everyday_CI

Meg Cychosz

11 January 2023

```
matches <- read.csv('dataframes/CI_TH_matches.csv') %>%
  select(-gender)

match_info <- matches %>% select(match, child_id)
```

```
# this code is shared for presentation purposes only
# the results are already constructed into csv files
# which are loaded in the following chunk

# get LENA measures
pre_its_df <- plyr::ldply( .data = list.files(pattern="*its_info.csv", # info about recording and child
                                              recursive=TRUE),
                           .fun = read.csv,colClasses=c("child_id"="character")) %>%
  select(-X, -DOB) %>%
  filter(child_id %in% matches$child_id) %>%
  mutate(endTimeSecs=case_when(child_id=='177RTP1' ~ "46214.05S",    # three participants paused their r
                               TRUE ~ "57599.99S")) %>%              # only one participant truly had a
                              mutate(corpus = substring(child_id, 4, 4)) # create a variable for corpus
                                                                     # note that endClockTime is wrong h
                                                                     # which is fine bc I'm not currentl

R <- pre_its_df %>%
  filter(corpus=='R' | corpus == 'J') %>% # timezone reported in GMT so we convert to EST and CST here
  mutate(startTimestamp = with_tz(ymd_hms(startClockTime, tz = "GMT"),"America/New_York"),
         endTimeStamp = with_tz(ymd_hms(endClockTime, tz = "GMT"),"America/New_York"))

its_df <- pre_its_df %>%
  filter(corpus!='R' & corpus !='J') %>%
  mutate(startTimestamp = with_tz(ymd_hms(startClockTime, tz = "GMT"),"America/Chicago"),
         endTimeStamp = with_tz(ymd_hms(endClockTime, tz = "GMT"),"America/Chicago")) %>%
  rbind(., R) %>%
    mutate(date = date(startTimestamp),
    startClockHours = hour(startTimestamp),
    startClockMinutes = minute(startTimestamp),
    startClockSeconds = second(startTimestamp),
    endClockHours = hour(endTimeStamp),
    endClockMinutes = minute(endTimeStamp),
    endClockSeconds = second(endTimeStamp)) %>%
    mutate(total_hrs=as.numeric(case_when(child_id=='177RTP1' ~ "12.83", # one participant < 16hr recor
                                TRUE ~ "16"))) %>%
    mutate(startClockTotalSeconds = ((startClockHours*60)*60)+(60*startClockMinutes)+startClockSeconds)

speech_df <- plyr::ldply( .data = list.files(pattern="*AN_timestamps.csv",
```

```
                                    recursive=TRUE, ignore.case = TRUE),
                    .fun = read.csv) %>%
  select(-X) %>%
  filter(duration<10) %>% # clips >10s are much more likely to be mislabeled
  mutate(corpus = substring(child_id, 4, 4)) %>%
  filter(child_id %in% matches$child_id) %>%
  mutate(hours = hour(seconds_to_period(seconds)),
         minutes = minute(seconds_to_period(seconds))) %>%
 merge(., its_df, by=c('corpus', 'child_id')) %>%
 mutate(avg_dB=avg_dB+97,
        peak_dB=peak_dB+97)

ctc_df <- plyr::ldply( .data = list.files(pattern="*CTC_timestamps.csv",
                                    recursive=TRUE, ignore.case = TRUE),
                    .fun = read.csv) %>%
  select(-X) %>%
  mutate(corpus = substring(child_id, 4, 4)) %>%
  filter(child_id %in% matches$child_id) %>%
  mutate(hours = hour(seconds_to_period(seconds)),
         minutes = minute(seconds_to_period(seconds))) %>%
 merge(., its_df, by=c('corpus', 'child_id'))

voc_df <- plyr::ldply( .data = list.files(pattern="*CHN_timestamps.csv",
                                      recursive=TRUE, ignore.case = TRUE),
                    .fun = read.csv,
                    colClasses=c("its_file_name"="character")) %>%
  select(-X) %>%
  mutate(corpus = substring(its_file_name, 4, 4)) %>%
  rename(child_id = its_file_name) %>%
  filter(child_id %in% matches$child_id) %>%
  mutate(hours = hour(seconds_to_period(seconds)),
         minutes = minute(seconds_to_period(seconds))) %>%
 merge(., its_df, by=c('corpus', 'child_id')) %>%
  mutate(secondsClock = startClockTotalSeconds+seconds,
         minutesClock = minute(seconds_to_period(secondsClock)),
         hoursClock = hour(seconds_to_period(secondsClock))) %>%
  mutate(avg_dB=avg_dB+97,
         peak_dB=peak_dB+97) %>% # scale intensity into something interpretable
  mutate(childUttLen=as.numeric(str_sub(childUttLen,2,-2)),
         childCryVfxLen=as.numeric(str_sub(childCryVfxLen,2,-2))) %>%
  filter(childUttLen!='0') %>% # 0 utt length indicates cries; remove them
  filter(childCryVfxLen=='0') # also remove the vocalizations that *contain* cries as we can't distingu
```

1 Read in data

```
# dataframe containing LENA data for the kids that are matched
its_df <- read.csv('dataframes/icphs_voc_its.csv') %>%
  merge(., matches, by=c('child_id')) %>% # merge with demo info
  select(-X.x, -X.y)

vocs <- read.csv('dataframes/icphs_voc_voc.csv') %>%
  merge(., matches, by=c('child_id')) %>%
```

```
  select(-X.x, -X.y)

speech <- read.csv('dataframes/icphs_speech.csv') %>%
  merge(., matches, by ='child_id') %>%
  select(-X.x, -X.y)

convo <- read.csv('dataframes/icphs_ctc.csv') %>%
  merge(., matches, by ='child_id') %>%
  select(-X.x, -X.y)
```

```
num_CI <- vocs %>%
  distinct_at(., vars(child_id, match)) %>%
  filter(match=='CI') %>%
  nrow()
print(paste('There should be 18 children w/ CIs and there are', num_CI))
```

```
## [1] "There should be 18 children w/ CIs and there are 18"
```

```
num_ha <- vocs %>%
  distinct_at(., vars(child_id, match)) %>%
  filter(match=='HA') %>%
  nrow()
print(paste('There should be 16 hearing age matches and there are', num_ha))
```

```
## [1] "There should be 16 hearing age matches and there are 16"
```

```
num_chrono <- vocs %>%
  distinct_at(., vars(child_id, match)) %>%
  filter(match=='chrono') %>%
  nrow()
print(paste('There should be 18 chronological age matches and there are', num_chrono))
```

```
## [1] "There should be 18 chronological age matches and there are 18"
```

```
num_CI_speech <- speech %>%
  distinct_at(., vars(child_id, match)) %>%
  filter(match=='CI') %>%
  nrow()
print(paste('There should be 18 children w/ CIs and there are', num_CI_speech))
```

```
## [1] "There should be 18 children w/ CIs and there are 18"
```

```
num_ha_speech <- speech %>%
  distinct_at(., vars(child_id, match)) %>%
  filter(match=='HA') %>%
  nrow()
print(paste('There should be 16 hearing age matches and there are', num_ha_speech))
```

```
## [1] "There should be 16 hearing age matches and there are 16"
```

```
num_chrono_speech <- speech %>%
  distinct_at(., vars(child_id, match)) %>%
  filter(match=='chrono') %>%
  nrow()
print(paste('There should be 18 chronological age matches and there are', num_chrono_speech))
```

```
## [1] "There should be 18 chronological age matches and there are 18"
```

```
# duration of segments
adult_dur <- speech %>%
  group_by(segment_type) %>%
  summarize(total_dur = sum(duration)) %>%
  summarize(total_dur_hour = (total_dur/60)/60)

voc_dur <- vocs %>%
  summarize(total_dur = sum(duration)) %>%
  summarize(total_dur_hour = (total_dur/60)/60)

convo_dur <- convo %>%
  summarize(total_dur = sum(duration)) %>%
  summarize(total_dur_hour = (total_dur/60)/60)

# counts of segments
voc_cts <- vocs %>%
  nrow()
```

2 Demo info

```
ci_device <- matches %>%
  filter(match=='CI') %>%
  count(device_config)
```

```
gender <- its_df %>%
  group_by(match) %>%
  count(gender) %>%
  spread(gender, n) %>%
  mutate(Gender=paste0(`FALSE`,',',M)) %>%
  select(match, Gender) %>%
  spread(key='match', value='Gender') %>%
  mutate(measure='Gender (F,M)')

demo <- its_df %>%
  group_by(match) %>%
  summarize(chrono_age = mean(age_mos),
            chrono_age_sd = sd(age_mos),
            chrono_age_min = min(age_mos),
            chrono_age_max = max(age_mos),
            mat_ed = mean(Maternal_education_level),
            mat_ed_sd = sd(Maternal_education_level),
            mat_ed_min = min(Maternal_education_level),
            mat_ed_max = max(Maternal_education_level)) %>%
```

```
    mutate_if(is.numeric, round, 2) %>%
  mutate(Chrono_age=paste0(chrono_age, '(', chrono_age_sd, ')', ',',chrono_age_min, '-', chrono_age_max
         Mat_ed=paste0(mat_ed,'(',mat_ed_sd,')',',',mat_ed_min,'-',mat_ed_max))

mat_ed <- demo %>%
  select(match, Mat_ed) %>%
  spread(match, Mat_ed) %>%
  mutate(measure='Maternal Education')

ci_demo <- its_df %>%
  filter(match=='CI') %>%
  summarize(ha_mean=mean(hearing_age),
            ha_sd=sd(hearing_age),
            ha_min=min(hearing_age),
            ha_max=max(hearing_age),
            implant_mean=mean(age_of_implantation),
            implant_sd=sd(age_of_implantation),
            implant_min=min(age_of_implantation),
            implant_max=max(age_of_implantation)) %>%
  mutate_if(is.numeric, round, 2) %>%
  mutate(`Hearing Age (mos)`=paste0(ha_mean,'(',ha_sd,')',',',ha_min,'-',ha_max),
         `Implant Age (mos)`=paste0(implant_mean,'(',implant_sd,')',',',implant_min,'-',implant_max)) %>
  select(`Hearing Age (mos)`,`Implant Age (mos)`) %>%
  gather(key='measure',value='stat',`Hearing Age (mos)`,`Implant Age (mos)`) %>%
  rename(CI=stat) %>%
  mutate(chrono='NA',
         HA='NA')


demo_tbl <- demo %>%
  select(match, Chrono_age) %>%
  spread(match, Chrono_age) %>%
  mutate(measure='Chrono. Age (mos)') %>%
  rbind(., gender) %>%
  rbind(., mat_ed) %>%
  rbind(., ci_demo) %>%
  select(measure, everything())

kable(demo_tbl, booktabs=T,
              caption= "Demographic and audiological information. Mean (SD), range.",
             row.names = FALSE,
       col.names = c(" ",
                     "Chrono. age matches",
                     "Cochlear implant",
                     "Hearing age matches")) %>%
  kable_styling() %>%
  kableExtra::kable_styling(latex_options = "hold_position")
```

Table 1: (#tab:make the demo info table)Demographic and audiological information. Mean (SD), range.

