

Parkinsons Disease

@Sarah Wilson

2021/04/06

Introduction

Parkinson's disease is a neurodegenerative disorder. Individuals with Parkinson's disease often suffer from stiffness, shaking, difficulty with coordination, and as the disease progresses it starts to impact aspects of speech. Specifically, the aspects of speech commonly affected by Parkinson's disease are hypophonia (lower volume), dysarthria (difficulty articulating words) and monotone (reduced pitch range).

Currently, Parkinson's disease is not diagnosed by a single test. Diagnosis is based upon medical history, symptoms and a neurological and physical examination which are all heavily based on motor functions. However, at the early stages of Parkinson's disease, it can be hard to detect motor deficits as they may not be severe. The use of an alternative biomarker, such as aspects of speech, may aid early diagnosis. Little et al. (2007) investigated this topic by assessing if two novel approaches (recurrence probability density entropy (RPDE) and Detrended Fluctuation Analysis (DFA)) which measure aspects of speech could help distinguish between participants with and without Parkinson's disease. The findings shown these novel approaches were able to classify individuals with and without Parkinson's disease better than traditional approaches of assessing aspects of speech.

The underlying importance of this research is to identifying better techniques to assess biomarkers of Parkinson's disease, so patient care can be improved and a better understanding of the disease can be gained. As I am greatly interested in neurodegenerative diseases and patient care, I have chosen to analyse and produce a visualisation for the data set produced by Little et al. (2007). This data set was retrieved from [Kaggle](#) authored by Debasis Samal based on the work by Little et al. (2007).

Little, M.A., McSharry, P.E., Roberts, S.J., Costello, D.A.E., Moroz, I.M. (2007). Exploiting Nonlinear Recurrence and Fractal Scaling Properties for Voice Disorder Detection. *BioMedical Engineering OnLine*, 6(23). <https://doi.org/10.1186/1475-925X-6-23>

Research Question for this Project

This data analysis and visualisation project will reassess if one of the new approaches (RPDE) tested by Little et al. (2007) can visually demonstrate differences between those with and without Parkinson's disease.

Preparation for the Visualisation

```
#set directory
setwd("C:/Parkinson-s-Disease-using-speech-as-a-biomarker-
main/wilson_psy64222/Parkinsosns disease detection")

#Load packages
library(readr)
library(tidyverse)

## -- Attaching packages ----- tidyverse
1.3.0 --

## v ggplot2 3.3.3      v dplyr  1.0.4
## v tibble  3.0.6      v stringr 1.4.0
## v tidyr   1.1.2      v forcats 0.5.1
## v purrr   0.3.4

## -- Conflicts -----
tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()

library(here) #built under R version 4.0.4

## here() starts at C:/Parkinson-s-Disease-using-speech-as-a-biomarker-
main/wilson_psy64222/Parkinsosns disease detection

#Load data
parkinsons_data <- read_csv("data/parkinsons_data.csv")

##
## -- Column specification -----
-----
## cols(
##   .default = col_double(),
##   name = col_character()
## )
## i Use `spec()` for the full column specifications.

head(parkinsons_data, 5) #first 5 rows of the data

## # A tibble: 5 x 24
##   name `MDVP:Fo(Hz)` `MDVP:Fhi(Hz)` `MDVP:Flo(Hz)` `MDVP:Jitter(%)`
##   <chr>      <dbl>      <dbl>      <dbl>      <dbl>
## 1 phon~      120.      157.      75.0      0.00784
## 2 phon~      122.      149.      114.      0.00968
## 3 phon~      117.      131.      112.      0.0105
## 4 phon~      117.      138.      111.      0.00997
## 5 phon~      116.      142.      111.      0.0128
## # ... with 19 more variables: `MDVP:Jitter(Abs)` <dbl>, `MDVP:RAP` <dbl>,
## # `MDVP:PPQ` <dbl>, `Jitter:DDP` <dbl>, `MDVP:Shimmer` <dbl>,
```

```
## # `MDVP:Shimmer(dB)` <dbl>, `Shimmer:APQ3` <dbl>, `Shimmer:APQ5` <dbl>,
## # `MDVP:APQ` <dbl>, `Shimmer:DDA` <dbl>, NHR <dbl>, HNR <dbl>, status
<dbl>,
## # RPDE <dbl>, DFA <dbl>, spread1 <dbl>, spread2 <dbl>, D2 <dbl>, PPE
<dbl>
```

