National College of Ireland

<BCs (Hons) in Computing>

<Data Analytics>

<Academic Year 2020/2021>

<Sergej Dikun>

<Student Number>

<x16150953@student.ncirl.ie>

<emCounter>

Final Technical Report

Contents

Executive Summary

In this research I have taken the existing APIs (of those available to public) and converted pre recorder volunteer speech to text, once this was done I have performed the speech analysis with the aid of Tableau visualisation. There is great potential in this research and while I've got a negative answer this time, I believe that further development of existing APIs (or development of a brand new – more precise API) would allow us to release such application and get all the advantage of its analysis.

1.0 Introduction

1.1. Background

I was always interested to see what the impact of the filler words on our speech is. I have not noticed that as much before I moved to Ireland, but once I did, I have started to notice an excessive use of "ummm", "emmm", "like" etc. in everyday talks between people.

1.2. Aims

With this project, I want to see if it is possible to perform an analysis based on recording made on a phone. The reason I have chosen the phone recording is to make it accessible to everyone. According to Statista in 2021 there will be 3.8 billion smartphones users and Dictaphone is one of the default features of the smartphone.

1.3. Technology

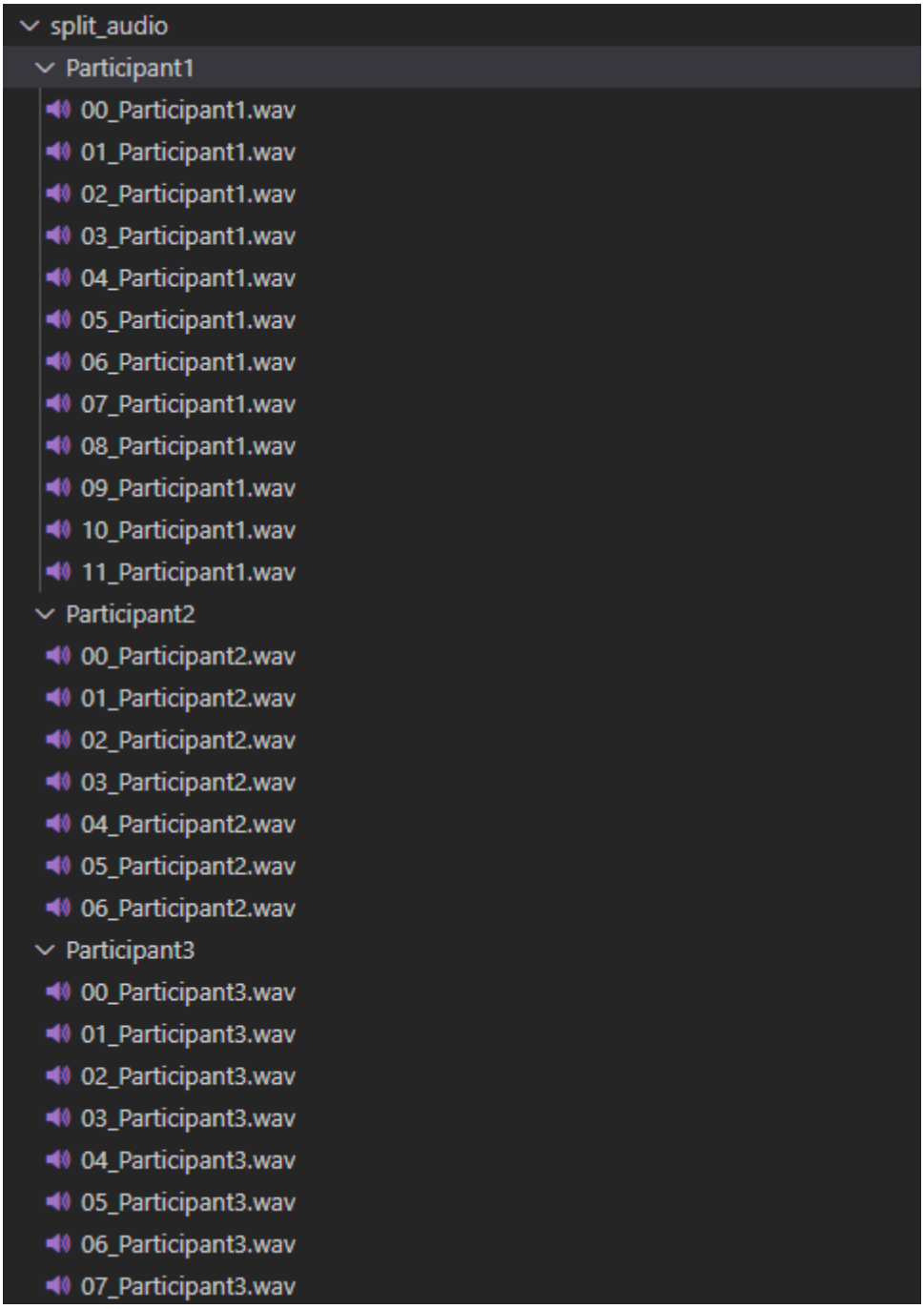
Technology use will be divided in to three stages:

- Data Gathering
  - Google Forms
  - Smartphone Dictaphone
- Data preparation (cleaning and converting)
  - Adobe Media Encoder - Converting all audio files to .wav format
  - Adobe Audition – Removing any personal data from audio file
  - Python – Converting the recording to text
- Data Analysis and Visualization
  - Python libraries – data analysis
  - Google API (Speech recognition and conversion to text)
  - Tableau – data visualization