	Chrono. age matches	Cochlear implant	Hearing age matches
Chrono. Age (mos)	46.28(10.8),32-66	47.72(9.84),31-65	35(12.71),17-52
Gender (F,M)	9,9	9,9	9,7
Maternal Education	6.22(1),3-7	6.11(1.02),3-7	6.25(1),3-7
Hearing Age (mos)	NA	31.28(14.3),8-54	NA
Implant Age (mos)	NA	16.44(9.7),7-45	NA

3 Vocalization analyses

3.1 Compute vocalizations

```
# summary statistics for each child
recording_vocs <- vocs %>%
  group_by(child_id) %>%
  summarize(normed_vocs = sum(childUttCnt)/total_hrs,
            avg_dur = mean(childUttLen)*1000,
            sd_dur = sd(childUttLen)*1000,
            mean_dB = mean(avg_dB),
            sd_dB =  sd(avg_dB)) %>%
  distinct(child_id, .keep_all = T) %>%
  merge(.,matches, by='child_id')

# the num of vocalizations for each child, for each hour of the day
hourly_vocs <- vocs %>%
  group_by(match, implanted, age_of_implantation, child_id, hours) %>%
  summarize(normed_hourly_vocs = sum(childUttCnt))

# summary statistics for each match
prep_voc_tbl <- vocs %>%
  group_by(match) %>%
  summarize(mean_dur = mean(childUttLen)*1000,
            sd_dur = sd(childUttLen)*1000,
            min_dur=min(childUttLen)*1000,
            max_dur=max(childUttLen)*1000,
            mean_dB = mean(avg_dB),
            sd_dB = sd(avg_dB),
            min_dB=min(avg_dB),
            max_dB=max(avg_dB)) %>%
  mutate_if(is.numeric, round, 2) %>%
    mutate(duration=paste0(mean_dur,'(',sd_dur,')',',',min_dur,'-',max_dur),
           intensity=paste0(mean_dB,'(',sd_dB,')',',',min_dB,'-',max_dB))

dur <- prep_voc_tbl %>% select(match,duration) %>% spread(match,duration)
intensity <- prep_voc_tbl %>% select(match,intensity) %>% spread(match,intensity)

voc_quantity <- vocs %>%
  group_by(match, child_id) %>%
  summarize(normed_vocs = sum(childUttCnt)/total_hrs) %>%
  ungroup() %>%
```

```
  group_by(match) %>%
  summarize(avg_normed_vocs = mean(normed_vocs),
            sd_normed_vocs = sd(normed_vocs),
            min_normed_vocs =min(normed_vocs),
            max_normed_vocs=max(normed_vocs)) %>%
  mutate_if(is.numeric, round, 2) %>%
  ungroup() %>%
  mutate(quantity=paste0(avg_normed_vocs,'(',
                         sd_normed_vocs,')',',',
                         min_normed_vocs,'-',
                         max_normed_vocs)) %>%
  select(match,quantity) %>%
  spread(match, quantity)

voc_tbl <- intensity %>%
  rbind(., voc_quantity) %>%
  rbind(., dur) %>%
  rownames_to_column(.,var = 'measure') %>%
  mutate(measure = case_when(measure=='1'~'intensity',
            measure=='2'~'num_vocs_hr',
            TRUE~'voc_dur'))
```

```
# compute the percentage of minutes in the child's day with >1 vocalization

time_steps <- rep(seq(60,57600,60),times=52) %>% as.data.frame()
time_steps$seconds <- time_steps$.
ids <- vocs %>% distinct(child_id)
ids_repeat <- rep(ids$child_id,960) %>% as.data.frame()
ids_repeat$child_id <- ids_repeat$.
time_steps_demo <- ids_repeat %>%
  arrange(child_id) %>%
  select(-.) %>%
  cbind(., time_steps) %>%
  select(child_id, seconds)

match_info <- matches %>% select(match, child_id)

pre_voc_consis <- vocs %>%
  select(child_id, seconds, duration, childUttCnt, childUttLen) %>%
  merge(., time_steps_demo, by=c('seconds', 'child_id'),all=TRUE) %>% # impute the missing seconds
  replace_na(list(duration = 0, childUttCnt=0, childUttLen=0)) %>% # replace the imputed time stamps wi
  merge(., match_info, by='child_id') # remerge to get complete df of addtl measures w/o na's


voc_consis <- pre_voc_consis %>%
  group_by(match, child_id, seconds) %>%
  summarize(vocalizations = sum(childUttCnt)) %>%
  ungroup() %>%
  mutate(contains_vocs = if_else((vocalizations > 0), 'TRUE', 'FALSE')) %>% # boolean if it contains vo
  ungroup() %>%
  group_by(child_id, contains_vocs) %>%
  tally() %>%
  mutate(perc_vocs = if_else(child_id=='177RTP1', n/770, n/960)) %>% #769.8 minutes in 12.83 hr recordi
```

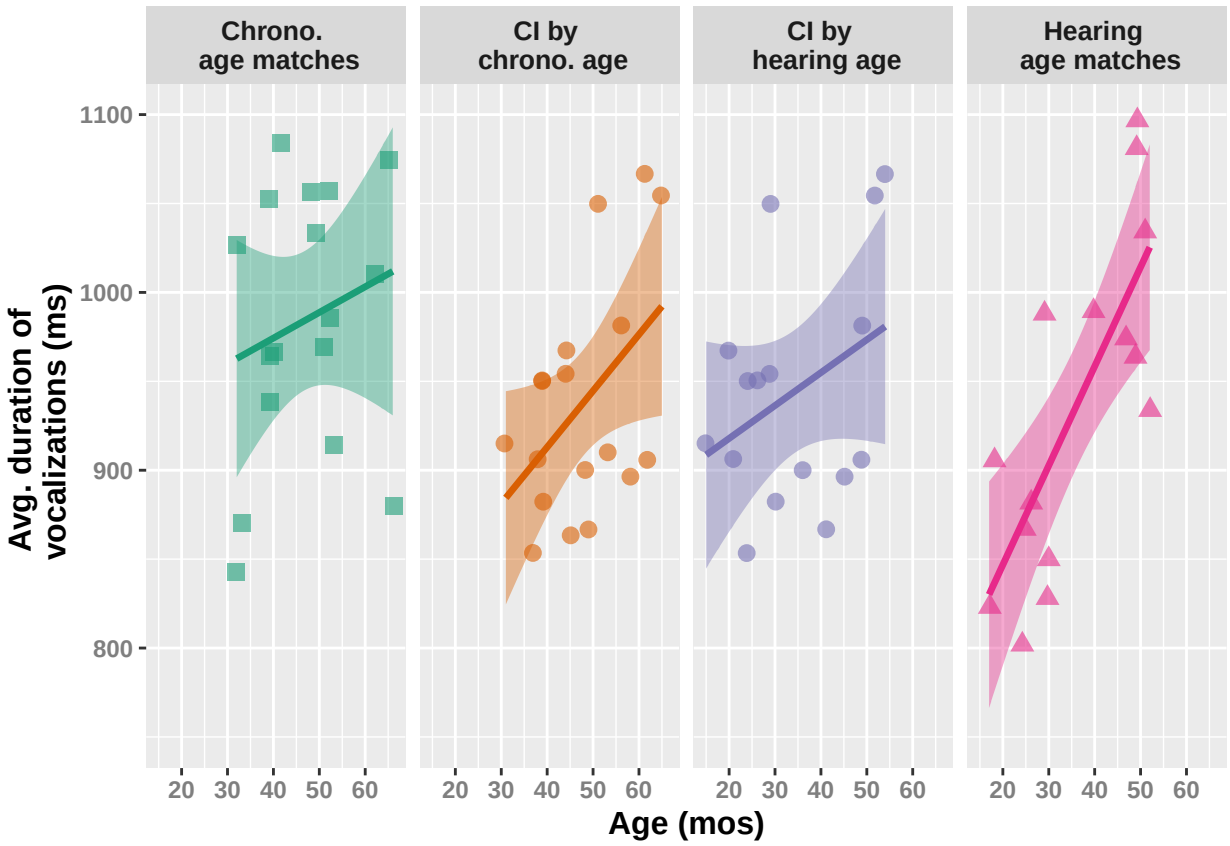
```
  filter(contains_vocs=='TRUE') %>%
  merge(., match_info, by='child_id')

voc_consis_tbl <- voc_consis %>%
  group_by(match) %>%
  summarize(mean_perc_vocs = mean(perc_vocs),
            sd_perc_vocs = sd(perc_vocs),
            min_perc_vocs = min(perc_vocs),
            max_perc_vocs = max(perc_vocs)) %>%
  mutate_if(is.numeric, round, 2) %>%
  mutate(voc_consistency = paste0(mean_perc_vocs,"(",sd_perc_vocs,")",",",min_perc_vocs,"-",max_perc_vo
  select(match, voc_consistency) %>%
  spread(match, voc_consistency) %>%
  rownames_to_column(.,var = 'measure')
```

```
# create a 4th "match" of CI kids to compute hearing age
ha_kids <- its_df %>%
  select(child_id, age_mos) %>%
  merge(., recording_vocs, by='child_id') %>%
  filter(match=='CI') %>%
  select(-age_mos, -match) %>%
  mutate(age_mos = hearing_age,
         match = 'CI_by_hearing_age') %>%
  filter(!hearing_age <= 12)  # remove the two kids who weren't matched by hearing age


voc_growth_tbl <- its_df %>%
  select(child_id, age_mos) %>%
  merge(., recording_vocs, by='child_id') %>%
  rbind(., ha_kids) %>%
  group_by(match) %>%
  do(voc_growth = lm(normed_vocs~age_mos, data=.),
     mod2 = cor(.$normed_vocs, .$age_mos, method = "pearson")) %>%
  mutate(slope = summary(voc_growth)$coeff[2],
         p_value = summary(voc_growth)$coeff[8],
         Pearson = mod2[1]) %>%
  select(match, slope, p_value, Pearson) %>%
  mutate_if(is.numeric, round, 2) %>%
  mutate(stats=paste0('B=',slope,",","p=",p_value, ",","r=", Pearson)) %>%
  select(-slope, -p_value, -Pearson) %>%
  spread(match, stats) %>%
  mutate(measure='Child voc. quantity growth')

voc_dur_growth_tbl <- its_df %>%
  select(child_id, age_mos) %>%
  merge(., recording_vocs, by='child_id') %>%
  rbind(., ha_kids) %>%
  group_by(match) %>%
  do(voc_growth = lm(avg_dur*1000~age_mos, data=.),
     mod2 = cor(.$avg_dur*1000, .$age_mos, method = "pearson")) %>%
  mutate(slope = summary(voc_growth)$coeff[2],
         p_value = summary(voc_growth)$coeff[8],
         Pearson = mod2[1]) %>%
```

