THE EFFECTS OF SONIFICATION ON INTERPRETATION OF GEO-VISUALIZATION OF COMPLEX DENGUE DATA

LILIAN LEE YEN WEI

FACULTY OF COMPUTING AND INFORMATICS
UNIVERSITI MALAYSIA SABAH
2018

THE EFFECTS OF SONIFICATION ON INTERPRETATION OF GEO-VISUALIZATION OF COMPLEX DENGUE DATA

LILIAN LEE YEN WEI

THESIS SUBMITTED IN FULFILLMENT FOR THE DEGREE OF MASTER OF SCIENCE

FACULTY OF COMPUTING AND INFORMATICS
UNIVERSITI MALAYSIA SABAH
2018

PUMS 99:1

UNIVERSITI MALAYSIA SABAH

BORANG PENGESAHAN TESIS

JUDUL : _____
_____
_____

IJAZAH : _____
_____

SAYA : _____ **SESI PENGAJIAN** : _____
(HURUF BESAR)

Mengaku membenarkan tesis *(LPSM/Sarjana/Doktor Falsafah) ini disimpan di Perpustakaan Universiti Malaysia Sabah dengan syarat-syarat kegunaan seperti berikut:-

1. Tesis adalah hak milik Universiti Malaysia Sabah.
2. Perpustakaan Universiti Malaysia Sabah dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan dibenarkan membuat salinan tesis ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. Sila tandakan (/)

☐ SULIT (Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di AKTA RAHSIA RASMI 1972)

☐ TERHAD (Mengandungi maklumat TERHAD yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan)

☐ TIDAK TERHAD

Disahkan oleh:

_____ _____
(TANDATANGAN PENULIS) (TANDATANGAN PUSTAKAWAN)

Alamat Tetap: _____
_____
_____
_____

_____
(NAMA PENYELIA)

TARIKH: _____ TARIKH: _____

Catatan:
*Potong yang tidak berkenaan.
*Jika tesis ini SULIT dan TERHAD, sila lampirkan surat daripada pihak berkuasa/organisasi berkenaan dengan menyatakan sekali sebab dan tempoh tesis ini perlu dikelaskan sebagai SULIT dan TERHAD.
*Tesis dimaksudkan sebagai tesis bagi Ijazah Doktor Falsafah dan Sarjana Secara Penyelidikan atau disertai bagi pengajian secara kerja kursus dan Laporan Projek Sarjana Muda (LPSM).

DECLARATION

I hereby declare that the material in this thesis is my own except for quotations, equations, summaries, and references, which have been duly acknowledged. Furthermore, I took reasonable care to ensure that the work is original, and, to the best of my knowledge, does not breach copyright law, and has not been taken from other sources except where such work has been cited and acknowledged within the text.

2nd Aug 2018

_____

Lilian Lee Yen Wei

MI 1511001 T

CERTIFICATION

NAME	**:**	**LILIAN LEE YEN WEI**
MATRIC NO.	**:**	**MI 1511001 T**
TITLE	**:**	**THE EFFECTS OF SONIFICATION ON INTERPRETATION OF GEO-VISUALIZATION OF COMPLEX DENGUE DATA**
MASTER	**:**	**MASTER OF SCIENCE (INFOMATION TECHNOLOGY)**
DATE OF VIVA	**:**	**8TH FEBRUARY 2018**

CERTIFIED BY,

SUPERVISOR **SIGNATURE**

Assoc. Prof Dr. Ag Asri Hj Ag. Ibrahim _____

ACKNOWLEDGEMENT

Foremost, I thank God for letting me through all the trials and difficulties. I have experienced your guidance and love day by day. Without you I would have not able to finish my Master Degree. Thank you, Lord.

I would like to express my sincere gratitude to my advisor Prof. Madya Dr. Ag Asri Ag Ibrahim, you have been a great mentor to me. I would like to thank you for your continuous support and motivation in encouraging my Master research study. Your advice and enthusiasm have inspired me all the time of research and writing of this thesis. I would also like to thank my committee members, Prof. Madya Dr.Jumat Bin Sulaiman, Dr. Rosnafisah Sulaiman, Prof. Madya Dr. Rayner Alfred, Dr. Lau Hui Keng, Dr. Norazlina Binti Khamis and Puan Winnie Binti Edward Alau. I also want to thank you for letting my thesis defence become a pleasant moment, and for your constructive yet brilliant comments and suggestions, thanks to you.

