The Impact of Audio Stimuli on Cognitive Flexibility: An Investigation of the Stroop Test

Karla Maria Garcia, Kevin O'Shea

Dept of Computer and Software Engineering
Technological University of the Shannon
Midlands Ireland
a00304376@student.ait.ie, a00304379@student.ait.ie

Abstract

The aim of the project is to measure and analyse the effect of audio stimuli on human cognitive ability. 48 members of the public were subjected to a variety of Stroop-like tests with a range of accompanying audio. Response times were catalogued and contrasted to examine the effects of audio on participants. There were five tests in total, including one control test to measure user reaction time. Of the remaining four tests, one had no audio while the other three had music, construction noise and spoken word. The majority of users improved as the test proceeded, leading to lower response times in general. Analysis of test data showed that there were some minor differences in performance between genders, implying that this could be an area worthy of further research.

Keywords

Stroopt test, data analysis, audio stimuli, control of attention, cognitive ability.

1. Introduction

Many modern day workplaces are environments with many outside influences and distractions. The effect of audio on cognitive ability is not known to a full extent. This investigation aims to shed some light on how human cognition is affected, either positively or negatively, by different audio stimuli. Cognitive ability is tested using the Stroop test with accompanying audio. User performance is measured to report success rate and response time. Test response times are analysed to observe performance during each of the four tests, in order to determine which audio environment is least/most beneficial to cognition performance.

2. Literature Review

According to [7], the Stroop Colour Word test, usually referenced only as Stroop test, is a psychological experiment to measure cognitive flexibility and attention, in which the processing of a stimulus is affected by the processing of a second stimulus. This results in an interference effect that can be easily triggered

by, for example, reading words which represent the name of a colour when the word text does not correspond to the name of said colour.

Jafari [1] mentions that the Stroop effect is associated with speed of processing and attention. Speed of processing is related to the fact that some individuals are faster at reading than naming colours, and attention is the ability to ignore some environmental distractions but not others. Consequently, the experiment can be affected by problems like environmental perturbations, the ability of individuals to become faster after repeated exposures, users' vision or hearing difficulties, language barriers for individuals not fluent in the test language, and slower reaction times that may score poorly.

The aim of this study is to investigate the correlation between different scenarios and response time, analysing how the environment affects the results. An application is implemented to perform Stroop tests, collect response times and success rate. Different audios are played in the background of each test in order to simulate everyday environmental distractions. User details such as age and gender are collected to help investigate any test performance correlations. Test order is random to avoid user memorization and all the output will be further analysed.

In order to understand how the Stroop test works and how to evaluate the results of this experiment, a research is conducted in the two main areas listed below:

- Stroop test
- Data Analysis

The Stroop test research consists of exploring previous studies aiming to have a better understanding of the experiment, the concepts involved, how it is applied, and what can be evaluated. The Data Analysis area details how the tests are executed to collect data, and also how data can be analysed and presented.

2.1. Stroop Test

The Stroop effect is a psychological experiment that was originally designed by John Ridley Stroop in 1935 [6]. The study consisted of individuals reading a sequence of the words red/green/brown/purple or naming the colours the words were written in. The words were not printed in the colour that they are named, instead they were printed in one of the other three colours. For example, the word red could be printed in green. Consequently, two stimuli are presented at the same name; the word and the colour.

Figure 1 - Original Stroop test words

Words were also printed in black and the colours were also presented in squares/swastikas. The reading stimuli was measured focusing on the time difference between reading the coloured word and reading the same word written in black ink. The naming of colour stimuli was measured considering the time difference between naming colours from words and same colours from squares/swastikas.

Stroop [6] concluded that the colour stimuli doesn't significantly affect the time for reading words, however the word stimuli caused an increase of 74.3% in naming colours.

As demonstrated in [2], by presenting individuals with stimuli that vary in two dimensions, the Stroop test proves that people have a natural capacity for selective attention and that some stimuli can be ignored, escaping attentional control.

According to [4], our life is full of distractions on a daily basis. Some can be easily ignored while others not so much. One example is auditory distraction that affects our cognitive process.

However, as mentioned earlier [1], there are studies that differ on how noise affects cognitive performance, with some arguing that this can bring about positive results, while others believing it to have a negative impact. This is the main factor that motivated this study; to investigate how the results of Stroop test can be affected due to external audio/distractions.

