

Access Wideband Audiology Immittance database using R and dplyr (Voss PI)

Nicholas Horton (nhorton@amherst.edu)

June 8, 2021

Introduction

This document is intended to describe how to access data from a MySQL database using R. It utilizes a database of wideband acoustic immittance variables from humans with normal hearing (see https://projectreporder.nih.gov/project_info_description.cfm?aid=8769352&icde=30039221&ddparam=&ddvalue=&ddsub=&cr=10&csb=default&cs=ASC for more details).

A relevant paper on the topic of data management and databases in R can be found at <http://chance.amstat.org/2015/04/setting-the-stage>.

Accessing data from a database using SQL commands

First I will demonstrate how to access data using SQL (structured query language) commands and the `dbGetQuery()` function. We begin by setting up a connection to the database.

```
library(mosaic)
library(tidyverse)
library(RMySQL)
con <- dbConnect(MySQL(), host = "scidb.smith.edu",
                  user = "waiuser", password = "smith_waiDB",
                  dbname = "wai")
```

Next a series of SQL queries can be sent to the database. These return R dataframes.

```
dbGetQuery(con, "SHOW TABLES")
```

```
##          Tables_in_wai
## 1          Codebook
## 2          Measurements
## 3 Measurements_pre2020
## 4          PI_Info
## 5          PI_Info_OLD
## 6          Subjects
## 7 Subjects_pre2020
```

```
dbGetQuery(con, "EXPLAIN PI_Info")
```

```
##          Field          Type Null Key Default Extra
## 1 Identifier varchar(50)  NO  PRI   <NA>
## 2 Year          int(11)   NO
## 3 Authors       text     NO
## 4 AuthorsShortList text     NO
## 5 Title         text     NO
```

```
## 6          Journal      text  NO      <NA>
## 7            URL       text  NO      <NA>
## 8      Abstract       text  NO      <NA>
## 9  DataSubmitterName  text  NO      <NA>
## 10 DataSubmitterEmail text  NO      <NA>
## 11      DateSubmitted  text  NO      <NA>
## 12          PI_Notes   text  NO      <NA>
```

```
dbGetQuery(con, "EXPLAIN Subjects")
```

```
##           Field          Type Null Key Default Extra
## 1      Identifier varchar(50)  NO PRI  <NA>
## 2      SubjectNumber  int(11)  NO PRI  <NA>
## 3      SessionTotal   int(11)  NO      <NA>
## 4      AgeFirstMeasurement float YES      <NA>
## 5      AgeCategoryFirstMeasurement varchar(50) YES      <NA>
## 6              Sex varchar(50)  NO      <NA>
## 7              Race varchar(50)  NO      <NA>
## 8              Ethnicity varchar(50)  NO      <NA>
## 9      LeftEarStatusFirstMeasurement varchar(50)  NO      <NA>
## 10     RightEarStatusFirstMeasurement varchar(50)  NO      <NA>
## 11              SubjectNotes      text YES      <NA>
```

```
dbGetQuery(con, "EXPLAIN Measurements")
```

```
##           Field          Type Null Key Default Extra
## 1      Identifier varchar(50)  NO PRI  <NA>
## 2      SubjectNumber  int(11)  NO PRI  <NA>
## 3      Session      int(11)  NO PRI  <NA>
## 4      Ear          varchar(50)  NO PRI
## 5      Instrument  varchar(50)  NO PRI
## 6      Age          float YES      <NA>
## 7      AgeCategory varchar(50)  YES      <NA>
## 8      EarStatus   varchar(50)  YES      <NA>
## 9      TPP         float YES      <NA>
## 10     AreaCanal   float YES      <NA>
## 11     PressureCanal float NO PRI  0
## 12     SweepDirection varchar(50)  NO PRI
## 13     Frequency   float NO PRI  0
## 14     Absorbance   float YES      <NA>
## 15     Zmag        float YES      <NA>
## 16     Zang        float YES      <NA>
```

```
ds <- dbGetQuery(con, "SELECT * from Measurements LIMIT 10")
```

```
ds
```

```
## Identifier SubjectNumber Session Ear Instrument Age AgeCategory EarStatus
## 1 Abur_2014 1 1 Left HearID 20 Adult Normal
## 2 Abur_2014 1 1 Left HearID 20 Adult Normal
## 3 Abur_2014 1 1 Left HearID 20 Adult Normal
## 4 Abur_2014 1 1 Left HearID 20 Adult Normal
## 5 Abur_2014 1 1 Left HearID 20 Adult Normal
## 6 Abur_2014 1 1 Left HearID 20 Adult Normal
## 7 Abur_2014 1 1 Left HearID 20 Adult Normal
## 8 Abur_2014 1 1 Left HearID 20 Adult Normal
## 9 Abur_2014 1 1 Left HearID 20 Adult Normal
```

```
## 10 Abur_2014          1      1 Left      HearID 20      Adult      Normal
##      TPP AreaCanal PressureCanal SweepDirection Frequency Absorbance      Zmag
## 1   -5  4.42e-05          0      Ambient    210.938  0.0333379 113780000
## 2   -5  4.42e-05          0      Ambient    234.375  0.0315705 103585000
## 3   -5  4.42e-05          0      Ambient    257.812  0.0405751  92951700
## 4   -5  4.42e-05          0      Ambient    281.250  0.0438399  86058000
## 5   -5  4.42e-05          0      Ambient    304.688  0.0486400  79492800
## 6   -5  4.42e-05          0      Ambient    328.125  0.0527801  73326200
## 7   -5  4.42e-05          0      Ambient    351.562  0.0583192  68793600
## 8   -5  4.42e-05          0      Ambient    375.000  0.0638881  64088600
## 9   -5  4.42e-05          0      Ambient    398.438  0.0687025  60200600
## 10  -5  4.42e-05          0      Ambient    421.875  0.0833181  56990900
##      Zang
## 1  -0.233504
## 2  -0.235778
## 3  -0.233482
## 4  -0.233421
## 5  -0.232931
## 6  -0.232837
## 7  -0.232115
## 8  -0.231642
## 9  -0.231356
## 10 -0.228356
```

