voice

easier audio analysis for digital phenotyping

Filipe J. Zabala

Supervisor: Giovanni A. Salum

Graduate Program in Psychiatry
Medical School · UFRGS

2025-09-16

Summary

1.Mental health

1. Mental health

- Some key challenges identified in literature

1. Mental health

- Some key challenges identified in literature
  - Heterogeneity of disorders

1. Mental health

- Some key challenges identified in literature
    - Heterogeneity of disorders
    - Lack of biomarkers

1. Mental health

- Some key challenges identified in literature
    - Heterogeneity of disorders
    - Lack of biomarkers
    - Subjective symptom reporting

1. Mental health

- Some key challenges identified in literature
  - Heterogeneity of disorders
  - Lack of biomarkers
  - Subjective symptom reporting
  - Categorical vs. dimensional issues

1. Mental health

- Some key challenges identified in literature
  - Heterogeneity of disorders
  - Lack of biomarkers
  - Subjective symptom reporting
  - Categorical vs. dimensional issues
- Phenotype quantification difficulty

1. Mental health

- Some key challenges identified in literature
  - Heterogeneity of disorders
  - Lack of biomarkers
  - Subjective symptom reporting
  - Categorical vs. dimensional issues
- Phenotype quantification difficulty
  - A significant barrier in mental health research

1. Mental health

- Some key challenges identified in literature
  - Heterogeneity of disorders
  - Lack of biomarkers
  - Subjective symptom reporting
  - Categorical vs. dimensional issues
- Phenotype quantification difficulty
  - A significant barrier in mental health research
- Current standard methods

1. Mental health

- Some key challenges identified in literature
  - Heterogeneity of disorders
  - Lack of biomarkers
  - Subjective symptom reporting
  - Categorical vs. dimensional issues
- Phenotype quantification difficulty
  - A significant barrier in mental health research
- Current standard methods
  - Primarily questionnaires and interviews

1. Mental health

- Some key challenges identified in literature
  - Heterogeneity of disorders
  - Lack of biomarkers
  - Subjective symptom reporting
  - Categorical vs. dimensional issues
- Phenotype quantification difficulty
  - A significant barrier in mental health research
- Current standard methods
  - Primarily questionnaires and interviews
- Emerging alternative

1. Mental health

- Some key challenges identified in literature
  - Heterogeneity of disorders
  - Lack of biomarkers
  - Subjective symptom reporting
  - Categorical vs. dimensional issues
- Phenotype quantification difficulty
  - A significant barrier in mental health research
- Current standard methods
  - Primarily questionnaires and interviews
- Emerging alternative
  - Daily life data, such as voice analysis

1. Mental health

- Some key challenges identified in literature
  - Heterogeneity of disorders
  - Lack of biomarkers
  - Subjective symptom reporting
  - Categorical vs. dimensional issues
- Phenotype quantification difficulty
  - A significant barrier in mental health research
- Current standard methods
  - Primarily questionnaires and interviews
- Emerging alternative
  - Daily life data, such as voice analysis
- Key advantages

1. Mental health

- Some key challenges identified in literature
  - Heterogeneity of disorders
  - Lack of biomarkers
  - Subjective symptom reporting
  - Categorical vs. dimensional issues
- Phenotype quantification difficulty
  - A significant barrier in mental health research
- Current standard methods
  - Primarily questionnaires and interviews
- Emerging alternative
  - Daily life data, such as voice analysis
- Key advantages
  - Non-invasive

1. Mental health

- Some key challenges identified in literature
  - Heterogeneity of disorders
  - Lack of biomarkers
  - Subjective symptom reporting
  - Categorical vs. dimensional issues
- Phenotype quantification difficulty
  - A significant barrier in mental health research
- Current standard methods
  - Primarily questionnaires and interviews
- Emerging alternative
  - Daily life data, such as voice analysis
- Key advantages
  - Non-invasive
  - Objective

1. Mental health

- Some key challenges identified in literature
  - Heterogeneity of disorders
  - Lack of biomarkers
  - Subjective symptom reporting
  - Categorical vs. dimensional issues
- Phenotype quantification difficulty
  - A significant barrier in mental health research
- Current standard methods
  - Primarily questionnaires and interviews
- Emerging alternative
  - Daily life data, such as voice analysis
- Key advantages
  - Non-invasive
  - Objective
  - Programmable

1. Mental health

- Some key challenges identified in literature
  - Heterogeneity of disorders
  - Lack of biomarkers
  - Subjective symptom reporting
  - Categorical vs. dimensional issues
- Phenotype quantification difficulty
  - A significant barrier in mental health research
- Current standard methods
  - Primarily questionnaires and interviews
- Emerging alternative
  - Daily life data, such as voice analysis
- Key advantages
  - Non-invasive
  - Objective
  - Programmable
  - Verifiable

1. Mental health

- Some key challenges identified in literature
  - Heterogeneity of disorders
  - Lack of biomarkers
  - Subjective symptom reporting
  - Categorical vs. dimensional issues
- Phenotype quantification difficulty
  - A significant barrier in mental health research
- Current standard methods
  - Primarily questionnaires and interviews
- Emerging alternative
  - Daily life data, such as voice analysis
- Key advantages
  - Non-invasive
  - Objective
  - Programmable
  - Verifiable
  - Scalable

Definitions

- Emotion

Definitions

- Emotion
  - An episode of interrelated, synchronized changes in the states of all or most of the five organismic subsystems in response to the evaluation of an external or internal stimulus event as relevant to major concerns of the organism (Scherer [2005])

Definitions

- Emotion
  - An episode of interrelated, synchronized changes in the states of all or most of the five organismic subsystems in response to the evaluation of an external or internal stimulus event as relevant to major concerns of the organism (Scherer [2005])
- Emotional valence

Definitions

- Emotion
  - An episode of interrelated, synchronized changes in the states of all or most of the five organismic subsystems in response to the evaluation of an external or internal stimulus event as relevant to major concerns of the organism (Scherer [2005])
- Emotional valence
  - The positive or negative quality of an emotion

2. Why another package

2. Why another package

Package	Rec/play	Conversion	Photo/Video	Visualization	Features	Sheet music	Remove vocals	Chords/scales	Diarization
audio	✓								
av	✓	✓	✓	✓					
EMU-SDMS	✓	✓		✓	✓				
gm	✓			✓		✓		✓	
karaoke							✓		
music	✓			✓				✓	
seewave	✓	✓		✓	✓				
signal		✓		✓	✓				
sound	✓								
tabr	✓			✓		✓			
tuneR				✓	✓				
voice		✓		✓	✓	✓		✓	✓
voiceR								✓	
wrassp					✓				
BirdNET-Analyzer	✓	✓		✓	✓				
Librosa				✓	✓			✓	
Parselmouth	✓	✓		✓	✓				
pyannote-audio	✓	✓		✓	✓				✓
SpeechBrain	✓	✓		✓	✓				✓

