

## *Pen to Print: Testing the Confounds of Handwriting to Text*

### **Introduction/Background:**

OCR's, otherwise known as optical character recognition systems, take images of text from physical documents and convert them into text that is machine readable.

OCR's begin with scanning the whole document, making sure the image is refined by smoothing the edges of letters and characters and removing any imperfections. Next, it converts the colors of the image to black and white through binarization. This helps the software distinguish the text from the background. Then, the OCR will identify characters by comparing the pixels of each scanned character to an existing database and create a character hypothesis list. Then it reduces errors through algorithms. Finally it will produce a digital text file.

In the following lab, the Pen to Print App will be tested in order to observe how OCR's function using the previous steps. Pen to Print is the first handwriting to text OCR app converting scanned handwritten notes into digital text available for edits, search and storage in any digital platform. This app will take in live photos or photos from one's camera roll. The scanner will align segment sections and scan each line of the handwritten text individually in order to provide a given result.

The purpose of this lab is to test how the quality of photos affects the accuracy of modern day OCR's accuracy and to determine what factors are the most determinant to the quality of each image.

**Materials:**

- Pen to Print App
- Stand/Selfie Stick
- Phone/Camera
- Protractor
- Loose Leaf Paper

**Procedure:**

1. All testing procedures will contain the same prompt for each test. The prompt provided is randomly generated in order to create a standardized sample for testing. Have three different participants of varying handwriting write out the following prompt neatly with pen onto a piece of loose leaf paper:

*“Was it a whisper or was it the wind? He wasn't quite sure. He thought he heard a voice but at this moment all he could hear was the wind rustling the leaves of the trees all around him. He stopped and listened more intently to see if he could hear the voice again. Nothing but the wind rustling the leaves could be heard. He was about to continue his walk when he felt a hand on his shoulder, and he quickly turned to see who it was. There was nobody there, but he heard the voice again.”*

Each participant will be tested on how the photo angle, photo brightness, and stroke of the image affects accuracy.

**2. Testing the angle of the photo**

- a. Use a stand to hold the camera in place and use a protractor to measure the angle of the camera.
- b. First, take a photo of the written out prompt with a flat angle (0 degrees)
- c. Next, take a photo of the written out prompt with a 45 degree angle from the left side, a 45 degree angle from the right, and a 45 degree angle from the bottom.
- d. Open the Pen to Print app and add a submission through the camera roll and observe the results.
- e. Record any errors including character changes or misalignment errors.
- f. Repeat steps a-g with each participant's handwriting.

### **3. Brightness of Picture Quality**

- a. Use a stand to take a photo of the handwritten prompt at a flat 0 degree angle.  
This will represent the photo at 0 brightness which is standard brightness.
- b. Open the Pen to Print app and add a submission through the camera roll using the flat image and observe the results.
- c. Record any errors including character changes or misalignment errors.
- d. Using the editing feature in the Photos app, adjust the brightness to 100 brightness.
- e. Repeat steps b and c.
- f. Then, use the editing feature in the Photos app, adjust the brightness to -100 brightness.
- g. Repeat steps b and c.

### **4. Testing the stroke**

- a. Take a photo of the original prompt written in pen.
- b. Open the Pen to Print app and add a submission through the camera roll using the flat image and observe the results.
- c. Record any errors including character changes or misalignment errors.
- d. Rewrite the prompt in pencil and take a cropped photo at a flat 0 degree angle.
- e. Open the Pen to Print app and add a submission through the camera roll using the flat image and observe the results.
- f. Record any errors including character changes or misalignment errors.