All the measures assessed by Little et al. (2007) can be seen in the data denoted by their abbreviations (see output above). As these abbreviations do not tell us what the measure assesses, I have renamed these variables to aid understanding of which aspect of speech these assessments measure.

```
#rename the data variables
parkinsons_data <- parkinsons_data %>%
  rename(average_vocal_frequency = `MDVP:F0(Hz)`,
         max_vocal_frequency = `MDVP:Fhi(Hz)`,
         min_vocal_frequency = `MDVP:Flo(Hz)`,
         diagnosis = status,
         fundamental_frequency = `Jitter:DDP`,
         amp = `Shimmer:DDA`,
         noise_tonal = NHR,
         complexity = RPDE,
         signal_frac_scale = DFA,
         nonlinear_fundamental_frequency = spread2
  )

#rename name to ppt_id by first reorganizing the information it holds.

#removing and separating out the strings in name
parkinsons_data <- parkinsons_data %>%
  separate(name, c(NA, NA, "group", "ppt"), sep = "_")

#recombining the third and forth string to make participant id
parkinsons_data <- unite(parkinsons_data,
                        ppt_id,
                        "group", "ppt",
                        sep = "_",
                        remove = TRUE,
                        na.rm = FALSE)
```

More detail of the renamed variables can be found in the code book accessible via the file path “~/wilson_psy64222/Parkinsons disease detection/scripts/codebook.pdf”.

Not all variables in this data set have been renamed. This is because some of the variables measure the same aspect of speech as other variables. So, I have decided to pick one measure for each aspect of speech and to remove the rest from the data set.

```
#selecting the variables I have renamed
parkinsons_data <- parkinsons_data %>%
  select(ppt_id,
         diagnosis,
```

```

    average_vocal_frequency,
    max_vocal_frequency,
    min_vocal_frequency,
    fundamental_frequency,
    amp,
    noise_tonal,
    complexity,
    signal_frac_scale,
    nonlinear_fundamental_frequency)
head(parkinsons_data, 5) #first 5 rows of data after rename and removal

## # A tibble: 5 x 11
##   ppt_id diagnosis average_vocal_f~ max_vocal_frequ~ min_vocal_frequ~
##   <chr>      <dbl>          <dbl>          <dbl>          <dbl>
## 1 S01_1         1          120.          157.           75.0
## 2 S01_2         1          122.          149.           114.
## 3 S01_3         1          117.          131.           112.
## 4 S01_4         1          117.          138.           111.
## 5 S01_5         1          116.          142.           111.
## # ... with 6 more variables: fundamental_frequency <dbl>, amp <dbl>,
## #   noise_tonal <dbl>, complexity <dbl>, signal_frac_scale <dbl>,
## #   nonlinear_fundamental_frequency <dbl>

