

# Access Wideband Audiology Immitance database using R and dplyr

*Nicholas Horton (nhorton@amherst.edu)*

*November 14, 2017*

## Introduction

This document is intended to describe how to access data from a MySQL database using R. It utilizes a database of wideband acoustic immitance variables from humans with normal hearing (see [https://projectreporter.nih.gov/project\\_info\\_description.cfm?aid=8769352&icde=30039221&ddparam=&ddvalue=&ddsub=&cr=10&csb=default&cs=ASC](https://projectreporter.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(RMySQL)
```

```
## Loading required package: DBI
```

```
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 Measurements
## 2 PI_Info
## 3 Subject
```

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

```
##      Field      Type Null Key Default Extra
## 1 Identifier varchar(20) YES      <NA>
## 2 PI_Year      int(11) YES      <NA>
## 3 PI           varchar(500) YES      <NA>
## 4 Affiliation  varchar(500) YES      <NA>
## 5 Email        varchar(30) YES      <NA>
## 6 Title        varchar(140) YES      <NA>
```

```
## 7      Pub   varchar(30) YES      <NA>
## 8      Date   char(20) YES      <NA>
## 9      URL   varchar(140) YES     <NA>
## 10     PI_Notes varchar(1500) YES  <NA>
```

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

```
## Identifier Sub_Number Session Left_Ear MEP Instrument Freq
## 1 Abur_2014      1      1      0 -5      1 210.938
## 2 Abur_2014      1      1      0 -5      1 234.375
## 3 Abur_2014      1      1      0 -5      1 257.812
## 4 Abur_2014      1      1      0 -5      1 281.250
## 5 Abur_2014      1      1      0 -5      1 304.688
## 6 Abur_2014      1      1      0 -5      1 328.125
## 7 Abur_2014      1      1      0 -5      1 351.562
## 8 Abur_2014      1      1      0 -5      1 375.000
## 9 Abur_2014      1      1      0 -5      1 398.438
## 10 Abur_2014     1      1      0 -5      1 421.875
## Absorbance      Zmag      Zang
## 1 0.0451375 110638000 -0.228113
## 2 0.0441247 100482000 -0.230561
## 3 0.0495935 90561100 -0.230213
## 4 0.0516088 83515500 -0.230959
## 5 0.0590836 77476800 -0.229652
## 6 0.0628038 71229100 -0.230026
## 7 0.0682962 66615500 -0.229576
## 8 0.0738373 61996200 -0.229327
## 9 0.0794857 58193600 -0.228984
## 10 0.0932373 54845900 -0.226507
```

## Accessing a database using dplyr commands

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

```
library(mosaic)
db <- src_mysql(dbname = "wai", host = "scidb.smith.edu", user = "waiuser",
               password="smith_waiDB")
Measurements <- tbl(db, "Measurements")
PI_Info <- tbl(db, "PI_Info")
Subject <- tbl(db, "Subject")
```

Let's explore the PI\_Info table.

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

```
## # Source:   lazy query [?? x 1]
## # Database: mysql 5.5.57-0ubuntu0.14.04.1 [waiuser@scidb.smith.edu:/wai]
##   total
##   <dbl>
## 1     5
```

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