Some R and Python packages for handling audio

2. Why another package

Package	Rec/play	Conversion	Photo/Video	Visualization	Features	Sheet music	Remove vocals	Chords/scales	Diarization
audio	✓								
av	✓	✓	✓	✓					
EMU-SDMS	✓	✓		✓	✓				
gm	✓			✓		✓		✓	
karaoke							✓		
music	✓			✓				✓	
seewave	✓	✓		✓	✓				
signal		✓		✓	✓				
sound	✓								
tabr	✓			✓		✓			
tuneR				✓	✓				
voice		✓		✓	✓	✓		✓	✓
voiceR								✓	
wrassp					✓				
BirdNET-Analyzer	✓	✓		✓	✓				
Librosa				✓	✓			✓	
Parselmouth	✓	✓		✓	✓				
pyannote-audio	✓	✓		✓	✓				✓
SpeechBrain	✓	✓		✓	✓				✓

Some R and Python packages for handling audio

2. Why another package

- Context

2. Why another package

- Context
  - Vocal features extraction

2. Why another package

- Context
    - Vocal features extraction
    - Audio files in various formats and quality levels

2. Why another package

- Context
  - Vocal features extraction
  - Audio files in various formats and quality levels
  - Technical

2. Why another package

- Context
  - Vocal features extraction
  - Audio files in various formats and quality levels
  - Technical
  - Labor-intensive

2. Why another package

- Context
    - Vocal features extraction
    - Audio files in various formats and quality levels
    - Technical
    - Labor-intensive
    - Code-heavy

2. Why another package

- Context
  - Vocal features extraction
  - Audio files in various formats and quality levels
  - Technical
  - Labor-intensive
  - Code-heavy
- Methodology

2. Why another package

- Context
  - Vocal features extraction
  - Audio files in various formats and quality levels
  - Technical
  - Labor-intensive
  - Code-heavy
- Methodology
  - `voice` package $\rightarrow$ Model-ready dataset

2. Why another package

- Context
  - Vocal features extraction
  - Audio files in various formats and quality levels
  - Technical
  - Labor-intensive
  - Code-heavy
- Methodology
  - `voice` package $\rightarrow$ Model-ready dataset
  - User-friendly

2. Why another package

- Context
  - Vocal features extraction
  - Audio files in various formats and quality levels
  - Technical
  - Labor-intensive
  - Code-heavy
- Methodology
  - `voice` package $\rightarrow$ Model-ready dataset
  - User-friendly
  - Streamlined

2. Why another package

- Context
  - Vocal features extraction
  - Audio files in various formats and quality levels
  - Technical
  - Labor-intensive
  - Code-heavy
- Methodology
  - `voice` package → Model-ready dataset
  - User-friendly
  - Streamlined
  - Low-code

3. What is `voice`

- Main deliverable: Free tool for general audio analysis

- Main deliverable: Free tool for general audio analysis
- Free, open-source toolkit available on CRAN and GitHub

3. What is `voice`

- Main deliverable: Free tool for general audio analysis
- Free, open-source toolkit available on CRAN and GitHub
- Designed to streamline audio analysis by integrating music theory and advanced computational techniques

3. What is voice

- Main deliverable: Free tool for general audio analysis
- Free, open-source toolkit available on CRAN and GitHub
- Designed to streamline audio analysis by integrating music theory and advanced computational techniques
- Based on

- Main deliverable: Free tool for general audio analysis
- Free, open-source toolkit available on CRAN and GitHub
- Designed to streamline audio analysis by integrating music theory and advanced computational techniques
- Based on
  - UNIX's ffmpeg, Homebrew, Miniconda, MuseScore, wget

- Main deliverable: Free tool for general audio analysis
- Free, open-source toolkit available on CRAN and GitHub
- Designed to streamline audio analysis by integrating music theory and advanced computational techniques
- Based on
  - UNIX's ffmpeg, Homebrew, Miniconda, MuseScore, wget
  - R's gm, music, reticulate, tabr, tidyverse, tuneR, wrassp

- Main deliverable: Free tool for general audio analysis
- Free, open-source toolkit available on CRAN and GitHub
- Designed to streamline audio analysis by integrating music theory and advanced computational techniques
- Based on
  - UNIX's ffmpeg, Homebrew, Miniconda, MuseScore, wget
  - R's gm, music, reticulate, tabr, tidyverse, tuneR, wrassp
  - Python's Parselmouth, pyannote-audio, pychord

- Main deliverable: Free tool for general audio analysis
- Free, open-source toolkit available on CRAN and GitHub
- Designed to streamline audio analysis by integrating music theory and advanced computational techniques
- Based on
  - UNIX's ffmpeg, Homebrew, Miniconda, MuseScore, wget
  - R's gm, music, reticulate, tabr, tidyverse, tuneR, wrassp
  - Python's Parselmouth, pyannote-audio, pychord
  - C's Praat

3. What is `voice`

- User-friendly functions

- User-friendly functions
  - `extract_features` builds data frames with state of art features from multiple audio files

3. What is `voice`

- User-friendly functions
  - `extract_features` builds data frames with state of art features from multiple audio files
  - `tag` attaches summarized audio features to datasets, supporting anonymization and privacy-aware analysis

- User-friendly functions
  - `extract_features` builds data frames with state of art features from multiple audio files
  - `tag` attaches summarized audio features to datasets, supporting anonymization and privacy-aware analysis
  - `diarize` identifies speaker segments

3. What is `voice`

- User-friendly functions
  - `extract_features` builds data frames with state of art features from multiple audio files
  - `tag` attaches summarized audio features to datasets, supporting anonymization and privacy-aware analysis
  - `diarize` identifies speaker segments
  - `splitw` splits the spoken parts into small blocks

- User-friendly functions
  - `extract_features` builds data frames with state of art features from multiple audio files
  - `tag` attaches summarized audio features to datasets, supporting anonymization and privacy-aware analysis
  - `diarize` identifies speaker segments
  - `splitw` splits the spoken parts into small blocks
- Novel contributions: Formant Removals

- User-friendly functions
  - `extract_features` builds data frames with state of art features from multiple audio files
  - `tag` attaches summarized audio features to datasets, supporting anonymization and privacy-aware analysis
  - `diarize` identifies speaker segments
  - `splitw` splits the spoken parts into small blocks
- Novel contributions: Formant Removals
  - Isolates fundamental frequency (F0) from formants

