CI_feedback_coartic

Meg Cychosz

28 April 2021

```
library('dplyr')
library('tidyr')
library('stringr')
library('purrr')
library('ggplot2')
library('lme4')
library('lmerTest')
library('stargazer')
library('kableExtra')
```

```
# acoustic measures
data <- read.csv("/Users/megcychosz/Google Drive/CI_feedback/data/production_data/RealWordRep_809E64MS2
demo_data <- read.csv("/Users/megcychosz/Google Drive/CI_feedback/data/match_participants/final_matched_

# test data
tp1_test_data <- read.csv("/Users/megcychosz/Google Drive/CI_feedback/data/testing_data/tp1_test_data.cs
tp3_test_data <- read.csv("/Users/megcychosz/Google Drive/CI_feedback/data/testing_data/participantinfo_
some_CI_test_data <- read.csv("/Users/megcychosz/Google Drive/CI_feedback/data/testing_data/UW_UMNCrossS
CI_gfta_min <- read.csv("/Users/megcychosz/Google Drive/CI_feedback/data/testing_data/ParticipantInfo_CI

# LENA data
tp1_lena_data <- read.csv("/Users/megcychosz/Google Drive/CI_feedback/data/testing_data/TimePoint1_LENA_
CI_lena_data1 <- read.csv("/Users/megcychosz/Google Drive/CI_feedback/data/testing_data/CochlearV1_hourl

# merge the test datasets together
CI_test_data <- CI_gfta_min %>%
  select(ResearchID, MinPair_ProportionCorrect, GFTA_Age,
         GFTA_RawCorrect, GFTA_NumTranscribed, GFTA_AdjCorrect, GFTA_Standard) %>%
  merge(., some_CI_test_data, by="ResearchID")

# some cleaning
data2 <- data %>%
  filter(Filename!='RealWordRep_638L28MS1_ML.WAV') %>% #remove 638 (HA match) because he repeated less
  mutate(Speaker=gsub("\\..*","",Speaker), # create speaker category
         Speaker=gsub('_[^_]*$', '', Speaker),
         CV_dur=phone_t1+phone_t2)

# merge with demographic and testing (GFTA/vocab) datasets
tp3_test_data_vars <- tp3_test_data %>% select(ParticipantIDLong, LateTalker, AAE, EVT_Raw, EVT_GSV,
                                               EVT_Standard, PPVT_Raw, PPVT_GSV, PPVT_Standard, GFTA_Ra
                                               GFTA_NumTranscribed, GFTA_Standard, GFTA_AdjCorrect, GFT
                                               MinPair_ProportionCorrect)
```

```
CI_test_data_vars <- CI_test_data %>% select(Speaker, LateTalker, AAE, EVT_Raw, EVT_Standard, EVT_GSV,
                                             PPVT_GSV, PPVT_Raw, PPVT_Standard, GFTA_RawCorrect,
                                             GFTA_NumTranscribed, GFTA_AdjCorrect, GFTA_Standard, GFTA_
                                             MinPair_ProportionCorrect)

data3 <- tp1_test_data %>%
  mutate(MinPair_ProportionCorrect = MinPairT1) %>%
  select(ParticipantIDLong, EVT_Standard,
         EVT_Raw, EVT_GSV, PPVT_Standard, PPVT_Raw,
         PPVT_GSV, LateTalker, AAE,
         GFTA_Age, GFTA_Standard, GFTA_NumTranscribed,
         GFTA_RawCorrect, GFTA_AdjCorrect,
         MinPair_ProportionCorrect) %>%
  rbind(., tp3_test_data_vars) %>%
  mutate(Speaker=ParticipantIDLong) %>%
  select(-ParticipantIDLong) %>%
  rbind(., CI_test_data_vars) %>%
  merge(., demo_data, by=c('Speaker')) %>%
  merge(., data2, by=c('Speaker'))

# create a variable to note which children with CIs have which matches
data4 <- data3 %>%
  mutate(condition = if_else(Speaker=="300E57MS2" | Hearing_age < 11, "chrono_only",
                             if_else(Speaker=="307E44MS1", "hearing_age_only",
                                     ifelse(hearing_status=='NH', "NH",
                                     "hearing_age_chrono")))) # all other children with CIs are matched

# how many children with NH are late talkers?
# one child contributed at tp1 and tp3, n=4
late <- data4 %>%
  distinct(Speaker, .keep_all = T) %>%
  filter(LateTalker=='1')

# are the late talkers within normal EVT ranges?
late_vocab <- data4 %>%
  distinct(Speaker, .keep_all = T) %>%
  filter(LateTalker=='1') %>%
  select(EVT_GSV, Speaker)

# are the late talkers within normal GFTA ranges?
late_gfta <- data4 %>%
  distinct(Speaker, .keep_all = T) %>%
  filter(LateTalker=='1') %>%
  select(GFTA_Standard, Speaker)

# how many children are received the task in AAE?
# same child at tp1 and tp3, so n=1
aae <- data4 %>%
  distinct(Speaker, .keep_all = T) %>%
  filter(AAE=='1')

# recordings are excluded (or weren't collected) from 14 total children from the kids w/ CIs and hearin
# calculate hourly estimates
```

```
lena_hourly <- tp1_lena_data %>% # remember we don't have LENA data for TP3
  rbind(., CI_lena_data1) %>%
  group_by(Subject) %>%
  mutate(recording_length_hours=sum(Duration)/3600,
         AWC_hourly = (sum(AWC_Actual)/recording_length_hours),
         CVC_hourly = (sum(CVC_Actual)/recording_length_hours),
         CTC_hourly = (sum(CTC_Actual)/recording_length_hours)) %>%
  distinct(Subject, .keep_all = T) %>%
  select(AWC_hourly, CVC_hourly, CTC_hourly, Subject, recording_length_hours) %>%
  filter(Subject!='006L' | Subject!='026L' & Subject!='076L' &
           Subject!='122L' & Subject!='655L' & Subject!='657L') # remove the 6 hearing age matches that

no_lena <- data4 %>%
  filter(match=='Chrono_age_match' | tp=='3' | Speaker =='800E65MS2') %>% # get chrono matches,  the 6
  mutate(AWC_hourly=NA,
         CVC_hourly=NA,
         CTC_hourly=NA,
         recording_length_hours=NA,
         preSubject=gsub("(L).*","\\1",Speaker),
         Subject=gsub("(A).*","\\1",preSubject))

data_w_lena <- data4 %>%
  filter(match!='Chrono_age_match') %>%
  filter(tp!='3') %>%
  filter(Speaker !='800E65MS2') %>% # don't have LENA data from TP3, one CI child
  mutate(preSubject=gsub("(L).*","\\1",Speaker),
         Subject=gsub("E.*","",preSubject)) %>%
  merge(., lena_hourly, by='Subject') %>%
  rbind(., no_lena) # recombine with tp3 data

check <- data_w_lena %>% # sanity check: there should be 28 children w/ CIs, 27 chrono age matches, and
  distinct_at(., vars(Speaker, match), .keep_all = T) %>%
  group_by(match) %>%
  count()

# write out stats; this dataset will remove all of the tp3 children, chrono age matches, children who d
# final lena dataset should include 20 children w/ CIs and 17 hearing age matches
descrip_lena_stats <- data_w_lena %>%
   filter(Speaker!='665L52FS4' & Speaker!='679L58MS6' & Speaker!='802E72FS3' & # remove children who re
           Speaker!='806E42MS1'& Speaker!='809E64MS2') %>%
  distinct_at(., vars(Speaker, match), .keep_all = TRUE) %>%
  filter(match!='Chrono_age_match') %>%
  filter(recording_length_hours != 'NA') %>%
  group_by(match) %>%
  summarize(avg_recording_length=round(mean(recording_length_hours),2),
            sd_recording_length=round(sd(recording_length_hours),2),
            range_recording_length=paste(round(min(recording_length_hours),2),"-",max(recording_length_
```

```
knitr::kable(descrip_lena_stats, caption = 'LENA recording length stats',
             booktabs=T,
             col.names = c("match", "mean", "sd", "range")) %>%
  kable_styling()
```

Table 1: (#tab:descrip lena stats table)LENA recording length stats

match	mean	sd	range
CI	15.81	0.42	14.42 - 16
HA_match	15.49	0.76	13.66 - 16

```
# what were the ages of the children with NH (HA matches) who completed recordings?
descrip_nh_lena_stats <- data_w_lena %>%
   filter(Speaker!='665L52FS4' & Speaker!='679L58MS6' & Speaker!='802E72FS3' & # remove children who re
            Speaker!='806E42MS1'& Speaker!='809E64MS2') %>%
  distinct_at(., vars(Speaker, match), .keep_all = TRUE) %>%
  filter(match=='HA_match') %>% # only ha matches
  filter(recording_length_hours != 'NA') %>% # remove ha matches we don't have recordings from; sanity
  summarize(mean_age = mean(Chrono_age),
            sd_age = sd(Chrono_age))
```

```
# write out data with demographics for vowel analysis
write.csv(data_w_lena, '/Users/megcychosz/Google Drive/CI_feedback/data/match_participants/coartic_w_le
```

```
data6 <- data_w_lena %>%
  mutate(manner = recode(Word,"sad"="fricative","chair"="affricate","cheese"="affricate",
                         "chicken"="affricate","rabbit"="approximant","rain"="approximant",
                         "rainbow"="approximant","raisins"="approximant","reading"="approximant",
                         "red"="approximant","rock"="approximant","rocking"="approximant",
                         "running"="approximant","sad"="fricative","sandbox"="fricative",
                         "sandwich"="fricative","scissors"="fricative","share"="fricative",
                         "sharing"="fricative","sheep"="fricative","shell"="fricative",
                         "ship"="fricative","shoe"="fricative","shoes"="fricative",
                         "shorts"="fricative","shoulder"="fricative","shovel"="fricative",
                               "shower"="fricative","sick"="fricative","sidewalk"="fricative",
                               "sink"="fricative","sister"="fricative","soap"="fricative",
                               "sock"="fricative","soup"="fricative","suitcase"="fricative",
                               "sun"="fricative","sunny"="fricative","walk"="approximant",
                               "washer"="approximant","watch"="approximant","water"="approximant",
                          "web"="approximant","wet"="approximant","wheel"="approximant",
                               "wind"="approximant","window"="approximant", "waiting"="approximant"),
         poa = recode(Word, "sad"="alveolar","chair"="alveopalatal","cheese"="alveopalatal",
                         "chicken"="alveopalatal","rabbit"="rhotic","rain"="rhotic",
                         "rainbow"="rhotic","raisins"="rhotic","reading"="rhotic",
                         "red"="rhotic","rock"="rhotic","rocking"="rhotic",
                         "running"="rhotic","sad"="alveolar","sandbox"="alveolar",
                         "sandwich"="alveolar","scissors"="alveolar","share"="postalveolar",
                         "sharing"="postalveolar","sheep"="postalveolar","shell"="postalveolar",
                         "ship"="postalveolar","shoe"="postalveolar","shoes"="postalveolar",
                         "shorts"="postalveolar","shoulder"="postalveolar","shovel"="postalveolar",
                               "shower"="postalveolar","sick"="alveolar","sidewalk"="alveolar",
                               "sink"="alveolar","sister"="alveolar","soap"="alveolar",
                               "sock"="alveolar","soup"="alveolar","suitcase"="alveolar",
                               "sun"="alveolar","sunny"="alveolar","walk"="labiovelar",
                               "washer"="labiovelar","watch"="labiovelar","water"="labiovelar",
                          "web"="labiovelar","wet"="labiovelar","wheel"="labiovelar",
```

```
                              "wind"="labiovelar","window"="labiovelar", "waiting"="labiovelar"),
       backness = recode(Word, "sad"="front","chair"="rhotic","cheese"="front",
                         "chicken"="front","rabbit"="front","rain"="front",
                         "rainbow"="front","raisins"="front","reading"="front",
                         "red"="front","rock"="back","rocking"="back",
                         "running"="back","sad"="front","sandbox"="front",
                         "sandwich"="front","scissors"="front","share"="front",
                         "sharing"="front","sheep"="front","shell"="front",
                         "ship"="front","shoe"="back","shoes"="back",
                         "shorts"="back","shoulder"="back","shovel"="back",
                               "shower"="diphthong","sick"="front","sidewalk"="diphthong",
                               "sink"="front","sister"="front","soap"="back",
                               "sock"="back","soup"="back","suitcase"="back",
                               "sun"="back","sunny"="back","walk"="back",
                               "washer"="back","watch"="back","water"="back",
                          "web"="front","wet"="front","wheel"="front",
                               "wind"="front","window"="front", "waiting"="front"))
```

1 demographic details for methods

```
# num of unique children in each condition
child_cts <- demo_data %>%
  mutate(preSubject=gsub("(L).*","\\1",Speaker)) %>%
  distinct(preSubject, .keep_all = T) %>%
  group_by(hearing_status) %>%
  count()
print(paste('# of unique children:', sum(child_cts$n))) # two children contributed data from the same t
```

```
## [1] "# of unique children: 72"
```

```
CI_child <- child_cts %>%
  filter(hearing_status=='CI')
print(paste('# of children with CIs:', CI_child$n))
```

```
## [1] "# of children with CIs: 28"
```

```
NH_child <- child_cts %>%
  filter(hearing_status!='CI')
print(paste('# of *unique* children with NH:', sum(NH_child$n)))
```

```
## [1] "# of *unique* children with NH: 44"
```

```
# gender distribution of unique children with NH
NH_gender <- demo_data %>%
  filter(hearing_status!='CI') %>%
  mutate(preSubject=gsub("(L).*","\\1",Speaker)) %>%
  distinct(preSubject, .keep_all = T) %>%
  group_by(Gender) %>%
  count()
```

```r
# device formation
CI_device <- demo_data %>%
  filter(match=='CI') %>%
  group_by(device_formation) %>%
  count()

CI_implant_age <- demo_data %>%
  filter(match=='CI') %>%
  summarize(avg_implant_age = mean(age_at_activation),
            sd_implant_age = sd(age_at_activation),
            min_implant_age = min(age_at_activation),
            max_implant_age = max(age_at_activation))
print(paste('avg age at implantation:', round(CI_implant_age$avg_implant_age,2)))
```

```
## [1] "avg age at implantation: 18.96"
```

```r
print(paste('sd:', round(CI_implant_age$sd_implant_age,2)))
```

```
## [1] "sd: 10.68"
```

```r
print(paste('range:', round(CI_implant_age$min_implant_age,2),'-',round(CI_implant_age$max_implant_age,
```

```
## [1] "range: 6 - 45 months"
```

```r
# gender distribution
CI_gender <- demo_data %>%
  filter(match=='CI') %>%
  group_by(Gender) %>%
  count()


# including the children grouped under *both* HA and chrono age matches; these are the actual #s of chi
dup_child_cts <- demo_data %>%
  group_by(match) %>%
  count()

chrono_child <- dup_child_cts %>%
  filter(match=='Chrono_age_match')
print(paste('# of chronological age matches, including duplicate children:', chrono_child$n))
```

```
## [1] "# of chronological age matches, including duplicate children: 27"
```

```r
ha_child <- dup_child_cts %>%
    filter(match=='HA_match')
print(paste('# of hearing age matches, including duplicate children:', ha_child$n))
```

```
## [1] "# of hearing age matches, including duplicate children: 25"
```

```r
# and their genders
chrono_child_gender <- demo_data %>%
  filter(match=='Chrono_age_match') %>%
  group_by(Gender) %>%
```

```
  count()

ha_child_gender <- demo_data %>%
  filter(match=='HA_match') %>%
  group_by(Gender) %>%
  count()
```

```
# age & SES stats
# do this separately for the two conditions because some kids were excluded
# from one condition (hearing age too young) but not the other condition
ha_stats <- demo_data %>%
  filter(Hearing_age > 11 & Speaker != '300E57MS2') %>% # remove the 3 children who weren't matched
  group_by(match) %>%
  filter(match != 'Chrono_age_match') %>%
  summarize(mean_ha = mean(Hearing_age),
            sd_ha = sd(Hearing_age),
            range_ha = paste(min(Hearing_age), ',', max(Hearing_age)),
            mean_mat_ed = mean(Mat_ed),
            sd_mat_ed = round(sd(Mat_ed),2))

chrono_stats <- demo_data %>%
  filter(Speaker != '307E44MS1') %>%# remove the child that wasn't matched
  group_by(match) %>%
  filter(match != 'HA_match') %>%
  summarize(mean_chrono = mean(Chrono_age),
            sd_chrono = sd(Chrono_age),
            range_chrono = paste(min(Chrono_age), ',', max(Chrono_age)),
            mean_mat_ed = mean(Mat_ed),
            sd_mat_ed = round(sd(Mat_ed),2))
```

