KINETIC TYPOGRAPHY VISUAL APPROACHES AS A LEARNING AID FOR ENGLISH INTONATION AND WORD STRESS

LIM CHEE HOOI

DISSERTATION SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF ARTS (VISUAL ARTS)

CULTURAL CENTRE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2018

UNIVERSITY OF MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: **Lim Chee Hooi**

Matric No**: RGB130017**

Name of Degree: **Master of Arts (Visual Arts)**

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

Kinetic Typography Visual Approaches As A Learning Aid For English Intonation And Word Stress

Field of Study: Kinetic Typography

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date:

Subscribed and solemnly declared before,

Witness's Signature Date:

Name: Dr. Emelia Ong Ian Li

Designation: Supervisor

KINETIC TYPOGRAPHY VISUAL APPROACHES AS A LEARNING AID FOR ENGLISH INTONATION AND WORD STRESS

ABSTRACT

Kinetic typography is a unique form of typographic art and technique of arranging text in animation to make the auditory aspects of a language visible. It expresses meaning of content in a unique and dynamic way. Past studies in this area often centers on the various applications of kinetic typography, but there are limited studies on the application of kinetic typography as a visual aid to represent the auditory aspects of English prosody, particularly in intonation and stress.

This research aims to explore and ascertain the effectiveness of kinetic typography as a visual learning aid for English intonation and stress. This study hypothesize that kinetic typography can help ESL students to capture the auditory variations in English intonation and stress more effectively. To validate the hypothesis, a pre-experimentation design has been assigned to acoustically analyze the correlation between the visual in kinetic typography and auditory in English speech. Typography characterization scheme designed by Wong (1995) has been adapted to match the two and used to develop a preliminary structure of kinetic typographic visual aid in an experiment. Twenty-five ESL college students have taken part in the exercise. The t-Test analysis result showed significant differences in the treatment of kinetic typography in learning English intonation and stress. However, due to the limitation of the research, many of the statistic cited and suggestions presented in this thesis are not complete solutions to the question considered. Rather, they are to be recognized as carefully formulated hypotheses, which need to be further investigated by other researchers.

Keywords: typography, kinetic typography, intonation, stress, rhythm, prosody.

PENDEKATAN KINETIK TYPOGRAFI VISUAL SEBAGAI ALAT BANTUAN MENGAJAR INTONASI DAN STRES PERKATAAN DALAM BAHASA INGGERIS

ABSTRAK

Tipografi kinetik ialah satu bentuk seni tipografi dan teknik menyusun teks dalam animasi unik yang membuat aspek-aspek pendengaran bahasa dapat dilihat. Ia menyatakan makna kandungan dalam satu cara yang unik dan dinamik. Terdapat pelbagai kajian mengenai pengaplikasian typografi kinetik yang telah dijalankan, tetapi terdapat kurang kajian yang mengaplikasikan typografi kinetik sebagai bantuan visual untuk mewakili aspek-aspek pendengaran prosodi bahasa Inggeris, terutamanya dari segi intonasi dan tekanan.

Penyelidikan ini bertujuan untuk menyelidik dan menentukan keberkesanan tipografi kinetik sebagai alat bantuan pembelajaran visual untuk intonasi dan tekanan dalam bahasa Inggeris. Hipotesis kajian ini ialah tipografi kinetik boleh membantu pelajar bahasa Inggeris sebagai bahasa kedua (ESL) untuk memahami variasi-variasi pendengaran dalam intonasi dan tekanan bahasa Inggeris dengan berkesan. Untuk mengesahkan hipotesis ini, satu pra-eksperimentasi telah dijalankan untuk menganalisa secara akustik, korelasi di antara visual di dalam tipografi kinetik dengan pendengaran dalam bahasa Inggeris. Skim pencirian tipografi yang direka oleh Wong (1995) telah disesuaikan untuk memadankan kedua-duanya dan ia juga digunakan untuk membangunkan satu struktur awal alat bantuan tipografi kinetik visual dalam satu eksperimen. Tiga puluh orang pelajar kolej ESL telah mengambil bahagian dalam eksperimen tersebut. Keputusan analisis t-Test menunjukkan perbezaan jelas dalam perlakuan tipografi kinetik dalam pembelajaran intonasi dan tekanan bahasa Inggeris. Akan tetapi, disebabkan oleh had penyelidikan, statistik dan cadangan yang disampaikan dalam tesis ini tidak dapat dianggap sebagai suatu penyelesaian yang menyeluruh. Sebaliknya, ia patut dikenali

sebagai rumusan hipotesis yang perlu dikaji dengan lebih lanjut oleh penyelidik-penyelidik yang lain.

Kata Kunci: tipografi, tipografi kinetik, intonasi, tekanan, ritma, prosodi

ACKNOWLEDGEMENTS

I would like to express my appreciation to Dr. Emelia Ong my research supervisor, for her careful and detailed guidance, enthusiastic encouragement and useful critique of the current study. I also would like to thank University Malaya and faculty staff for their support during my candidature.

Furthermore, I would like to thank Saito University College for supporting and facilitating my research study and the student participants who gave me their time by participating in several experiments.

For my family, I dedicate this work to them. Without their support, I would not have been able to complete this thesis.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLE

LIST OF SYMBOLS AND ABBREVIATIONS

ESL	:	English As Second Language
3D	:	Three Dimensional
2D	:	Two Dimensional
dB	:	Decibel
Hz	:	Hertz
RSVP	:	Random Serial Visual Presentation
VA	:	Visual Action
VT	:	Visual Technique

LIST OF APPENDICES

CHAPTER 1: INTRODUCTION

This chapter aims to provide the background and rationale of the study which are framed within the field of visual representation of kinetic typography, a character animation methodology of typographic art and technique to make the auditory aspects of a language visible (Uekita, Sakamoto, & Furukata, 2000). This experimental study of kinetic typography as an English prosody visual learning aid is an attempt to visualize various auditory dimensions in English prosody, in particular intonation and word stress. Intonation is the patterning of pitch changes in utterances, the melody of the voice during the speech with the variations in the pitch, length and loudness of spoken utterances, and Word stress is the emphasis, weight, and intensity, the degree of force and loudness in spoken English. This study explores the role of technology in kinetic typography as a potential source of visual output for prosody features in spoken English and as an alternative linguistic learning tool. The research strives to provide a structural analysis of kinetic typography in representing the auditory features by adapting the typography characterization scheme proposed by Wong (1995). This chapter also contains the problem statement, the purpose of the study, research questions, conceptual framework, research design and the limitations of this study. The last section will cover the significance of the study.

1.1 Research Background

Many researchers acknowledged that the computer has changed the way people work, learn, communicate, and play. Thus, technology has become common learning tools for students, teachers, and research scientists throughout the world, including many individuals who study and work at home (Gündüz, 2005, p. 197).

Goodall (2014) claims that "such trends are leading to morphing literacy practices, altering the way texts are written, read, interpreted and experienced" (p. 1). The impact

of the early development of computer technology throughout the 20th century has significantly influenced the processes and representational form of the written communication. Throughout the years, the omnipresence of information and communication technologies have measurable impacted the way people present the structure, organization and appearance of textual content (Goodall, 2014).

In the mid-1980s, the Massachusetts Institute of Technology (MIT) begun to take an interest in the research field of kinetic typography. One of the significant research in that era was a study by Cooper (YEAR) on the Prototypes of Multi-Dimensional Information Displays Incorporating Type that was presented in the Visible Language Workshop. While, in 1994, the Information Landscape project, which displays typography as a virtual landscape, was presented and it created a strong awareness and interest in the graphic design community (Staples, 2000).

Thus, the influence of digital technologies has greatly motivated change which progressively transformed the conventional written form to a more interactive visual content in communication; there is a need to explore these new visual elements, which can be adapted into the conventional media (Cheung, 2011).

The current generation is constantly exposed to all sorts of visual stimuli. The Project for Excellence in Journalism stated, "there is a growing list of media platforms and modes of information gathering. Sensory technologies such as podcasts, YouTube, and mobile phones did not exist many years ago" (Buehner, 2009, p. 7). Recent studies claim that the omnipresence of Web 2.0 and Web 3.0 have led to the popularity and increase of interest in accessing various Web visuals and applications, including the accessibility of free and downloadable typefaces which could be used in many platforms of communication. Moreover, the young people of today's digital revolution are bombarded

by highly-sensory media such as Youtube, which in turn serve as the main medium of communication of the current generation (Keane & Fam, 2006).

Hence, visual learning aids have become a well-established, popular form of learning method, and it is effective way to learn spoken English by observing the visual animation and information during the learning process. These visual elements provide incentives for the learners because it attracts and maintain their attention in the learning process well (Malamed, 2015). Children books like Geronimo Stilton, a children book series published by Scholastic Inc. (Stilton, 2011), contain words, such as shadows, scared, wedding, slam which are actually graphically modified typography form to compensate for the missing parts of the tone of voice in the written text. With reference to figure 1.1, the word "SHADOW" has an actual visual representation of shadow, "SCARED" was represented by a brush form to show aggressiveness, which is further enhanced by the colour red. While, the word "Wedding" looks inviting and softer in tone with its organic forms, and the word "SLAM" has a condense and tall structure to emulate the loudness of the sound. So, from the visual point of view, the combination of graphical words and images echoes the comments by Cook-Gumperz & Gumperz, who asserted the importance of intonation in children's reading process by employing graphical concepts to emphasis it (Cook-Gumperz, Corsaro, & Streeck, 1986).

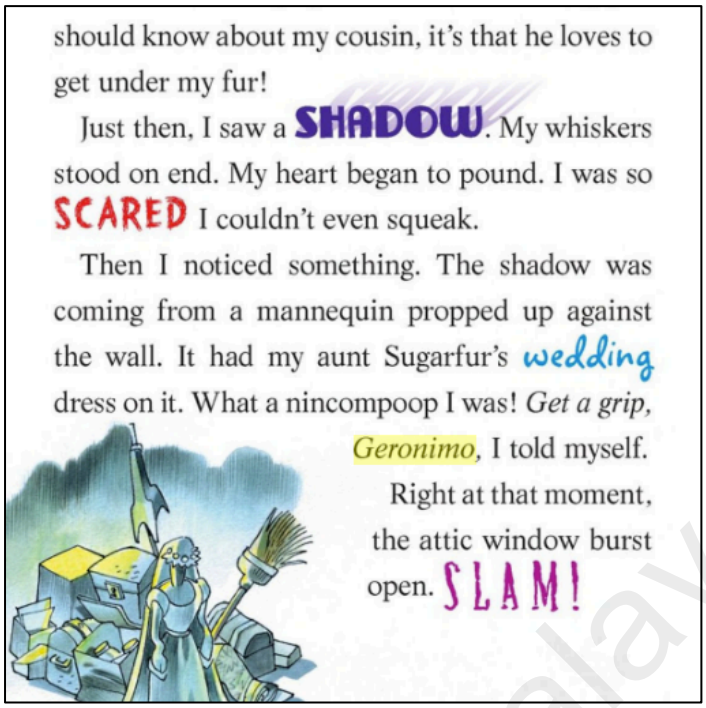


Figure 1.1: A page from Geronimo Stilton and the Kingdom of Fantasy 1: The Kingdom of Fantasy (Stilton, 2011)

Visual approaches are important in teaching English prosody, especially in regards to intonation. However, there is no transparent or generally accepted way of representing intonation in the written system besides the use of punctuations, which is not the best method in representing the modulation of voice in intonation (Dziubalska-Kołaczyk & Przedlacka, 2008).

Applications such as Nativox and BBC's Learn English Audio and Video have applied audio-visual approaches in their learning system (Learn English Audio and Video, 2017). In addition, the widespread use of information and communication technologies (ICTs) in computers, smartphones and a range of electronic-based platforms in literacy practices have changed the way texts are written, read, interpreted and experienced (Goodall, 2014).

ESL learners have long benefitted from the use of ICT technology in their learning process especially in the development of their language skills, namely listening, reading, writing, and speaking skills. Listening and speaking skills are the two very crucial skills

applied in the spoken language. The spoken language has different features from written language because it has prosodic features like intonation, pause and stress to indicate nuances of meaning (Jahandarie, 1999). Zhuang (2015) found that the use of a computer program by ESL learners effectively improves their ability to use English intonation in communication. There are also past studies which claims that the intonation and the supra-segmental features in spoken English have significant effects on the proficiency and comprehensibility of the English language (Derwing, Thomson, & Munro, 2006; Kang, Rubin, Pickering, & Rang, 2014). However, Zhuang (2015) asserted that "studies on the effectiveness of teaching suprasegmentals have not reached consensus on how intonation teaching can be effectively taught" (p. ii).

A considerable amount of literature have the stressed the importance of intonation in effective communication (Halliday, 1967; Pierrehumbert, 1981). However, the studies and methodologies of how intonation can be effectively taught to ESL learner are insufficient. Thus, there is a need for language teachers and ESL learners to have an effective intonation learning aid for advance ESL English intonation teaching and learning (Zhuang, 2015).

Pronunciation has always been an integral part of communication and it should have been incorporated into classroom teaching and learning activities. According to Gilakjani & Sabouri (2016), "ESL teachers should pay attention to the needs and abilities of learners and include pronunciation into their oral skills and other English classes and focus on both segmental and supra-segmental features" (p. 972). Improvement of English pronunciation is a long-term process and needs a significant amount of time, as bad habits in spoken English tend to be difficult to overcome. Thus, the capabilities of using different techniques and strategies for teaching pronunciation is important. When the teacher has obtained the skills in teaching techniques and strategies, they will feel much

more confident and comfortable in teaching ESL students to improve their pronunciation, which includes the supra-segmental elements of intonation in a word stress. A recent study suggests that listening and speaking to native and non-native speakers can increase the confidence of ESL students (Gilakjani & Sabouri, 2016, p. 972). Despite the identification of this need, there is limited progress in the study of kinetic typography as a visual aid and its effectiveness in representing the auditory aspect of English prosody, especially concerning intonation and stress.

Perhaps the most disadvantageous aspect of traditional reading of text on paper is that it is, in fact, a static visual language which embeds various shapes of letterforms in a passive two-dimensional medium. However, with the advancement of digital technologies, it has enabled designers to transform the representation of text from static typography to an animated and more engaging form in kinetic typography. Letterforms no longer in the layout of the two-dimensional surface but they can now virtually move fluidly within a three-dimensional virtual space and time. To date, many traditional language-teaching methodologies are adhering to the principles of two-dimensional visual composition of text and information, and the dynamic time-based visual communications with kinetic typography has not been explored (Heller, 2004).

New communication technologies have made the sharing of knowledge to be possible in myriad methods. Thus, in recent years, kinetic typography has become omnipresent because of its appearance in various commercial and non-commercial media channels such as advertisements and Youtube videos. Due to kinetic typography's dynamics and flexibility in visual and text communication, it has been featured in various platforms and environments such as on mobile devices, at home, in school and public spaces (Bachfischer & Robertson, 2005). Recent studies also found that kinetic typographic representation could better reflect the tone, intonation, rhythm and speed of reading which

makes it an ideal method for experiments that explore the relationship between reading and speaking (Brownie, 2015).

There are past studies on the various applications of kinetic typography, but most of the studies only focused on the visual representation of emotive expression and communication (James Ohene-Djan, 2007; J. Lee, Jun, Forlizzi, & Hudson, 2006; Rashid, 2008). For example,, Ohene-Djan (2007) has developed emotional subtitles with kinetic typography, while Rashid (2008)'s study was on kinetic typographic emotional closed-captioning for deaf viewers. Whereas, Lee et al. (2002) conducted a preliminary study on using kinetic typography to convey emotions in text-based interpersonal communication through emoticons. Other common applications of kinetic typography are its abundant use in screen media, visual narration in advertising, and as a form of artistic expression in a poem or song's lyrics. Additionally, there are several attempts have been made to explore the use of kinetic typography to enhance the communication of text message by using kinetic typography as a visual communication tool to deliver messages more effectively (Bodine & Pignol, 2003; Lee & et al., 2002; Uekita et al., 2000).

Typographical characters as a medium to communicate information have existed for years. Today, with the advancement of the computer and communication technology, kinetic typography as a visual representation in new media has gained rapid developments in many visual media productions. This advancement and development of technology have brought in a new education platform in digital devices such a smartphones and tablets (Veletsianos, 2016, p. ix). In these platforms, visual communication plays an important part because it can convey much more information in a shorter period of time than other forms of communication (Jukes, McCain, & Crockett, 2010). Thus, kinetic typography as a new visual communication form that consists of a hybrid of text, visuals, and animation could play an important role as a digital visual learning aid, particularly

for time-based spoken English. According to Hillner (2009), "Conventional text presents words simultaneously and the pace of verbal expression, as we know it from spoken language, is lost. But through the careful definition of perceptual challenges, kinetic typography seeks to reintroduce the time factor to the communication of words" (p. 167).

1.2 Problem Statement

Kinetic typography gives the opportunity to add dynamic expressiveness in another visual form. There is limited research in the area of design solutions using computer-based visual aids that takes advantage of the dynamic visual expression of kinetic typography.

Kinetic typography has changed the way we write and read. Instead of words staying static and to be passively read on paper, the dynamic feature of kinetic typography's visual and motion properties have enabled it to represent the rich expression in the utterance of the speaker (Hostetler, 2006).

A knowledge of the structure of sound systems and the acoustic feature and the method of speech sound is indispensable in the teaching and learning of foreign languages (Brosnahan, Brosnahan, & Malmberg, 1976, p. 9). However, the conventional fixed-size written text remains typographically constrained in terms of its expressiveness due to space and movement (Wong, 1995). Even though there are a some English prosody books which are accompanied by an audio CD as an auditory learning aid, the transcript and audio are not visually linked, so it does not establish an effective visual recognition of auditory properties in the taught content.

The standard IPO model which was used as the English intonation transcript was developed by the Dutch, but it has also been used for the description of intonation in other languages, such as English, German and Russian (Hart, Collier, & Cohen, 2006). As

Arvanti (2011) stated, "in this system, the main elements rise and fall, a choice justified on perceptual grounds. The IPO researchers noticed that pitch changes were realized more slowly than possible by laryngeal control" (p. 1) (as determined by the studies of c nm n, and Sundberg, 1979), and Hart et al. (2006) concluded that "the purpose of this slow execution must be to give listeners the perception of pitch movement rather than of a jump in pitch" (p. 71). In the analysis of Arvanti (2011):

> "Rises and falls are composed of four "perceptual features": pitch direction, timing relative to syllable boundaries, a rate of pitch change and excursion size. Also, rises and falls combined into larger configurations or contours; e.g. a rise-plateau-rise creates a "hat-pattern' while a rise-fall creates a "pointed-hat." The complication of the pitch movement was difficult to represent by lengthy explanation and dependent on the IPO transcript".

Intonation is not easily learnt by reading the standard IPO transcript. Ladd (1978) has stated that "the tone units (or tunes or tone groups) can span entire utterances but are also decomposed into smaller parts, the pre-head, head, nucleus and tail" (p.16), which is the division proposed by Crystal, 1972, and O'Connor & Arnold, 1973. Further details and references will be presented in Chapter 2.

By visualizing the whole structure into a visual form will help the ESL learner in the process of learning intonation. One of the reason ESL learners have difficulty in understanding English intonation is due to the emphasis placed on the teaching of its structural analysis rather than on its communicative value in ESL programmes. The description of tone in intonation such as rising and falling, developed with many complex configurations such as 'fall-rise', 'falling to mid', and 'low rising' could further confuse

the ESL learner, because they are not native speakers who uses the language as primary language of communication (Atoye, 2005).

To date, there is an absence of research which tests the effectiveness of kinetic typography as a visual form to represent the auditory component in English prosody. Naturally, there is little research in the development of structural composition in matching the visual component of kinetic typography with the auditory component of intonation and stress, which could be developed into a workable visual learning aid for ESL students. For instance, to ascertain the effectiveness of conveying emotions like happiness, Lee et al. (2006) have used a pre-design composition composed by designers to generate a kinetic typography expression with pre-defined emotions and a neutral-content sentence. However, the design composition from the designer could be subjective and does not provide a systematic framework for the design.

Rude (2012) study suggested 2D and 3D prosodic fonts to be used in prosodic writing system that encodes prosodic data in the 3D geometry of letter strings and this design technique helps viewers visualize auditory features in an intuitive way (See Figure 1.2). However, Rude's design technique has limited accuracy in the manipulation of visual dimensions in typography. Due to its static form, it is also limited in representing the dynamic real-time changes of pitch and loudness in English prosody. In contrast, the time-based temporal expressiveness of kinetic typography is more advantageous because it can represent the time factor in spoken English.

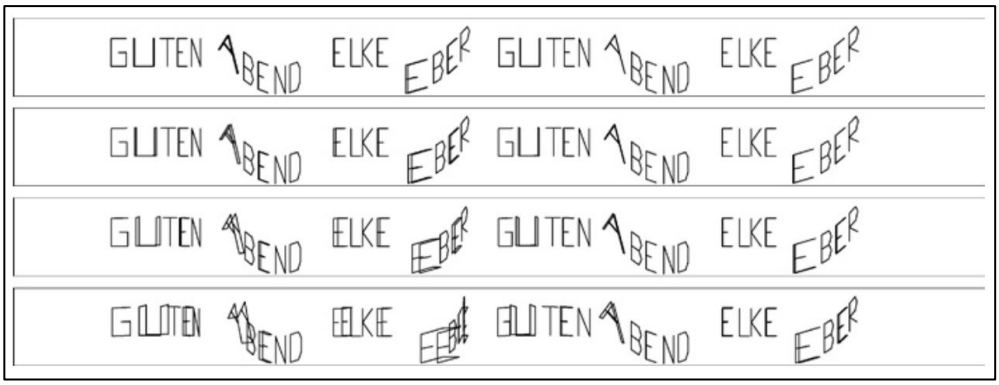


Figure 1.2: Adding Z-dimension (depth of letter) for showing tension (Rude, 2012)

Many scholars have identified the limitation of written text in representing English speech. According to Gibson & Levin (1975), the writing system has its limitation in communication. The auditory aspect in spoken English like stress, pitch, rate, pauses, voice qualities, and a host of other sound patterns could not be defined in the written words to communicate a message as well as attitudes and feelings. Thus, learning English prosody with the conventional system like IPA transcript in the textbook has been progressively replaced by the advancement of technologies. Computer-Assisted Language Learning (CALL), a term used by teachers and students to describe the use of computers as part of a language course. (Levy, Blin, Siskin, & Takeuchi, 2011) has evolved into a more engaging multimedia learning material that comes in a different format such as mp3, mp4, WMV etc. Today. ESL students could download these materials into their mobile phones for learning purposes. Online English speech training software is also widely available nowadays in learning English speech ("Nativox - Improve your English", 2017; "Learn English Audio and Video", 2017).

The hybrid attribute of kinetic typography, which has the combination of visual and motion representation in one single domain, could build upon this technological advancement as part of the evolution of the teaching pedagogy which will be beneficial to ESL learner. Typography has a central role in kinetic typography that is relevant to the present. According to Ambrose & Harris (2005), the visual dimension of typography is

one of the most influential elements that control the characteristics and emotional quality of the design. Together with motion elements, kinetic typography can effectively convey a speaker's tone of voice, qualities of character, and affective qualities of text (Ford, 1997). To date, the study and application of kinetic typography are limited to the context of new media communications, poetic narration, instant messaging, music video, title sequences and commercial advertising, but there are no known visual representation of auditory in English speech.

1.3 Research Objective

The research objective of this study aims to to analyze the effectiveness of kinetic typography in representing the prosody feature of English intonation and stress.

1.4 Research Questions

This study aims to investigate the following research questions:

1. How do we match the visual form in kinetic typography with the auditory structure in English prosody?
2. Can kinetic typography make a significant difference in the recognition of English intonation and stress?

1.5 Conceptual Framework

This thesis presents a preliminary study aimed at filling the gap of limited research in applications of kinetic typography as a visual learning aid in English prosody. Figure 1.3 illustrates the conceptual framework of this research. It is targeted at non-native-speakers of English among Malaysian. Thus, this study centers upon English as a Second Language (ESL) students in a local college in Malaysia. A kinetic typographic visual aid was developed to test the effectiveness of visual representation functionality in auditory

aspects of English prosody and this study is focused on the intonation and stress in English prosody.