```
## Identifier PI_Year PI
## 1 Rosowski_2012 2012 John J. Rosowski
## 2 Abur_2014 2014 Defne Abur; Nicholas J. Horton; Susan E. Voss
## 3 Shahnaz_2006 2006 Navid Shahnaz; Karin Bork
## 4 Voss_1994 1994 Susan E. Voss
## 5 Werner_2010 2010 Douglas Keefe
##
## 1 Eaton-Peabody Laboratory, Massachusetts Eye and Ear Infirmary, Boston; Department of Otology and L
## 2
## 3
## 4
## 5
## Email
## 1 John_Rosowski@meei.harvard.edu
## 2 svoss@smith.edu
## 3 nshahnaz@audiospeech.ubc.ca
## 4 svoss@smith.edu
## 5 Douglas.Keefe@boystown.org
## Title
## 1 Ear-Canal Reflectance, Umbo Velocity, and Tympanometry in Normal-Hearing Adults
## 2 Intrasubject Variability in Power Reflectance
## 3 Wideband Reflectance Norms for Caucasian and Chinese Young Adults
## 4 Measurement of acoustic impedance and reflectance in the human ear canal
## 5 Ear-Canal Wideband Acoustic Transfer Functions of Adults and Two- to Nine-Month-Old Infants
## Pub Date
## 1 Ear & Hearing 11/06/2015
## 2 J. Am Acad Audiol 08/24/2016
## 3 Ear & Hearing 08/24/2016
## 4 Journal of the Acoustical Soci 02/16/2017
## 5 Ear and Hearing 9/1/2017
##
## 1 http://www.ncbi.nlm.nih.gov/pub
## 2 N
## 3 http://journals.lww.com/ear-hearing/Abstract/2006/12000/Wideband_Reflectance_Norms_for_Caucasian_
## 4 http://www.ncbi.nlm.nih.gov/pub
## 5 https://www.ncbi.nlm.nih.gov/pub
##
## 1 HearID (Mimosa Acoustics); \nNormal Criteria as follows: \n(1) There was no history of significant
## 2
## 3
## 4
## 5
```

Let's explore the Subjects table.

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

```
## # Source: lazy query [?? x 1]
## # Database: mysql 5.5.57-0ubuntu0.14.04.1 [waiuser@scidb.smith.edu:/wai]
## total
```

```
## <dbl>
## 1 394
```

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

```
## # A tibble: 394 x 12
##   Identifier Sub_Number Session_Total Age Female Race Ethnicity
##   <chr>      <chr>      <int> <int> <int> <int>    <int>
## 1 Rosowski_2012      3          1  30     1     5        2
## 2 Rosowski_2012      6          1  29     0     5        2
## 3 Rosowski_2012     11          1  64     1     5        2
## 4 Rosowski_2012     12          1  42     1     5        2
## 5 Rosowski_2012     14          1  24     0     5        2
## 6 Rosowski_2012     15          1  32     1     5        2
## 7 Rosowski_2012     17          1  22     0     5        2
## 8 Rosowski_2012     18          1  33     1     5        2
## 9 Rosowski_2012     21          1  36     1     3        2
## 10 Rosowski_2012    22          1  33     0     5        2
## # ... with 384 more rows, and 5 more variables: Left_Ear_Status <int>,
## #   Right_Ear_Status <int>, Left_Ear_Area <dbl>, Right_Ear_Area <dbl>,
## #   Sub_Notes <chr>
```

Let's explore the Measurements table.

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

```
## # Source: lazy query [?? x 1]
## # Database: mysql 5.5.57-0ubuntu0.14.04.1 [waiuser@scidb.smith.edu:/wai]
##   total
##   <dbl>
## 1 128998
```

Let's download the data from a given subject

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

```
## # A tibble: 6 x 11
##   Identifier Sub_Number Session Left_Ear MEP Instrument Freq
##   <chr>      <int>    <int>    <int> <dbl>      <int>    <dbl>
## 1 Rosowski_2012      3        1        1    NA          1 210.938
## 2 Rosowski_2012      3        1        1    NA          1 234.375
## 3 Rosowski_2012      3        1        1    NA          1 257.812
## 4 Rosowski_2012      3        1        1    NA          1 281.250
## 5 Rosowski_2012      3        1        1    NA          1 304.688
## 6 Rosowski_2012      3        1        1    NA          1 328.125
## # ... with 4 more variables: Absorbance <dbl>, Zmag <dbl>, Zang <dbl>,
## #   SessionNum <fctr>
```

Finally we can plot the results

```
onesubj <- mutate(onesubj,  
  Ear=ifelse(Left_Ear==1, "Left", "Right"))  
ggplot(data = onesubj, aes(x = Freq, y = Absorbance)) + geom_point() +  
  aes(colour = Ear) + scale_x_log10() + labs(title="Absorbance by ear Rosowski subject 3")
```

