**GREGG SHORTHAND TRANSLATOR USING OPTICAL CHARACTER RECOGNITION**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_

A Thesis Presented to the Faculty of the

College of Engineering and Architecture

**BOHOL ISLAND STATE UNIVERSITY**

Main Campus, Tagbilaran City

\_\_\_\_\_\_\_\_\_\_\_\_\_\_

In Partial Fulfillment of the Requirements of the Degree

**Bachelor of Science in Computer Engineering**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Jovihanni Caseñas

Ronnie Ibale

Christian Paul Vertulfo

Jazmin Joy Fullante

July 2022

**TABLE OF CONTENTS**

**TITLE PAGE** ……………………………………………………………………..….. i

**APPROVAL SHEET** ...…………………………………………………………..….. ii

**TABLE OF CONTENTS** …………………………………………………………..... iii

**LIST OF FIGURES AND TABLES** ………………………………………………… iv

**Chapter**

1. **INTRODUCTION**

1.1 Background of the Study …………………………………….. 1

1.2 Statement of the Problem ……………………………………. 3

1.3 Objectives of the Study ………………………………………. 4

1.4 Scope and Limitation …………………………………………. 5

1.5 Significance of the Study …………………………………….. 7

1.6 Theoretical Background ……………………………………… 8

1.7 Operational Definition of Terms ……………………………. 14

**2 REVIEWS OF RELATED LITERATURE**

1.1 Introduction ……………. ……………………………………. 15

1.2 Related Literature ……… …………………………………… 17

1.3 Related Studies ……….……………………………………... 19

**3 METHODOLOGY**

3.1 Introduction ……………. ……………………………………. 24

3.2 Experimental Design …… ………………………………….. 24

3.3 Research Procedure ……….……………………………….. 37

3.4 Site Variables ………. ………………………………………. 38

3.5 Environment and Participants………………………………. 38

3.6 Researcher’s Instruments…………………………………… 38

3.7 Data Gathering ….……………………………………........... 39

3.8 Statistical Treatment ……….………………………….......... 42

3.9 Proposed Gantt Chart of Design Process ……….……….. 43

**REFERENCE LIST** …………………………………………………………………. 44

**APPENDIX A**

Letter of Permission for Fare Survey……………………………...52

**APPENDIX B**

Technical Design ……………………………………………………53

Estimated Bill of Materials …….……………………………………55

**Chapter 1**

**INTRODUCTION**

* 1. **Background of the Study**

Shorthand is a system for rapid writing that uses symbols or abbreviations for letters, words, or phrases (Russon, 2017). The process of writing in shorthand is called stenography, which derives from two Greek words, stenos means “narrow”, and graphein means “to write”. It was also called brachygraphy and tachygraphy, which respectively mean short writing and swift writing. Throughout the centuries, shorthand has been written in systems based on orthography, phonetics, and arbitrary symbols while improving the speed and brevity of writing compared to the conventional method of writing. This will allow a well-trained person in the system to write approximately as fast as someone speaks.

As the world evolves, the shorthand system is also evolving. There were several shorthand systems introduced, but the majority of them were only temporary. Among the modern systems, Gregg shorthand is the most popular and efficient shorthand system. It was invented by John Robert Gregg, who originally called it Light-Lined Phonography and published it in England as a pamphlet in 1888. Gregg’s system used the curvilinear motion of longhand writing while employing phonetic rather than alphabetic spelling (Norman, 2022). Pen strokes of Gregg shorthand are formed as straight lines, ellipses, or curved lines in varying sizes; each shape is assigned to a specific letter sound. These shapes are then joined together to form whole words according to the same basic principle of writing in cursive longhand (Farrer, 2022). Since its first publication, it has been adapted into many languages ​​including Afrikaans, Chinese, Esperanto, French, German, Hebrew, Irish, Italian, Japanese, Polish, Portuguese, Russian, Spanish, Thai, and Tagalog.

The benefits of shorthand writing have enhanced its general acceptability and recognition in the world of business. Knowledge of shorthand can contribute to the development of administration and office skills by speeding up transcription, improving accuracy, and enhancing listening skills. The skill is essential in business offices, courtrooms, government offices, committee meetings, or for those in office-based roles where recording and note-taking are required. Normal human speech is too fast for the average person to write comprehensive transcriptions in longhand cursive without missing significant information. Thus, shorthand remains in use to this day. People with several years of Gregg shorthand experience are often able to transcribe at rates of over 200 words per minute (Farrer, 2022). As a result, professionals who are experts in Gregg shorthand can record spoken presentations with much greater accuracy and completeness.

