Data Science Final project

Laura Conover, University of South Florida, Department of Communication Sciences and Disorders

# Abstract

Human speech sounds are full of variation. Some of this variation is phonemic - to English speakers, the difference between the unvoiced ‘p’ sound in ‘pug’ and the voiced ‘b’ sound in ‘bug’ completely changes the meaning of the resulting word. On the other hand, not all variation within speech sounds carry meaning, such as the contrast between the aspirated ‘p’ sound in ‘pot’ and the unaspirated ‘p’ sound in ‘spot’ for English speakers. Although the type of contrast that is considered important varies between languages (aspiration contrasts are meaningful in languages such as Hindu, Thai, and Punjabi), it is a general rule that adults have a difficult time perceiving non-phonemic contrasts. It is possible, however, that certain acoustic markers may interact with individual listener characteristics to make some non-phonemic contrasts more easily perceptible for some listeners than others. The following study examines such interactions.

# Acknowledgements

Much thanks to Richard Layton for his [tutorial](https://rmarkdown.rstudio.com/articles_docx.html) on a work-around for formatting when the papaja package was too frustrating.

# Introduction (and disclaimer)

The data selected comes from a study of individuals’ perceptions of both spontaneous and volitional laughter produced by young adult female American English speakers. The discrepency between this description and what was mentioned in the abstract is obvious. Clearly, this is not at all the data set that will be used in the final study. However, the structure of the data and variables are identical to the structure of the study’s data: responses to a 2A-FC test, categorical listener variables, and continuous acoustic variables. Therefore, this data is a good stand-in until actual data collection is complete. Similarly, not all of the descriptive text is finalized. I hope that what is written here and in the folder’s README file will provide sufficient explanation.

# Methods

The original data includes two csv files: one detailing information on the acoustic features of the laugh stimuli (including spontaneous/volitional condition and duration, voicing, pitch, and intensity measures), and the other detailing information on the listeners and perceptions (including basic demographic data such as gender, country of origin, English fluency, and education, as well as whether each stimulus was perceived as volitional or spontaneous). This dataset has no missing values; it is possible that any cases with missing values were deleted listwise prior to publication. Even so, these data files are somewhat messy. Participant IDs were reused at different sites, and stimulus files were renamed between analyses. Additionally, there are a number of variables that are not relevant to the current project. As such, the data were slightly cleaned and reorganized prior to analysis.

Selected variables for this analysis include: Laugh condition (spontaneous or volitional), accuracy of listener’s judgement on a trial, total laugh duration in seconds, the average pitch of the laugh in logHz, and the listener’s country of origin.

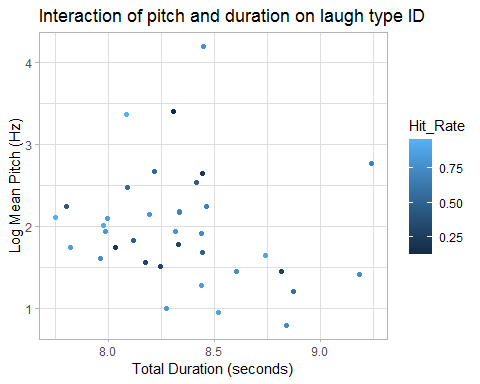
## Data analysis

The average accuracy by listener’s country of origin is presented in Table 1 below. Although the scores do vary from country to country, the range of scores only reaches from 0.5621693 to 0.6856725, with an average accuracy of 0.6371769.

Table 1. The average hit rate of laugh identifications by listener’s country of origin

| country | accuracy |
| --- | --- |
| Australia | 0.65 |
| Austria | 0.67 |
| Canada | 0.61 |
| China | 0.64 |
| Ecuador/Shuar | 0.57 |
| India | 0.59 |
| Indonesia | 0.64 |
| Iran | 0.63 |
| Japan | 0.69 |
| Korea | 0.68 |
| Netherlands | 0.68 |
| Peru (Rural) | 0.61 |
| Peru (Urban) | 0.65 |
| Qatar | 0.64 |
| Samoa | 0.56 |
| Singapore | 0.62 |
| Slovakia | 0.64 |
| South Africa/Zulu | 0.60 |
| Spain | 0.65 |
| Turkey | 0.67 |
| USA | 0.67 |

The influence of acoustic features on hit rates - or, rather, the apparent lack thereof - can be seen in Figure 1:



Note that, in Table 2 below, there is no significant difference in either pitch or duration, or their interaction, between the two stimulus conditions. That is, there is no acoustic difference between spontaneous and volitional laughter within these two measures.

