

University of Ottawa

GNG 2101[C]: Group 2.3

Deliverable B – Needs Identification

Submitted by

[Team 2.3, Image Descriptor]

[Shahid Awati, 300015213]

[Serigne Sourang, 300040876]

[Jathushan Karthigesar, 300060617]

[Valentin Mugabo, 300038960]

Date: January 23, 2022

University of Ottawa

1 Table of Contents

Abstract	ii
List of Acronyms	iii
1 Introduction.....	4
2 Client Statements and Observations	4
3 Customer Needs	5
4 Problem Statement	6
5 Metrics	6
5.1 Functional/Non-Functional or Constraint Metrics	6
5.2 Metrics and Units linked to Client Needs	6
6 Competitor Benchmarking.....	7
7 Target Specifications	8
7.1 Marginal and Ideal Value Reasoning	9
8 Conclusions	10
9 Bibliography	11

Abstract

This document summarizes the specifications of our application based on the initial interaction we have had with the client. The customer needs have been determined and clearly identified. Research has been performed about the available products that are similar to the application that we are trying to build, and some benchmarking has been done to that effect based on the metrics obtained from the customer needs. Finally, the target specifications were determined based on benchmarking and some reasons were provided for the choices that were made.

List of Acronyms

Acronym	Definition
WPM	Words per minute
s	seconds

1 Introduction

In Deliverable B, the objective was to empathize with the client to understand and gather all their problems that will be prioritized based on the client's needs. After interpreting the needs, a problem statement was defined prior to listing the target specifications and metrics in a tabular manner. Finally, benchmarking was done in comparison to three mobile applications being Live Text, Image to Speech and Talkie OCR.

2 Client Statements and Observations

Client Statements / Observations	Interpreted Need
I am looking for something that will allow me to read the text on images.	The application will be able to read out loud the text on an image within a reasonable response time.
Products on the market are not easy to use and they are costly.	The application will be easy to use and of minimal cost.
It would be nice to also have the possibility to take pictures.	The application will allow the user to access their camera to take pictures.
The file types should be JPEG and PDF.	The application will allow the user to select common file types from their device.
I set the reading speed at 80%.	The reading speed of the application will be convenient for the user.
It would be nice to have more than one voice.	The application will propose different voices to choose from.
English is the most important. But it would be great to also have French and Spanish.	The application will consider more languages to accommodate more users.

3 Customer Needs

Number	Need	Importance (1-5 Scale)
1	The application will be able to read out loud the text on an image with a reasonable response time.	5
2	The application will be easy to use and of minimal cost.	5
3	The application will allow the user to select common file types from their device.	5
4	The reading speed of the application will be convenient for the reader.	4
5	The application will allow the user to access their camera to take pictures.	3
6	The application will propose different voices to choose from.	2
7	The application will consider more languages to accommodate more users.	1

Legend

5: Extremely important

4: Very Important

3: Important

2: Somewhat important

1: Not important at all

4 Problem Statement

The problem is transcribing the text on an image that is already on a device for people with partial or complete visual impairment. A solution to this problem would be to create a user-friendly mobile application that can transcribe the texts on the image verbally at the user's convenient reading speed on iOS and Android platform.

5 Metrics

5.1 Functional/Non-Functional or Constraint Metrics

Metric Descriptor	Unit	Functional/Non-Functional or Constraint
Response time	Seconds (s)	Functional
Monthly Subscription Cost	CAD (\$)	Non-Functional
Number of image files supported	N/A	Functional
Speech Speed	Word per minute	Functional
Number of voices	N/A	Functional
Number of languages	N/A	Functional
App OS Compatibility	N/A	Constraint

5.2 Metrics and Units linked to Client Needs

Metric #	Need #	Metric	Importance	Unit
1	1 & 4	Response time	4	s
2	2	Monthly Subscription Cost	2	\$
3	3 & 5	Number of image files supported	5	N/A

4	4	Speech Speed	4	Words per minute (WPM)
5	6	Number of voices	2	NA
6	7	Number of languages	3	NA
7	2	App OS Compatibility	5	NA

6 Competitor Benchmarking

The three competitors that were chosen to benchmark against include Live Text, Talkie OCR and Image to Speech within the chosen metrics based on the client needs.