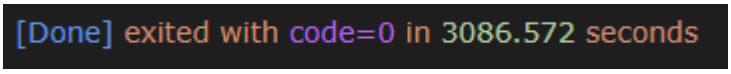
2.0 Data

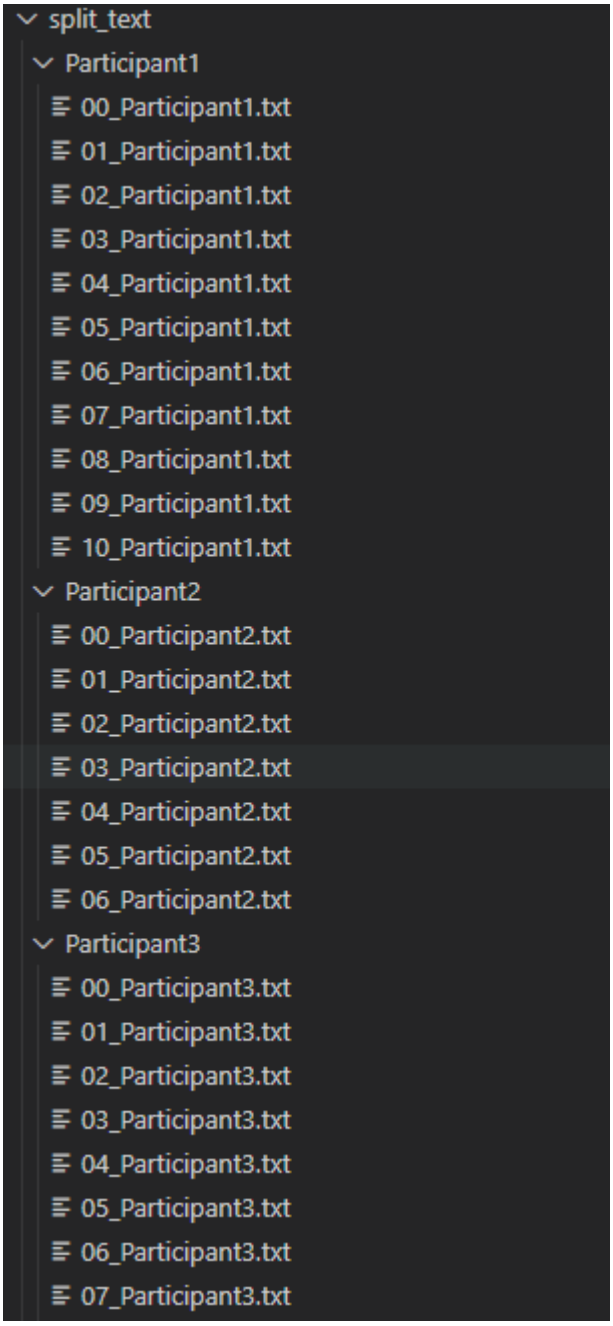
Data was gathered through the Google form application where volunteers have recorded their speech and uploaded an audio file to the link. Once it was uploaded, I would use Adobe Audition together with Adobe Media Converter to anonymise (delete any data that can be used to identify anyone) data and once it was anonymised I used the Python scripts that I wrote to:

1) Break down the recording of each participant to 1 minute segments (this is needed for the Google speech recognition to work as the free version only works on the 60 second recording and alternatives (AWS, Microsoft) only allow this to be done in their cloud platforms where the cost can vary.)

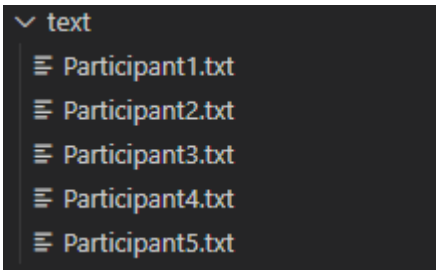


2) Extract speech from the audio recording to text file and place it in the designated folder. This part takes the most of the resource and time. In my particular case it took about 50 minutes to extract roughly 3 hours of text from an audio recording:

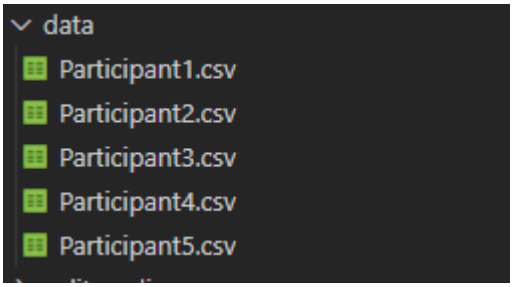


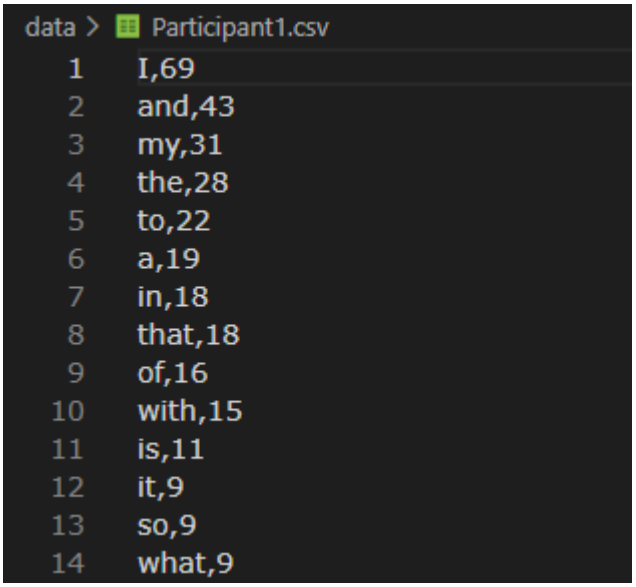


3) Concatenate the text files to a 1 big file for each of the participants in to one big file



4) Convert the file to .csv and with the use of the nltk library count the most popular words (in order to reduce the number of single words I took the top 100 for each of the participants)





3.0 Methodology

For this analysis we are using exploratory statistics as the question that we are trying to answer is if it is possible to perform speech analysis to identify filler words using this methodology. For that I have built dashboards in tableau for each of the participants to visualise and display the statistics of their speech.