Since the original Stroop was presented, many other variations have been developed. For example, the Auditory Stroop tests, where audio is introduced to the experiment. Steven et al [5] investigated gender associations in adults and children using Stroop tests, in which participants were asked to categorise the gender of voices while being presented with gender-stereotypical words (e.g., dress) and names (e.g., Peter). Dittrich and Stahl [3] also conducted an Auditory Stroop experiment in which an auditory tone is played just before the visual Stroop stimuli. These studies introduced audio to the experiment, but they did not evaluate the impact of trivial noises, like conversations, music, traffic, etc, on traditional Stroop test results.

3. Test Configuration

In this study, the Stroop test consists of two questions regarding users' age and gender, a control test and four different environmental tests. Fig 2 illustrates the initial screen. The stats button presents the statistics for each test.

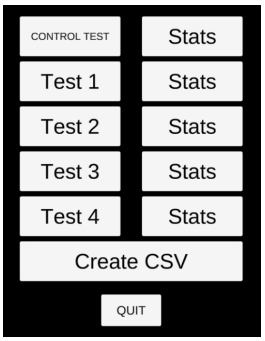


Figure 2 - Initial screen

At the beginning of the experiment, a control test is given to ensure the examinee's comprehension about how tests work and also to determine the participant's baseline reaction time. This baseline result is used to offset the performance of each of the following four Stroop tests. The control test does not present incongruence options, it shows only one option per time, in which the word colour matches with the colour name, as shown in Fig 3. No audio is played during the control test and the colour is presented 20 times for selection.

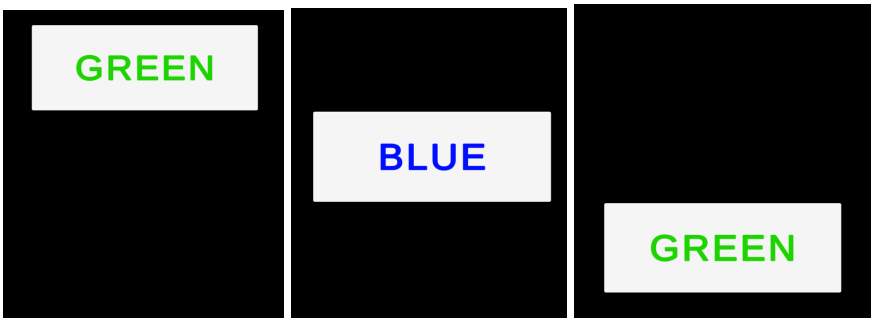


Figure 3 - Series of control tests

The four Stroop tests show a list of three words, where two of them are incongruent and only one is congruent. The examinee should select the congruent option. For each test a list of words, as illustrated in Fig 4, is presented 20 times varying the words positions and colours and accompanied by different audio stimuli. Different audios are played for each test:

- Test 1: No audio,
- Test 2: Classical music,
- Test 3: Construction noise,
- Test 4: The words green, blue and red are spoken in a random order.

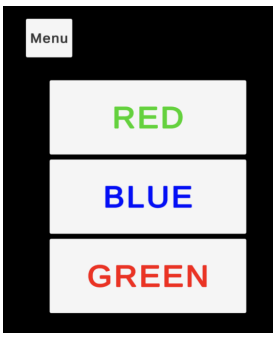


Figure 4 - Test example

As shown on Fig 5, once the user finishes the tests, the user should click on ***Create CSV*** to provide identification, age and gender to save data on a csv file. If multiple users run the test sequentially data will be appended to the same file.

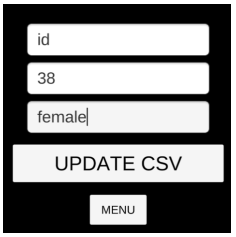


Figure 5 - Required fields

For each participant the application records the demographic, response time, and success rate of each test.

3.1. Data Collection

The application was made available via Google Drive and the link was sent to users via email and other communication tools, like Slack and WhatsApp, with a brief explanation and steps to be executed. This is the message sent to the users:

Hello, can I have 5 minutes of your time? It's for an experiment of my master degree.
I'd ask you to run the application shared in the link below then click on Control Test and Start.
You have to select the word colour that matches with the colour name.
Repeat the same steps for test1, test2, test3 and test4.
When all tests are done, then click in Create CSV, fill in the required data and click in Update CSV, you can add any name to the field name.
The CSV file will be generated in the same folder where you saved the application.
Please send me this csv.