3. What is `voice`

- User-friendly functions
  - `extract_features` builds data frames with state of art features from multiple audio files
  - `tag` attaches summarized audio features to datasets, supporting anonymization and privacy-aware analysis
  - `diarize` identifies speaker segments
  - `splitw` splits the spoken parts into small blocks
- Novel contributions: Formant Removals
  - Isolates fundamental frequency (F0) from formants
  - Improves feature interpretability in models

3. What is `voice`

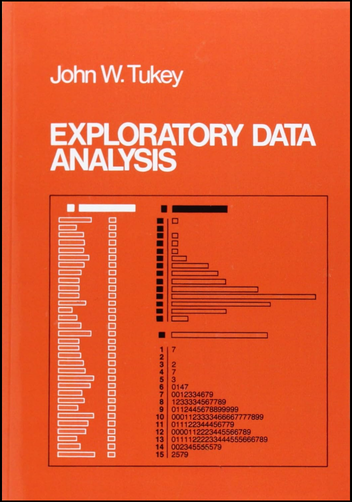
- User-friendly functions
  - `extract_features` builds data frames with state of art features from multiple audio files
  - `tag` attaches summarized audio features to datasets, supporting anonymization and privacy-aware analysis
  - `diarize` identifies speaker segments
  - `splitw` splits the spoken parts into small blocks
- Novel contributions: Formant Removals
  - Isolates fundamental frequency (F0) from formants
  - Improves feature interpretability in models
  - Preliminary results indicate Formant Removals among the most important variables

3. What is `voice`

- User-friendly functions
  - `extract_features` builds data frames with state of art features from multiple audio files
  - `tag` attaches summarized audio features to datasets, supporting anonymization and privacy-aware analysis
  - `diarize` identifies speaker segments
  - `splitw` splits the spoken parts into small blocks
- Novel contributions: Formant Removals
  - Isolates fundamental frequency (F0) from formants
  - Improves feature interpretability in models
  - Preliminary results indicate Formant Removals among the most important variables
- Allows the use of music notation and theory (ongoing work)

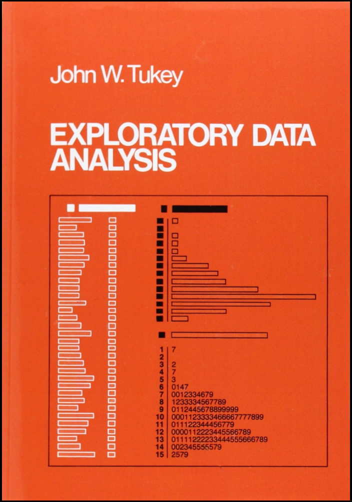
4. x-number summaries

4. x-number summaries



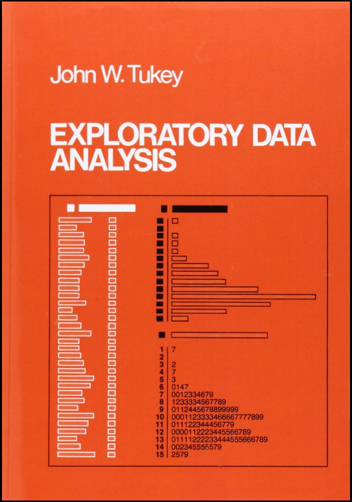
- Tukey (1977)

4. x-number summaries



- Tukey (1977)
  5-number summary

4. x-number summaries

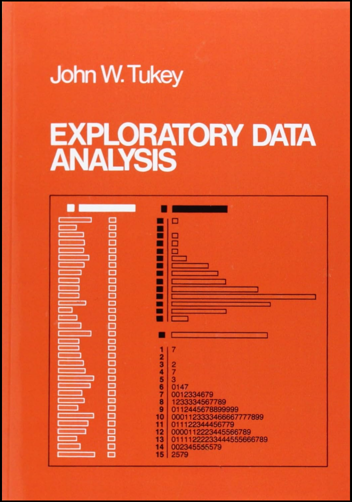


- Tukey (1977)

  5-number summary

  1. Minimum (0th percentile)

4. x-number summaries

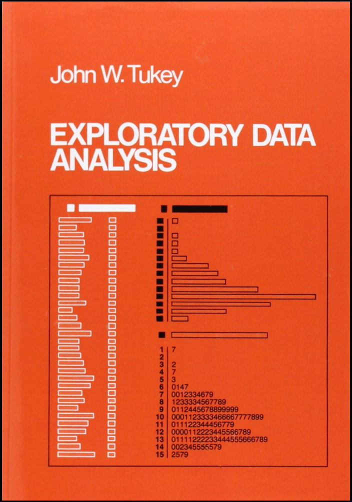


- Tukey (1977)

  5-number summary

  1. Minimum (0th percentile)
  2. 1st quartile (25th perc.)

4. x-number summaries

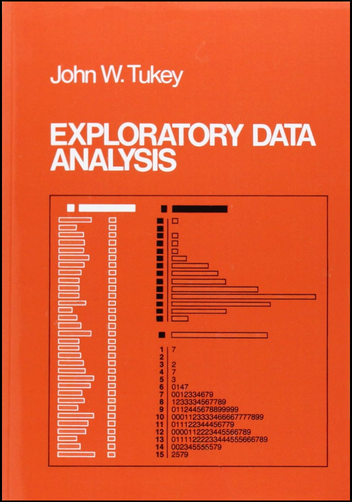


- Tukey (1977)

  5-number summary

  1. Minimum (0th percentile)
  2. 1st quartile (25th perc.)
  3. Median (50th perc.)

4. *x*-number summaries

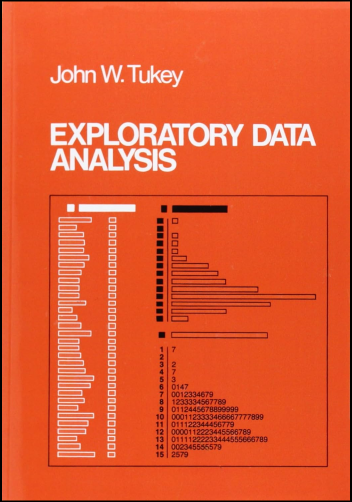


- Tukey (1977)

  5-number summary

  1. Minimum (0th percentile)
  2. 1st quartile (25th perc.)
  3. Median (50th perc.)
  4. 3rd quartile (75th perc.)

4. *x*-number summaries

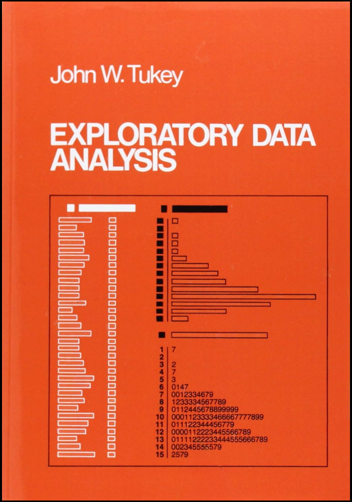


- Tukey (1977)

  5-number summary

  1. Minimum (0th percentile)
  2. 1st quartile (25th perc.)
  3. Median (50th perc.)
  4. 3rd quartile (75th perc.)
  5. Maximum (100th perc.)

