voice

easier audio analysis for digital phenotyping

Filipe J. Zabala · Giovanni A. Salum Jr.

Graduate Program in Psychiatry
Medical School · UFRGS

2023-07-25

- Tool for voice analysis

- Tool for voice analysis
- Free software available on CRAN

- Tool for voice analysis
- Free software available on CRAN
- Based on

- Tool for voice analysis
- Free software available on CRAN
- Based on
  - UNIX's ffmpeg, Homebrew, Miniconda, MuseScore, wget

- Tool for voice analysis
- Free software available on CRAN
- Based on
  - UNIX's ffmpeg, Homebrew, Miniconda, MuseScore, wget
  - R's gm, music, reticulate, tabr, tidyverse, tuneR, wrassp

UFRGS
PPGPSIO Programa de Pós-Graduação em
Psiquiatria e Ciências do Comportamento

1. What is `voice`

- Tool for voice analysis
- Free software available on CRAN
- Based on
  - UNIX's ffmpeg, Homebrew, Miniconda, MuseScore, wget
  - R's gm, music, reticulate, tabr, tidyverse, tuneR, wrassp
  - Python's Parselmouth, pyannote-audio

- Tool for voice analysis
- Free software available on CRAN
- Based on
  - UNIX's ffmpeg, Homebrew, Miniconda, MuseScore, wget
  - R's gm, music, reticulate, tabr, tidyverse, tuneR, wrassp
  - Python's Parselmouth, pyannote-audio
  - C's Praat

- Tool for voice analysis
- Free software available on CRAN
- Based on
  - UNIX's ffmpeg, Homebrew, Miniconda, MuseScore, wget
  - R's gm, music, reticulate, tabr, tidyverse, tuneR, wrassp
  - Python's Parselmouth, pyannote-audio
  - C's Praat
- Recommended for digital phenotyping

- Tool for voice analysis
- Free software available on CRAN
- Based on
  - UNIX's ffmpeg, Homebrew, Miniconda, MuseScore, wget
  - R's gm, music, reticulate, tabr, tidyverse, tuneR, wrassp
  - Python's Parselmouth, pyannote-audio
  - C's Praat
- Recommended for digital phenotyping
  - Non-invasive

- Tool for voice analysis
- Free software available on CRAN
- Based on
  - UNIX's ffmpeg, Homebrew, Miniconda, MuseScore, wget
  - R's gm, music, reticulate, tabr, tidyverse, tuneR, wrassp
  - Python's Parselmouth, pyannote-audio
  - C's Praat
- Recommended for digital phenotyping
  - Non-invasive
  - Objective

- Tool for voice analysis
- Free software available on CRAN
- Based on
  - UNIX's ffmpeg, Homebrew, Miniconda, MuseScore, wget
  - R's gm, music, reticulate, tabr, tidyverse, tuneR, wrassp
  - Python's Parselmouth, pyannote-audio
  - C's Praat
- Recommended for digital phenotyping
  - Non-invasive
  - Objective
  - Programmable

- Tool for voice analysis
- Free software available on CRAN
- Based on
    - UNIX's ffmpeg, Homebrew, Miniconda, MuseScore, wget
    - R's gm, music, reticulate, tabr, tidyverse, tuneR, wrassp
    - Python's Parselmouth, pyannote-audio
    - C's Praat
- Recommended for digital phenotyping
    - Non-invasive
    - Objective
    - Programmable
    - Verifiable

- User-friendly functions, e.g.

UFRGS Programa de Pós-Graduação em
PPGPSIQ Psiquiatria e Ciências do Comportamento

- User-friendly functions, e.g.
  - `extract_features` to build data frames with state of art features from multiple audio files

- User-friendly functions, e.g.
  - `extract_features` to build data frames with state of art features from multiple audio files
  - `tag` to attach summarized columns from audio files to canonical data

UFRGS
PPGPSIQ Programa de Pós-Graduação em
Psiquiatria e Ciências do Comportamento

2. Why another package

- User-friendly functions, e.g.
    - extract_features to build data frames with state of art features from multiple audio files
    - tag to attach summarized columns from audio files to canonical data
    - diarize to map who spoke when

- User-friendly functions, e.g.
  - `extract_features` to build data frames with state of art features from multiple audio files
  - `tag` to attach summarized columns from audio files to canonical data
  - `diarize` to map who spoke when
  - `splitw` to split the spoken parts into small blocks

- User-friendly functions, e.g.
  - `extract_features` to build data frames with state of art features from multiple audio files
  - `tag` to attach summarized columns from audio files to canonical data
  - `diarize` to map who spoke when
  - `splitw` to split the spoken parts into small blocks
- Three new features proposed