```

The Visualisation

First I have made a stable graph showing the difference in average fundamental frequency by complexity in those with and without Parkinson's disease.

```

#visualizing the data
p <- ggplot(data = (parkinsons_data %>% group_by(diagnosis == 1)),
  aes(x = parkinsons_data$average_vocal_frequency,
    y = parkinsons_data$complexity,
    colour = factor(diagnosis == 0,
      labels = c("Healthy",
        "Parkinson's Disease"))))

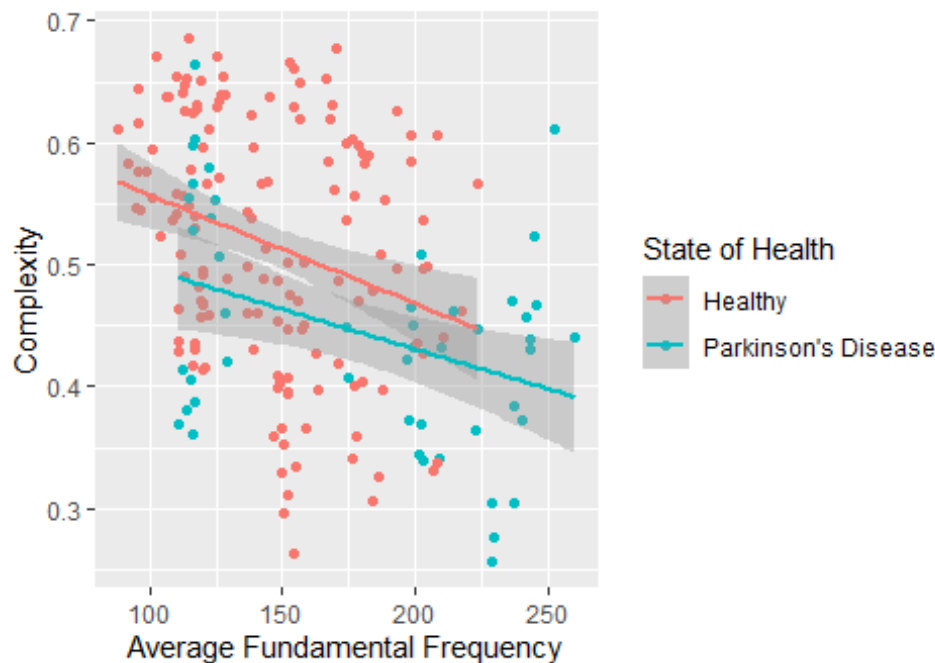
p +geom_point() +
  geom_smooth(method = "lm") +
  labs(title = "Effects of Parkinsons Disease on Aspects of Speech",
    subtitle = "Plot of Average Frequency by Complexity of Speech",
    x = "Average Fundamental Frequency",
    y = "Complexity",
    colour = "State of Health")

## `geom_smooth()` using formula 'y ~ x'

```

Effects of Parkinsons Disease on Aspects of Speech

Plot of Average Frequency by Complexity of Speech



```
#save the graph
ggsave(filename = "static_plot.jpg",
        plot = last_plot(),
        path = "C:/Parkinson-s-Disease-using-speech-as-a-biomarker-
main/wilson_psy64222/Parkinsons disease detection/figs")

## Saving 5 x 4 in image
## `geom_smooth()` using formula 'y ~ x'
```