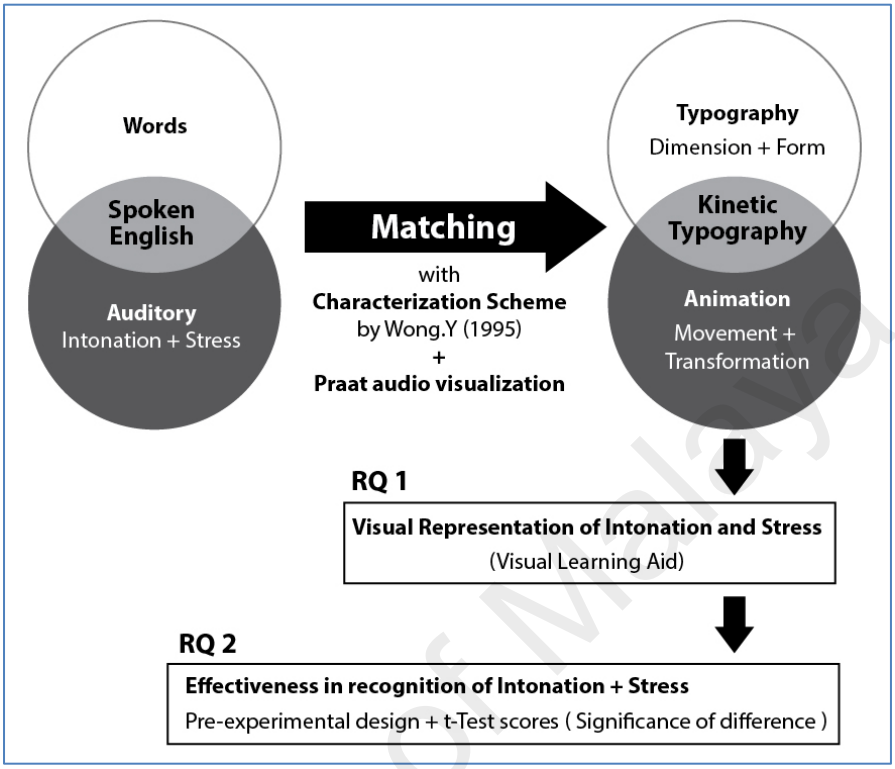


Figure 1.3: Conceptual framework of kinetic typography as a visual learning aid

1.6 Research Methodology

This research will be presented into two parts to answer the research questions in section 1.4. First is to find out the structural visual representation components of kinetic typography to match the prosody component in English intonation and word stress. Thus, the relationship between the auditory aspect of spoken English and the visual representation component in kinetic typography will be analysed using an expert's approach. The Characterization Scheme designed by Wong (1995) will be adapted for the visual composition of kinetic typography.

In the second part, a pre-experimental design will be carried out to ascertain the effectiveness and accuracy of visual perception and interpretation of kinetic typography. The comparative results in t-Test will be presented upon the completion of the

experimentation. The details of the matching structure and the method of evaluation on the effectiveness in the experiments will be presented in Chapter 3 on this thesis.

1.7 Scope and Limitation of the Study

For the pre-experimental design, the sampling population in this study are the pre-selected ESL students from Saito College due to the limitation of location and scope of the research. To comply with the time frame and scale of this research, the focus of the analysis was limited to two dependent variables, pitch and amplitude of voice, which were assessed in the pre-experimental design.

Even though this research topic extends along the domain of linguistic study, the primary focus is on the visual communication aspects of kinetic typography. The linguistic auditory analysis of this study was limited to the basic component of English intonation and stress. Other components of English pronunciation and prosody system such as vowel, syllables, rhythm and gestures were not included in this study.

1.8 Significance of the Study

Past studies on intelligibility or incomprehensibility of intonation were subjective, as they were often based on the listeners' judgment and ratings and lack data and numbers as parameters. This experimental based research aims to present the measurable impact of kinetic typography, and it is hoped that the data obtained from the experiments of this study will provide insights to future developments in the typography field. Understanding of intonation will assist ESL students to communicate with increased clarity. Hahn (2004) found that the use of the primary stress accurately will significantly affect listeners' understanding of communication. While, Zhang (2009) claims that ESL learners were able to improve their communication by improving the recognition of word stress and putting the right prominence on the syllables in words and sentences.

This research will contribute to the development of kinetic typography as a visual aid to represent prosody component in English speech. It could serve as a preliminary study to understand the effectiveness of kinetic typography in representing prosodic components in spoken English. The study's theoretical framework on kinetic typography visual hierarchy is based on the Characterization Scheme (Wong 1995) and with the study's preliminary framework of matching structure, it could serve as a reference for further research.

By providing data on the effectiveness pertinent to kinetic typography as a visual representation of English intonation and stress, designers and educators can envisage a rich picture of a new digital visual element for the future development of visual learning techniques.

CHAPTER 2: HISTORICAL BACKGROUND AND LITERATURE REVIEW

This chapter will present a historical background and literature review of articles and studies in typography and kinetic typography. This is particularly in the correlation of kinetic typography visual form and the auditory in intonation and stress, which serves as a basis for the initial development of the intonation visual aid in this research's experiment.

2.1 Historical Background

To understand the characteristic of modern time-based typography like kinetic typography, it is crucial to look into the review of the modern achievements in printed typography, past studies on the ways pioneers or significant figures of modern typographer who started to explore and move away from the conventional typographic system in order to create an innovative typographic that were ahead of their time (Hillner, 2009). With regards to chronological timeline, ancient China has been employed movable type much earlier than the European has but it was proven less useful there. This is due to the structure of Chinese writing system that contains tens of thousands of distinct characters. While the European's Latin alphabet found its advantages with the structure of translates the sounds of speech into a small set of marks, making it well-suited to mechanization in printing compared to the Chinese writing system. (Lupton, 2010). According to Hilner (2006),

> "in the mid-fourteenth century, Johannes Gutenberg's method of printing with movable type has reinforced the convention of writing in straight horizontal lines, from top left to bottom right due to its printing technology. However, a group of artists, Christian Morgenstern and Stéphane Mallarmé, Guillaume Apollinaire began to experiment and rebel against those conventional rules during the turn of the nineteenth century" (p. 12).

They introduced the revolutionary idea of visualizing poetic writing by simulating the tone and voice in the poem. This bold attempt to blend text and image into visually bold typographic compositions inspired many forms of typographic art expression at those time, which involves the art movement of futurism, Dadaism and even constructivism (Hillner, 2006). The significant similarity of the artistic expression in the form of typography has come with the form of representation of tone and rhythm.

In 1984, the Apple Macintosh computer introduced and it paved the way to popular key technologies and concepts that revolutionize the conventional typographic age. With the launch of a desktop publishing software known as PageMaker in 1985, Macintosh introduced the WYSIWYG (an acronym for "what you see is what you get") to the public and its associated technologies like bitmapped fonts and dot-matrix printing (Loxley, 2006). Hilner (2009) also claims that besides the historical invention of movable type and early twentieth-century avant-garde art movements, the digital revolution in typography during the 1980s and 1990s were the influencing factor in radical change in typographic communication, Currently, designers are have more freedom to do what they want rather than being constrained by the arrangement of metal that were dictated by old printing technologies. Thus, at the beginning of 1990s, graphic designers reacted to the International Style and a wave of revolution paved the way to break away from the constraints of the grid patterns. But in favour of a creative experimentation and with the advancement of the computer technology which was initiated by Macintosh, the only limitation of the typography expression of using the type is creativity and imagination which made way for more expressive, sophisticated and creative form of typography. The visual form of typography has become an important part of the message to communicate the contents directly rather than just its conveyor according to the sign and signifier theory in semiotic (Ambrose, et al, 2006, Ambrose & Harris, 2011)

2.2 Typography Visual Representation

Typography is one of the important roles in kinetic typography that is relevant to the aspect of time (Hostetler, 2006). A thorough understanding of typography applications serves to direct the designer in choosing a suitable typeface that facilitates the enhancement of expression and communication of the message. Every typeface has its own aesthetic and different visual expressive qualities. This complicated visual representation is evident in the letterform that are based on the individual forms and visual characteristics. (Hostetler, 2006).

According to Hostetler (2006), typography itself has two roles in functionality. First is to represent a concept, and the second is to be presented as a visual form. The correlation and interplay of conceptual meaning and visual form brings in the balance of both function and expression in the content.

On the role of representing a concept, the letterform of typography can express a different meaning. The physical characteristics of letterform, such as light or bold, round or square, short or long, wide or narrow, slim or thick can make their impression with its form and shape. The changes of this physical appearance will generate the different expression of beautiful, delightful, fresh, ugly, angry, formal, casual, loose or stiff according to the transformation of the forms(Wong, 1995). The type also serves to represent the expressions of various actions. With the manipulation of shape and form, letterforms can appear to emulate the movement of the walk, run, jump, hide, float, sink, run, climb, fly, rise, crash or hang. They can be loud or quiet, surprised, or appeared to be crying or shouting. When letters are combined as word, it separates from the type and has to mean as soon as they are combined with words or sentences. (Hostetler, 2006). On the other hand, when a typeface is perceived as a visual form, it has no longer solely read as words with meaning. The typeface has been manipulated in the form of distortion,

texture, scale, and has been extruded into another space that has become an active live stage which brings in a new dimension on the new visual platform (Carter, Meggs, Day, Maxa, & Sanders, 2014).

Currently, type is an essential visual form present in our daily life. Typography legends as well as many design leaders consider typography as a matter of visuals and not merely a system or tool to aid communication. The main reason is that it was proven that visual elements are capable of drawing a viewer to accompanying the body of text. Thus, it is worthwhile to examine which typographical treatments are have most visual impact and such examination may lead to further research into increasing reading through this visual impact (Buehner, 2009).

In the aspect of design, typography systematically visualizes language intended for texts to communicate with the specific audiences. Whether designing a book of poetry or a poster for a museum, typography's job is to allow for the reading of language (as text) and to communicate the meaning of the information thus. Challenges as to how readers engaged with typography and the act of reading have surfaced from time to time in the history of typographic practice. Interesting and curious work has emerged from these. However, the primary purpose of typography is to enhance the act of reading texts, not to hinder and confuse the reader. It is also important for typography to respond appropriately with the intention from the author. Both the content of and the context for which a text is intended must form important considerations for the designer and it can be exemplified in the use of bold typefaces and capitalizations to enhance the meaning of a word (Harkins, 2011). But in today's culture, the use of these forms can be misunderstood. When someone writes in all capital letters, it will look and/ or sound to many recipients that they were being shouted at (Lupton, 2010). This could be because lower case words appear to be more approachable due to its natural and organic form of many circular

shapes and lines that create the feel of comfortability, pleasant, playful and positive feelings. Unlike uppercase words, which are mainly comprised of rigid and harsh lines and sharp shapes, which are used usually used to portray feelings of frustration, loudness, rage, anger or pain (Ponce, 2012). However, there are no definite rules of controlling the usage of bold and capitalization in typography. In this case, the visual representation of typography form not only played important parts in our perception of tone but also the expression of meaning in the written language.

As mentioned in the previous section, significant figures in typographic production, from Gutenberg, the moveable type inventor to the development of the first desktop publishing software PageMaker by Macintosh. The form of visual representation in typography has rapidly developed and advanced in this century. However, over usage of this technology might create too much noise, and instead of enhancing on the clarity of communication, the additional component might lead to confusion. As mentioned by Crystal (2006), the advancement of the range of typographic treatment, colour variation that far exceeds the pen, the typewriter and the early word processor, which allow wider choice of options which were not available on conventional publishing, like animated text and multimedia supported sound and video - does not mean that everyone can manipulate these technologies correctly. There are many techniques and guides that need to be considered to avoid illegibility, visual confusion, over-ornamentation and many other possible problems. Crystal (2006) also claims that the problem of ignorant or over usage of typography technologies is called graphic translatability and without proper consideration and understanding the content of the message, the over usage of computer techniques and multimedia elements will cause frustration that generates the issue of legibility and visual perception in the reading process.

In the past few decades, visual representation of typography has always been related to the language, as both are the essential tools for communication. Researchers in graphic design have investigated a variety of approaches to find out if there is a language of visual imagery which mirrors verbal language. Booth-Clibborn & Baroni (1980) claimed that they have found a universal syntax emerges and indeed graphic design itself is a language. The following section, will discuss the relationship between the visual representation of typography and language.

2.2.1 Typography and Language

Long before the development of writing, human-produced visual representations of objects surrounding their living environment. The pre-historic records of it such as cave painting in Spain, France and the Sahara Dessert can be dated back from 12,000 to 40,000 years old. These paintings represent concepts rather than the words itself (Finegan, 2007). According to Marcus (1987), writing and drawing were similar activities to seek and record, describe, explain, and even arouse the readers. The elements used were concurrently lines, objects, positional markers, ideograms, pictograms and phonograms in a two or three-dimensional space or background. In their earlier forms, writing and drawing required simplification through abstraction from the complicated world in which the forms represented considerable realizations of visual concept (Marcus, 1987).

Writing can make language visible and permanent and it is one of the humanity's greatest inventions. Visible language, the visual medium of language expression, typography, markings, and writings are part of a visible language that expresses concepts and emotions. The development of visible language records a movement from iconic to abstract forms, from a full two-dimensional or possible three-dimensional array of signs to a linear organization, from a variety of elements to a more limited set (Woods, Emberling, Teeter, & of Chicago. Oriental Institute, 2010).

Visible language transformation was influenced by the materials available to a given culture as well as communicative tasks. Historically, many systems and forms have come into being as visible languages changed and evolved. For example, The Egyptian way of picture writing known as hieroglyphics, which means 'carving' that originated 5,000 years ago, used both abstract and iconic representational symbols organized into vertical and horizontal groups of the mark (Sarkar, 2008). Phaestos disk which is a 3,700-year-old Cretan ritual document has simple phonetic characters on its fired clay and were made up of 45 different stamps which suggests that the script was syllabic (Daniels & Bright, 1996). Thus, this proves that the historical developments of visual language have certain expressive qualities of language that have been decreased due to simplicity. Furthermore, they lack of the expressiveness of suprasegmental feature in language due to its limitation of space and technology of the time. The syntax of these signs is limited and individual characters are the convention (Marcus, 1987).

Letters used in the writing system by themselves do not communicate by itself or have any meaning other than to represent a sound or phoneme. It was not until the letters were combined and placed together that led to the creation of the written word that is communicative and meaningful. However, one of the main limitations of the Romanized alphabet compared to the Chinese letterform is that in most cases, a word does not look like the idea or expression it represents (Woolman & Bellantoni, 2000). However, the enormous choice of typefaces available has enabled the designer to improve the weakness of the visual form and style. (Marcus, Hanseman, Gustafson, & Marcus, 1987) .

In the past 2,000 years, typographers, painters, sculptors, writers and poetries have sought different methods to enrich the symbolism in typography, with the motivation of searching for an effective visual form to communicate the complexity in the language contents (Williams, 1967). At the beginning of the twentieth century, one can name the

work of Marinetti and the Futurists, which brings in the form of visual presentation of poetic language in exploring new forms of expressions (Bartram, 2005), and also the artists and the designers of the Bauhaus, as well as the typographic collages of the Cubists. The experimental typographic forms presented was a new visual logic of signs emerged in which the lack of symmetrical composition, the tilted orientation of typography, the abstract figure and space relationship, with the use of photographic imagery on the typographic domain, the outcome of these compositions were extraordinary innovative in their expressions. (Marcus et al., 1987).

According to Drucker (1994) and Kunz (1998), typographers and designers stressed that type is a matter of visuals rather than just as a content and communication alone. In other words, the subjective appearance of a font must be considered along with the objective content of the words (Drucker, 1994; Kunz, 1998). The typefaces as a subjective appearance suggest emotions and reactions in readers by matching the actual content that they are used to be set upon. However,, with the advancement of digital technology, it has allowed a larger group of people to publish their amateur work. Subsequently, it led to an exploitation of the aesthetic, creative side of typeface design and perhaps less emphasis on the typographic rules of legibility (Buehner, 2009).

The essential component of visible language element is typography. According to Hillner (2009), typography is a visual representation of text information (p. 12). In which it includes all the design choices entrenched in selecting a type on a web page, printed page or on the television screen. Type can be determined by elements such as font or typeface, size, spacing between lines of type like sentences and contrast of scales. In general, there are two common kinds of typefaces, which categorised as Serif and Sans Serif. A serif typeface has small lines projecting from the ends of the letters, while a Sans Serif typeface does not have these lines (Ernest-Moriarty & Ernst, 1984). The size of type

is measured in vertical height or points (72 points in one inch), and the size is measured from common baseline. The weight of type refers to the density of letter such as bold, light, and heavy (Ernst, 1984).

It is important to understand the characteristic of the typeface in typography as this research will include the experimentation by manipulating the various dimension of form in typeface itself. Typography had been identified by its ability to communicate meaning. Meanings of the typeface can be predicted through common cultural experience because the exposure to the use of fonts to makes meanings through their visual appearance (Cranny-Francis, 2005). Many researchers and experts on the type and document design often attribute typeface persona to its physical appearance. For example, Benson (1985) suggested that sans serif faces are typically perceived to have "a cleaner, more modern look than serif type" (Benson, 1985, p. 37). Kostelnick & Roberts (1998) similarly found sans serif type more technical than serif type, possibly because of the former's "clean, machine-like look of modernism" (Kostelnick, 1990). As the geometric and monoline form factor of sans serif typeface does suggest this attribute. Shape and weight are also seen as contributing to the characteristic of the typeface. For instance, Parker (1997) suggested that typefaces with rounded serifs are typically "friendly," whereas typefaces with squared serifs are more "official" in tone (Parker & Berry, 1998, p. 62). Optically typefaces that are lighter in width and thickness of stroke are seen as delicate, gentle and feminine, while heavier typefaces with thicker stroke contrary look tough, aggressive, and masculine (Strizver, 2010). Sassoon (1993) created a typeface specially designed for children learning how to read, suggested that smoother, more flowing shapes with longer ascenders and descenders add a "juvenile and friendly atmosphere to the letterforms" (Sassoon, 1993, p. 160). Halliday (1978) proposed the usage of typography with the metafunctions in which she mentioned that typography could be utilized ideationally to represent meanings, actions and ideas and to establish relationships and emotions

(Leeuwen, 2006). Similarly, McLean (2000) first proposed the possibility that a particular typeface or font may help to express a feeling or mood in addition to the meaning inherent in the sign vehicle itself. As we could see from the above, typography form and shape have been identified as major contributing factors for the recognition of meaning in the language system.

Besides the traditional characteristics of weight, size, position and colour, graphic designers have additionally created a various classification of typographic characteristics to consider when designing contemporary texts. Krause (2007) and Samara (2007) have offered different conceptual categories in their taxonomies of fonts, including moods, concepts, energy, elegance, technology and specific eras (Krause, 2006; Samara, 2007). Though these taxonomies did not focus exclusively on meaning potentials, they were designed to help graphic designers to select appropriate fonts for specific texts and a book's overall appearances.

According to Kara (2009), many authors, artists and graphic designers have experimented with typography that expresses the tone of voice. A few notable is inclusive of Edgar Allan Poe and French graphic designer, Robert Mass, whom pursued the idea that letter-face shaped tone and meaning in the language". Author and typographer Jan White has mentioned in her book Editing by Design: "Type is speech made visible" (White, 2003, p. 93). According to White (2003), the visible form in typography can reflect the loudness by boldness, whispering by smallness, shouting by size, emphasis by contrast(p. 98). Typography can crystallize a tone of voice in language. It can be raised or lowered; it can appear to shout – or can appear to whisper (White, 2003). Typography can be considered as the vehicle that visualizes spoken language. Through the transformation of letterforms in space, a visual creator can reproduce this voice to convey a nuanced impression of language as it is planned to be heard.

2.2.1.1 Typography Design and Language

Writing produces visual image in mind of the person, such as different shapes, sizes, forms and placement of letters on a page contribute to the meaning of the message we produced. It creates significant statements which are not available in the spoken language (Leonardo & Drucker, 2008). Swiss linguist Ferdinand de Saussure recognized that for literate individuals, the written word carries power and influence impact due to the understanding of information through the reading. As the result of written encoding and rationalization, a sense of authority and security exists within the printed word that does not always occur in spoken words. While in the creative sector of graphic design rarely directly interact with the field of linguistics, there was indeed a significant relationship between written and spoken representation exists in between the two domains, which has a great potential for research and study within a design framework. Slovakian type designer Peter Bil'ak (2004), who is well-known for his contribution of bridging the parts of linguistics into graphic design has stated that language is studied systematically by linguists, and the main interest of linguist in spoken languages and in the problems of analyzing them as they operate at a given point in time. However, linguists are hardly extended their interest into the visible representation of language. This ignorant happened due to their misunderstanding of typography as an artificial element and thus considered it as a secondary subject to spoken language (Bilak & Bailey, 2004). On the other hand, typographers are the one who put emphasis on the appearance of type in print and other reproduction technologies. They often have competent knowledge of design theories and principles of composition, colour, proportions, forms and paper whichever related to design. Yet they have often neglected or have little knowledge on the language which they represent. Thus, letterforms and the visual representation of language through typography means has motivated the present research interests into the correlation

between language and visual, auditory, representational and other medium. (DiPietro, 1999, p. 1).

Language and visual representation are the rich variations in sound combinations and visual forms of writing systems. Writing, which is the visual representation of speech, relocates oral speech into the visual modality and facilitates an investigative process between speech and graphic form. Ferdinand de Saussure's (2013) theory of semiology states in Principle: "The bond between signifier and the signified is arbitrary." The field of linguistics is built on the presupposition that the relationship between the signifier (sound or image) in speech and the signified (meaning, idea or concept), has no natural, straightforward or perceptible pattern (de Saussure & Harris, 2013).

In conclusion, besides the relationship between typography and language, this research intends to explore the relationship between speech and its representation via designed the form and shape of letterform and the movements in virtual space of kinetic typography. This thesis is not an attempt to resolve the current language and visual relationship represented by the speech and letterforms. Instead, it is a visual and graphical exploration of this significant relationship that enables ESL students to learn and communicate confidently with the recognition of both visual and auditory aspects. (DiPietro, 1999). The prototype and visual experimentation of visual learning aid developed in this investigation involved the creation of a typography placement and an animation technique. The visual experimentation is graphic and dynamic in nature, and draw from ideas of sound visualization, phonetic theory in a dynamic form.

2.3 Kinetic Typography

The following section will discuss on the continued evolvement of typography to kinetic typography.

2.3.1 The Evolvement of Typography Expressions

Historically, long before the innovative invention of desktop published in the year of 1984 by Macintosh (Loxley, 2006), "Dadaists and Futurists in the 1920s and 1930s and later Concrete Poetry during the 1950s and 1960s have started to freed type from the mechanical grid of reproduction" (Matthias Hillner, 2007, p. 62). As socio-political initiatives in Zurich in 1915, the Dada movement was emerged in which constituted an anti-functional approach to typography, which was reduced to a visual form of expression without the revelation of meaningful information (Matthias Hillner, 2009).

The Futurists and Dadaists challenged the vertical and horizontal conventions typography design on the printed page, which clarity and efficiency in reading was the important feature in typographic compositions. The visual representation of typography depicting the picture form become an anomaly and has become controversial to typographic convention. These included dynamic compositions with different typefaces, sizes and weights. Like Lissitzky, Werkman explored type as concrete visual form as well as typography communication (Meggs & Purvis, 2016). These art movements have brought in the new form of visual communication to express the tone of voice.

The good example of the influence is Massin's The Bald Soprano (1964) book which presented Eugene Ionesco's absurdist play in words and images (Figure 2.1). His visually rich rendering technique of the spoken word as a way to bring the intonation and timing of the play into the prints. The reader could read different actors speak their lines in a much more engaging way. It shows what dynamic typography could achieve long before the arrival of computers' intervention in typography manipulation. (Palacio & Vit, 2011)

Figure 2.1: The Bald Soprano Centre Spread (Masin, 1964)

Another example of the usage of typography to visually suggest sounds and movements, as in Lucio Vienna's 1917 piece, "Circo Equestre." (Boyarski, n.d.) Around the turn of the century, many writers and poets began experimenting with typographic layout and bringing figurative verse, a new information presentation with variety of other unique typographic forms into the realm of scholarly literature. (Sally et al., 2014)

In the exploration on the expressiveness of typography as a visual representation, the study of visual poetry is necessary as there was the type of visual representation broke off with the conventions of traditional horizontal layout and composition of typography, well before modern and postmodern forms of typography emerged. In the reader's eye, the visual poem was primarily a confusing piece of information which is not as legible as the conventional reading of poem, the reader cannot tell if one is looking at an image-like text or a text-like image (Hillner, 2009). However, the experimental text composition in visual poetry has significantly brought in the new way to visually suggest sounds and movements, as in Lucio Venna's 1917 piece, "Circo Equestre." (Spencer, 2004)

During turn of the century, many writers and poets began experimenting with typographic layout that led to the introduction of figurative verse that consists of new information presentation with a variety of other unique written forms into the realm of scholarly literature (Sally et al., 2014). "Visual poetry has controversially reversed and

translated the written word back into an image. The written word becomes the primary signifying system, and the image becomes the secondary signifying system in the reading process, which considered as a revolutionary step in the history and the context of visual communication" (Matthias Hillner, 2009, p. 13). The awareness of this revolution information presentation as its visual representation impact is significant for the further development of kinetic typography with the development of computational technology at the end of nineteenth century. According to Barbara Brownie, kinetic poetry is the next step in the advancement of a medium that has already transformed over its long history.