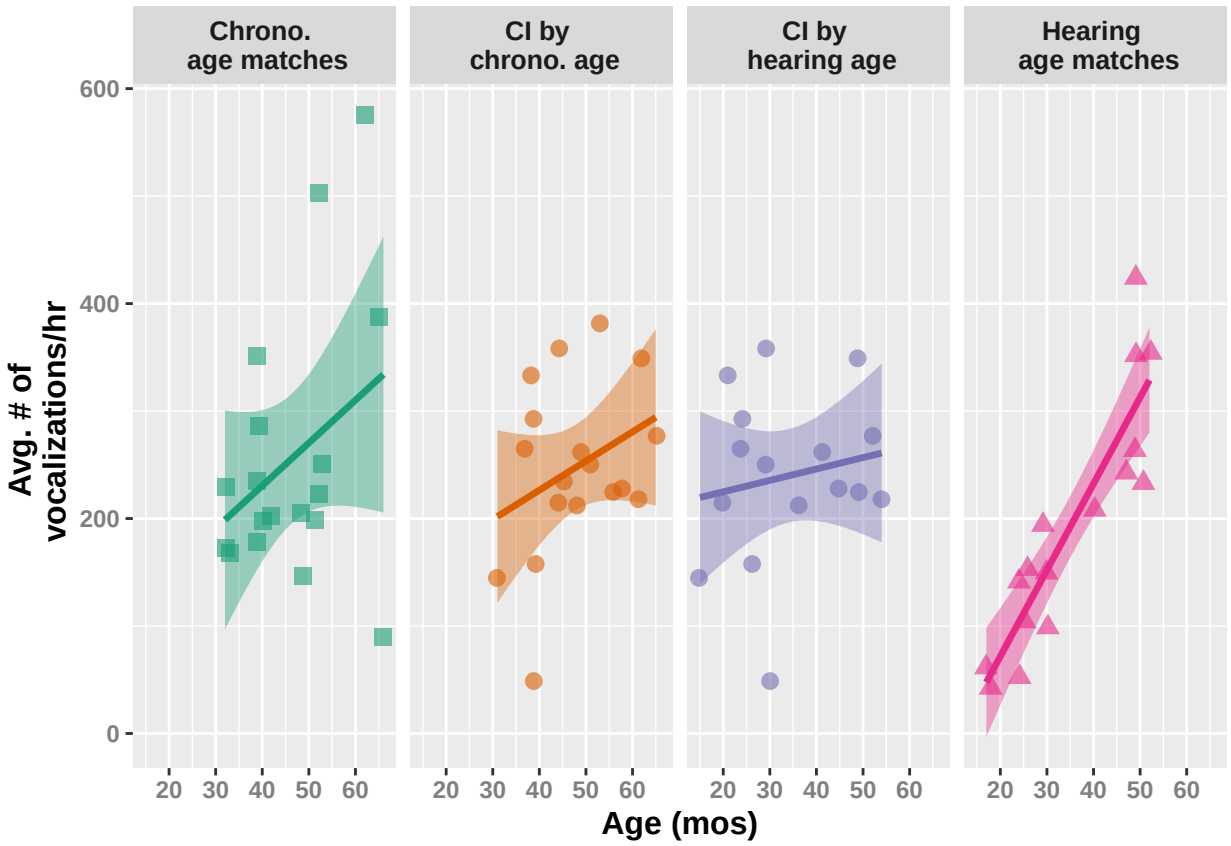
```
  select(match, slope, p_value, Pearson) %>%
  mutate_if(is.numeric, round, 2) %>%
  mutate(stats=paste0('B=',slope,",","p=",p_value, ",","r=", Pearson)) %>%
  select(-slope, -p_value, -Pearson) %>%
  spread(match, stats) %>%
  mutate(measure='Child voc. duration growth (ms)')
```

3.2 Visualize vocalizations

```
its_df %>%
  select(child_id, age_mos) %>%
  merge(., recording_vocs, by='child_id') %>%
  rbind(., ha_kids) %>%
mutate(match=recode(match,
                    chrono='Chrono. \n age matches',
                    CI='CI by \n chrono. age',
                    CI_by_hearing_age='CI by \n hearing age',
                    HA='Hearing \n age matches')) %>%
ggplot(., aes(age_mos, avg_dur)) +
  geom_jitter(aes(color=match, fill=match, shape=match),size=2.8,alpha=.6,width = .3) +
  geom_smooth(aes(fill=match, color=match), method = "lm",size=1.2) +
  facet_grid(~match) +
  scale_y_continuous(limits = c(750, 1100)) +
  scale_color_brewer(palette="Dark2") +
  scale_fill_brewer(palette="Dark2") +
  scale_shape_manual(values=c(15,16,16,17)) +
  xlab("Age (mos)") +
  ylab("Avg. duration of \n vocalizations (ms)") +
  theme(axis.title = element_text(face ="bold", size=12),
        legend.position = "none",
        axis.text = element_text(face="bold", color='gray50', size=9),
        strip.text=element_text(face='bold', size=10))
```




```
its_df %>%
  select(child_id, age_mos) %>%
  merge(., recording_vocs, by='child_id') %>%
  rbind(., ha_kids) %>%
  mutate(match=recode(match,
                      chrono='Chrono. \n age matches',
                      CI='CI by \n chrono. age',
                      CI_by_hearing_age='CI by \n hearing age',
                      HA='Hearing \n age matches')) %>%
ggplot(., aes(age_mos, normed_vocs)) +
  geom_jitter(aes(color=match, fill=match, shape=match),size=2.8,alpha=.6,width = .3) +
  geom_smooth(aes(fill=match, color=match), method = "lm",size=1.2) +
  scale_shape_manual(values=c(15,16,16,17)) +
  facet_grid(~match) +
  scale_color_brewer(palette="Dark2") +
  scale_fill_brewer(palette="Dark2") +
  xlab("Age (mos)") +
  ylab("Avg. # of \n vocalizations/hr") +
    theme(axis.title = element_text(face ="bold", size=12),
        legend.position = "none",
        axis.text = element_text(face="bold", color='gray50', size=9),
        strip.text=element_text(face='bold', size=10))
```



3.3 Model vocalizations

```
# ---------- QUANTITY ------------
# repeated measures
# intensity
vocs$match <- relevel(factor(vocs$match), ref = "CI")
voc_intensity_m0 <- lmer(avg_dB~ + (1 | child_id), data=vocs)
voc_intensity_m1 <- lmer(avg_dB~match + (1 | child_id), data=vocs)
anova(voc_intensity_m0,voc_intensity_m1)
```

```
## Data: vocs
## Models:
## voc_intensity_m0: avg_dB ~ +(1 | child_id)
## voc_intensity_m1: avg_dB ~ match + (1 | child_id)
##                  npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## voc_intensity_m0    3 963736 963766 -481865   963730
## voc_intensity_m1    5 963739 963789 -481865   963729 0.6216  2     0.7329
```

```
# duration
hourly_vocLen_m0 <- lmer(childUttLen*1000~ + (1 | child_id), data=vocs)
hourly_vocLen_m1 <- lmer(childUttLen*1000~ match+ (1 | child_id), data=vocs)
anova(hourly_vocLen_m0,hourly_vocLen_m1)
```

```
## Data: vocs
## Models:
```

```
## hourly_vocLen_m0: childUttLen * 1000 ~ +(1 | child_id)
## hourly_vocLen_m1: childUttLen * 1000 ~ match + (1 | child_id)
##                  npar     AIC     BIC   logLik deviance  Chisq Df Pr(>Chisq)
## hourly_vocLen_m0    3 2620905 2620935 -1310450  2620899
## hourly_vocLen_m1    5 2620902 2620952 -1310446  2620892 6.9481  2    0.03099 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summary(hourly_vocLen_m1)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: childUttLen * 1000 ~ match + (1 | child_id)
##    Data: vocs
##
## REML criterion at convergence: 2620868
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.8378 -0.5909 -0.2715  0.2922 30.3097
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  child_id (Intercept)   8178    90.43
##  Residual             377986   614.81
## Number of obs: 167130, groups:  child_id, 52
##
## Fixed effects:
##             Estimate Std. Error     df t value Pr(>|t|)
## (Intercept)   937.56      21.49  48.25   43.64   <2e-16 ***
## matchchrono    59.25      30.38  48.24    1.95    0.057 .
## matchHA       -19.46      31.40  48.74   -0.62    0.538
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) mtchch
## matchchrono -0.707
## matchHA     -0.684  0.484
```

```
# hourly measures
hourly_vocs$match <- relevel(factor(hourly_vocs$match), ref = "CI")
hourly_vocs$hours <- as.factor(hourly_vocs$hours)
hourly_vocs_m0 <- lmer(normed_hourly_vocs~ + (1 | child_id) + (1|hours), data=hourly_vocs)
hourly_vocs_m1 <- lmer(normed_hourly_vocs~ match+ (1 | child_id) + (1|hours), data=hourly_vocs)
anova(hourly_vocs_m0,hourly_vocs_m1)
```

```
## Data: hourly_vocs
## Models:
## hourly_vocs_m0: normed_hourly_vocs ~ +(1 | child_id) + (1 | hours)
## hourly_vocs_m1: normed_hourly_vocs ~ match + (1 | child_id) + (1 | hours)
##                 npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## hourly_vocs_m0    4 8970.1 8988.1 -4481.0   8962.1
## hourly_vocs_m1    6 8971.8 8998.9 -4479.9   8959.8 2.2808  2     0.3197
```

```
# ---------- CONSISTENCY ------------
voc_consis2 <- its_df %>% select(age_mos, child_id) %>% merge(., voc_consis, by='child_id')
voc_consis2$match <- relevel(factor(voc_consis2$match), ref = "CI")
voc_consis_m0 <- lm(perc_vocs~age_mos, data=voc_consis2)
voc_consis_m1 <- lm(perc_vocs~ age_mos + match, data=voc_consis2)
anova(voc_consis_m0,voc_consis_m1) # no
```

```
## Analysis of Variance Table
##
## Model 1: perc_vocs ~ age_mos
## Model 2: perc_vocs ~ age_mos + match
##   Res.Df     RSS Df Sum of Sq      F Pr(>F)
## 1     50 0.79041
## 2     48 0.78237  2 0.0080337 0.2464 0.7826
```

```
voc_consis_m2 <- lm(perc_vocs~ age_mos*match, data=voc_consis2)
anova(voc_consis_m0,voc_consis_m2) # no
```

```
## Analysis of Variance Table
##
## Model 1: perc_vocs ~ age_mos
## Model 2: perc_vocs ~ age_mos * match
##   Res.Df     RSS Df Sum of Sq      F Pr(>F)
## 1     50 0.79041
## 2     46 0.71175  4  0.078656 1.2709 0.2951
```

4 Input analyses

4.1 Compute input statistics and make tables

```
# summary statistics for each child
recording_speech <- speech %>%
  group_by(match, child_id) %>%
  summarize(normed_words = sum(wordCount)/total_hrs, # avg. number of words/hr
            normed_speech = sum(duration)/total_hrs) %>%  # avg. # of seconds of speech/hr
  distinct(child_id, .keep_all = T)

prep_speech_quantity_tbl <- recording_speech %>%
  group_by(match) %>%
  summarize(mean_normed_words = mean(normed_words), # avg. number of words/hr
            sd_normed_words = sd(normed_words),
            min_normed_words = min(normed_words),
            max_normed_words = max(normed_words),
            mean_normed_speech = mean(normed_speech),  # avg. # of seconds of speech/hr
            sd_normed_speech = sd(normed_speech),
            min_normed_speech = min(normed_speech),
            max_normed_speech = max(normed_speech)) %>%
  mutate_if(is.numeric, round, 2)

intensity_stat <- speech %>%
  group_by(match) %>%
  summarize(mean_dB = mean(avg_dB), # group-level average
```

```
            sd_dB = sd(avg_dB), # group-level variance
            min_dB = min(avg_dB),
            max_dB = max(avg_dB)) %>%
  mutate_if(is.numeric, round, 2)

word_stat <- prep_speech_quantity_tbl %>%
  mutate(word_quantity = paste0(mean_normed_words,"(",
                                sd_normed_words,")",",",
                                min_normed_words,"-",
                                max_normed_words)) %>%
  select(match, word_quantity) %>%
  spread(match, word_quantity)

speech_stat <- prep_speech_quantity_tbl %>%
  mutate(input_quantity = paste0(mean_normed_speech,"(",
                                 sd_normed_speech,")",",",
                                 min_normed_speech,"-",
                                 max_normed_speech)) %>%
  select(match, input_quantity) %>%
  spread(match, input_quantity)

speech_quantity_tbl <- intensity_stat %>%
  mutate(input_intensity = paste0(mean_dB,"(",sd_dB,")",",",min_dB,"-",max_dB)) %>%
  select(match, input_intensity) %>%
  spread(match, input_intensity) %>%
  rbind(., word_stat) %>%
  rbind(., speech_stat) %>%
  rownames_to_column(.,var = 'measure') %>%
  mutate(measure = case_when(measure=='1'~'intensity',
            measure=='2'~'num_words_hr',
            TRUE~'mins_speech_hr'))


# the num of words and amount of speech from adults for each child, for each hour of the day
# hourly speech refers to the avg. num of seconds of speech input each hour
hourly_speech <- speech %>%
  group_by(match, implanted, child_id, age_mos, hours) %>%
  summarize(normed_hourly_words = sum(wordCount),
            normed_hourly_speech = sum(duration)) %>%
  filter(normed_hourly_speech>3) # remove all hours with less than 3 seconds of speech

# now choose the highest vocal activity hour
# for each child
high_word_hour <- hourly_speech %>%
  group_by(child_id) %>%
  arrange(desc(normed_hourly_words)) %>%
  slice(n=1) %>%
  select(-normed_hourly_speech, -hours)

high_hour <- hourly_speech %>%
  group_by(child_id) %>%
  arrange(desc(normed_hourly_speech)) %>%
  slice(n=1) %>%
```

```
  select(-normed_hourly_words, -hours) %>%
  merge(., high_word_hour, by=c('child_id', 'match'))

# summary statistics for each match
all_speech_quantity_tbl <- speech %>%
  group_by(match, child_id) %>%
  summarize(normed_words = sum(wordCount)/total_hrs,
            normed_speech = sum(duration)/total_hrs) %>%
  ungroup() %>%
  merge(., high_hour, by=c('child_id', 'match')) %>% # with info about the measures from the highest vo
  group_by(match) %>%
  summarize(avg_highhour_words = mean(normed_hourly_words),
            sd_highhour_words = sd(normed_hourly_words),
            avg_highhour_speech = mean(normed_hourly_speech),
            sd_highhour_speech = mean(normed_hourly_speech),
            avg_normed_words = mean(normed_words),
            sd_normed_words = sd(normed_words),
            avg_normed_speech = mean(normed_speech),
            sd_normed_speech = sd(normed_speech))

kable(all_speech_quantity_tbl, booktabs=T,
              caption= "Speech input statistics, by hearing group",
             row.names = FALSE) %>%
  kable_styling() %>%
  kableExtra::kable_styling(latex_options = "hold_position")
```