Some tests play audio, so a headset is required.
Thank you very much for your time.
https://drive.google.com/file/d/1qVMy-mg_7NOad7jwyrY8tleFNiB7PKJY/view?usp=share_link

During the data collection process MacOS users reported issues with saving the CSV file that was not created on some machines. The users were oriented to take screenshots of the statistics reported via the ***Stats*** button and return them. After one week, 48 users returned, all files were merged and the data collection was concluded.

Fig 6 shows a sample of the original CSV files obtained from users. At this figure the first column was removed to avoid exposing participants as some users provided their name.

25	M	0	20	0.55	0.53	0.53	0.43	0.53	0.43	1.08	0.47	0.50	0.47	0.57	0.62	0.57	0.47	0.52	0.58	0.57	0.52	0.55	0.45
25	M	1	19	1.63	0.93	1.10	1.03	0.88	1.10	2.15	2.45	2.22	1.73	0.95	0.83	0.90	1.28	1.28	1.65	1.20	1.25	1.92	1.15
25	M	2	20	0.77	0.93	1.30	1.33	0.98	0.93	0.92	1.30	1.25	1.03	1.45	1.20	1.10	1.07	1.47	0.98	1.43	1.07	1.03	1.20
25	M	3	19	1.25	1.67	1.33	1.22	2.20	1.33	1.48	1.43	1.30	1.12	1.28	1.32	1.13	0.98	1.32	0.93	1.85	1.05	1.53	1.52
25	M	4	20	2.10	1.17	0.88	1.48	1.43	1.48	1.20	2.50	1.98	1.43	1.42	1	1.48	2.60	1.77	2.27	0.88	1.30	1.17	1.88
36	M	0	20	0.60	0.65	0.58	0.67	0.52	0.75	0.70	0.60	0.67	0.67	0.72	0.67	0.80	0.72	0.87	0.82	0.62	0.67	0.73	0.78
36	M	1	20	2.28	2.32	1.77	1.33	3.63	1.57	2.93	1.27	1.17	2.65	1.60	2.05	3.77	2.58	1.70	1.23	1.48	1.32	1.30	3.03
36	M	2	20	4.37	1.83	2.32	7.15	2.87	1.30	2.07	2.57	1.25	1.27	1.35	1.12	1.20	1.72	2.20	1.07	1.82	2.03	1.37	1.72
36	M	3	20	1.62	4.68	1.83	1.12	1.82	0.93	1.33	3.53	2.58	3.23	2.70	1.72	3.67	1.75	1.28	2.37	1.07	2.17	1.13	1.97
36	M	4	20	3.60	1.33	2.07	2.35	3.18	1.45	1.43	1.20	1.50	1.35	0.97	1.38	1.40	2	3.97	1.55	1	2.05	1.70	2.82

Figure 6 - CSV response file before sanitising

A manual process was performed to have data in a good format for analysis:

- Added headers to each column as the file provided by the application did not add it automatically.
- Formatted *gender* field to either 'male' or 'female'.
- Add a new column called *Nationality* adding the corresponding country of each participant.
- Identify the fastest response time on the control test to measure the individual physical response time of users with a computer mouse. This time was subtracted from the response times for tests 1 to 4 in order to establish the actual cognition response time.
- Replaced users' name provided in the first column by a sequential identifier.

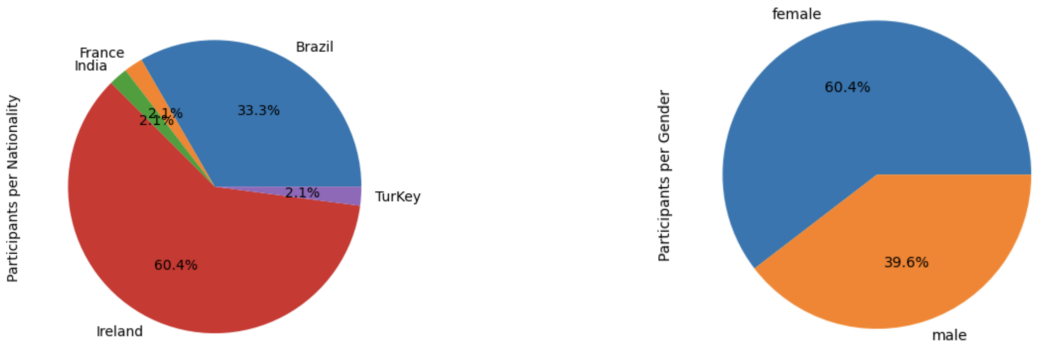
Fig 7 shows data after sanitising.

