What we need to do

- Finding data source: listed below.
- Data processing: resources listed below.
- Prediction / classification: gender, age, and potentially accent.

Data source

- https://www.kaggle.com/mozillaorg/common-voice

Additional candidates:

- Google's audio from youtube, with labels ("female singing" "child" etc): https://research.google.com/audioset//ontology/human_voice_1.html
- Speech Accent Archive (2000 files) https://www.kaggle.com/rtatman/speech-accent-archive

Resources

1. Intro materials:

Comment (word of mouth info:

> *voice is datafied in a few ways like wavelets and binary. so features are pretty much like time series analysis and anomaly detection. so for example.... https://medium.com/@ageitgey/machine-learning-is-fun-part-6-how-to-do-speech-recognition-with-deep-learning-28293c162f7a Take a wavelet of saying some words; in the example above this is "hello". then, the key is a series of numbers as in time series and the value is the actual word.*
> *then features can be extracted by*
> - *fast fourier transformation,*
> - *linear predictive cepstral coefficients,*
> - *mel frequency cepstral analysis, etc.*
>
> *If you would like to form a sentence with a few words then you could use hidden markov models or recurrent neural network etc*

Paper "Techniques for feature extraction in speech recognition system" has a list of features & explanations https://arxiv.org/pdf/1305.1145.pdf

- *Linear predictive analysis (LPC)*
- *Linear predictive cepstral coefficients (LPCC)*
- *perceptual linear predictive coefficients (PLP)*
- *Mel-frequency cepstral coefficients (MFCC)*
- *Power spectral analysis (FFT)*
- *Mel scale cepstral analysis (MEL)*
- *Relative spectra filtering of log domain coefficients (RASTA)*
- *First order derivative (DELTA)*

Conclusion section has a list highlighting what each

Packages:

TuneR package doc: https://cran.r-project.org/web/packages/tuneR/tuneR.pdf
Seewave package doc: http://rug.mnhn.fr/seewave/

- Tutorial https://www.r-project.org/conferences/useR-2009/slides/Sueur+Aubin+Simonis.pdf

A very short introduction to sound analysis for those who like elephant trumpet calls or other wildlife sound (good intro of sound data + what FFT/DFT feature is)
https://cran.r-project.org/web/packages/seewave/vignettes/seewave_analysis.pdf

NYU lecture ppt on Sound Classification (no R code. theories/formula. I just skimmed.)
http://www.nyu.edu/classes/bello/ACA_files/8-classification.pdf

Via CLIO:
(For searching papers:) **IEEE Transactions on Speech and Audio Processing** (This place has a lot of related papers from Google search results)
http://ieeexplore.ieee.org.ezproxy.cul.columbia.edu/xpl/RecentIssue.jsp?punumber=89

Fundamentals of Speaker Recognition (a PDF book via columbia library, just as a reference. Cited by the most popular paper, "Musical genre classification of audio signals", in the journal above.)
https://clio.columbia.edu/catalog/9390290?counter=3

- Ch5 Signal Processing of Speech and Feature Extraction: **5.3 and 5.4** has feature explanations.

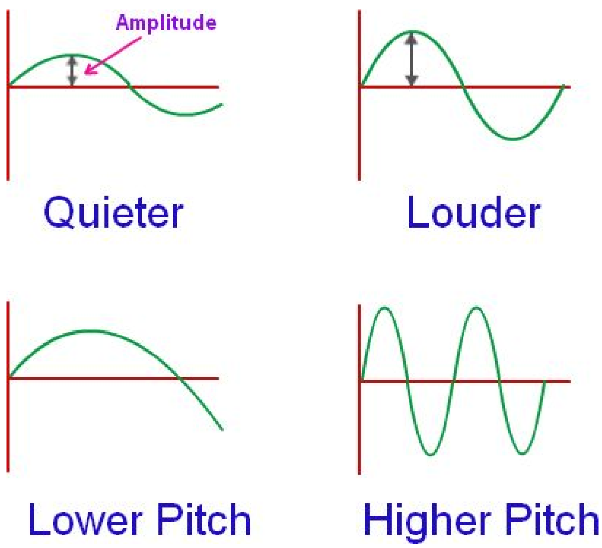
Practical/quick R tutorial:
Basic Sound Processing with R (quick r code tutorial. Analysis, not classification)
http://samcarcagno.altervista.org/blog/basic-sound-processing-r/
Intro to Sound Analysis with R (quick r code tutorial. No classification)
https://www.r-bloggers.com/intro-to-sound-analysis-with-r/