Table 2: (#tab:compute input stats)Speech input statistics, by hearing group

match	avg_highhour_words	sd_highhour_words	avg_highhour_speech	sd_highhour_speech	avg_normed_wo
chrono	3567.413	850.8975	877.0062	877.0062	1286.
CI	3617.534	1227.9349	875.6365	875.6365	1488.
HA	3531.419	1042.9615	893.0836	893.0836	1371.

```
# compute the percentage of minutes in the child's day with > 1 AW

time_steps <- rep(seq(60,57600,60),times=52) %>% as.data.frame()
time_steps$seconds <- time_steps$.
ids <- speech %>% distinct(child_id)
ids_repeat <- rep(ids$child_id,960) %>% as.data.frame()
ids_repeat$child_id <- ids_repeat$.
time_steps_demo <- ids_repeat %>%
  arrange(child_id) %>%
  select(-.) %>%
  cbind(., time_steps) %>%
  select(child_id, seconds)

match_info <- matches %>% select(match, child_id)

pre_input_consis <- speech %>%
  select(child_id, seconds, duration, wordCount, clip_onset) %>%
```

```
  merge(., time_steps_demo, by=c('seconds', 'child_id'),all=TRUE) %>% # impute the missing seconds
  replace_na(list(duration = 0, wordCount=0)) %>% # replace the imputed time stamps with 0 adult words
  merge(., match_info, by='child_id') # remerge to get complete df of addtl measures w/o na's

input_consis <- pre_input_consis %>%
  group_by(match, child_id, seconds) %>%
  summarize(adult_words = sum(wordCount)) %>%
  ungroup() %>%
  mutate(contains_words = if_else((adult_words > 0), 'TRUE', 'FALSE')) %>% # boolean if it contains wor
  ungroup() %>%
  group_by(child_id, contains_words) %>%
  tally() %>%
  mutate(perc_words = if_else(child_id=='177RTP1', n/770, n/960)) %>% #769.8 minutes in 12.83 hr record
  filter(contains_words=='TRUE') %>%
  merge(., match_info, by='child_id')

speech_consis_tbl <- input_consis %>%
  group_by(match) %>%
  summarize(mean_perc_words = mean(perc_words),
            sd_perc_words = sd(perc_words),
            min_perc_words = min(perc_words),
            max_perc_words = max(perc_words)) %>%
  mutate_if(is.numeric, round, 2) %>%
  mutate(input_consistency = paste0(mean_perc_words,"(",sd_perc_words,")",",",min_perc_words,"-",max_pe
  select(match, input_consistency) %>%
  spread(match, input_consistency) %>%
  rownames_to_column(.,var = 'measure')
```

```
# create a 4th "match" of CI kids to compute hearing age
ha <- matches %>% select(child_id, hearing_age)
ha_speech <- its_df %>%
  select(child_id, age_mos) %>%
  merge(., recording_speech, by=c('child_id')) %>%
  merge(., ha, by='child_id') %>%
  filter(match=='CI') %>%
  select(-age_mos, -match) %>%
  mutate(age_mos = hearing_age,
         match = 'CI_by_hearing_age') %>%
  filter(!hearing_age<=12) %>%
  select(-hearing_age)


input_growth_tbl <- its_df %>%
  select(child_id, age_mos) %>%
  merge(., recording_speech, by='child_id') %>%
  rbind(., ha_speech) %>%
  group_by(match) %>%
  do(speech_growth = lm(normed_words~age_mos, data=.),
     mod2 = cor(.$normed_words, .$age_mos, method = "pearson")) %>%
  mutate(slope = summary(speech_growth)$coeff[2],
         p_value = summary(speech_growth)$coeff[8],
         Pearson = mod2[1]) %>%
  select(match, slope, p_value, Pearson) %>%
```

```
  mutate_if(is.numeric, round, 2) %>%
  mutate(stats=paste0('B=',slope,",","p=",p_value, ",","r=", Pearson)) %>%
  select(-slope, -p_value, -Pearson) %>%
  spread(match, stats) %>%
  mutate(measure='Adult word growth')
```

4.2 Model input

```
# ---------- QUANTITY ------------
# repeated measures
# intensity
speech$match <- relevel(factor(speech$match), ref = "CI")
speech_intensity_m0 <- lmer(avg_dB~ + (1 | child_id), data=speech)
speech_intensity_m1 <- lmer(avg_dB~match + (1 | child_id), data=speech)
anova(speech_intensity_m0,speech_intensity_m1)
```

```
## Data: speech
## Models:
## speech_intensity_m0: avg_dB ~ +(1 | child_id)
## speech_intensity_m1: avg_dB ~ match + (1 | child_id)
##                     npar     AIC     BIC  logLik deviance  Chisq Df Pr(>Chisq)
## speech_intensity_m0    3 1121616 1121646 -560805  1121610
## speech_intensity_m1    5 1121618 1121669 -560804  1121608 1.4138  2     0.4932
```

```
speech_intensity <- lmer(avg_dB~match + (1 | child_id), data=speech)

# hourly measures
# minutes
hourly_speech$match <- relevel(factor(hourly_speech$match), ref = "CI")
hourly_speech$hours <- as.factor(hourly_speech$hours)
hourly_speech_m0 <- lmer(normed_hourly_speech~ + (1 | child_id) + (1|hours), data=hourly_speech)
hourly_speech_m1 <- lmer(normed_hourly_speech~ match+ (1 | child_id) + (1|hours), data=hourly_speech)
anova(hourly_speech_m0,hourly_speech_m1)
```

```
## Data: hourly_speech
## Models:
## hourly_speech_m0: normed_hourly_speech ~ +(1 | child_id) + (1 | hours)
## hourly_speech_m1: normed_hourly_speech ~ match + (1 | child_id) + (1 | hours)
##                   npar    AIC    BIC  logLik deviance Chisq Df Pr(>Chisq)
## hourly_speech_m0    4 9351.2 9369.3 -4671.6   9343.2
## hourly_speech_m1    6 9353.6 9380.7 -4670.8   9341.6 1.624  2      0.444
```

```
hourly_speech_m2 <- lmer(normed_hourly_speech~ age_mos+ (1 | child_id) + (1|hours), data=hourly_speech)

# words
hourly_mins_m0 <- lmer(normed_hourly_words~ + (1 | child_id) + (1|hours), data=hourly_speech)
hourly_mins_m1 <- lmer(normed_hourly_words~ match+ (1 | child_id) + (1|hours), data=hourly_speech)
anova(hourly_mins_m0,hourly_mins_m1)
```

```
## Data: hourly_speech
## Models:
## hourly_mins_m0: normed_hourly_words ~ +(1 | child_id) + (1 | hours)
```

```
## hourly_mins_m1: normed_hourly_words ~ match + (1 | child_id) + (1 | hours)
##                npar   AIC   BIC  logLik deviance  Chisq Df Pr(>Chisq)
## hourly_mins_m0    4 11250 11268 -5621.0    11242
## hourly_mins_m1    6 11253 11280 -5620.3    11241 1.4309  2      0.489
```

```
# does the overall amount of speech change as children age?
hourly_mins_m2 <- lmer(normed_hourly_words~ age_mos+ (1 | child_id) + (1|hours), data=hourly_speech)



# ---------- CONSISTENCY ------------
input_consis2 <- its_df %>% select(age_mos, child_id) %>% merge(., input_consis, by='child_id')
input_consis2$match <- relevel(factor(input_consis2$match), ref = "CI")
speech_consis_m0 <- lm(perc_words~age_mos, data=input_consis2)
speech_consis_m1 <- lm(perc_words~ age_mos + match, data=input_consis2)
anova(speech_consis_m0,speech_consis_m1) # no
```

```
## Analysis of Variance Table
##
## Model 1: perc_words ~ age_mos
## Model 2: perc_words ~ age_mos + match
##   Res.Df     RSS Df Sum of Sq      F Pr(>F)
## 1     50 0.57049
## 2     48 0.54406  2  0.026437 1.1662 0.3202
```

```
speech_consis_m2 <- lm(perc_words~ age_mos*match, data=input_consis2)
anova(speech_consis_m0,speech_consis_m2) # no
```

```
## Analysis of Variance Table
##
## Model 1: perc_words ~ age_mos
## Model 2: perc_words ~ age_mos * match
##   Res.Df     RSS Df Sum of Sq      F Pr(>F)
## 1     50 0.57049
## 2     46 0.51661  4  0.053881 1.1994 0.3239
```

```
# minutes

# words
```