2 write out data for vowel analysis

```
# TODO : write out demographic data to analyze vowels separately
```

3 data pre-processing coarticulation

```
# select only word-initial CV sequences
data7 <- data6 %>%
  filter(Word !='') %>%# remove empty annotations
  mutate(var = paste(Word,Previous)) %>%
  filter(var != 'sister IH1' & var!='shorts T'
         & var!='shorts OW1' & var!='shower AW1'
         & var!='suitcase K' & var!='suitcase EY1'
         & var!='sidewalk D' & var!='washer AH1'
         & var!='sandwich D' & var!='sandwich IH2'
         & var!='sharing EY1' & var!='chair AY1'
         & var!='sandbox K' & var!='share AY1'
         & var!='watch AH1' & var!='suitcase EY2'
         & var!='chair EY1')
```

```r
# for coartic analysis, select only the words that were repeated at all timepoints
subdata <- data7 %>%
  filter(Word=='sandwich' | Word=='scissors' | Word=='share' |
           Word=='sharing' | Word=='sheep' | Word=='shoe' | Word=='shoes' |
           Word=='shorts' | Word=='shovel' | Word=='shower' | Word=='sick' |
           Word=='sink' | Word=='soup' |
         Word=='sun')
```

```r
subdata2 <- subdata %>%
  filter(Analysis!='DONTUSE') %>% # remove words unable to segment
  filter(Repetition=='1' |
           (Repetition==2 & Analysis=='USE') |
           (Repetition==3 & Analysis=='USE')) %>% # only use second/third repetitions when marked to do
  filter(Analysis!='VOWELONLY' & Analysis!='VowelOnly') # remove words marked to only do vowel analysis

# get some summary stats on words removed; only reflects words removed from both coartic and vowel anal
rmvd_stats <- subdata %>%
  group_by(match) %>%
  mutate(total_words = n()) %>%
  filter(Analysis!='DONTUSE') %>% # remove words unable to segment
  filter(Repetition=='1' |
           (Repetition==2 & Analysis=='USE') |
           (Repetition==3 & Analysis=='USE')) %>% # only use second/third repetitions when marked to do
  group_by(match) %>%
  summarize(percen_used = (n()/total_words)*100) %>%
  distinct()
```

```r
# convert structure of spectral measurements at edges to something computable
# remove brackets
subdata2$Spectrum <- gsub( ']', '', subdata2$Spectrum)
subdata2$Spectrum <- gsub( '[ ', '', subdata2$Spectrum, fixed = TRUE) # open bracket denotes regex so f

# convert measurements to string
subdata2$variable_sep <- str_extract_all(subdata2$Spectrum, "[-0-9\\.]+")

# for euclidean distance and raw distance, convert to numeric:
subdata2$spec_vector <- lapply(subdata2$variable_sep , FUN = as.numeric)
subdata2 <- as.data.frame(subdata2)


# --------- option to find raw difference/euclidean between vectors --------
if(any(grepl("package:plyr", search()))) detach("package:plyr") else message("plyr not loaded")
library('dplyr')


# calculate raw difference (sanity check) and euc distance between vectors
diff_df <- subdata2 %>%
  group_by(Speaker, match, Word, Word_duration) %>% # IMPORTANT: always group by Speaker + match b/c tw
  #mutate(raw_diff = map2(spec_vector, lead(spec_vector), `-`)) %>% # sanity check (note to take absolu
  mutate(euc_dist = map2(spec_vector, lead(spec_vector), function(x, y) sqrt(sum((x-y) ^ 2)))) %>%
  as.data.frame()

# remove NA rows where measurement was made upon but not stored
```

```
df.final <- subset(diff_df, euc_dist != '0')
df.final$euc_dist <- as.numeric(df.final$euc_dist)

# sanity check - no word*speaker*Word_duration should have more than 1 row
df_ct <- df.final %>%
  group_by(Speaker, match, Word, Word_duration) %>%
  count()
```

```
set.seed(123)
# some children still have 3+ repetitions of individual words
# in those case, randomly select 2 utterances
df.final2 <- df.final %>%
  #distinct_at(vars(Speaker, match, Word, Word_duration), .keep_all = T) %>% # get each speaker's indiv
  add_count(Speaker, match, Word)

tworeps <- df.final2 %>% filter(n=='1' | n=='2')

df.final3 <- df.final2 %>%
  filter(n=='4' | n=='3') %>%
  group_by(Speaker, match, Word) %>%
  sample_n(2) %>% # sample two elicitations
  rbind(., tworeps) # put back together

# what % of words by hearing condition were *first* repetitions?
reps <- df.final3 %>%
  group_by(match) %>%
  mutate(total_reps = n()) %>%
  group_by(match, Repetition) %>%
  summarize(percen_rep = (n()/total_reps)*100) %>%
  distinct()
```

4 coarticulation analysis

4.1 coartic by hearing status

```
coartic_tbl <- df.final3 %>%
  ungroup() %>%
  select(poa, match, euc_dist) %>%
  group_by(match, poa) %>%
  summarize(mean_coartic = round(mean(euc_dist),2),
            sd_coartic = round(sd(euc_dist),2),
            range_coartic = paste(round(min(euc_dist),2),"-",round(max(euc_dist),2))) %>%
  mutate(stats=paste(mean_coartic,"(",sd_coartic,")",range_coartic)) %>%
  select(-mean_coartic, -sd_coartic, -range_coartic) %>%
  spread("poa", "stats") %>%
  mutate(match=recode(match, "CI"="CIs", "HA_match"="Hearing age matches", "Chrono_age_match"="Chronolo
```

```
knitr::kable(coartic_tbl, caption = 'Mean spectral distance between C-V by hearing status and consonant
             booktabs=T,
             col.names = c("Hearing status", "[s-V]", "[SH-V]")) %>%
  kable_styling() %>%
  add_header_above(c(" " = 1, "Spectral distance" = 2))
```

Table 2: Mean spectral distance between C-V by hearing status and consonant

	Spectral distance	
Hearing status	[s-V]	[SH-V]
Chronological age matches	11.95 (4.26) 2.15 - 30.24	11.9 (3.46) 3.11 - 31.81
CIs	10.93 (2.95) 5.46 - 29.15	10.95 (3.4) 2.64 - 30.1
Hearing age matches	11.92 (3.98) 6.32 - 29.27	11.48 (3.8) 5.6 - 28.77

We first computed coarticulation, or the Mel spectral distance between phones, by hearing status, within each target CV sequence, where a larger distance indicates less coarticulatory overlap. Descriptive statistics in Table 2 show differences in coarticulation by hearing status: children with CIs coarticulate more within [s-V] and [S-V] sequences than both groups of children with NH. A linear mixed effects model was fit to predict the Mel spectral distance between phones in each target CV sequence (model summary in Table **??**). The baseline model included random effects for **Word** and **Speaker**. **Word duration** was additionally added to control for the effect of speaking rate on coarticulation and **Child Chronological Age** was added to control for age-related changes in coarticulation unrelated to the other variables of interest. The effect of **Hearing Status** improved upon this model fit: children with CIs coarticulated significantly more within CV sequences than their chronological age-matched peers. They tended to coarticulate more than their hearing age matches as well (Figure **??**), but this effect was not significant under an alpha value of .05, suggesting that children with CIs pattern coarticulatorily closer to their hearing age matches. Neither **Place of Articulation** ([s] versus [S]) nor its interaction with **Hearing Status** improved upon model fit.

```
center_scale <- function(x) {
  scale(x, scale = FALSE)
}
coartic_model_data <- df.final3 %>%
  mutate(Word_duration = center_scale(Word_duration))
coartic_model_data$Chrono_age_centered <- coartic_model_data$Chrono_age

coartic_model_data$match <- factor(coartic_model_data$match, ordered = FALSE )
coartic_model_data$match <- relevel(coartic_model_data$match, ref = "CI")
baseline <- coartic_model_data %>%
  lmer(euc_dist ~ +
         (1|Speaker) +
         (1|Word),
       data=.)

m1 <- coartic_model_data %>%
  lmer(euc_dist ~ Word_duration +
         (1|Speaker) +
         (1|Word),
       data=.)
anova(baseline, m1) # improves
```

```
## Data: .
## Models:
## baseline: euc_dist ~ +(1 | Speaker) + (1 | Word)
## m1: euc_dist ~ Word_duration + (1 | Speaker) + (1 | Word)
##          npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## baseline    4 6177.8 6198.1 -3084.9   6169.8
## m1          5 6176.3 6201.6 -3083.2   6166.3 3.5059  1    0.06115 .
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
m1a <- coartic_model_data %>%
  lmer(euc_dist ~ Chrono_age_centered +
         Word_duration +
         (1|Speaker) +
         (1|Word),
       data=.)
anova(m1, m1a) # doesn't improve but we keep to control for age-related changes
```

```
## Data: .
## Models:
## m1: euc_dist ~ Word_duration + (1 | Speaker) + (1 | Word)
## m1a: euc_dist ~ Chrono_age_centered + Word_duration + (1 | Speaker) +
## m1a:     (1 | Word)
##     npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## m1     5 6176.3 6201.6 -3083.2   6166.3
## m1a    6 6178.0 6208.4 -3083.0   6166.0 0.2871  1     0.5921
```

```
coartic_model_data$match <- factor(coartic_model_data$match, ordered = FALSE )
coartic_model_data$match <- relevel(coartic_model_data$match, ref = "CI")
m2 <- coartic_model_data %>%
  lmer(euc_dist ~ match +
         Chrono_age_centered +
         Word_duration +
         (1|Speaker) +
         (1|Word),
       data=.)
anova(m1a, m2) # improves
```

```
## Data: .
## Models:
## m1a: euc_dist ~ Chrono_age_centered + Word_duration + (1 | Speaker) +
## m1a:     (1 | Word)
## m2: euc_dist ~ match + Chrono_age_centered + Word_duration + (1 |
## m2:     Speaker) + (1 | Word)
##     npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## m1a    6 6178.0 6208.4 -3083.0   6166.0
## m2     8 6176.9 6217.4 -3080.4   6160.9 5.1229  2    0.07719 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
m2a <- coartic_model_data %>%
  lmer(euc_dist ~ match*Chrono_age_centered +
         Word_duration +
         (1|Speaker) +
         (1|Word),
       data=.)
anova(m2, m2a) # doesn't improve
```

```
## Data: .
## Models:
## m2: euc_dist ~ match + Chrono_age_centered + Word_duration + (1 |
```

```
## m2:     Speaker) + (1 | Word)
## m2a: euc_dist ~ match * Chrono_age_centered + Word_duration + (1 |
## m2a:     Speaker) + (1 | Word)
##     npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## m2     8 6176.9 6217.4 -3080.4   6160.9
## m2a   10 6179.2 6229.8 -3079.6   6159.2 1.6616  2     0.4357
```

```
m3 <- coartic_model_data %>%
  lmer(euc_dist ~ poa +
         match +
         Chrono_age_centered +
         Word_duration +
         (1|Speaker) +
         (1|Word),
       data=.)
anova(m2, m3) # doesn't improve; no effect of s~SH
```

```
## Data: .
## Models:
## m2: euc_dist ~ match + Chrono_age_centered + Word_duration + (1 |
## m2:     Speaker) + (1 | Word)
## m3: euc_dist ~ poa + match + Chrono_age_centered + Word_duration +
## m3:     (1 | Speaker) + (1 | Word)
##    npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## m2    8 6176.9 6217.4 -3080.4   6160.9
## m3    9 6178.9 6224.4 -3080.4   6160.9 0.0195  1      0.889
```

```
m4 <- coartic_model_data %>%
  lmer(euc_dist ~ poa*match +
         Chrono_age_centered +
         Word_duration +
         (1|Speaker) +
         (1|Word),
       data=.)
anova(m2, m4) # doesn't improve; no effect of s~SH dependent on hearing status
```

```
## Data: .
## Models:
## m2: euc_dist ~ match + Chrono_age_centered + Word_duration + (1 |
## m2:     Speaker) + (1 | Word)
## m4: euc_dist ~ poa * match + Chrono_age_centered + Word_duration +
## m4:     (1 | Speaker) + (1 | Word)
##    npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## m2    8 6176.9 6217.4 -3080.4   6160.9
## m4   11 6181.9 6237.5 -3079.9   6159.9 1.0386  3     0.7919
```

```
# final model; make a table
coartic_model_data %>%
  lme4::lmer(euc_dist ~ match + Chrono_age_centered + Word_duration + (1|Speaker) + (1|Word), data=.) %
  stargazer(., header=FALSE,
            dep.var.caption = "",
            dep.var.labels.include = FALSE,
            type = "latex",
```

```
            star.cutoffs=c(0.05,0.01,0.001),
            star.char = c("*", "**", "***"),
            title="Model predicting Mel spectral distance by hearing status",
            digits = 2,
            ci = TRUE,
            style = "all",
            order=c(5,2,1,3,4),
            covariate.labels = c("Intercept", "Hearing age matches", "Chronological age matches",  "Chi
```

```
##
## \begin{table}[!htbp] \centering
##   \caption{Model predicting Mel spectral distance by hearing status}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
##  Intercept & 10.86$^{***}$ \\
##   & (8.64, 13.09) \\
##   & t = 9.57 \\
##   & p = 0.00 \\
##   Hearing age matches & 0.83 \\
##   & ($-$0.18, 1.85) \\
##   & t = 1.61 \\
##   & p = 0.11 \\
##   Chronological age matches & 0.90$^{*}$ \\
##   & (0.09, 1.72) \\
##   & t = 2.18 \\
##   & p = 0.03 \\
##   Child chronological age & 0.003 \\
##   & ($-$0.03, 0.04) \\
##   & t = 0.15 \\
##   & p = 0.89 \\
##   Word duration & 1.97 \\
##   & ($-$0.09, 4.03) \\
##   & t = 1.87 \\
##   & p = 0.07 \\
##  \hline \\[-1.8ex]
## Observations & 1,163 \\
## Log Likelihood & $-$3,082.78 \\
## Akaike Inf. Crit. & 6,181.55 \\
## Bayesian Inf. Crit. & 6,222.02 \\
## \hline
## \hline \\[-1.8ex]
## \textit{Note:}  & \multicolumn{1}{r}{$^{*}$p$<$0.05; $^{**}$p$<$0.01; $^{***}$p$<$0.001} \\
## \end{tabular}
## \end{table}
```