Table 2. Model of Laugh Condition by Pitch and Duration

| predictor | estimate | ci | statistic | p.value |
| --- | --- | --- | --- | --- |
| Intercept | 34.41 | $[-60.53$, $139.33]$ | 0.70 | .486 |
| Pitch | -4.04 | $[-16.44$, $7.23]$ | -0.69 | .489 |
| Duration | -39.42 | $[-104.27$, $12.31]$ | -1.34 | .179 |
| Pitch $\times$ Duration | 4.70 | $[-1.46$, $12.41]$ | 1.35 | .177 |

However, the human perceptual system is modified by more factors than just acoustic information. As seen in Table 3 below, the listener’s country of origin does impact whether or not the listener can correctly perceive a stimulus as spontaneous or volitional, , , .

Table 3. Analysis of Variance of Hit Rate by Country

| predictor | estimate | ci | statistic | p.value |
| --- | --- | --- | --- | --- |
| Intercept | 0.65 | $[0.63$, $0.68]$ | 48.95 | < .001 |
| CountryAustria | 0.02 | $[-0.02$, $0.06]$ | 1.06 | .287 |
| CountryCanada | -0.05 | $[-0.08$, $-0.01]$ | -2.41 | .016 |
| CountryChina | -0.01 | $[-0.05$, $0.03]$ | -0.47 | .637 |
| CountryEcuador/Shuar | -0.09 | $[-0.12$, $-0.05]$ | -4.56 | < .001 |
| CountryIndia | -0.06 | $[-0.10$, $-0.02]$ | -3.21 | .001 |
| CountryIndonesia | -0.01 | $[-0.05$, $0.02]$ | -0.65 | .514 |
| CountryIran | -0.02 | $[-0.05$, $0.02]$ | -1.05 | .294 |
| CountryJapan | 0.03 | $[0.00$, $0.07]$ | 1.97 | .049 |
| CountryKorea | 0.03 | $[0.00$, $0.06]$ | 1.87 | .061 |
| CountryNetherlands | 0.02 | $[-0.01$, $0.06]$ | 1.41 | .158 |
| CountryPeru Rural | -0.05 | $[-0.08$, $-0.01]$ | -2.38 | .018 |
| CountryPeru Urban | 0.00 | $[-0.04$, $0.04]$ | 0.05 | .961 |
| CountryQatar | -0.02 | $[-0.05$, $0.02]$ | -0.82 | .410 |
| CountrySamoa | -0.09 | $[-0.13$, $-0.06]$ | -5.12 | < .001 |
| CountrySingapore | -0.03 | $[-0.07$, $0.00]$ | -1.86 | .063 |
| CountrySlovakia | -0.01 | $[-0.04$, $0.02]$ | -0.79 | .429 |
| CountrySouth Africa/Zulu | -0.05 | $[-0.08$, $-0.02]$ | -3.41 | .001 |
| CountrySpain | 0.00 | $[-0.04$, $0.04]$ | 0.06 | .949 |
| CountryTurkey | 0.02 | $[-0.02$, $0.06]$ | 1.06 | .287 |
| CountryUSA | 0.02 | $[-0.01$, $0.05]$ | 1.21 | .227 |

# Session Information

devtools::session\_info()