Despite the advancement and modification of Gregg shorthand or other shorthand systems, students still have difficulty learning shorthand. Student experience tension, inability of students to retain what they have learnt, poor English language background, low vocabulary knowledge, lack of career guidance, students attitude to shorthand, large class etc. (Afribary, 2018). In addition to this concern, learning and comprehending shorthand will be more challenging for stenography newbies or even non-shorthand writers.

To address this problem, the researchers will propose a Gregg Shorthand Translator using Optical Character Recognition. Optical Character Recognition (OCR) is the electronic conversion of handwritten content, printed text, or image-only digital documents into a machine-readable and searchable digital data format (Callaghan, 2021). The application of OCR in this study will translate Gregg shorthand into longhand, wherein each Gregg shorthand stroke has a corresponding English word. The system is mainly used for image processing and recognition of characters.

* 1. **Statement of the Problem**

Learning shorthand is like learning a new foreign language. This means that it requires extensive learning and practice. The art of shorthand has long been a skill that can be quite tedious. Gregg shorthand usually takes several months or more than a year to master. Therefore, many individuals are still unable to read this shorthand writing. For this reason, the researchers will be creating a system that would address this problem.

Particularly, the researchers aim to answer these questions:

1. How to design a system for Gregg Shorthand Translator using Optical Character Recognition?

2. How accurate would the system be in translating Gregg Shorthand stroke into its corresponding English word?

3. How fast would the system translate a Gregg Shorthand stroke into its corresponding English word?

* 1. **Objectives of the Study**

This study aims to create a Gregg shorthand translation system by using optical character recognition.

At the culmination of the study, the researchers aimed to achieve the following:

1. To be able to design and develop a Gregg shorthand translator using optical character recognition.

2. To be able to recognize Gregg shorthand strokes.

3. To be able to translate Gregg shorthand stroke into its corresponding English word.

4. To be able to create a system that will help stenography newbies and non-shorthand writers.

**1.4 Scope and Limitation**

The research scope and limitation are enumerated as follows:

**1.4.1 Scope:**

* + 1. **Limitations:**
  1. **Significance of the Study**
  2. **Theoretical Background**

**1.6.1 Gregg Shorthand Stenography**

Gregg Shorthand, invented by John Robert Gregg on the late 19th century, is a writing system which utilizes curves, lines, hooks, and loops to deliver a handwriting speed of up to more than 200 words per minute. As technology engulfs the field of speedwriting, handwritten Gregg shorthand has slowly been replaced by stenotypes. Also known as a shorthand machine or steno writer, this device requires simultaneously pressing a combination of keys to spell out words or phrases. Although commonly used in the modern courtrooms, stenotypes are not widely accessible in the country due to its price and absence of local manufacturers.

In the Philippines, the Gregg shorthand is still widely used in court rooms and journalists. What makes it preferable than other existing shorthand writing systems is its phonetic system and intuitive nature which mimics human’s natural language processing. Phonetically superfluous or redundant letters from the alphabet are deleted such as “C” and “Q” which is mostly used with the “K” and “S” sound. Phonemes with combination of several letters such as “th” is represented by one character. Even silent letters such as “gh” in “thorough” is omitted, as well as unstressed vowels.

**1.6.2 Optical Character Recognition**

**1.6.3 K-Nearest Neighbors**

**Gregg Shorthand Stenography** (Dionis A. Padilla, Nicole Kim U. Vitug, Julius Benito S. Marquez, 2020)

“…Transcribing shorthand writing is time-consuming and sometimes confusing because of a lot of characters or words to be transcribed…”

**Optical Character Recognition** (Arindam Chaudhuri, Krupa Mandaviya, Pratixa Badelia, Soumya K. Ghosh, 2017)

“…process of classification of optical patterns contained in a digital image…”

**K-Nearest Neighbors** (Lishan Wang, 2019)

“…has been applied to text categorization in early research strategies and is one of the highly operational methods…”

**PHILIPPINE SUPREME COURT, ADMINISTRATIVE CIRCULAR NO. 24-90**

“…All stenographers are required to transcribe all stenographic notes and to attach the transcripts to the record of the case not later than twenty (20) days from the time the notes are taken …”

**THE 1987 CONSTITUTION OF THE REPUBLIC OF THE PHILIPPINES, ARTICLE XIV Section 10**

“…The State shall give priority to research and development, invention, innovation, and their utilization…”

Gather training data from Gregg shorthand dictionary

Develop and test K-Nearest Neighbor model for Optical Character Recognition

Design of Desktop App for Gregg Shorthand Translator

Implementation of Gregg Shorthand Translator

**Figure N Theoretical and Conceptual Framework**

**1.7 Operational Definition of Terms**