users select an appointment



Users begin a session

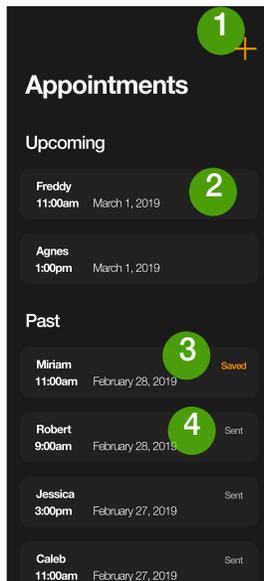


End recording to end session

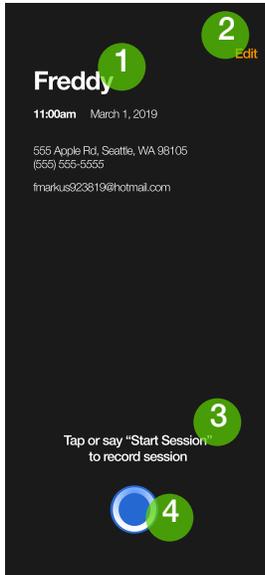


User updates the summary information and send's it to the client

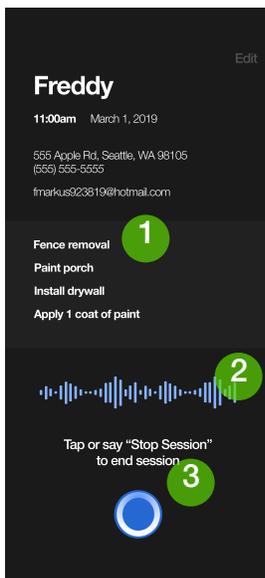
Main Screens: Annotated Screens



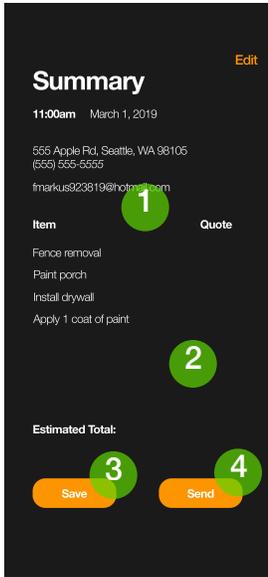
- 1 The "+" symbol in the top right corner allows the user to input upcoming appointments.
- 2 Each list item includes the client's name and date/time of appointment. List items can be selected to show the client's profile.
- 3 Past appointments have "Saved" in orange on the right side when the session summary has not been sent to the client
- 4 Past appointments have "Sent" in grey on the right side when the session summary has been sent to the client.



- 1 Client profiles include their name, appointment time/date, address, phone number and email address
- 2 Users are able to edit client details by selecting the "Edit" on the top right corner
- 3 Users can say "start session" to start the recording
- 4 Alternatively, users can tap on the circle button to start the recording.



- 1 As keywords are said, ESTi will generate a list of items that were discussed during the session. This list will be updated in real time and can be viewed on the screen during the session
- 2 As the recording is happening, there will be a sound wave that moves across the screen to provide visual feedback that the recording is occurring
- 3 Once the session is over the user can either say "end session" or tap the button to stop the recording



- 1 The app then generates a summary of scope of work, allowing the user to update quantity and estimates
- 2 As the user is updating information, estimated total will auto-generate
- 3 Users can save their session to be edited later
- 4 Alternatively, users select send if they have reviewed the summary are satisfied with it's contents and the summary will be sent to the user's email.