4. x-number summaries



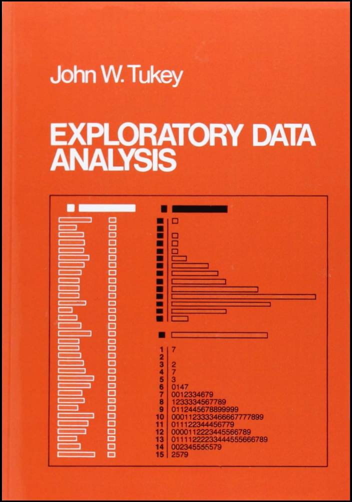
- Tukey (1977)

  5-number summary

  1. Minimum (0th percentile)
  2. 1st quartile (25th perc.)
  3. Median (50th perc.)
  4. 3rd quartile (75th perc.)
  5. Maximum (100th perc.)

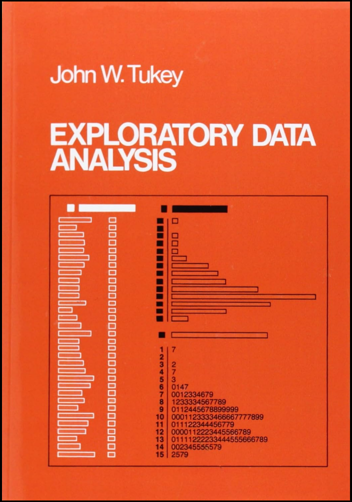
- Zabala & Salum (2025b)

4. x-number summaries



- Tukey (1977)
  5-number summary
  1. Minimum (0th percentile)
  2. 1st quartile (25th perc.)
  3. Median (50th perc.)
  4. 3rd quartile (75th perc.)
  5. Maximum (100th perc.)
- Zabala & Salum (2025b)
  6-number summary

4. *x*-number summaries



- Tukey (1977)

  5-number summary

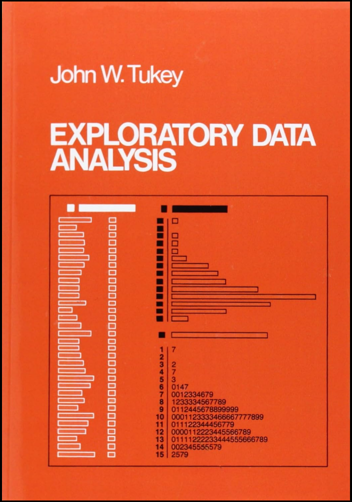
  1. Minimum (0th percentile)
  2. 1st quartile (25th perc.)
  3. Median (50th perc.)
  4. 3rd quartile (75th perc.)
  5. Maximum (100th perc.)

- Zabala & Salum (2025b)

  6-number summary

  1. Mean

4. *x*-number summaries



- Tukey (1977)

  5-number summary

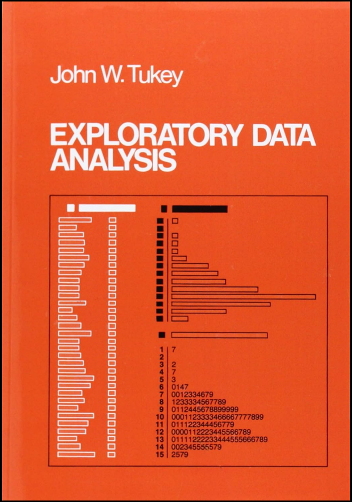
  1. Minimum (0th percentile)
  2. 1st quartile (25th perc.)
  3. Median (50th perc.)
  4. 3rd quartile (75th perc.)
  5. Maximum (100th perc.)

- Zabala & Salum (2025b)

  6-number summary

  1. Mean
  2. Standard deviation

4. x-number summaries



- Tukey (1977)

  5-number summary

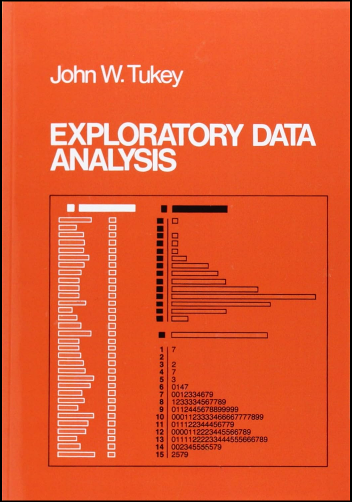
  1. Minimum (0th percentile)
  2. 1st quartile (25th perc.)
  3. Median (50th perc.)
  4. 3rd quartile (75th perc.)
  5. Maximum (100th perc.)

- Zabala & Salum (2025b)

  6-number summary

  1. Mean
  2. Standard deviation
  3. Coefficient of variation

4. x-number summaries



- Tukey (1977)

  5-number summary

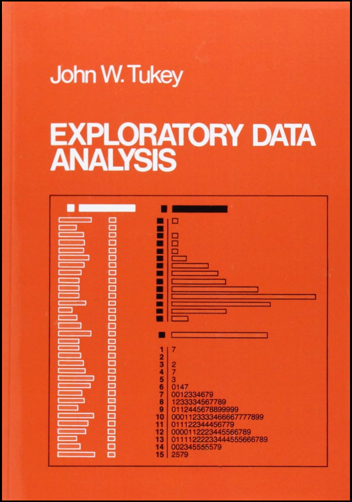
  1. Minimum (0th percentile)
  2. 1st quartile (25th perc.)
  3. Median (50th perc.)
  4. 3rd quartile (75th perc.)
  5. Maximum (100th perc.)

- Zabala & Salum (2025b)

  6-number summary

  1. Mean
  2. Standard deviation
  3. Coefficient of variation
  4. Median

4. *x*-number summaries



- Tukey (1977)

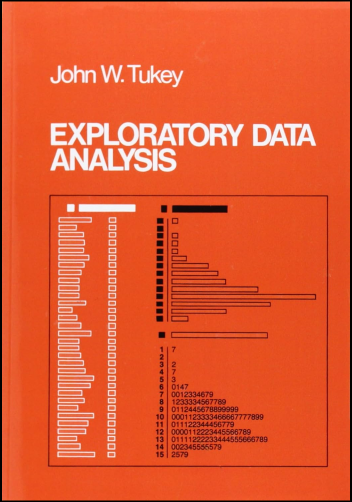
  5-number summary

  1. Minimum (0th percentile)
  2. 1st quartile (25th perc.)
  3. Median (50th perc.)
  4. 3rd quartile (75th perc.)
  5. Maximum (100th perc.)