4.2 Coarticulation by hearing age in children with CIs

```
# children with CIs where we expect an effect of hearing age, but not chrono age
ci_model_data <- df.final3 %>%
  filter(hearing_status=='CI') %>%
  mutate(Word_duration = center_scale(Word_duration))
```

```r
ci_model_data$Hearing_age_centered = center_scale(ci_model_data$Hearing_age)
ci_model_data$Chrono_age_centered = center_scale(ci_model_data$Chrono_age)

age_m <- ci_model_data %>%
  lmer(euc_dist ~ Word_duration +
         (1|Speaker) +
         (1|Word),
       data=.)

age_mI <- ci_model_data %>%
  lmer(euc_dist ~ Hearing_age_centered +
         Word_duration +
         (1|Speaker) +
         (1|Word),
       data=.)
anova(age_m, age_mI) # significantly improves
```

```
## Data: .
## Models:
## age_m: euc_dist ~ Word_duration + (1 | Speaker) + (1 | Word)
## age_mI: euc_dist ~ Hearing_age_centered + Word_duration + (1 | Speaker) +
## age_mI:     (1 | Word)
##        npar    AIC    BIC  logLik deviance Chisq Df Pr(>Chisq)
## age_m     5 1911.9 1931.5 -950.95   1901.9
## age_mI    6 1909.8 1933.3 -948.90   1897.8 4.093  1    0.04306 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

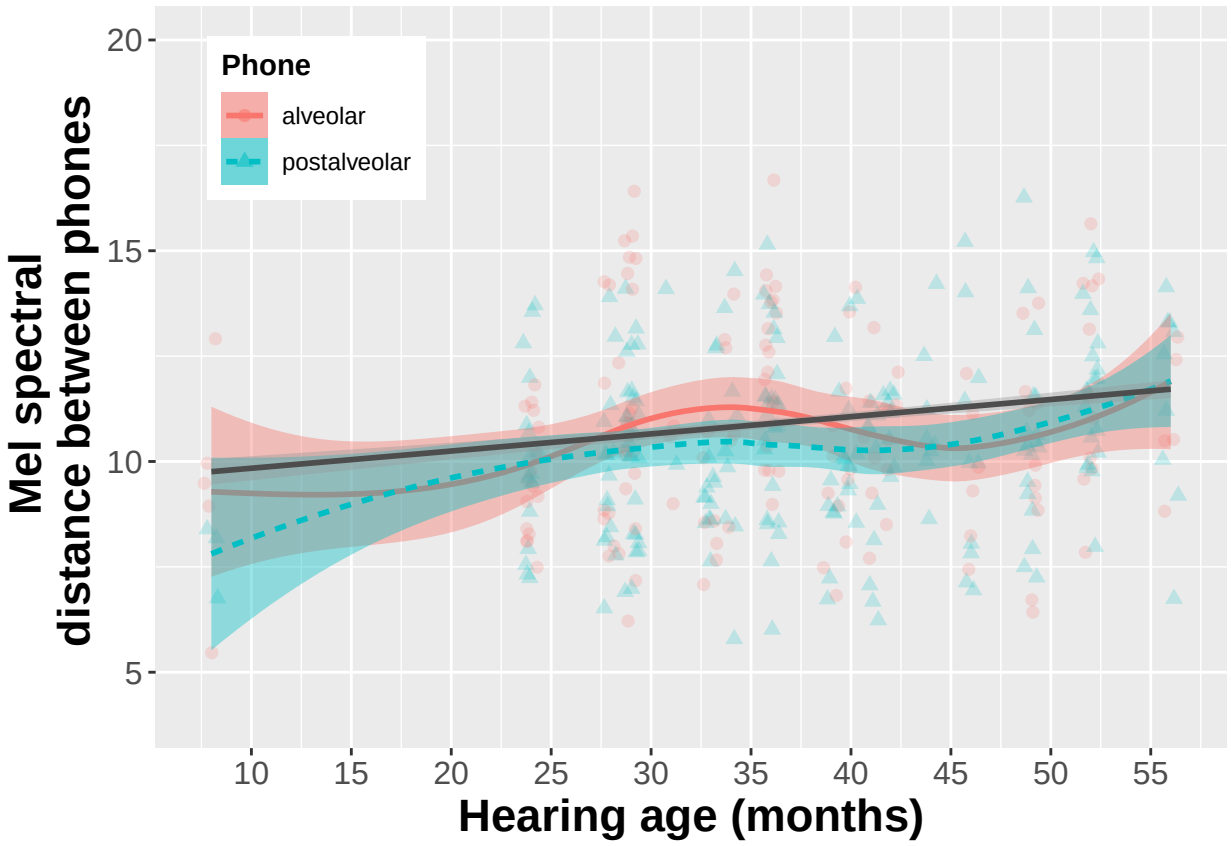
```r
hearing_age_model <- summary(age_mI)

age_mII <- ci_model_data %>%
  lmer(euc_dist ~ Chrono_age_centered +
         Word_duration +
         (1|Speaker) +
         (1|Word),
       data=.)
chrono_log_test <- anova(age_m, age_mII) # no effect
pvalue <- chrono_log_test$`Pr(>Chisq)`[2]

age_mIII <- ci_model_data %>%
  lmer(euc_dist ~ Hearing_age_centered +
         Word_duration +
         poa +
         (1|Speaker) +
         (1|Word),
       data=.)
anova(age_mI, age_mIII) # no effect of poa
```

```
## Data: .
## Models:
## age_mI: euc_dist ~ Hearing_age_centered + Word_duration + (1 | Speaker) +
## age_mI:     (1 | Word)
## age_mIII: euc_dist ~ Hearing_age_centered + Word_duration + poa + (1 |
```

```
## age_mIII:     Speaker) + (1 | Word)
##          npar    AIC    BIC logLik deviance  Chisq Df Pr(>Chisq)
## age_mI      6 1909.8 1933.3 -948.9   1897.8
## age_mIII    7 1911.8 1939.2 -948.9   1897.8 0.0012  1     0.9727
```

We next evaluated the role of hearing versus chronological age upon the children with CIs' coarticulation. A linear mixed effects model, with random effects of **Word** and **Speaker** and a fixed effect of **Word duration**, was fit to predict the degree of coarticulation for the children with CIs. There was a significant effect of **Hearing Age** (β=0.04, t=2.03, p=0.05), but not **Chronological Age** (model comparison with and without **Chronological age**: χ^2=0.88, df=1, p=0.35), indicating that it was the children with CIs' increased *hearing* experience, and not other chronological age-related maturity such as physiological development or domain general fine motor control, that best predicted the degree of their coarticulation (Figure **??**. There was again no significant effect of **Place of Articulation**.

```
# fit the model
ci_data <- df.final3 %>%
  filter(hearing_status=='CI')

hearing_age_m <- ci_data %>%
  lme4::lmer(euc_dist ~ Hearing_age +
         Word_duration +
         (1|Speaker) +
         (1|Word),
       data=.)
ci_data$fit <- predict(hearing_age_m)

ci_data %>% # TODO: note in the methods that the 7mo hearing age CI child (803E41FS1) was removed becau
  mutate(Phone=poa) %>%
  ggplot(., aes(x=Hearing_age, y=euc_dist),color=Phone, shape=Phone) +
  ylim(4,20) +
  geom_jitter(size=2, alpha=.2, aes(color=Phone, shape=Phone, fill=Phone)) +
  geom_smooth(aes(fill=Phone, color=Phone, lty=Phone)) +
  geom_smooth(method="lm", aes(y=fit), size=1,color="grey30") + # the actual model fit
  xlab("Hearing age (months)") +
  ylab("Mel spectral \n distance between phones") +
  scale_x_continuous(breaks=seq(5,60,5)) +
  theme(axis.text=element_text(size=12),
      axis.title=element_text(size=17,face="bold"),
      legend.title = element_text(size=11, face="bold"),
      legend.position = c(.15,.85),
       legend.background = element_rect(fill="white",
                                  size=0.5, linetype="solid")) +
  guides(colour = guide_legend(override.aes = list(alpha = .55)))
```



4.3 Coarticulation by vocabulary size and daily speech practice

To further evaluate the role of hearing status on coarticulation, we examined the interaction of hearing status with two known predictors of child coarticulation—vocabulary size (EVT-2) and hourly child vocalization count (average number of times per hour that the child vocalized in their at-home recording)—and two novel measures that we predicted might predict coarticulation: articulation skill (GFTA-2) and minimal pair discrimination ability. Descriptive statistics of these predictors by hearing status are listed in Table **??**. For the vocabulary and articulation scores, we report growth scale values (vocabulary) and standard scores (articulation) which are transformations of raw scores that grow linearly with age.

```
vocab_gsv_stats <- df.final3 %>% # 76 *unique* EVT scores, but 78 total as two children are included in
  ungroup() %>%
  distinct_at(., vars(Speaker,match), .keep_all = T) %>%
  mutate(EVT_GSV=as.numeric(as.character(EVT_GSV))) %>%
  select(match, EVT_GSV) %>%
  group_by(match) %>%
  summarize(EVT_GSV_stats=paste(round(mean(EVT_GSV, na.rm=T),2),
                                "(",
                                round(sd(EVT_GSV, na.rm=T),2),
                                ")",
                                round(min(EVT_GSV, na.rm=T),2),
                                "-",
                                round(max(EVT_GSV, na.rm=T),2)))

vocab_standard_stats <- df.final3 %>% # 76 *unique* EVT scores, but 78 total as two children are includ
  ungroup() %>%
  distinct_at(., vars(Speaker,match), .keep_all = T) %>%
```

```
  mutate(EVT_Standard=as.numeric(as.character(EVT_Standard))) %>%
  select(match, EVT_Standard) %>%
  group_by(match) %>%
  summarize(EVT_Standard_stats=paste(round(mean(EVT_Standard, na.rm=T),2),
                                     "(",
                                     round(sd(EVT_Standard, na.rm=T),2),
                                     ")",
                                     round(min(EVT_Standard, na.rm=T),2),
                                     "-",
                                     round(max(EVT_Standard, na.rm=T),2)))

gfta_stats <- df.final3 %>% # 72 total gfta scores
  ungroup() %>%
  distinct_at(., vars(Speaker,match), .keep_all = T) %>%
  mutate(GFTA_Standard=as.numeric(GFTA_Standard)) %>%
  select(match, GFTA_Standard) %>%
  group_by(match) %>%
  summarize(GFTA_stats=paste(round(mean(GFTA_Standard, na.rm=T),2),
                                     "(",
                                     round(sd(GFTA_Standard,na.rm=T),2),
                                     ")",
                                     round(min(GFTA_Standard,na.rm = T),2),
                                     "-",
                                     round(max(GFTA_Standard,na.rm = T),2)))


minpair_stats <- df.final3 %>% #
  ungroup() %>%
  filter(match!='Chrono_age_match') %>%
  distinct_at(., vars(Speaker,match), .keep_all = T) %>%
  select(match, MinPair_ProportionCorrect) %>%
  group_by(match) %>%
  summarize(minpair_stats=paste(round(mean(MinPair_ProportionCorrect, na.rm=T),2),
                                     "(",
                                     round(sd(MinPair_ProportionCorrect,na.rm=T),2),
                                     ")",
                                     round(min(MinPair_ProportionCorrect,na.rm = T),2),
                                     "-",
                                     round(max(MinPair_ProportionCorrect,na.rm = T),2))) %>%
  rbind(., c("Chrono_age_match", "NA")) #  add an NA row for chrono kids who didn't complete this


# get LENA data for HA matches and CI kids (don't have for chrono matches)
new_lena_stats <- data_w_lena %>%
  ungroup() %>%
  filter(Speaker!='665L52FS4' & Speaker!='679L58MS6' & Speaker!='802E72FS3' & # remove children who rec
           Speaker!='806E42MS1'& Speaker!='809E64MS2') %>%
  filter(match!='Chrono_age_match' & CVC_hourly!='NA') %>% # only HA and CIs
  distinct_at(., vars(Speaker,match), .keep_all = T) %>%
  select(match, CVC_hourly) %>%
  group_by(match) %>%
  summarize(CVC_stats=paste(round(mean(CVC_hourly),2),
                            "(",
```

```
                            round(sd(CVC_hourly),2),
                            ")",
                            round(min(CVC_hourly),2),
                            "-",
                            round(max(CVC_hourly),2))) %>%
  rbind(., c("Chrono_age_match", "NA")) #  add an NA row for chrono kids



# merge tables
predic_tbl <- new_lena_stats %>%
  merge(., vocab_gsv_stats, by="match") %>%
  merge(., vocab_standard_stats, by="match") %>%
  merge(., gfta_stats, by="match") %>%
  merge(., minpair_stats, by="match") %>%
  mutate(match=recode(match, "CI"="CIs", "HA_match"="Hearing age matches", "Chrono_age_match"="Chronolo
```

```
# write out
knitr::kable(predic_tbl, caption = 'Task statistics by hearing status',
             booktabs=T,
             col.names = c("Hearing status", "mean (SD) range", "mean (SD) range",  "mean (SD) range",
   column_spec(., column = 1:6, width = ".4in") %>%
  kable_styling(full_width=F,
                latex_options = "hold_position") %>%
  add_header_above(c(" " = 1,
                     "Hourly voc. count" = 1,
                     "EVT-2 GSVs" = 1,
                     "EVT-2 Standard Score" = 1,
                     "GFTA-2 Standard Score" = 1,
                     "Discrim. prop. correct" = 1))
```

```
# vocab
vocab_model_data <- df.final3 %>%
  filter(EVT_GSV!='NA') %>% # participants 310 (CI) and 390A (chrono) didn't complete the EVT
  mutate(Word_duration = center_scale(Word_duration))
vocab_model_data$EVT_GSV_centered <- center_scale(vocab_model_data$EVT_GSV)
vocab_model_data$Chrono_age_centered <- center_scale(vocab_model_data$Chrono_age) # TODO: make sure thi

# start where we left off above
vocabI <- vocab_model_data %>%
  lmer(euc_dist ~ match +
         Chrono_age_centered +
         Word_duration +
         (1|Speaker) +
         (1|Word),
       data=.)

