

Tutorial to reproduce cluster analysis on the Papuan Malay spontaneous data.

Version 03-2021

This tutorial is meant as a practical guide to reproduce the contour clustering example in the accompanying article (Kaland, 2021, Section 3). For an in-depth description of the clustering procedures, see the manual that accompanies the article.

Files needed (from <https://constantijnkaland.github.io/contourclustering/>):

- contour_clustering_GUI.R
- pmy_spontaneous_phrases.csv

Software needed:

- R (<https://cran.r-project.org/>)
- RStudio (<https://rstudio.com/>)

R-packages needed:

- ggplot2
- reshape
- reshape2
- dplyr
- shiny

Package installation instructions:

- <https://cran.r-project.org/doc/manuals/r-release/R-admin.html#Installing-packages>

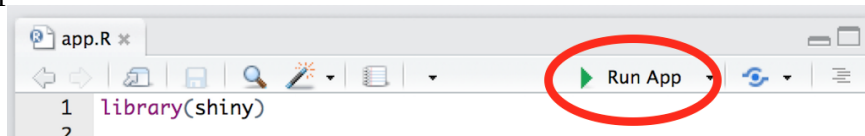
Download all of the above files, and install all the software and packages before continuing this tutorial.

Reference:

Kaland, C.C.L. (2021). Contour clustering: A field-data-driven approach for documenting and analysing prototypical f0 contours. *Journal of the International Phonetic Association*.

1. Loading the contour clustering graphical user interface (GUI)

- Run R-studio
- From the R-studio main screen: File > Open file > Select 'contour_clustering_GUI.R'
- Run app:



2. Load the datafile and prepare the data for clustering

- Load 'pmy_spontaneous_phrases.csv':

http://127.0.0.1:4983 | Open in Browser

Choose datafile

Browse... file selected

Separator
☒ comma ☐ tab

StringsAsFactors
☐ true ☒ false

File encoding
☒ UTF-8 ☐ UTF-16

SkipNul
☐ true ☒ false

Before uploading: select correct file properties above.

Status | Data (long) | Dendrogram | Table | Plot | Data (wide)

- Select speaker correction method 'Standardise':

http://127.0.0.1:5960 | Open in Browser

Choose datafile

Browse... pearstory_pmy.csv

Upload complete

File uploaded. (sep=comma, SaF=F, enc=utf8, skN=F)

☐ clean data (remove NA and f0 errors)

Allowed % change after octave jump correction:
10

Correct speaker differences:
3: standardise

Apply to data

Dendrogram

☐ Keep files after exit

- Apply selected options to data:

http://127.0.0.1:5960 | Open in Browser

Choose datafile

Browse... pearstory_pmy.csv
Upload complete

File uploaded. (sep=comma, SaF=F, enc=utf8, skN=F)

☐ clean data (remove NA and f0 errors)

Allowed % change after octave jump correction:
10

Correct speaker differences:
3: standardise

Apply to data

Dendrogram

☐ Keep files after exit

3. Start the cluster analysis.

- Obtain the dendrogram:

http://127.0.0.1:5960 | Open in Browser

Choose datafile

Browse... pearstory_pmy.csv

Upload complete

File uploaded. (sep=comma, SaF=F, enc=utf8, skN=F)

☐ clean data (remove NA and f0 errors)

Allowed % change after octave jump correction:

10

Correct speaker differences:

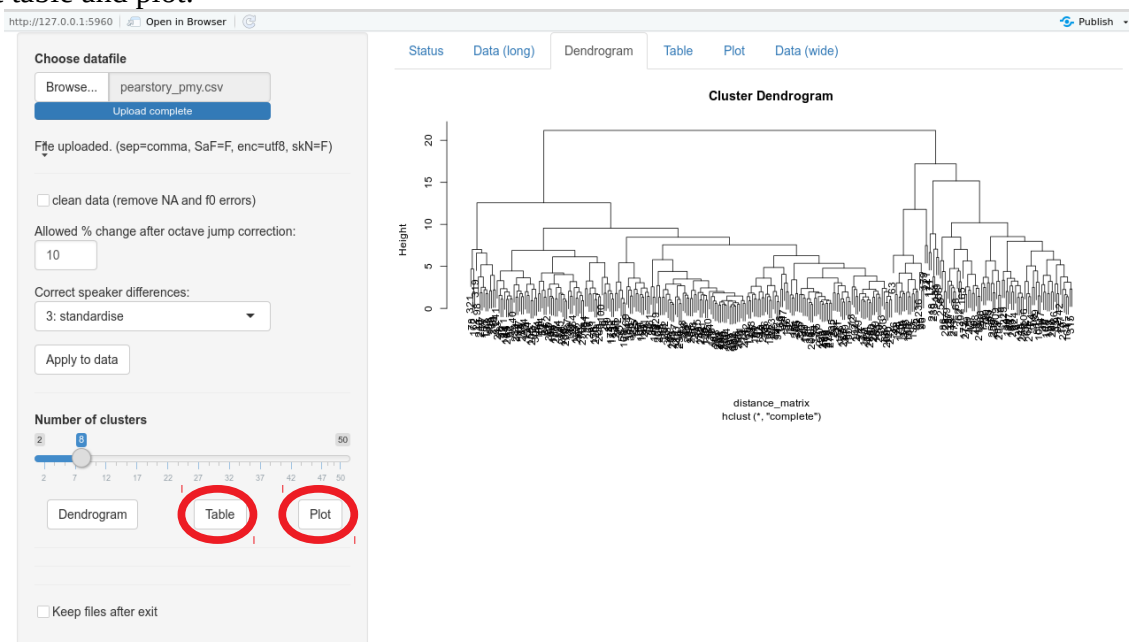
3: standardise

Apply to data

Dendrogram

☐ Keep files after exit

- Inspect table and plot:



- Set number of clusters to 25 and apply subsetting until no clusters are flagged (2x):

http://127.0.0.1:5960 Open in Browser

Choose datafile

Browse... pearstory_pmy.csv

Upload complete

File uploaded. (sep=comma, SaF=F, enc=utf8, skN=F)

☐ clean data (remove NA and f0 errors)

Allowed % change after octave jump correction:

10

Correct speaker differences:

3: standardise

Apply to data

Number of clusters

2 25 50

Dendrogram Table Plot

Remove these clusters:

<input type="checkbox"/> 1	<input type="checkbox"/> 2	<input type="checkbox"/> 3	<input type="checkbox"/> 4	<input type="checkbox"/> 5	<input type="checkbox"/> 6	<input type="checkbox"/> 7	<input type="checkbox"/> 8
<input type="checkbox"/> 9	<input type="checkbox"/> 10	<input type="checkbox"/> 11	<input type="checkbox"/> 12	<input type="checkbox"/> 13	<input type="checkbox"/> 14	<input checked="" type="checkbox"/> 15	
<input type="checkbox"/> 16	<input type="checkbox"/> 17	<input checked="" type="checkbox"/> 18	<input checked="" type="checkbox"/> 19	<input checked="" type="checkbox"/> 20	<input checked="" type="checkbox"/> 21	<input checked="" type="checkbox"/> 22	
<input checked="" type="checkbox"/> 23	<input checked="" type="checkbox"/> 24	<input checked="" type="checkbox"/> 25					

Apply subsetting

Generate textgrids

☐ Keep files after exit

- Set number of clusters back to 9 and obtain plot (identical to Figure 3d in Kaland 2021).

http://127.0.0.1:5960 | Open in Browser

Choose datafile

Browse... pearstory_pmy.csv
Upload complete

File uploaded. (sep=comma, SaF=F, enc=utf8, skN=F)

☐ clean data (remove NA and f0 errors)

Allowed % change after octave jump correction:
10

Correct speaker differences:
3: standardise

Apply to data

Number of clusters

2 9 50
2 7 17 22 27 32 37 42 47 50

Dendrogram Table Plot

Remove these clusters:

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8
☐ 9

Apply subsetting

Generate textgrids

☐ Keep files after exit