- Zabala & Salum (2025b)

  6-number summary

  1. Mean
  2. Standard deviation
  3. Coefficient of variation
  4. Median
  5. Interquartile range



- Tukey (1977)
  5-number summary
  1. Minimum (0th percentile)
  2. 1st quartile (25th perc.)
  3. Median (50th perc.)
  4. 3rd quartile (75th perc.)
  5. Maximum (100th perc.)
- Zabala & Salum (2025b)
  6-number summary
  1. Mean
  2. Standard deviation
  3. Coefficient of variation
  4. Median
  5. Interquartile range
  6. Median absolute deviation

5. Article 1

5. Article 1

- voice: A Comprehensive R Package for Audio Analysis

- voice: A Comprehensive R Package for Audio Analysis
  - Journal of Open Source Software (JOSS)

- voice: A Comprehensive R Package for Audio Analysis
    - Journal of Open Source Software (JOSS)
    - Published 30 July 2025

- voice: A Comprehensive R Package for Audio Analysis
  - Journal of Open Source Software (JOSS)
  - Published 30 July 2025
  - Presents the `voice` package

5. Article 1

- voice: A Comprehensive R Package for Audio Analysis
  - Journal of Open Source Software (JOSS)
  - Published 30 July 2025
  - Presents the `voice` package
  - Statement of need, main features, and example applications

- voice: A Comprehensive R Package for Audio Analysis
  - Journal of Open Source Software (JOSS)
  - Published 30 July 2025
  - Presents the `voice` package
  - Statement of need, main features, and example applications
  - Highlights the package's performance and availability

1. Extract features

1.1 Load packages and audio files

```
# packs
library(voice)
library(tidyverse)

# get path to audio file
wavDir <- list.files(system.file('extdata', package = 'wrassp'),
                     pattern = glob2rx('*.wav'), full.names = TRUE)
```

1.2 Extract features

```
# minimal usage
M <- voice::extract_features(wavDir)
glimpse(M)
#> Rows: 1,196
#> Columns: 13
#> $ section_seq      <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16…
#> $ section_seq_file <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16…
#> $ wav_path         <chr> "/Library/Frameworks/R.framework/Versions/4.5-arm64/R…
#> $ f0               <dbl> NA, NA, NA, NA, NA, NA, NA, 115.8593, 108.9439, 107.4…
#> $ f1               <int> NA, NA, NA, NA, 185, 260, 254, 277, 261, 231, 177, 19…
#> $ f2               <int> 1854, 1886, 1749, 1888, 1962, 1973, 2026, 2037, 2130,…
#> $ f3               <int> NA, 2893, 2676, 2659, 2639, 2676, 2993, 2932, 3016, 2…
#> $ f4               <int> 3113, 3708, 3509, 3658, 3248, 3239, 3830, 3479, 3561,…
#> $ f5               <int> 4191, 4678, 4502, 4331, 3653, 3836, 4602, 4585, 4720,…
#> $ f6               <int> 5226, 5659, 5035, 5177, 5208, 5146, 5233, 5390, 5366,…
#> $ f7               <int> 6077, 6725, 6526, 6518, 6493, 6567, 6603, 6532, 6510,…
#> $ f8               <int> 6675, NA, NA, NA, 7681, 7751, 7803, NA, 7835, 7614, 7…
#> $ gain             <dbl> 21.63347, 22.76034, 28.52825, 29.67069, 36.25124, 43.…
```

2. Tag

```
# creating Extended synthetic data
E <- dplyr::tibble(subject_id = c(1,1,1,2,2,2,3,3,3), wav_path = wavDir)
E
#> # A tibble: 9 × 2
#>   subject_id wav_path
#>        <dbl> <chr>
#> 1          1 /Library/Frameworks/R.framework/Versions/4.5-arm64/Resources/libra…
#> 2          1 /Library/Frameworks/R.framework/Versions/4.5-arm64/Resources/libra…
#> 3          1 /Library/Frameworks/R.framework/Versions/4.5-arm64/Resources/libra…
#> 4          2 /Library/Frameworks/R.framework/Versions/4.5-arm64/Resources/libra…
#> 5          2 /Library/Frameworks/R.framework/Versions/4.5-arm64/Resources/libra…
#> 6          2 /Library/Frameworks/R.framework/Versions/4.5-arm64/Resources/libra…
#> 7          3 /Library/Frameworks/R.framework/Versions/4.5-arm64/Resources/libra…
#> 8          3 /Library/Frameworks/R.framework/Versions/4.5-arm64/Resources/libra…
#> 9          3 /Library/Frameworks/R.framework/Versions/4.5-arm64/Resources/libra…
```

```r
# minimal usage
voice::tag(E)
#> # A tibble: 9 × 7
#>   wav_path    f0_tag_mean f0_tag_sd f0_tag_vc f0_tag_median f0_tag_iqr f0_tag_mad
#>   <chr>             <dbl>     <dbl>     <dbl>         <dbl>      <dbl>      <dbl>
#> 1 /Library/…         85.4      17.6     0.206          76.1       29.4       7.53
#> 2 /Library/…         85.4      15.6     0.183          80.1       27.8      14.4
#> 3 /Library/…         84.6      13.0     0.154          78.8       23.9      14.0
#> 4 /Library/…         84.8      14.5     0.171          79.1       28.1      11.9
#> 5 /Library/…         86.0      14.7     0.170          78.7       30.0      11.0
#> 6 /Library/…         82.9      15.6     0.188          74.8       23.8       4.78
#> 7 /Library/…         78.2      16.2     0.207          73.5       13.4       6.82
#> 8 /Library/…         84.5      14.5     0.172          78.1       17.8       8.95
#> 9 /Library/…         81.0      12.2     0.151          75.9       23.1       9.14

# canonical data
voice::tag(E, groupBy = 'subject_id')
#> # A tibble: 3 × 7
#>   subject_id f0_tag_mean f0_tag_sd f0_tag_vc f0_tag_median f0_tag_iqr f0_tag_mad
#>        <dbl>       <dbl>     <dbl>     <dbl>         <dbl>      <dbl>      <dbl>
#> 1          1        85.1      15.3     0.180          78.3       26.8      11.9
#> 2          2        84.6      14.9     0.176          76.4       28.3       7.97
#> 3          3        81.0      14.6     0.180          75.6       21.6       8.68
```