vocabII <- vocab_model_data %>%
  lmer(euc_dist ~ EVT_GSV_centered +
         match +
         Chrono_age_centered +
         Word_duration +
         (1|Speaker) +
```

Table 3: Task statistics by hearing status

	Hourly voc. count	EVT-2 GSVs	EVT-2 Standard Score	GFTA-2 Standard Score	Discrim. prop. correct
Hearing status	mean (SD) range	mean (SD) range	mean (SD) range	mean (SD) range	mean (SD) range
Chronological age matches	NA	146.31 (11.94) 126 - 165	122.81 (16.42) 90 - 151	95.29 (13.59) 69 - 113	NA
CIs	248.7 (89.15) 50.12 - 387.33	117 (25.69) 42 - 148	94.69 (20.48) 43 - 126	70.78 (18.47) 39 - 104	0.69 (0.15) 0.38 - 0.97
Hearing age matches	213.86 (95.13) 31.25 - 376.25	126.21 (18.99) 85 - 160	119.42 (18.98) 84 - 160	88.57 (12.96) 73 - 116	0.69 (0.15) 0.43 - 0.93

```
         (1|Word),
       data=.)
anova(vocabI, vocabII) # not sig
```

```
## Data: .
## Models:
## vocabI: euc_dist ~ match + Chrono_age_centered + Word_duration + (1 |
## vocabI:     Speaker) + (1 | Word)
## vocabII: euc_dist ~ EVT_GSV_centered + match + Chrono_age_centered + Word_duration +
## vocabII:     (1 | Speaker) + (1 | Word)
##         npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## vocabI     8 6033.3 6073.6 -3008.7   6017.3
## vocabII    9 6033.4 6078.7 -3007.7   6015.4 1.9551  1      0.162
```

```
vocab_model_data$match <- factor(vocab_model_data$match, ordered = FALSE )
vocab_model_data$match <- relevel(vocab_model_data$match, ref = "CI")
vocabIII <- vocab_model_data %>%
  lmer(euc_dist ~ EVT_GSV_centered*match +
         Chrono_age_centered +
         Word_duration +
         (1|Speaker) +
         (1|Word),
       data=.)
anova(vocabII, vocabIII) #  significant
```

```
## Data: .
## Models:
## vocabII: euc_dist ~ EVT_GSV_centered + match + Chrono_age_centered + Word_duration +
## vocabII:     (1 | Speaker) + (1 | Word)
## vocabIII: euc_dist ~ EVT_GSV_centered * match + Chrono_age_centered + Word_duration +
```

```
## vocabIII:     (1 | Speaker) + (1 | Word)
##          npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## vocabII     9 6033.4 6078.7 -3007.7   6015.4
## vocabIII   11 6031.9 6087.3 -3004.9   6009.9 5.4721  2    0.06482 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
vocab_model <- summary(vocabIII)

# sig effect of vocab for CI kids, but no effect for hearing age or chrono age matches

# **** Exploratory ******
# maybe you have to get past a certain articulatory point before vocab can predict your outcomes?
# model the interaction of GFTA and vocab
#ci_kids <- vocab_model_data %>% filter(match=='CI' & GFTA_Standard!='NA')
vocab_GFTA <- vocab_model_data %>%
  filter(match!='Chrono_age_match') %>%
  ungroup() %>%
  group_by(match) %>%
  mutate(med_gfta = median(GFTA_Standard,na.rm = T)) %>% # select just the top half of CI kids and top
  filter(GFTA_Standard!='NA') %>%
  group_by(match) %>%
  filter(GFTA_Standard > med_gfta) %>%
  select(-med_gfta) %>%
  #rbind(., ci_kids) %>% # it's not fair to include all the CI kids because there's more of them and th
  lmer(euc_dist ~ EVT_GSV_centered +
       Chrono_age_centered +
         Word_duration +
         (1|Speaker) +
         (1|Word),
       data=.) # no effect of vocab on coartic when combining just the top half of CI kids and HA kids

# maybe you have to get past a certain vocabulary size before vocab can predict your outcomes? check wi
vocab_size <- vocab_model_data %>%
  filter(match!='CI') %>%
  filter(Chrono_age>=41) %>% #(the mean age in Cychosz et al. 2021)
  lmer(euc_dist ~ EVT_GSV_centered +
         Word_duration +
         (1|Speaker) +
         (1|Word),
       data=.)

# try lumping all of the NH kids together, and limit it to the range studied in Cychosz et al. 2021
vocab_size2 <- vocab_model_data %>%
  filter(match=='CI') %>%
  #filter(EVT_GSV>=100) %>% # only one kid in Cychosz et al had a score lower than this
  lmer(euc_dist ~ EVT_GSV_centered +
         Chrono_age_centered +
         Word_duration +
         (1|Speaker) +
         (1|Word),
       data=.)
```

```r
# final model, make a table
vocab_model_data %>%
  lme4::lmer(euc_dist ~ EVT_GSV_centered*match + Chrono_age_centered + Word_duration +  (1|Speaker) + (
  stargazer(., header=FALSE,
            dep.var.caption = "",
            dep.var.labels.include = FALSE,
            type = "latex",
            star.cutoffs=c(0.05,0.01,0.001),
            star.char = c("*", "**", "***"),
            title="Effect of hearing status and vocabulary on Mel spectral distance",
            digits = 2,
            ci = TRUE,
            style = "all",
            order=c(8,1,2,3,4,5,6,7),
            covariate.labels = c("Intercept", "EVT-2 score", "Chronological age matches", "Hearing age m
                                 "EVT-2 score*Chrono age match", "EVT-2 score*Hearing age match"))
```

```
##
## \begin{table}[!htbp] \centering
##   \caption{Effect of hearing status and vocabulary on Mel spectral distance}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lc}
## \\[-1.8ex]\hline
## \hline \\[-1.8ex]
##  Intercept & 11.35$^{***}$ \\
##   & (10.38, 12.32) \\
##   & t = 22.89 \\
##   & p = 0.00 \\
##   EVT-2 score & 0.03$^{*}$ \\
##   & (0.002, 0.05) \\
##   & t = 2.13 \\
##   & p = 0.04 \\
##   Chronological age matches & 0.21 \\
##   & ($-$0.93, 1.35) \\
##   & t = 0.36 \\
##   & p = 0.72 \\
##   Hearing age matches & 0.45 \\
##   & ($-$0.68, 1.59) \\
##   & t = 0.79 \\
##   & p = 0.44 \\
##   Child chronological age & 0.003 \\
##   & ($-$0.04, 0.05) \\
##   & t = 0.15 \\
##   & p = 0.89 \\
##   Word duration & 1.97 \\
##   & ($-$0.10, 4.04) \\
##   & t = 1.86 \\
##   & p = 0.07 \\
##   EVT-2 score*Chrono age match & $-$0.004 \\
##   & ($-$0.06, 0.05) \\
##   & t = $-$0.14 \\
##   & p = 0.89 \\
##   EVT-2 score*Hearing age match & $-$0.04$^{*}$ \\
##   & ($-$0.09, $-$0.005) \\
```

```
##   & t = $-$2.19 \\
##   & p = 0.03 \\
##  \hline \\[-1.8ex]
## Observations & 1,134 \\
## Log Likelihood & $-$3,016.76 \\
## Akaike Inf. Crit. & 6,055.52 \\
## Bayesian Inf. Crit. & 6,110.89 \\
## \hline
## \hline \\[-1.8ex]
## \textit{Note:}  & \multicolumn{1}{r}{$^{*}$p$<$0.05; $^{**}$p$<$0.01; $^{***}$p$<$0.001} \\
## \end{tabular}
## \end{table}
```

As before, we followed a forward-building model procedure to predict coarticulation within CV sequences with a baseline model containing random effects of **Speaker** and **Word** and fixed effects of **Word duration** and **Chronological Age**. As LENA recordings were only collected from the children with CIs and their hearing age matches, models with **Hourly Child Vocalization Count** only include those children. We did not find effects of **Articulation Skill** or **Minimal Pair Discrimination Ability**, or their interaction with hearing condition, upon children's coarticulation, so we excluded those variables from further analysis. For models with vocabulary size, the interaction of **Vocabulary Size** with **Hearing Status** improved upon the baseline model fit, indicating that the relationship between expressive vocabulary and coarticulation varied by hearing group (**??**; model summary included in Appendices). There was a significant, positive relationship between expressive vocabulary and degree of coarticulation for the children with CIs (EVT-2 score: β=0.03, t=2.13, p=0.04), but no reliably significant effects of vocabulary on coarticulation for either NH group, perhaps due to the NH groups' limited vocabulary score ranges. We elaborate upon this possibility in the Discussion.

```
# LENA
lena_model_data <- df.final3 %>% # TODO missing one hearing age match child
  filter(CVC_hourly!='NA') %>% # only looking at HA matches and CIs
  filter(Speaker!='665L52FS4' & Speaker!='679L58MS6' & Speaker!='802E72FS3' & # recorded 3+ months afte
           Speaker!='806E42MS1'& Speaker!='809E64MS2') %>%
  filter(tp!='3') %>%
  mutate(Word_duration=center_scale(Word_duration))
lena_model_data$EVT_GSV_centered <- center_scale(lena_model_data$EVT_GSV)
lena_model_data$CVC_hourly_centered <- center_scale(as.numeric(lena_model_data$CVC_hourly))
lena_model_data$CTC_hourly_centered <- center_scale(as.numeric(lena_model_data$CTC_hourly))
lena_model_data$AWC_hourly_centered <- center_scale(as.numeric(lena_model_data$AWC_hourly))
lena_model_data$Chrono_age_centered <- center_scale(lena_model_data$Chrono_age)

lenaI <- lena_model_data %>%
  lmer(euc_dist ~ match +
         Chrono_age_centered +
         Word_duration +
         (1|Speaker) +
         (1|Word),
       data=.)

lenaII <- lena_model_data %>%
  lmer(euc_dist ~ CVC_hourly_centered +
         Chrono_age_centered +
         match +
         Word_duration +
         (1|Speaker) +
```

```
        (1|Word),
      data=.)
anova(lenaI, lenaII) # not sig
```

```
## Data: .
## Models:
## lenaI: euc_dist ~ match + Chrono_age_centered + Word_duration + (1 |
## lenaI:     Speaker) + (1 | Word)
## lenaII: euc_dist ~ CVC_hourly_centered + Chrono_age_centered + match +
## lenaII:     Word_duration + (1 | Speaker) + (1 | Word)
##        npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## lenaI     7 3009.2 3039.5 -1497.6   2995.2
## lenaII    8 3010.5 3045.1 -1497.3   2994.5 0.7163  1     0.3974
```

```
lenaIII <- lena_model_data %>%
  lmer(euc_dist ~ CVC_hourly_centered*match +
        Chrono_age_centered +
        Word_duration +
        (1|Speaker) +
        (1|Word),
      data=.)
anova(lenaI, lenaIII) # interaction doesn't improve
```

```
## Data: .
## Models:
## lenaI: euc_dist ~ match + Chrono_age_centered + Word_duration + (1 |
## lenaI:     Speaker) + (1 | Word)
## lenaIII: euc_dist ~ CVC_hourly_centered * match + Chrono_age_centered +
## lenaIII:     Word_duration + (1 | Speaker) + (1 | Word)
##         npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## lenaI      7 3009.2 3039.5 -1497.6   2995.2
## lenaIII    9 3012.4 3051.2 -1497.2   2994.4 0.8814  2     0.6436
```

```
# vocab*CVC
lenaIV <- lena_model_data %>%
  lmer(euc_dist ~ EVT_GSV_centered*match +
        Chrono_age_centered +
        Word_duration +
        (1|Speaker) +
        (1|Word),
      data=.)

lenaV <- lena_model_data %>%
  lmer(euc_dist ~ CVC_hourly_centered*EVT_GSV_centered*match +
        Chrono_age_centered +
        Word_duration +
        (1|Speaker) +
        (1|Word),
      data=.)
anova(lenaIV, lenaV) # neither CVC or CTC improve upon vocab-only model
```

```
## Data: .
## Models:
```

```
## lenaIV: euc_dist ~ EVT_GSV_centered * match + Chrono_age_centered + Word_duration +
## lenaIV:     (1 | Speaker) + (1 | Word)
## lenaV: euc_dist ~ CVC_hourly_centered * EVT_GSV_centered * match + Chrono_age_centered +
## lenaV:     Word_duration + (1 | Speaker) + (1 | Word)
##        npar    AIC    BIC  logLik deviance Chisq Df Pr(>Chisq)
## lenaIV    9 3011.8 3050.7 -1496.9   2993.8
## lenaV    13 3016.1 3072.2 -1495.0   2990.1  3.78  4     0.4366
```

```
# no effect of vocalization count

# EXPLORATORY: let's look at just the kids w/ NH
lena_HA <- lena_model_data %>%
  filter(match=='HA_match') %>%
  lmer(euc_dist ~ CVC_hourly_centered +
         Chrono_age_centered +
         Word_duration +
         (1|Speaker) +
         (1|Word),
       data=.)  # no effect

# let's look at just the kids with CIs
lena_CI <- lena_model_data %>%
  filter(match=='CI') %>%
  lmer(euc_dist ~ CVC_hourly_centered + # exclude chrono age because correlated
         Word_duration +
         (1|Speaker) +
         (1|Word),
       data=.)  # no effect

# now let's look at just the kids with CIs who have 3+ years of hearing experience (median hearing age
lena_CI2 <- lena_model_data %>%
  filter(match=='CI') %>%
  filter(Hearing_age >35) %>%
  lmer(euc_dist ~ CVC_hourly_centered + # exclude chrono age because correlated
         Word_duration +
         (1|Speaker) +
         (1|Word),
       data=.)

lena_CI3 <- lena_model_data %>%
  filter(match=='CI') %>%
  lmer(euc_dist ~ CVC_hourly_centered*Hearing_age +
         Word_duration +
         (1|Speaker) +
         (1|Word),
       data=.) # no effect when we code hearing age continuously

# do a model that includes all kids but with a variable that has them split
split_model_data <- lena_model_data %>%
  filter(match=='CI') %>%
  mutate(ha_group=if_else(Hearing_age > 35, "older hearing age", "younger hearing age"))

split_model_data$ha_group <- factor(split_model_data$ha_group, ordered = FALSE )
split_model_data$ha_group <- relevel(split_model_data$ha_group, ref = "older hearing age")
```

```
lena_CI3a <- split_model_data %>%
  lmer(euc_dist ~ ha_group + # exclude chrono age because correlated
         Word_duration +
         (1|Speaker) +
         (1|Word),
       data=.)