2. Features info & functions

- FFT (fast fourier transformation)
  - FFT function is in stats package.
  - Wikipedia: An FFT algorithm computes the discrete Fourier transform (DFT) of a sequence, or its inverse (IFFT).
  - Fourier Transform: A R Tutorial http://www.di.fc.ul.pt/~jpn/r/fourier/fourier.html
- MFCC (mel frequency cepstral analysis)
  - MFCC function in TuneR package
  - MFCC explanation: http://practicalcryptography.com/miscellaneous/machine-learning/guide-mel-frequency-cepstral-coefficients-mfccs/ (no R example)

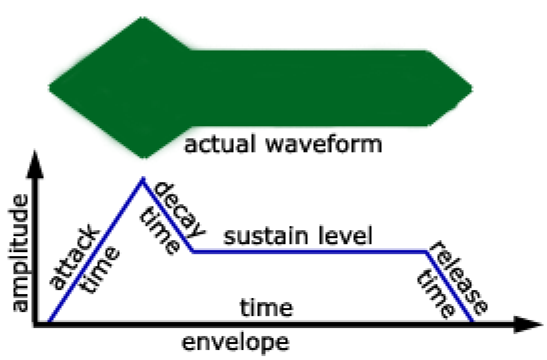
    - Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques https://arxiv.org/pdf/1003.4083.pdf (no R example)
- LPCC (linear predictive cepstral coefficients)
    - http://practicalcryptography.com/miscellaneous/machine-learning/tutorial-cepstrum-and-lpccs/ (no R example)

Amplitude



http://physics.tutorvista.com/waves/amplitude-of-a-wave.html

Amplitude envelope



http://www.indiana.edu/~emusic/etext/acoustics/chapter1_amplitude5.shtml

Frequency spectrum:

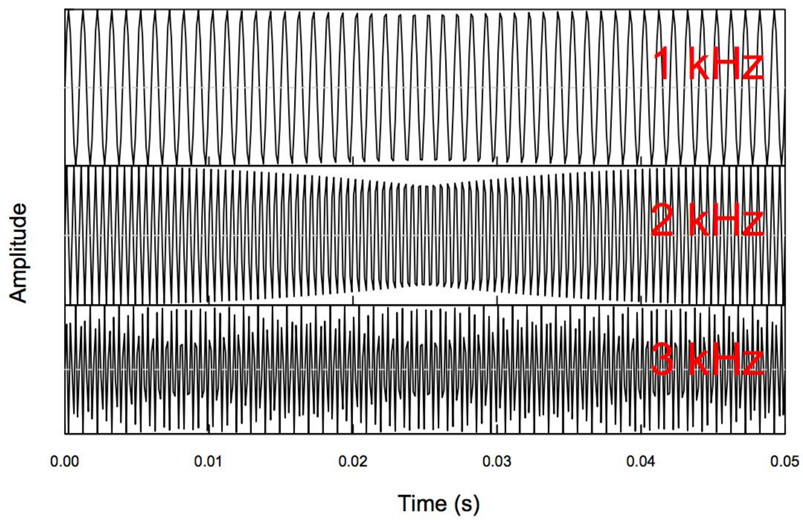


Figure 8: Decomposition of the time wave s into three sine functions. See figure 7.

https://cran.r-project.org/web/packages/seewave/vignettes/seewave_analysis.pdf