Application 1: Live Text

- Recognize useful information in photos or online images, be it a phone number or an email id.
- One can also use the Live Text feature in the Camera app by pointing the device camera at a photo or image with text, and then tapping the indicator icon.
- Users will get options to either copy-paste the text, look it up or translate or even make a call, in a case of a phone number.
- All iPhones with A12 Bionic chipset or later can support this.

Application 2: Talkie OCR

- Automatically recognize the input Language and speak your accent
- Built-in Translation Service
- Generating PDF on the fly, copying device clipboard and sharing of scanned documents

- State of the art OCR processing algorithm powered by PixLab

Application 3: Image to Speech

- Application that allows you to read any text aloud
- Translation services

Metric #	Need #	Unit	Metric	Importance	Live Text	Talkie OCR	Image to Speech
1	1, 4	s	Response time	4	Not specified	Not specified	Not specified
2	2	\$	Monthly Subscription Cost	2	0	4.49	0
3	3, 5	NA	Types of image files supported	5	All image file types	All image file types, PDF	All image file types
4	4	NA	Speech Speed	4	Not specified	Not specified	Not specified
5	6	#	Number of voices	2	6	1	1
6	7	#	Number of languages	3	7	45	27
7	2	NA	App OS Compatibility	5	Apple	Apple and Android	Apple

7 Target Specifications

METRIC #	Metric	UNIT	MARGINAL VALUES	IDEAL VALUES
1	Response time	s	5	3
2	Monthly Subscription Cost	\$	5	0
3	Types of image files supported	N/A	JPEG, PNG, PDF	All image file types, PDF
4	Speech Speed	%	< 300 WPM	< 250 WPM
5	Number of voices	N/A	3	5

6	Number of languages	N/A	5	20
7	App OS Compatibility	N/A	IOS and Android	IOS and Android

7.1 Marginal and Ideal Value Reasoning

Metric #1: The response time refers to the gap between the user pressing the “Read” button and the application converting the text to speech. In this case the marginal value is chosen to be 5 seconds considering a one-page pdf document. An ideal value that we want to reach is at least 3 seconds to process the text to speech output to satisfy the user’s attention span.

Metric #2: The monthly subscription cost has a marginal value of \$5 monthly because Talkie OCR is charging their users about the same cost. Likewise, the ideal value of the monthly subscription cost is to be free for our users because the competition is offering a free service, while the client’s requirement was to make a low-cost application.

Metric #3: The types of image files supported has been set to a marginal value of supporting JPEG, PNG and PDF since this will be most types of files present in a device’s photo gallery. The ideal value was chosen to support all types of image files, including PDF documents so the application has no restrictions towards the user.

Metric #4: The speech speed metric has a marginal value of less than 300 WPM since this resembles the reader as a “good reader” with a comprehension level of 80% (Refer to Table #). The ideal value of speech speed is set to be less than 250 WPM so that all readers up to the level of “good reader” can use our application.

Metric #5: The number of voices is one of the metrics that will allow the user to change their speaker between 3 different choices as a marginal value. The ideal value to meet is set to be 5 different choices of speaker accents because it will help the user in understanding the transcriptions at a higher level. In terms of competition, Live Text can output the translations in 6 different accents.

Metric #6: The number of languages supported by the application has a marginal value of 5 so that it can cover the main languages including English, French, Spanish, Hindi, Mandarin. The client writes and reads in Hindi so this language must be included in the marginal value. Otherwise, the ideal value to meet would be the top 20 most spoken languages in the world. This would allow our product to compete with Live Text and Image to Speech.

Metric #7: The app OS compatibility is marginally valued at supporting both Apple and Android operating systems since there are many users on both platforms globally. Ideally, the application should be available on both platforms to target more clients and compete with Talkie OCR who currently support both platforms.

8 Conclusions

This deliverable summarizes the needs of the client, defines the problem statement, outlines the different metrics, benchmarking and different specifications. The implications of creating such an application will be revolutionary as there are hardly any applications that offer the same features we have proposed. More importantly, the application will be very beneficial for many people who have visual impairments. Further discussion will have to be made with the clients to learn more about some of the difficulties that the team might not be aware of.

9 Bibliography

- A5. (2019, November 5). *Image to speech*. App Store. Retrieved January 24, 2022, from <https://apps.apple.com/us/app/image-to-speech/id1484770181>
- Eddine, M. C. (2020, October 2). *Talkie OCR - image to speech*. App Store. Retrieved January 24, 2022, from <https://apps.apple.com/us/app/talkie-ocr-image-to-speech/id1512795289>