In the long tradition of print in visual representation cannot present sounds and movies with its two-dimensional platform, the time-based form is required. The sound is naturally a time-based. It does not carry any physical form, the projection of sound, movement, and sound rhythm has to be strategically projected, transferred, and represented through the juxtaposition of meaning and the strategic use of space in print media. The non-time-based form allows viewers to control the time when they look at the content but leaves the sound, rhythm, and pace of the content to our own subjective imagination. It is about feeling the forming in our mind (Cheung, 2011, p. 24).

Comparing the differences between print-base and screen-based communication, it became clear that the former deals with the simultaneous presentation of words on a page, with ink printed on paper. In contrary, on the screen-based platform, it has many ways to control the visual representation of words, the simulation of tone and characteristic is much clearer and closer to the spoken word than the printed word (Matthias Hillner, 2009). In the next section, various aspects of this domain are discussed.

In the late 1970s and early 1980s, researchers and programmers, notably at MIT and Stanford, began to develop innovative ways to describe an image letter digitally. A new medium emerged to force the issue of the screen to the forefront, which was the CD-

ROM. The multi-functional and visual rich interactive multimedia created a new venue for displaying words and graphics. It introduced new technical and aesthetic issues in the design (Staples, 2000),

Beginning in the early 1980s, digital technology has radically influenced typographic design (Bigelow & Day, 1983). The advancement of this technology enabled designers to create and manipulate letters in new ways, offering new options for crafting letterforms and "outputting" them, whether in the medium of toner particles on paper, or pixels on a screen (Staples, 2000). Moving forward, digital typography's innovations through the early 1990s were largely dependent on the digital technologies and consistent attitudes that revised the image of the printed word.

As the invention of digital multimedia technology promises immediacy and interactivity in communication, visual and sound in motion graphics have stretched the range of design possibilities in creative advertising, social media, and telecommunication beyond the visual aspect. The experience of visual has been greatly enriched by sound as a dynamic and multi-dimensional solution in the presentation of multimedia.

The ability to record and transfer visual and sound with new media has granted the designers to have a more dominion control in manipulating a viewer's experience of time and space. This control allows time-based form to become the groundwork that establishes many interactive, multi-sensory and interdisciplinary applications. However, the understanding of conventional design theory in print media is still important and adequate to effectively approach time-based media (Lister, 2009).

The ways time and the perception of time can be manipulated for an effective design in the visual communication is crucial. In time-based media, this is time activation. The element of sound in multimedia is one of the effective methods to the active time design as it does not only capture the structure of an omnipresent time-based form but also makes changes in time visibly measurable and comparable (Cheung, 2011).

Communication has changed dramatically in the last couple of decades as graphical interfaced computers have become common tools in the hands of young creatives and talented professional designers (Buehner, 2009). Form represents both the static and dynamic aspects of function. Exploring the state of equilibrium and opposition between stasis and dynamics in visual communication facilitates our understanding of interactivity and the multi-sensory dynamics in the digital era. Dynamic form underlies a life force that can be internalized as "movement" and externally represented as "change."

2.3.2 Definition of Kinetic Typography

Kinetic typography is a character animation methodology which operated by computer. Motion character implies its language together with feeling and tone of voice. (Uekita, Y. Sakamoto, J. & Furukata, M., 2000). Professor Dan Boyarski defined kinetic typography as type in sequence and movement, transition, appearance, and performance which have the potential to communicate meaning (Samara, 2004). There is a difference in between reading on static written words and reading text on the screen. To recognize a single moment in time-based presentation, an audience will have to prepare to expect the reoccurrence of a similar pattern or variation presenting in the future. In kinetic typography, the visual form will start to evolve on both visual and aural structures to communicate new meaning. Form grew from the need to communicate meaning. (Cheung, 2011).

The emergence of evolving computer technologies has allowed designers to experiment with kinetic typography. Furthermore, with the developments of the internet and video sharing channel like YouTube, it has made kinetic typography into a new form of communication tools (Rabinowitz, 2015). The visual representation of how words could be seen as its sounds is the unique function in kinetic typography (Brownie, 2015). What truly intrigues are the connection to spoken word and often the emotional response

from those in the audience. Kinetic typography visual form can resemble the rhythm, loudness, and other auditory aspects of spoken words. Kinetic typography is a form of live performance and controlling and manipulating time and duration is important in a live performance. While, visual representation of the voice in words suggests that we hear with our eyes.

According to Hillner (2009), each letter or word or graphic element in kinetic typography is an individual actor that needs direction on its visual transformation in the transformation process of entrance, performance, and exit. For example, a word makes its entrance with a slow fade, taking about two seconds to fade in. Then its performance is to remain in place, vibrating and wiggle for a few seconds. Finally, the word exits in a slow fade. The whole duration of time span is closed to four seconds. These processes of the work performed in the three stages produce behaviour and personality. (Brownie, 2015)

Kinetic typography is always associated with screen-based media and synonymous with motion typography. The type animation and presentation are retrained within the computer screen and the screen border act as a reference of the performing stage for kinetic typography. According to Hilner (2009), “we need to differentiate between motion typography (the type that moves) and transitional typography, which is a phenomenon of gradually changing typography form” (p. 35). Another term called temporal typography can be used synonymously with transitional typography to characterize “typographic forms which change dynamically over time. Changes in type size, weight and position can be used to enhance the expressiveness of typographic messages” (Wong, 1995). The hierarchy between these definitions could be seen in Figure 2.2

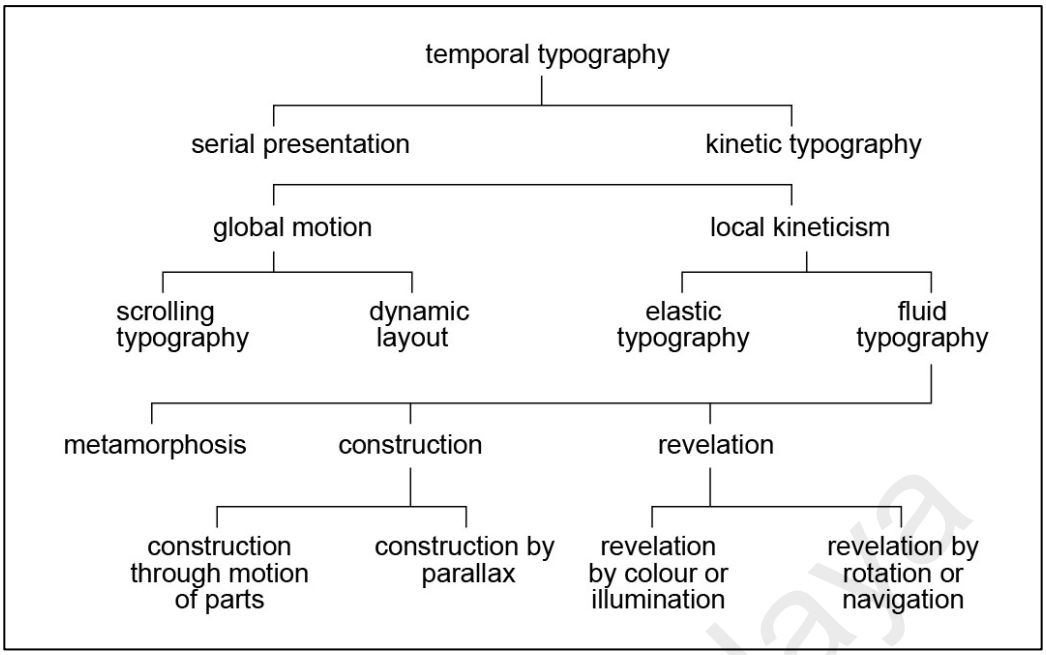


Figure 2.2: Diagram of Kinetic Classification (Stake, 2015)

2.3.3 Kinetic Typography – Visual Communication

The visual communication in kinetic typography is different from the conventional visual communication of static typography in print media. According to Waller (1987),

> "In printed text, the writers and readers are seen as communicating with substitutes in the form of the text, beyond which they may imagine a writer or readers whose identity is understood by the content, arrangement, and style of the typography. The three most important stages of writing, publishing, and reading model are account for three kinds of structure, which typically are overlaid in the same printed document" (p.140).

Waller (1987) also has categorized it to topic structure, artifact structure, and access structure. "The topic structure includes those typographic effects whose purpose is to display information about the author's argument, Artifact structure represents those features of a typographic display that results from the physical nature of the document or display and its production technology" (Waller, 1987, p. 140). The artifact in printed document included aids to interacting with the text as a topic. Typography in printed text is often used "to define the status of different 'voices' in the conversation with

components such as quotations, glosses, pedagogical devices. The complete model suggests that a typically printed text exhibits a combination of topic, artefact and access structures" (p. 141). The process of visual communication for a written language with a "context-bound and audience-related model in which typography plays a key role as a visual metaphor, can be demonstrated by the progression of topic structure to access structure through artefact" (Waller, 1987, p. 152).

However, in kinetic typography, the structure of the writer and reader relation is changed. Kinetic typography does not operate on the exact characteristic of legibility like conventional printed typography with the three main structural as spoken earlier. Instead, the emphasis has shifted to the motion and visual representation of the typography itself due to its expressiveness. According to Hillner (2009), kinetic typography challenges "the viewing process during the recognition stage that precedes to the reading. The viewer becomes a reader precisely when typographic forms become legible. This transitional stage follows the point in the duration of time where information is identified as typography" (p. 49). The flash of identification is the very instant moment when we consider text virtually typographical". Depending on the conditions and the recognition capabilities of the individual reader, this understanding of the content will occur during the process of transition in kinetic typography (Hillner, 2006 & Hilner, 2009).

Previous studies have reported that typography carries both information and communication function. Designer shapes language with type and makes words alive and communicative (Cullen, 2012). However, Hillner (2009) stated "there was some confusion between information design and communication design which were derived from the over-emphasis on the creation of information designs. Information is produced for communication but not all information communicates" (p. 54). He also claims that information in the perspective of communication theory is nothing "but 'noise' that is subject to the interpretation of the information recipient. The viewer's reconstruction of

meaningful content is a complex process" (p.55). Hillner (2009) considered typographic communication as a perception process that precedes these reactive activities in which we can differentiate between three stages: a) the cognition of data (registration of signals); b) the recognition of information (classification of information) and c) the process of decoding (for example, reading). "These three stages usually happened in a close succession and will often be felt as immediate. Situations, where we misperceived information, make us aware of the complexity of the perception process that relates to reading" (p.54). The ambiguity that is often integral to strong design solutions could provoke a moment of rest as it invites the viewers to examine the work more closely" (Matthias Hillner, 2009, p. 54).

In kinetic typography reading process, first, in the stage of cognitive of data, information must first be identified as such. This means that it must be filtered from random irrelevant noise in signals. Hilnner (2009) claims that "during this initial cognition stage, signals that appeared to be potentially significant are selected as a primary consideration, establishing a focus of attention. This stage of cognitive is particularly important in the environment of transitional information such as kinetic typography". This is due to the time-span for recognizing of information is very constraint and limited as this is the nature of transitional information. The viewers need to keep adjusting their focus of attention rapidly in order to understand the message.

The second stage of the reading process in kinetic typography is the stage of classification, following the cognition of information, a process of classified action takes place. Here the viewer will "separate typographic from graphics or photographic information. The images on the screen media are scanned differently to texts"(Matthias Hillner, 2007, p. 2). When confronted with a foreign writing system, we realized that this process of classification is much more difficult than when we are looking at a written

language with which we are familiar. For instance, the Chinese scripts scanned in a different way compared to the Roman alphabet. Furthermore, images are generally scanned in a much more random way than plaintexts (Mathias Hillner, 2006). Thus, in the preparation of reading process of kinetic typography, we need to determine the appropriate visual communication code by testing different possible scanning patterns from the viewer. Kinetic typography turns this process of classification stage into a recognition challenge in which "readers need to readjust their scanning pattern in this reading process. Viewer can never predict on what kind of information, visual of text and images will flow in during this time-based information display on the screen" (p. 40). As for the last stage in reading on kinetic typography is call "decoding" (Mathias Hillner, 2006). It is very similar to the structure of accessibility in the reading process in printed text mention by Waller (1987). Once the viewer aware of the presence of information after the cognition and classification processes, the recipient can begin to decode and interpret the information in the visual communication of kinetic typography. These three stages were the complete process of cycle of reading in kinetic typography (Matthias Hillner, 2009, p. 57).

Hillner (2009) acknowledged that typography is essentially a hybrid form of text and image with motion, which is similar to the multimedia component. From a design perspective, a typographic message in kinetic typography is not only written but rather a piece of visual composition. The organization of words, sentences and column of text lends the text an image-like eminence. Due to its architectural characteristics of metalinguistic, kinetic typography is scanned as well as read. In the comprehensive study of Hillner (2009) of kinetic typography, he concluded that one of the key advantages of transitional typography is that emotive expression can be enhanced by the definition of the graphics transition. "A conventional written word with standard normal font can become more expressive due to the way it changes over time in time-based typography.

Time-based typography also is more memorable as it does not accelerate the process of reading but instead it usually slows it down to let the audience to contemplate" (p. 55) or engage themselves in the pace of reading with graphic transition in kinetic typography".

As discussed earlier, kinetic typography is a hybrid of image and text in motion. Thus, it is important to understand the sequence of frames in order to understand motion in kinetic typography. In animation, keyframes and in- between frames are two different type of frames. Keyframes occur at the beginning and at the end of a movement and register the changes in the sequence of a movement or a story. In-between frames occur between the two keyframes and serve to support the main movement. By adjusting the number of in-between frames in the animation of kinetic typography, the object will appear to move slower, fluidity or quicker, or snappier. The manipulation on the number of frames between two keyframes can produce a variety of emotional reactions for the viewer because of the speed and the motion in kinetic typography animation (Williams, 2012).

In general, fast and snappy movement generates a more powerful and dominant impact. It generates a greater intensity of much aggressive feelings like anger, excitement, surprise, hate, emulation, obsession, loudness, dynamism, tension or fear. The speed of motion in kinetic typography plays a key role in the emotions evoke by the visual experience. Movement with faster speed evokes intensity of emotions, whereas slow movement is better at depicting emotions that are more contemplative in mood, creates a feeling of peacefulness where the viewer feels relax, serenity, calm or joy (Hostetler, 2006). In kinetic typography visual representation, the manipulation on the number of in-between frames could radically affect the pace of the narrative content, and also set the appropriate tone and pace in the spoken words.

Wong (1995), a scholar from Massachusetts Institute of Technology (MIT) studied in the field of Media Arts and Sciences has presented the characterization scheme in her thesis titled: "Temporal Typography: Characterization of time-varying typographic form" with an attempt at providing a set of concepts and terminology which allows for the description of typographic expressions.

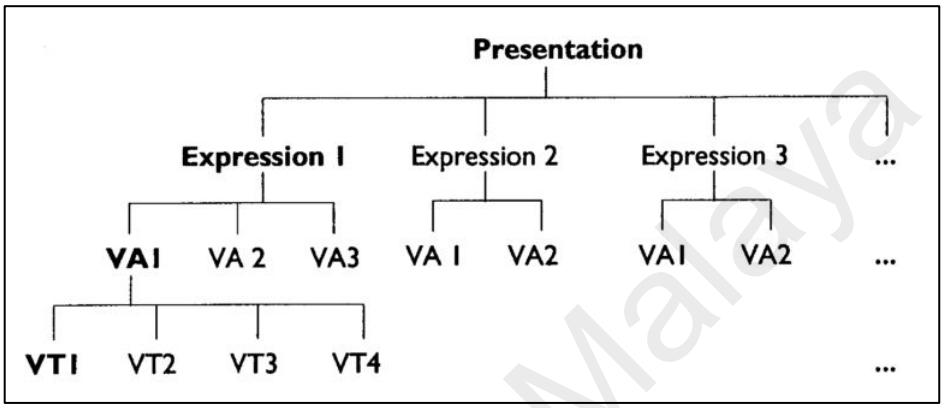


Figure 2.3: Structural hierarchy of characterization scheme

The characterization scheme identifies the descriptive terminology, which characterizes style from the point of view in kinetic typography. The scheme consists of two basic main components, structural description and behavioural description. The structural description provides a conceptual framework for building kinetic typography analytically with the physical description of the form. The behavioural description provides a set of terms for describing the temporal actions of dynamic typographic forms. The physical description of the forms are base on the notion of dimension, including the dynamic physical attributes such as colour, weight, size, and position in kinetic typography. Terminology provided in this structural scheme are at three conceptual levels, Visual Technique, Visual Action and Presentation (Wong, 1995).

The level one is on Visual Technique (VT). At the Visual Technique level, manner describes how a physical typographic dimension changes over time. For example, the dimension of size can change in a "nonstop" or "speeded" manner. Manner is either

created by a fixed sequence or generated by an algorithm. Figure 2.3 shows categories of terms which characterized the behaviour of temporal forms.

According to Wong (1995), Manner in level one is determined by what direction a dimension changes toward and at what speed. Each typographic dimension can have a range of graphical treatments. Direction describes to which end of the range the change is shifting towards. For example, size can either increase or decrease and colour value can get lighter or darker.

When a particular typographic dimension is described, the description of direction can have levels of specificity. For example, fade is a particular term describing transparency. The direction of transparency is a two-direction description because it can be described both as fading in and fading out. Terminology is more specific when change is in a particular direction. For example, appearing and disappearing is one-direction descriptions.

Speed describes how fast the dimension changes. It is defined by the relationship between range and duration. Range describe an interval of graphical treatment and duration describe an interval of time in which the change occurs. For example, we perceived the speed to be very slow if a small change in size occurs over a long period of time (Wong, 1995).

The level two is on Visual Action (VA). One way to characterize the manner of a Visual Action is from the point of view of formal identity. A form can be perceived to have certain identity. For example, a particular polygon may be identified as a "square". Similarly, a particular combination of dimensions gives a typographic form its formal identity. When a form changes dynamically over time, its formal identity may be perceived as changed or preserved. A "square" may be perceived to have a different

identity (i.e., "triangle") or the same identity at the end of a change. The manner in which the formal identity changes can be divided into two basic categories, variation and transformation. Wong (1995) stated that:

> "The term variation refers to changes within the form; it does not alter the way we perceived the formal identity. For example, when text is scaled up proportionally, we recognize the end shape as the same except it is larger. In a similar vein, a square scaled up is just a larger square. In general, dimensions, which we do not perceive as having discrete states such as colour value, saturation and transparency, are transitions. The term transformation denotes changes in which results in a perceived shift in the formal identity. I have further divided this category into two sub-categories: transition and deformation. Transition refers to a closed non-linear change from one state to another, each being perceptually distinct from the other. For example, a change from one hue to another or from one font to another. Helvetica and Times are two different typefaces which can be morphed but the two end states are perceived as having different formal identities. This can be compared to a change in a form such as from a square to a triangle. Deformation refers to an open-ended change of the form through distortion. A letterform may be stretched and distorted to a point of losing its original identity" (p. 21).

The sequence shows a piece of text being distorted over time. This movement is generated by a mathematical model simulating the physically based motion of springs (Wong, 1995).

Due to the expressiveness of moving type and its ability to simulate the pace of the spoken words, it was brought to public attention through the means of popular medias, from television and cinema, music videos and commercials to the new social media like

YouTube and Vimeo. Especially on broadcast media, the use of moving type fitted well due to the demand on the structure of short message, slogans or punch-lines which were constrained by the specific duration spot purchase by the advertiser. With the kinetic typography, the broadcast media has obtained new creative tools to penetrate audiences in new ways, constantly begging for short by great attention from the audience. Michael Worthington described expressive typography in general as broadcast type in his essay 'Entranced by Motion, Seduced by Stillness', and stated that by using moving type in broadcast media "the story is read to us in a particular voice" (Worthington, 1999, p. 39) - the reader had brought himself into a visual narrative with kinetic typography, he or she becomes a viewer instead of reader. (Heller, 2004). There were other aspects why kinetic typography could emerged as a powerful tool in the creative media industry. According to Woolman (2005), kinetic typography has intrinsic, embedded meaning and the ability to inform, entertain and emotionally affect the audiences with the ability to expressing emotions, moods, personal characteristics, and tone of voices. One of the earlier examples of adapted with this functionality of kinetic typography can be found in the title sequence of the Alfred Hitchcock film Psycho (Lee, Forlizzi & Hudson, 2002), where erratic lettering and movement of text on screen successfully communicated the unsettling nature, mood and tone of voice of the classic horror film.

As mentioned in the previous section, many researchers have stressed on the key advantages of transitional typography in its emotive expressions and effectiveness in visual communication which could be enhanced by its visual transition effect (Hillner, 2009, p. 39). This includes the manipulation of form in relation to the content, according to the tone of voice, emotion and character are conveyed in the visual and motion of kinetic typography (Shannon Ford, Jodi Forfizzi, Suguru Ishizaki, 1997). These advantages have made kinetic typography become one of the powerful tools in creative media industry which has most commonly encountered in our daily commercials, motion

graphics, music clips and film title sequences (Turgut, 2012, p. 583). At present, usage of kinetic type with sound is common in screen medias like movies, television, and computer-based advertisings(Forlizzi, Lee, & Hudson, 2003). However, few writers have been able to draw on any structured research into the field of educational learning aid using this technology.

2.3.4 The Application of Kinetic Typography

The idea that typography conveying a visual and tonal message is not a new one. The term of kinetic typography dates back to the late 1950s. An early pioneer, Saul Bass, an award-winning film-maker and graphic designer who first used the technique in Alfred Hitchcock's film North by Northwest in 1959(Horak, 2014). Saul Bass once mentioned his initial thoughts of animated movie title was an attempt to set the mood and to prime the underlying core of the film's story, which was to express the story in some metaphorical way. He saw the titles as a way of conditioning the audience so that when the film began, viewers would already have an emotional resonance with it. (Haskin & Bass, 1996).

Movie title sequences and opening credits play a significant role in the feature presentations of the introduction of the movie. They set the emotion, mood and facilitate the audience's experience and entrance into the world created by the film. The dynamic typography representation in film credits and title sequences are one of the best tools for introducing new usages of typography to a receptive viewer (Vartanian, 2003). Bellantoni and Woolman (1999) even suggested that film title design is the mother of all moving typography, which is now common in music and art videos (Woolman & Bellantoni, 2000). The title sequence was one of the first instances where animation, typography, music and graphic design all integrated together to set the mood and the story

in the opening sequence of the film. This brought in the beginning of modern kinetic typography (Ponce, 2012).

Although early works of Saul Bass were designed using conventional production methods for cinema-like multiple exposures and cell animation, the shift from analogue to computational devices expanded the way moving type, could use in, further nurturing expressive typography with the more creative approach (Bachfischer & Robertson, n.d.). For a recent film example, there was Edgar Wright's Scott Pilgrim vs. the World (2010), the film has the unique narrative kinetic typography of the cinematic adaptation (McFarlen Savannah & Colette Dom Inique Elliott James Gladman, 2014). As mentioned, kinetic typography was used to highlight particular scenes from movies. It could elevate the visual expression in the particular scene in which would make the visual content much more engaging.

With continuous advancements of new technology in animation and visual effect, the award-winning tactical creative studio, MK12 Studios, has produced many title sequences and adverting by including with their unique approach to narrative structure and experimental storytelling with juxtapositions of kinetic typography, live action and graphic design. For instance, in their Lexus Hybrid Technologies commercial and opening title sequence for the film "Stranger than Fiction" (Figure 2.4), motion graphic tracking feature has applied onto kinetic typography in the narration and storytelling. The interaction between the text and live action created an interesting visual landscape which makes the presentation of information more effective and engaging (MK12, 2012).