4.0 Analysis

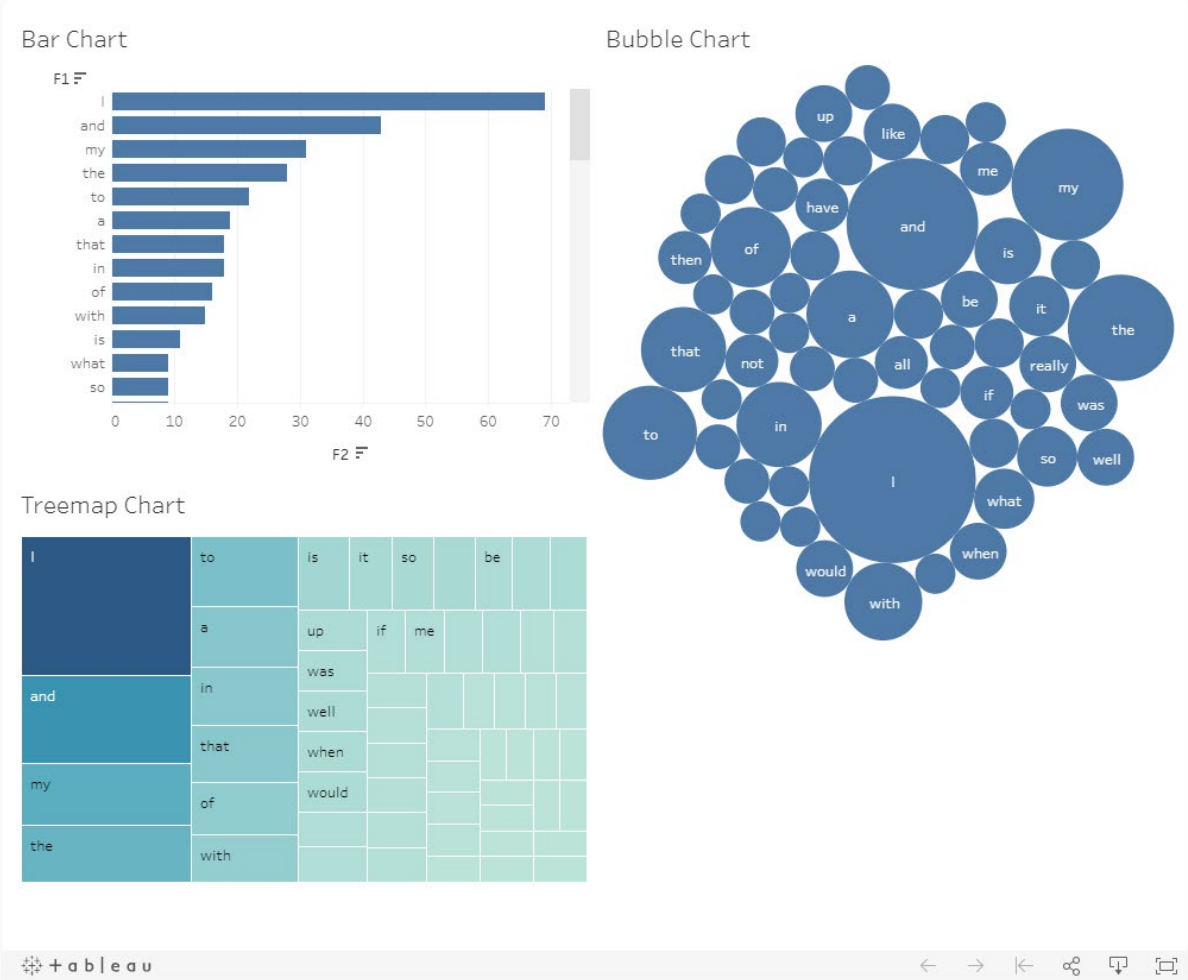
For the analysis I've explored the data in Tableau and the dashboards that I have built to identify the filler words.

For each of the participants, I have published the following interactive dashboards:

Participant 1:

Interactive dashboard link:

https://public.tableau.com/profile/sergej.dikun#!/vizhome/emCounter-Participant1/Dashboard1?publish=yes

