Real-Time Subtitles

Project Team: Ashton Gerhardt (8821135), Hannah Berthold-Brush (8635989), Nick Broadbent (8709720), Philippe Lalonde (8341617), Timi Owoturo (8606957)

April 15th, 2018

TABLE OF CONTENTS

I. ABSTRACT

In our Introduction to product development and management class, we got the opportunity to develop a real world device while following the steps taken by startups and established businesses alike that were simultaneously shown to us in class. Our device is “Real-Time Subtitles” and we worked on both the hardware and software side of it, however developing it was only a fraction of the work to have our final product. We were able to meet and work with a client through the semester to understand and prioritize the clients requests. In this report, we will go in depth in the full process taken to ensure our prototype is exactly how we want it to be.

II. List of Table and Figures

III. Introduction

Hearing is a powerful sense that we take for granted everyday. It allows us to communicate easily with one another. So, It's a very important part of our communication, which is a problem for those hard of hearing. People that are hard of hearing have difficulty communicating in their everyday lives as they usually rely on reading lips or having things written down. This makes each day a challenge. So, in order to solve this problem, our client, who is hard of hearing, asked us to develop a device which would transcribe a verbal conversation in real-time. We made this device which uses microphones to get the voice input, then presents the speech as a message in a chat style app on the screen. What makes our design different is it is portable, while most solutions to this problem require a computer or a real person to do the job. This is important because carrying a laptop everywhere isn't always practical, especially if you want to talk to someone in the street, and hiring a person to do the job is expensive. We followed the design thinking design process in order to better understand the problem and empathise with our client.

IV. Need Identification and Product Specification Process

Problem Statement

After our initial meeting with our client, we felt that we had a much better understanding of her needs and what she was looking for in a device. Based on this new understanding we were able to develop the following problem statement that we tried to follow for the rest of the design process.

Our client, Marie-Claire, is hard of hearing and has difficulty understanding what the other person has said in conversations. She needs a non-restrictive way to translate the other person's voice to text, allowing for more natural conversations.

Benchmarking

Before we began brainstorming and developing possible solutions we did some benchmarking to determine what products were already established and on the market. This way we could see what was already available to our client, and why these options do not work for her.

With this knowledge we could start to develop our own solution while keeping in mind these already available products, and potentially expanding on or incorporating some of their best features into our solution. Some of the designs we discovered were:

CART - Communication Access Realtime Translation:

Cart is a live word for word transcription of speech to text so that people who are hard of hearing can read (on a laptop) a script of what is being said in a group or at appointments. Clients usually have to book CART transcribers. These are people who write spoken words into text using either a stenotype machine, notebook computer and real-time software. This text is then displayed to the client. Clients can use CART at specific events like meetings, classrooms, courtrooms, religious services, conventions and conferences, personal appointments and civic events

Q-System:

The Q System is a live speech to text translator for the hard of hearing. The Q System is a wearable device that displays text (from speech) on a screen/display and also helps graph emotions from the person currently talking to them. When a person is speaking their voice is captured by the array microphone, transported to the band via Bluetooth, and translated into text using voice recognition software. The text appears on a screen which is embedded into the palm of the user's band.

Features of the device as listed on the page:

- Include Touch speech compression combinations to represent sounds that a user is missing based on a hearing test.
- Allows users bookmark parts of the transcript in real-time, e.g dates, times, or important moments for future reference.

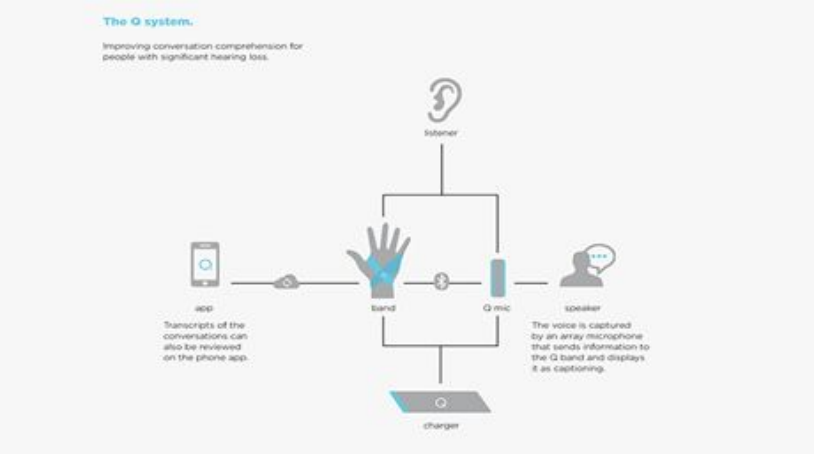

Figure 1: Benchmarking - The Q System

Metrics

In order to make our end product more convenient:

- The system must operate for 8 hours on a single charge
- The weight of the device
- The charging port must always be accessible
- It must be able to turn on and display text within 20 seconds of activation
- Every option should be accessed with no more than 3 clicks

For our product to be more aesthetically appealing:

- All colours should be beige, grey, black or white
- The product should not show any exposed circuitry
- There must not be any flashing lights
- The device must be silent
- If there are microphones, they should fit within a 5cm cube

For our product to have a better usability

- The text must be visible for the client
- The user should be able to distinguish who is talking
- The device should display the text no more than 3 seconds after it was said

Target Specifications

- ➢ Multiple Wireless Microphones (color coded) → will be beneficial to be able to transcribe each person's voice to the device. Being wireless it will help with the mobility for everyone.
- ➢ Use without Wifi → No matter where the client is, the client is able to use the product, and not have to rely on wifi.
- ➢ Transcribe French and English → With the client being bilingual, it is beneficial for the product to be able to transcribe both French and English.
- ➢ Functional → The client is able to use the product with little to no confusion.

V. Conceptual designs

Our first task of our prototype was to create something we can easily compare, improve on and show our client at a very lost cost. This is what we came up with:

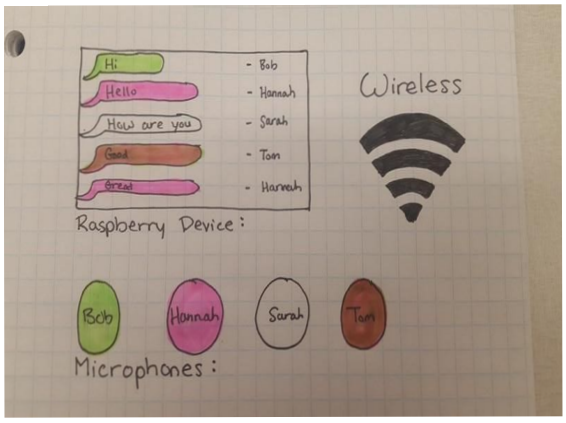

Figure 2: Conceptual Design

This was a couple phases in but some of our biggest ideas originated from this phase because of how easy it was to change stuff. For example, the coloured text, the ability to set the names of the microphone and an early model of the UI. Once we agreed on it, we were ready to translate it into Android.

VI. Project Planning and Feasibility Study

The project plan consists of the tasks performed, milestones reached, Scrum Diagrams and Materials used.

Our first step was to separate our tasks in an even matter and balance our forces. This is what we came up with:

1. Order/Purchase Raspberry Pi, LCD Screen and Microphones – Hannah + 2 hour
2. Measure the dimensions of the Raspberry Pi and LCD screen and design an appropriate

casing on paper for the devices - Hannah and Ashton, 1-2 hours

3. Transfer the design concept for the case to the 3D Printing software - Hannah and Ashton, 2-4 hours
4. 3D print the case - Hannah and Ashton + 1-3 hours
5. Make a non-functional UI – Nicholas + 2 hours
6. Integrate the microphones with the device – Timi + 5 hours
7. Transfer the audio to text with the library – Philippe + 4 hours
8. Combine the code and the designed UI– Nicholas + 4 hours
9. Measure the dimensions of the device and design a stand – Hannah and Ashton + 1-2 hours
10. Transfer the design concept for the stand to the 3D Printing software - Hannah and Ashton + 2-4 hours
11. 3D print the stand – Hannah and Ashton + 1-3 hours
12. Put all the components of the device together – Philippe + 2 hours

We also had to set ourselves some timelines to ensure that we completed the project before our deadline in a reasonable manner. This was our result that we followed quite well and was effective:

Feb.15 – Have all the materials ordered

Feb 25 – Have the case designed for the Raspberry Pi and LCD Screen

Feb 20– Have a non-functional UI agreed on by everyone

March 2– Have printed the case

March 5– Have the stand designed for the device

March 9– Have the raspberry pi be able to do speech to text

March 16– Have the software and hardware completed and agreed on

March 20– Have printed the stand

We had a scrum approach to the project that we altered a little bit by calling out what we were working on without actually following the official rules of scrum but this still let us finish the project in an effective manner.