ENGINEERING DESIGN DETAILS

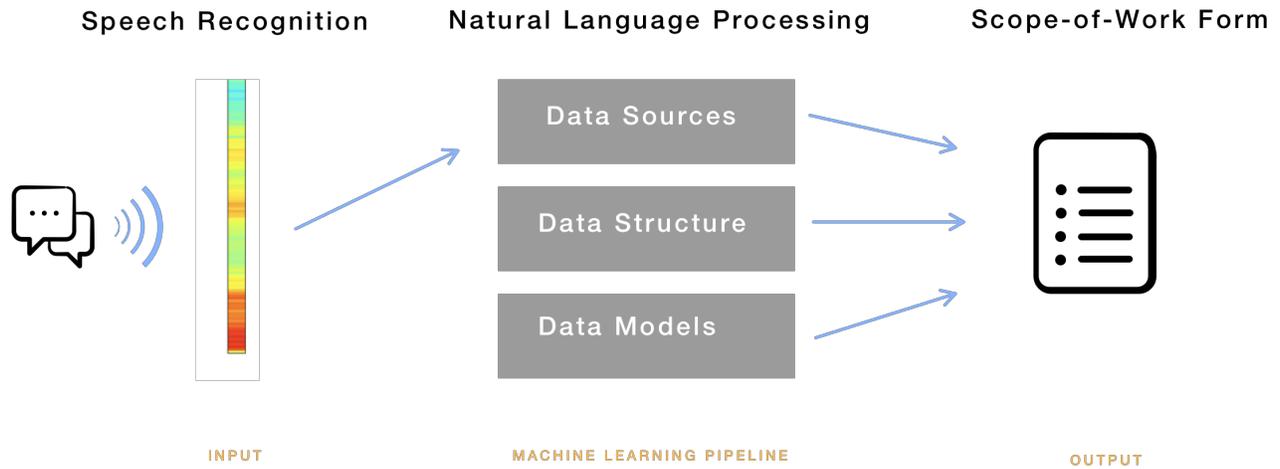


Figure 2. The Structure of ESTi's system.

System Overview

ESTi's system implementation is comprised of speech recognition and natural language processing (NLP). ESTi utilizes an API to complete the task of translating speech to text. NLP is employed to determine and analyze the intent and entities embedded in the dialogue exchange between the user and the client.

Speech Recognition

The first component of ESTi's system is speech recognition. The user's mobile phone microphone is used to convert the physical sound into an electrical signal and then into digital data with an analog-to-digital converter. ESTi speech recognition ability relies on the Hidden Markov Model (HMM), an established approach especially known for recognizing the phoneme sequences of utterances which are used as labels for speech synthesis [3].

Gathering Representative Audio

ESTi aims to extend speech recognition to target the type of conversation that most commonly takes place on the site of the client's residential property. ESTi trains speech recognition models to adapt to the user's and user's client's speaking style, expressions or unique vocabulary. ESTi sources a speech processing API with samples of scenario-specific conversational utterances. A collection of samples consisting of 450 conversations will make up the origination library and will support language models with the task of reducing word error rate [4].

Speech Processing

Several speech processing application programming interfaces (API) can address transforming human speech to text. Google Cloud Speech API, Amazon Transcribe and Microsoft Azure Bing Speech API are a few examples. ESTi requires a service that can work in both batch and real-time modes and requires a system that supports customization in the form of providing a list of possible words to be recognized. At times, ESTi will have access to an active internet connection. However, sometimes the user may not always have internet access, therefore Esti requires this flexibility from an API.

Natural Language Processing

ESTi utilizes an NLP service to find insights and relationships between the real-time stream of audio received and the residential construction domain vocabulary and terminology.

Data Sources

ESTi sources data from construction price indexes and commercial costbooks, such as those published by Craftsman Book Company, a publisher of over hundreds of technical and professional references for the construction industry. One such example is National Building Cost Manual [5],[6] which is used by adjusters and appraisers in the residential construction field to work up reliable budget estimates based on actual materials, design features, class of construction, area, shape, wall height, number of floors and support requirements.