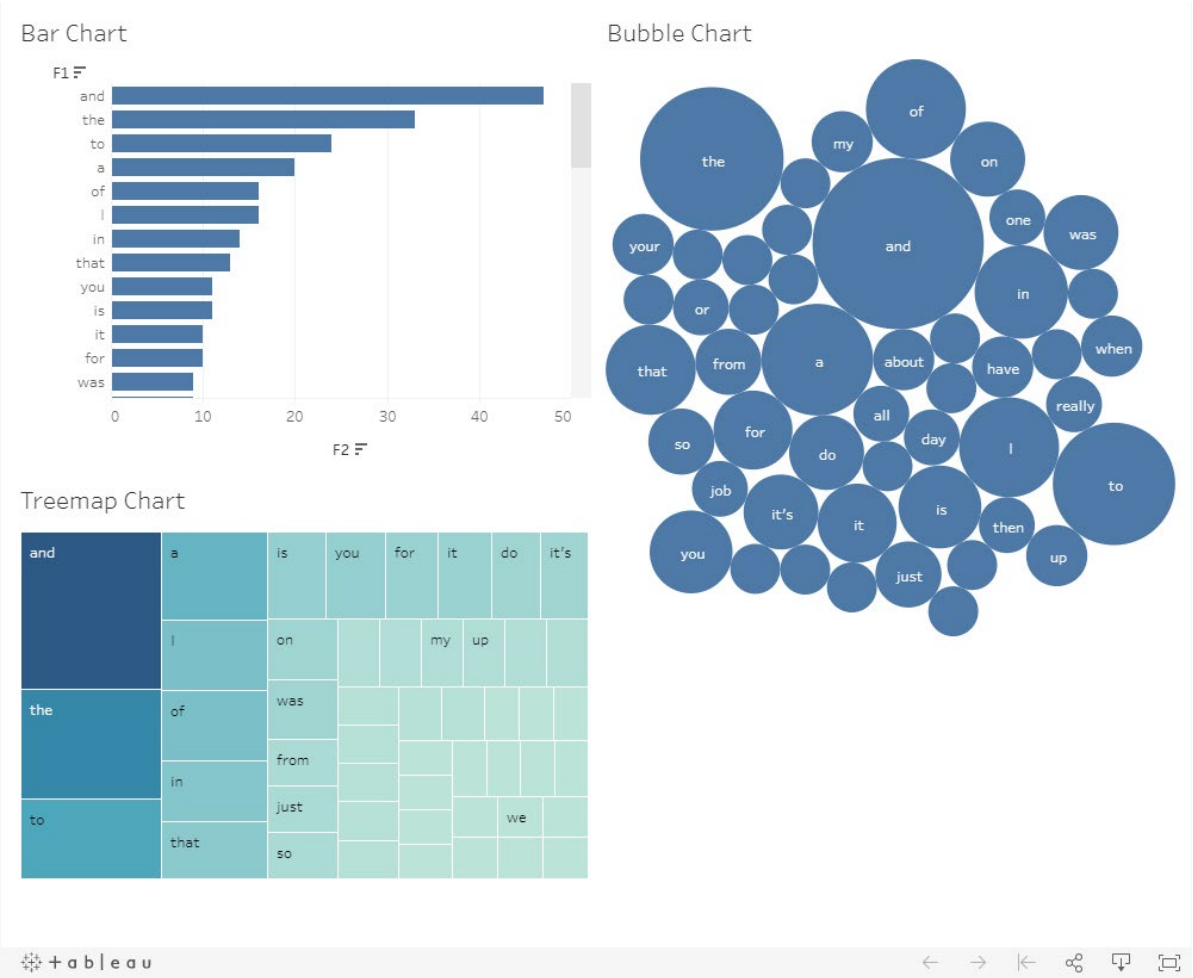


For this participant we can see that top 3 words used were “I”, “And” and “My”. I have not noticed any significant usage of any other words and can say that the participants speech was clear. Since I knew the participant personally, I have shared the finding and they have later shared that they have started to notice the “I” in the speech much more often which lead to further self evaluation thus improvement of the speech. I can proudly say that this application has helped that one individual.

Participant 2:

Interactive dashboard link:

https://public.tableau.com/profile/sergej.dikun#!/vizhome/emCounter-Participant2/Dashboard1?publish=yes

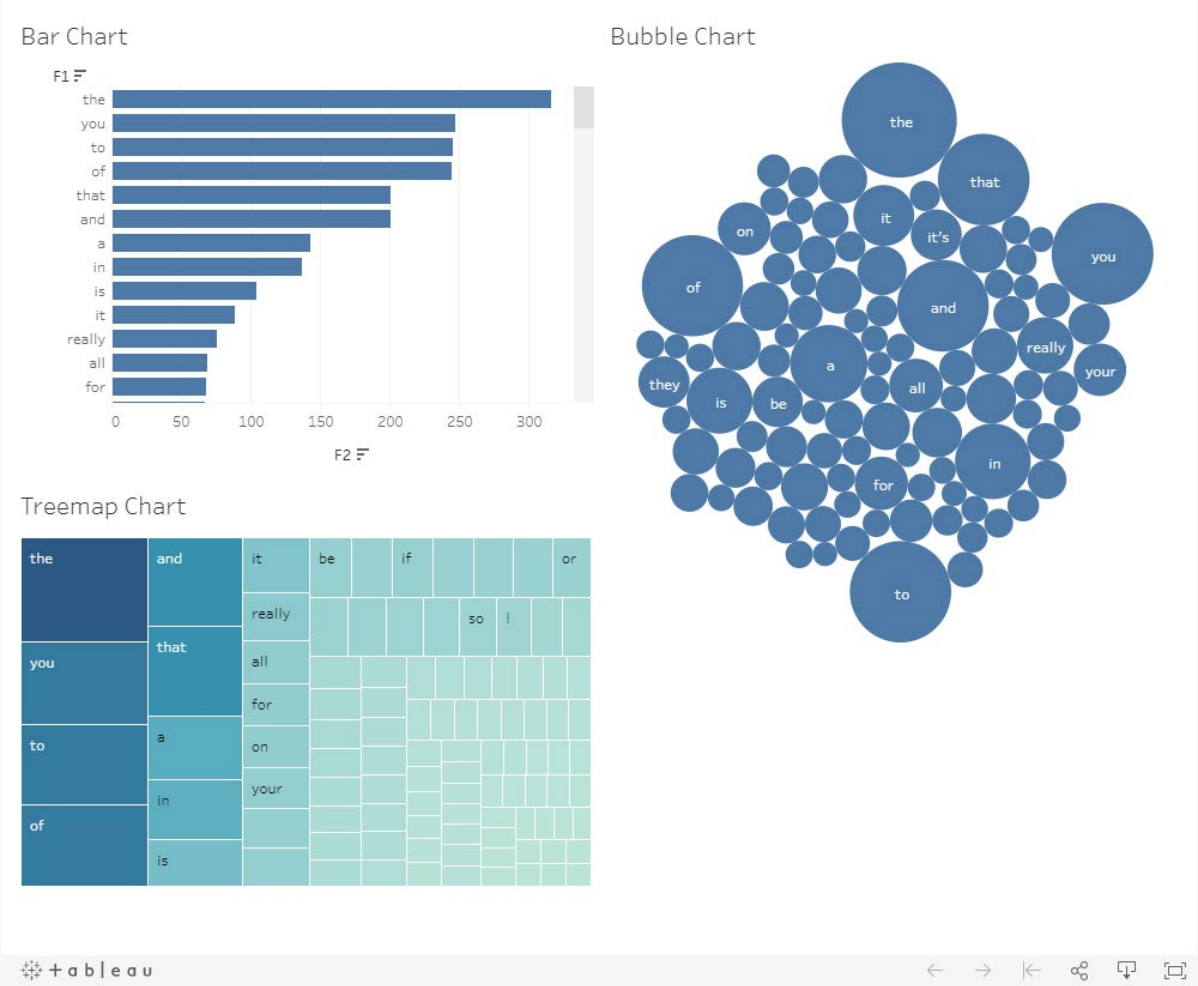


In this case we can see that the top words are “and” and “the”, it seems that in comparison to all other words “and” is being overused which could indicate that this can be a filler word. For this participant I would recommend further speech therapist consultation to check if there is an issue that can be worked on.