5 Convo turn analyses

5.1 Compute convo turns

```
# summary statistics for each child
recording_convo <- convo %>%
  group_by(child_id) %>%
  summarize(normed_turns = sum(convo_count)/total_hrs,
            avg_dur = mean(convo_count),
            sd_dur = sd(convo_count)) %>%
  distinct(child_id, .keep_all = T)
```

```
# the num of vocalizations for each child, for each hour of the day
hourly_turns <- convo %>%
  group_by(match, implanted, child_id, hours) %>%
  summarize(normed_hourly_turns = sum(convo_count))

# summary statistics for each match
turn_quantity_tbl <- convo %>%
  group_by(match, child_id) %>%
  summarize(normed_turns = sum(convo_count)/total_hrs) %>%
  ungroup() %>%
  group_by(match) %>%
  summarize(avg_normed_turns = mean(normed_turns),
            sd_normed_turns = sd(normed_turns),
            min_normed_turns = min(normed_turns),
            max_normed_turns = max(normed_turns)) %>%
  mutate_if(is.numeric, round, 2) %>%
  mutate(turn_quantity = paste0(avg_normed_turns,"(",sd_normed_turns,")",",",min_normed_turns,"-",max_n
  select(match, turn_quantity) %>%
  spread(match, turn_quantity) %>%
  rownames_to_column(.,var = 'measure')
```

```
# compute the percentage of epochs (5-min chunks) in the child's day with > 1 CT
time_steps <- rep(seq(1,192,1),times=52) %>% as.data.frame()
time_steps$epochs <- time_steps$.
ids <- convo %>% distinct(child_id)
ids_repeat <- rep(ids$child_id,192) %>% as.data.frame()
ids_repeat$child_id <- ids_repeat$.
time_steps_demo <- ids_repeat %>%
  arrange(child_id) %>%
  select(-.) %>%
  cbind(., time_steps) %>%
  select(child_id, epochs)

pre_convo_consis <- convo %>%
  mutate(epochs=floor(seconds/300)) %>% #round down to the nearest integer
  select(child_id, epochs, seconds, convo_count, clip_onset) %>%
  merge(., time_steps_demo, by=c('epochs', 'child_id'),all=TRUE) %>% # impute the missing epochs
  replace_na(list(convo_count=0)) %>%
  merge(., match_info, by='child_id')# remerge to get complete df of addtl measures w/o na's

convo_consis_check <- pre_convo_consis %>%
  group_by(match, child_id, epochs) %>%
  summarize(turns = sum(convo_count)) %>%
  ungroup() %>%
  mutate(contains_turns = if_else((turns > 0), 'TRUE', 'FALSE')) %>% # boolean if it contains turns
  ungroup() %>%
  group_by(child_id, contains_turns) %>%
  tally() %>%
  mutate(perc_turns = if_else(child_id=='177RTP1', n/154, n/192)) %>% #153.96 epochs in 12.83 hr record
  filter(contains_turns=='TRUE') %>%
  merge(., match_info, by='child_id')

# report stats for speech table
```

```
convo_consis_tbl <- convo_consis_check %>%
  group_by(match) %>%
  summarize(avg_convo_consis = mean(perc_turns),
            sd_convo_consis = sd(perc_turns),
            min_convo_consis = min(perc_turns),
            max_convo_consis = max(perc_turns)) %>%
  mutate_if(is.numeric, round, 2) %>%
  mutate(convo_consis=paste0(avg_convo_consis,"(",sd_convo_consis,")",",",min_convo_consis,"-",max_conv
  select(match,convo_consis) %>%
  spread(match, convo_consis) %>%
  rownames_to_column(.,var = 'measure')
```

```
# create a 4th "match" of CI kids to compute hearing age
match <- matches %>% select(child_id, match, hearing_age)

ha_kids_ctc <- its_df %>%
  select(child_id, age_mos) %>%
  merge(., recording_convo, by='child_id') %>%
  merge(., match, by='child_id') %>%
  filter(match=='CI') %>%
  select(-age_mos, -match) %>%
  mutate(age_mos = hearing_age,
         match = 'CI_by_hearing_age') %>%
  filter(!hearing_age<=12) %>%
  select(-hearing_age)

ctc_growth_tbl <- its_df %>%
  select(child_id, age_mos, match) %>%
  merge(., recording_convo, by='child_id') %>%
  rbind(., ha_kids_ctc) %>%
  group_by(match) %>%
  do(ctc_growth = lm(normed_turns~age_mos, data=.),
     mod2 = cor(.$normed_turns, .$age_mos, method = "pearson")) %>%
  mutate(slope = summary(ctc_growth)$coeff[2],
         p_value = summary(ctc_growth)$coeff[8],
         Pearson = mod2[1]) %>%
  select(match, slope, p_value, Pearson) %>%
  mutate_if(is.numeric, round, 2) %>%
  mutate(stats=paste0('B=',slope,",","p=",p_value, ",","r=", Pearson)) %>%
  select(-slope, -p_value, -Pearson) %>%
  spread(match, stats) %>%
  mutate(measure='Convo. turn growth')
```

5.2 Model turns

```
# ---------- QUANTITY ------------
# hourly measures
# turns
hourly_turns$match <- relevel(factor(hourly_turns$match), ref = "CI")
hourly_turns$hours <- as.factor(hourly_turns$hours)
hourly_turns_m0 <- lmer(normed_hourly_turns~ + (1 | child_id) + (1|hours), data=hourly_turns)
hourly_turns_m1 <- lmer(normed_hourly_turns~ match+ (1 | child_id) + (1|hours), data=hourly_turns)
```

```
anova(hourly_turns_m0,hourly_turns_m1)
```

```
## Data: hourly_turns
## Models:
## hourly_turns_m0: normed_hourly_turns ~ +(1 | child_id) + (1 | hours)
## hourly_turns_m1: normed_hourly_turns ~ match + (1 | child_id) + (1 | hours)
##                 npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## hourly_turns_m0    4 6959.5 6977.4 -3475.7   6951.5
## hourly_turns_m1    6 6963.2 6990.1 -3475.6   6951.2 0.3015  2     0.8601
```

```
# ---------- CONSISTENCY ------------
convo_consis_check2 <- its_df %>% select(age_mos, child_id) %>% merge(., convo_consis_check, by='child_
convo_consis_check2$match <- relevel(factor(convo_consis_check2$match), ref = "CI")
convo_consis_m0 <- lm(perc_turns~age_mos, data=convo_consis_check2)
convo_consis_m1 <- lm(perc_turns~ age_mos + match, data=convo_consis_check2)
anova(convo_consis_m0,convo_consis_m1) # no
```

```
## Analysis of Variance Table
##
## Model 1: perc_turns ~ age_mos
## Model 2: perc_turns ~ age_mos + match
##   Res.Df     RSS Df Sum of Sq      F Pr(>F)
## 1     50 0.74424
## 2     48 0.71800  2  0.026235 0.8769 0.4226
```

```
voc_consis_m2 <- lm(perc_turns~ age_mos*match, data=convo_consis_check2)
anova(convo_consis_m0,voc_consis_m2) # no
```

```
## Analysis of Variance Table
##
## Model 1: perc_turns ~ age_mos
## Model 2: perc_turns ~ age_mos * match
##   Res.Df     RSS Df Sum of Sq      F Pr(>F)
## 1     50 0.74424
## 2     46 0.66108  4  0.083161 1.4466 0.2339
```

6 Contingency

6.1 Compute contingency

```
for_cont <- vocs %>%
  select(child_id, onset, offset, segment_type, avg_dB) %>%
  rename(clip_onset = onset,
         clip_offset = offset)

# what percentage of vocs are contingent
contingent_df <- speech %>%
  select(child_id, clip_onset, clip_offset, segment_type, avg_dB) %>%
  rbind(., for_cont) %>%
  arrange(child_id, clip_offset) %>%
  group_by(child_id) %>%
  mutate(speech_lag = clip_offset - lag(clip_offset, default = clip_offset[1])) %>% # calculate lag tim
```

```
  mutate(contingent = if_else(speech_lag <=2, "Y", "N"))

total_chn <- contingent_df %>%
  filter(segment_type=='CHN') %>%
  count(contingent) %>% # note that this is not the correct contingent-noncontingent count; it's just t
  group_by(child_id) %>%
  mutate(total_vocs = sum(n)) %>%  # compute the denominator (total CHN vocs)
  distinct(child_id, .keep_all = T) %>%
  select(child_id, total_vocs)

contingent_df_lag <- contingent_df %>%
  filter(contingent=='Y' & segment_type=='CHN' & (lag(segment_type=='FAN')|lag(segment_type=='MAN'))) #


contingent_df2 <- contingent_df_lag %>%
  count(contingent) %>% # this is the correct count of contingent vocs
  merge(., total_chn, by='child_id') %>%
  mutate(perc_contingent = (n/total_vocs)*100) %>%
  merge(its_df, by='child_id')

contingent_df_CI <- contingent_df2 %>% filter(match=='CI')# get the kids with CIs to get hearing age ki

contingent_final <- contingent_df_CI %>%
  select(-match, -age_mos) %>%
  mutate(age_mos=hearing_age) %>%
  filter(!hearing_age<=12) %>%
  mutate(match='CI_by_hearing_age') %>%
  rbind(contingent_df2)
```

```
# make a dataframe containing the timestamps for all child-adult vocal interactions
# within a 5-second 'contingent' window
# computing over all vocalizations, regardless of temporal window, results in lots and lots of meaningl
# outliers e.g. 50seconds between adult and child vocalization
for_temp <- vocs %>%
  select(child_id, onset, offset, segment_type, hours) %>%
  rename(clip_onset = onset,
         clip_offset = offset)

time_vocs <- speech %>%
  select(child_id, clip_onset, clip_offset, segment_type, hours) %>%
  rbind(., for_temp) %>%
  arrange(child_id, clip_offset) %>%
  group_by(child_id) %>%
  mutate(speech_lag = clip_offset - lag(clip_offset, default = clip_offset[1])) %>% # calculate lag tim
  mutate(contingent = if_else(speech_lag <=2, "Y", "N")) %>%
  filter(contingent=='Y' & segment_type=='CHN' & (lag(segment_type=='FAN')|lag(segment_type=='MAN'))) %
  merge(its_df, by='child_id') %>%
  group_by(child_id,hours) %>%
  mutate(avg_lag = mean(log(speech_lag))) %>%
  distinct_at(., vars(child_id, hours), .keep_all = T)

time_vocs_CI <- time_vocs %>% filter(match=='CI') # get the kids with CIs to get hearing age kids
```