My sincere thanks also go the doctors, nurses, environmental Health Officer and to all staff of the Ministry of Health Labuan F. T. and Ministry of Health Sabah. All of you have been there to support me when I did my task assessment and data collection for my Master Thesis.

A special thanks to my family. Words cannot express how grateful I am to my mother, father, sisters, and brother for all of the sacrifices that you've made on my behalf. I would also like to thank my beloved husband, Ray Dhirendra Alasa. Thank you for Compassion love and support. To my beloved son Michael Ryan Ray Alasa, my apology if I have not given enough attention on you, thanks for being such a good boy always cheering me up.

Lilian Lee Yen Wei

2nd Aug 2018

ABSTRACT

The effective way for analysing complex data sets, it's the ability to represent the overall pattern of data in a single presentation by providing alternative input options. While vast of work for data mapping focusing on visual presentations, the work of this thesis suggest that data sonification can be better aids towards complex data interpretation especially when comes to complement visual representation. In this project, an ambient sound framework was designed so it can be applied to display complex dengue data. Parameter mapping method was applied to filter data with temporal factor and represent it with sound parameters. Four sound parameters were used: onset, volume, pitch and tempo. A simulation prototype was developed using open source JavaScript library: leaflet map, leaflet timeline and audiosynth.js. Data collected from experiment through task assessment were analysed using PASW. Results from the tested framework shows that there is positive significance statistic difference in the level of data interpretation when sound was used for complex dengue data representation.

ABSTRAK

KESAN SONIFIKASI TERHADAP INTERPRETASI DATA DENGGI YANG KOMPLEKS DALAM GEO-VISUALISASI

Cara yang berkesan untuk menganalisis set data yang rumit adalah dengan menyediakan alternatif bagi input pilihan untuk mentafsir keseluruhan corak data dalam persembahan tunggal. Kebanyakkan kerja penyelidikan dalam pemetaan data hanya tertumpudalam persembahan visual, kerja-kerja tesis ini mencadangkan bahawa sonifikasi data boleh menjadi altrenatif yang baik ke arah penafsiran data yang rumit terutama dalam membantu serta melengkapi kaedah persembahan secara visual. Dalam projek ini, satu model kerja menggunakan bunyi ambien direka supaya dapat digunakan untuk membantu kaedah visual memaparkan data denggi yang kompleks. Kaedah pemetaan parameter digunakan untuk menapis data jenis temporal dan mewakilkannya dengan parameter bunyi. Empat parameter bunyi digunakan: permulaan, kelantangan, padang dan tempo. Prototaip simulasi telah dibangunkan menggunakan sumber terbuka JavaScript library: leaflet map library dan audiosynth.js. Data yang dikumpul dari eksperimen melalui penilaian tugasan dianalisa menggunakan PASW. Keputusan dari model kerja yang diuji menunjukkan terdapat perbezaan statistik positif dalam tahap tafsiran data apabila bunyi digunakan untuk perwakilan data denggi yang kompleks.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF PHOTOS

LIST OF ABBREVIATIONS

DF	Dengue Fever
DHF	Dengue Haemorrhagic Fever
GIS	Geographic Information System
CDC	Centre of Disease Control
GVD	Geographical Visual Display
HCI	Human Computer Interaction
2-D	Two Dimensional
3-D	Three Dimensional
EDA	Exploratory Data Analysis
SMR	Standard Mortality ratio
eDengue	Dengue Surveillance and Information System
eNotifikasi	Outbreak Management and Notification System
SPWD	Dengue Outbreak Management System
SMS	Short messaging System
MCDCD	Multiple Component Dynamic Cartography Display
D3	Data-driven Document
ESTAT	Exploratory Spatial-Temporal Analysis Toolkit
STIS	Space-time Information System
API	Application Programming Interface
TC	Traditional Cartographic
G	Geo-visualization
GM	General Map
PGC	Prior Geo-collaboration
GC	Geo-collaboration
MU	Mash-ups
HMU	Hybrid Mash-ups
TM	Thematic Map
CT	Clustering
CL	Colouring