**Data:**

**Participant #1 (Neat):**

*Angle-*

Level in Degrees	Amount of Numerical Errors	Notes/Observations
0 degrees (FLAT)	0 errors	
45 degrees (left)	1 error	
45 degrees (right)	3 errors	It could not properly align the first few lines
45 degrees (from bottom)	0 errors	

*Brightness-*

Level in Brightness	Amount of Numerical Errors
0 brightness	0 errors
+100 brightness	0 errors

-100 brightness	0 errors
-----------------	----------

*Stroke (Pen vs. Pencil)-*

<b>Stroke</b>	<b>Amount of Numerical Errors</b>
Pen	0 errors
Pencil	0 errors

**Participant #2 (Kalene)**

*Angle-*

<b>Level in Degrees</b>	<b>Amount of Numerical Errors</b>	<b>Notes/Observations</b>
0 degrees (FLAT)	2 errors	-Capitalization of w in wind  -Capitalization on S of shoulder
45 degrees (left)	6 errors	-repetition of the word "again" - "W" in whisper is capitalized - "the" instead of he - "to" is removed - "Nothing" is omitted - there is misaligned
45 degrees (right)	5 errors	-W capitalized in whisper - "shoulder, and he quickly" placed before the "turned to see who it was" (lines were not aligned correctly - "to" read the line after "see if he could hear the voice heard again" - S in "shoulder" capitalized

		- “on” is placed a line after shoulder (misaligning of text)
45 degrees (from bottom)	2 errors	- 'W' capitalized in whisper - 'W' in wind capitalized

*Brightness-*

<b>Level in Brightness</b>	<b>Amount of Numerical Errors</b>	<b>Notes/Observations</b>
0 brightness	2 errors	-Capitalization of w in wind -Capitalization on S of shoulder
+100 degrees	3 errors	-Cap of W in whisper -Cap of W in wind -Cap of S in shoulder
-100 degrees	5 errors	-Cap of W in whisper -Switched lines, “the leaves all” appears before ll he could hear was the wind rustling -Period after intently -W in wind rustling -S in shoulder,

*Stroke-*

Stroke	Amount of Numerical Errors	Notes/Observation
Pen	2 errors	-Capitalization of w in wind  -Capitalization on S of shoulder
Pencil	5 errors	- “wind” in the first line was removed - “sure” spelled “sore” - “wind” spelled wrong - ”rustling” spelled wrong -was spelled incorrectly -”walk” spelled incorrectly  *Most occurrences where there was ‘w’ character there was an error

**Participant #3 (Christina)-**

*Angle-*

Level in Degrees	Amount of Numerical Errors	Notes/Observations
0 degrees (FLAT)	2 errors	--Extra period after “again”  -“Felt” spelled “selt”
45 degrees (left)	3 errors	---Extra period after “again” -“Felt” spelled “selt”  -Alignment off
45 degrees (right)	4 errors	--Extra period after “again”  -“Felt” spelled “selt”

		- Alignment error -Alignment error
45 degrees (from bottom)	3 errors	--Extra period after “again” -“Felt” spelled “self” - inconsistent lining

*Brightness-*

<b>Level in Brightness</b>	<b>Amount of Numerical Errors</b>	<b>Notes/Observations</b>
0 brightness	2 errors	-Extra period after “again” -“Felt” spelled “self”
+100 brightness	2 errors	-An additional period after “again” -”felt” is incorrectly spelt as “self”
-100 brightness	3 errors	-An additional period after “again” - “Continue” is omitted -Last line “voice again” is omitted.



### *Stroke-*

<b>Stroke</b>	<b>Amount of Numerical Errors</b>	<b>Notes/Observation</b>
Pen	2 errors	-Extra period after “again” -“Felt” spelled “selt”
Pencil	1 errors	- “felt” spelled as “selt”

### **Analysis:**

Some causes for lower accuracy rates were low lighting, angling the photo to the left and right, and using pencil rather than pen. For the angle factor, providing an image that was tilted 45 degrees to the left or right produced more errors than the photos that were taken at a 0 degree angle above the page and a 45 degree angle from the bottom. For brightness, a decrease in brightness created the most errors. For stroke, the results from the document written in pencil were slightly less accurate than the results from the document written in pen. Overall, the program did not produce a large number of errors.

### **Conclusion:**

Factors such as the image’s angle, brightness, and writing utensil used can impact the accuracy of the OCR results. OCR programs use processes to provide users with accurate results. They must refine the image and use binarization in order to allow the algorithms to accurately read the text that was provided. Therefore, although it is generally important to provide OCR software with high quality images, modern day OCR technology is capable of accurately deciphering a variety of image qualities.