

投稿類別：資訊類

篇名：

探討匿名聊天是否可分辨對方性別

作者：

劉子豪。內湖高中。高三 7 班

郭育廷。內湖高中。高三 7 班

指導老師：

蔡明男老師

壹、前言

一、動機：

現在市面上有許多匿名配對的聊天軟體或網站，使用者時常會臆測配對對象的性別，即使對方坦承性別，也因完全匿名而不知是否為真，那我們能不能透過文字的特性得知對方正確的性別呢？如果可以，又是以什麼樣的根據做判斷的呢？為了得到答案，我們設計了這個實驗，以圖靈測試的變形模式來探討這個問題。

二、實驗所需：

- (一) 80 名受試者
- (二) 2 間資訊教室(各 40 台電腦)
- (三) 自製聊天軟體
- (四) 資料庫

三、自製聊天軟體所需功能

- (一) 確保一對一匿名聊天
- (二) 詢問受試者問題並回傳資料庫
- (三) 達成聊天五分鐘且達 50 則訊息後跳出調查視窗
- (四) 實驗說明

四、實驗流程設計

- (一) 每間電腦教室兩名工作人員，為每台電腦設定好聊天室軟體。
- (二) 確認兩間教室軟體配對正確。
- (三) 將受試者依 40/40 人分為兩組，分派到兩間電腦教室。
- (四) 座位以隨機方式分配入座。
- (五) 和受試者說明實驗流程。
- (六) 實驗開始，受試者開始匿名聊天。
- (七) 檢查聊天紀錄，是否在中途有透露性別，過濾無效資料。
- (八) 將有效資料進行資料分析。
- (九) 報告撰寫。

貳、正文

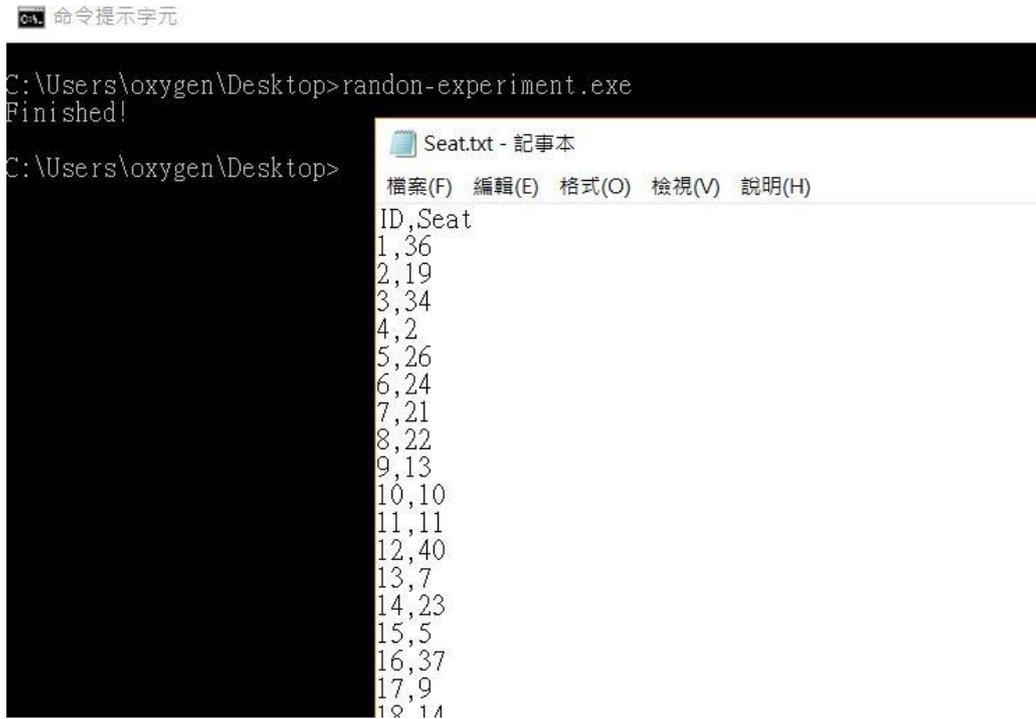
一、我們使用 Visual Basic.Net 開發，目的在於利用其快速的使用者介面設計與網路程式設計，避免程式開發上的過多的困擾，以提高效率。隨機座位配置採用 C++ mt19937 亂數演算法產生，接收回應的伺服器建置採用 Linux + Apache + MySQL + PHP 的架構，伺服器系統資訊如下:

- (一) Operating System: Ubuntu 16.04 LTS
- (二) HTTP Server: Apache2
- (三) PHP 5.6
- (四) MariaDB 10.3.5

編號	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	編號	班級	座號	姓名	受試結果	性別	是否交陪	實驗代碼	分配座位	座位IP	測試對象IP	分配座位	測試對象IP	是否正確	是否異常
2	78				1	0		A-01	10.1.2.1		C-36	10.1.4.36	0	0	
3	4				1	0		A-02	10.1.2.2		C-19	10.1.4.19	0	0	
4	65				1	1		A-03	10.1.2.3		C-34	10.1.4.34	1	0	
5	54				1	1		A-04	10.1.2.4		C-02	10.1.4.2	1	0	
6	15				1	1		A-05	10.1.2.5		C-26	10.1.4.26	1	0	
7	31				1	1		A-06	10.1.2.6		C-24	10.1.4.24	1	0	
8	73				0	1		A-07	10.1.2.7		C-21	10.1.4.21	0	0	
9	21				1			A-08	10.1.2.8		C-22	10.1.4.22	0	1	
10	17				0	0		A-09	10.1.2.9		C-13	10.1.4.13	1	0	
11	10				1	0		A-10	10.1.2.10		C-10	10.1.4.10	0	0	
12	75							A-11	10.1.2.11		C-11	10.1.4.11	0	1	
13	68					1		A-12	10.1.2.12		C-40	10.1.4.40	0	1	
14	9				0	0		A-13	10.1.2.13		C-07	10.1.4.7	1	0	
15	18				1	1		A-14	10.1.2.14		C-23	10.1.4.23	1	0	
16	63				0	1		A-15	10.1.2.15		c-05	10.1.4.5	0	0	
17	30				0	1		A-16	10.1.2.16		C-37	10.1.4.37	0	0	
18	67					0		A-17	10.1.2.17		c-09	10.1.4.9	0	1	
19	23					1		A-18	10.1.2.18		C-14	10.1.4.14	0	1	
20	2				0	0		A-19	10.1.2.19		C-04	10.1.4.4	1	0	
21	33				0	0		A-20	10.1.2.20		C-30	10.1.4.30	1	0	
22	7				1	0		A-21	10.1.2.21		C-08	10.1.4.8	0	0	
23	8				0	1		A-22	10.1.2.22		C-15	10.1.4.15	0	0	
24	47					1		A-23	10.1.2.23		C-17	10.1.4.17	0	1	

圖一、亂數分配後的實驗資料與座位表

匿名聊天是否可分辨對方性別



```
C:\Users\oxygen\Desktop>random-experiment.exe
Finished!
C:\Users\oxygen\Desktop>
```

```
Seat.txt - 記事本
檔案(F) 編輯(E) 格式(O) 檢視(V) 說明(H)
ID,Seat
1,36
2,19
3,34
4,2
5,26
6,24
7,21
8,22
9,13
10,10
11,11
12,40
13,7
14,23
15,5
16,37
17,9
18,14
```

圖二、C++ 亂數分配程式

二、軟體使用介紹

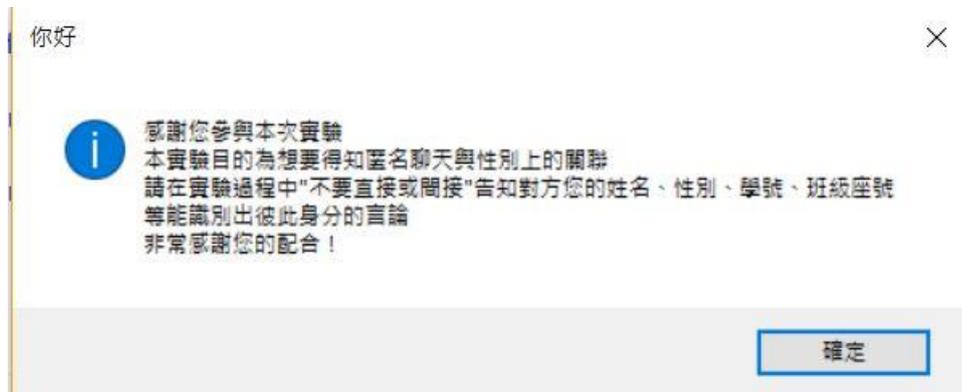
- (一) 軟體採用組態檔(Config file)的方式進行設定，組態檔內各項設定以分號區隔，格式如下：

```
配對者IP;我方PORT;對方PORT;伺服器接介URL;我方實驗代號;對方實驗代號;對話數量臨界值;時間限制(sec)
Example:
10.0.1.1;7000;7001;https://web.oxygentw.net/api/ChatExperiment/api.php;Chat01;Chat02;50;300
```