# ***** final model *****
lena_CI4 <- split_model_data %>%
  lmer(euc_dist ~ CVC_hourly_centered*ha_group + # exclude chrono age because correlated
         Word_duration +
         (1|Speaker) +
         (1|Word),
       data=.)
anova(lena_CI3a, lena_CI4) # improves under alpha
```

```
## Data: .
## Models:
## lena_CI3a: euc_dist ~ ha_group + Word_duration + (1 | Speaker) + (1 | Word)
## lena_CI4: euc_dist ~ CVC_hourly_centered * ha_group + Word_duration + (1 |
## lena_CI4:     Speaker) + (1 | Word)
##           npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## lena_CI3a    6 1435.1 1456.8 -711.54   1423.1
## lena_CI4     8 1433.8 1462.8 -708.91   1417.8 5.2469  2    0.07255 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
cvc_model <- summary(lena_CI4)


# TODO: GFTA, perception, and vocab are not significant in models with CVC; fit models showing that tho
# with CVC*ha_group
```

Hourly Child Vocalization Count from the LENA recordings did not improve upon the baseline model: there was no effect of daily speech practice upon the degree of coarticulation for children with CIs or their hearing age-matched peers. This null result was unexpected given the effect of child vocalization frequency upon coarticulatory development that we found in our previous work on four-year-olds with NH. For the children with NH in this study, we hypothesized that the null result could stem from the limited number of hearing age matches studied here or from the smaller range of hourly child vocalization frequency, a point that we return to in the Discussion.

For the children with CIs, we hypothesized that there could be another source of the null result. We considered that for children with CIs, post-implantation it may take time to learn to incorporate auditory feedback into speech routines and as such, we might not see an effect of vocal output frequency upon speech production outcomes immediately or even several months after implantation. We explored this idea by performing a median split upon the children with CIs by their hearing age (median hearing age = 36 months). We then fit the model as before but with the interaction of the binary variable **Hearing Age Group** (<>35 months hearing experience) and **Hourly Child Vocalization Count**. (Note that the models with the **Hearing Age Group** parameter did not include **Chronological Age** as the parameters not independent.) While the **Hearing Age Group*Hourly Child Vocalization Count** interaction was not significant in the model, **Hourly Child Vocalization Count** was significant with the reference level "> 35 months hearing experience" (β=0.01, t=2.2, p=0.045), but not significant with the reference level "<= 35 months hearing experience", indicating that the effects of vocal output upon degree of coarticulation tend to appear in the children with more than three years hearing experience, but not children with less than three years experience.

```r
# GFTA
gfta_model_data <- df.final3 %>%
  filter(GFTA_Standard!='NA') %>% # participants XX didn't complete GFTA
  mutate(Word_duration = center_scale(Word_duration))
gfta_model_data$GFTA_Standard_centered <- center_scale(gfta_model_data$GFTA_Standard)
gfta_model_data$Chrono_age_centered <- center_scale(gfta_model_data$Chrono_age)

gftaI <- gfta_model_data %>%
  lmer(euc_dist ~ match +
         Chrono_age_centered +
         Word_duration +
         (1|Speaker) +
         (1|Word),
       data=.)

gftaII <- gfta_model_data %>%
  lmer(euc_dist ~ GFTA_Standard_centered +
         match +
         Chrono_age_centered +
         Word_duration +
         (1|Speaker) +
         (1|Word),
       data=.)
anova(gftaI, gftaII) # no improvement
```

```
## Data: .
## Models:
## gftaI: euc_dist ~ match + Chrono_age_centered + Word_duration + (1 |
## gftaI:     Speaker) + (1 | Word)
## gftaII: euc_dist ~ GFTA_Standard_centered + match + Chrono_age_centered +
## gftaII:     Word_duration + (1 | Speaker) + (1 | Word)
##        npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## gftaI     8 5686.5 5726.3 -2835.3   5670.5
## gftaII    9 5688.4 5733.2 -2835.2   5670.4 0.0771  1     0.7812
```

```r
gftaIII <- gfta_model_data %>%
  lmer(euc_dist ~ GFTA_Standard_centered*match +
         Chrono_age_centered +
         Word_duration +
         (1|Speaker) +
         (1|Word),
       data=.)
anova(gftaI, gftaIII) # no improvement
```

```
## Data: .
## Models:
## gftaI: euc_dist ~ match + Chrono_age_centered + Word_duration + (1 |
## gftaI:     Speaker) + (1 | Word)
## gftaIII: euc_dist ~ GFTA_Standard_centered * match + Chrono_age_centered +
## gftaIII:     Word_duration + (1 | Speaker) + (1 | Word)
##         npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## gftaI      8 5686.5 5726.3 -2835.3   5670.5
## gftaIII   11 5688.8 5743.6 -2833.4   5666.8 3.7015  3     0.2956
```

```r
gftaIV <- gfta_model_data %>%
  filter(match=='CI') %>%
  lmer(euc_dist ~ GFTA_Standard_centered +
         Chrono_age_centered +
         Word_duration +
         (1|Speaker) +
         (1|Word),
       data=.)

# no effect of GFTA on overall coartic, no interaction with match, and no effect in models just w/ CI k
```

```r
# min pair discrimination
# TODO: remove the kids who completed less than a third of trials (<10)
# note that the tp3 kids didn't do the min pair task
minpair_model_data <- df.final3 %>%
  filter(MinPair_ProportionCorrect!='NA') %>% # TODO: check which participants didn't complete this tas
  filter(match!='Chrono_age_match') %>% # mostly don't have data from them
  mutate(Word_duration = center_scale(Word_duration))
minpair_model_data$MinPair_ProportionCorrect_centered <- center_scale(minpair_model_data$MinPair_Propor
minpair_model_data$Chrono_age_centered <- center_scale(minpair_model_data$Chrono_age)

minI <- minpair_model_data %>%
  lmer(euc_dist ~ match +
         Chrono_age_centered +
         Word_duration +
         (1|Speaker) +
         (1|Word),
       data=.)

minII <- minpair_model_data %>%
  lmer(euc_dist ~ MinPair_ProportionCorrect_centered +
         match +
         Chrono_age_centered +
         Word_duration +
         (1|Speaker) +
         (1|Word),
       data=.)
anova(minI, minII)  # no improvement
```

```
## Data: .
## Models:
## minI: euc_dist ~ match + Chrono_age_centered + Word_duration + (1 |
## minI:     Speaker) + (1 | Word)
## minII: euc_dist ~ MinPair_ProportionCorrect_centered + match + Chrono_age_centered +
## minII:     Word_duration + (1 | Speaker) + (1 | Word)
##       npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## minI     7 3379.3 3410.5 -1682.6   3365.3
## minII    8 3381.1 3416.8 -1682.6   3365.1 0.1515  1     0.6971
```

```r
minIII <- minpair_model_data %>%
  lmer(euc_dist ~ MinPair_ProportionCorrect_centered*match +
         Chrono_age_centered +
         Word_duration +
```

```
        (1|Speaker) +
        (1|Word),
      data=.)
anova(minI, minIII)  #  no interaction of perceptual experience with match
```

```
## Data: .
## Models:
## minI: euc_dist ~ match + Chrono_age_centered + Word_duration + (1 |
## minI:     Speaker) + (1 | Word)
## minIII: euc_dist ~ MinPair_ProportionCorrect_centered * match + Chrono_age_centered +
## minIII:     Word_duration + (1 | Speaker) + (1 | Word)
##        npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## minI      7 3379.3 3410.5 -1682.6   3365.3
## minIII    9 3381.9 3422.0 -1681.9   3363.9 1.4118  2     0.4937
```

```
# let's look at just the kids w/ CIs who we think there may be an effect of perceptual ability on
minCI <- minpair_model_data %>%
  filter(match=='CI') %>%
  lmer(euc_dist ~ MinPair_ProportionCorrect_centered +
        EVT_GSV +
        Chrono_age_centered +
        Word_duration +
        (1|Speaker) +
        (1|Word),
      data=.)
summary(minCI)
```

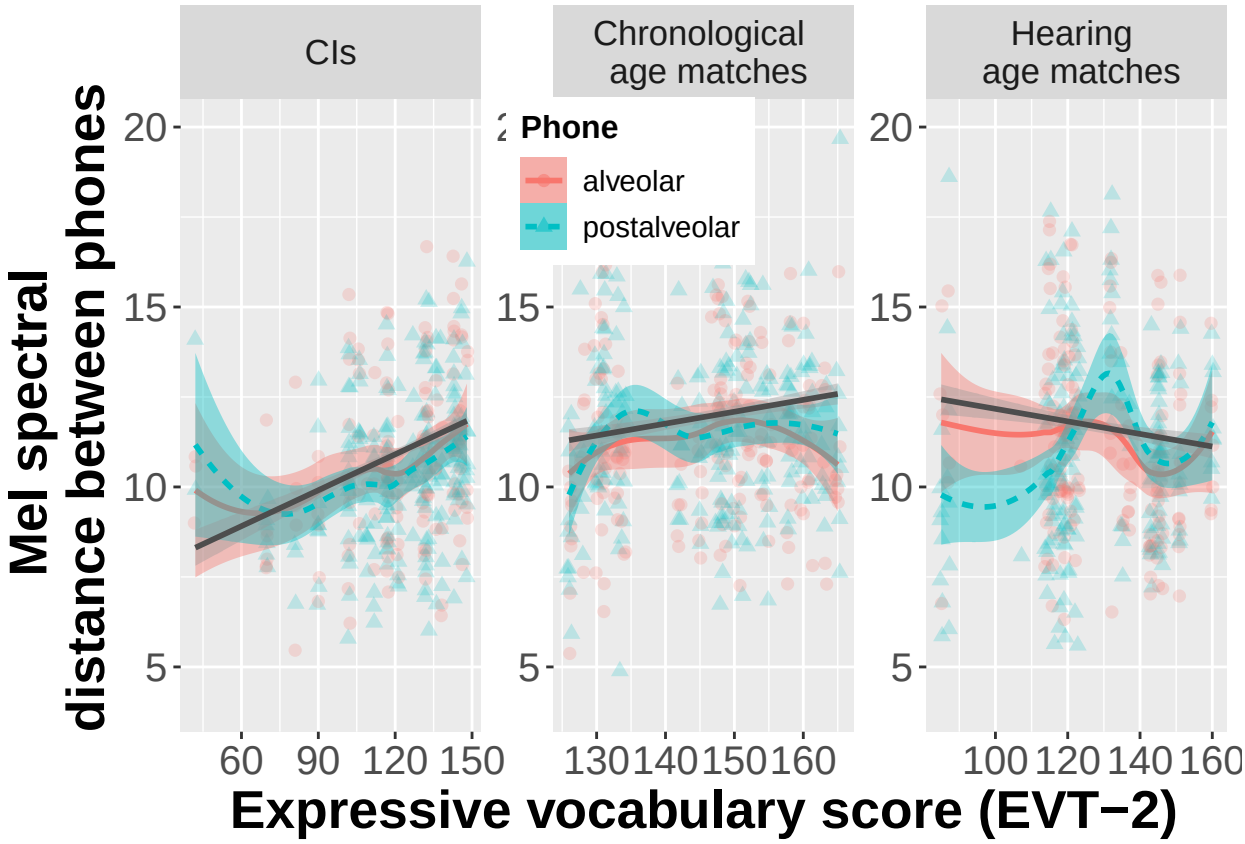
```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula:
## euc_dist ~ MinPair_ProportionCorrect_centered + EVT_GSV + Chrono_age_centered +
##     Word_duration + (1 | Speaker) + (1 | Word)
##    Data: .
##
## REML criterion at convergence: 1837
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.5924 -0.5218 -0.1518  0.3101  6.0121
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Speaker  (Intercept) 0.1705   0.4129
##  Word     (Intercept) 0.7502   0.8661
##  Residual             9.1160   3.0193
## Number of obs: 359, groups:  Speaker, 26; Word, 14
##
## Fixed effects:
##                                    Estimate Std. Error         df t value
## (Intercept)                        6.825216   1.199258  32.250296   5.691
## MinPair_ProportionCorrect_centered -1.141984   1.558458  20.381512  -0.733
## EVT_GSV                            0.034306   0.009601  27.070500   3.573
## Chrono_age_centered                0.005725   0.023819  18.865122   0.240
```

```
## Word_duration                         1.936397   1.885997 319.433596   1.027
##                                     Pr(>|t|)
## (Intercept)                         2.59e-06 ***
## MinPair_ProportionCorrect_centered  0.47205
## EVT_GSV                             0.00135 **
## Chrono_age_centered                 0.81264
## Word_duration                       0.30533
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) MP_PC_ EVT_GS Chrn__
## MnPr_PrprC_  0.506
## EVT_GSV     -0.952 -0.441
## Chrn_g_cntr -0.118 -0.443 -0.067
## Word_duratn  0.000  0.000  0.000  0.000
```

```
# no effect of min discrim just in kids with CIs

minIV <- minpair_model_data %>%
  filter(match=='CI') %>%
  lmer(euc_dist ~ MinPair_ProportionCorrect_centered*EVT_GSV +
         Chrono_age_centered +
         Word_duration +
         (1|Speaker) +
         (1|Word),
       data=.)
# no effect of discrim ability dependent upon vocab
```

```
vocabIII <- vocab_model_data %>%
  lmer(euc_dist ~ EVT_GSV*match +
         Chrono_age_centered +
         Word_duration +
         (1|Speaker) +
         (1|Word),
       data=.)
vocab_model_data$fit <- predict(vocabIII)

# write out the figure manually ot adjust panel size
#jpeg("/Users/megcychosz/Google Drive/CI_feedback/analysis/results/2_feedback_results_files/figure-late
vocab_model_data %>%
  #group_by(Speaker, EVT_GSV, match, poa) %>%
  #summarize(median_euc_dist = median(euc_dist)) %>%
  mutate(match=(recode(match,"Chrono_age_match"="Chronological \n age matches",
                       "CI"="CIs", "HA_match"="Hearing \n age matches")),
         Phone=poa) %>%
  ggplot(., aes(x=EVT_GSV, y=euc_dist)) +
  ylim(4,20) +
  geom_jitter(size=2, alpha=.2, aes(color=Phone, shape=Phone, fill=Phone)) +
  geom_smooth(aes(fill=Phone, color=Phone, lty=Phone)) +
  geom_smooth(method="lm", aes(y=fit), size=1,color="grey30") + # the actual model fit
  facet_wrap(~match, scales="free") +