Participant 3:

Interactive dashboard link:

https://public.tableau.com/profile/sergej.dikun#!/vizhome/emCounter-Participant3/Dashboard1?publish=yes

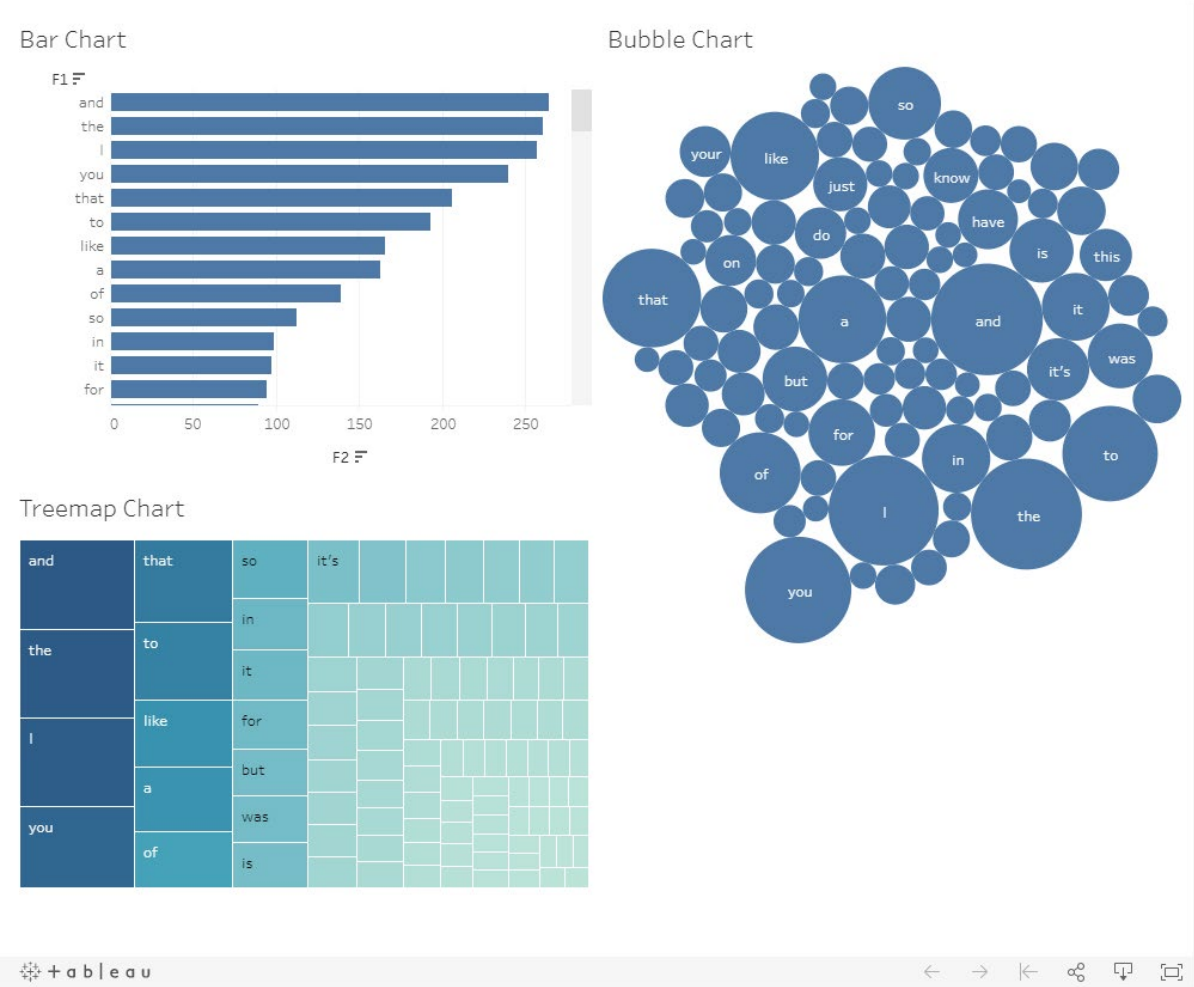


For this participant the top words are “the”, “you”, “to”, “of”. It seems like “the” stands out and is overused which could indicate the this is the filler word for that particular individual. I would suggest to consult with the speech therapist and work with then on correcting this.

Participant 4:

Interactive dashboard link:

https://public.tableau.com/profile/sergej.dikun#!/vizhome/emCounter-Participant4/Participant4?publish=yes