SB	Symbolization
HL	Highlighting
HSB	Hybrid Symbolization
HLA	Highlighting and Animation
SDLC	Software Development Life Cycle
MO	Medical Officer
EHO	Environmental Health Officer
AEHO	Assistant of Environmental Health Officer
EO	Entomologist Officer
IT	Information Technology
ADMIN	Administration
4V	Four visual elements
4V2S	Four visual elements and two sound elements
2V2S	Two visual elements and two sound elements
4S	Four Sound elements
T1	Task One
T2	Task Two
T3	Task Three
T4	Task Four
T5	Task Five
ANOVA	Analysis of Variance
SOM	Self-Organizing Map

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Overview

Over the years, there are increasing numbers of dengue cases reported every day. In 2015, the number of cases reported in Malaysia was 120,836 cases with the number of death was 336 compared to year 2013 where only 43,346 dengue cases reported with only 92 death cases. In 2016, dengue outbreak stills a main concern to the society where 100,357 cases reported with 237 death cases (Idengue Remote Sensing, 1997-2016). Dengue fever (DF) and the potentially fatal dengue haemorrhagic fever (DHF) have been identified as main contributed to the number of death in Malaysia it has been an epidemic in Malaysia for long decade (Ghee, 1993).Malaysia was reported to have higher case fatality rates (4.67%) compared to neighbour countries like Thailand and Indonesia, with the fatality rates of 0.3% and 0.5% respectively (Nor Azura Husin, 2006). There are different types of Dengue cases which can be divided into five sero-substype: Den-1, Den-2, Den-3, Den-4 and the latest type of Dengue is named as Den-5 which categorized as unknown. Den-2 is the most common types as it appeared in major vector of dengue known as *aedesaegypti*. There are two type of dengue vector: *aedesaegypti* and *aedesalbopictus*. When a single case reported, it takes seven days' interval for it to become an outbreak when there are second case reported within 200 cubic metres at a local compound. Furthermore, when the third cases reported at the same locality within seven days from the second case, the category of cases will change into an uncontrolled outbreak. Worst happened when 4th cases were reported at the same locality within seven days from the 3rd cases, then the area will be declared as dengue hotspot. The numbers of cases will soar when rainy seasons start as it increases mosquito breeding grounds which lead to the need of full coverage from Dengue outbreak surveillance.