  xlab("Expressive vocabulary score (EVT-2)") +
```

```
  ylab("Mel spectral \n distance between phones") +
  #labs(col='Place of articulation') +
  theme(axis.text=element_text(size=15),
      axis.title=element_text(size=20,face="bold"),
      legend.title = element_text(size=12, face="bold"),
      legend.text = element_text(size=11),
      strip.text.x = element_text(size=13),
      legend.position = c(.43,.87),
      legend.background = element_rect(fill="white",
                                       size=0.5, linetype="solid")) +
  guides(colour = guide_legend(override.aes = list(alpha = .55)))
```

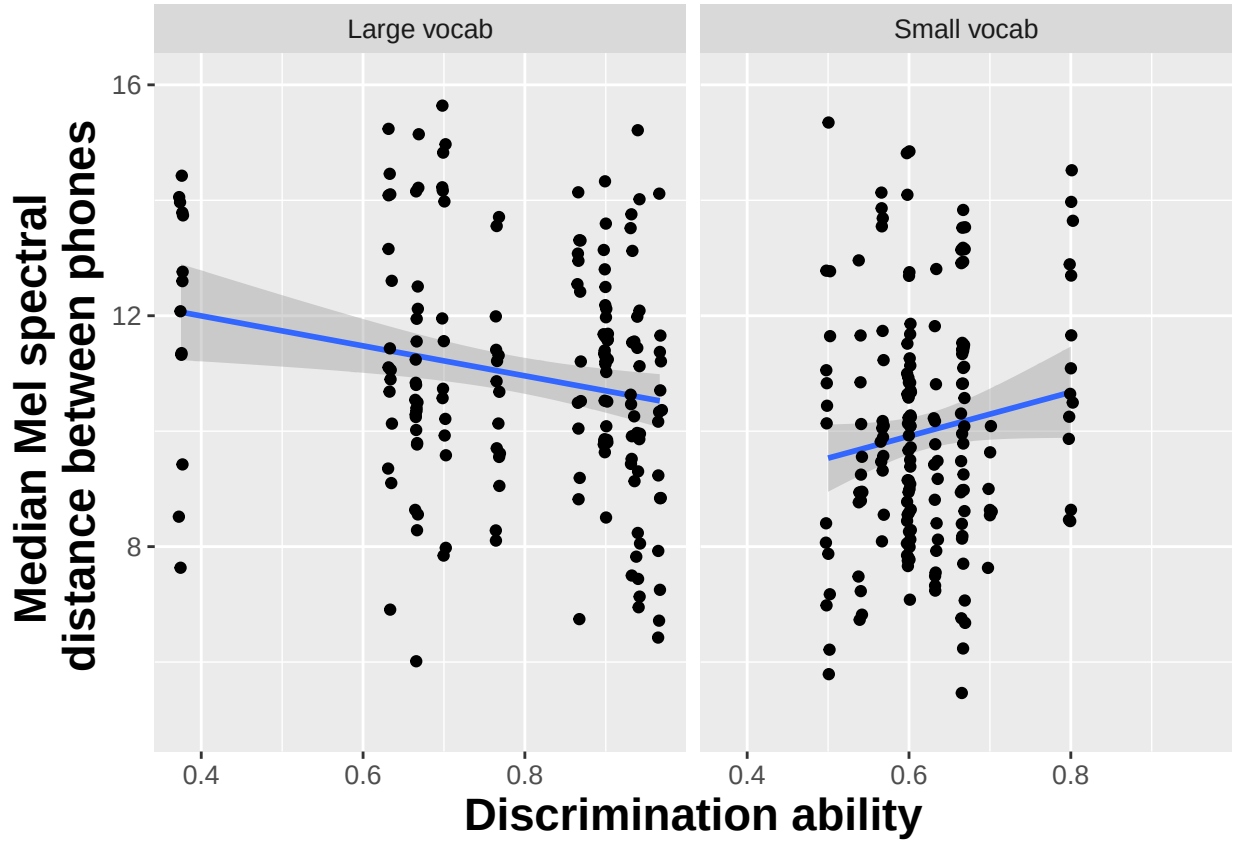



```
#dev.off()
```

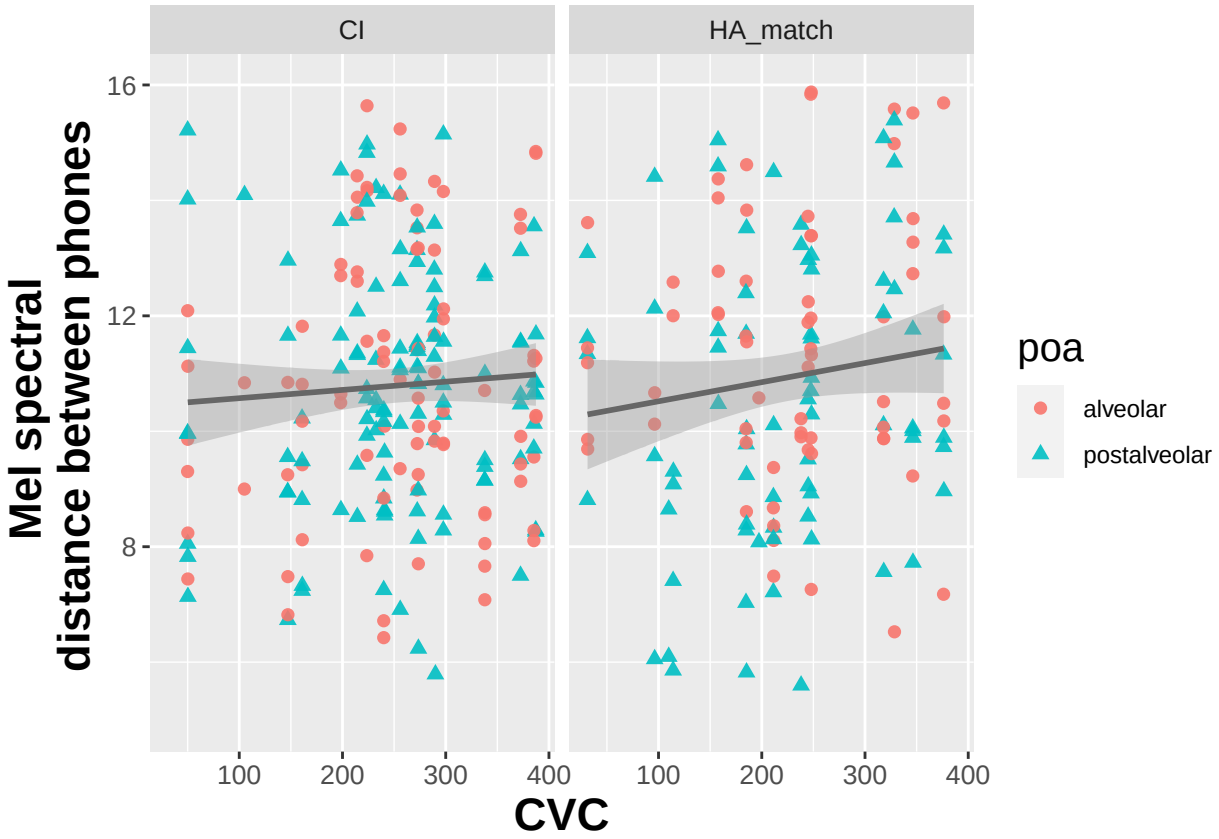
5 not using

```
# split the CI kids on their vocab scores
med_evt <- df.final3 %>%
  filter(EVT_GSV!='NA') %>%
  filter(match=='CI') %>%
  #group_by(match) %>%
  ungroup() %>%
  summarize(med_evt=median(as.numeric(EVT_GSV)))
```

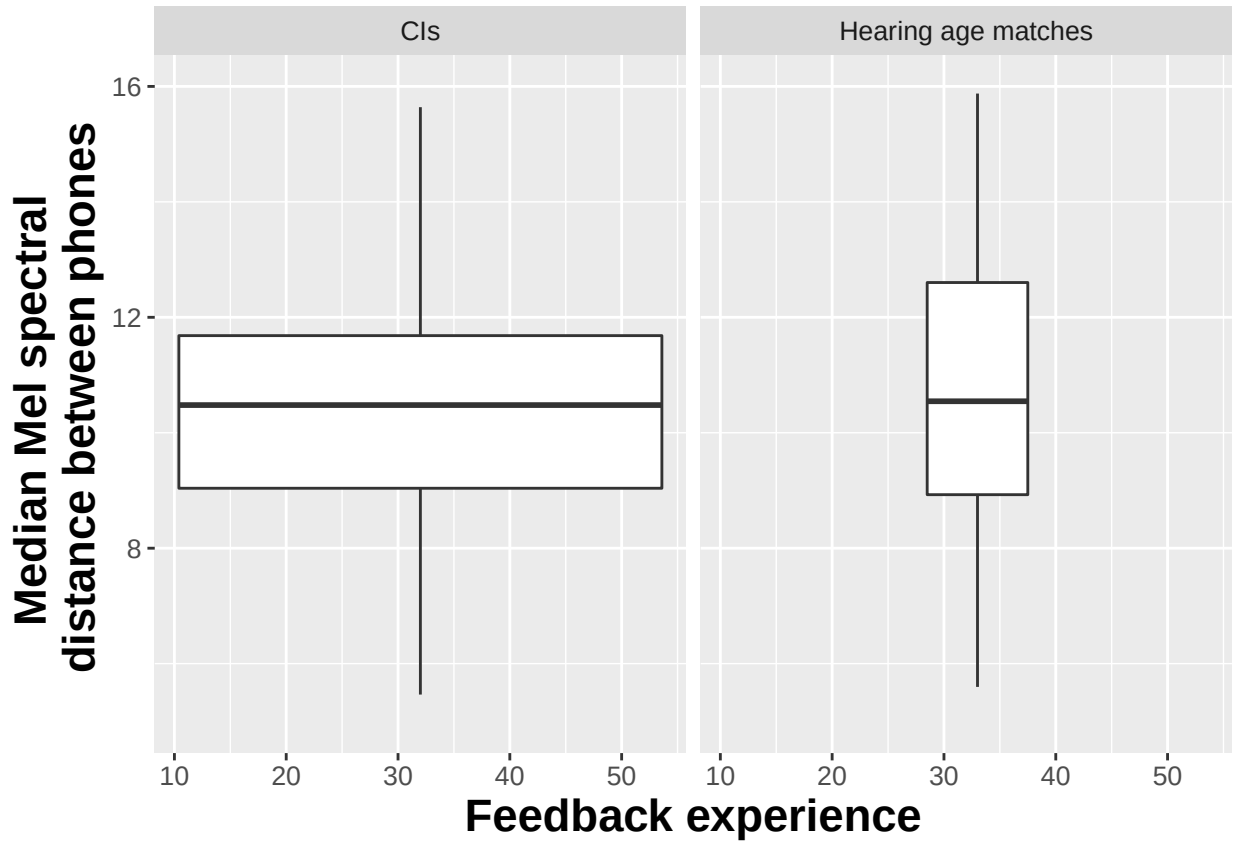
```r
# CIS
CI_evt<- df.final3 %>%
  filter(hearing_status=='CI') %>%
  filter(EVT_GSV!='NA') %>%
  mutate(vocab_group = if_else(EVT_GSV < 127,
                               "Small vocab",
                               "Large vocab"))

 CI_evt %>%
  #group_by(Speaker, CVC_hourly, feedback_experience, match, poa) %>%
  #summarize(median_euc_dist = median(euc_dist)) %>%
  mutate(Phone=poa) %>%
  #filter(match=='CIs') %>%
  ggplot(., aes(x=MinPair_ProportionCorrect, y=euc_dist)) +
  ylim(5,16) +
  #geom_boxplot(aes()) +
  geom_smooth(method="lm") +
   geom_jitter() +
  facet_wrap(~vocab_group) +
  xlab("Discrimination ability") +
  ylab("Median Mel spectral \n distance between phones") +
  theme(axis.text=element_text(size=10),
      axis.title=element_text(size=17,face="bold"),
      legend.title = element_text(size=15),
      strip.text.x = element_text(size=10))
```



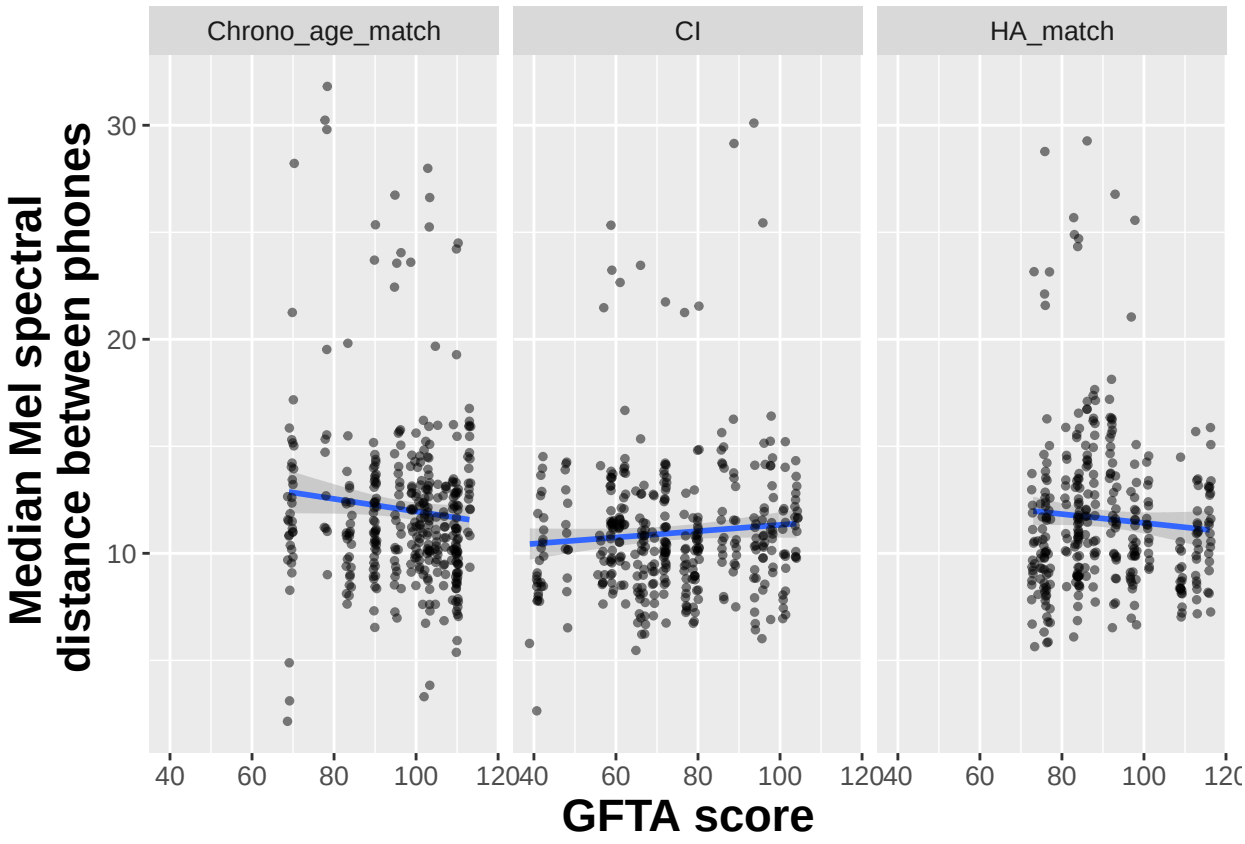
```
#665L52FS4 completed recording 6 months later 679L58MS6 completed recording 4 months alter
#802E72FS3 did it 4 months later; 806E42MS1 did it 3 months later
#809E64MS2 did LENA recording 4 months later

lena_model_data %>%
  filter(match!='Chrono_age_match') %>%
   #group_by(Speaker, poa) %>% # don't need to group by match bc only ha matches
  #summarize(median_euc_dist = median(euc_dist)) %>%
  #merge(., df.final3, by=c("Speaker", "poa")) %>%
  distinct_at(., vars(Speaker, poa), .keep_all = T) %>%
  #mutate(match=(recode(match,"CI"="CIs", "HA_match"="Hearing age matches")),
   #      Phone=poa,
    #     CVC_hourly=as.numeric(CVC_hourly)) %>%
  ggplot(., aes(x=CVC_hourly, y=euc_dist)) +
  ylim(5,16) + # TODO add a note saying that one outlier was removed
  geom_jitter(size=2, alpha=.9, aes(color=poa, shape=poa)) +
  geom_smooth(method='lm', color='gray40') +
  facet_wrap(~match) +
  xlab("CVC") +
  ylab("Mel spectral \n distance between phones") +
  theme(axis.text=element_text(size=10),
      axis.title=element_text(size=17,face="bold"),
      legend.title = element_text(size=15),
      strip.text.x = element_text(size=10))
```