I have then built on this visualisation by adding an animation effect, turning this stable graph into a gif.

```
#packages needed for animation
install.packages("gganimate", repos = "http://cran.us.r-project.org")

## Installing package into 'C:/Users/wisar/OneDrive/Documents/R/win-
library/4.0'
## (as 'lib' is unspecified)

## package 'gganimate' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\wisar\AppData\Local\Temp\RtmpkRUK17\downloaded_packages
install.packages("gifski", repos = "http://cran.us.r-project.org")
```

```

## Installing package into 'C:/Users/wisar/OneDrive/Documents/R/win-
library/4.0'
## (as 'lib' is unspecified)

## package 'gifski' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\wisar\AppData\Local\Temp\RtmpkRUK17\downloaded_packages

install.packages("av", repos = "http://cran.us.r-project.org")

## Installing package into 'C:/Users/wisar/OneDrive/Documents/R/win-
library/4.0'
## (as 'lib' is unspecified)

## package 'av' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\wisar\AppData\Local\Temp\RtmpkRUK17\downloaded_packages

#Load the packages
library(gganimate)

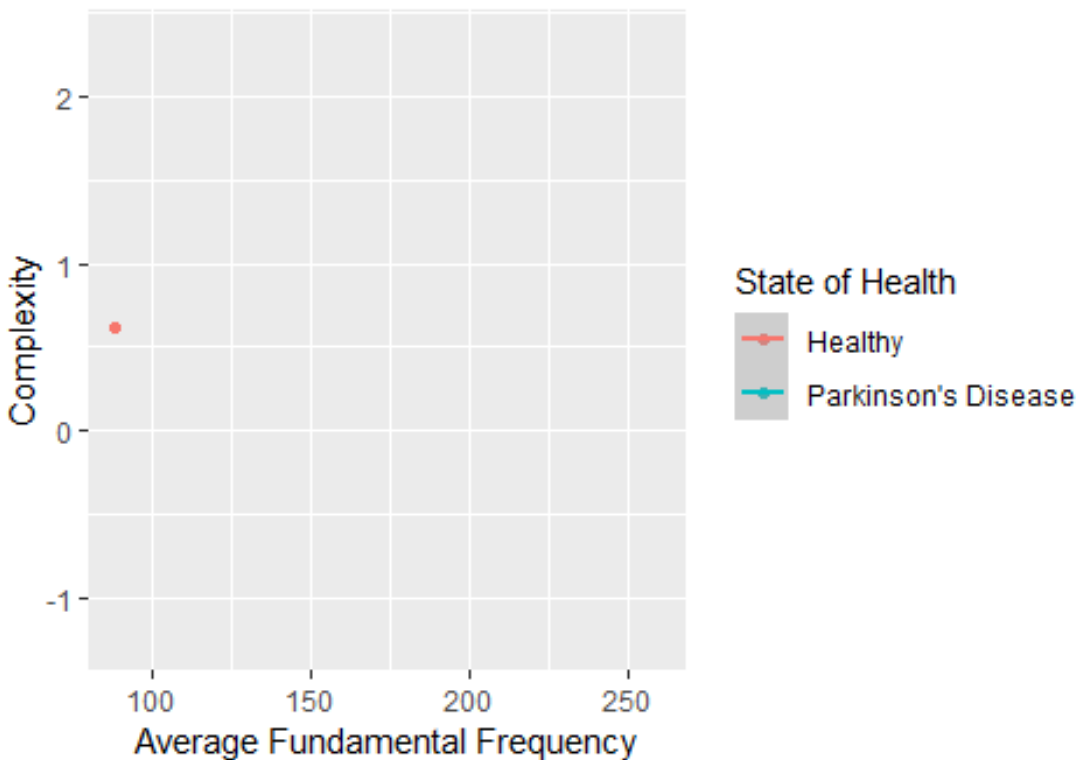
#the animated visualization
p + transition_reveal(parkinsons_data$average_vocal_frequency) +
  geom_point() +
  geom_smooth(method = "lm") +
  labs(title = "Effects of Parkinsons Disease on Aspects of Speech",
        subtitle = "Plot of Average Frequency by Complexity of Speech",
        x = "Average Fundamental Frequency",
        y = "Complexity",
        colour = "State of Health")

## `geom_smooth()` using formula 'y ~ x'

```

Effects of Parkinsons Disease on Aspects of Speech

Plot of Average Frequency by Complexity of Speech



```
#save the animated graph
anim_save(filename = "animated_plot.gif",
          plot = last_plot(),
          path = "C:/Parkinson-s-Disease-using-speech-as-a-biomarker-
main/wilson_psy64222/Parkinsosns disease detection/figs") #saved in figs
file.
```

If the animated graph does not play in this PDF, please see the file path “C:/Parkinson-s-Disease-using-speech-as-a-biomarker-main/wilson_psy64222/Parkinsosns disease detection/figs” and the image is called aimed_plot.

Summary

The graphs above show participants without Parkinson’s Disease (labeled as healthy) have greater non linear dynamical complexity in their speech compared to participants with Parkinsons Disease. Although, in the stable graph the line of best fit shows clear distinctions between participants with and without Parkinson’s disease however this clear distinction disappears when looking at the animated graph. Additionally, looking at the animated graph, variation is greater among those without Parkinson’s disease compared to participants with Parkinson’s disease. This reduced variation may aid diagnosis if set values can be determined. Future research could therefore investigate if using RPDE to

measure a non linear dynamical complexity is a viable tool to implement in Parkinson's disease diagnosis.

One issue to raise about this data set is that it did not come directly from the paper published by Little et al. (2007). Therefore, the data set used in this project may be invalid due to errors produced by Debasis Samal (author of the document on Kaggle).