5. Diarize

```r
# download
url0 <- 'https://github.com/filipezabala/voiceAudios/raw/main/wav/sherlock0.wav'
wavDir <- normalizePath(tempdir())
download.file(url0, paste0(wavDir, '/sherlock0.wav'), mode = 'wb')
```

Diarization can be performed to detect speaker segments (i.e., 'who spoke when').

```r
# diarize
voice::diarize(fromWav = wavDir, toRttm = wavDir, token = 'YOUR_TOKEN')
#> Time difference of 27.49933 secs
```

The `voice::diarize()` function creates Rich Transcription Time Marked (RTTM)[1] files, space-delimited text files containing one turn per line defined by NIST - National Institute of Standards and Technology. The RTTM files can be read using `voice::read_rttm()`.

```
# read_rttm
(rttm <- voice::read_rttm(wavDir))
#> $doremi.rttm
#>      type   file chnl  tbeg  tdur ortho stype       name conf slat
#> 1 SPEAKER doremi    1 0.031 5.805  <NA>  <NA> SPEAKER_00 <NA> <NA>
#>
#> $sherlock0.rttm
#>      type      file chnl   tbeg  tdur ortho stype       name conf slat
#> 1 SPEAKER sherlock0    1  0.908 5.231  <NA>  <NA> SPEAKER_00 <NA> <NA>
#> 2 SPEAKER sherlock0    1  6.933 6.463  <NA>  <NA> SPEAKER_00 <NA> <NA>
#> 3 SPEAKER sherlock0    1 13.565 8.674  <NA>  <NA> SPEAKER_00 <NA> <NA>
```

Finally, the audio waves can be automatically segmented.

```
# split audio wave
voice::splitw(fromWav = wavDir, fromRttm = wavDir, to = wavDir)
#> TOTAL TIME 0.262 SECONDS
dir(wavDir, pattern = '.[Ww][Aa][Vv]$')
#> [1] "doremi_split_1.wav"    "doremi.wav"            "sherlock0_split_1.wav"
#> [4] "sherlock0_split_2.wav" "sherlock0_split_3.wav" "sherlock0.wav"
```

6. Article 2

6. Article 2

- Predicting sex and emotional valence automatically from voice

6. Article 2

- Predicting sex and emotional valence automatically from voice
  - Submitted to Journal of Biomedical Informatics (JBI-Elsevier)

- Predicting sex and emotional valence automatically from voice
    - Submitted to Journal of Biomedical Informatics (JBI-Elsevier)
    - Historical remarks

6. Article 2

- Predicting sex and emotional valence automatically from voice
  - Submitted to Journal of Biomedical Informatics (JBI-Elsevier)
  - Historical remarks
  - RAVDESS and CREMA-D open datasets

6. Article 2

- Predicting sex and emotional valence automatically from voice
  - Submitted to Journal of Biomedical Informatics (JBI-Elsevier)
  - Historical remarks
  - RAVDESS and CREMA-D open datasets
  - Binary Logistic (BL), Support Vector Machines (SVM), Random Forest (RF), and BART models

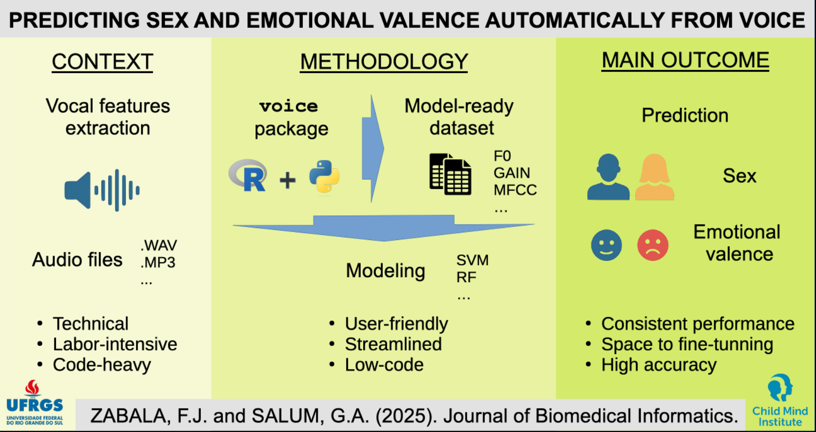
6. Article 2

- Predicting sex and emotional valence automatically from voice
  - Submitted to Journal of Biomedical Informatics (JBI-Elsevier)
  - Historical remarks
  - RAVDESS and CREMA-D open datasets
  - Binary Logistic (BL), Support Vector Machines (SVM), Random Forest (RF), and BART models
  - Models achieve accuracy statistically superior to the No Information Rate, the largest proportion of the observed classes

6. Article 2

- Predicting sex and emotional valence automatically from voice
  - Submitted to Journal of Biomedical Informatics (JBI-Elsevier)
  - Historical remarks
  - RAVDESS and CREMA-D open datasets
  - Binary Logistic (BL), Support Vector Machines (SVM), Random Forest (RF), and BART models
  - Models achieve accuracy statistically superior to the No Information Rate, the largest proportion of the observed classes
  - RF and SVM consistently demonstrate strong performance across all evaluated variables and quality measures

- Predicting sex and emotional valence automatically from voice
  - Submitted to Journal of Biomedical Informatics (JBI-Elsevier)
  - Historical remarks
  - RAVDESS and CREMA-D open datasets
  - Binary Logistic (BL), Support Vector Machines (SVM), Random Forest (RF), and BART models
  - Models achieve accuracy statistically superior to the No Information Rate, the largest proportion of the observed classes
  - RF and SVM consistently demonstrate strong performance across all evaluated variables and quality measures
  - Emotional valence classification is feasible but may require hyperparameter optimization



Graphical abstract

	RAVDESS	CREMA-D
Rows/audios	1440	7442
Speakers	24	91
Language	English	English
Sex (M/F)		
#	720/720	3930/3512
%	50/50	52.8/47.2
Emotional Valence (−/+)		
#	672/768	5084/2358
%	53.3/46.7	68.3/31.7

Datasets summary

- Models

- Models
    - BL - Binary Logistic

6. Article 2

- Models
  - BL - Binary Logistic
  - RF - Random Forests

- Models
  - BL - Binary Logistic
  - RF - Random Forests
  - SVM - Support Vector Machine

6. Article 2

- Models
  - BL - Binary Logistic
  - RF - Random Forests
  - SVM - Support Vector Machine
  - BART - Bayesian Additive Regression Trees

6. Article 2

- Models
  - BL - Binary Logistic
  - RF - Random Forests
  - SVM - Support Vector Machine
  - BART - Bayesian Additive Regression Trees
- 70%-30% train-test split

- Models
  - BL - Binary Logistic
  - RF - Random Forests
  - SVM - Support Vector Machine
  - BART - Bayesian Additive Regression Trees
- 70%-30% train-test split
- 1,000 runs