Accessing a database using dplyr commands

Alternatively, a connection can be made to the server by creating a series of dplyr table objects.

```
db <- DBI::dbConnect(RMySQL::MySQL(), dbname = "wai", host = "scidb.smith.edu", user = "waiuser",
                    password="smith_waiDB")
Measurements <- tbl(db, "Measurements")
PI_Info <- tbl(db, "PI_Info")
Subjects <- tbl(db, "Subjects")
```

```
PI_Info %>% collect() %>% summarise(total = n())
```

Let's explore the PI_Info table.

```
## # A tibble: 1 x 1
##   total
##   <int>
## 1     17
```

```
PI_Info %>% collect() %>% data.frame() # collect() is a bad idea when dealing with large tables!
```

```
##      Identifier Year
## 1     Abur_2014 2014
## 2     Feeney_2017 2017
## 3     Groon_2015 2015
## 4     Hunter_2016 2016
## 5     Keefe_2017 2017
## 6     Lewis_2015 2015
## 7     Lewis_2018 2018
## 8       Liu_2008 2008
## 9   Merchant_2010 2010
```

```

## 10 Rosowski_2012 2012
## 11 Shahnaz_2006 2006
## 12 Shaver_2013 2013
## 13 Sun_2016 2016
## 14 Voss_1994 1994
## 15 Voss_2010 2010
## 16 Voss_2016 2016
## 17 Werner_2010 2010
##
## 1
## 2 M. Patrick Feeney, Douglas H. Keefe, Lisa L. Hunter, Denis F. Fitzpatrick, Angela C. Garinis,
## 3 Katherine A. Groon, Daniel M. Rasetshwane, Judy G. Keefe,
## 4 Lisa L. Hunter, Douglas H. Keefe, M. Patrick Feeney,
## 5 Douglas H. Keefe, Kelly L. Archer, Kendra K. Schmid, Denis F. Fitzpatrick,
## 6
## 7
## 8 Yi-Wen Liu, Chris A. Sanford, John C. Ellison, Denis F. Fitzpatrick,
## 9 Gabrielle R. Merchant,
## 10 John J. Rosowski, Hideko H. Nakajima, Mohamad A. Hamade, Lorice Mahfoud, Gabrielle R. Merchant,
## 11
## 12
## 13
## 14
## 15 Susan E. Voss, Modupe F. Adegoke, Nicholas J. Horton, Kevin N. Shelton,
## 16 Susan E. Voss, Barbara S. Herrmann, Nicholas J. Horton,
## 17 Lynne A. Tyler
##
## AuthorsShortList
## 1 Abur et al.
## 2 Feeney et al.
## 3 Groon et al.
## 4 Hunter et al.
## 5 Keefe et al.
## 6 Lewis and Neely
## 7 Lewis
## 8 Liu et al.
## 9 Merchant et al.
## 10 Rosowski et al.
## 11 Shahnaz and Bork
## 12 Shaver and Sun
## 13 Sun
## 14 Voss and Allen
## 15 Voss et al.
## 16 Voss et al.
## 17 Werner et al.
##
## 1
## 2 Normative wideband reflectance, equivalent admittance at the tympanic membrane, and acoustic
## 3 Air-leak effects
## 4 Longitudinal development of wideband reflectance tympanometry
## 5 Identifying Otosclerosis with Aural Acoustical Tests of Absorbance, Group Delay, Acoustic Reflex
## 6 Non-invasive estimation of middle ear admittance
## 7 The area discontinuity between probe and ear canal as a source of power-reflectance
## 8 Wideband absorbance tympanometry using pressure sweeps: System development and validation
## 9 Normative Reflectance and Transmittance Measurements on Human Ears