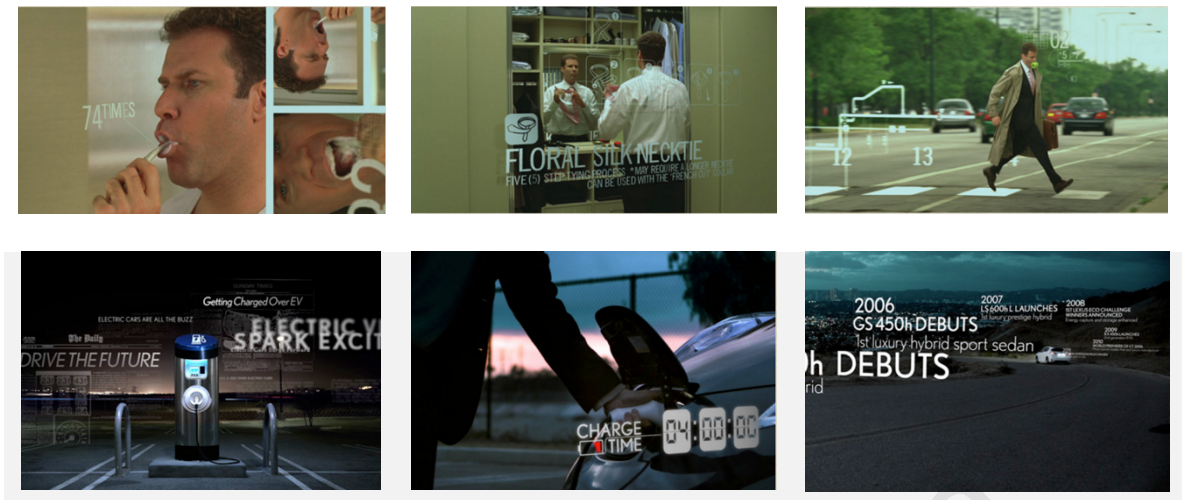


Figure 2.4: MK12 Stranger than Fiction opening title sequence (top) and Lexus Hybrid Technologies (bottom)

Today, it is common to see kinetic typography appearance and usage is not just in films but also in television shows, commercials and many explanatory videos, as it has become one of the captivating ways to gain people's attention. By manipulating the size, shape, rotation, movements and adding animation to text, kinetic type designers have proved that kinetic typography can able to capture emotion and tone of voice much better than with static print-based typography (Ponce, 2012).

2.3.5 Kinetic Typography and Auditory

Many studies have shown that Kinetic typography is connected to the tone or intonation that it carries as it is to the words and letterforms. Four main qualities of the tone of voice are pitch, amplitude, duration, and timbre. All these qualities can be measured except timbre. Pitch measured by frequency, amplitude measured by decibel, and duration measured by time. Visually these qualities can be reflected in typography visual elements which include font weight, size, letter spacing, font width and font posture (Fellows, 2009; Rosenberger, 1999; Rude, 2012). Kostelnick (1990) ascertained that the visual language of typography and other elements not only able to represent visual texture, but it also expresses the invisible tone and mood, it could express serious, conversational, low-key, energetic, highly technical, or user-friendly in rhetoric stance

(p. 199). Kostelnick and Roberts (1998) have also mentioned that type can "sound" serious, funny, formal, friendly, personable, and technical (Brumberger, 2003, p. 208).

Besides tone of voice and intonation, the pace in kinetic typography representation can reflect the rhythm and speed of reading. It is ideal for experiments that explore the relationship between reading and speaking (Brownie, 2015). The representation of pace and rhythm can change our perception of time. Similarly, letter spacing and the usage of typographic arrangement affect our reading time. When letters are set far apart from each other, the isolation makes them look more independent according to the gestalt principle of proximity. When more space is given to the form, the figure gains more attention as an individual interval. Building upon these visual principles, repeating or varying this interval consistently throughout the given space creates a sense of pace and rhythm. These changes occurred due to our time perception as a viewer (Cheung, 2011).

As stated by a renowned typographer and poetry, Robert Bringhurst, in The Elements of Typography Style, "Typography is the craft of endowing human language with a durable visual form, and thus with an independent existence" (Bringhurst, 2004, p. 11). Spoken about language and visual form, there is a similarity between spoken languages and music. Both are temporal form in their notational systems with a visual cue of supplemental meaning through visual representation (Heller, 2004).

Several attempts have been made to explore ways to use animated text to represent the auditory content for the deaf and hearing impaired people (James Ohene-Djan, 2007). As in James Ohene-Djan1 et al.'s thesis titled Emotional Subtitles: "A System and potential applications for deaf and hearing impaired people" had mentioned about the usage of colour code and change of the typefaces in the subtitle to help the deaf and hearing impaired. According to James Ohene-Djan et al., deaf people cannot determine how something has been said. Another study has been carried out with the adaptation of the

similar form of subtitles to gain access to audio content on television and film presentations. These studies have shown that by customized presentation of subtitles that depict the emotions behind the words used on screen provides viewers with the ability to personalize and adapt their interaction with subtitles, so as to assist them in their learning of emotive contents and the tone of voice (James Ohene-Djan, 2007).

In 2006, Rashid et al. (2008) have suggested that animating captions is one way of capturing more sound information contained in television and film. Studies have been conducted where animated text characterized emotions contained in music, speech, and sound effects compared with regularly closed captions for the same content. The difficult of hearing participants responded positively to the moving captions, which has provided improved access to the emotive information contained in the content. Besides this, video caption has also been explored in representing the emotional content in music to examine their impact on readability, user attitudes, preferences and their understanding of the content to the deaf and hearing the impaired audience. In these studies, viewer's reactions were examined and their understanding of songs presented with animated text lyrics. The study showed that participants were able to identify the videos as songs and understood the emotional intention of the content displayed and did not interfere with the readability (Mori & Fels, 2009; Vy et al., 2008). However, the main weakness of the study is the lack of quantitative data on the evaluation of result.

Even though kinetic typography had more success in representing linguistic features such as volumes (loudness or softness of voice), which can be represented by changes in size, position, colour, or weight of text. Tempo is easily represented with kinetic typography, by controlling the speed of the RSVP (Rapid Serial Visual Presentation) display, pauses, and fading. However, in kinetic typography, the issues in the time-based presentation of text (Ford, Forlizzi, & Ishizaki, 1997) found that it is difficult to represent

paralinguistic features of English speech by kinetic form alone. Other limitations of visual representation of sound also been explored. French philosopher, Jacques Derrida, has shared in his study that although the alphabet represents sound, it cannot function without silent marks and space.

The sound is produced by a vibrating object, transmitting waves produced by high and low-pressure areas of molecules through an elastic medium, such as air. Our ears detect variations in air pressure and translate them into signals that our brain understands (Allstetter, 1991). The perception of noise is a chaotic group of random sounds, which lack in organization. Conversely, melody could be described as frequency or pitch of sound that has been structured by their wavelength and the duration of time, present in a two-dimensional acoustic field(Heller, 2004). In view of the definition of melody in sound and intonation as the melody in the language, therefore, could be analyzed as a structured wavelength in the duration of time, which will enable the mapping of structural visual form in the in the units of amplitude and frequency in this research.

An attempt of visualizing sound also been examined in sign language. For example, auditorily perceived intonational melodies, the visual intonational arrays in sign language provide a subtle, intricately structured, and meaningful accompaniment to the words and sentences of the language. The usage of facial expression in sign language corresponds to intonation in spoken language has been suggested by some researchers (Nespor & Sandler, 1999; Padden, 1990; Reilly, McIntire, & Bellugi, 1990). There was a study that provides evidence showing that specific actions of the brows and eyes in a sign language function and pattern much like intonational melodies of spoken language (Dachkovsky & Sandler, 2009, p. 288). With the Brow Raise component function as comparable to high tones and the component, Squint instructed the addressee to retrieve information that is not readily accessible and characterizes relative clauses, topics, and other structures.

Even though in these studies of visual cue components did not involve any forms of visual representation of kinetic typography, the study has presented an identical connection between prosody component and visual representation of motion elements.

On the other hand, the dynamism of kinetic typography also has the visual representation functionality in rhythm, pitch and melody. The review of previous studies in the correlations of kinetic typography and music will help to identify the right instruments in the auditory experiments of this research.

Numerous visual attributes of kinetic typography can cross-reference with auditory attributes of music. For example, loudness can visualize as scale, rhythm can be represented by the repeated bounce or orientation movement of different visual elements, the musical contour can be represented as text on path with curvature shape and pitch can be associated with the height of shape or even luminance. The only difference is that sound originates in time-based form and a melodic song is a rendition of time-based design (Cheung, 2011, p. 11). Through the synchronization of sight and sound in music, the constraints and strengths in visual language can be re-discovered.

Music and kinetic typography shared the same feature as temporal experiences and forms of communication, expressing ideas and thought. Although these two domains occur as a separate entity, both shared a similarity in their expression of dimensions, structure, motion and time. Originally, based on the rhythms of speech, music is a form of storytelling that was structured by phrasing acoustic information. (Heller, 2004). Spoken language and music have considerable parallels and similarities as auditory-based communication systems (Thaut, 2013, p. 68). Thaut (2013) has identified two shared functions of music, speech and language. First are the aural and production features in spoken language and musical vocalization in singing. Second is the ability of communicative functions in the auditory modality.

With this overlapping feature in music and language and due to the popularity of video sharing website like YouTube, kinetic typography was also used for the popular official lyric video like Cee-Lo Green's "It is Ok" (Cee Lo Green, 2010). The way the words appeared in different sizes and positions along the animation and movement of the text created a great engaging effect (Figure 2.5). This approach helped the song gone viral.

Figure 2.5: Official lyric video by Cee-lo Green and Sabrina Carpenter

Many other artists have done the same. For example, kinetic typography videos like "Thumbs" by Sabrina Carpenter (Sabrina, 2017). The video goes extra step by using clever animation to emphasise further the lyrics, which fit right with the style of the song. In the video, the "Shape Of You" by Ed Sheeran appeared in bubble/floating movement as an accurate representation of the way it is sung (Ed Sheeran, 2017). Additionally, a stop motion style and RSVP effect on the lyric were used to represent the pop style music. It is this kind of animation that helps bring to life the simple lyrics and adds a new level of emotive content in the video that makes the viewer understand the meaning of the song.

This section has reviewed the key areas of visual communication in kinetic typography and its applications. The following section will discuss the correlation of linguistic and aspect of English intonation and stress in this research.

2.4 English Prosody

This section will present the fundamental auditory component in English prosody particularly on English intonation and stress. Understanding of English prosody is critical

in this research in speech language processing, and to utilize the technology for building speech synthesis and spoken English learning system (Silverman et al., 1992)

Linguists traditionally dealt with segmental aspects of language – the segments themselves includes phonemes, morphemes, words and sentences. Since most actual utterances contain the paralanguage, which is not strictly verbal, but it has the function which contributes to both meaning and to the structure of the spoken words. Bolinger (1986) has proposed that paralanguage has a counterpart in written language, In spoken language, the expression of body language like pointing, waving, shrugging and facial expression of winking, smiling and exciting are all non-controversial paralinguistic since they are not phonological in nature (Dwight Bolinger & Le Merton Bolinger, 1986). Auditory attribute that is phonological which could be found as stress, rhythm, tone and pitch are usually defined as prosodic or suprasegmental in linguistic. Suprasegmental denoting a feature of an utterance, with the aspect of speech that is define group of sounds rather than strings of single component like consonantal and vocalic. For example (in English) stress and intonation (Stevenson, 2010; Zsiga, 2012). The analysis features used in the experiments in this research have taken both aspects of paralinguistic and prosodics into consideration.

Prosody is fundamental to the interface between speech and natural language processing technologies. It accounts for more changeability in the speech signal and conveys most of the information to recover the intended meaning of an utterance which hardly expressed in a conventional writing system. It is also crucial in relates both the acoustic and text aspect in the utterance for a complete understanding of speech (Silverman et al., 1992).

In linguistics, the term prosody can be traced back to the Greeks, from Ancient Greek προσῳδία refers to features of speech specifically to the tone of melodic accent which

characterized full words. It has been associated with the melodic features of spoken language (Couper-Kuhlen, 1986, p. 1). According to Ladd and Cutler (21013), there are two broad categories in the study of prosody. One is Concrete and another one is Abstract. Concrete defines prosody in its physical term involve pitch, duration and intensity, While, "Abstract" defines prosody from the point in linguistic structure (Cutler & Ladd, 2013, p. 27) Those elements of language that are not independent phonetic segments but are properties of syllables and larger units of speech which include intonation, tone, stress, and rhythm (Cutler & Ladd, 2013). These prosody units of an utterance add an expressive dimension to the communication process. Speakers can supplement their utterance with elements of communication to reflect their emotional state. , the form of the utterance which could be presented as a statement, question, order or other elements of spoken words that may not be determined by grammar or by choice of vocabulary.

There are two attributes in the study of prosody aspects of speech which could distinguish between auditory measures (subjective impressions produced in the mind of the listener) and acoustic measures (physical properties of the sound wave that may be measured objectively). The prosodic analysis is rarely performed nowadays on a purely auditory basis. A typical procedure for analyzing the prosody is to record the speech on a computer device, and then align and measure with a frequency parameter in a visual representation of auditory like oscillogramme, spectral display (Hirst & Cristo, 1998, p. 84). Software like Praat used in this research is one of the instruments which can generate this type of specific, measurable frequency (Dr Will Styler, 2017).

In auditory prosody terms, the major variables are the pitch of the length of sounds, loudness and timbre. As for acoustic terms, the major variables will include the parameter of frequency measured in hertz, duration of time measured in milliseconds or seconds and intensity level measured in decibels. There is a visual representation of sound that

shows as spectral characteristics, which makes the distribution of energy in different parts of the audible frequency range visible. Further development of these variables is exploited in the linguistic functions of intonation, stress, rhythm, tempo and loudness.

Prosody features are said to be suprasegmental, a feature of speech that extends over more than a single speech sound, they are properties of units of speech larger than the individual segment (Nunan & Carter, 2001). It is crucial to distinguish between the characteristic in a person's voice (for instance their habitual pitch range) with a variable in prosody that been utilized to communicate meaning. For instance, the use of different pitch in communicating statements and questions (D Crystal & Quirk, 1964, p. 10).

Having defined what it meant by prosody in English linguistic, the discussion will now move on to the specific features in prosody system which will be analyzed and experimented in this research, English intonation and stress.

2.4.1 Intonation

Intonation refers to the patterning of pitch changes in utterances. Alternatively, the melody of the voice during the speech with the variations in the pitch, length and loudness of spoken utterances (Wennerstrom, 2001). However, intonation is a complex linguistic system that interacts with grammatical, pragmatic, and effective levels of language descriptions (Halliday, 1967, p. 12).

In intonation, the pitch is defined as the relative height of speech sounds as perceived by a listener and is what we heard when we refer to a voice being "high" or "low". The varying pitch levels throughout an utterance form is what we hear as intonation, the "falling" or "rising" of the voice (Cruttenden, 1997, p. 4). The pitch pattern analysis on this paper is nuclear tone approach model expounded by Cruttenden (1997) which is the

most recent version of a long British tradition of representing intonation in terms of contours such as fall and rise-fall (Snow & Balog, 2002, p. 1026).

Although intonation is primarily a matter of pitch variation, it is important to be aware on the functions attributed to intonation such as the expression of attitudes and emotions, or highlighting aspects of the grammatical structure. Crystal (1976), for example, says that "...intonation is not a single system of contours and levels, but the product of the interaction of features from different prosody systems – tone, pitch-range, loudness, rhythmically and tempo in particular."

For general purposes, the International offers the two intonation marks shown in the box at the head of this article. Global rising and falling intonation are marked with a diagonal arrow rising left-to-right [↗] and falling left-to-right [↘], respectively. These may be written as part of a syllable or separated with space when they have a broader scope. In many descriptions of English, the following intonation patterns are distinguished (Roach, 2009):

- Rising Intonation means the pitch of the voice rises over time [↗];
- Falling Intonation means that the pitch falls with time [↘];
- Dipping or Fall-rise Intonation falls and then rises [↘↗];
- Rise-fall Intonation rises and then falls [↗↘].

Pitch, rhythm, stress, speed, loudness, vocal setting, and the realization of particular phonemes work together to create a constellation of phonological characteristics that, although complex, is immediately recognizable to native speakers as signalling a specific meaning, mood, or attitude towards the listener or the topic (Evans & Jones, 1994)

Intonation is a synonym of speech melody (Dwight Bolinger & Le Merton Bolinger, 1986). Listener conscious of melody in music as a tune, written music shows the melody from left to right. Similarly, speech makes use of pitch in more than one dimension to present a different statement in the speech, which will create the impression of mood and emotion. Author D Bolinger (1986) has mentioned that the relation of music and intonation by connecting level tone, glide tone (rise-fall, fall-rise) in English. Regarding this, rhythm is another critical component of music. In English speech, rhythm is created between stress and unstress in the syllables. Thus, the next discussion will be on stress – another component of English prosody.

Intonation can further define as the use of pitch and the suprasegmental features to convey discourse-level meaning, similar to tone in that the same variable, pitch in manipulated. With the application of intonation, the lexical of the word being spoken is not changed by pitch pattern but the status of the item in the discourse is changed (Zsiga, 2012). According to Arvanti (2011), the term intonation also used to refer to phrasal tonal patterns. While, the terms pitch, accent and tone are traditionally used to refer to lexical tonal specifications. In languages like English, the phrasal level is presented with the complex interplay between metrical structure and prosodic phrasing. These measurable factors determine where and how pitch movements will be performed (Arvaniti, 2011).

The role of meaning in the analysis of intonation by linguistic scholar could classify as the component of intonation. Tone units (or tunes or tone groups) are not only span entire utterances but are also decomposed into smaller parts, the pre-head, head, nucleus and tail (D Crystal, 1976; Cutler & Ladd, 2013). According to Arvanti (2011),

> "The nucleus, defined as the pitch movement on the stressed syllable of the most important word of the utterance, is the only required element of a tune. The unstressed syllables following it is the tail, while the stretch covering all syllables

from the first stressed syllable to the nucleus is the head. Any unstressed syllables preceding the head forms the pre-head. As can be surmised, not all melodies include all four primitives, while particular primitives can span arbitrary lengths of an utterance. These elements do not combine entirely freely. For example, O'Connor & Arnold (1973) distinguished seven nuclear tones, four types of heads and two types of pre-heads, but only twenty types of tone groups, instead of the 56 that all possible combinations of primitives would produce" (p. 765).

In general, all verbal languages use pitch to express emotional and other paralinguistic information to convey a more complex meaning in the intonation. Prior studies have found that there were still many functional components in spoken English which are not able to represent by the current IPA transcript system. For example, the concept of illocutionary acts was presented into the linguistics study by the philosopher John L.Austin in his investigation of the various facets of speech acts (Austin & Sbisà, 1975, p. 129). In Austin's framework, there were two different components in the speech itself, the term locution is what was said, illocution is what was meant, and perlocution is what happened as a result (Austin & Sbisà, 1975, p. 122). We can detect locution in written words but not in the case of illocution without the indication of intonation. Language philosopher and linguists have the critical role of intonation as an indicator of illocutionary force continually. There was hypothesis claimed that a given intonation marking always occurs when a particular illocution is present. (Couper-Kuhlen, 1986, p. 163). By understanding the illocution in intonation, the communication will be much more meaningful. This brings in another important function of intonation, which is the expression of emotions and attitudes. However, Stankiewicz (1964) has argued that languages have many ways of expressing emotive meaning, including the use of lexical items, grammatical forms and constructions (Couper-Kuhlen, 1986, p. 173). In the study of Lieberman/Michaels (1962) has obtained a naturalistic recording of eight different

"emotional modes" based on a set of neutral statements. Another scholar, Udall (1960,1964), has taken four neutral sentences representing different grammatical categories including the "statement", "yes-no question", "wh-question" and "command", and input 16 different intonational contours upon them synthetically. His finding concluded that by assigning different intonation on this neutral statement would result in different emotion perception from the participants (Uldall, 1964). From this experimentation, it can be seen that the correlation between emotional perception and intonation. However, although there were measurable auditory elements like pitch and loudness in intonation, till to date, there are no reliable instruments which can measure these subtle phenomena of emotive expression in a reliable way (Hart et al., 2006).

All languages have to use the pitch and pitch modulation differently. Certain languages have an accentual system based on pitch differences, not stress difference as English did. Their syllables carry varying levels of pitch differences which are able to distinguish the meaning of the word (Arvaniti, 2011, p. 1). Kuhlen (1986) state that

> "these syllables are lexically specified for tone and tonal changes affect lexical meaning wherein languages often referred to as pitch accent languages – such as Japanese, Swedish, or Serbian – tone operates in a similar fashion, except that at most one syllable in each word is lexically specified for tone. In both tone and pitch accent languages, additional tonal patterns are specified at the phrasal level (p. 117).

The focus of this research is in languages without lexical tonal specifications, since it is the intonation of these languages that been mostly examined.

2.4.1.1 Suprasegmentals

Based on past literatures, suprasegmental is defined in this study as a broad term referring to the features that extend over one single sound segment in an utterance included pitch rate, pausing, stress and pitch. In other words. According to Brazil, the term suprasegmental means "units. Intonation, in this study, refers to the manipulation of the features of prominence, pitch, and tone" (Brazil, 1997b).

After defining the terms of suprasegmental and intonation, it is important to understand the linguistic characteristics of intonation. This sub-section briefly reviews the previous theories of the components of intonation. There are various theories on ways to describe the fundamental components of intonation, including the prosodic or contour approach (Halliday, 1967), Bolinger's theory of pitch accent (Dwight Bolinger & Le Merton Bolinger, 1986), the phonetic approach, and the discourse intonation (Brazil, 1997a).

2.4.2 The Description of Intonation and Approach

Willems (1982) stated that "most studies of English intonation in the past may effect be divided into two geographically oriented main streams, namely a British and American school" (p. 22). The British school group and subdivide the tunes and tone groups according to sentence mode. It described intonation in term of overall units of the sentence and do not emphasize on the internal details of contour composition (Willems, 1982, p. 22). Many researchers have treated intonation contours as gestalts, such as Bolinger (1951), Jones (1972), Cooper & Sorensen (1981), and Arvaniti et. al (2009). They claimed that that "it is most often the case that intonational contours are seen as holistically and directly reflecting certain functional or structural aspects of speech, such as the distinction between questions and statements or that between levels of phrasing" (p. 4).

According to Jones (1962) – who in the last edition of An Outline of English Phonetics followed several earlier intonational analyses and the further development from Armstrong & Ward (1926) in a more precisely formulate on the falling and rising tunes, have concluded that English has two basic tunes. These are a fall and rise respectively, Crystal (1976) has mentioned that "their dichotomy of falling tunes and rising tunes referring to the general shape of an intonation contour centered on a specific nuclear type" (p.201), the rising and falling tunes may be spread over a large number of syllables (Armstrong & Ward, 1967; David Crystal, 1969; Jones, 1962). On the other hand, Bolinger (1951) also concludes his critique of level-based analyses and quoted that "it is the intonation to which a sentence owes its character as a configuration. Another work fall onto this development are Cooper & Sorensen (2012), both scholars have presented a series of experiments in which intonation contours are treated as a whole and the highest pitch is taken to directly represent the levels of wording (Cooper & Sorensen, 2012).

Configurational approaches have been widely used for at present British English intonation courses for ESL learner for several reasons. First, their appeal is to be intuitive to the learner: F0 contours are relatively continuous, as Bolinger (1951) put it, "intonation comes to us in snatches of voiced sound (on the vowels and on consonants)" (p. 215). Secondly, the relationship between visual form of contour and function looks natural which has been argued on its similarity on the biological reinforcements of pitch contour production. The use of rising F0 for questions and falling F0 for statements is supported by certain general trends. (Arvaniti, 2011).

Other studies of intonation contour such as Superpositional models, which used accentuation to refer to pitch movements (accents) on stressed syllables, and intonation to refer to the general course of F0, "the rise and fall of pitch as it occurs along the speech chain". As for American School approach on intonation, early level-based models, the

model which descriptions of intonation by means of level tones date from the American structuralists (Pike, 1979; Trager & Smith, 2009). As mentioned earlier, intonation is analysed by means of four levels, extra-high, high, mid and low in this system.

2.4.2.1 Prosodic or contour approach

The prosodic approach was proposed by the British linguists. As Zhuang (2015) stated, "this approach uses the term of tones, sense group, nucleus, pitch range, and relative pitch height to analyze the intonation system". Sweet (2014) identified 5 types of tone choices, "Intonation or tone is either level, rising, or falling…The compound rise or falling-rising tone may be heard in take care! When used to warn, the compound fall or rising-falling tone may be heard in oh! when expressing sarcasm" (p. 228). O'Connor and Arnold (1973) have followed this approach and further refined and developed the prosodic approach which could be classified as one of the Six types of intonation patterns, including "low fall, high fall, rise-fall, low rise, high rise, and fall-rise" by distinguishing between the phonetic and phonological level using an analysis of intonation and connecting the function of tone choice. (O'Connor, Arnold, & Arnold, 1973). While tone groups have been defined as a grouping of tunes which is used to convey the speaker's attitude, however, many of the tone groupings found it difficult to identify the one to one correspondent patterns of attitudinal and intonation contour (Chun, 2002).