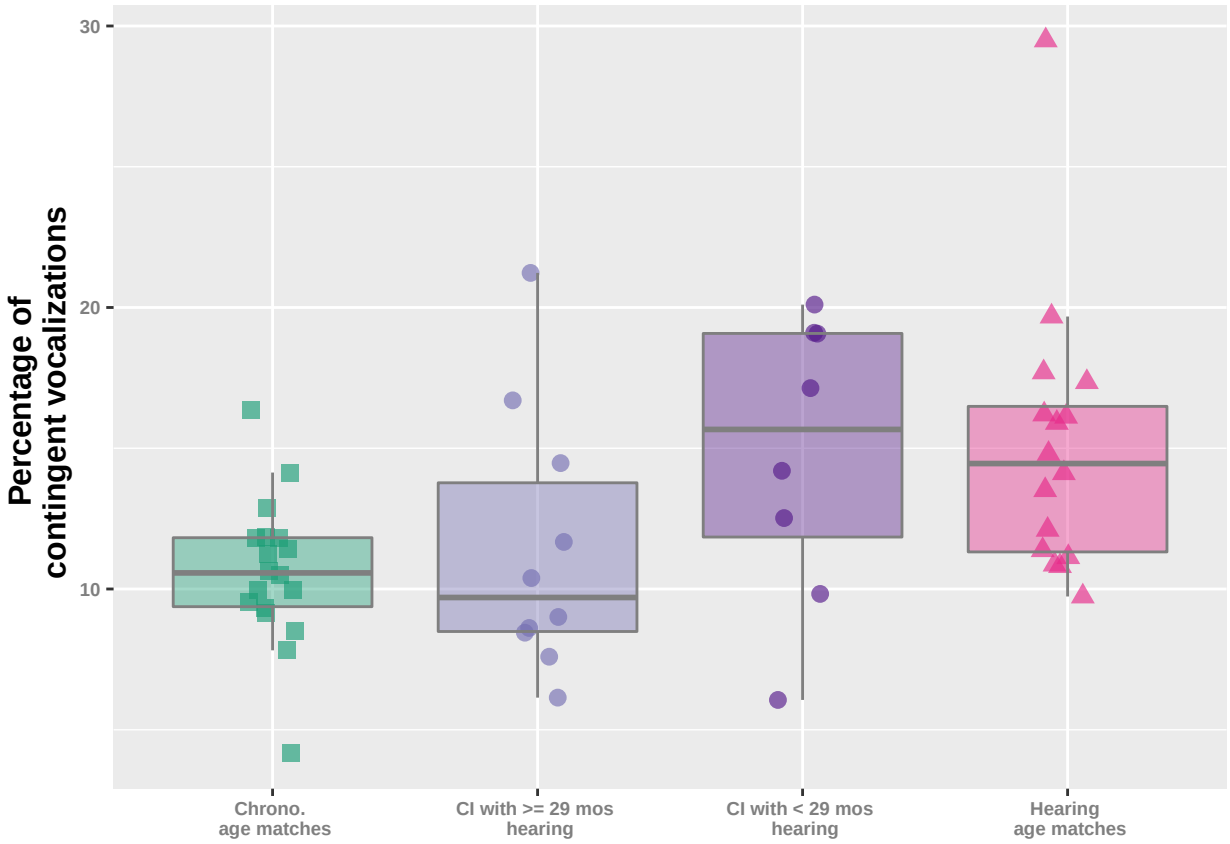
```
time_vocs_final <- time_vocs_CI %>%
  select(-match, -age_mos) %>%
  mutate(age_mos=hearing_age) %>%
  filter(!hearing_age<=12) %>%
  mutate(match='CI_by_hearing_age') %>%
  rbind(time_vocs)
```

```
# classify each adult vocalization as high vs low intensity
med_meas <- contingent_df %>%
  filter(segment_type=='FAN' | segment_type=='MAN') %>% # only compute median over input
  group_by(child_id) %>%
  summarize(med_dB=median(avg_dB))

intens <- contingent_df %>%
  merge(., med_meas, by='child_id') %>%
  group_by(child_id) %>%
  mutate(adult_loudness = if_else(avg_dB>med_dB, "loud", "soft"))

total_voc_df <- total_chn %>% distinct_at(., vars(child_id,total_vocs)) #dataframe containing total num

all_intens_vocs <- intens %>%
  group_by(child_id) %>%
  mutate(adult_loudness=lag(adult_loudness)) %>% # put the adult voc classification in the same row as
  mutate(adult_voc_dB=lag(avg_dB)) %>% # put the adult voc dB measurement in the same row as the child
  select(-avg_dB) %>%
  filter(contingent=='Y' & segment_type=='CHN' & (lag(segment_type=='FAN')|lag(segment_type=='MAN'))) %
  group_by(child_id,adult_loudness) %>%
  add_count(contingent) %>% # this is the count of contingent vocs in response to loud versus soft care
  merge(., total_voc_df, by='child_id') %>% # merge with dataframe containing total # of CHN
  group_by(child_id, adult_loudness) %>%
  mutate(perc_contingent_loudness = (n/total_vocs)*100) %>% # compute the percentage of contingent in r
  merge(., its_df, by='child_id')

intens2 <- all_intens_vocs %>%
  distinct_at(., vars(child_id, adult_loudness), .keep_all = T) %>%
  select(-adult_voc_dB, -med_dB, -speech_lag,  -clip_offset, -clip_onset) # clean up
```

6.2 Visualize contingency

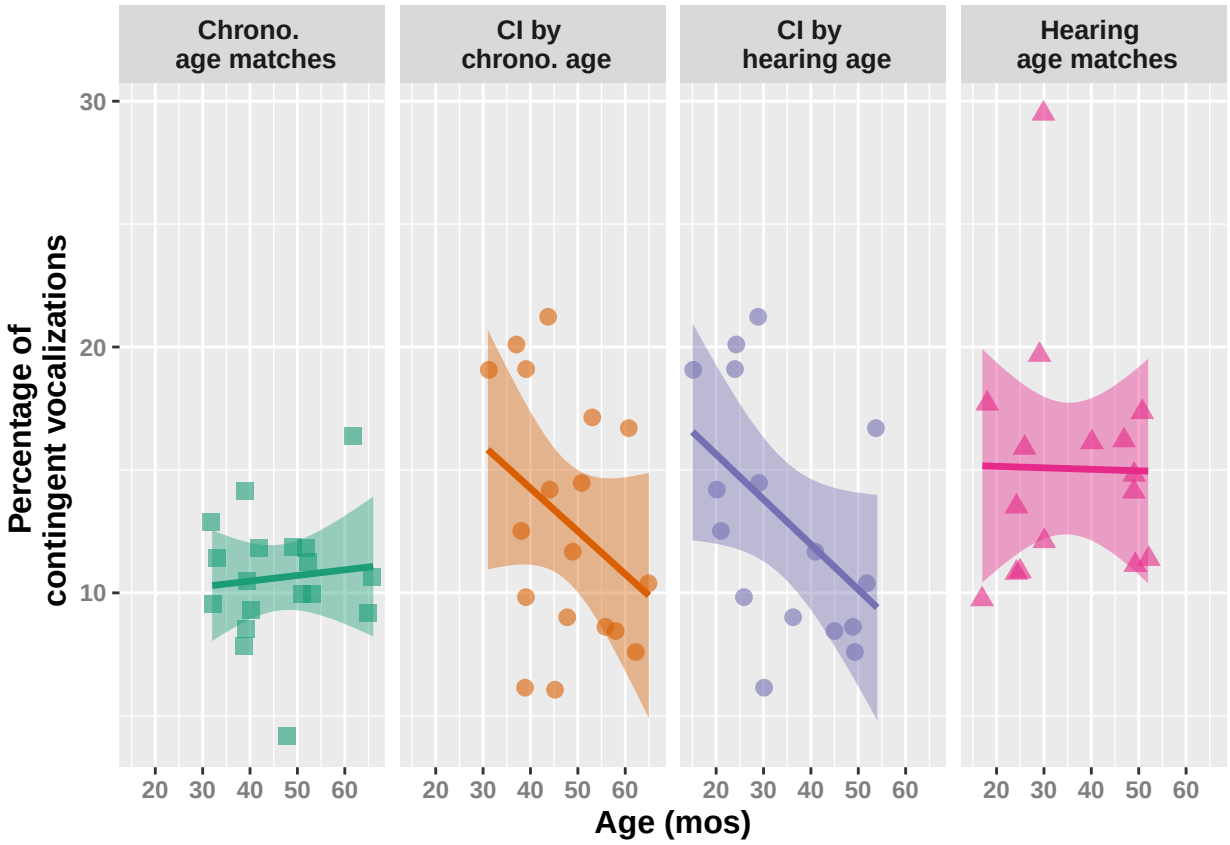
```
# do a binary split among the CI kids to examine the effects of hearing experience
# upon vocal contingency

cont_boxplot <- contingent_final %>%
  filter(match=='CI') %>%
  mutate(med_hearing_age = median(hearing_age)) %>%
  mutate(match=if_else(hearing_age >= med_hearing_age, 'more', 'less')) %>%
  select(-med_hearing_age)

cont_boxplot2 <- contingent_final %>%
  filter(match=='HA' | match=='chrono') %>%
  rbind(., cont_boxplot)
```

```
cont_boxplot2 %>%
  mutate(match=factor(match, levels = c("chrono", "more", "less", "HA"))) %>%
   mutate(match=recode(match,
                       chrono='Chrono. \n age matches',
                       more='CI with >= 29 mos \n hearing',
                       less='CI with < 29 mos \n hearing',
                       HA='Hearing \n age matches')) %>%
ggplot(., aes(match, perc_contingent)) +
  geom_jitter(aes(color=match, fill=match,shape=match),size=2.8,width = .1,alpha=.65) +
  geom_boxplot(aes(fill=match), alpha=.4, color='gray50', outlier.shape = NA) +
  scale_color_manual(values=c("#1B9E77", "#7570B3", "purple4", "#E7298A"))+
  scale_fill_manual(values=c("#1B9E77", "#7570B3", "purple4", "#E7298A")) +
  scale_shape_manual(values=c(15,16,16,17)) +
  #xlab("Hearing experience") +
  ylab("Percentage of \n contingent vocalizations") +
  #ggtitle("Contingent vocalizations, by hearing group") +
  theme(legend.position = "none",
        axis.title.y = element_text(face ="bold", size=12),
        axis.title.x = element_blank(),
        axis.text = element_text(face="bold", color='gray50', size=7),
        plot.title = element_text(face="bold", size=16))
```



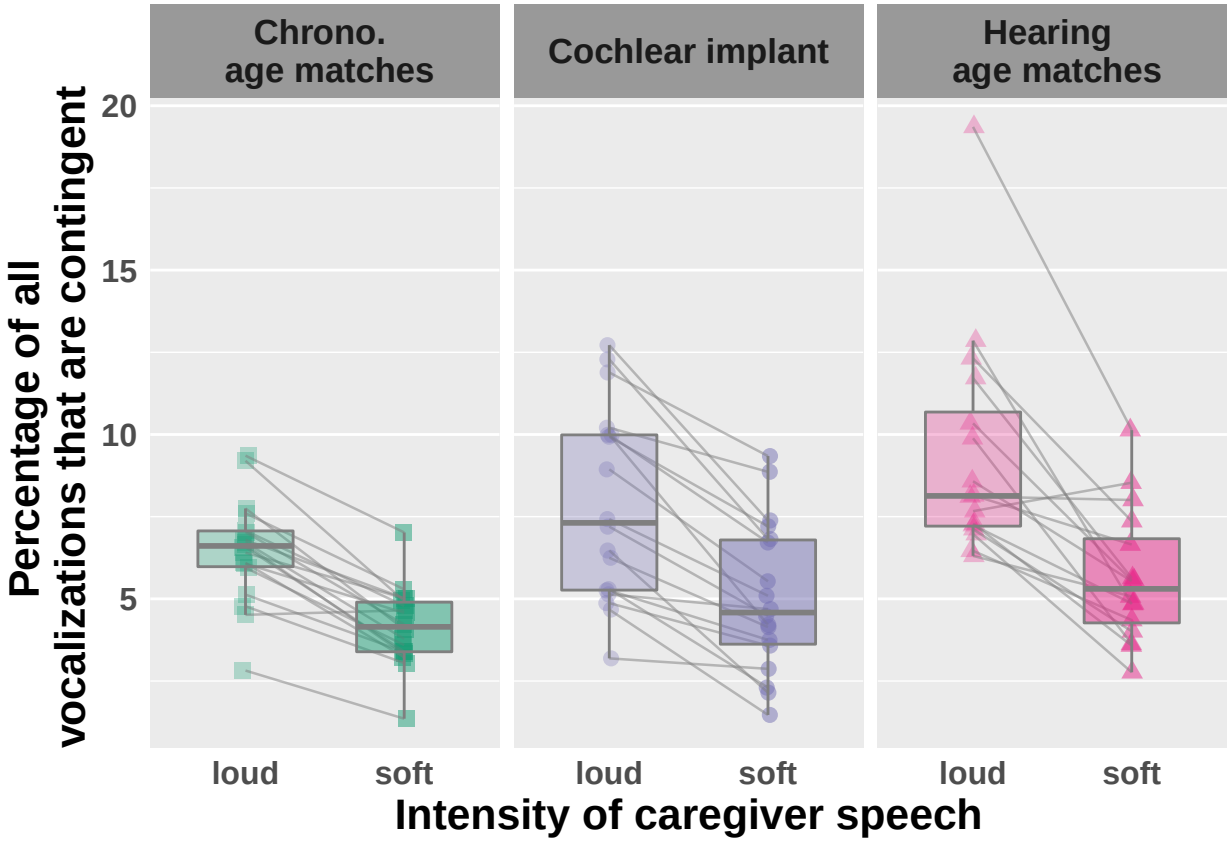

```
contingent_final %>%
  mutate(match=recode(match,
                       chrono='Chrono. \n age matches',
                       CI='CI by \n chrono. age',
```

```
                      CI_by_hearing_age='CI by \n hearing age',
                      HA='Hearing \n age matches')) %>%
ggplot(., aes(age_mos, perc_contingent)) +
  geom_jitter(aes(color=match, fill=match, shape=match),size=2.8,alpha=.6,width = .3) +
  geom_smooth(aes(fill=match, color=match), method = "lm",size=1.2) +
  facet_grid(~match) +
  scale_color_brewer(palette = "Dark2") +
  scale_fill_brewer(palette = "Dark2") +
  scale_shape_manual(values=c(15,16,16,17)) +
  xlab("Age (mos)") +
  ylab("Percentage of \n contingent vocalizations") +
    theme(axis.title = element_text(face ="bold", size=12),
        legend.position = "none",
        axis.text = element_text(face="bold", color='gray50', size=9),
        strip.text=element_text(face='bold', size=10))
```