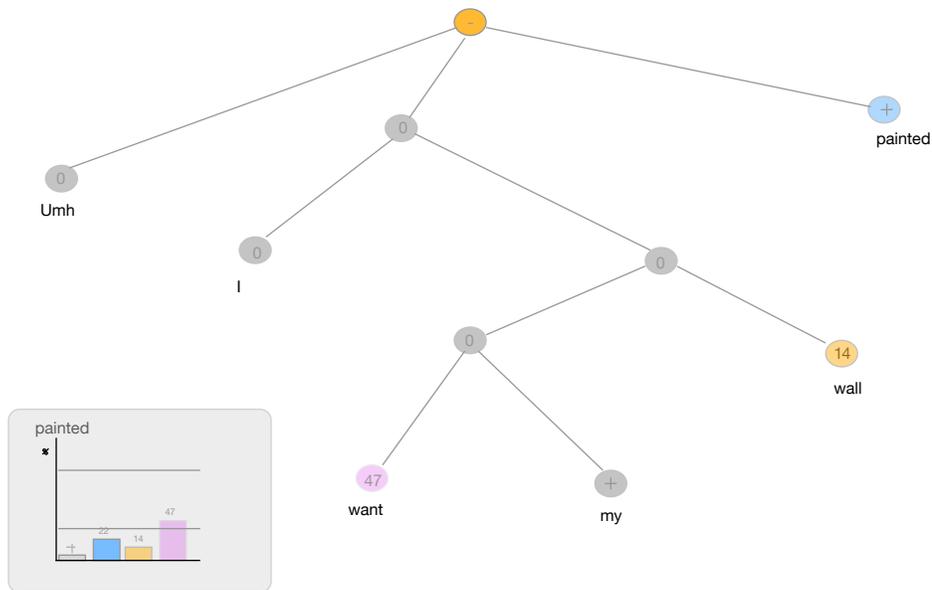


Figure 3. Semantic Analysis Tree

Data Structure

ESTi utilizes a language model to formulate a description of what the client wants the user to do. A corresponding meaning is determined by using the $M = \langle T, R, I \rangle$ structure. In this case, terms is represented by T, the relations between these terms: R, and their interpretation: I. The diagram above displays a type of dependency within a sampled phrase, "Umh I want my wall painted."

Data Models

ESTi maps a generic model (speech recognition content) to a custom model (SOW form) to populate the SOW form. ESTi employs the Natural Language Toolkit (NLTK) to comprehend these dependencies and also allows for the development of a customizable model. The NLTK is referenced for the following functionalities: 1) Classification, 2) Tokenization, 3) Stemming, 3) Tagging, 4) Parsing, 5) Semantic reasoning. For a complete description of the cookbook functionalities, please refer to NLTK [7].

Further Development

ESTi can partner with other cost estimation softwares to extend the scope of work parameters to include pricing based on labor, materials and time factors. Softwares such as Sigma Estimates provide proposals and quotations with detailed insights and reporting capabilities for project stakeholders, yet a voice-interface has yet to be integrated [8]. Evaluating the technical feasibilities and examining the potential value of interactions of this nature on a mobile device is ripe for further understanding.

References

1. Botha Purushotham, "Designing Personal Assistant Software for Task Management using Semantic Web Technologies and Knowledge Databases" May 2013. <http://web.mit.edu/smadnick/www/wp/2013-11.pdf>
2. Forecasting Industry Resources," presented by A.R. Crosby at the Institution of Chemical Engineers in London, November 4, 1981. https://www.cmu.edu/cee/projects/PMbook/05_Cost_Estimation.html
3. Aizawa, Y, Kato, M., & Kosaka, T. (2016). Many-to-many voice conversion using hidden Markov model-based speech recognition and synthesis. *Journal of the Acoustical Society of America*, 140(4), 2964-2965. <https://asa-scitation-offcampus.lib.washington.edu/doi/abs/10.1121/1.4969167>
4. Veton Këpuska, & Gamal Bohouta. (2017). Comparing Speech Recognition Systems (Microsoft API, Google API And CMU Sphinx). *International Journal of Engineering Research and Applications*, 7(3), 20-24. https://www.researchgate.net/publication/314938892_Comparing_Speech_Recognition_Systems_Microsoft_API_Google_API_And_CMU_Sphinx
5. Bureau of Labor and Statistic <https://www.bls.gov/cpi/data.htm>
6. Moselle, Ben. (2019). 2019 National Building Cost Manual. Craftsman Book Company. 2017 national
7. Bird, Steven, Edward Loper and Ewan Klein (2009), *Natural Language Processing with Python*. O'Reilly Media Inc. <https://www.nltk.org>
8. Sigma Estimates - Estimating, bid and documentation software ...<https://sigmaestimates.com/>