Built upon the grammatical unit by O'Connor & Arnold (1973), the Halliday (1967) system includes various forms of five intonation patterns, including "low falling, high rising or high falling-rising, mid rising, mid (rising)-falling-rising, and low (falling)-rising-falling" (Fox, 2002, p. 313). According to Chun (2002), "each of the five intonation patterns had different grammatical meanings, demonstrating different meanings such as certainty, doubt or confirmation' (p.20). However, the contour analysis method proposed by the prosodic approach had a problem in determining the number of different pitch

contours, which were too significant in a given language. This has caused difficulties in teaching and learning intonation with this approach.

2.4.2.2 Bolinger's theory of pitch accent

With extensive knowledge in the parametric nature of the prosodics-phonetic structuring, Bolinger (1958) has coined the term "pitch accent". According to Chun (2002), it is "a prosodic element that is simultaneously a marker of prominent and building block of intonation contours" (p. 21). The stress has an important role in the intonation-contour approach. The term "pitch accent refers to the prosodic pattern that marked both the stress and intonation contours" (Chun, 2002, p. 21). In this case, pitch and stress are intonation features which are related, as pitch could be the main indication of stress(Chun, 2002). Compared to British prosodic approaches, the focus of approach is not as straightforward. In summary, Bolinger (1958) has retained three major types of pitch accents: Accent A (falling), Accent B (rising), and Accent C ("a kind of anti-accent A") (D Bolinger, 1958). These accent-parts are named as "profiles". The major contribution of Bolinger's pitch accent theory was its combination of stress and pitch, which distinguished this approach from the British prosodic approach (Zhuang, 2015)

2.4.2.3 Phonetic approach

Another theoretical approach to intonation from the American school of thought is the phonemic approach. According to Chun's (2002) analysis, "most American analyses have involved four pitch levels (1– 4) and three terminal junctures (the pitch direction on the last syllable of an intonation-group) falling, rising, and level" (p. 24). Bloomfield (1933) adapted the concept of the phoneme, instead of intonation contour, to describe the differences in language pitch. Five phonemes (pitch contours) and their meanings were identified as "fall" (used for statement), "rise" (used for yes/no question), "lesser rise" (used for wh-question), "exclamatory pitch" (used for emotions such as anger, surprise,

sneering and likeness), and "pause-pitch" (used in the middle of a sentence showing continuation)" (Bloomfield, 1933, p. 115) . Bloomfield (1933) used conventional punctuation mark such as periods, question marks as a means to represent intonation. Another representative of American phonetic approach is Pike (1945). His analysis of pitch contour as a series of target values which are connected by the transitional function is similar to the approach used by Bloomfield (1933) (Pierrehumbert, 1981). Pike (1945) found that the categorization of falling, rising, falling-rising was insufficient to describe the actual intonation features effectively. Thus, he has classified the four types of pitch levels to serve as a basic building block for intonation contour: extra high, high, mid and low. According to Pike, these four levels are sufficient to handle all differences of meaning in English. However, by ignoring the pitch level contributed at the end point, beginning points and direction point of the pitch contour, the level indicator could be meaningless in certain complex situations. Thus it was difficult to accurately divide the overall pitch range into four levels (Crystal, 1976).

In summary, the development of traditional theoretical treatments of British contour or tune approach and the American phonemic approaches are both focused on the sentence-level phenomena (Brazil, 1997b; Couper-Kuhlen, 1986). Having defined what is meant by intonation and listed the available theoretical approaches, we will now move on to discuss another prosodic component, stress

2.4.3 Stress

There are many definitions of stress by different scholars, which include emphasis, weight, and intensity in spoken English. Some considered it as the degree of force, the degree of loudness, etc. (Crystal, 1976, p. 113). Stress is a complex phenomenon in the acoustic term. For a primarily physiological definition as stated by Bloomfield (1993), articulatory stress by means consist of pumping more breath and bringing vocal chords

tighter and using the muscles more vigorously in speech. According to Bolinger (1985), as in auditory, it is important to understand that it is not the stressed syllable of the whole word get more emphatic than other syllables, rather a syllable is the one that will get the special emphasis when the word is emphasized in (Dwight Bolinger & Le Merton Bolinger, 1986).

Stress is also the comparative force with which the separate syllables of a sound-group are uttered. There are three main degrees of stress, strong, half-strong or medium, the weak stress being unmarked (David Crystal, 1969).

When we talk about stressing something, we might refer it to a special emphasis on the specific word in the sentence. Acoustically, at least three primary components of stress can be identified, which are the intensity, duration and pitch (Couper-Kuhlen, 1986, p. 22).

2.4.4 Melody and Pitch

Another aspect of correlation of tone and pitch to language is the melody in music. Music and language shared the same key perceptual auditory mechanism. Peretz and Coltheart suggested that mechanisms supporting pitch versus rhythm organization may contribute to the differences between music and language (Coltheart, Sartori, & Job, 1987; Peretz & Zatorre, 2003). According to Rebuchat (2012), "temporal structure is clearly integral to both music and language. Researchers have reported that an infant's brains extract both music and language structural auditory properties, such as intensity, duration, amplitude and frequency, together with the correlations and dependencies between these auditory parameters" (p. 294).

Western music notation offers a unique representation of music composition. The fleeting phenomenon of rhythm and melody are delineated in lines of the stave and notes.

We can clearly visualize the rise and fall in melody as the pitch goes up as the melody progresses. Connecting the music notes in a line, the rise and fall of the melody can be visualized by a contour line. This visualization is identical to the intonation contour in language. However, the sound patterns of languages differ significantly as arbitrarily different sound patterns can code for the same proposition. Contrary, the sound pattern in music do not map arbitrarily into communication content (Rebuschat, 2012). Patel (2007) has reported, "it is clear that the spoken and musical melodies have noticeable differences in terms of structures and perceptions of communication". For example, "one rarely notices or remembers spoken pitch patterns but it is common to find oneself humming or singing along with the popular musical tune. This is because musical melodies are designed to be aesthetically pleasing to the audible range, while pitch or intonation patterns in spoken language are not" (p. 348). Instead, "pitch variations in speech serve communicative functions such as distinguishing the meaning of the words as lexical items in tone languages, making prosodic restrictions to indicate the current status of the words in the utterances by indicating a speaker's emotional state in a non-tonal language like English" (A D Patel, 2007, p. 177). As can be seen, the function of spoken pitch patterns is stressed on the practical communication purpose instead of becoming a central focus of attention in the pitch itself, which is contrary to the pleasant melody to the listeners like music. This indication is important in the mapping process of assigning visual properties in kinetic typography visual learning aid. The understanding helps to avoid over emphasis of the pitch and becomes a musical melody instead of speech.

Over the time, number of musicians have found themselves captivated by spoken melodies and have invested considerable efforts in treating these melodies in a musical framework. Perhaps one of the best known in this regard is Joshua Steel. Steele wanted to capture the prosody of talented actors and orators. So that "the types of modern

elocution may be transmitted to posterity as accurately as we have received the musical compositions of Corelli" (Steele, 1779)

The essay briefly described the "prosogram" model and suggested a few ways in which it might use to examine music-language relations. As Kassler mention, the prosogram model "converts a sentence's fundamental frequency contour into a sequence of discrete tones and glides" and this sequence is mean "to represent a listener's perception of pitch in connected speech. This article briefly described the prosogram and suggested a few ways in which it can be used to compare the structure of spoken and musical melodies" (p. 166).

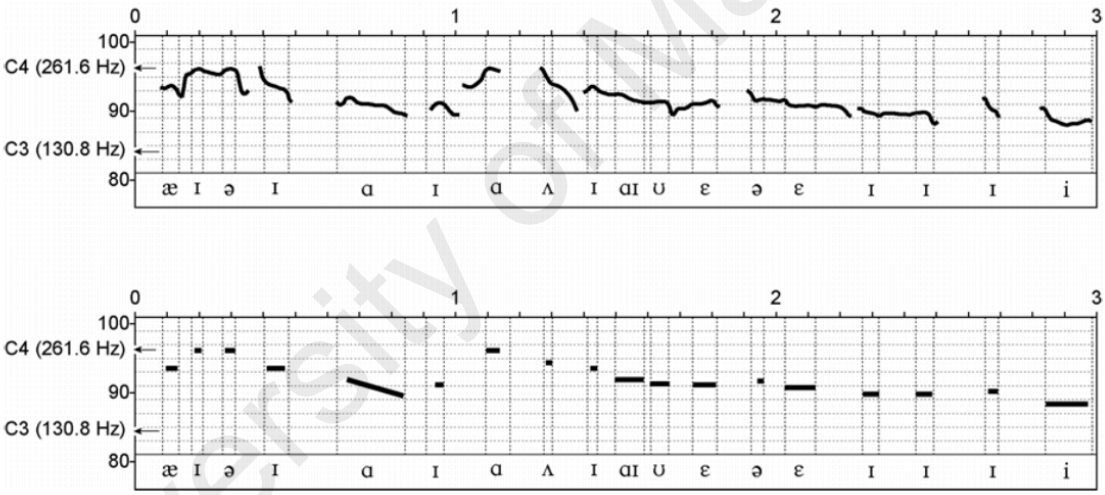


Figure 2.6: Prosogram in comparing speech and music

These were the illustration of the prosogram using the British English Sentence "Having a big car is not something I would recommend in this city" as uttered by a female speaker. According to the specification mentioned by Patel (2007) on both graphs, the "horizontal axis along the top "shows time in seconds, while the vertical axis shows semitones relative to 1 Hz (thus 90 st corresponds to 181.02 Hz)". "Horizontal lines on the grids show "2- semitone increments, and arrows mark the frequencies corresponding to musical C3 and C4 for reference" (p. 167). Characters along the bottom of each graph

are the International Phonetic Alphabet (IPA) symbols for the vowels in this sentence. According to Patel (2007), "the temporal onset and offset of each vowel are indicated by vertical dashed lines above that vowels' IPA symbol. The upper graph shows the original fundamental frequency (Fo) contour of the sentence, while the lower graph shows the prosogram" (p. 167). In this case, the prosogram was based on the sentence's vowels, and required the onset and offset time of each vowel as input. (Different phonological units, such as syllables or syllable rimes, could also use as input. The more recent version of the prosogram does automatic segmentation of speech into syllable-like units based on patterns of voicing and loudness.) In this example, the prosogram has assigned level of tones to all vowels save for the vowel in "car", which was assigned a glide. The prosogram is freely available for download.

Figure 2.6 reveals why the prosogram is useful to those who are interested in comparing speech and music. The perceptual representation of intonation produced by the prosogram is very music-like as it consists mostly of level pitches. These pitches are displayed on a semitone axis, in accordance with the logarithmic nature of pitch perception in speech (Nolan, 2003) [1]. An important point about the prosogram is that the pitches it assigns to syllables are not constrained to follow any particular musical scales. Thus, the pitches of the prosogram can "fall between the cracks" of the Western chromatic scale, and the intervals between pitches can have non-integer values (e.g., 4.2 semitones). (Aniruddh D Patel, 2006, p. 166).

2.4.5 Computer Assisted Language Learning (CALL)

As this research has included computer technology as a screen-based language learning aid, it will be necessary to look into the literature review on the development of computer assisted language learning (known hereafter as CALL). Language support is recognized as a key element in language education. The use of technology in the 20th

century has enabled and facilitated the delivery of useful resources and support to ESL students. In this section, we will discuss the application of CALL. It is a term used by instructors and students to describe the use of computers as part of a language course development (Hardisty & Windeatt: 1989). It is traditionally described as a means of "any process in which a learner uses a computer and, as a result, improves his or her language" (Beatty, 2013, p. 7).

Education technologies are one of the most important technologies developed in many countries around the world. It has been over 50 years since the introduction of computer-assisted language learning (CALL) which has changed how ESL students are learning their second language (B Zou, 2012). The technologies, which first entered the language educational system in the late 1950s, is still continuing to evolve and develop. Today, the technologies in computer-assisted language learning have become more powerful, faster, user-friendly, more convenient and inexpensive, and with better capacity in terms of processing speed and storage of data. With the advancement, a computer software is now able to handle multimedia contents such as sound, video, and animation more efficiently. The development of technologies has provided incredible opportunities and great support to education all over the world.

At the end of the 20th century, the computer-mediated communication and the Internet have reshaped the use of computers for language learning. With the internet today which is 'always-on", coupled with the speed of broadband technologies, digital content of language learning can be delivered to many computer devices like laptop, netbook, tablets and smartphone by streaming video and digital content in real time without the limitation of time and location (Thomas, Reinders, & Warschauer, 2012). However, there are limitations to CALL technologies, according to Thomas, Reinders and Warschauer (2012), " While technologies can provide greater access to information and education

content, facilitate the forming of online communities and aiding learners in constructing their own learning environments, the need to find the appropriate socio-technical infrastructure remains" (p. 2). Dhaif (1989) has also claimed that language learners "can never replace the real classroom interaction that takes place between a student and his or her fellow students or his or her teacher in any learning situation " (p. 4). He has continued to point out "the advantages of CALL program in language learning. Firstly, it is more interesting and could present the language content through interactivities of multimedia contents; secondly, it is self-accessible and able to adapt to the capacity and condition of the learner; thirdly, it is a more engaging learning experience with multimedia contents which involves visual and sound. Lastly, learning with a computer means the learner will have better computer skill which is an essential skill in modern societies (Dhaif, 1989).

The language laboratories, which were founded in the 1970s under the influence of the Audiolingual Method, has marked audio as one of the important parts in computer assisted language learning (CALL). According to Hummel (2013), " the Audiolingual Method was strongly influenced by the psychological framework predominant at the time, in particular behaviourist view that considered learning to occur as the result of habituation between a stimulus and response" (p.110). Basic exercises in Audiolingual Method (ALM) include repeatedly pronouncing the same sentence or words. With the increased involvement of audio component in computers, Crystal (2003) stated that "Microcomputers used as word processors complement the audio facilities, enabling the interactive teaching of all four language skills reading, listening, speaking and writing" (p. 377).

In order to design materials and courses that meet their learners' need of these four language skills, the collaborations among field-specific experts such as linguists and

programmers are important by utilizing the abundance of multimedia materials (Stanley, 2013). As mentioned, new developments in science and technologies have evolved from print to audio, to animation and finally to the age of high-speed Internet streaming connectivity. As a result, CALL is today not only used as a learning aid in developed countries to provide supplementary language skill practices in writing, reading, speaking and listening, it is also able to be used at any remote parts of the world. However, no matter how advanced digital technologies are, it is important to note that language educators need to be trained in operating CALL technologies so it could be used effectively as an instructional aid to the ESL students. (Gündüz, 2005, p. 194; Bin Zou, 2013).

In the field of language learning, CALL has played an important role in the learning of oral skill. Oral communication is very important in the language learning process. According to Chastain (1988) "Speaking is using background and linguistic knowledge to create an oral message that will be meaningful in communication to the intended audience" (p. 272). In the language education of today's classrooms, there is considerable emphasis on speaking skill in which ESL learners could use the language to communicate with each other. Jones & Fortescue (1998) has mentioned that these activities include "simulations, role-plays and discussions. Computer simulations can provide a motivating stimulus for such work, as they offer both a focus for oral activity and a continually changing scenario for learners to talk about" (p. 63). According to previous studies, oral skills can also be improved through videos by studying the dialogue. Based on evidence from several studies, Gambier, Caimi, & Mariotti (2014) has pointed out that "the practice of watching subtitled videos has turned out to be adaptable to all second language learners' levels with the main objective of enhancing their receptive skills" (p. 9)., The conversation, setting and cultural atmosphere will help the student to have a better understanding of the context of the conversation between the characters.

. The main advantage of computer simulations is that they are very motivating to the ESL learning due to the use of multisensory audiovisual and interactivity. CALL is also able to process and give learners instant correction feedback. According to the study of Lee (2008) of the web-based language learning software MyET (My English Tutor), the function of recording and giving the learner immediate and detailed feedback in real time is important. The software MyET is able to tell students about their mistakes and provides some improvement suggestions in verbal and visual ways. Lee (2008) has stated that the software is "able to tell them how to pronounce English phonemes. When learners produce English sounds and record them, the waveform, spectrum and some indications are shown on the screen along with comparison between the student and native pronunciation" (p. 7). The software also enables students to choose different native speakers from different English-speaking countries as their model. This function provides language learners "an authentic" and "native-like" environment to learn English. The study also showed the advantages of self- access and self-monitor system which students can evaluate their own performance based on their own decisions (S. T. Lee, 2008, p. 7). Higgins (1988) has suggested that the computer's main value is as an environment which allows language experiments to be carried out (Higgins, 1988).

The rise of computer-mediated communication and the internet has reshaped the use of computers for language-learning at the end of the 20th century. As stated by Gündüz (2005), "previously, computers used in language teaching were limited to text. Simple simulations and exercises. Technological and pedagogical developments now allow us to integrate computer technology into the language learning process" (p. 212). The integration of multimedia programs with speech-recognition and self-correction capability can immerse students into a more engaging environment for language learning experience. With the Internet, the function of CALL technologies has transited from an information processing tool to a communication tool in society and classroom. ESL

Learners can now communicate with other learners or native speakers of a specific language in the world. Thus, teaching with technology requires not only adaptation of the technology but also a significant shift in the way of thinking by both practitioners and administrator (Stockwell, 2012).

It is important to choose and prepare carefully the materials, lesson planning and classroom management to avoid the ineffectiveness in the usage of computer as a language learning tool, as designing CALL activities could include different integration of advanced technologies. One must take into account a sizeable number of variables, such as learner and instructor accessibility, ease of use of application, platform dependencies. This includes the expertise and support that is available for the given technology. According to Higgins (1998), "the value of CALL is that it allows a richer form of language exploration and play that has ever been possible. The use of computers is compatible with a variety of approaches, methods and techniques of learning and teaching" (p.212).

2.4.5.1 Mobile-Assisted Language Learning (MALL)

Beside CALL, there has been a change of trend in recent years which have shown an increase of interest in using computers and mobile apps for foreign language teaching and learning. Considering the popularity and accessibility of mobile technologies in language learning and teaching, to date there has been a significant increasing amount of second language research on mobile-assisted language learning (MALL) which is an emerging tool in computer-assisted language learning (CALL) research (Son, 2014). As a research domain, Son (2014) has stated that "MALL is in a preliminary, yet evolving, stage; most MALL-oriented research seems to investigate language learners' perceptions of MALL". Though some studies point out the challenges and disadvantages of MALL that result from the limitations of mobile technology and the passive recipient role of the learner in

language learning process (Chapelle & Sauro, 2017). Son (2014) pointed out that " a decade of endeavors to implement MALL has shown that it can provide language learners with opportunities to negotiate meaning and to engage with comprehensible input and output" (p. 48)

As discussed earlier, the implementation of modern tools can be enhanced by creating innovative ways to use mobile devices in language learning. With the mobility of smaller devices like tablet and smartphone, "mobile apps have encouraged exciting opportunities for personalised and learner-centred environments with flexible access to learning materials anytime and anywhere" (Guo, 2014, p. 3). According to Guo (2014) the advantages of mobile-assisted language learning (MALL) is that it has flexibility and ubiquitous access to the mobile apps. The user can learn not only at home, office but anywhere with their mobile device. As MALL is an emergence from CALL, Conway et. al. (2015) stated that Laptops may "increasingly combine with the ubiquitous mobile phone and iPad for learning. Several types of interactions are possible, including touch and voice base interaction" (p. 180). Though some studies pointed out the challenges and disadvantages of MALL that result from the limitations of mobile technology, as we see from this section, there is a greater demand from the new generation for the increased access and flexibility of MALL. Therefore, it is important to share and develop the knowledge in this sector where technology could make a positive impact on the instructional material and process.

2.4.6 Spoken English Visual Learning Aid

As mentioned earlier in Chapter 1, the static printed text has a limitation on representing spoken English on prosody features like intonation and stress. In the area of children language learning, kinetic typography has been studied for the cognition and memorization of vocabularies by stimulating and strengthen the sensitivity of the children

in their perception and learning ability (Lau & Chu, 2015, p. 36). These studies have shown the evidence of functionality of kinetic typography on its visual expressiveness in subtitling.

Many apps have been developed to help the ESL learners to improve their spoken English. The Nativox apps has claimed itself as the only method that allows the users to see the intonation patterns while listening to the speech ("Nativox - Improve your English", 2017). There are three steps in the learning systems with a video showing a real person speaking. First is to listen to the audio from the native speaker and check the melodic line. Second is to repeat, imitate by recording and check on the melodic line. Lastly is to listen and repeat until one obtains the right pronunciation skill (Figure 2.7). The main drawback of this apps is that the text and melodic line are separated. Thus readers would not have any visual representation of exact words or syllables on learning intonation and stress.

Figure 2.7: Nativox Apps

Other apps available in App Store have shared the similar audio-visual approaches. This includes BBC's Learn English Audio and Video apps which claimed to improve English listening skills with the best podcasts. Their supporting tools include videos

offered on the Learn English website on topics such as UK food and culture, everyday life and famous stories and poems. Extra features such as a moving audio script and pitch control have been added in the apps to help with listening and understanding the podcasts. The audio script, glossary and pitch control will help the learners to take control of their learning ("Learn English Audio and Video", 2017). The video and audio have made the learning more interesting and engaging. However, the focus on prosody aspect has not been emphasized in these apps.

On the other hand, Gilbert suggested the use of a prosody pyramid in the teaching method. According to J. Gillbert, the English prosodic system can be illustrated visually with a pyramid shape. (Figure 2.8)

The base of the system is the thought group, a group of words that may be a short sentence, a clause, or a phrase within a longer sentence (Chafe 1970, Bolinger 1989, Brown 1990, Cauldwell 1992). Within the base unit, there is a focus word which is the most important word in the thought group. Within the focus word, one syllable is given the main stress. That syllable functions as the peak of information within the thought group. It is sometimes called as the nucleus, or the peak.

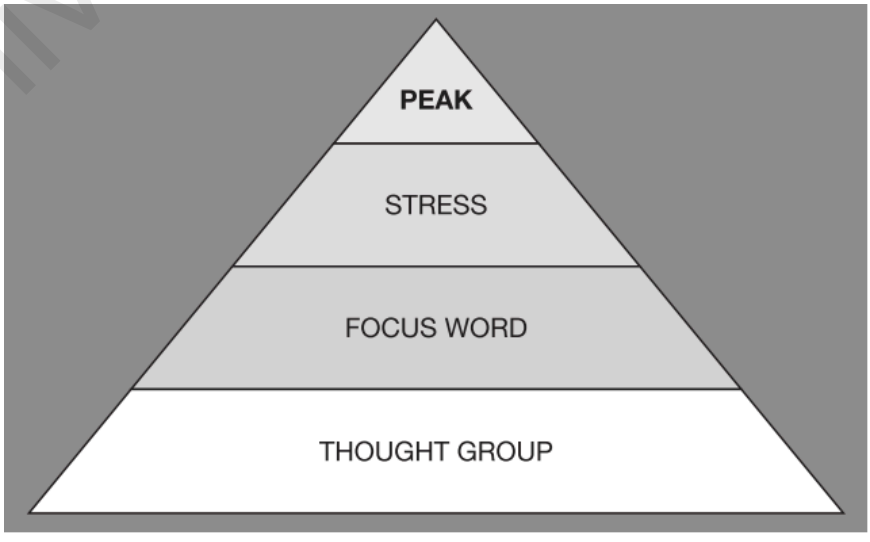


Figure 2.8: Prosody Pyramid

Besides digital visual learning aid, there are other visual learning methods for intonation and stress for the deaf and hearing-impaired community (Figure 2.9). For instance, Underhill (2005) suggested a method of "using fingers with connected speech. When apart, the fingers represent words spoken separately, e.g. 'Hold .. up .. your .. left .. hand.' When together, they represent the words spoken as one flow": 'Hold up your left hand.'" ". The hand movement has multiple functions in helping the learner to recognise separate words or stream of speech, longer phrases and grouping words" (Underhill, 2005, p. 164).

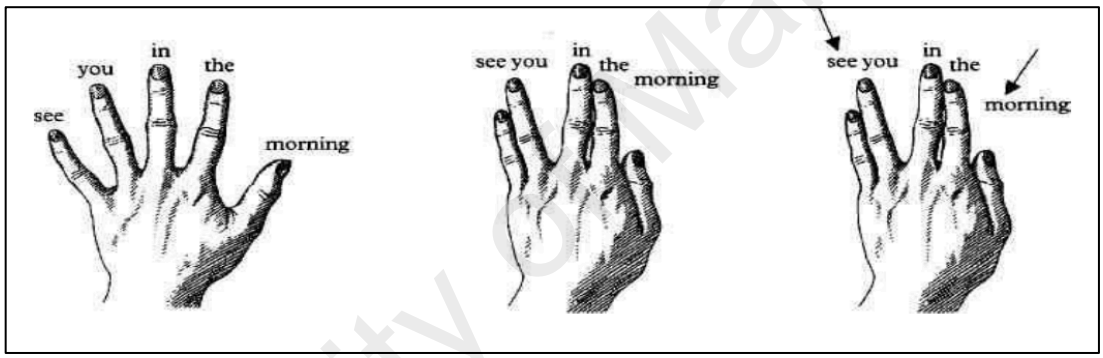


Figure 2.9: Using finger as visual cue connecting to speech (Underhill, 2005)

Underhill (2005) also suggested five kinds of activity that could be included in the intonation syllabus which are:-

1. signs and symbols to represent intonation;
2. sensitization to intonation;
3. listening to pre-recorded cassettes;
4. intonation from a printed text;
5. learners' spontaneous speaking.