```r
# perform a median split by CVC which is a proxy for auditory feedback experience and see if there is a
med_CVC <- df.final3 %>%
  filter(CVC_hourly!='NA') %>%
  group_by(match) %>%
  summarize(med_CVC=median(as.numeric(CVC_hourly)))

# CIS
CI_auditory <- df.final3 %>%
  filter(hearing_status=='CI') %>%
  filter(CVC_hourly!='NA') %>%
  mutate(feedback_experience = if_else(CVC_hourly < 256,
                               "Less auditory \n feedback group",
                               "More auditory \n feedback group"))
# hearing age matches
HA_auditory <- df.final3 %>%
  filter(match=='HA_match') %>%
  filter(CVC_hourly!='NA') %>%
  mutate(feedback_experience = if_else(CVC_hourly < 245,
                               "Less auditory \n feedback group",
                               "More auditory \n feedback group")) %>%
  rbind(., CI_auditory)


 HA_auditory %>%
  #group_by(Speaker, CVC_hourly, feedback_experience, match, poa) %>%
  #summarize(median_euc_dist = median(euc_dist)) %>%
  mutate(match=(recode(match,"CI"="CIs", "HA_match"="Hearing age matches")),
         Phone=poa,
         CVC_hourly=as.numeric(CVC_hourly)) %>%
  #filter(match=='CIs') %>%
  ggplot(., aes(x=Hearing_age, y=euc_dist)) +
  ylim(5,16) + # TODO add a note saying that one outlier was removed
  geom_boxplot(aes()) +
  #geom_jitter() +
  facet_wrap(~match) +
  xlab("Feedback experience") +
  ylab("Median Mel spectral \n distance between phones") +
  theme(axis.text=element_text(size=10),
      axis.title=element_text(size=17,face="bold"),
      legend.title = element_text(size=15),
      strip.text.x = element_text(size=10))
```

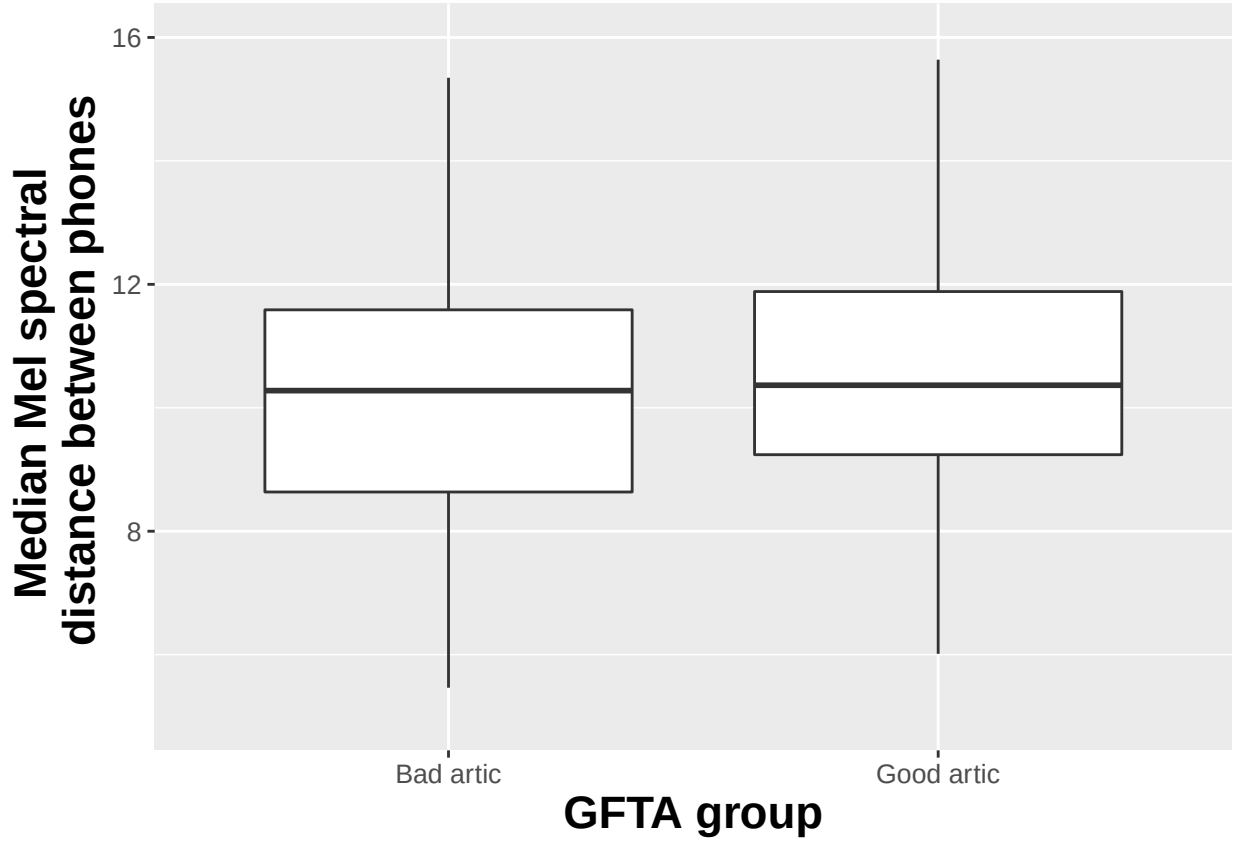



```
gfta_model_data %>%
  ungroup() %>%
  #distinct(Speaker, .keep_all = T) %>%
 #group_by(Speaker, match, poa, MinPair_ProportionCorrect) %>%
 #summarize(median_euc_dist = median(euc_dist)) %>%
 mutate(Phone=poa) %>%
 #filter(match=='CIs') %>%
 ggplot(., aes(x=GFTA_Standard, y=euc_dist)) +
 #ylim(5,16) +
 #geom_boxplot(aes()) +
 geom_smooth(method="lm") +
  geom_jitter(size=1, alpha=.5) +
 facet_wrap(~match) +
 xlab("GFTA score") +
 ylab("Median Mel spectral \n distance between phones") +
 theme(axis.text=element_text(size=10),
     axis.title=element_text(size=17,face="bold"),
     legend.title = element_text(size=15),
     strip.text.x = element_text(size=10))
```



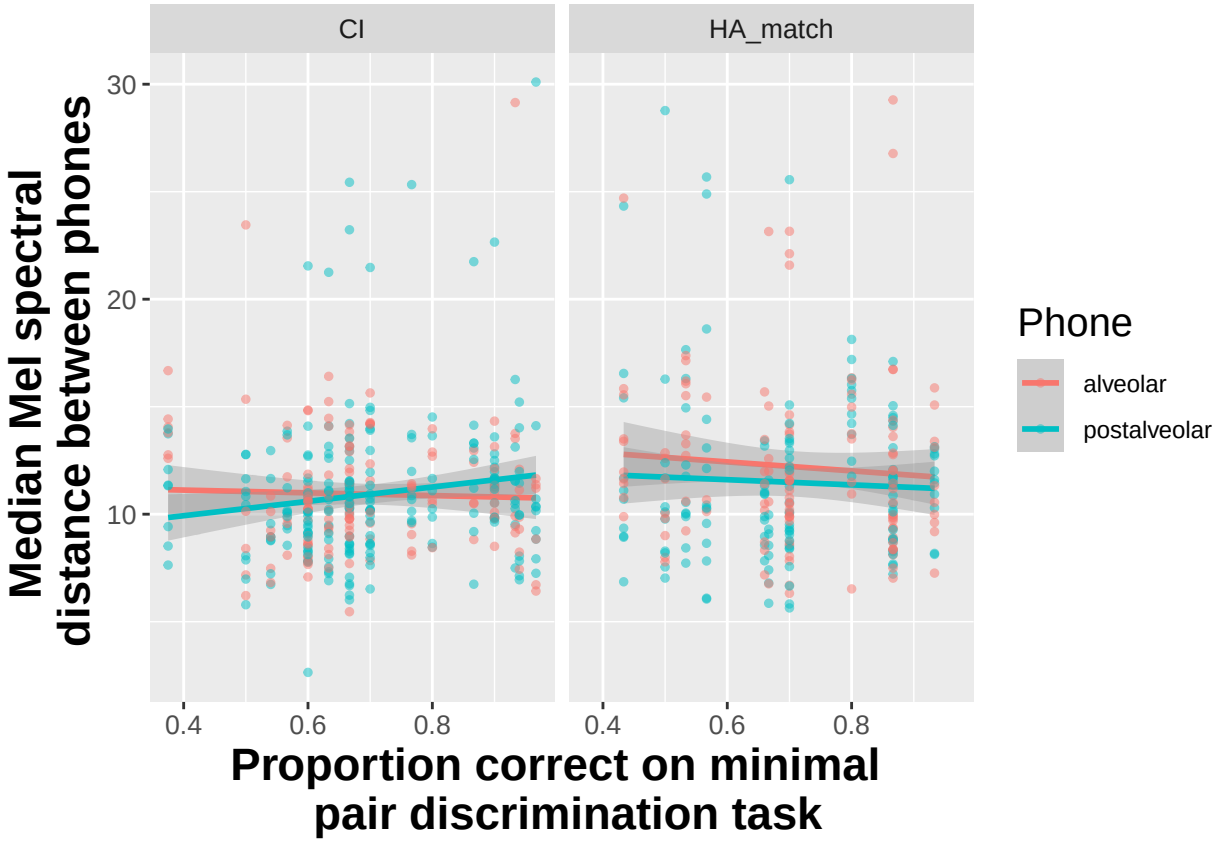

```
# split the CI kids on their GFTA scores
# absolutely no coartic difference based on GFTA scores
med_gfta <- df.final3 %>%
  filter(GFTA_Standard!='NA') %>%
  filter(match=='CI') %>%
  #group_by(match) %>%
  ungroup() %>%
  summarize(med_gfta=median(as.numeric(GFTA_Standard)))

# CIS
CI_gfta<- df.final3 %>%
  filter(hearing_status=='CI') %>%
  filter(GFTA_Standard!='NA') %>%
  mutate(gfta_group = if_else(GFTA_Standard < 72,
                              "Bad artic",
                              "Good artic"))

 CI_gfta %>%
  #group_by(Speaker, CVC_hourly, feedback_experience, match, poa) %>%
  #summarize(median_euc_dist = median(euc_dist)) %>%
  mutate(Phone=poa) %>%
  #filter(match=='CIs') %>%
  ggplot(., aes(x=gfta_group, y=euc_dist)) +
  ylim(5,16) +
  geom_boxplot(aes()) +
  #geom_jitter() +
  #facet_wrap(~match) +
```

```
 xlab("GFTA group") +
 ylab("Median Mel spectral \n distance between phones") +
 theme(axis.text=element_text(size=10),
     axis.title=element_text(size=17,face="bold"),
     legend.title = element_text(size=15),
     strip.text.x = element_text(size=10))
```



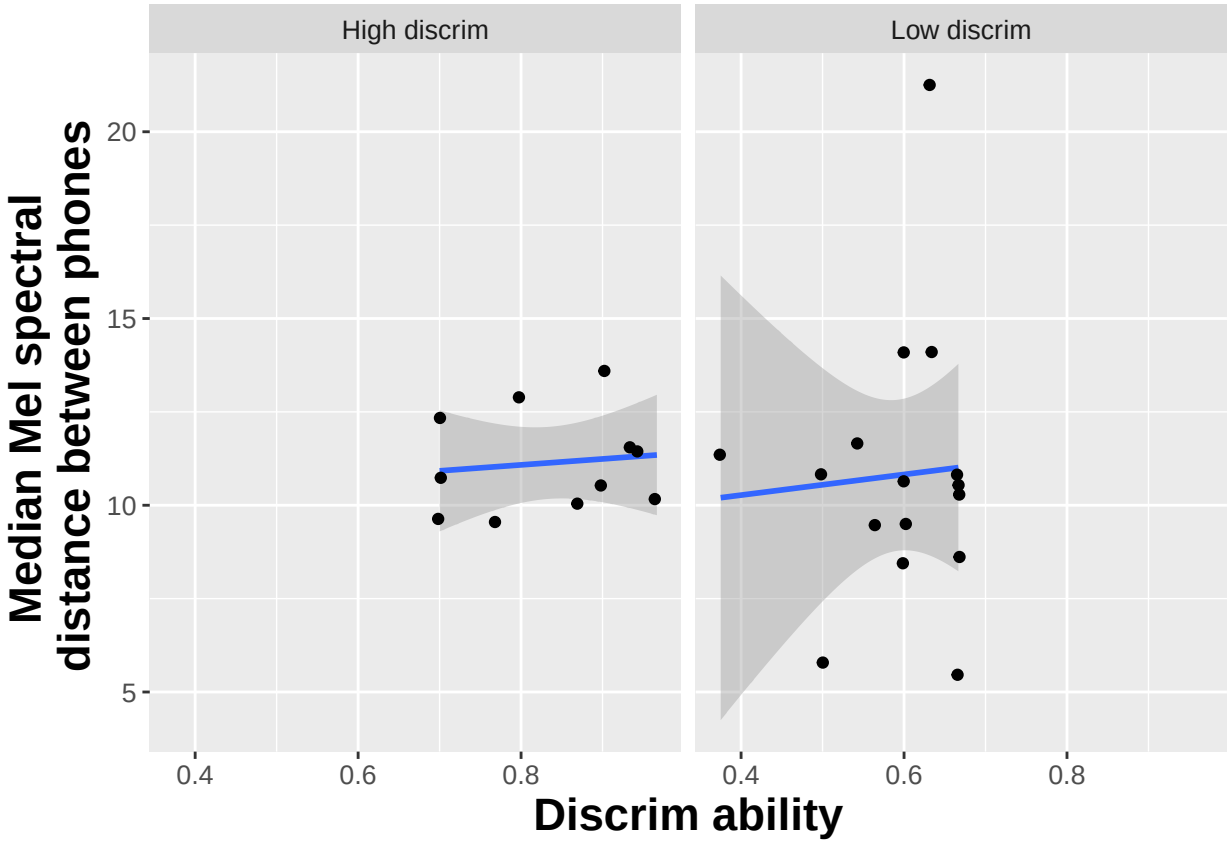

```
minpair_model_data %>%
  ungroup() %>%
 filter(match!='Chrono_age_match') %>% # no data from this group
  #distinct(Speaker, .keep_all = T) %>%
 #group_by(Speaker, match, poa, MinPair_ProportionCorrect) %>%
 #summarize(median_euc_dist = median(euc_dist)) %>%
 mutate(Phone=poa) %>%
 #filter(match=='CIs') %>%
 ggplot(., aes(x=MinPair_ProportionCorrect, y=euc_dist)) +
 #ylim(5,16) +
 #geom_boxplot(aes()) +
 geom_smooth(aes(color=Phone),method="lm") +
  geom_jitter(aes(color=Phone),size=1, alpha=.5) +
 facet_wrap(~match) +
 xlab("Proportion correct on minimal \n pair discrimination task") +
 ylab("Median Mel spectral \n distance between phones") +
 theme(axis.text=element_text(size=10),
     axis.title=element_text(size=17,face="bold"),
     legend.title = element_text(size=15),
```