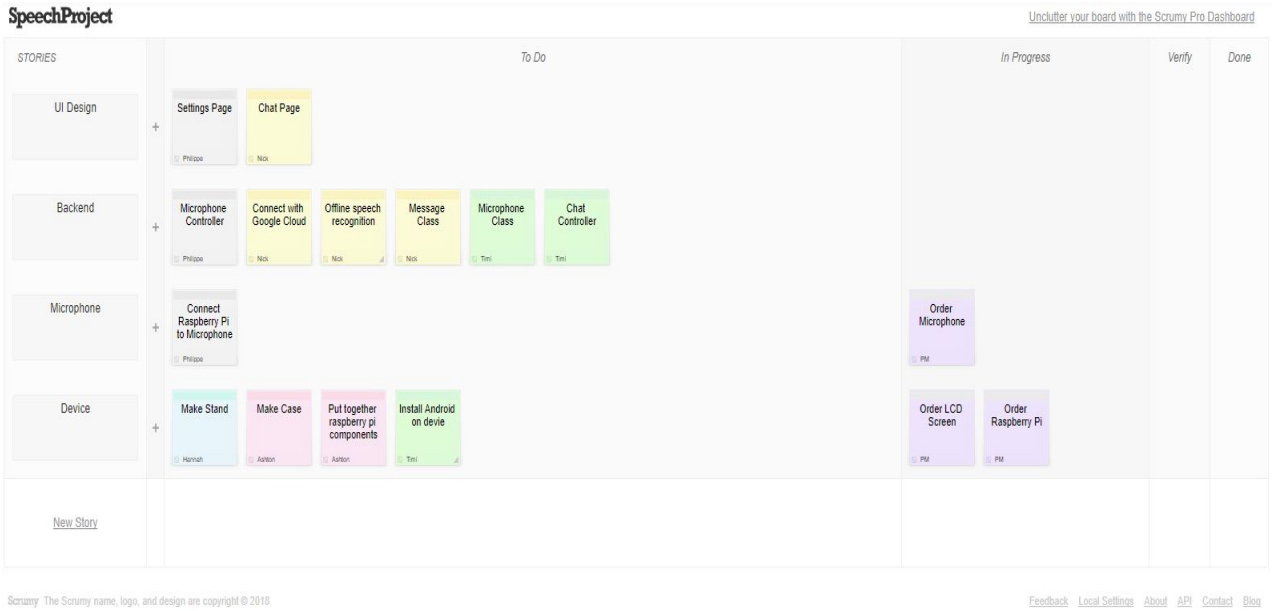

Figure 3: Scrum Diagram

Item Number	Part Name	Description	Quantity	Unit Cost	Extending Cost
1	Raspberry Pi	The main device that will help with speech to text	1	$30+	$30+
2	LCD screen	The Display. This is the device that will display text and the UI to Marie	1	$100	$104
3	Case	Its function is to hold the Raspberry pi and LCD screen together	1	Build Ourselves	$0
4	Microphone	The device that takes in speech and sends it to the Raspberry Pi	2	$20	$40
5	Stand	Its function is to make the device stand upright.	1	Build Ourselves	$0

Table 1: Materials, Cost, and Descriptions

With each item of purchase there is a justified reason. The Raspberry pi board is one of the important parts of the end product that will help with the transcribing. The LCD screen will visually show the transcriptions. A case would be to hold and keeping all of the components together. The microphones are key to transcribe what is being said. A stand is an accessory for when it is a long conversation or if the conversation is during dinner the client will be able to read the conversation without always holding the device.

VII. Prototyping, Testing and Customer Validation

It was very important for us to satisfy our client's needs in almost every step of the development. She represents our customers and can bring third party opinions that proved to be valuable. She also led us in the right direction whenever we encountered a fork in the road such as if we should have a device or make our project simply reliant on software. We met with our client numerous times to validate our product and after we weren't satisfied with the final prototype we showed her on our final meeting, we arranged to meet with her one final time with our functional prototype, with the hardware that we were missing the three previous times. This was the best way to validate our prototype and know what to include in the next model. We also wanted to test our prototype against our own standards so the metrics and it compared very well. All that's left to see is if our client likes the product and uses it on a daily basis.

VIII. Final Solution

Our final product that we are quite proud of is a tablet device running android os which runs the speech to text application. Microphones connected to the device collect the voice input, which the application will display as a chat conversation. Each microphone is coloured and the user can also assign names to them in the app, which will appear tagged on each message, to easily identify who's speaking. Another feature is the app starts up automatically when the device does.