Visual learning always plays an important role in pedagogy and in neurobiology. We can learn why pictures both resemble and differ from other text. When viewing on the picture, our brain process the information quicker than words. (Kenney, 2010, p. 178) .

CHAPTER 3: RESEARCH DESIGN AND METHODOLOGY

3.1 Introduction

This chapter highlights the research design and methodology used in the data selection, data collection and data analysis of the study. The first section covers the selection of the participants, followed by the type of data collected in this study. The third section describes the collection and organization of data and the last part explains the procedures for data analysis.

The methodology was designed to acoustically analyse the prosody properties of English in relation to the kinetic typography visual representation based on the descriptive characterization scheme proposed by Wong (1995). According to Wong, the scheme of characterization has two parts, structural description and behavioural description. Structural description is a physical description of the form based on the notion of dimension and behaviour description describes the manner of 'how' the form changes in kinetic typography. Table 3.1 shows how the visual component is matched to the measurement of the auditory in English speech with both descriptions in the character scheme.

3.2 Research Design

Table 3.1: Research Design

Research Question	Development	Source of Data	Method of Collection	Method of Data Analysis
Q1: 1. How do we match the visual form in kinetic typography with the auditory structure in English prosody? Q2: 2. Can kinetic typography make a significant different in the recognition of English intonation and stress.	Matching of visual representation in kinetic typography with auditory of English intonation Adapting characterization scheme develop by Y. Wong (1995)	25 Saito students average age at 23 years old – 8 Malay – 7 Indian – 10 Chinese All non fluent speakers Homogenous in – Education background – Profession – Age	Learning spoken English sentence – "Where do you live?" Learning medium: – Printed copy with IPS transcript – Kinetic typography on screen	Praat version 6.0.23 (Boersma & Weenink, 2016) Measurements: Pitch and amplitude differences between "do" and "live" in the sentence Statistics form t-Test - Means - Variance - t Statistic - P-Value

3.3 Research Methodology

A research methodology was developed to answer the research question 1 and 2 as stated in Section 1.4. The first research question seeks to find out the structure of matching visual representation properties in kinetic typography and the prosody auditory component of English intonation and word stress. The relationship between the auditory and the visual representation component in kinetic typography will be analysed using an expert approach in order to obtain a precise and matching result. The details of the matching structure will be presented in Chapter 4, which will include the measurement of intonation frequency and amplitude of stress in auditory aspect. Visual composition analysis will include Visual Technique, Visual Action, visual expression and presentation, adapting the Characterization Scheme of temporal typography design by Wong (1995). This scheme aligns with the design solution needed for the matching of the English auditory component in this research. Different dimension of kinetic typography visual representation will be explored to match the pitch in English intonation and amplitude in English stress.

The second research question relates to the effectiveness of kinetic typography as a visual learning aid for English intonation and stress. A pre-experimental design was used to ascertain the effectiveness and accuracy of visual perception and interpretation of the visual approaches of kinetic typography. The comparative results in t-Test were presented upon the completion of the experimentation, with the independent variable of visual representation of spoken English and the dependent variable from the learning outcome of the sampling.

3.4 Sampling Selection

The sampling population is from Saito College, whereby 30 college students were selected based on probability sampling method with a known probability of participants

selection method advocated by Kalton (1983). The non-probability random sampling method (Kalton, 1983) is not suitable for visual representation of language research because random selection will most likely pick up participants who rarely or proficient in spoken English causing unreliable results. The two main criteria of the selected sample aren that they are of a non-proficient English speech background and age to be between 19 to 21 years old. All participants have completed their secondary or high school education at a Malaysian National Secondary School or Chinese Independent High School. The ethnicity of the participants are 9 Malays, 14 Chinese and 7 Indians. The distribution of gender is 16 males and 14 females. Ethnicity and gender independent variable are at the least control in this research as the main interest of this research were on the effectiveness of intonation perception among the young college students regardless of ethnic background influences. However, this could be a separate topic for further discussion in future on the development of this research to study about the gender and ethnicity influences. This sampling method will ensure the equality of the sampling background and minimizes the threat of the background differences in pretest and posttest data comparison. The advantage of this method is that it guarantees the inclusion of the type of people according to the research need (Kumar, 2011). The sampling size was decided according to the sampling size speculated by G-power[1], a computer statically power analysis online software. The speculated sampling size is 27 (Figure 3.1). The calculation is based on the one directional hypothesis of that kinetic typography is able to represent English prosody, with 0.05 on Type Error 1, 0.8 on Type Error 2 and moderate Effect size of 0.5. Thus, with the consideration within this computational result, the number of sampling was set to 30 participants.

*Power is a tool to compute statistical power analyses for many different t tests, F tests, $\chi 2$ tests, z tests and some exact tests. G*Power can also be used to compute effect sizes and to display graphically the results of power analyses.

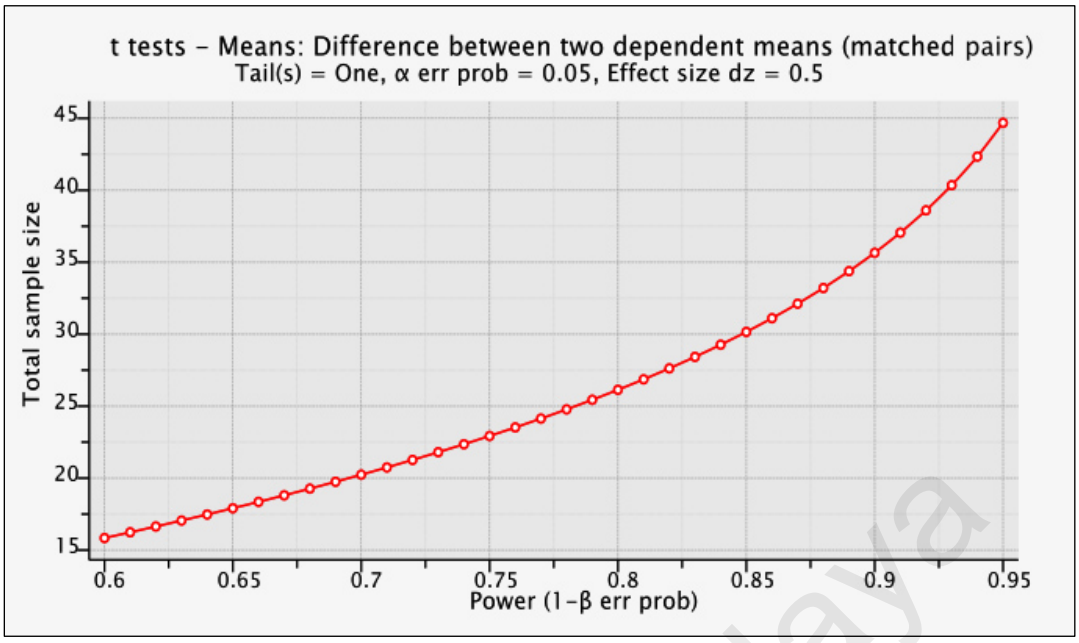


Figure 3.1: Using GPower, the speculated sample size is 27

3.5 Methods

The scope and purpose of the research design of this study are to evaluate causal relationships between intervention and outcome with exploratory approaches. Therefore, this study employs Pre-Experimental Design, using a One-Group Pretest-Posttest Design type. A pretest sampling was recorded and observed before the intervention. A post-test will be observed and recorded with the intervention of kinetic typography visual learning aid. Differences from the pretest to posttest are the outcomes of interest, which are presumed to be the result of the intervention (Salkind, 2010, p. 1085). The data obtained in this research can be discussed descriptively and analyzed using paired-sample t-test with the experiment outcome in scale as internal or ratio variable (Thyer, 2012, p. 167). This will enable the test of the hypothesis, by developing an experimentation to reach the conclusion of the relationship between independent and dependent variable (Cash, Stanković, & Štorga, 2016, p. 212). The pretest and posttest experiments will obtain two dependable variable O_1 and O_2 with one treatment of independent variable of X in posttest which is the treatment of kinetic typography visual aid. The dependent variable O_1 is the outcome in pretest by reading on regularly printed sentence with IPA mark, and the

dependent variable O_2 is the effect of reading with the independent variable of kinetic typography visual aid in the posttest (Table 3.2).

Table 3.2: Pre-experimental design

Group	Pretest	Treatment	Posttest
E	**O_1**	**X**	**O_2**

3.6 Hypotheses

This research study explored the effects of kinetic typography visual representation in its relationship with the auditory aspect of English prosody in computer environment learning. Exploring kinetic typography in an experimental with the objective of providing an insight of how the visual form of kinetic typography has an impact on English prosody learning processes. The use of kinetic typography as a visual learning aid may potentially affect the English intonation and stress prosody-learning processes of ESL students. It is believed that the treatment of kinetic typography visual representation form as visual aid will create a positive learning outcome in term of recognition of pitch and loudness in English prosody. Properties of visual form and movement of typography are also of interest in this research.

Therefore, the research hypothesis of this study was that kinetic typography as a visual learning aid can generate a statistically significant difference in helping the ESL students to recognize English intonation and stress in comparison to conventional textbook reading method. To test the hypothesis, the study conducted comparison and analysis of the pretest and posttest result with the standard native intonation to retrieve the data on the differences between the pre and postest results.

3.7 Instrumentation and Calibration

3.7.1 Kinetic Typography Visual Representation

A visual aid was used as a medium to test the effectiveness of kinetic typography visual representation form in its relationship with English prosody. The first step in creating kinetic typography visual aid is to create an animation that would represent the auditory aspect of English intonation and stress. Thus, a framework that was based on a descriptive structural development (Wong, 1995) was designed with references to two types of information, that are; 1) English prosody and 2) Typography characterization scheme based on the 4 level descriptive structural developed by Wong (1995) (Figure 3.2), which includes Visual Action (VA), Visual Technique (VT), Expression and Presentation.

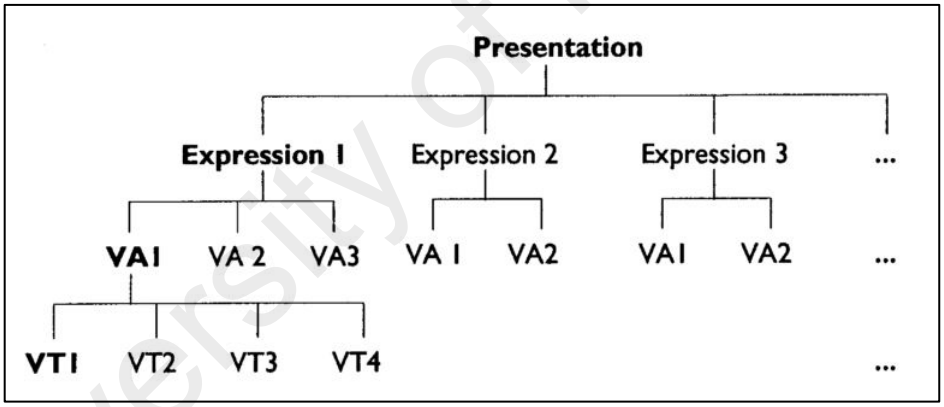


Figure 3.2: Structural hierarchy of characterization scheme

Visual Action: One way to characterize the manner of a Visual Action is from the point of view of formal identity. A form can be perceived to have certain identity. For example, a particular polygon may be identified as a "square". Similarly, a particular combination of dimensions gives a typographic form its formal identity. When a form changes dynamically over time, its formal identity may be perceived as changed or preserved. A "square" may be perceived to have a different identity (i.e., "triangle") or the same identity at the end of a change. The manner in which the formal identity changes

can be divided into two basic categories, variation and transformation. Wong (1995) stated that:

> "The term variation refers to changes within the form; it does not alter the way we perceived the formal identity. For example, when text is scaled up proportionally, we recognize the end shape as the same except it is larger. In a similar vein, a square scaled up is just a larger square. In general, dimensions, which we do not perceive as having discrete states such as colour value, saturation and transparency, are transitions. The term transformation denotes changes in which results in a perceived shift in the formal identity. I have further divided this category into two sub-categories: transition and deformation. Transition refers to a closed non-linear change from one state to another, each being perceptually distinct from the other. For example, a change from one hue to another or from one font to another. Helvetica and Times are two different typefaces which can be morphed but the two end states are perceived as having different formal identities. This can be compared to a change in a form such as from a square to a triangle. Deformation refers to an open-ended change of the form through distortion. A letterform may be stretched and distorted to a point of losing its original identity" (p. 21).

The sequence shows a piece of text being distorted over time. This movement is generated by a mathematical model simulating the physically based motion of springs (Wong, 1995).

The information of English prosody under this framework will include the intonation description, IPA (International Phonetic Alphabet) transcript, the dimension of intonation and stress. As for the section of kinetic typography, the information will include the typography characterization scheme's conceptual structure in a four-level hierarchical,

which will use to categorize and describe the typographic presentations. The four levels of descriptive presentation are Visual Technique, Visual Action, Expression and Presentation (Figure 3.2). The adaptation of characterization scheme will allow the kinetic typography visual aid to be build up from the basic unit (Visual Technique) to the complete presentation. Based on the hierarchy of the scheme, kinetic typography visual design will start from Visual Technique, followed by Visual Action unit which is the combination of form and movement in Visual Techniques. Then an Expression, a composition description of how was the movement feel and finally, the presentation which represents a meaningful composition by a set of Expression. (Table 3.3)

Table 3.3: Visual and auditory mapping

English prosodic informations				
Phrase	Where	do	you	live
Intonation Description	Fall	level	level	Fall rise
IPA (International Phonetic Alphabet) transcript	↘			↘↗
Frequency (Hz)	137hz-226hz-153hz	173hz	178hz	234hz-134hz-218hz
Stress	Stress	unstress	unstress	Stress
Decible (dB)	85dB	81dB	81dB	83dB
Kinetic Typography Visual Composition				
Visual Technique :				
Form	Bigger typeface, extended down	smaller typeface	smaller typace	bigger typeface, extend up,
Composition	Go down	Stay	Stay	Go up
Visual Action :(form+composition of visual technique)	"Where" move down - "do" scale up and return to original-"you' scale up and return to original state - "live" move up, extended up vertically then back to original shape at higher position..			
Expression (composition of one or more visual action)	"Where" Fall down organically , "do" and you subtlely pop up in sequence and finally "live"bounce up agressively			
Presentation (Composition of expression)	The animation falls according to the voice on the word "where" and rises again towards the end of the question. This indicates to the listener that the speaker is aware that the litsener should know the answer.(Kelly,G 2000)			

The sentence “Where do you live?” and the audio track used in this research were taken from the book “How to Teach Pronunciation” published by Pearson Education Limited, which was written by Gerald Kelly, and the Compact Disc accompanying the book contains examples of sounds and sentences from the text. According to Kelly (2000), this book was written for all teachers of English who wish to improve and develop their knowledge and skills in teaching pronunciation. The book covers the topic of both intonation and stress, so selected sentences from Kelly’s book was used to demonstrate this two features. Thus, this book provided a suitable context for this research and a written permission to be allowed to use this book for this study has been obtained from the publisher.

According to the author’s description, the expression of the spoken words “Where do you live” shown as the speaker has asked the question for the second time which the person is having already been given the information once but having forgotten it. The intonation of the voice falls on the word “where” and rises again towards the end of the question indicates to the listener that the speaker is aware that they should know the answer. (Kelly, 2000, p. 3). The acoustic element of this sentence will be captured and analyzed in the following section.

3.7.2 Audio Analysis

Analyzation of the audio is necessary to explain and detect the intonation contour and the acoustic feature in a sentence is shown as a pitch modulation, which is the changes in pitch that are described as the rises and falls. (Gooden, Drayden & Beckman, 2009) (Arvaniti, 2011, p. 1). Thus, an audio recording and analysis software known as Praat (Boesrma & Weenink, 2011), which was designed and developed by Paul Boersma from The Institute of Phonetic Sciences at the University of Amsterdam, was used in this research to capture the audio data of standard native English and sampling of the 30

participants. Praat ('talk' in Dutch), enables acoustic analysis and annotation speech data and is used widely for phonetic and phonological researches (Allan, 2013, p. 127). Pratt has been used in many linguistic researches which centers upon intonation such as a study by Ladd & Schulman in 2003 for pitch accent in English, and stress study the dialect of Papiamentu by Remijsen & van Heuven in 2005 (p. 128).

Analyzing the structure of pitch modulations and the primitives that make up pitch contours in spoken English are a challenging process because the connections to the segmental material are harder to identify and associated meanings are not easy to be located. Moreover, they deal with information structure and pragmatic interpretation rather than lexical meaning.

Arvaniti (2011) highlighted that there are at least three main points about intonation. First, the shape of intonational contours with a given pragmatic interpretation can vary noticeably, depending on the segmental material with which they are uttered. Such differences are related to the overall prosodic structure of the utterance with which the contours are associated, including the number of syllables and the position of stresses. Second, contours do not have a persistent meaning either within or across speech. The interpretation of the content may well depend on lexical and other elements that accompany the use of the tone. The audio analysis of intonation and stress in the recorded audio aims to capture these properties and explain the connection between intonation and meaning to map the visualization elements based on the result of the analysis. Many researchers have treated F0 contours as a configuration that is a general pitch movement that incorporates entire expressions of speech and has a constant meaning in the content of speech. In other intonation analysis models, melodies are seen as being composed of primitives of dynamic tones, such as local rises and falls, or level tones, such as high, mid and low semantics. It is most often the case that intonational contours are seen as

holistically and directly reflecting certain functional or structural aspects of speech, such as the distinction between questions and statements or that between levels of phrasing (Arvaniti, 2011, p. 3).

The main purpose of using Praat software is due to its capability of creating publication-quality spectrographic graphics with a general set of tools for analyzing, synthesizing and manipulating speech (Podesva, Podesva, & Sharma, 2014, p. 375). The graphical function in Praat enables the researchers to visualize speech with waveform, pitch curve and intensity curve, which is important to construct functional kinetic typography visual aid by mapping on all these auditory-visual with kinetic typography animation visual representation. (Figure 3.3)

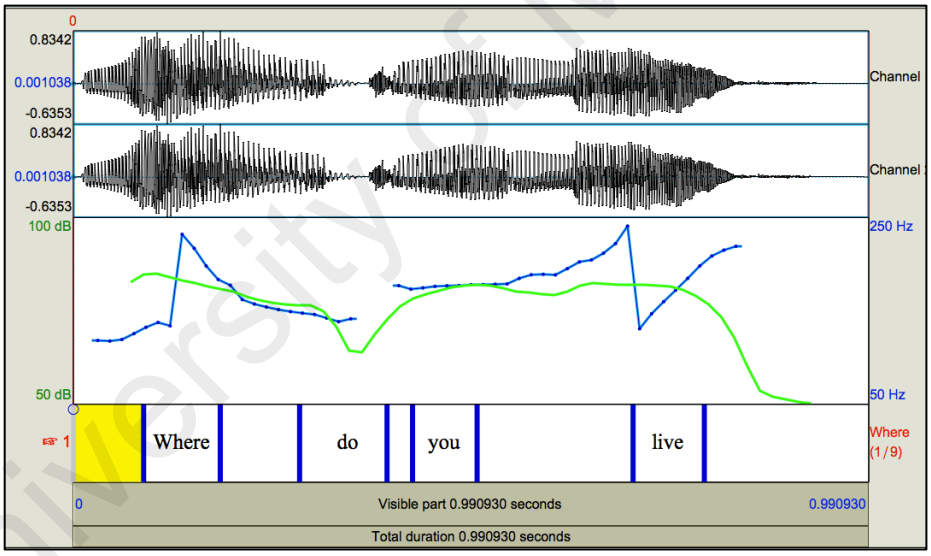


Figure 3.3: Praat auditory analysis

To use kinetic typography as a visual representational form of intonation and stress in this sentence, we need to match, measure and transform the parameter obtained from Praat into a meaningful visual and animation with kinetic typography in another software, Adobe After Effect. The matching will only involve two typographic dimensions, which is form and composition (Figure 3.4).

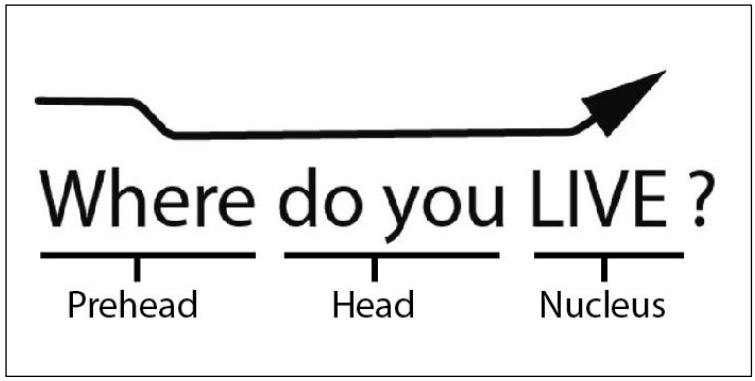


Figure 3.4: Tone unit

As mentioned in chapter 2, one of the components of intonation is the tone unit. The sentence of "Where do you live?" could be divided into three tone units which are Pre-head, Head and Nucleus (Arvaniti, 2011) and were constructed according to this expert method. The visual auditory and measurement of pitch and amplitude in Praat will be used to match the position and size of kinetic typography animation by matching the Physical dimension of typography form parameter framework proposed by Wong (1995) as shown in Figure 3.5. The results of this study are further discussed in Chapter 4.

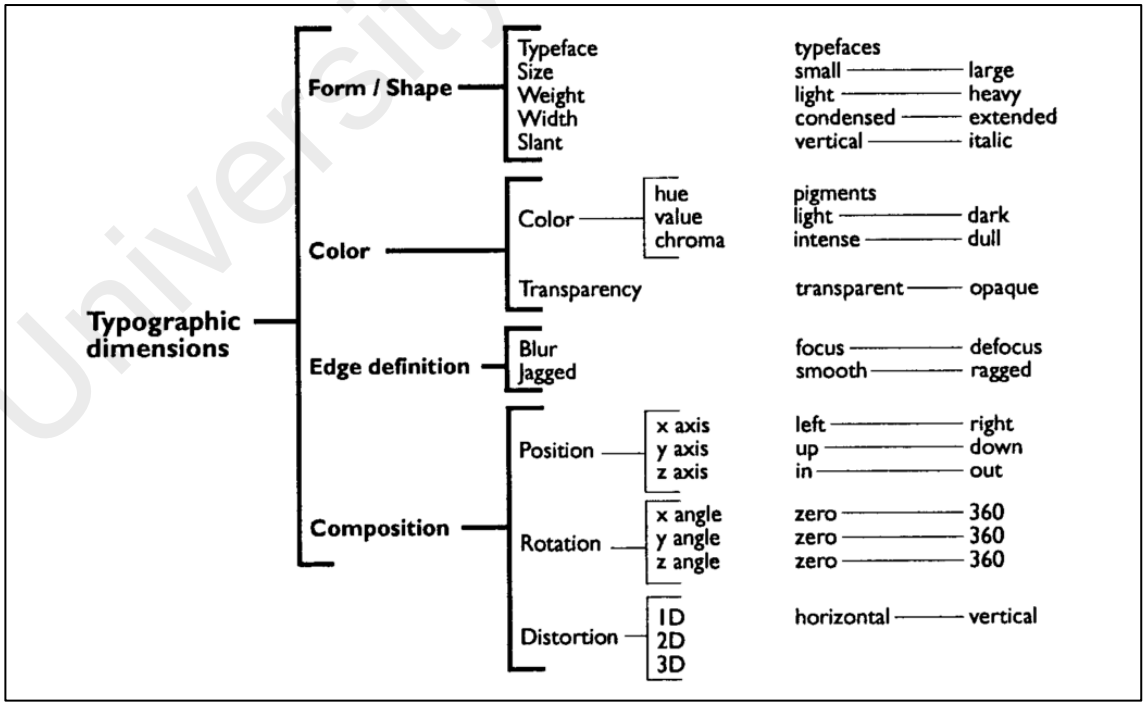


Figure 3.5: Physical dimension of typography form (Wong, 1995)

3.7.3 Pilot Test

Upon the completion of the kinetic typography visual aid in animation and visual composition, before the actual experiment, a pilot test has been carried out with five students from Saito College. A computer lab was set up as a recording booth, with a 14-inch MacBook Pro to display the kinetic typography as a visual learning aid. The built-in MacBook Pro microphone was held 15cm from the speaker's mouth to capture the participant reading voice. An open source software Praat was used to visualise the recorded pretest and posttest intonation patterns.