```
      strip.text.x = element_text(size=10))
```



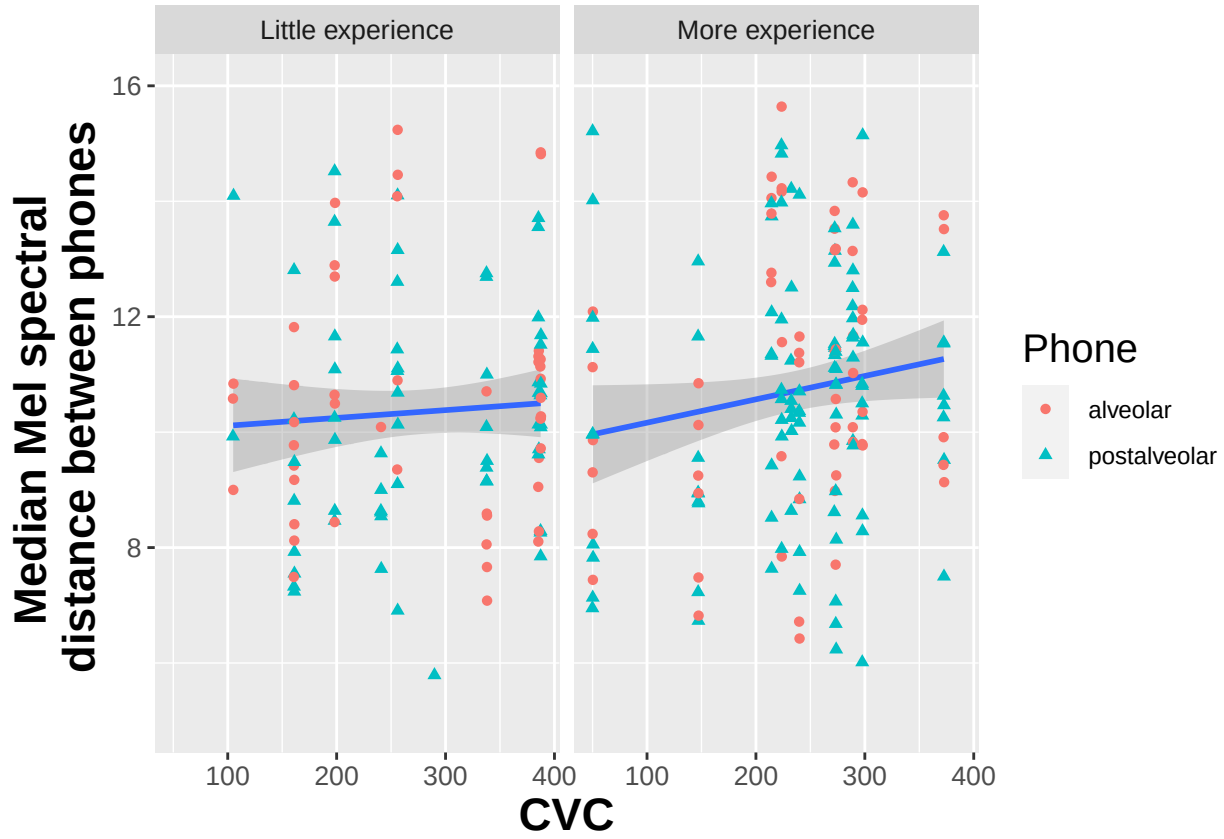

```
# split the CI kids on their minpair discrim scores
med_min_pair <- df.final3 %>%
  filter(MinPair_ProportionCorrect!='NA') %>%
  filter(match=='CI') %>%
  filter(Speaker!='304E48FS2') %>% # this kid listened to less than 10 trials
  #group_by(match) %>%
  ungroup() %>%
  summarize(med_min_pair=median(as.numeric(MinPair_ProportionCorrect)))

# CIS
CI_min_pair<- df.final3 %>%
  filter(hearing_status=='CI') %>%
  filter(MinPair_ProportionCorrect!='NA') %>%
  mutate(min_pair_group = if_else(MinPair_ProportionCorrect < .667,
                                  "Low discrim",
                                  "High discrim"))

 CI_min_pair %>%
   ungroup() %>%
   distinct(Speaker, .keep_all = T) %>%
  #group_by(Speaker, match, poa, MinPair_ProportionCorrect) %>%
  #summarize(median_euc_dist = median(euc_dist)) %>%
  mutate(Phone=poa) %>%
  #filter(match=='CIs') %>%
```

```
  ggplot(., aes(x=MinPair_ProportionCorrect, y=euc_dist)) +
  #ylim(5,16) +
  #geom_boxplot(aes()) +
  geom_smooth(method="lm") +
   geom_jitter() +
  facet_wrap(~min_pair_group) +
  xlab("Discrim ability") +
  ylab("Median Mel spectral \n distance between phones") +
  theme(axis.text=element_text(size=10),
      axis.title=element_text(size=17,face="bold"),
      legend.title = element_text(size=15),
      strip.text.x = element_text(size=10))
```



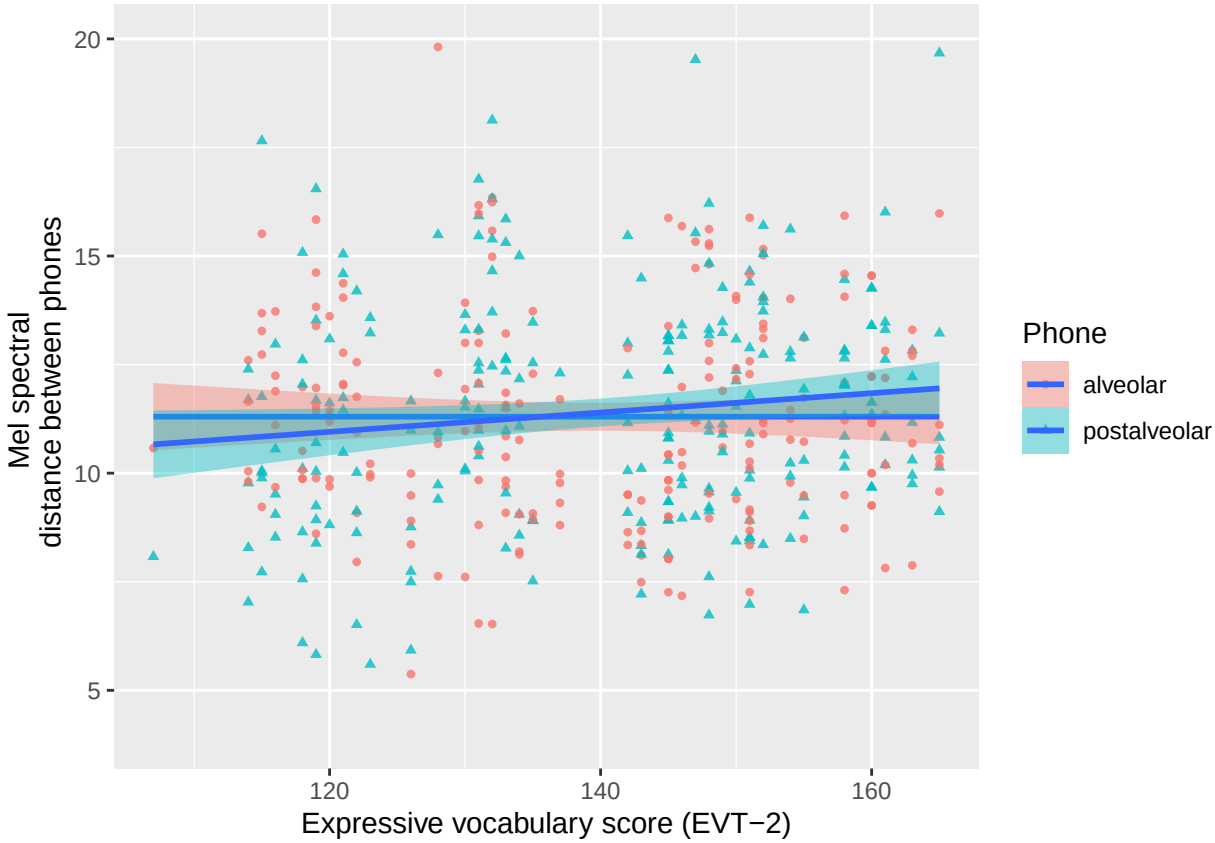

```
med_ha <- lena_model_data %>%
  filter(match=='CI') %>%
  #group_by(match) %>%
  ungroup() %>%
  summarize(med_ha=median(as.numeric(Hearing_age)))

# CIS
CI_ha <- lena_model_data %>%
  filter(hearing_status=='CI') %>%
  mutate(ha_group = if_else(Hearing_age < 36,
                            "Little experience",
                            "More experience"))
```

```
CI_ha %>%
  group_by(Speaker, CVC_hourly, ha_group, poa) %>%
 # summarize(median_euc_dist = median(euc_dist)) %>%
  mutate(Phone=poa) %>%
  #filter(match=='CIs') %>%
  ggplot(., aes(x=CVC_hourly, y=euc_dist)) +
  ylim(5,16) +
  #geom_boxplot(aes()) +
  geom_smooth(method="lm") +
  geom_jitter(aes(shape=Phone,color=Phone)) +
  facet_wrap(~ha_group) +
  xlab("CVC") +
  ylab("Median Mel spectral \n distance between phones") +
  theme(axis.text=element_text(size=10),
      axis.title=element_text(size=17,face="bold"),
      legend.title = element_text(size=15),
      strip.text.x = element_text(size=10))
```




```
vocab_model_data %>%
  #group_by(Speaker, EVT_GSV, match, poa) %>%
  #summarize(median_euc_dist = median(euc_dist)) %>%
  mutate(match=(recode(match,"Chrono_age_match"="Chronological \n age matches",
                       "CI"="CIs", "HA_match"="Hearing \n age matches")),
         Phone=poa) %>%
  filter(match!='CIs') %>%
  mutate(med_euc_dist = median(euc_dist)) %>%
  distinct(Speaker, .keep_all = T) %>%
```

```
  filter(EVT_GSV>100) %>%
  ggplot(., aes(x=EVT_GSV, y=euc_dist)) +
  ylim(4,20) +
  geom_point(size=1.2, alpha=.8, aes(color=Phone, shape=Phone, fill=Phone)) +
  #facet_wrap(~Speaker, scales="free") +
  geom_smooth(aes(fill=Phone),method='lm') +

  xlab("Expressive vocabulary score (EVT-2)") +
  ylab("Mel spectral \n distance between phones") #+
```




```
  #labs(col='Place of articulation') +
  theme(axis.text=element_text(size=15),
      axis.title=element_text(size=20,face="bold"),
      legend.title = element_text(size=12, face="bold"),
      legend.text = element_text(size=11),
      strip.text.x = element_text(size=13),
      legend.position = c(.43,.87),
      legend.background = element_rect(fill="white",
                                       size=0.5, linetype="solid")) +
  guides(colour = guide_legend(override.aes = list(alpha = .55)))
```

```
## List of 8
##  $ axis.title       :List of 11
##   ..$ family       : NULL
##   ..$ face         : chr "bold"
##   ..$ colour       : NULL
```

```
##   ..$ size         : num 20
##   ..$ hjust        : NULL
##   ..$ vjust        : NULL
##   ..$ angle        : NULL
##   ..$ lineheight   : NULL
##   ..$ margin       : NULL
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi FALSE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ axis.text        :List of 11
##   ..$ family       : NULL
##   ..$ face         : NULL
##   ..$ colour       : NULL
##   ..$ size         : num 15
##   ..$ hjust        : NULL
##   ..$ vjust        : NULL
##   ..$ angle        : NULL
##   ..$ lineheight   : NULL
##   ..$ margin       : NULL
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi FALSE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ legend.background:List of 5
##   ..$ fill         : chr "white"
##   ..$ colour       : NULL
##   ..$ size         : num 0.5
##   ..$ linetype     : chr "solid"
##   ..$ inherit.blank: logi FALSE
##   ..- attr(*, "class")= chr [1:2] "element_rect" "element"
##  $ legend.text      :List of 11
##   ..$ family       : NULL
##   ..$ face         : NULL
##   ..$ colour       : NULL
##   ..$ size         : num 11
##   ..$ hjust        : NULL
##   ..$ vjust        : NULL
##   ..$ angle        : NULL
##   ..$ lineheight   : NULL
##   ..$ margin       : NULL
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi FALSE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ legend.title     :List of 11
##   ..$ family       : NULL
##   ..$ face         : chr "bold"
##   ..$ colour       : NULL
##   ..$ size         : num 12
##   ..$ hjust        : NULL
##   ..$ vjust        : NULL
##   ..$ angle        : NULL
##   ..$ lineheight   : NULL
##   ..$ margin       : NULL
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi FALSE
```

```
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ legend.position  : num [1:2] 0.43 0.87
##  $ strip.text.x     :List of 11
##   ..$ family       : NULL
##   ..$ face         : NULL
##   ..$ colour       : NULL
##   ..$ size         : num 13
##   ..$ hjust        : NULL
##   ..$ vjust        : NULL
##   ..$ angle        : NULL
##   ..$ lineheight   : NULL
##   ..$ margin       : NULL
##   ..$ debug        : NULL
##   ..$ inherit.blank: logi FALSE
##   ..- attr(*, "class")= chr [1:2] "element_text" "element"
##  $ colour           :List of 21
##   ..$ title        : list()
##   .. ..- attr(*, "class")= chr "waiver"
##   ..$ title.position: NULL
##   ..$ title.theme  : NULL
##   ..$ title.hjust  : NULL
##   ..$ title.vjust  : NULL
##   ..$ label        : logi TRUE
##   ..$ label.position: NULL
##   ..$ label.theme  : NULL
##   ..$ label.hjust  : NULL
##   ..$ label.vjust  : NULL
##   ..$ keywidth     : NULL
##   ..$ keyheight    : NULL
##   ..$ direction    : NULL
##   ..$ override.aes  :List of 1
##   .. ..$ alpha: num 0.55
##   ..$ nrow         : NULL
##   ..$ ncol         : NULL
##   ..$ byrow        : logi FALSE
##   ..$ reverse      : logi FALSE
##   ..$ order        : num 0
##   ..$ available_aes : chr "any"
##   ..$ name         : chr "legend"
##   ..- attr(*, "class")= chr [1:2] "guide" "legend"
##  - attr(*, "class")= chr [1:2] "theme" "gg"
##  - attr(*, "complete")= logi FALSE
##  - attr(*, "validate")= logi TRUE
```

```
  #mutate(environment=paste(poa,backness)) %>%
  #mutate(environment=recode(environment, "alveolar back"="su", "alveolar front" = "si", "postalveolar
 # filter(Word=='scissors' | Word=='sheep' | Word=='ship' | Word=='shell' | Word=='shoe' | Word=='shoul
```