id	age	gender	Nationality	test_no	successful	t1	t2	t3	t4	t5	t6	t7	t8	t9	t10	t11	t12	t13	t14	t15	t16	t17	t18	t19	t20
1	52	male	Ireland	0	20	1.45	0.57	0.53	0.6	0.66	0.5	0.66	0.63	0.56	0.6	0.67	0.5	0.64	0.6	0.61	0.69	0.52	0.71	0.59	0.57
1	52	male	Ireland	1	20	2.65	2.2	1.81	3.12	3.76	2.72	1.46	1.99	2.71	3.2	2.75	1.81	1.69	1.75	2.47	1.42	2.09	2.04	3.52	2.71
1	52	male	Ireland	2	20	2.41	2.22	1.18	1.56	2.02	1.65	2.67	3.96	2.73	1.87	1.08	3.08	1.67	1.05	2.72	1.83	1.68	1.27	1.98	1.33
1	52	male	Ireland	3	20	2.64	1.68	0.87	1.45	2.04	2.86	2.96	0.95	2.24	1.03	2.43	1.11	1.55	4.52	1.84	1.49	1.77	1.86	3.63	2.26
1	52	male	Ireland	4	20	3.02	0.84	2.67	0.99	2.79	1.55	2.89	1.46	2.06	1.81	1.04	1.51	1.11	1.75	1.25	2.93	1.16	2.68	2.35	1.37
2	42	male	Ireland	0	20	0.75	0.77	0.71	0.78	0.75	0.68	0.75	0.73	0.85	0.77	0.75	0.76	0.69	0.66	0.73	0.89	0.8	0.79	0.87	0.8
2	42	male	Ireland	1	20	1.32	0.9	1.24	0.89	1.04	1.25	0.88	0.95	0.94	0.71	0.88	1.08	2.03	1.5	1.76	1.92	1.88	1.18	1.15	1.71
2	42	male	Ireland	2	20	1.65	0.88	0.84	1.03	2.11	0.86	1.14	1.42	1.12	1.47	1.06	1.19	0.89	1.33	1.58	1.29	1.6	0.81	1.12	1.24
2	42	male	Ireland	3	19	1.97	1.35	1.67	1.96	1.2	1.23	1.35	1.37	1.55	1.42	1.24	2.3	0.83	2.08	1.04	0.73	0.72	1.22	1.23	2.85
2	42	male	Ireland	4	20	1.59	1.83	1.7	3.32	0.87	1.67	1.1	1.15	0.91	0.81	0.95	1.05	1.15	0.79	1.15	0.93	1.05	1.65	2.32	1.04

Figure 7 - CSV response file after sanitising

4. Results

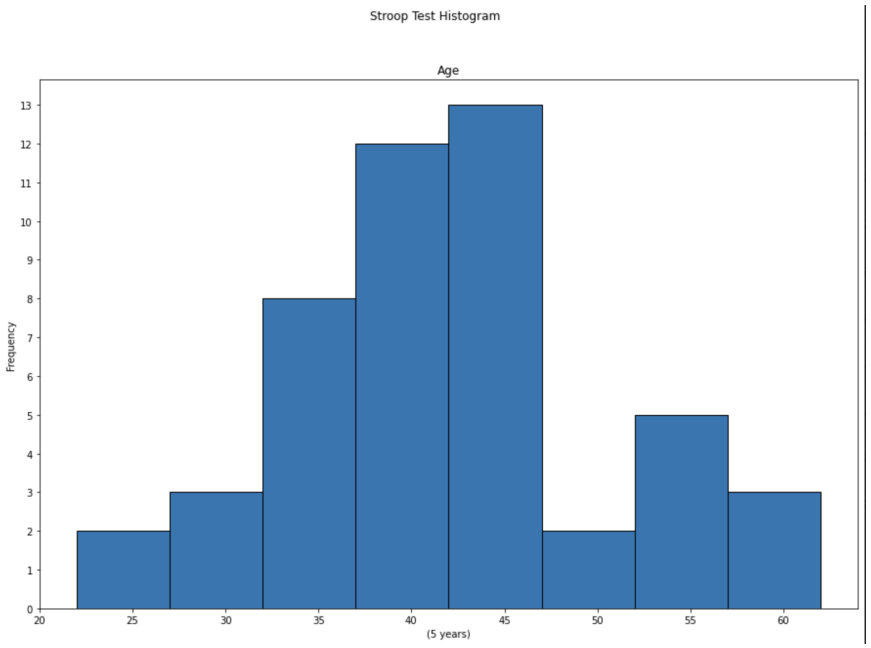
Data analysis was performed by using Python as programming language, Anaconda[1], JupiterLab[2] and Matplotlib[3] and the first step was to explore data. An initial analysis identified that the majority of participants are female and from Ireland, as we can see on Fig 8.