Participants will first listen to the audio track of the selected sentence, and then read out the sentence in a printed copy with IPA mark. After a short break, participants continued with the posttest by reading with kinetic typography visual aid on screen. Both audio data from pre-ttest and posttest were recorded into a computer hard disk. There was a total of three sentences to be read by the participants which include "Opportunity" and "I need a pen, a pencil and some paper" which were used to minimize the repetitive memory threat in the experiment.

Upon the completion of experimentation in the pilot test, it was found that if the audio track is played before the sentence reading, it would influence the experiment negatively. This is because, after listening to the audio track, participants seem to be distracted and lose their focus on the reading of visual representation in kinetic typography. Thus, it is observed that participants were affected either by their auditory sensory or visual sensors in the process of experiment. Therefore, the pilot test led to the elimination of the audio track in both pretest and posttest during the actual pre-experimental design. This ensures a better accuracy of visual recognition effectiveness in the test result.

Another improvement of kinetic typography visual aid through the observation of pilot study includes the manipulation of typography dimension. Instead of over stretching and

distorting the letterform to represent high pitch and louder amplitude which affect the readability of text. The visual representation with position movement of a text to above or below the baseline according to the visual pitch shown in Praat software is much more accurate and comfortable to read.

Another observation found via the pilot study is that kinetic typography visual aid looks unnatural without applying the animation techniques of ease in and out. Thus, the ease-in and ease-out animations were applied to the composition of kinetic typography visual aid in the pre and posttest. As in using keyframes and in-between frames, an essential consideration is the speed or rhythm of motion and the adjustment which is referred to as "easing" in motion-based media was use. When use the "easing" technique, the motion of objects will appear to have a more natural movement instead of a mechanical movement. The movement was derived from the effect that gravity has on objects in the real environment. In animation technique, there are two primary methods of easing which are "easing in" and "easing out". "Easing in" technique makes the objects start slowly in the beginning and it accelerates towards the end of a movement. "Easing out" makes the object to start quickly in the beginning and slows down towards the end of a movement. The designer uses positive and negative values to adjust the easing in and out appearance of the objects. The easing in technique allows the situation to start quietly and prepare the viewers before accelerating its speed to improve the visual rhythm and dynamism. While, the easing out technique begins with more tension but gradually reducing the tension towards the end. In both cases, easing makes use of the tone of the motion and supports the sense of visual play (Hostetler, 2006). By applying these techniques to the final composition, the visual aid looks more organic and natural in its visual representation of native spoken English (Figure 3.6).

Figure 3.6: Pilot study of kinetic typography with stretching form.

3.7.4 Pretest

In the actual experiment, a pretest and posttest sampling data collection were conducted. A similar recording booth as in pilot test was set up. Students were given a simple briefing on the procedure of both experimentations. During the pretest, participants were given a piece of a printed sentence with an arrow indicator as shown in the book and all cap on the word "LIVE" to indicate the tone unit of Nucleus and word stress (Figure 3.7). Participants were instructed to read aloud according to what they have seen. After they have attempted the first reading, a brief explanation of the indicator was explained accordingly. Next, participants were allowed to read it twice to familiarize themselves with the sentence, and then only the third reading was recorded and saved into Praat for further analysis of the dependent variable.

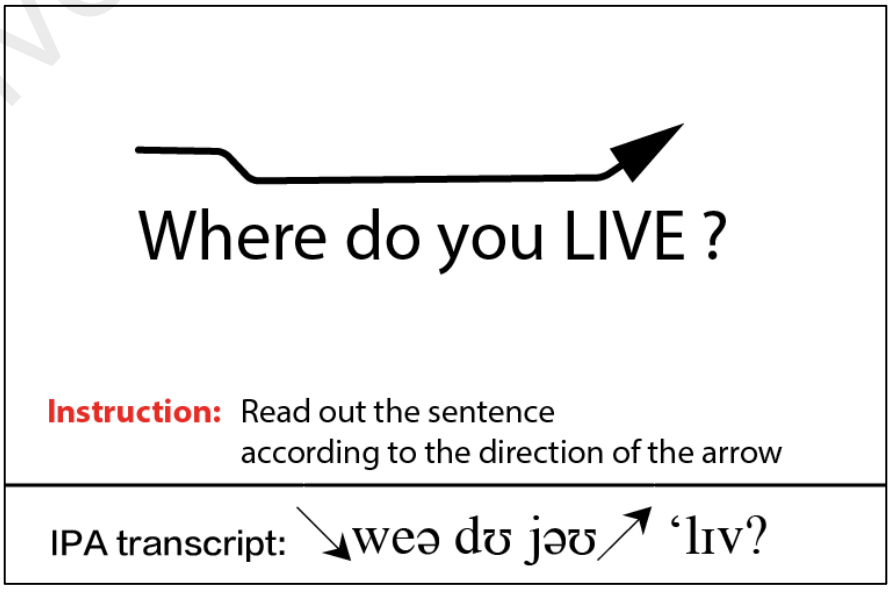


Figure 3.7: Pretest reading sheet

3.7.5 Posttest

In posttest experimental design, kinetic typography visual aid act as a treatment to obtain the second group of the dependent variable. A brief definition on the visual cue in kinetic typography was presented to the participant, e.g. higher position for a higher pitch, a lower position for lower pitch and bigger text is louder than the smaller text (Figure 3.8). The participant was asked to observe the movement and shape of kinetic typography on screen and intuitively read with their interpretation of pitch and loudness from what they have seen after the pre-count of three. Similar to the pretest experimentation, the participant was allowed to read it two times for familiarization and only the third reading was recorded into Praat for further analysis.

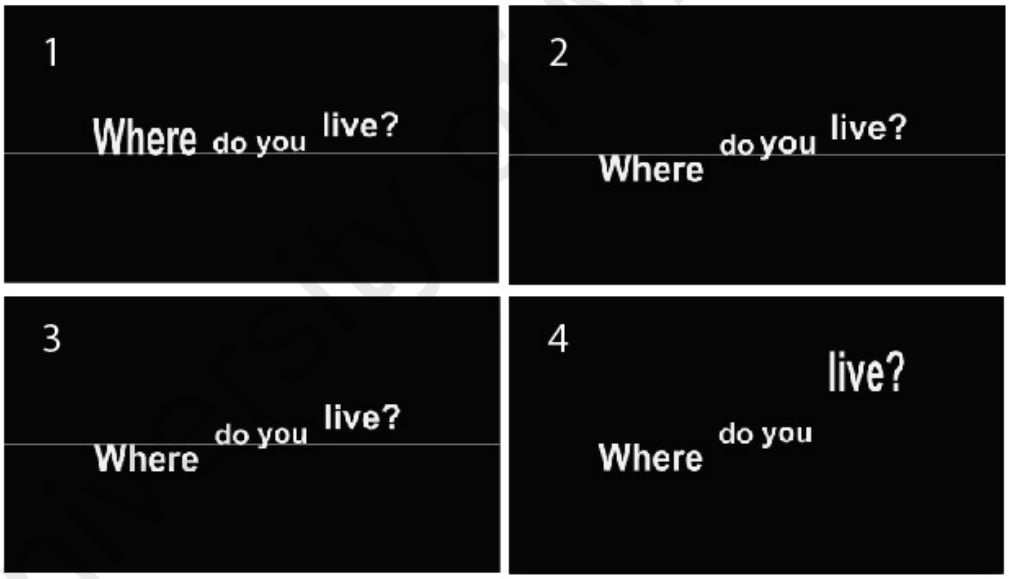


Figure 3.8: Kinetic typography visual aid

CHAPTER 4: RESULTS AND DISCUSSION

4.1 Introduction

This chapter will discuss on the quantitative data collected to construct a visual learning aid, as well as the results of experiments carried out on 30 participants from Saito College (9 Malays, 14 Chinese and 7 Indians), which were analyzed based on the comparison results between the pretest and posttest with the standard native speaker intonation. Quantitative data were analyzed using a paired samples t-test for both pretest and posttest data. Total pitch (Hz) differences between the two words "do" and "live" were analyzed first in the pretest session. Next, a post-test analysis with the same words on the pitch (Hz) measurement were conducted. Then, the analysis of the comparison result between pretest-standard and post-test standard will be discussed. The final section of this chapter will discuss the findings of the data analysis.

4.2 Analysis of Quantitative Data for Research Question 1

The research question 1 "How do we match the visual form in kinetic typography with the auditory structure in English prosody?" has been partially answered in Chapter three in section 3.7. It demonstrated shows the method of matching Praat auditory visualization in spoken English with the visual dimension of kinetic typography. This section will further discuss and analyze the visualization matching of kinetic typography with quantitative data obtained from Praat software.

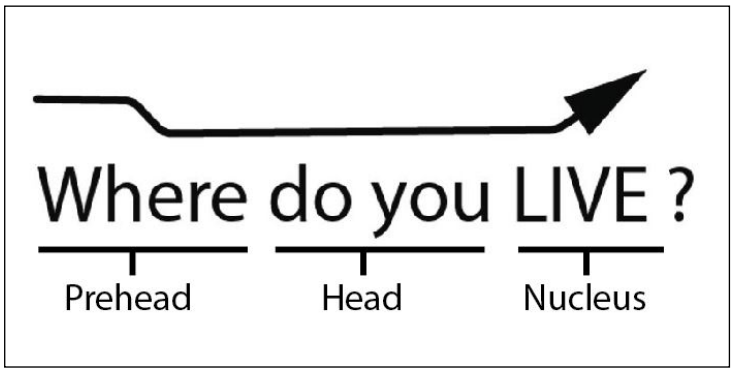


Figure 4.1: Tone unit

As mentioned in chapter 3, section 3.7.2, the components of intonation is the tone unit. There are three tone units in the sentence of "Where do you live?" which is Pre-head, Head and Nucleus (Arvaniti, 2011) (Figure 4.1). The method used to construct the visualization is according to this Arvaniti's method. The first composition starts with the word "Where" that has only one syllable and as the pre-head of the tone unit, it contains the complex tone structure of Rise-Fall. The auditory pitch measurement from 0 sec to 0.34 sec shows that it was rise from 137hz to 226hz and then fall aggressively to 153hz. The animation to match this information in Visual Technique was a bigger typeface according to the intensity of the voice in 85db. The form has been extended up to match the rising pitch which was scaled up from 137hz to 226hz. Then, The word has a moving down form to match the falling pitch measured at 226hz to153hz. The complete expression description for the word "Where" falls organically and bounces up subtlety at the end (Figure 4.2).

Figure 4.2: Descriptive animation diagram

The second unit composes of "do you", which has two syllables and act as Head of the tone unit. It contains a simple tone structure of level with minimal changes on the pitch.

The auditory pitch differences measurement from 0.37 sec to 0.60 sec shows that the pitch has not much change at 173hz to 178hz. The animation to match this information in Visual Technique is a smaller typeface according to the intensity of the voice in 81db. The form was scaled up slightly and returned to its original state to match the subtle changes from 173hz to 178hz, both words stayed at the baseline level to show the expression of pop up effect without any aggressive movement as in the Pre-head.

The last tone unit compose of "live" that has only one syllable which functions as a focus word and it is the most important word in the whole sentence. This syllable is given the main stress and function as the peak of information within the thought group (Meagher-Alkan, 2010). This tone unit contains the complex tone structure of Fall-rise. The auditory pitch measurement from 0.62 sec to 0.83 sec shows that it fell from 234hz to 134hz and then rise steeply to 218hz. The animation to match this information in Visual Technique is similar to Pre-head words "where" with bigger typeface according to the amplitude of the voice in 83db. Even though there is a fall in the pitch measurement from 234hz to 134hz, but the duration of fall is too short to be animated. Thus in actual animation, the visible action moves up from baseline and extended up in shape transformation to match the more prominent rising tone from 134hz to 218hz. The complete expression for the word "live" is an aggressive bounce up effect in organic style.

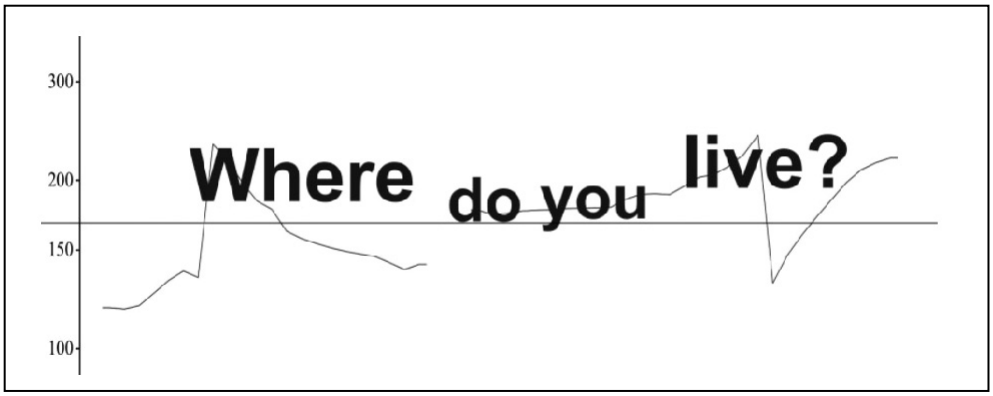


Figure 4.3: Adobe After Effect animation

The mapping method which used Wong (1995) scheme of characterization scheme that is based on structural description and behavioural description have proven to be effective in designing the kinetic typography visual learning aid analytically as shown in the result of recognition of intonation and stress in this experiment. It was believed to be an improvement over conventional IPA reading without the help of kinetic typography visual learning aid. However, the parameter setting in visual and motion of kinetic typography visual learning aid is not precise in its details due to the limitation of one to one mapping in the numeric scale setting, instead, the characteristics and movement of the animation were based on the visual form of intonation contour and amplitude scale shown in Praat (Figure 4.3).

4.3 Analysis of Quantitative Data for Research Question 2

The quantitative data in next section will continue to answer research question 2: "Can kinetic typography make a significant difference in recognition of English intonation and stress?".

There were two categories of finding which will answer the question in the significance of differences with the treatment of kinetic typography.

The first category of data analysis is to see the t-Test result in the pitch differences between "do' and "live" from the reading of 30 preselected participants. The word "do" was selected as an anchor word for comparison with the rise-fall structure of "live" due to its neutral status as level pitch and moderate amplitude. The mean pitch and amplitude was recorded on the word "do" as level tone and the maximum pitch and amplitude was measured for the word "live". The differences between the two words were captured as the dependent variable of treatment results. The independent variable was the various participants in both pretest and posttest, and the dependent variable as the separate result

shown in before the treatment in pretest and after the treatment of kinetic typography visual aid in the posttest.

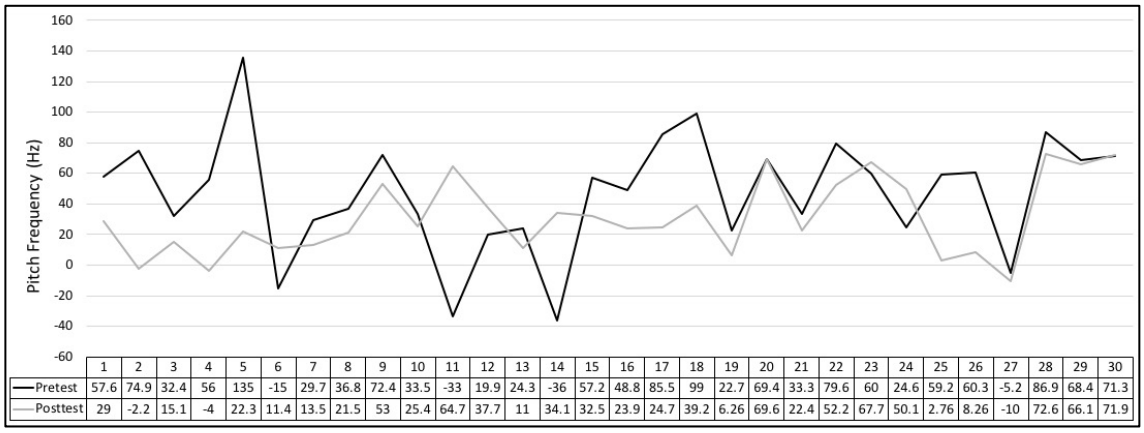
After the experiment, the mean score of the treatment result is 31.07hz in posttest, which is 33.82% lower than the mean score of the pretest at 46.95hz, before the intervention. Comparatively, with the treatment of kinetic typography visual aid, the pitch measurement is closer to the standard score. With the null hypothesis mean differences of 0, the computation t-Test calculation of P-Value is 0.021, and the t Statistic is 2.12, which is bigger than the t Critical one-tail at 1.70, this show there is a significant difference between these two tests. The overall pretest group has significantly lower performance in its accuracy of recognizing the pitch differences within "do and "live" as shown in Table 4.1.

It appears that the participants have higher variance score at 1437.07Hz in recognizing the English intonation at pretest experiments compared to the posttest Variance score of 609.36Hz (Figure 4.4). Thus, the lesser variance score in posttest shown a positive impact on a better consistency and accuracy in English intonation recognition.

Table 4.1: Significance of hypothesis Pretest-Posttest pitch(Hz) t-Test

t-Test: Paired Two Sample for Means

	Pretest (Hz)	*Posttest (Hz)*
Mean	46.95	31.07
Variance	1437.07	609.36
Observations	30.00	30.00
Pearson Correlation	0.19	
Hypothesized Mean Difference	0.00	
df	29.00	
t Stat	2.12	
P(T<=t) one-tail	0.02	
t Critical one-tail	1.70	
P(T<=t) two-tail	0.04	
t Critical two-tail	2.05	



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Pretest	57.6	74.9	32.4	56	135	-15	29.7	36.8	72.4	33.5	-33	19.9	24.3	-36	57.2	48.8	85.5	99	22.7	69.4	33.3	79.6	60	24.6	59.2	60.3	-5.2	86.9	68.4	71.3
Posttest	29	-2.2	15.1	-4	22.3	11.4	13.5	21.5	53	25.4	64.7	37.7	11	34.1	32.5	23.9	24.7	39.2	6.26	69.6	22.4	52.2	67.7	50.1	2.76	8.26	-10	72.6	66.1	71.9

Figure 4.4: Pretest-posttest pitch differences compare with Standard

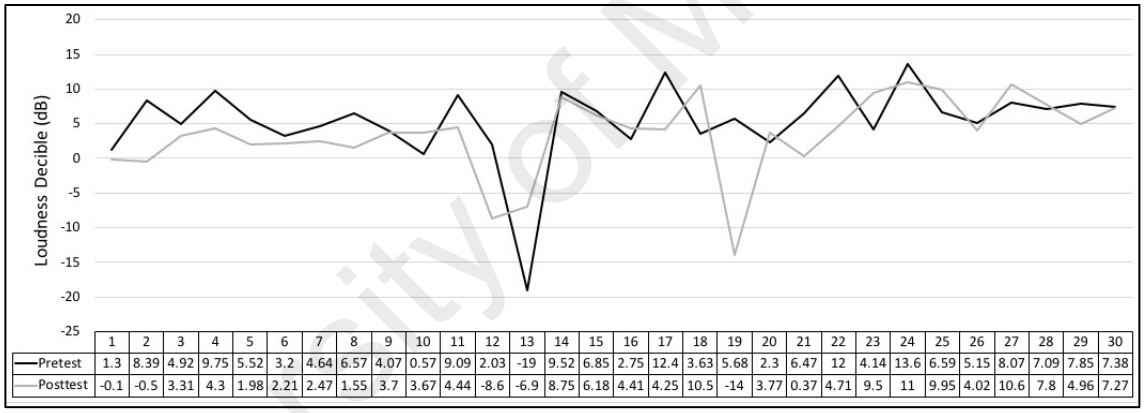
The second categories of data analysis are to see the t-Test result in the stress differences between "do' and "live". The measurement method is similar to the measurement method using for the experiment of the pitch, but this time it measures the amplitude in dB for the intensity or loudness of the audio. The experiment means score of the treatment result is 3.52dB in posttest, which is 34.93% lower than the mean score of the pretest at 5.41dB, before the intervention. With comparison to the treatment of kinetic typography visual aid the pitch measurement is also closer to the standard score. With the null hypothesis mean differences of 0, the computation t-Test calculation of P-Value is 0.04, and the t Statistic is 1.77, which is bigger than the t Critical one-tail at 1.70, it shows that the significant differences between these two tests are valid according to the computation calculation. The overall pretest group significantly has a lower performance in its accuracy in recognizing the stress (loudness) differences within "do and "live" as shown in Figure 4.5. It was also found that the participants have a higher variance score at 31.62dB in recognizing the English stress at pretest experiments compared with the posttest variance score of 31.54dB (Table 4.2). In comparison to the variance score in pitch, the variance score in the word stress posttest shown a minor improvement in the reduction of inaccuracy in English stress recognition.

Table 4.2: Significance of hypothesis Pretest-Posttest stress(dB) t-Test

Significant of hypothesis: Pretest-Posttest stress(dB) t-Test

t-Test: Paired Two Sample for Means

	Pretest(dB)	*Posttest(dB)*
Mean	5.41	3.52
Variance	31.62	31.54
Observations	30.00	30.00
Pearson Correlation	0.46	
Hypothesized Mean Difference	0.00	
df	29.00	
t Stat	1.77	
P(T<=t) one-tail	0.04	
t Critical one-tail	1.70	
P(T<=t) two-tail	0.09	
t Critical two-tail	2.05	



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Pretest	1.3	8.39	4.92	9.75	5.52	3.2	4.64	6.57	4.07	0.57	9.09	2.03	-19	9.52	6.85	2.75	12.4	3.63	5.68	2.3	6.47	12	4.14	13.6	6.59	5.15	8.07	7.09	7.85	7.38
Posttest	-0.1	-0.5	3.31	4.3	1.98	2.21	2.47	1.55	3.7	3.67	4.44	-8.6	-6.9	8.75	6.18	4.41	4.25	10.5	-14	3.77	0.37	4.71	9.5	11	9.95	4.02	10.6	7.8	4.96	7.27

Figure 4.5: Pretest-posttest stress differences compare with Standard

In comparing the significant differences of recognition in English intonation and word stress, out of 30 participants, only fie posttest results shown negative performance on a better recognition of intonation pitch (participant number 11, 20, 23, 27, 30). While the remaining 25 students showed significant positive results in the recognition of the intonation pitch. In term of the word stress recognition comparison, the impact of improvement of recognition in this aspect is not as good as recognition of pitch intonation. Out of 30 participants, there were 21 positive posttest results with a better performance of recognition of word stress, but the number of negative results was higher compared to

the pitch intonation recognition (Figure 4.5). A total of nine participants has shown negative result upon the treatments of kinetic typography visual learning aid (participant number 10, 12, 18,19, 20, 23, 25, 27, 28). Thus, with the results shown in the above mentioned tests, it can be concluded that the intonation pitch recognition in the visual representation of kinetic typography is more effective than the recognition of word stress, which is the loudness and intensity of spoken English. These findings further support the idea of physical characteristics of letterform, such as light or bold, round or square, short or long, wide or narrow, slim or thick can make their impression with its form and shape as asserted by Wong (1995).

Based on the results obtained from the computation t-Test in this experiment, the results indicate that with the intervention of kinetic typography visual aid, it could significantly reduce the confusion of auditory quality in the learning process of English prosody and able to improve participants' interpretation of English intonation and stress. In this study, kinetic typography visual learning aid was found to influence the effectiveness of learning English intonation and stress and this result corresponded to the finding of R. Rashid (2008) which asserted that animated text is a promising method in representing non-dialogue sound information.

There were no apparent threats to internal validity except for a potential test threat caused by the repetitive test effects in the experimental arrangements. This may have occurred due to the familiarization of the sentence and increase of comfort level in the experimentation process. As mentioned in the pilot test section, the memory threats of repetition have been minimized by the addition of in different phrases as a placebo in between the selected phrase "Where do you live?"

CHAPTER 5: CONCLUSION

Kinetic typography has an essential role as a communication tool in the world of digital media. Since its inception, little has been done to extend its usage in education and language learning. Thus, this research presents an alternative visual representation form in English prosody compared to the conventional method that currently exists and widely used at present. This study has investigated the process of developing a functional kinetic typography visual aid and carried out an experimentation to test the visual representation effectiveness on English intonation and stress among a group of non-native speakers of Malaysian college students. The correlation between the visual form of kinetic typography and English intonation and stress have been structurally presented. Moreover, this study has produced results which corroborate its findings with past studies in this field, in term of the functionality of kinetic typography and typography itself as a visual form for spoken English (Matthias Hillner, 2009; Rashid, 2008; Wong, 1995), and it has also provided a substantial amount of data to enable further research in the development of kinetic typography as a visual learning aid.