- User-friendly functions, e.g.
  - extract_features to build data frames with state of art features from multiple audio files
  - tag to attach summarized columns from audio files to canonical data
  - diarize to map who spoke when
  - splitw to split the spoken parts into small blocks
- Three new features proposed
  - Formant Removal

- User-friendly functions, e.g.
  - extract_features to build data frames with state of art features from multiple audio files
  - tag to attach summarized columns from audio files to canonical data
  - diarize to map who spoke when
  - splitw to split the spoken parts into small blocks
- Three new features proposed
  - Formant Removal
  - Formant Cumulated Removal

- User-friendly functions, e.g.
  - extract_features to build data frames with state of art features from multiple audio files
  - tag to attach summarized columns from audio files to canonical data
  - diarize to map who spoke when
  - splitw to split the spoken parts into small blocks
- Three new features proposed
  - Formant Removal
  - Formant Cumulated Removal
  - Formant Position Removal

- User-friendly functions, e.g.
  - `extract_features` to build data frames with state of art features from multiple audio files
  - `tag` to attach summarized columns from audio files to canonical data
  - `diarize` to map who spoke when
  - `splitw` to split the spoken parts into small blocks
- Three new features proposed
  - Formant Removal
  - Formant Cumulated Removal
  - Formant Position Removal
- Allows the use of music notation and theory

- Predicting sex

- Predicting sex
  - Median accuracy greater than 90%

- Predicting sex
  - Median accuracy greater than 90%
  - Median F1 greater than 89%

- Predicting sex
  - Median accuracy greater than 90%
  - Median F1 greater than 89%
  - Median kappa greater than 55%

- Predicting sex
  - Median accuracy greater than 90%
  - Median F1 greater than 89%
  - Median kappa greater than 55%
  - Median precision greater than 93%

- Predicting sex
  - Median accuracy greater than 90%
  - Median F1 greater than 89%
  - Median kappa greater than 55%
  - Median precision greater than 93%
- Kaggle's BirdCLEF 2023

- Predicting sex
  - Median accuracy greater than 90%
  - Median F1 greater than 89%
  - Median kappa greater than 55%
  - Median precision greater than 93%
- Kaggle's BirdCLEF 2023
  - Simple solution

- Predicting sex
  - Median accuracy greater than 90%
  - Median F1 greater than 89%
  - Median kappa greater than 55%
  - Median precision greater than 93%
- Kaggle's BirdCLEF 2023
  - Simple solution
  - No need of huge processing

- Predicting sex
  - Median accuracy greater than 90%
  - Median F1 greater than 89%
  - Median kappa greater than 55%
  - Median precision greater than 93%
- Kaggle's BirdCLEF 2023
  - Simple solution
  - No need of huge processing
- Vocal signature

- Predicting sex
  - Median accuracy greater than 90%
  - Median F1 greater than 89%
  - Median kappa greater than 55%
  - Median precision greater than 93%
- Kaggle's BirdCLEF 2023
  - Simple solution
  - No need of huge processing
- Vocal signature
  - Nine speakers clearly discriminated

- Predicting sex
  - Median accuracy greater than 90%
  - Median F1 greater than 89%
  - Median kappa greater than 55%
  - Median precision greater than 93%
- Kaggle's BirdCLEF 2023
  - Simple solution
  - No need of huge processing
- Vocal signature
  - Nine speakers clearly discriminated
  - $62 \times 6 = 372$ summarized features

- Predicting sex
  - Median accuracy greater than 90%
  - Median F1 greater than 89%
  - Median kappa greater than 55%
  - Median precision greater than 93%
- Kaggle's BirdCLEF 2023
  - Simple solution
  - No need of huge processing
- Vocal signature
  - Nine speakers clearly discriminated
  - $62 \times 6 = 372$ summarized features
  - First 3 principal components explains 87% of the total variability

UFRGS
PPGPSIQ Programa de Pós-Graduação em
Psiquiatria e Ciências do Comportamento

3. Some results

- Predicting sex
  - Median accuracy greater than 90%
  - Median F1 greater than 89%
  - Median kappa greater than 55%
  - Median precision greater than 93%
- Kaggle's BirdCLEF 2023
  - Simple solution
  - No need of huge processing
- Vocal signature
  - Nine speakers clearly discriminated
  - $62 \times 6 = 372$ summarized features
  - First 3 principal components explains 87% of the total variability
- Putting voice on musical sheet

- Predicting sex
  - Median accuracy greater than 90%
  - Median F1 greater than 89%
  - Median kappa greater than 55%
  - Median precision greater than 93%
- Kaggle's BirdCLEF 2023
  - Simple solution
  - No need of huge processing
- Vocal signature
  - Nine speakers clearly discriminated
  - $62 \times 6 = 372$ summarized features
  - First 3 principal components explains 87% of the total variability
- Putting voice on musical sheet
  - voice vignette