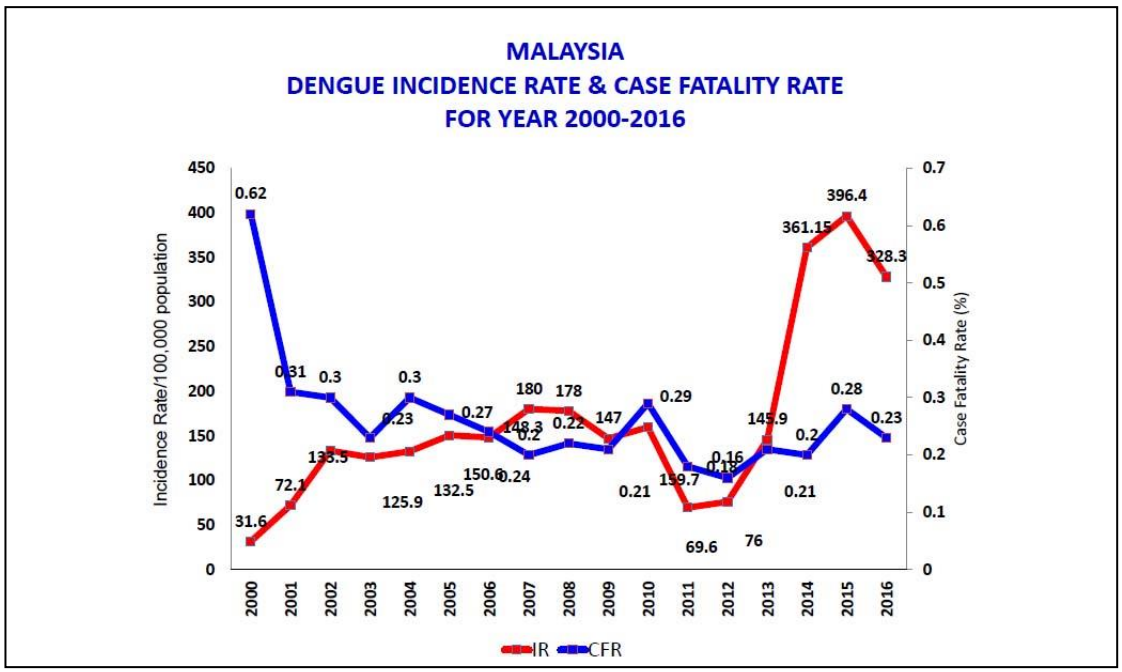

Figure 1.1 : Dengue Statistic

Source : (MOH M. &., 1997-2016)

As the volume and complexity of infectious disease data increases, public health professionals must synthesize highly disparate data to facilitate communication with the public and inform decisions regarding measures to protect the public's health (Carroll, 2014). Geo-visualization has seen much interest nowadays in not only amongst computer science community, but also in areas where practical solutions are required through significant representation applications. However, Geo-visualization does not limit only to visual representation but also finding the possibility alternative approach that may assist towards multivariate data representation. Sonic representation approaches has arisen from many visualization tools in multiple domain– including voice over and redundant aural cues on cartographic communication (MacEachren A. M., 1991) (McGee, 2009), auditory graph (Bertin, 1981), audio cartography (McGranaghan, 1987), sound variables mapping (Krygier J. B., 1994), sound extension for data layers (Bearman, 2012), and synthesizing results in Geo-visualization (Robinson A. C., 2004) for instance. This research has resulted in the development of several geo-visualization tools that have seen as the practical success.

This practical visualization supports various principles such as Metaphor and Functionality (Kraak M.-J. , 2003) (Slocum T. A., 2001), Support Data Exploration (Zhang, 2010), Multivariate and Big Data Handling [14], Web 2.0 and Geo-collaboration (Jern, 2006) (Roth, 2008).Extended design of these Geo-visualization tools has been applied and cited in many related researchers. With the target to provide the future researcher a better guideline on complex data representation designs.

1.2 Problem Statement

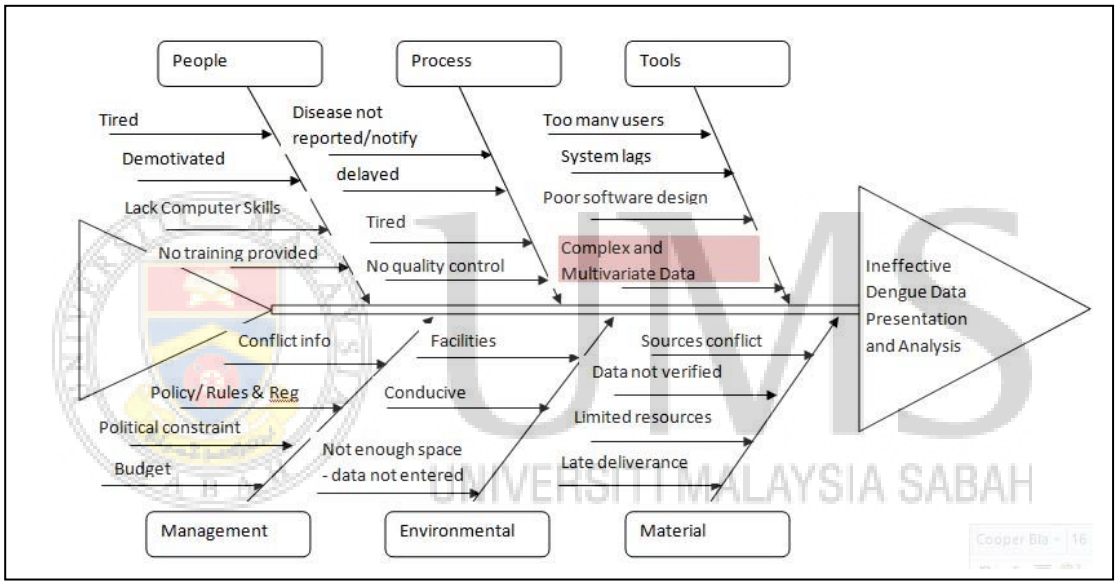

Figure 1.2 : Ishikawa Diagram

Source : Adapted from Kaoru Ishikawa (1968)