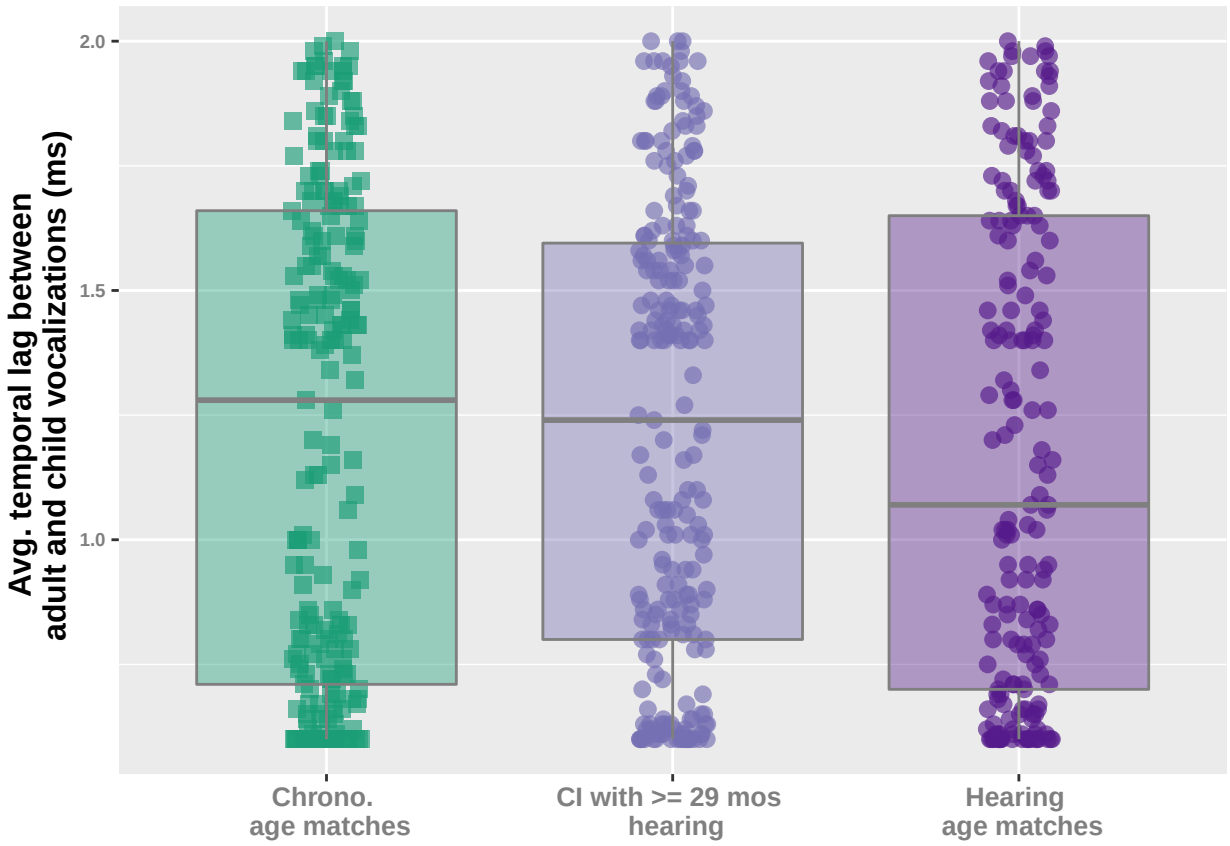

```
# are kids more likely to respond to louder speech? does this vary by hearing group?
intens2 %>%
    mutate(match=recode(match,
                      chrono='Chrono. \n age matches',
                      CI='Cochlear implant',
                      HA='Hearing \n age matches')) %>%
ggplot(., aes(x=adult_loudness, y=perc_contingent_loudness)) +
    geom_point(aes(fill=match,shape=match,group=child_id,color=match, alpha=adult_loudness),
                size=2.5, position=position_jitter(0.02)) +
    geom_line(aes(group = child_id), color="gray50",size = .45, alpha = 0.5) +
```

```
    geom_boxplot(aes(fill=match, alpha=adult_loudness),notch=FALSE,size=.5, outlier.shape = NA,
                 width=0.6,color="gray50", position=position_dodge(.6),) +
    facet_wrap(~match) +
   scale_fill_manual(values=c("#1B9E77", "#7570B3", "#E7298A"))+
   scale_color_manual(values=c("#1B9E77", "#7570B3", "#E7298A"))+
     scale_alpha_manual(values=c(.3, .5))+
  scale_shape_manual(values=c(15,16,17)) +

  labs(x="Intensity of caregiver speech",y="Percentage of all \n vocalizations that are contingent") +
  guides(alpha="none", fill = "none",color="none",shape="none") +
  theme(legend.position = c(.8, .8),
        axis.ticks = element_blank(),
        legend.title=element_text(face="bold", size=13),
        legend.text=element_text(face="bold", size=9),
        axis.text = element_text(face ='bold', size=12),
        axis.title = element_text(face ='bold', size=16),
        panel.grid.major.x = element_blank(),
        strip.text.x = element_text(face = "bold",size=13),
        strip.background = element_rect(fill = "gray60", size = 1))
```




```
temp_boxplot <- time_vocs_final %>%
  filter(match=='CI') %>%
  mutate(med_hearing_age = median(hearing_age)) %>%
  mutate(match=if_else(hearing_age >= med_hearing_age, 'more', 'less')) %>%
  select(-med_hearing_age)
```

```
temp_boxplot2 <- time_vocs_final %>%
  filter(match=='HA' | match=='chrono') %>%
  rbind(., temp_boxplot)

temp_boxplot2 %>%
  mutate(match=factor(match, levels = c("chrono", "more", "less", "HA"))) %>%
   mutate(match=recode(match,
                       chrono='Chrono. \n age matches',
                       more='CI with >= 29 mos \n hearing',
                       less='CI with < 29 mos \n hearing',
                       HA='Hearing \n age matches')) %>%
   #group_by(match,child_id) %>%
   #summarize(mean_speech_lag=mean(speech_lag*1000)) %>%
ggplot(., aes(match, speech_lag)) +
  geom_jitter(aes(color=match, fill=match,shape=match),size=2.8,width = .1,alpha=.65) +
  geom_boxplot(aes(fill=match), alpha=.4, color='gray50', outlier.shape = NA) +
  scale_color_manual(values=c("#1B9E77", "#7570B3", "purple4", "#E7298A"))+
  scale_fill_manual(values=c("#1B9E77", "#7570B3", "purple4", "#E7298A")) +
  scale_shape_manual(values=c(15,16,16,17)) +
  ylab("Avg. temporal lag between \n adult and child vocalizations (ms)") +
  #ggtitle("Contingent vocalizations, by hearing group") +
  theme(legend.position = "none",
        axis.title.y = element_text(face ="bold", size=12),
        axis.title.x = element_blank(),
        axis.text.y = element_text(face="bold", color='gray50', size=7),
        axis.text.x = element_text(face="bold", color='gray50', size=10),
        plot.title = element_text(face="bold", size=16))
```



6.3 Model contingency

```
# ------- look at percentage contingent -------
# look at just three groups
contingent_final_3 <- contingent_final %>% filter(match!='CI_by_hearing_age')
contingent_final_3$match <- relevel(factor(contingent_final_3$match), ref = "chrono")
m1 <- lm(perc_contingent~match, data=contingent_final_3)
summary(m1)
```

```
##
## Call:
## lm(formula = perc_contingent ~ match, data = contingent_final_3)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.8452 -2.9846 -0.3205  1.7434 14.4340
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   10.622      1.005  10.566 3.13e-14 ***
## matchCI        2.283      1.422   1.606  0.11479
## matchHA        4.434      1.465   3.025  0.00395 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.265 on 49 degrees of freedom
```

```
## Multiple R-squared:  0.1578, Adjusted R-squared:  0.1234
## F-statistic: 4.589 on 2 and 49 DF,  p-value: 0.0149
```

```
# what's the correlation between age and perc contingent for kids with CIs?
ci_for_cor <- contingent_final_3 %>% filter(match=='CI')
ci_cor <- cor.test(ci_for_cor$age_mos, ci_for_cor$perc_contingent)
ci_cor_aoi <- cor.test(ci_for_cor$age_of_implantation, ci_for_cor$perc_contingent)

# look at HA, chrono, and more vs. less hearing experience groups
cont_boxplot2$match <- relevel(factor(cont_boxplot2$match), ref = "HA")
ha_model <- lm(perc_contingent~match, data=cont_boxplot2)
summary(ha_model) # the kids with more experience differ from the hearing age matches, as expected
```

```
##
## Call:
## lm(formula = perc_contingent ~ match, data = cont_boxplot2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.6923 -2.8033 -0.4043  2.2772 14.4340
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  15.0559     1.0472  14.377  < 2e-16 ***
## matchchrono  -4.4337     1.4393  -3.081  0.00342 **
## matchless    -0.3039     1.8138  -0.168  0.86766
## matchmore    -3.6287     1.6886  -2.149  0.03671 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.189 on 48 degrees of freedom
## Multiple R-squared:  0.2042, Adjusted R-squared:  0.1544
## F-statistic: 4.105 on 3 and 48 DF,  p-value: 0.01133
```

```
cont_boxplot2$match <- relevel(factor(cont_boxplot2$match), ref = "chrono")
chrono_model <- lm(perc_contingent~match, data=cont_boxplot2)
summary(chrono_model) # the kids with less experience differ from the chrono age matches, as expected
```

```
##
## Call:
## lm(formula = perc_contingent ~ match, data = cont_boxplot2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.6923 -2.8033 -0.4043  2.2772 14.4340
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  10.6222     0.9873  10.758 2.19e-14 ***
## matchHA       4.4337     1.4393   3.081  0.00342 **
## matchless     4.1298     1.7799   2.320  0.02463 *
## matchmore     0.8050     1.6521   0.487  0.62830
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 4.189 on 48 degrees of freedom
## Multiple R-squared:  0.2042, Adjusted R-squared:  0.1544
## F-statistic: 4.105 on 3 and 48 DF,  p-value: 0.01133
```

```
cont_m0 <- lm(perc_contingent~age_mos, data=contingent_final_3)
cont_m1 <- lm(perc_contingent~age_mos + match, data=contingent_final_3)
cont_m2 <- lm(perc_contingent~age_mos*match, data=contingent_final_3)

anova(cont_m0, cont_m1)
```

```
## Analysis of Variance Table
##
## Model 1: perc_contingent ~ age_mos
## Model 2: perc_contingent ~ age_mos + match
##   Res.Df     RSS Df Sum of Sq      F  Pr(>F)
## 1     50 1000.73
## 2     48  880.49  2    120.23 3.2773 0.04633 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(cont_m1, cont_m2)
```

```
## Analysis of Variance Table
##
## Model 1: perc_contingent ~ age_mos + match
## Model 2: perc_contingent ~ age_mos * match
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     48 880.49
## 2     46 840.10  2    40.394 1.1059 0.3396
```

```
# ------ look at temporal dynamics -------
# for all kids?
time_vocs_final2 <- time_vocs_final %>% filter(match!='chrono')
time_vocs_final2$match <- relevel(factor(time_vocs_final2$match), ref = "CI")
m0_lag <- lmer(speech_lag*1000~match + (1 | child_id), data=time_vocs_final2)
m1_lag <- lmer(speech_lag*1000~ match+age_mos+(1|child_id), data=time_vocs_final2)
anova(m0_lag,m1_lag) # no improvement
```

```
## Data: time_vocs_final2
## Models:
## m0_lag: speech_lag * 1000 ~ match + (1 | child_id)
## m1_lag: speech_lag * 1000 ~ match + age_mos + (1 | child_id)
##        npar    AIC    BIC  logLik deviance Chisq Df Pr(>Chisq)
## m0_lag    5 9434.7 9456.9 -4712.4   9424.7
## m1_lag    6 9436.4 9463.1 -4712.2   9424.4 0.298  1     0.5851
```

```
m2_lag <- lmer(speech_lag*1000~ match*age_mos+(1|child_id), data=time_vocs_final2)
anova(m1_lag,m2_lag) # no improvement
```

```
## Data: time_vocs_final2
## Models:
## m1_lag: speech_lag * 1000 ~ match + age_mos + (1 | child_id)
```

```
## m2_lag: speech_lag * 1000 ~ match * age_mos + (1 | child_id)
##        npar    AIC    BIC  logLik deviance Chisq Df Pr(>Chisq)
## m1_lag    6 9436.4 9463.1 -4712.2   9424.4
## m2_lag    8 9439.8 9475.3 -4711.9   9423.8 0.624  2      0.732
```

```
# what's the association between age and temporal synchrony for kids with CIs?
ci_temp_for_model <- time_vocs_final %>% filter(match=='CI')
ci_temp_m <- lmer(speech_lag*1000~hearing_age + (1|child_id), data=ci_temp_for_model)
summary(ci_temp_m)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: speech_lag * 1000 ~ hearing_age + (1 | child_id)
##    Data: ci_temp_for_model
##
## REML criterion at convergence: 3470.4
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.40200 -0.90924  0.01766  0.81871  1.71995
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  child_id (Intercept)   1418    37.66
##  Residual             206984   454.95
## Number of obs: 231, groups:  child_id, 18
##
## Fixed effects:
##              Estimate Std. Error        df t value Pr(>|t|)
## (Intercept) 1225.0001    76.0482   14.3570  16.108 1.36e-10 ***
## hearing_age   -0.1716     2.2176   14.3767  -0.077    0.939
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr)
## hearing_age -0.912
```

```
# look at HA, chrono, and more vs. less hearing experience groups for temporal synchrony
temp_boxplot2$match <- relevel(factor(temp_boxplot2$match), ref = "HA")
ha_temp_model <- lmer(speech_lag*1000~match+(1|child_id), data=temp_boxplot2)
summary(ha_temp_model) # the kids with more experience differ from the hearing age matches, as expected
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: speech_lag * 1000 ~ match + (1 | child_id)
##    Data: temp_boxplot2
##
## REML criterion at convergence: 9687.8
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.30862 -1.00017 -0.02111  0.89944  1.72260
##
```

```
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  child_id (Intercept)      0     0.0
##  Residual             224470   473.8
## Number of obs: 641, groups:  child_id, 52
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)  1183.86      34.46  638.00  34.352   <2e-16 ***
## matchchrono    23.42      46.94  638.00   0.499    0.618
## matchmore      36.14      46.47  638.00   0.778    0.437
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) mtchch
## matchchrono -0.734
## matchmore   -0.742  0.544
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