圖三、軟體組態檔格式與範例

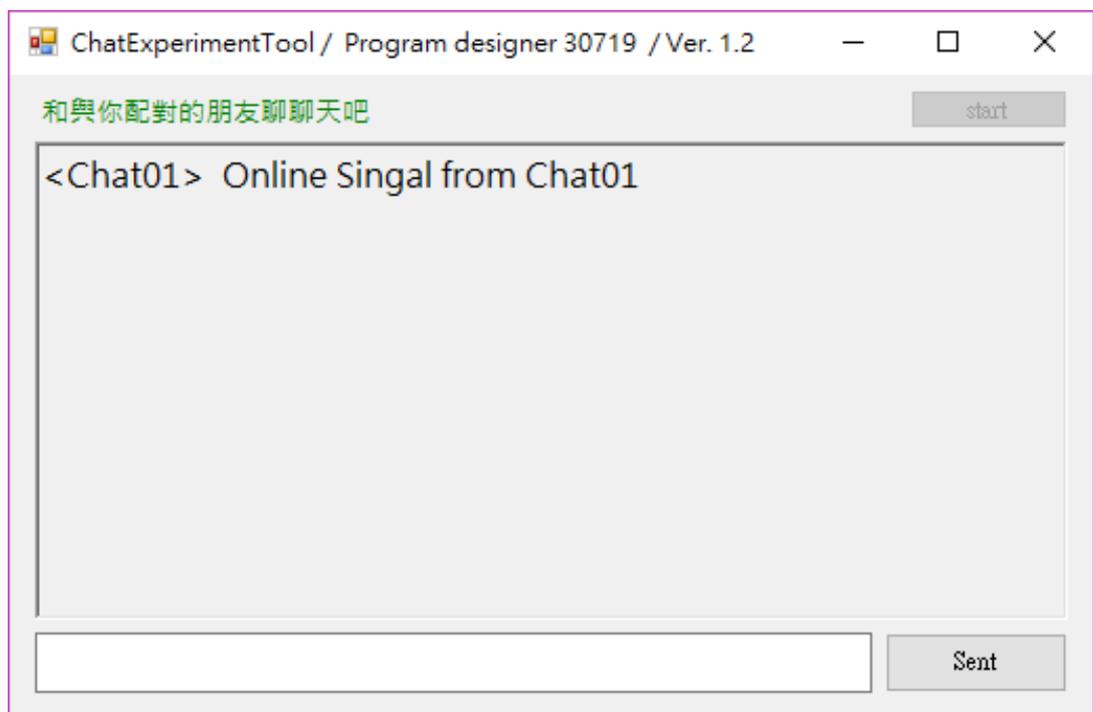
匿名聊天是否可分辨對方性別

- (二) 軟體啟動時會顯示 MessageBox，向受試者介紹實驗並告訴他們一些注意事項。



圖四、軟體啟動時的實驗說明

- (三) 接著會進入主要聊天視窗，在實驗工作人員確定兩間教室內受試者皆已就緒時通知受試者按下右上角的 **start button**，使用者此時被允許送出訊息到配對的受試者，同時計時器啟動，計算實驗進行時間。