From Figure 1.1, many problems can contribute to the ineffective Dengue Data Presentation and Analysis. In Ministry of Health Malaysia, *eDengue* version 2 (MOH, Dengue Surveillance and Information System, 2012) is used to monitor case notified within areas and states in the form of lists and tables presentation. *Dengue* (MOH M. &., 1997-2016), a website platform provided from Ministry of Science and Innovative, Remote Sensing which corporate with Ministry of Health are also used to provide annual statistics of dengue in Malaysia and current dengue hotspots in map presentations. While, *Dengue Outbreak Management System* (MOH, 2014)which

integrated between *eDengue* and *eNotifikasi* are used for online real-time disease mapping purposes in peninsular Malaysia connecting *IDengue* for community, provides information dissemination (SMS) and allowed outbreak prediction within focal areas.

The issues and problem that found in the current system used are described as follows:

1. "Over 65 of data need to be displayed at one time."
   From the dataset collected from *eDengue* Version2 system (MOH, 2012), given by Environmental Health Officer from Centre of Disease Control State Health Department of Sabah, there were more than 65 variables that need to be presented for the purpose of effective analysis. The current disease mapping used in *SPWD* (MOH, 2014)provides visualization of cluster disease with multivariate information which required user's attention to select only few data and focus towards presentation before can interpret what information was being presented.
2. "Adapted online tool required more actions to access information."
   Like other CDC department in the world, local CDC also applied the use of online ArcGIS to trace and analysis more complex dengue data. Available online tool like ArcGIS, GeoVista, CartoDB, Google Map and other free mapping tool required user to click, toggle, zoom, pop-up and query before information can be accessed. Because of the limit amount of complex data can be shown visually in an effective way, reduction of screen size on portable devices, some data maybe overlapped and distorted over visualization presentation that may effect on how user analyse data.
3. "Information overload."
   Too many information presented together in multi-screen may causes fatigue, missed of information or wrong interpretation. Thus, a real-time monitoring, adaptable and support input from multi-sensory is needed to solve complex data presentation. Because of the way the human brain processes information, it is faster for people to grasp the meaning of many data points when they are displayed in charts and graphs rather than poring over piles of spreadsheets or reading pages and pages of reports. Visualizations help people see things that

were not obvious to them before. Even when data volumes are very large, patterns can be spotted quickly and easily. Visualizations convey information in a universal manner and make it simple to share ideas with others. Data are like soil and easily can be retrieved from multi-sources, when a suitable design of representation is created, data can be easily interpreted and remember by users.

Given the issues and problem stated above, the motivation for selecting data sonification as domains for the research described in this thesis are as follow:

1. With the increasing dengue cases throughout the years, many efforts have been made to establish an effective presentation tool. However, the current system used in local health department required multi-link views in few separated windows. But when more multivariate data need to be presented at one time at a single screen, information is overlapped with one another and caused visual distraction. If users are fatigue, some data might be left out or hidden from view. Besides, more actions might be needed, user will need to click, toggle or enlarge screen to access additional information. There is therefore potential for human error resulting in different interpretation being conveyed in different presentation reports. This will eventually decrease the effectiveness of analysis and data interpretation.
2. Little work has been conducted with respect to data sonification. Although the use of multiple visualization tools in extracting data from long dataset is undeniable effective, but there is lack of study shown these visualization tools are appropriately designed to overcome human cognitive issues for the ability to interpret multivariate and complex data at one time. Ironically, many researchers have been done to find a better approach that able to represent vast amount of data in single screen for better analysis and representation of complex data. The work here is to structure this data in ways that will improve the performance of data interpretation and analysis. As mentioned in (Yau, 2011), the effective way for analysing a complex data sets, it's the ability to represent the overall pattern of data in a single presentation.