```

```

## 10 Ear-canal reflectance, umbo velocity, and
## 11 Wideband reflectance norms f
## 12 Wideband energy reflectance measurements: Effects of negative middle ear pressure and application
## 13 Wideband acoustic immittance: Normative study and test-retest reliability of
## 14 Measurement of acoustic impedance a
## 15 Posture systematically alters ear-c
## 16 Reflectance Measures from Infant Ears With Normal Hearing a
## 17 Ear-canal wideband acoustic transfer functions of adul
## Journal
## 1 J Am Acad Audiol
## 2 Ear Hear
## 3 Ear Hear
## 4 Hear Res
## 5 J Am Acad Audiol
## 6 J Acoust Soc Am
## 7 J Acoust Soc Am
## 8 J Acoust Soc Am
## 9 Ear Hear
## 10 Ear Hear
## 11 Ear Hear
## 12 J Acoust Soc Am
## 13 J Speech Lang Hear Res
## 14 J Acoust Soc Am
## 15 Hear Res
## 16 Ear Hear
## 17 Ear Hear
##
## 1 https://www.ncbi.nlm.nih.gov
## 2 https://www.ncbi.nlm.nih.gov
## 3 https://journals.lww.com/ear-hearing/fulltext/2015/01000/Air_Leak_Effects_on_Ear_Canal_Acoustic_A
## 4 https://pubmed.ncbi.nlm
## 5 https://pubmed.ncbi.nlm.n
## 6 https://asa.scitation.org/doi/abs/
## 7 https://doi.org/
## 8 https://www.ncbi.nlm.nih.gov
## 9 https://pubmed.ncbi.nlm.n
## 10 http://www.ncbi.nlm.nih.gov
## 11 http://journals.lww.com/ear-hearing/Abstract/2006/12000/Wideband_Reflectance_Norms_for_Cauc
## 12 "\nhttps://asa.scitation.org/doi/full/1
## 13 https://www.ncbi.nlm.nih.gov
## 14 https://asa.scitation.org/doi/abs
## 15 https://www.ncbi.nlm.nih.gov
## 16 https://pubmed.ncbi.nlm.n
## 17 https://www.ncbi.nlm.nih.gov
##
## 1
## 2 "<p> <strong> Objectives: </strong> Wideband acoustic immittance (WAI) measures such as\npressure
## 3
## 4
## 5
## 6
## 7
## 8
## 9