(a) *(b)*

Figure 8 - Participants by nationality (a) and gender (b)

The age of the majority of participants is between 35 and 45 as shown on Fig 9.



[1] Distribution of the Python and R programming languages for scientific computing. https://www.anaconda.com
[2] Web-based user interface for Project Jupyter. https://jupyterlab.readthedocs.io/en/stable
[3] Library for creating visualisations in Python. https://matplotlib.org

Figure 9 - Participant's age

For each test, the first two response times were disregarded as too often this led to unreliable response times. This is typically down to users not beginning the test right away, perhaps due to distractions. For this reason, statistics were calculated on the final 18 response times of each test.

Table 1 presents the main metrics of the response time for each response regarding test 1, 2, 3 and 4. The metrics min and max, that can vary between 0.7s and 31s, and the standard deviation indicate that some participants could be distracted with other tasks when running the test.

	t1	t2	t3	t4	t5	t6	t7	t8	t9	t10	t11	t12	t13	t14	t15	t16	t17	t18	t19	t20
mean	3.3	2.2	2.0	2.1	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.1	2.0	2.0	2.0	2.0	2.1	2.1	2.1	2.1
std	2.9	1.0	0.9	1.2	0.8	0.8	0.7	0.7	0.8	0.8	0.9	0.8	0.9	0.8	0.8	0.9	1.3	1.0	0.8	1.0
min	0.7	0.8	0.7	0.8	0.8	0.8	0.8	0.8	0.8	0.7	0.8	0.8	0.8	0.7	0.8	0.0	0.7	0.8	1.0	0.6
25 %	1.9	1.5	1.5	1.4	1.5	1.4	1.4	1.4	1.4	1.4	1.4	1.5	1.4	1.4	1.4	1.5	1.5	1.4	1.5	1.5
50 %	2.6	2.0	1.8	1.8	1.8	1.8	1.8	1.9	1.8	1.8	1.7	1.8	1.7	1.8	1.7	1.8	1.8	1.9	1.9	1.9
75 %	3.6	2.6	2.4	2.6	2.4	2.5	2.4	2.4	2.4	2.4	2.3	2.4	2.4	2.3	2.3	2.3	2.4	2.5	2.4	2.4
max	31	6	6	12	7	5	5	6	8	5	9	5	5	5	5	8	14	7	6	8

Table 1 - Response time metrics

After removing the first two attempts for each test the average of the response time among tests is between 1.2 seconds and 1.6 seconds. In Fig 10 we can see that user response times tend to decrease as the test proceeds, with a slight increase on test 4 (spoken word audio). One reason for improved response times may be that users become familiar with the test structure as the test proceeds.