According to the pre-posttest of stress and pitch result in Chapter 4, Section 4.2, participants have higher variance score at 31.62dB in recognizing the English stress at pretest experiments compared to the posttest variance score of 31.54dB. In comparison of the variance score in pitch, the variance score in the word stress posttest shown a minor improvement in the reduction of inaccuracy in English stress recognition. This means that the impact of improvement of recognition of word stress is not as impactful as the recognition of pitch intonation. Thus, based on the stress and pitch results, it can be concluded that the cognitive result of pitch learning is better than the recognition of amplitude.

Even though there was some minor negative impact on the recognition of both pitch intonation and amplitude in word stress, the computational calculation on the significance of differences on the t-Test has been established as a valid finding. However, many of the statistic and suggestions presented in this research are not a complete solution to the questions considered, as iterated by Hillner (2009), "the viewer's reconstruction of meaningful content is a complex process" (p.55). Rather, they are to be recognized as carefully formulated hypotheses, which need to be further investigated by other researchers.

As an extension of this thesis research, a more specific and precise visual mapping of intonation and word stress is definitely an area to be investigated further, especially to improve on the word stress visual representation.

Due to the limitation of the scope of this research, the experimentation only allowed one sentence as subject matter for the experiment while the other two words and one sentence was placed in as a placebo in this research to minimize the memory threats during the experiment. In order to establish a better scale of results on the effectiveness of kinetic typography visual representation, more sentences should be included in the test. Furthermore, the number of participants could be done on a larger scale and various ethnicity background of the participants could be consider in future researches to measure and determine if the independent variable of ethnicity have a correlation with the result on recognition of the reproduction of the intonation and word stress in English. Other suprasegmentals of English prosody like attitudinal function which was not included in this study could be further investigated as well.

For the aspect of language extension in this research, it is hoped that other researchers could find the most effective combinations based on the current visual-auditory mapping system and test them in other languages, especially in tonal spoken language like

Mandarin. Xing (2006) claims that Mandarin tones have a significantly wider pitch range than a non-tonal language such an English whereby there was some common confusion which arised between the second and third tone, and the first and fourth tone. Thus, studies in such aveneues can be further investigated to enrich the study of spoken languages.

The results of this research on kinetic typography as a visual learning aid are encouraging and it strongly suggests that kinetic typography visual approach is a considerable way to represent the intonation and word stress in English prosody. It also asserts that kinetic typography may contribute and provide better progression in the learning of spoken English. The researcher of this study strongly hopes that the results acquired from his study would provide valuable knowledge and insights to the creative industry as well as the academic community, with regards to the application of kinetic typography as a visual learning aid.

REFERENCES / BIBLIOGRAPHY

Allan, K. (2013). *The Oxford Handbook of the History of Linguistics*. Oxford: Oxford University Press.

Allstetter, W. D. (1991). *Speech and Hearing*. NY: Chelsea House Publishers.

Ambrose, G., & Harris, P. (2011). *The Fundamentals of Typography*. *Lausanne: AVA Publishing*. Lausanne: AVA Publishing SA.

Armstrong, L. E., & Ward, I. C. (1967). *Handbook of English Intonation*. W. Heffer.

Arvaniti, A. (2011). The Representation of Intonation. In *The Blackwell Companion to Phonology, 5 Volume Set*. WSX: John Wiley & Sons.

Atoye, R. (2005). Non-native perception and interpretation of English intonation. *Nordic Journal of African Studies*.

Austin, J. L., & Sbisà, M. (1975). *How to Do Things with Words*. Harvard University Press.

Bachfischer, G., & Robertson, T. (2005). From movable type to moving type - evolution in technological mediated typography (pp. 1–13). Hobart: Apple University Consortium.

Bartram, A. (2005). *Futurist Typography and the Liberated Text*. Yale University Press.

Beatty, K. (2013). *Teaching & Researching: Computer-Assisted Language Learning*. Taylor & Francis.

Benson, P. J. (1985). Writing visually: Design considerations in technical publications. *Technical Communication*, 35–39.

Bigelow, C., & Day, D. (1983). Digital Typography. *Scientific American*, *249*(2), 106–119.

Bilak, P., & Bailey, S. (2004). *Dot Dot Dot 8*. Princeton Architectural Press.

Bloomfield, L. (1933). *Language*. University of Chicago Press.

Bodine, K., & Pignol, M. (2003). Kinetic typography-based instant messaging. *CHI 03 Extended Abstracts on Human Factors in Computing Systems CHI 03*, 914. https://doi.org/10.1145/765891.766067

Bolinger, D. (1958). A Theory of Pitch Accent in English. *WORD*, *14*(2–3), 109–149.

Bolinger, D., & Le Merton Bolinger, D. (1986). *Intonation and Its Parts: Melody in Spoken English*. Stanford University Press.

Booth-Clibborn, E., & Baroni, D. (1980). *The Language of Graphics*. Thames and Hudson.

Boyarski, D. (n.d.). Kinetic Typography : a new direction for communication design. Retrieved September 20, 2017, from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.113.675&rep=rep1&type=pdf

Brazil, D. (1997a). *The communicative value of intonation in English*. Cambridge: Cambridge Univ. Press.

Brazil, D. (1997b). *The Communicative Value of Intonation in English Book*. Cambridge University Press.

Bringhurst, R. (2004). *The elements of typographic style*. Hartley & Marks, Publishers.

Brosnahan, L. F., Brosnahan, L. F., & Malmberg, B. (1976). *Introduction to Phonetics*. Cambridge University Press.

Brownie, B. (2015). *Transforming Type: New Directions in Kinetic Typography*. Bloomsbury Publishing.

Buehner, T. M. (2009). *A Q-Study of Display Type ' S Visual Impact in College Textbooks*. South Dakota State University.

Carter, R., Meggs, P. B., Day, B., Maxa, S., & Sanders, M. (2014). *Typographic Design: Form and Communication*. Wiley.

Cash, P., Stanković, T., & Štorga, M. (2016). *Experimental Design Research: Approaches, Perspectives, Applications*. Springer International Publishing.

Chapelle, C. A., & Sauro, S. (2017). *The Handbook of Technology and Second Language Teaching and Learning*. Wiley.

Cheung, H. Y. P. (2011). *Representing Sight and Sound in Design Media: A Cyclical Time-based Model*. Arizona State University.

Chun, D. M. (2002). *Discourse Intonation in L2: From Theory and Research to Practice*. Amsterdam: J. Benjamins.

Coltheart, M., Sartori, G., & Job, R. (1987). *The Cognitive Neuropsychology of Language*. Erlbaum.

Cook-Gumperz, J., Corsaro, W. A., & Streeck, J. (1986). *Children's Worlds and Children's Language*. Walter de Gruyte.

Cooper, W. E., & Sorensen, J. M. (2012). *Fundamental Frequency in Sentence Production*. Springer New York.

Couper-Kuhlen, E. (1986). *An Introduction to English Prosody*. Edward Arnold.

Cranny-Francis, A. (2005). *Multimedia : Texts and Contexts*. London, GBR: SAGE Publications Inc. (US).

Cruttenden, A. (1997). *Intonation*. Cambridge University Press.

Crystal, D. (1969). *Prosodic Systems and Intonation in English*. Cambridge University Press.

Crystal, D. (1976). *Prosodic Systems and Intonation in English*. Cambridge University Press.

Crystal, D. (2006). *Language and the Internet*. Cambridge University Press.

Crystal, D., & Quirk, R. (1964). *Systems of prosodic and paralinguistic features in English*. Mouton.

Cullen, K. (2012). *Design Elements, Typography Fundamentals: A Graphic Style Manual for Understanding How Typography Affects Design*. Rockport Publishers.

Cutler, A., & Ladd, D. R. (2013). *Prosody: Models and Measurements*. Springer Berlin Heidelberg.

Dachkovsky, S., & Sandler, W. (2009). Visual Intonation in the Prosody of a Sign Language. *Language and Speech*, *52*(2–3), 287–314. https://doi.org/10.1177/0023830909103175

Daniels, P. T., & Bright, W. (1996). *The World's Writing Systems*. Oxford University Press.

de Saussure, F., & Harris, R. (2013). *Course in general linguistics*. Bloomsbury Publishing.

Derwing, T. M., Thomson, R. I., & Munro, M. J. (2006). English pronunciation and fluency development in Mandarin and Slavic speakers. *System*, *34*(2), 183–193.

Dhaif, H. A. (1989). Can computers teach languages. *English Teaching Forum*, *27*(3), 17–19.

DiPietro, B. (1999). *DO YOU SEE WHAT I'M SAYING. Read*. York University, TO.

Dr. Will Styler. (2017). Praat: doing Phonetics by Computer. (2017). Retrieved January 22, 2017, from http://www.fon.hum.uva.nl/praat/

Dziubalska-Kołaczyk, K., & Przedlacka, J. (2008). *English Pronunciation Models: A Changing Scene*. Peter Lang.

Ernest-Moriarty, S. B., & Ernst, S. B. (1984). *The ABCs of Typography: A Practical Guide to the Art & Science of Typography*. Art Direction Book Co.

Fellows, K. S. (2009). *Typecast: the voice of typography*. University of Iowa.

Finegan, E. (2007). *Language: Its Structure and Use*. Cengage Learning.

Ford, S., Forlizzi, J., & Ishizaki, S. (1997). Kinetic typography: issues in time-based presentation of text. In *CHI'97 extended abstracts on Human factors in computing systems* (pp. 269–270). https://doi.org/10.1145/1120212.1120387

Forlizzi, J., Lee, J., & Hudson, S. (2003). The kinedit system: affective messages using dynamic texts. In *Proceedings of the conference on Human factors in computing*. https://doi.org/10.1145/642611.642677

Fox, A. (2002). *Prosodic Features and Prosodic Structure: The Phonology of "Suprasegmentals."* OUP Oxford.

Gilakjani, A. P., & Sabouri, N. B. (2016). How Can EFL Teachers Help EFL Learners Improve Their English Pronunciation? *Journal of Language Teaching and Research*, *7*(5), 967–972. https://doi.org/10.17507/jltr.0705.18

Goodall, E. (2014). Impacts of Technological Change Upon Narrative. *Text: Journal of Writing and Writing Courses*, *18*(1).

Gündüz, N. (2005). Computer Assisted Language Learning. *Journal of Language and Linguistic Studies*, *1*(2), 193–214. https://doi.org/10.17263/jlls.24391

Guo, H. (2014). *Analysing and evaluating current mobile applications for learning English speaking*. University of London.

Halliday, M. A. K. (1967). *Intonation and grammar in British English*. De Gruyter.

Harkins, M. (2011). *Basics Typography 02: Using Type*. Lausanne: AVA Publishing SA.

Hart, J. T., Collier, R., & Cohen, A. (2006). *A Perceptual Study of Intonation: An Experimental-Phonetic Approach to Speech Melody*. Cambridge University Press.

Haskin, P., & Bass, S. (1996). "Saul, Can You Make Me a Title?": Interview with Saul Bass. *Source: Film Quarterly*, *50*(1), 10–17.

Heller, S. (2004). *Education of a typographer*. NY: Allworth Press.

Higgins, J. (1988). *Language, learners, and computers: human intelligence and artificial unintelligence*. Longman.

Hillner, M. (2006). "Virtual Typography": Time Perception in Relation to Digital Communication. *Leonardo Electronic Almanac*, *14*, 1–6.

Hillner, M. (2007). *The poetics of transition: the ambiguous characteristics of virtual typography*. *Royal College of Art*. Royal College of Art.

Hillner, M. (2009). *Virtual Typography*. Lausanne: AVA Publishing.

Hirst, D., & Cristo, A. Di. (1998). *Intonation in British English*. *Intonation Systems. A Survey of Twenty Languages*. NY: Cambridge University Press.

Horak, J. C. (2014). *Saul Bass: Anatomy of Film Design*. University Press of Kentucky.

Hostetler, S. C. (2006). *Integrating Typography and Motion in Visual Communication*. *Integrating Typography and Motion in Visual Communication*. University of Northern Iowa, IA.

Jahandarie, K. (1999). *Spoken and Written Discourse: A Multi-disciplinary Perspective*. Ablex Pub.

James Ohene-Djan, J. W. K. C.-S. (2007). *Emotional subtitles: A system and potential applications for deaf and hearing impaired people*. *Assistive Technology for All Ages*. University of London.

Jones, D. (1962). *An Outline of English Phonetics: With 116 Illustrations and with Appendices*. W. Heffer & sons.

Jukes, I., McCain, T., & Crockett, L. (2010). *Understanding the Digital Generation: Teaching and Learning in the New Digital Landscape*. SAGE Publications.

Kang, O., Rubin, D. O. N., Pickering, L., & Rang, O. (2014). Measures suprasegmental and judgments of accentedness of language learner proficiency in oral English. *The Modern Language Journal*, *94*(4), 554–566. https://doi.org/10.1111/j.1540-4781.2010.01091.x

Keane, L., & Fam, K.-S. (2006). Media: Through the eyes of young adults. *Journal of Promotion Management*, *11*(2–3), 155–174.

Kelly, G. (2000). *How To Teach Pronunciation*. (J. Harmer, Ed.). Essex: Bluestone Press.

Kenney, K. (2010). *Visual Communication Research Designs*. Taylor & Francis.

Kostelnick, C. (1990). The rhetoric of text design in professional communication. *Technical Writing Teacher*, *17*(3), 189–202.

Krause, J. (2006). *Type Idea Index: The Designer's Ultimate Tool for Choosing and Using Fonts Creatively*. F+W Media.

Kumar, R. (2011). *Research Methodology*. London: SAGE Publications.

Lee, & et al. (2002). The Kinetic Typography Engine: An Extensible System for Animating Expressive Text. In *UIST '02 Proceedings of the 15th annual ACM symposium on User interface software and technology*. PA: Carnegie Mellon University. https://doi.org/10.1145/571985.571997

Lee, J., Jun, S., Forlizzi, J., & Hudson, S. E. (2006). Using kinetic typography to convey emotion in text-based interpersonal communication. *Proceedings of the 6th ACM Conference on Designing Interactive Systems - DIS '06*. NY: ACM Press. https://doi.org/10.1145/1142405.1142414

Lee, S. T. (2008). Teaching Pronunciation of English Using Computer Assisted Learning Software : An Action Research Study in an Institute of Technology in Taiwan Submitted by. *English Language and Literature Studies*, *3*(1), 219. https://doi.org/10.5539/ells.v3n1p111

Leeuwen, T. Van. (2006). Towards a semiotics of typography. *Information Design Journal*, *14*(2), 139–155.

Leonardo, S., & Drucker, J. (2008). Letterpress Language: Typography as a Medium for the Visual Representation of Language. *Leonardo*, *41*(1), 66–74.

Levy, M., Blin, F., Siskin, C. B., & Takeuchi, O. (2011). WorldCALL: International Perspectives on Computer-Assisted Language Learning.

Lister, M. (2009). *New Media: A Critical Introduction*. Routledge.

Loxley, S. (2006). *Type: The Secret History of Letters*. I. B. Tauris.

Lupton, E. (2010). *Thinking with Type: A Critical Guide for Designers, Writers, Editors, & Students*. NY: Prinsceton Architectural Press.

Malamed, C. (2015). *The Learning Designer's Visual Design Book: How to Design Instruction Like a Pro*. NJ: John Wiley & Sons.

Marcus, A. (1987). Diagrammatic Visible Language: An Investigation of Visual Logic. *Leonardo*, *20*(1), 9–15. https://doi.org/10.2307/1578204

Marcus, A., Hanseman, D. J., Gustafson, E. F., & Marcus, A. (1987). Diagrammatic Visible Language: An Investigation of Visual Logic. *Leonardo*, *20*(1), 9–15. https://doi.org/10.2307/1578204

Meagher-Alkan, C. (2010). *Teaching Pronunciation Using the Prosody Pyramid*. Cambridge University Press.

Meggs, P. B., & Purvis, A. W. (2016). *Meggs' History of Graphic Design*. NJ: John Wiley & Sons.

Mori, J., & Fels, D. I. (2009). Seeing the music can animated lyrics provide access to the emotional content in music for people who are deaf or hard of hearing? In *2009 IEEE Toronto International Conference Science and Technology* (pp. 951–956). TO.

Nespor, M., & Sandler, W. (1999). Prosody in Israeli sign language. *Language and Speech*, *42*(2–3), 143–176.

Nunan, D., & Carter, R. (2001). *The Cambridge Guide to Teaching English to Speakers of Other Languages*. Cambridge University Press.

O'Connor, J. D., Arnold, G. F., & Arnold, G. F. (1973). *Intonation of colloquial English: a practical handbook*. Longman.

Padden, C. (1990). The relation between space and grammar in ASL verb morphology. *Sign Language Research: Theoretical Issues*, 118–132.

Palacio, B. G., & Vit, A. (2011). *Graphic Design, Referenced: A Visual Guide to the Language, Applications, and History of Graphic Design*. Rockport Publishers.

Parker, R. C., & Berry, P. (1998). *Looking Good in Print*. Coriolis Group Books.

Patel, A. D. (2006). An Empirical Method for Comparing Pitch Patterns in Spoken and Musical Melodies: A Comment on J.G.S. Pearls "Eavesdropping with a Master: Leos Janáček and the Music of Speech." *Empirical Musicology Review*, *1*(3), 166–169.

Patel, A. D. (2007). *Music, Language, and the Brain*. Oxford University Press.

Peretz, I., & Zatorre, R. J. (2003). *The Cognitive Neuroscience of Music*. OUP Oxford.

Pierrehumbert, J. (1981). Synthesizing intonation. *The Journal of the Acoustical Society of America*, *70*(4).

Pike, K. L. (1979). *The Intonation of American English*. Greenwood Publishing Group.

Podesva, R. J., Podesva, R., & Sharma, D. (2014). *Research Methods in Linguistics*. Cambridge University Press.

Ponce, D. A. (2012). *Type Set: Exploring the Effects of Making Kinetic Typography Interactive*. BiblioBazaar.

Rabinowitz, T. (2015). *Exploring Typography*. Cengage Learning.

Rashid, R. (2008). *Representing Emotions with Animated Text*. University of Toronto, TO.

Rebuschat, P. (2012). *Language and Music as Cognitive Systems*. OUP Oxford.

Reilly, J. S., McIntire, M., & Bellugi, U. (1990). The acquisition of conditionals in American Sign Language: Grammaticized facial expressions. *Applied Psycholinguistics*, *11*(04), 369–392.

Roach, P. (2009). *English phonetics and phonology: A practical course*. Cambridge University Press.

Rosenberger, T. (1999). Prosodic Font : Translating speech into graphics. *CHI 99 Extended Abstracts on Human Factors in Computing Systems*, (May), 252–253. https://doi.org/10.1145/632716.632872

Rude, M. (2012). Prosodic Writing with 2D- and 3D-fonts : An approach to integrate pronunciation in writing systems, 233–252.

Salkind, N. (2010). *Encyclopedia of Research Design*. CA: SAGE. https://doi.org/10.4135/9781412961288

Samara, T. (2004). *Typography Workbook: A Real-World Guide to Using Type in Graphic Design*. Rockport Publishers.

Samara, T. (2007). *Design Elements: A Graphic Style Manual*. Rockport Publishers.

Sarkar, N. N. (2008). *Art and Print Production*. Oxford University Press.

Sassoon, R. (1993). *Computers and Typography*. Intellect Books.

Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M. M., Wightm, Wightman, C., … Hirschberg, J. (1992). ToBI: A standard for labeling English prosody. *Second International Conference on Spoken Language Processing*, (October), 867–870. https://doi.org/10.1515/cllt-2012-0011

Son, J. B. (2014). *Computer-Assisted Language Learning: Learners, Teachers and Tools*. Cambridge Scholars Publisher.

Spencer, H. (2004). *Pioneers of Modern Typography*. MIT Press.

Stake, L.-A. F. (2015). *Temporal typography*. UMT.

Stanley, G. (2013). *Integrating technology into secondary English language teaching*. *Innovations in learning technologies for English language teaching*.

Staples, L. (2000). Typography and the Screen: A Technical Chronology of Digital Typography, 1984–1997. *Design Issues*, *16*(3), 19–34. https://doi.org/10.1162/07479360052053306

Steele, J. (1779). *Prosodia Rationalis: Or, An Essay Towards Establishing the Melody and Measure of Speech, to be Expressed and Perpetuated by Peculiar Symbols*. J. Nichols.

Stevenson, A. (2010). *Oxford Dictionary of English*. OUP Oxford.

Stilton, G. (2011). *Geronimo Stilton and the Kingdom of Fantasy #1: The Kingdom of Fantasy*. Scholastic Incorporated.

Stockwell, G. (2012). *Computer-Assisted Language Learning: Diversity in Research and Practice*. Cambridge University Press.

Strizver, I. (2010). *Type Rules!: The Designer's Guide to Professional Typography*. Wiley.

Thaut, M. (2013). *Rhythm, Music, and the Brain: Scientific Foundations and Clinical Applications*. Taylor & Francis.

Thomas, M., Reinders, H., & Warschauer, M. (2012). *Contemporary Computer-Assisted Language Learning*. Bloomsbury Publishing.

Thyer, B. A. (2012). *Quasi-Experimental Research Designs*. Oxford University Press, USA.

Trager, G. L., & Smith, H. L. (2009). *An Outline of English structure*. CA: Рипол Классик.

Turgut, O. P. (2012). Kinetic typography in movie title sequences. *Procedia-Social and Behavioral Sciences*, *51*, 583–588.

Uekita, Y., Sakamoto, J., & Furukata, M. (2000). The method of kinetic typography communication. *Smc 2000 Conference Proceedings. 2000 Ieee International Conference on Systems, Man and Cybernetics. "cybernetics Evolving to Systems, Humans, Organizations, and Their Complex Interactions" (Cat. No.0*, *1*. https://doi.org/10.1109/ICSMC.2000.885030

Uldall, E. (1964). Dimensions of meaning in intonation. London: Longman.

Underhill, A. (2005). *Sound foundations : learning and teaching pronunciation*. OX: Macmillan.

Vartanian, I. (2003). *Typo Graphics: The Art and Science of Type Design in Context*. RotoVision.

Veletsianos, G. (2016). *Emergence and Innovation in Digital Learning: Foundations and Applications*. Athabasca University Press.

Vy, Q. V, Mori, J. A., Fourney, D. W., & Fels, D. I. (2008). EnACT: A software tool for creating animated text captions. In *International Conference on Computers for Handicapped Persons* (pp. 609–616).

Waller, R. (1987). *The typographic contribution to language*. University of Reading.

Wennerstrom, A. (2001). *The Music of Everyday Speech: Prosody and Discourse Analysis*. Oxford University Press.

White, J. V. (2003). *Editing by Design: For Designers, Art Directors, and Editors : the Classic Guide to Winning Readers*. Allworth Press.

Willems, N. (1982). *English Intonation from a Dutch Point of View*. De Gruyter.

Williams, R. (2012). *The Animator's Survival Kit: A Manual of Methods, Principles and Formulas for Classical, Computer, Games, Stop Motion and Internet Animators*. Farrar, Straus and Giroux.

Wong, Y. Y. (1995). *Temporal typography : characterization of time-varying typographic forms*. Massachusetts Institute of Technology.

Woods, C., Emberling, G., Teeter, E., & of Chicago. Oriental Institute, U. (2010). *Visible Language: Inventions of Writing in the Ancient Middle East and Beyond*. Oriental Institute of the University of Chicago.

Woolman, M., & Bellantoni, J. (2000). *Moving Type: Designing for Time and Space*. RotoVision SA.

Worthington, M. (1999). Entranced by Motion, Seduced by Stillness. *Eye Magazine*, *9*, 28–39.

Xing, J. Z. (2006). *Teaching and Learning Chinese as a Foreign Language: A Pedagogical Grammar*. Hong Kong University Press.

Zhuang, Y. (2015). EVALUATING INTONATION TEACHING Evaluating the Effectiveness of Teaching Intonation to Learners in an Intensive English Program. *EVALUATING INTONATION TEACHING*, (May).

Zou, B. (2012). *Explorations of Language Teaching and Learning with Computational Assistance*. PA: Information Science Reference.

Zou, B. (2013). *Computer-Assisted Foreign Language Teaching and Learning: Technological Advances: Technological Advances*. Hershey, PA: Information Science Reference.

Zsiga, E. C. (2012). *The Sounds of Language: An Introduction to Phonetics and Phonology*. Wiley.