UFRGS
PPGPSIQ Programa de Pós-Graduação em
Psiquiatria e Ciências do Comportamento

4. Current research

- Link with other areas

- Link with other areas
  - Linguistics (Prosody, Cláudia Brescancini, PUCRS, Brazil)

- Link with other areas
  - Linguistics (Prosody, Cláudia Brescancini, PUCRS, Brazil)
  - Quantitative Biology (Birds sounds, Rafael de Andrade Moral, Maynooth University, Ireland)

- Link with other areas
  - Linguistics (Prosody, Cláudia Brescancini, PUCRS, Brazil)
  - Quantitative Biology (Birds sounds, Rafael de Andrade Moral, Maynooth University, Ireland)
  - Pulmonology (Chronic cough, Cesar A. Keller, Mayo Clinic, USA)

- Link with other areas
  - Linguistics (Prosody, Cláudia Brescancini, PUCRS, Brazil)
  - Quantitative Biology (Birds sounds, Rafael de Andrade Moral, Maynooth University, Ireland)
  - Pulmonology (Chronic cough, Cesar A. Keller, Mayo Clinic, USA)
- Automatically identify speakers

- Link with other areas
  - Linguistics (Prosody, Cláudia Brescancini, PUCRS, Brazil)
  - Quantitative Biology (Birds sounds, Rafael de Andrade Moral, Maynooth University, Ireland)
  - Pulmonology (Chronic cough, Cesar A. Keller, Mayo Clinic, USA)
- Automatically identify speakers
  - Proportion of speech on audio by speaker

- Link with other areas
  - Linguistics (Prosody, Cláudia Brescancini, PUCRS, Brazil)
  - Quantitative Biology (Birds sounds, Rafael de Andrade Moral, Maynooth University, Ireland)
  - Pulmonology (Chronic cough, Cesar A. Keller, Mayo Clinic, USA)
- Automatically identify speakers
  - Proportion of speech on audio by speaker
  - Summary statistics by speaker

- Link with other areas
  - Linguistics (Prosody, Cláudia Brescancini, PUCRS, Brazil)
  - Quantitative Biology (Birds sounds, Rafael de Andrade Moral, Maynooth University, Ireland)
  - Pulmonology (Chronic cough, Cesar A. Keller, Mayo Clinic, USA)
- Automatically identify speakers
  - Proportion of speech on audio by speaker
  - Summary statistics by speaker
  - Probabilities by speaker

- Link with other areas
  - Linguistics (Prosody, Cláudia Brescancini, PUCRS, Brazil)
  - Quantitative Biology (Birds sounds, Rafael de Andrade Moral, Maynooth University, Ireland)
  - Pulmonology (Chronic cough, Cesar A. Keller, Mayo Clinic, USA)
- Automatically identify speakers
  - Proportion of speech on audio by speaker
  - Summary statistics by speaker
  - Probabilities by speaker
- `plot_note` function to intuitively plot frequencies/notes

- Link with other areas
  - Linguistics (Prosody, Cláudia Brescancini, PUCRS, Brazil)
  - Quantitative Biology (Birds sounds, Rafael de Andrade Moral, Maynooth University, Ireland)
  - Pulmonology (Chronic cough, Cesar A. Keller, Mayo Clinic, USA)
- Automatically identify speakers
  - Proportion of speech on audio by speaker
  - Summary statistics by speaker
  - Probabilities by speaker
- `plot_note` function to intuitively plot frequencies/notes
- Chords progression (e.g. major/minor)

- Link with other areas
  - Linguistics (Prosody, Cláudia Brescancini, PUCRS, Brazil)
  - Quantitative Biology (Birds sounds, Rafael de Andrade Moral, Maynooth University, Ireland)
  - Pulmonology (Chronic cough, Cesar A. Keller, Mayo Clinic, USA)
- Automatically identify speakers
  - Proportion of speech on audio by speaker
  - Summary statistics by speaker
  - Probabilities by speaker
- `plot_note` function to intuitively plot frequencies/notes
- Chords progression (e.g. major/minor)
- Build a dashboard to allow easy interaction

- Link with other areas
  - Linguistics (Prosody, Cláudia Brescancini, PUCRS, Brazil)
  - Quantitative Biology (Birds sounds, Rafael de Andrade Moral, Maynooth University, Ireland)
  - Pulmonology (Chronic cough, Cesar A. Keller, Mayo Clinic, USA)
- Automatically identify speakers
  - Proportion of speech on audio by speaker
  - Summary statistics by speaker
  - Probabilities by speaker
- `plot_note` function to intuitively plot frequencies/notes
- Chords progression (e.g. major/minor)
- Build a dashboard to allow easy interaction
  - Allow user to send, link and record audios

UFRGS
PPGPSIQ
Programa de Pós-Graduação em
Psiquiatria e Ciências do Comportamento

THANK YOU!