圖五、主要聊天視窗

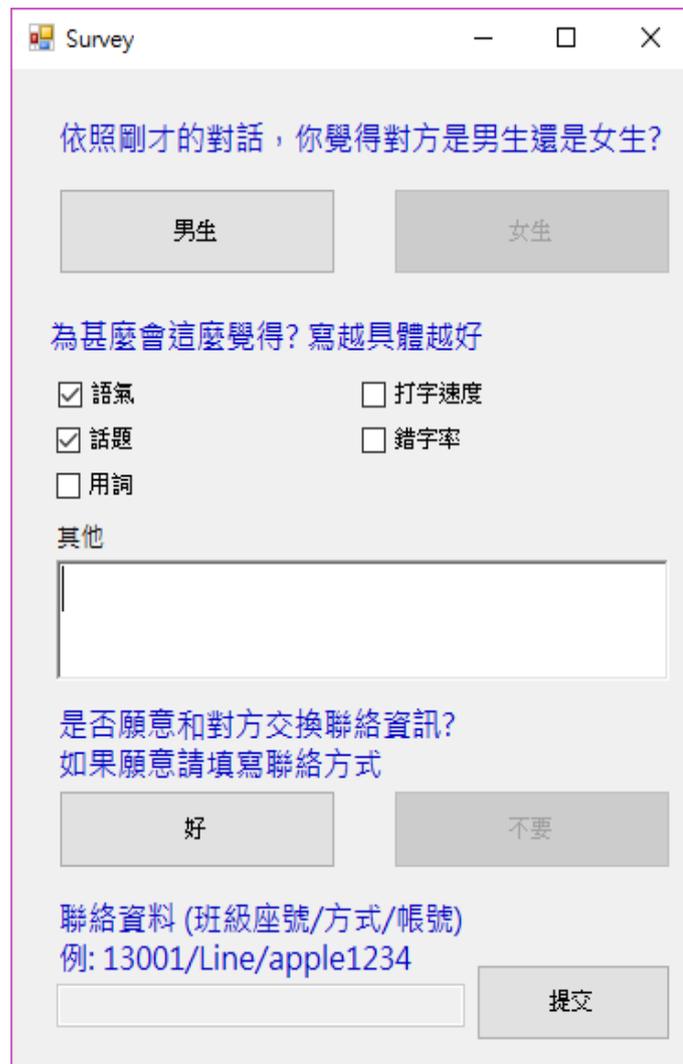
- (四) 聊天訊息數達到 50 則且聊天時間達到 300 秒，軟體會終止聊天並顯示回

匿名聊天是否可分辨對方性別

應表單介面，此舉能防範受試者大量傳送無意義的訊息而導致很快達到聊天訊息數限制，也能防止受試者只是乾坐在那等時間到，完全沒有進行對談。

(五) 回應表單主要收集兩項資訊，性別和判斷理由，藉此我們可以得知判斷的正確率與影響判斷的主要原因，交換聯絡方式的部分是由於實驗軟體設計上會突然中斷聊天(我們並沒有告知受試者還剩多少時間)受試者可能尚有想對配對者說的話，因此給受試者自由選擇是否願意交換聯絡方式，當雙方都願意時將由實驗工作人員事後知會受試者。

(六) 事後我們使用 Python 做 JSON 解析資料庫資料並統計結果，使用 Google Sheet 做資料整理與圖表輸出。



The image shows a survey form window titled "Survey". The form contains the following elements:

- Question: 依照剛才的對話，你覺得對方是男生還是女生?
- Buttons: 男生 (Male) and 女生 (Female).
- Question: 為甚麼會這麼覺得? 寫越具體越好
- Checkboxes: 語氣 (checked), 打字速度, 話題 (checked), 錯字率, 用詞.
- Text input: 其他 (Other).
- Question: 是否願意和對方交換聯絡資訊? 如果願意請填寫聯絡方式
- Buttons: 好 (Yes) and 不要 (No).
- Text input: 聯絡資料 (班級座號/方式/帳號) 例: 13001/Line/apple1234
- Button: 提交 (Submit).

圖六、問卷填答介面

匿名聊天是否可分辨對方性別

```
1 import json
2
3 file = open(r"C:\Users\oxygen\Desktop\聊天實驗解析\result.json", "r")
4 jsondata = file.read()
5 file.close()
6
7 data = json.loads(jsondata)
8
9 list = ["t"]
0 #檔案前處理
1 for x in range(60):
2     #print(data[x]["sexuality"])
3     if(data[x]["exchange"] == "1" ):
4         print(data[x]["objectID"] + " -> " + data[x]["contact"] + " -> " + data[x]["TestobjectID"])
5
6
7
```