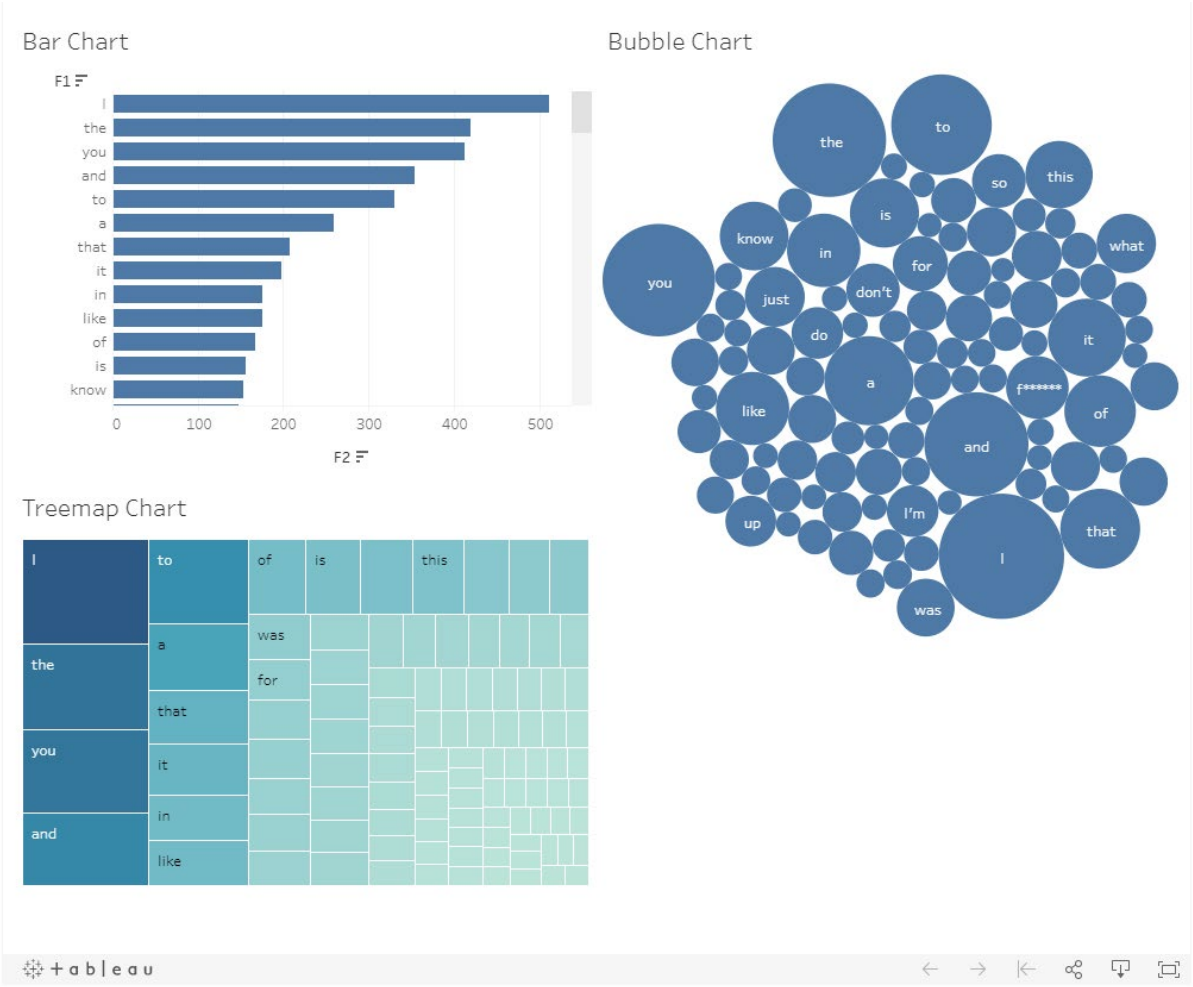


In this participants speech we can see same popular words as the other ones used, but we can also see that none of the words are standing out thus I can not claim that the participant is using any of the filler words. What I can claim is that the participants speech is clear.

Participant 5:

Interactive dashboard link:

https://public.tableau.com/profile/sergej.dikun#!/vizhome/emCounter-Participant5/Participant5?publish=yes



This participant is overusing “I” word, this may not mean that the participant is using it as a filler word, but it could mean that this participant is selfish. So perhaps not the speech therapy but a psychology consultation might be in order.

5.0 Results

Unfortunately for my research, currently existing algorithms are excluding “non real” world from the speech meaning that if someone uses “emm”, “umm”, “omm” in their speech, it will not be transferred to text thus it is not possible to get a complete analysis done. However, if one would use real words as fillers, then this would be easily identifiable as you can see in the examples above.

6.0 Conclusions

It is not possible to provide a full analysis of one’s speech with current technology, but further research can allow this to be done with AI implementation. Saying that, the application “as is” is near of a commercial usage as it is still providing valuable information to the participant. For

example, it turned out that the most common words that was used among all of the participants was "I". Is it good or not? This would be a different type of research, but it is definitely something I wish I would've found out before about my speech.

7.0 Further Development or Research

I can see this being used in medical field. This would include speech therapy and perhaps diagnose of the Alzheimer's (by analysing the speech and pauses between the speech).

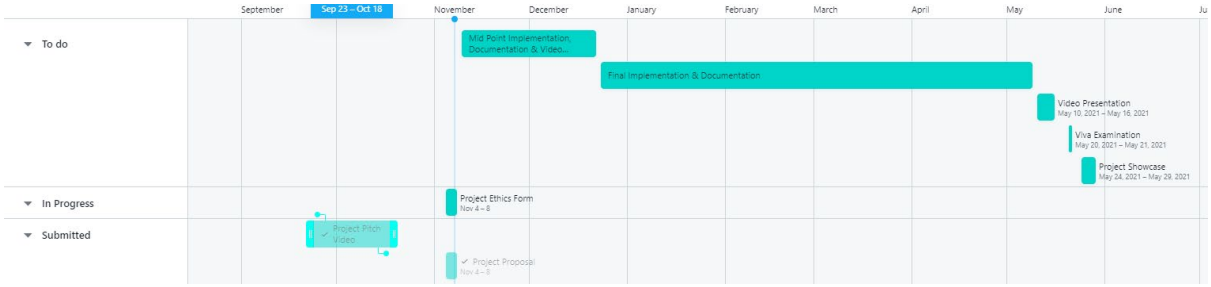
8.0 References

Statista. 2020. *Smartphone Users 2020 | Statista*. [online] Available at: <https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/> [Accessed 21 December 2020].

9.0 Appendices

This section should contain information that is supplementary to the main body of the report.