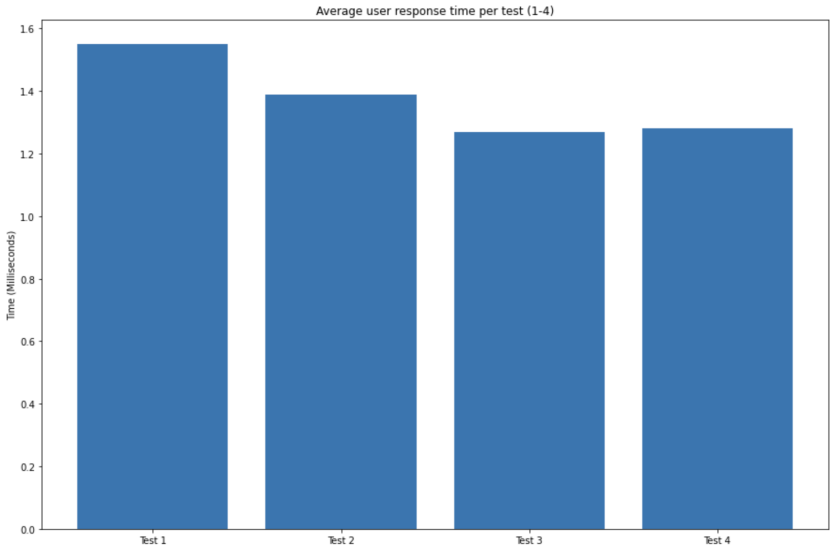


Figure 10 - Average user response time per test

As illustrated in Fig 11, response times were analysed by gender for each test 1 to 4. For tests 1 (no audio) and 2 (music) males achieved lower average response times than females. This trend was reversed however in tests 3 (construction noise) and 4 (spoken words), with females performing better than their male counterparts. The biggest discrepancy in response times was on test 3, where females registered an average of 60 milliseconds quicker than males.

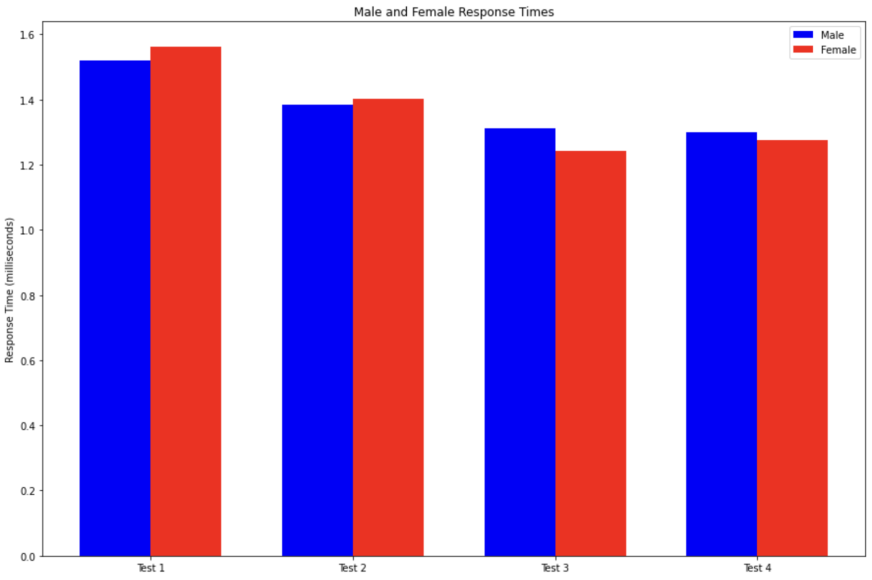


Figure 11 - Average user response time per test by gender

As illustrated in Fig 12, users performed best in test 1, which is the only test without accompanying audio, indicating that regardless of the audio type, it affects the control of attention. For tests 2, 3 and 4 users achieved an identical average success rate which was lower than test 1.

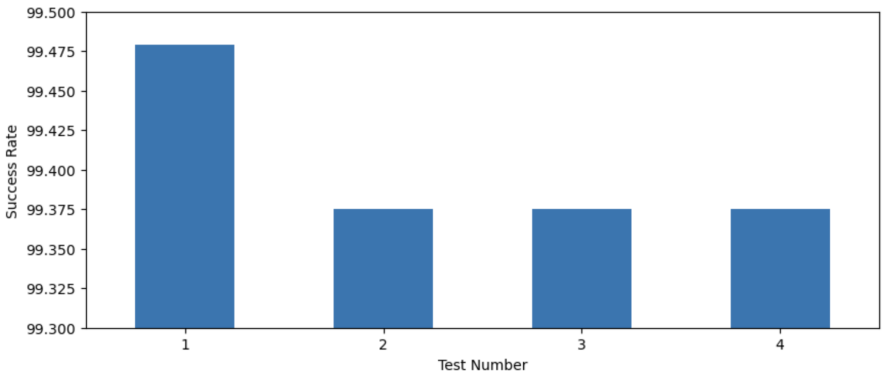


Figure 12 - Success rate per test

On average, females achieved the highest success rate in test 1 (99.8%), while males scored best in test 4(100%). Conversely, the lowest success rate for females was in test 4 (98.9%), while males achieved the lowest score for test 1(98.9%). These findings may suggest that men and women's cognitive abilities can be affected by differing environmental audio stimuli.