```

```

## 10
## 11
## 12
## 13
## 14
## 15
## 16
## 17
##
##           DataSubmitterName
## 1           Susan Voss
## 2 M. Patrick Feeney; Douglas H. Keefe
## 3           Steve Neely
## 4           Lisa Hunter
## 5           Douglas Keefe
## 6           James Lewis
## 7           James D. Lewis
## 8           Douglas Keefe
## 9           Susan Voss
## 10          John Rosowski
## 11          Navid Shahnaz
## 12          Xiao-Ming Sun
## 13          Xiao-Ming Sun
## 14          Susan Voss
## 15          Susan Voss
## 16          Susan Voss
## 17          Douglas Keefe
##
##           DataSubmitterEmail DateSubmitted
## 1           svoss@smith.edu      8/24/16
## 2 Patrick.Feeney@va.gov; Douglas.Keefe@boystown.org 6/7/18
## 3           Stephen.Neely@boystown.org 6/18/19
## 4           Lisa.Hunter@cchmc.org 7/16/2020
## 5           Douglas.Keefe@boystown.org 9/1/2020
## 6           jdlewis@uthsc.edu     10/10/18
## 7           jdlewis@uthsc.edu     8/10/20
## 8           Douglas.Keefe@boystown.org 6/26/18
## 9           svoss@smith.edu      11/1/20
## 10          John_Rosowski@meei.harvard.edu 11/6/15
## 11          nshahnaz@audiospeech.ubc.ca 8/24/16
## 12          xiao-ming.sun@wichita.edu 10/6/18
## 13          xiao-ming.sun@wichita.edu 10/31/17
## 14          svoss@smith.edu      2/8/17
## 15          svoss@smith.edu      6/5/18
## 16          svoss@smith.edu      12/1/20
## 17          Douglas.Keefe@boystown.org 9/1/17
##
## 1
## 2
## 3
## 4
## 5
## 6
## 7
## 8
## 9

```

```
## 10 "HearID (Mimosa Acoustics); \nNormal Criteria as follows: \n(1) There was no history of significant hearing loss in either ear."
## 11
## 12
## 13
## 14
## 15
## 16
## 17
```

```
Subjects %>% collect() # be careful with collect() with large tables!
```

Let's explore the Subjects table.

```
## # A tibble: 930 x 11
##   Identifier SubjectNumber SessionTotal AgeFirstMeasure~ AgeCategoryFirs~ Sex
##   <chr>          <int>         <int>         <dbl> <chr>          <chr>
## 1 Abur_2014           1             7             20 Adult          Fema~
## 2 Abur_2014           3             8             19 Adult          Fema~
## 3 Abur_2014           4             7             21 Adult          Fema~
## 4 Abur_2014           6             8             21 Adult          Fema~
## 5 Abur_2014           7             5             20 Adult          Fema~
## 6 Abur_2014           8             5             19 Adult          Fema~
## 7 Abur_2014          10             5             19 Adult          Fema~
## 8 Feeney_20~         103             2             NA Adult          Unkn~
## 9 Feeney_20~         104             2             NA Adult          Unkn~
## 10 Feeney_20~        106             2             NA Adult          Unkn~
## # ... with 920 more rows, and 5 more variables: Race <chr>, Ethnicity <chr>,
## #   LeftEarStatusFirstMeasurement <chr>, RightEarStatusFirstMeasurement <chr>,
## #   SubjectNotes <chr>
```

```
Measurements %>% summarise(total = n())
```

Let's explore the Measurements table.

```
## Error in .local(conn, statement, ...): could not run statement: Unknown column '' in 'field list'
mtcars %>% summarise(total = n())
```

```
## total
## 1 32
```

```
onesubj <-
  Measurements %>%
  filter(Identifier == "Rosowski_2012", SubjectNumber == 3) %>%
  collect %>%
  mutate(SessionNum = as.factor(Session))
head(onesubj)
```

Let's download the data from a given subject

```
## # A tibble: 6 x 17
##   Identifier SubjectNumber Session Ear Instrument Age AgeCategory EarStatus
##   <chr>          <int>         <int> <chr> <chr>         <dbl> <chr>          <chr>
## 1 Rosowski_2~           3             1 "Lef~ HearID           30 Adult          Normal
```

```
## 2 Rosowski_2~          3      1 "Lef~ HearID      30 Adult      Normal
## 3 Rosowski_2~          3      1 "Lef~ HearID      30 Adult      Normal
## 4 Rosowski_2~          3      1 "Lef~ HearID      30 Adult      Normal
## 5 Rosowski_2~          3      1 "Lef~ HearID      30 Adult      Normal
## 6 Rosowski_2~          3      1 "Lef~ HearID      30 Adult      Normal
## # ... with 9 more variables: TPP <dbl>, AreaCanal <dbl>, PressureCanal <dbl>,
## #   SweepDirection <chr>, Frequency <dbl>, Absorbance <dbl>, Zmag <dbl>,
## #   Zang <dbl>, SessionNum <fct>
```

Finally we can plot the results

```
ggplot(data = onesubj, aes(x = Frequency, y = Absorbance)) + geom_point() +
  aes(colour = Ear) + scale_x_log10() + labs(title="Absorbance by ear Rosowski subject 3")
```