9.1. Project Plan



9.2. Ethics Approval Application (only if required)

Copy of the submitted proposal (submitted on November 7th 2020)

National College of Ireland

Human Participants Ethical Review Application Form

All parts of the below form must be completed. However in certain cases where sections are not relevant to the proposed study, clearly mark NA in the box provided.

Part A: Title of Project and Contact Information

Name

Sergej Dikun

Student Number (if applicable)

X16150953

Email

X16150953@student.ncirl.ie

Status:

Undergraduate ☐
Postgraduate ☐
Staff ☐

Supervisor (if applicable)

David O'Dwyer

Title of Research Project

emCounter

Category into which the proposed research falls (see guidelines)

Research Category A ☐
Research Category B ☐
Research Category C ☐

Have you read the NCI Ethical Guidelines for Research with Human Participants?

Yes ☐
No ☐

Please indicate any other ethical guidelines or codes of conduct you have consulted

GDPR

Has this research been submitted to any other research ethics committee?

Yes ☐
No ☐

If yes please provide details, and the outcomes of this process, if applicable:

Is this research supported by any form of research funding?

Yes ☐
No ☐

If yes please provide details, and indicate whether any restrictions exist on the freedom of the researcher to publish the results:

Part B: Research Proposal

Briefly outline the following information (not more than 200 words in any section).

Proposed starting date and duration of project

01/12/2020 duration of 3 months to gather enough data

The rationale for the project

Final Year Project

The research aims and objectives

Research aims to answer the question to whether it is possible to analyse speech through existing speech to text libraries in Python.

The research design

Participants will receive a few open ended questions. We will take the recordings, remove any personal data on them and then use them to analyse the speech (mainly to find the filler words used by the participant). We will however ask for participants age group to allocate them to an age group for better understanding of data.

The research sample and sample size
Please indicate the sample size and your justification of this sample size. Describe the age range of participants, and whether they belong to medical groups (those currently receiving medical treatment, those not in remission from previous medical treatment, those recruited because of a previous medical condition, healthy controls recruited for a medical study) or clinical groups (those undergoing non-medical treatment such as counselling, psychoanalysis, in treatment centres, rehabilitation centres, or similar, or those with a DSM disorder diagnosis).

18+ any and all participants are welcome. There are no specific requirements for any of the groups.

If the study involves a MEDICAL or CLINICAL group, the following details are required:

a) **Do you have approval from a hospital/medical/specialist ethics committee?**
   If YES, please append the letter of approval. Also required is a letter from a clinically responsible authority at the host institution, supporting the study, detailing the support mechanisms in place for individuals who may become distressed as a result of participating in the study, and the potential risk to participants.
   If NO, please detail why this approval cannot or has not been saught.
b) **Does the study impact on participant's medical condition, wellbeing, or health?**
   If YES, please append a letter of approval from a specialist ethics committee.
   If NO, please give a detailed explanation about why you do not expect there to be an impact on medical condition, wellbeing, or health.

The nature of any proposed pilot study. Pilot studies are usually required if a) a new intervention is being used, b) a new questionnaire, scale or item is being used, or c) established interventions or questionnaires, scales or items are being used on a new population. If no such study is planned, explain why it is not necessary.

The project can be used as a pilot study for future in depth research and further development.

The methods of data analysis. Give details here of the analytic process (e.g. the statistical procedures planned if quantitative, and the approach taken if qualitative. It is not sufficient to name the software to be used).

We have not covered that in our studies yet so I will not be able to give a full answer, but I will be using data visualisation and will be builting plots to display the findings.

Study Procedure

Please give as detailed an account as possible of a participant's likely experience in engaging with the study, from point of first learning about the study, to study completion. State how long project participation is likely to take, and whether participants will be offered breaks. Please attach all questionnaires, interview schedules, scales, surveys, and demographic questions, etc. in the Appendix.

In order to have a better quality of data I will inform participants that it will be a speech analysis, I will also inform them that their names or any details will not be gathered and thus disclosed anywhere. After that, they will record their answers to few open-ended questions and send the recording to me either through email or through virtual drive. Once I have the recording, I will clean it of any personal data and then run it through analytics tool to add it to the dataset.

Part C: Ethical Risk

Please identify any ethical issues or risks of harm or distress which may arise during the proposed research, and how you will address this risk. Here you need to consider the potential for physical risk, social risk (i.e. loss of social status, privacy, or reputation), outside of that expected in everyday life, and whether the participant is likely to feel distress as a result of taking part in the study. Debriefing sheets must be included in the appendix if required. These should detail the participant's right to withdraw from the study, the statutory limits upon confidentiality, and the obligations of the researcher in relation to Freedom of Information legislation. Debriefing sheets should also include details of helplines and avenues for receiving support in the event that participants become distressed as a result of their involvement in this study.

Participants will make the recordings from the comfort of their home and will be able to choose not to send the recording if they wish so.

Do the participants belong to any of the following vulnerable groups?
(Please tick all those involved).

- □ Children;
- □ The older old (85+)
- □ People with an intellectual or learning disability
- □ Individuals or groups receiving help through the voluntary sector
- □ Those in a subordinate position to the researchers such as employees
- □ Other groups who might not understand the research and consent process
- □ Other vulnerable groups

How will the research participants in this study be selected, approached and recruited? From where will participants be recruited? If recruiting via an institution or organisation other than NCI please attach a letter of agreement from the host institution agreeing to host the study and circulate recruitment advertisements/email etc.

Anyone who will volunteer to participate and will have a smartphone that can record sound (or video, as I can extract audio from it as well)

What inclusion or exclusion criteria will be used?

I will not know anything about the participants except for their age group thus the only exclusion will be minors.

How will participants be informed of the nature of the study and participation?

Through the online form.

Does the study involve deception or the withholding of information? If so, provide justification for this decision.

n/a

What procedures will be used to document the participants' consent to participate?