5. Conclusion

Stroop Test is a psychological experiment, originally presented in 1935, that measures individuals' cognitive flexibility and attention. Since then, many variations have been developed, including the Auditory Stroop test, where audio is introduced.

Some studies include audio as a stimulus where the participant needs to hear the audio to respond to the test, while some other variations add a tone before or during the Stroop stimuli. However, they do not evaluate how trivial noises affect, either positively or negatively, the results of Stroop tests.

This study aimed to investigate how the sounds present in individuals' lives affect their control of attention and impact the performance of daily tasks. The success rate indicates that regardless of the type of audio it can negatively affect the control of attention, as users performed better on test 1 (no audio). Regarding the different types of audio, the results denote that there is no difference among them as response time and success rate are quite similar.

Considering that the results show some marginal differences between gender-based response times, there may be grounds for further investigation into how men and women are affected by audio stimuli.

It should be mentioned that in most cases participants performed the tests without supervision, which could affect the results, since there is no guarantee that the user was not distracted by other stimuli.

6. Future work

Application was developed to run on two different Operational Systems, macOs and Windows, aiming to cover a big number of users. However many users faced issues when running the application on macOs as the CSV file was not created due to a privileges issue.

In a next experiment, it should be considered to implement a Web Application where data would be saved directly to an online database. This improvement removes the dependency of the user local environment and avoids the work of having to merge all files.

Future analysis could be carried out on new and existing data in order to research if there is a link to performance times and native speaking language. There is also room to increase the difficulty of the Stroop test application to include four colour options instead of just three.

The experiment could be repeated in a controlled environment where participants are supervised, for example, in a dedicated room where it is possible to ensure identical testing conditions, free from outside distractions.

Another interesting analysis based on the Stroop test is to evaluate attention control according to the audience. An experiment can be performed to analyse the attention control of participants diagnosed with Autism Spectrum Disorder (ASD) and participants not diagnosed with any disorder.

7. References

[1] Jafari MJ at el. The Effect of Noise Exposure on Cognitive Performance and Brain Activity Patterns. Open Access Maced J Med Sci. 2019 Aug 30;7(17):2924-2931. doi: 10.3889/oamjms.2019.742. PMID: 31844459; PMCID: PMC6901841.

[2] J. D. Cohen at el. "On the control of automatic processes: A parallel distributed processing account of the Stroop effect," Psychol. Rev., vol. 97, (3), pp. 332-361, 1990. https://www.proquest.com/scholarly-journals/on-control-automatic-processes-parallel/docview/614320835/se-2. doi: https://doi.org/10.1037/0033-295X.97.3.332 (accessed Feb. 12, 2023).

[3] K. Dittrich and C. Stahl, "Nonconcurrently presented auditory tones reduce distraction," Attention, Perception and Psychophysics, vol. 73, (3), pp. 714-719, 2011. https://www.proquest.com/scholarly-journals/nonconcurrently-presented-auditory-tones-reduce/docview/920257828/se-2 (accessed Feb. 12, 2023).

[4] Lavie, N. (2010). Attention, Distraction, and Cognitive Control Under Load. Current Directions in Psychological Science, 19(3), 143–148. doi: https://doi-org.ezproxy.ait.ie/10.1177/0963721410370295

[5] Steven B. Most, Anne Verbeck Sorber, Joseph G. Cunningham. Auditory Stroop reveals implicit gender associations in adults and children. Journal of Experimental Social Psychology, Volume 43, Issue 2, 2007, Pages 287-294, ISSN 0022-1031. doi: https://doi.org/10.1016/j.jesp.2006.02.002.

[6] Stroop, J. R. (1935). Studies of interference in serial verbal reactions. Journal of Experimental Psychology, 18(6), 643–662. doi: https://doi.org/10.1037/h0054651.

[7] V. Mueller et al., "The Stroop Competition: A Social-Evaluative Stroop Test for Acute Stress Induction," 2022 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI), Ioannina, Greece, 2022, pp. 1-4, doi: 10.1109/BHI56158.2022.9926835.