6. Article 2

- Models
  - BL - Binary Logistic
  - RF - Random Forests
  - SVM - Support Vector Machine
  - BART - Bayesian Additive Regression Trees
- 70%-30% train-test split
- 1,000 runs
- 9 quality measures evaluated

- Models
  - BL - Binary Logistic
  - RF - Random Forests
  - SVM - Support Vector Machine
  - BART - Bayesian Additive Regression Trees
- 70%-30% train-test split
- 1,000 runs
- 9 quality measures evaluated
- Results presented with Tukey's 5-number summary

NIR: No Information Rate

NIR: No Information Rate

- The largest proportion of the observed classes

NIR: No Information Rate

- The largest proportion of the observed classes
- Smaller AccuracyPValue, stronger the hypothesis Accuracy $>$ NIR

NIR: No Information Rate

- The largest proportion of the observed classes
- Smaller AccuracyPValue, stronger the hypothesis Accuracy $>$ NIR

Under H_0 : Accuracy $\leq$ NIR *vs* H_1 : Accuracy $>$ NIR,

NIR: No Information Rate

- The largest proportion of the observed classes
- Smaller AccuracyPValue, stronger the hypothesis Accuracy $>$ NIR

Under H_0 : Accuracy $\leq$ NIR *vs* H_1 : Accuracy $>$ NIR,

$$\begin{aligned} \text{AccuracyPValue} &= P(X \geq TP + TN | \text{Accuracy} = \text{NIR}) \\ &= \sum_{i=TP+TN}^{n} \binom{n}{i} \text{NIR}^{i} (1 - \text{NIR})^{n-i} \end{aligned}$$

	Accuracy					Sensitivity					Specificity				
	0%	25%	50%	75%	100%	0%	25%	50%	75%	100%	0%	25%	50%	75%	100%
RAVDESS															
BART	0.127	0.776	0.842	0.927	0.998	0.002	0.757	0.971	1.000	1.000	0.007	0.738	0.946	1.000	1.000
RF	0.100	0.806	0.885	0.933	0.998	0.125	0.883	0.967	0.992	1.000	0.100	0.760	0.908	0.983	1.000
SVM	0.056	0.841	0.912	0.950	1.000	0.056	0.897	0.983	1.000	1.000	0.106	0.820	0.938	0.994	1.000
BL	0.360	0.729	0.792	0.842	0.965	0.274	0.751	0.871	0.953	1.000	0.319	0.690	0.812	0.906	1.000
CREMA-D															
BART	0.284	0.857	0.907	0.929	0.978	0.129	0.859	0.940	0.978	1.000	0.179	0.859	0.931	0.974	1.000
RF	0.609	0.888	0.910	0.928	0.969	0.554	0.909	0.947	0.972	1.000	0.479	0.844	0.903	0.944	0.998
SVM	0.654	0.913	0.934	0.948	0.986	0.578	0.920	0.955	0.980	1.000	0.548	0.897	0.940	0.972	1.000
BL	0.630	0.881	0.905	0.922	0.974	0.549	0.882	0.926	0.957	0.999	0.555	0.864	0.913	0.948	1.000

	Pos Pred Value					Neg Pred Value					F1				
	0%	25%	50%	75%	100%	0%	25%	50%	75%	100%	0%	25%	50%	75%	100%
RAVDESS															
BART	0.126	0.698	0.947	1.000	1.000	0.125	0.710	0.972	1.000	1.000	0.005	0.771	0.842	0.930	0.998
RF	0.155	0.729	0.915	0.988	1.000	0.168	0.835	0.965	0.992	1.000	0.222	0.804	0.887	0.940	0.998
SVM	0.169	0.779	0.938	0.996	1.000	0.157	0.853	0.982	1.000	1.000	0.107	0.832	0.916	0.953	1.000
BL	0.173	0.653	0.822	0.922	1.000	0.160	0.699	0.868	0.958	1.000	0.296	0.735	0.795	0.848	0.965
CREMA-D															
BART	0.288	0.861	0.937	0.978	1.000	0.200	0.837	0.934	0.978	1.000	0.229	0.864	0.910	0.932	0.979
RF	0.389	0.851	0.915	0.957	0.999	0.327	0.883	0.941	0.971	1.000	0.560	0.894	0.917	0.933	0.971
SVM	0.423	0.894	0.947	0.978	1.000	0.340	0.898	0.952	0.979	1.000	0.595	0.918	0.937	0.951	0.987
BL	0.439	0.865	0.922	0.959	1.000	0.325	0.855	0.919	0.957	0.999	0.609	0.887	0.910	0.926	0.976

	Detection Rate					Kappa					AccuracyPValue				
	0%	25%	50%	75%	100%	0%	25%	50%	75%	100%	0%	25%	50%	75%	100%
RAVDESS															
BART	0.002	0.375	0.431	0.481	0.554	0.001	0.572	0.689	0.854	0.996	0.000	0.000	0.000	0.000	1.000
RF	0.000	0.373	0.475	0.498	0.604	0.000	0.620	0.760	0.863	0.996	0.000	0.000	0.000	0.000	1.000
SVM	0.000	0.375	0.481	0.498	0.604	0.000	0.685	0.821	0.897	1.000	0.000	0.000	0.000	0.000	1.000
BL	0.123	0.360	0.410	0.448	0.565	0.076	0.473	0.591	0.684	0.929	0.000	0.000	0.000	0.000	1.000
CREMA-D															
BART	0.106	0.450	0.476	0.492	0.528	0.050	0.719	0.813	0.857	0.956	0.000	0.000	0.000	0.000	1.000
RF	0.249	0.454	0.495	0.526	0.590	0.306	0.775	0.818	0.854	0.937	0.000	0.000	0.000	0.000	1.000
SVM	0.249	0.458	0.498	0.529	0.603	0.329	0.823	0.868	0.894	0.971	0.000	0.000	0.000	0.000	1.000
BL	0.248	0.448	0.482	0.511	0.567	0.303	0.761	0.810	0.843	0.949	0.000	0.000	0.000	0.000	1.000