I will create a google form with conditions of participation to which they will need to agree.

Can study participants withdraw at any time without penalty? If so, how will this be communicated to participants?

Yes, they can withdraw at any time. I expect that no more than 50% of people who will sign up will actually send their recordings.

If vulnerable groups are participating, what special arrangements will be made to deal with issues of informed consent/assent?

Please include copies of any information letters, debriefing sheets, and consent forms with the application.

Part D: Confidentiality and Data Protection

Please indicate the form in which the data will be collected.

□ Identified □ Potentially Identifiable **□ De-Identified**

What arrangements are in place to ensure that the identity of participants is protected?

No such information will be gathered, if participants will include any information, it will be deleted.

Will any information about illegal behaviours be collected as part of the research process? If so, detail your consideration of how this information will be treated.

No

Please indicate any recording devices being used to collect data (e.g. audio/video).

Audio (inbuilt audio recording device in any smart phone)

Please describe the procedures for securing specific permission for the use of these recording devices in advance.

Audio recordings will be deleted after data will be extracted from them

Please indicate the form in which the data will be stored.

□ Identified □ Potentially Identifiable **□ De-Identified**

Who will have responsibility for the data generated by the research?

I will go through data and make sure that it does not contain any personal information.

Is there a possibility that the data will be archived for secondary data analysis? If so, has this been included in the informed consent process? Also include information on how and where the data will be stored for secondary analytic purposes.

No, data will not be kept for econdary analyss, new research will require new data collection.

If not to be stored for secondary data analysis, will the data be stored for 5 years and then destroyed, in accordance with NCI policy?

□ Yes □ No

Dissemination and Reporting

Please describe how the participants will be informed of dissemination and reporting (e.g. submission for examination, reporting, publications, presentations)?

They will be informed in advance that data is for analysis in college project.

If any dissemination entails the use of audio, video and/or photographic records (including direct quotes), please describe how participants will be informed of this in advance.

n/a

Part E: Signed Declaration

I confirm that I have read the NCI Ethical Guidelines for Research with Human Participants, and agree to abide by them in conducting this research. I also confirm that the information provided on this form is correct (Electronic signature is acceptable).

Signature of Applicant __Sergej Dikun_

Date __24/05/2021__

Signature of Supervisor (where appropriate):

Date _____

Any other information the committee should be aware of?

9.3. Reflective Journals

10.0 Objectives

With the project we are trying to find the answer if it is possible to analyse speech of the individual through the voice to text methodology. I will invite individuals to record their answers to the open-ended questions. None of the questions will have any personal data and will only be used to analyse the speech. We will need their age just for the better sorting to appropriate age group.

11.0 Background

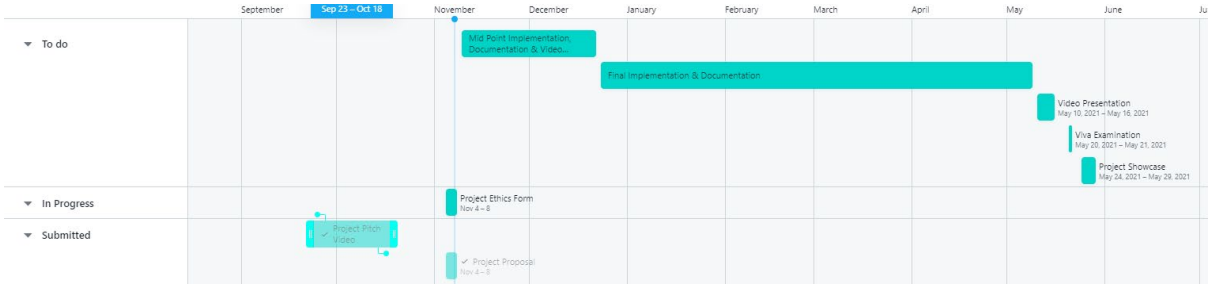
Initially the idea came to me when I was listening to a teenager speaking on the radio and saying the word “like” nearly after each other word. While this was a bit funny, it was also interesting to find out how would the speech change should we delete that word from the conversation. After that the idea was evolving as I have started to notice the filler words in nearly every conversation. Some would use it more often, some would use it less. Some time has passed and the idea evolved in to something that would potentially help speech therapists around the world to use it as a tool that would help to display people the issues that they have when they speak and how. I have proposed this as an idea in my video project pitch and it was accepted. Initially the question was about analysis itself (as in to detect the actual percentage of the filler words was being used) but after the conversation with the supervisor we agreed that better phrasing of the question would be if it is possible to do such analysis using current technologies.

12.0 Technical Approach

For this project we will require multiple devices as well as some of the software to be used in addition to coding and analysis methodologies.

1. Recording devices. It was initially considered to create an application to record the sound, but it may be to much work for something that is already available on the market thus the decision is to use phones recording abilities to record the sound.
2. After sound is created, participants will have a choice to either email the file to me or to upload it to the shared drive.
3. Once I will receive the file I will need to make sure that the file is anonymised and no personal data is on it. For that I will use Adobe Audition software. (I have found an Audio Analysis Library https://github.com/tyiannak/pyAudioAnalysis but for now I will keep it as optional as I want to make sure that Audio to Text is set and works before I try out the Audio Analysis. Please consider is as an option for future development)
4. Once data is clear, I will use Python to convert the data from Audio to text and then to sort/clean the data to get it ready for analysis.
5. In addition to Python, I will use Tableau for the visualisation of the data.

13.0 Project Plan



14.0 Technical Details

Programming language that will be used for this project: Python

Libraries:

- SpeechRecognition https://pypi.org/project/SpeechRecognition/
- Pandas https://pandas.pydata.org/

15.0 Evaluation

The system will need to successfully capture the filler words, recording will also be manually checked to make sure that those are being captured correctly. Once the data is collected and be sufficient for analysis, I will consider it as successful.