圖七、Python 分析資料庫資料

```
2018/3/29 下午 03:03:51 <Chat60> Online Singal from Chat60
2018/3/29 下午 03:04:08 <Chat33> (沒有朋友((
2018/3/29 下午 03:04:33 <Chat60> hi
2018/3/29 下午 03:04:39 <Chat33> 嘻嘻嘻嘻~~~
2018/3/29 下午 03:04:59 <Chat33> 不能打注音文 嗚嗚嗚
2018/3/29 下午 03:05:23 <Chat33> 你可以給我一點熱情嗎 q q
2018/3/29 下午 03:05:26 <Chat60> 為何啊!
2018/3/29 下午 03:05:38 <Chat33> 沒 沒 (
2018/3/29 下午 03:05:43 <Chat60> 熱情
2018/3/29 下午 03:05:55 <Chat60> 要怎麼給
2018/3/29 下午 03:06:12 <Chat33> 給我一點回應嘛 嗚嗚
2018/3/29 下午 03:06:22 <Chat33> ;W;
2018/3/29 下午 03:06:31 <Chat60> 怎樣的回應
2018/3/29 下午 03:06:51 <Chat33> 沒就隨便聊 (
2018/3/29 下午 03:06:55 <Chat60> 像你過的好嗎
2018/3/29 下午 03:07:09 <Chat33> 段考爆炸了 嗚嗚嗚嗚;;;
2018/3/29 下午 03:07:26 <Chat60> 沒關西我也不差
2018/3/29 下午 03:07:36 <Chat33> 數學十國文只有60耶哇喔...
2018/3/29 下午 03:08:07 <Chat33> ; ;
2018/3/29 下午 03:08:09 <Chat60> 我到現在只有國文及個
2018/3/29 下午 03:08:29 <Chat33> 啊我是數學十英文... ((
2018/3/29 下午 03:08:44 <Chat33> 我國文有及格wwwww
2018/3/29 下午 03:09:08 <Chat60> 我數學家英文也只有70左右
2018/3/29 下午 03:09:16 <Chat33> 你數學還好嗎XDDDDD
2018/3/29 下午 03:09:35 <Chat33> 比我好一點wwww
2018/3/29 下午 03:09:41 <Chat60> 那你平常喜歡做啥事
2018/3/29 下午 03:09:56 <Chat60> 我數學也才27
2018/3/29 下午 03:10:00 <Chat33> 我，我喜歡打手遊 ( (
2018/3/29 下午 03:10:01 <Chat60> 哈哈
2018/3/29 下午 03:10:08 <Chat33> 我21wwwwwwwwww
2018/3/29 下午 03:10:18 <Chat60> 玩啥遊戲啊
2018/3/29 下午 03:10:26 <Chat60> 西西
2018/3/29 下午 03:10:27 <Chat33> FGO (?)
2018/3/29 下午 03:10:47 <Chat33> 西?
2018/3/29 下午 03:11:13 <Chat60> 我沒有太常玩哪個
2018/3/29 下午 03:11:22 <Chat33> 喔喔XDD
2018/3/29 下午 03:11:23 <Chat60> 嘻嘻
2018/3/29 下午 03:11:29 <Chat33> XDD
2018/3/29 下午 03:11:38 <Chat60> XDDDD
```

圖八、匿名對話過程節錄

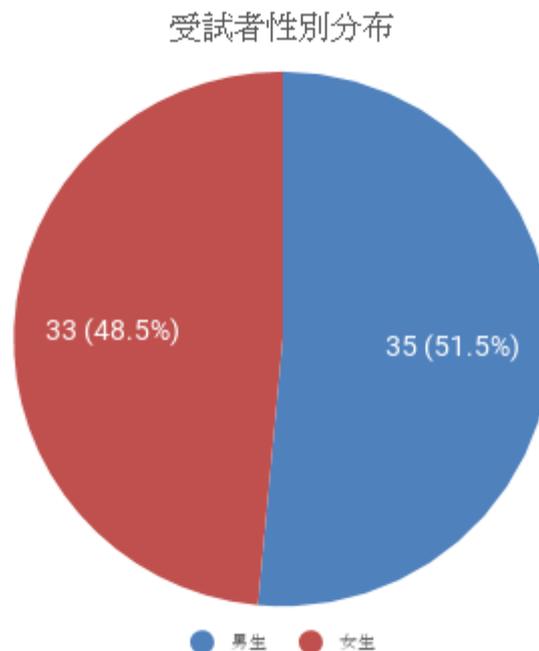
參、結論

一、實際實驗狀況

實際實驗過程並不順利，因為實驗受試者招募非常倉促，我們在 20 分鐘的下課裡臨時被告知要進行實驗，老師找了兩個班級作為受試者，但因時間緊迫，受試者未能完全理解實驗步驟與方法，兩間教室的座位配置也十分混亂，並非實驗設計中的規劃，甚至造成多位受試者沒有成功配對的問題，也因指導老師需要使用教室上課而要求受試者以「洗版」的方式完成實驗，造成實驗品質極度不佳。