## - Session info ---------------------------------------------------------------  
## setting value   
## version R version 3.6.2 (2019-12-12)  
## os Windows 8.1 x64   
## system x86\_64, mingw32   
## ui RTerm   
## language (EN)   
## collate English\_United States.1252   
## ctype English\_United States.1252   
## tz America/New\_York   
## date 2021-03-29   
##   
## - Packages -------------------------------------------------------------------  
## package \* version date lib source   
## assertthat 0.2.1 2019-03-21 [1] CRAN (R 3.6.2)   
## backports 1.2.0 2020-11-02 [1] CRAN (R 3.6.3)   
## base64enc 0.1-3 2015-07-28 [1] CRAN (R 3.6.0)   
## broom 0.7.2 2020-10-20 [1] CRAN (R 3.6.3)   
## callr 3.5.1 2020-10-13 [1] CRAN (R 3.6.3)   
## cellranger 1.1.0 2016-07-27 [1] CRAN (R 3.6.2)   
## cli 2.2.0 2020-11-20 [1] CRAN (R 3.6.3)   
## colorspace 2.0-0 2020-11-11 [1] CRAN (R 3.6.3)   
## crayon 1.3.4 2017-09-16 [1] CRAN (R 3.6.2)   
## data.table 1.13.2 2020-10-19 [1] CRAN (R 3.6.3)   
## DBI 1.1.0 2019-12-15 [1] CRAN (R 3.6.2)   
## dbplyr 2.0.0 2020-11-03 [1] CRAN (R 3.6.3)   
## desc 1.2.0 2018-05-01 [1] CRAN (R 3.6.2)   
## devtools 2.3.2 2020-09-18 [1] CRAN (R 3.6.3)   
## digest 0.6.27 2020-10-24 [1] CRAN (R 3.6.3)   
## dplyr \* 1.0.2 2020-08-18 [1] CRAN (R 3.6.3)   
## ellipsis 0.3.1 2020-05-15 [1] CRAN (R 3.6.3)   
## evaluate 0.14 2019-05-28 [1] CRAN (R 3.6.2)   
## fansi 0.4.1 2020-01-08 [1] CRAN (R 3.6.2)   
## farver 2.0.3 2020-01-16 [1] CRAN (R 3.6.2)   
## flextable \* 0.6.0 2020-12-01 [1] CRAN (R 3.6.3)   
## forcats \* 0.5.0 2020-03-01 [1] CRAN (R 3.6.3)   
## fs 1.5.0 2020-07-31 [1] CRAN (R 3.6.3)   
## gdtools \* 0.2.2 2020-04-03 [1] CRAN (R 3.6.3)   
## generics 0.1.0 2020-10-31 [1] CRAN (R 3.6.3)   
## ggplot2 \* 3.3.2 2020-06-19 [1] CRAN (R 3.6.3)   
## glue 1.4.2 2020-08-27 [1] CRAN (R 3.6.3)   
## gtable 0.3.0 2019-03-25 [1] CRAN (R 3.6.2)   
## haven 2.3.1 2020-06-01 [1] CRAN (R 3.6.3)   
## here \* 1.0.0 2020-11-15 [1] CRAN (R 3.6.3)   
## hms 0.5.3 2020-01-08 [1] CRAN (R 3.6.2)   
## htmltools 0.5.0 2020-06-16 [1] CRAN (R 3.6.3)   
## httr 1.4.2 2020-07-20 [1] CRAN (R 3.6.3)   
## jsonlite 1.7.1 2020-09-07 [1] CRAN (R 3.6.3)   
## knitr 1.30 2020-09-22 [1] CRAN (R 3.6.3)   
## labeling 0.4.2 2020-10-20 [1] CRAN (R 3.6.3)   
## lifecycle 0.2.0 2020-03-06 [1] CRAN (R 3.6.3)   
## lubridate 1.7.9.2 2020-11-13 [1] CRAN (R 3.6.3)   
## magrittr 2.0.1 2020-11-17 [1] CRAN (R 3.6.3)   
## MASS 7.3-51.4 2019-03-31 [2] CRAN (R 3.6.2)   
## memoise 1.1.0 2017-04-21 [1] CRAN (R 3.6.2)   
## modelr 0.1.8 2020-05-19 [1] CRAN (R 3.6.3)   
## munsell 0.5.0 2018-06-12 [1] CRAN (R 3.6.2)   
## officer 0.3.15 2020-11-01 [1] CRAN (R 3.6.3)   
## papaja \* 0.1.0.9942 2020-04-27 [1] Github (crsh/papaja@b0a224a)  
## pillar 1.4.7 2020-11-20 [1] CRAN (R 3.6.3)   
## pkgbuild 1.1.0 2020-07-13 [1] CRAN (R 3.6.3)   
## pkgconfig 2.0.3 2019-09-22 [1] CRAN (R 3.6.2)   
## pkgload 1.1.0 2020-05-29 [1] CRAN (R 3.6.3)   
## prettyunits 1.1.1 2020-01-24 [1] CRAN (R 3.6.2)   
## processx 3.4.5 2020-11-30 [1] CRAN (R 3.6.3)   
## ps 1.5.0 2020-12-05 [1] CRAN (R 3.6.2)   
## purrr \* 0.3.4 2020-04-17 [1] CRAN (R 3.6.3)   
## R6 2.5.0 2020-10-28 [1] CRAN (R 3.6.3)   
## Rcpp 1.0.5 2020-07-06 [1] CRAN (R 3.6.3)   
## readr \* 1.4.0 2020-10-05 [1] CRAN (R 3.6.3)   
## readxl 1.3.1 2019-03-13 [1] CRAN (R 3.6.2)   
## remotes 2.2.0 2020-07-21 [1] CRAN (R 3.6.3)   
## reprex 0.3.0 2019-05-16 [1] CRAN (R 3.6.2)   
## rlang 0.4.9 2020-11-26 [1] CRAN (R 3.6.3)   
## rmarkdown 2.5 2020-10-21 [1] CRAN (R 3.6.3)   
## rprojroot 2.0.2 2020-11-15 [1] CRAN (R 3.6.3)   
## rstudioapi 0.13 2020-11-12 [1] CRAN (R 3.6.3)   
## rvest 0.3.6 2020-07-25 [1] CRAN (R 3.6.3)   
## scales 1.1.1 2020-05-11 [1] CRAN (R 3.6.3)   
## sessioninfo 1.1.1 2018-11-05 [1] CRAN (R 3.6.2)   
## stringi 1.5.3 2020-09-09 [1] CRAN (R 3.6.2)   
## stringr \* 1.4.0 2019-02-10 [1] CRAN (R 3.6.2)   
## svglite \* 1.2.3.2 2020-07-07 [1] CRAN (R 3.6.3)   
## systemfonts 0.3.2 2020-09-29 [1] CRAN (R 3.6.3)   
## testthat 3.0.0 2020-10-31 [1] CRAN (R 3.6.3)   
## tibble \* 3.0.4 2020-10-12 [1] CRAN (R 3.6.3)   
## tidyr \* 1.1.2 2020-08-27 [1] CRAN (R 3.6.3)   
## tidyselect 1.1.0 2020-05-11 [1] CRAN (R 3.6.3)   
## tidyverse \* 1.3.0 2019-11-21 [1] CRAN (R 3.6.2)   
## usethis 1.6.3 2020-09-17 [1] CRAN (R 3.6.3)   
## uuid 0.1-4 2020-02-26 [1] CRAN (R 3.6.3)   
## vctrs 0.3.5 2020-11-17 [1] CRAN (R 3.6.3)   
## withr 2.3.0 2020-09-22 [1] CRAN (R 3.6.3)   
## xfun 0.19 2020-10-30 [1] CRAN (R 3.6.3)   
## xml2 1.3.2 2020-04-23 [1] CRAN (R 3.6.2)   
## yaml 2.2.1 2020-02-01 [1] CRAN (R 3.6.2)   
## zip 2.1.1 2020-08-27 [1] CRAN (R 3.6.3)   
##   
## [1] C:/Users/Me/Documents/R/win-library/3.6  
## [2] C:/Program Files/R/R-3.6.2/library