IX. Business model

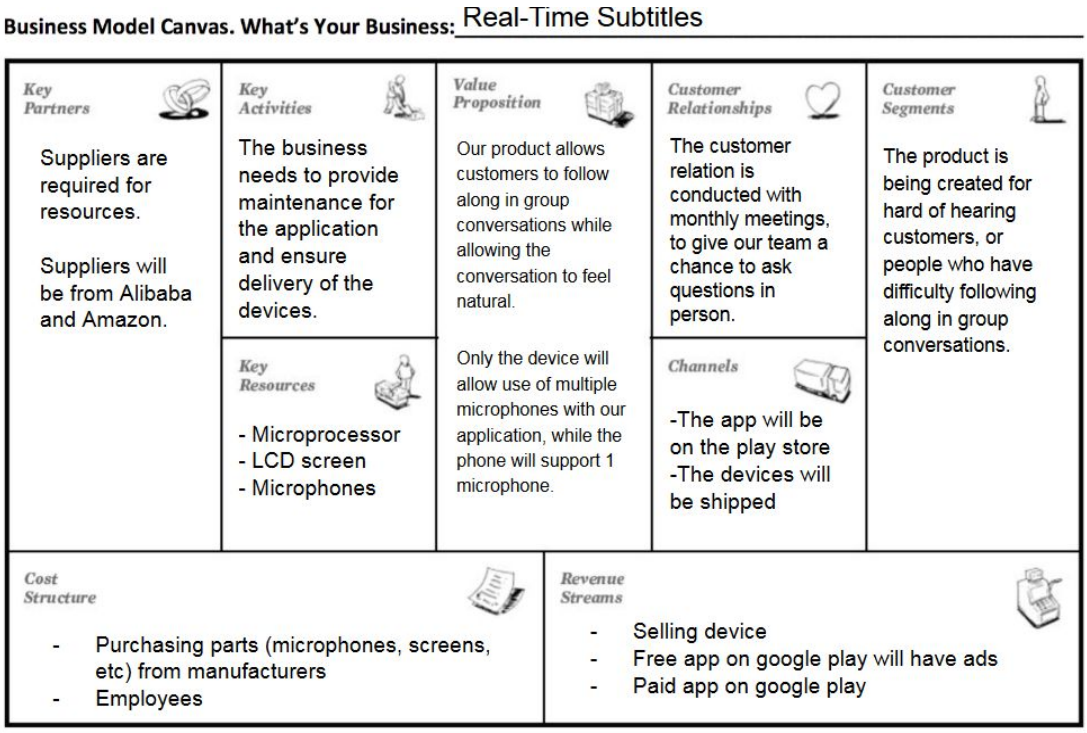

Figure 4: Business Model

X. Economic analysis

An obvious problem with our current economic situation is with our hardware. We went very general with a raspberry pi connected to a screen. This gives us a lot more abilities then we need and consequently also raises the price quite drastically (we were over double the budget). Luckily, if we were to mass produce the device, we would be able to cut the price under the 100$ mark by ordering in bulk and more specific or even custom parts with only the features we need. We believe that selling our device for 150$ is reasonable and we hypothesise that we can make them for around 100$. We also assume that our operating expenses for the computers, maintenance outside of our personal time, electricity etc. would be about 2000$ a year. We hypothesise that our interest is of 12% a year calculated annually and we sell 1.4x the unites of the first year in the second and 10x in the third. With this hypothesis we can write out this equation: (n*50-2000)+(1.4n*50-2000)(1.12)+ (10n*50-2000)(1.12)^2= operating income for the first 3 years where n is the amount of units sold in the first year. If we want an operating

income of 0 after 3 years to calculate our break even point, we need sell 9 units the first year, 13 the second and 90 in the 3rd. This seems more than reasonable and absolutely achievable if we put the effort in our marketing making this product worth pursuing.

XI. Conclusion

Finally, after going through all the steps of the design process we've reached the validation stage. From our problem statement, our client needs a non-restrictive way to translate the other person's voice to text, allowing for more natural conversations. Our device translates voice to text well and the device has long cords for the microphones, so that it doesn't restrict the other person much. Our client liked the length of the microphone cords. Based on the problem statement, we have solved the problem and the solution is valid.

There are still a few steps needed for us to be absolutely satisfied with our product. The next steps to take are to start looking into filing our patent, cutting back the price to meet our 100$ mark, look into making a more reliable case, switch to wireless microphones and have the ability to add more languages. These are all fairly small steps to have a product ready for production and at the time, we are all satisfied with our prototype.

XII. Bibliography

- **Dragon NaturallySpeaking - world's best-selling speech recognition software. (n.d.). Retrieved April 10, 2018, from https://www.nuance.com/dragon.html**

- **Speech to text transcription (CART – Communication Access Realtime Translation). (n.d.). Retrieved April 11, 2018, from http://www.chs.ca/services/speech-text-transcription-cart-communication-access-realtime-translation**

- **Android.speech.SpeechRecognizer. (2018, April 03). Retrieved April 11, 2018, from https://developer.android.com/reference/android/speech/SpeechRecognizer.html**