二、實驗數據分析

(一) 受試者性別比例相當，實驗結果不會受到大量相同性別而產生誤差。



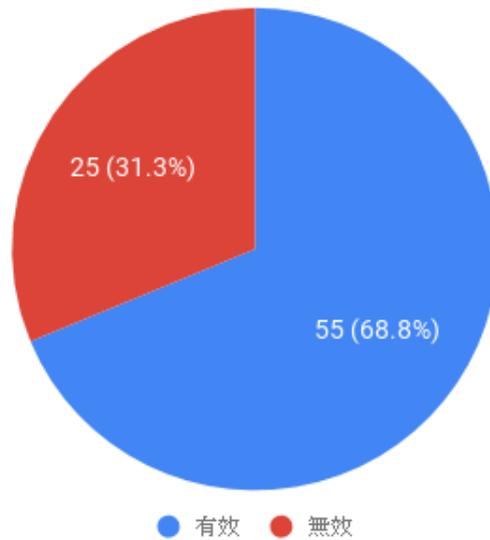
圖九、受試者性別分布圓餅圖

(二) 回應的有效率不高，僅有總回答的 75%，主因是實驗過程過於混亂，且準備時間不周以及與老師們協調有問題。無效樣本說明：

- 1、電腦連不上網
- 2、部分受試人員未出席，導致其對面的其他受試者出現無人可聊的窘境
- 3、雙方連線失敗

匿名聊天是否可分辨對方性別

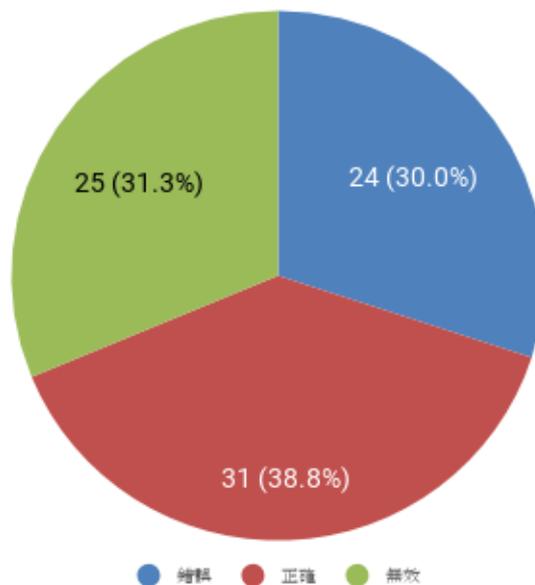
受試者回應有效率



圖十、受試者回應有效率圓餅圖

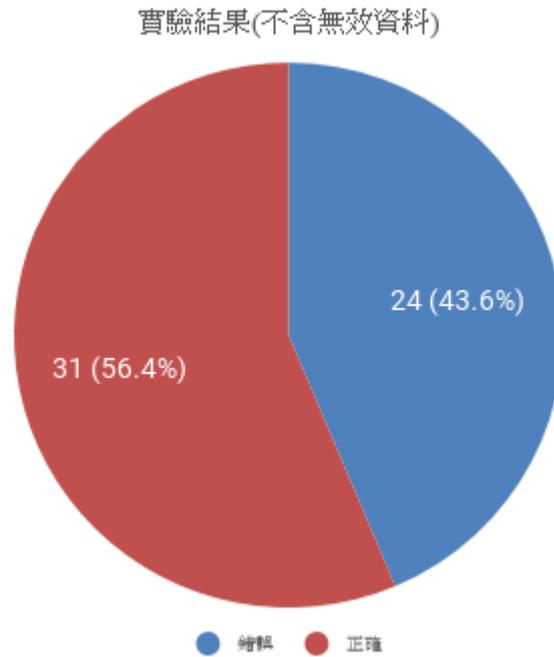
(三) 實驗結果扣除無效資料後我們發現受試者猜對對方性別的比例為 56.4% 略超過一半，實驗結果並不能明顯顯示出匿名聊天是否能得知對方性別，畢竟即使受試者全部憑感覺猜測，正確的期望值亦為 50%，雖然正確率較錯誤率高，我們依然認為匿名聊天不能正確得知對方性別。