Summary of SEX classification performance over 1,000 simulations

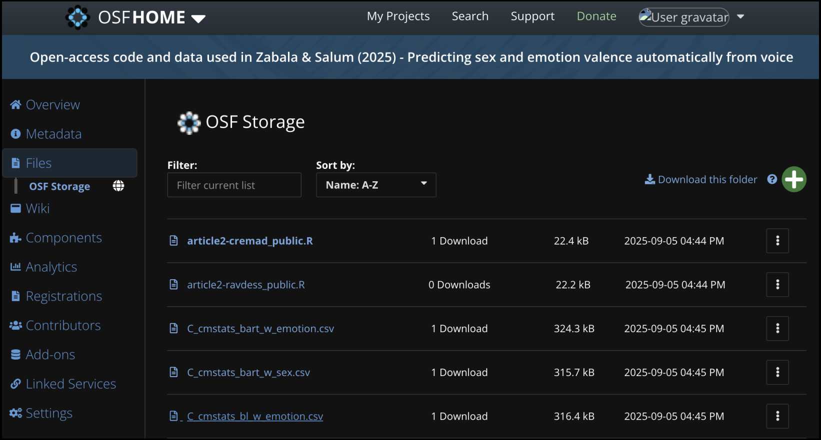
	Accuracy					Sensitivity					Specificity				
	0%	25%	50%	75%	100%	0%	25%	50%	75%	100%	0%	25%	50%	75%	100%
RAVDESS															
BART	0.606	0.698	0.717	0.735	0.804	0.625	0.707	0.727	0.746	0.828	0.536	0.674	0.705	0.737	0.826
RF	0.613	0.698	0.719	0.738	0.796	0.684	0.766	0.789	0.812	0.895	0.442	0.603	0.638	0.674	0.795
SVM	0.635	0.710	0.731	0.748	0.802	0.711	0.789	0.812	0.832	0.898	0.469	0.598	0.638	0.674	0.795
BL	0.487	0.575	0.600	0.621	0.704	0.449	0.574	0.602	0.625	0.703	0.455	0.562	0.598	0.625	0.728
CREMA-D															
BART	0.633	0.669	0.677	0.685	0.716	0.573	0.625	0.636	0.646	0.684	0.645	0.751	0.766	0.781	0.835
RF	0.704	0.725	0.731	0.736	0.764	0.902	0.935	0.944	0.952	0.970	0.181	0.255	0.275	0.294	0.373
SVM	0.718	0.740	0.746	0.753	0.775	0.863	0.900	0.912	0.920	0.952	0.309	0.376	0.393	0.409	0.492
BL	0.535	0.584	0.601	0.645	0.716	0.495	0.581	0.607	0.661	0.790	0.461	0.577	0.604	0.627	0.692

	Pos Pred Value					Neg Pred Value					F1				
	0%	25%	50%	75%	100%	0%	25%	50%	75%	100%	0%	25%	50%	75%	100%
RAVDESS															
BART	0.622	0.717	0.739	0.760	0.834	0.583	0.673	0.693	0.712	0.798	0.635	0.714	0.732	0.749	0.819
RF	0.623	0.693	0.714	0.733	0.815	0.596	0.701	0.724	0.747	0.846	0.654	0.732	0.749	0.765	0.818
SVM	0.635	0.698	0.720	0.740	0.817	0.635	0.724	0.747	0.768	0.853	0.685	0.746	0.762	0.776	0.825
BL	0.520	0.605	0.629	0.653	0.742	0.452	0.542	0.567	0.590	0.668	0.484	0.592	0.615	0.637	0.711
CREMA-D															
BART	0.792	0.846	0.854	0.862	0.891	0.446	0.486	0.494	0.503	0.539	0.682	0.721	0.729	0.736	0.766
RF	0.715	0.732	0.736	0.741	0.765	0.585	0.670	0.696	0.721	0.789	0.810	0.824	0.827	0.830	0.846
SVM	0.742	0.759	0.763	0.768	0.792	0.593	0.654	0.673	0.695	0.761	0.811	0.826	0.831	0.835	0.850
BL	0.709	0.751	0.766	0.786	0.834	0.353	0.394	0.413	0.456	0.544	0.592	0.657	0.675	0.718	0.777

	Detection Rate					Kappa					AccuracyPValue				
	0%	25%	50%	75%	100%	0%	25%	50%	75%	100%	0%	25%	50%	75%	100%
RAVDESS															
BART	0.333	0.377	0.388	0.398	0.442	0.205	0.394	0.432	0.468	0.606	0.000	0.000	0.000	0.000	0.001
RF	0.365	0.408	0.421	0.433	0.477	0.216	0.389	0.430	0.469	0.587	0.000	0.000	0.000	0.000	0.000
SVM	0.379	0.421	0.433	0.444	0.479	0.259	0.412	0.454	0.490	0.602	0.000	0.000	0.000	0.000	0.000
BL	0.240	0.306	0.321	0.333	0.375	-0.028	0.147	0.198	0.239	0.409	1.000	1.000	1.000	1.000	1.000
CREMA-D															
BART	0.392	0.426	0.434	0.441	0.465	0.244	0.336	0.350	0.364	0.419	0.000	0.403	0.714	0.922	1.000
RF	0.616	0.638	0.644	0.650	0.662	0.260	0.277	0.362	0.377	0.369	0.000	0.000	0.000	0.000	0.013
SVM	0.591	0.615	0.622	0.628	0.648	0.263	0.325	0.341	0.358	0.419	0.000	0.000	0.000	0.000	0.000
BL	0.338	0.397	0.414	0.452	0.541	0.064	0.148	0.181	0.249	0.391	1.000	1.000	1.000	1.000	1.000

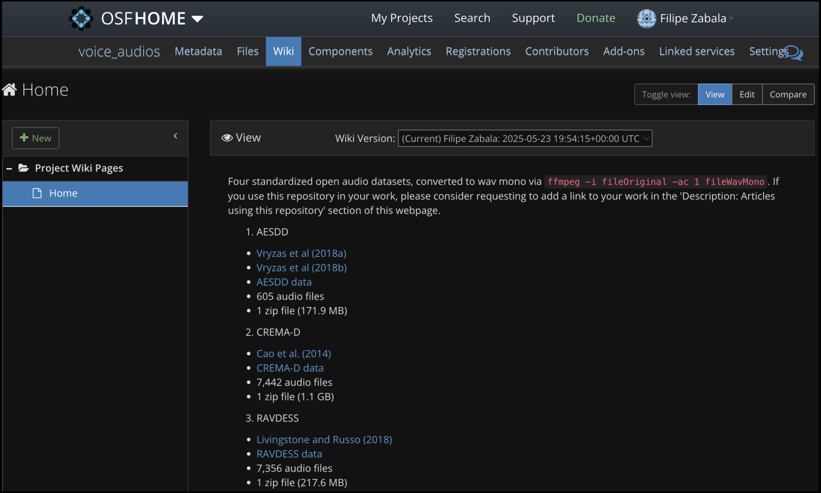
Summary of EMOTIONAL VALENCE classif. performance over 1,000 simulations

6. Article 2



https://osf.io/ahrbs/files/osfstorage

6. Article 2



https://osf.io/9g73a/wiki/home/

Thank you!

filipezabala.com