實驗結果



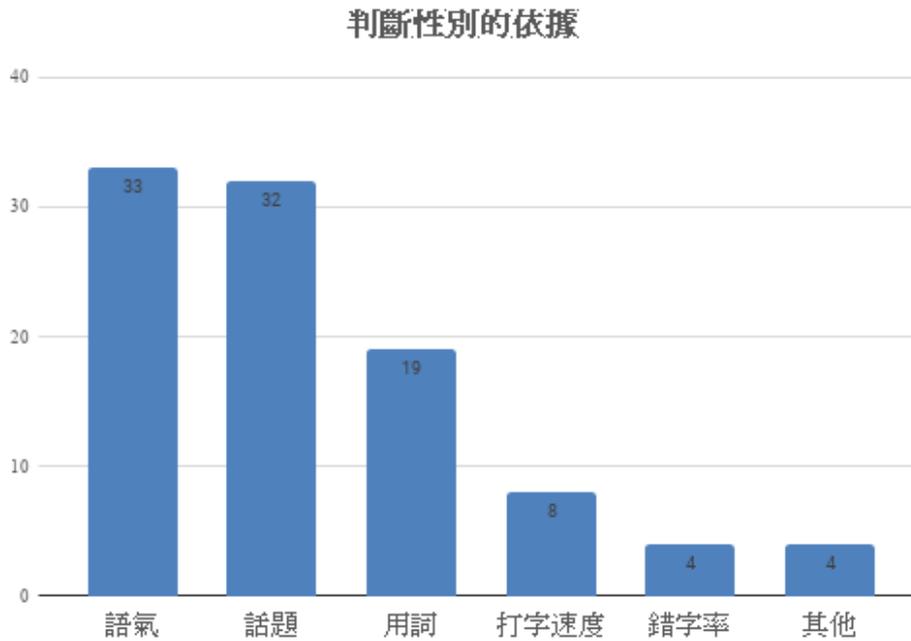
圖十一、實驗結果(含無效回應)圓餅圖

匿名聊天是否可分辨對方性別



圖十二、實驗結果圓餅圖

- (四) 此圖表是根據上一張圖表的整理，也就是包含答對與答錯的所有有效測資的整理，因為分辨理由可供多選，故數量相加後遠超過有效測資的人數。其中語氣與話題占了如此高的比例，主要是由於部分受試男性在實驗開始後，不斷出現不恰當的話題與語尾助詞，這也是多數人能快速猜出對象性別的理由之一，而這也是現今許多匿名聊天軟體的亂象。語氣與話題在日常生活中原本就是男女差異較大的部分，這和實驗數據吻合，因此我們認為受試者**主要透過語氣與話題來進行性別判斷**，但因表單選項是我們設計的，可能有引導作答之疑慮，這是我們實驗設計上的疏失。



圖十三、判斷性別的依據長條圖

(五) 檢討與困難

1. 受試者招募不易
2. 實驗過程過於匆忙
3. 未掌握實驗控制變因
4. 師生協調不良

(六) 修正與發展

1. 透過簡易獎賞吸引受試者參與
2. 實驗場勘與實際演練
3. 選擇充裕時間時進行實驗
4. 招募更多受試者以建立更精確的數據
5. 將聊天時間長短加入操作變因來實驗

(七) GitHub

1. <https://github.com/oxygen-TW/ChatExperimentTool/>

肆、引註資料

1. <https://tw.answers.yahoo.com/question/index?qid=20090226000015KK07933>
2. https://en.wikipedia.org/wiki/Turing_test
3. https://www.w3schools.com/php/php_mysql_connect.asp
4. <http://kuma-uni.blogspot.com/2012/06/jsonpythonjson.html>
5. Visual Basic.Net 網路程式發展實務，廖文良，碁峰出版，2001/10