```
temp_boxplot2$match <- relevel(factor(temp_boxplot2$match), ref = "chrono")
chrono_temp_model <- lmer(speech_lag*1000~match+(1|child_id), data=temp_boxplot2)
summary(chrono_temp_model) # the kids with less experience differ from the chrono age matches, as expec
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: speech_lag * 1000 ~ match + (1 | child_id)
##    Data: temp_boxplot2
##
## REML criterion at convergence: 9687.8
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -1.30862 -1.00017 -0.02111  0.89944  1.72260
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  child_id (Intercept)      0     0.0
##  Residual             224470   473.8
## Number of obs: 641, groups:  child_id, 52
##
## Fixed effects:
##             Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)  1207.29      31.87  638.00  37.881   <2e-16 ***
## matchHA       -23.42      46.94  638.00  -0.499    0.618
## matchmore      12.71      44.58  638.00   0.285    0.776
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##           (Intr) mtchHA
## matchHA   -0.679
## matchmore -0.715  0.485
```

```
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

```
center_scale <- function(x) {
    scale(x, scale = FALSE)
}
all_intens_vocs <- all_intens_vocs %>%
  mutate(adult_voc_dB_centered = center_scale(adult_voc_dB))

#not using these
intens2$match <- relevel(factor(intens2$match), ref = "CI")
intens_m0 <- lm(perc_contingent_loudness~adult_loudness, data=intens2)
intens_m1 <- lm(perc_contingent_loudness~match+adult_loudness, data=intens2)
anova(intens_m0,intens_m1)
```

```
## Analysis of Variance Table
##
## Model 1: perc_contingent_loudness ~ adult_loudness
## Model 2: perc_contingent_loudness ~ match + adult_loudness
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1    102 625.64
## 2    100 542.15  2    83.483 7.6992 0.0007762 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
intens_m2 <- lm(perc_contingent_loudness~match*adult_loudness, data=intens2)
anova(intens_m1,intens_m2) # no interaction
```

```
## Analysis of Variance Table
##
## Model 1: perc_contingent_loudness ~ match + adult_loudness
## Model 2: perc_contingent_loudness ~ match * adult_loudness
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    100 542.15
## 2     98 533.51  2    8.6462 0.7941 0.4549
```

```
# repeated measures model:
# hearing status?
all_intens_vocs$match <- relevel(factor(all_intens_vocs$match), ref = "CI")
lag_m0 <- lmer(speech_lag*1000~ + (1 | child_id), data=all_intens_vocs)
lag_m1 <- lmer(speech_lag*1000~ match + (1 | child_id), data=all_intens_vocs)
anova(lag_m0,lag_m1)
```

```
## Data: all_intens_vocs
## Models:
## lag_m0: speech_lag * 1000 ~ +(1 | child_id)
## lag_m1: speech_lag * 1000 ~ match + (1 | child_id)
##        npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## lag_m0    3 325152 325176 -162573   325146
## lag_m1    5 325154 325193 -162572   325144 2.2063  2     0.3318
```

```r
# age?
lag_m2 <- lmer(speech_lag*1000~ age_mos + (1 | child_id), data=all_intens_vocs)
anova(lag_m0,lag_m2) # no improvement
```

```
## Data: all_intens_vocs
## Models:
## lag_m0: speech_lag * 1000 ~ +(1 | child_id)
## lag_m2: speech_lag * 1000 ~ age_mos + (1 | child_id)
##        npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## lag_m0    3 325152 325176 -162573   325146
## lag_m2    4 325152 325183 -162572   325144 2.2086  1     0.1372
```

```r
lag_m3 <- lmer(speech_lag*1000~ age_mos*match + (1 | child_id), data=all_intens_vocs)
anova(lag_m0,lag_m3) # no improvement
```

```
## Data: all_intens_vocs
## Models:
## lag_m0: speech_lag * 1000 ~ +(1 | child_id)
## lag_m3: speech_lag * 1000 ~ age_mos * match + (1 | child_id)
##        npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## lag_m0    3 325152 325176 -162573   325146
## lag_m3    8 325153 325216 -162568   325137 9.2057  5     0.1011
```

```r
intens_m3 <- lmer(speech_lag*1000~adult_voc_dB_centered + (1 | child_id), data=all_intens_vocs)
intens_m4 <- lmer(speech_lag*1000~adult_voc_dB_centered + match + (1 | child_id), data=all_intens_vocs)
anova(intens_m3, intens_m4)
```

```
## Data: all_intens_vocs
## Models:
## intens_m3: speech_lag * 1000 ~ adult_voc_dB_centered + (1 | child_id)
## intens_m4: speech_lag * 1000 ~ adult_voc_dB_centered + match + (1 | child_id)
##           npar    AIC    BIC  logLik deviance  Chisq Df Pr(>Chisq)
## intens_m3    4 325078 325110 -162535   325070
## intens_m4    6 325080 325128 -162534   325068 2.0713  2      0.355
```

```r
intens_m5 <- lmer(speech_lag*1000~adult_voc_dB_centered*match + (1 | child_id), data=all_intens_vocs)
anova(intens_m3, intens_m5)
```

```
## Data: all_intens_vocs
## Models:
## intens_m3: speech_lag * 1000 ~ adult_voc_dB_centered + (1 | child_id)
## intens_m5: speech_lag * 1000 ~ adult_voc_dB_centered * match + (1 | child_id)
##           npar    AIC    BIC  logLik deviance Chisq Df Pr(>Chisq)
## intens_m3    4 325078 325110 -162535   325070
## intens_m5    8 325078 325142 -162531   325062  8.22  4    0.08384 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(intens_m5)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
```

```
## Formula: speech_lag * 1000 ~ adult_voc_dB_centered * match + (1 | child_id)
##    Data: all_intens_vocs
##
## REML criterion at convergence: 325037.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.5349 -1.0080 -0.1107  0.8852  1.9751
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  child_id (Intercept)   1158    34.04
##  Residual             215115   463.81
## Number of obs: 21500, groups:  child_id, 52
##
## Fixed effects:
##                                    Estimate Std. Error        df t value
## (Intercept)                        1185.476      9.832    42.899 120.577
## adult_voc_dB_centered                -3.235      0.938 17768.311  -3.449
## matchchrono                          -6.875     14.128    46.246  -0.487
## matchHA                             -20.091     14.589    44.982  -1.377
## adult_voc_dB_centered:matchchrono    -3.370      1.361 16826.200  -2.476
## adult_voc_dB_centered:matchHA        -1.730      1.371 18139.183  -1.262
##                                   Pr(>|t|)
## (Intercept)                        < 2e-16 ***
## adult_voc_dB_centered             0.000564 ***
## matchchrono                       0.628812
## matchHA                           0.175269
## adult_voc_dB_centered:matchchrono 0.013285 *
## adult_voc_dB_centered:matchHA     0.206941
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) ad__B_ mtchch mtchHA ad__B_:
## adlt_vc_dB_ -0.017
## matchchrono -0.696  0.012
## matchHA     -0.674  0.012  0.469
## adlt_vc_B_:  0.012 -0.689  0.016 -0.008
## adlt__B_:HA  0.012 -0.684 -0.008 -0.016  0.472
```

```
# interaction means that the effect of loudness on speech lag differs by hearing status
# centering dB means that the intercept is the speech lag for CI group when dB = 0, or the mean dB
# adult_voc_dB_centered: adding 1 dB decreases the speech lag by 3.235 ms *for CI kids* (so a slope of
# smaller lags for chrono and hearing matches than CI kids
# the slope of relationship between dB and lag for chrono kids is -3.235 + -3.370 (so a slope of -6.605
# the slope of the relationship between dB and lag for ha kids is -3.235 + -1.730 (so a slope of -4.965
```

```
# create model summary
intens_m5_tbl <- rbind(tidy(intens_m5,
                        effects = c("fixed"),
                        conf.int = TRUE)) %>%
  select(-effect) %>%
  mutate(term=recode(term,"(Intercept)"="Intercept",
```

```
                      "adult_voc_dB_centered"="Adult speech intensity (dB)",
                      "matchchrono"="Match:Chronological",
                      "matchHA"="Match:Hearing Age",
                      "adult_voc_dB_centered:matchchrono"="Adult speech intensity*Match:Chronological",
                      "adult_voc_dB_centered:matchHA"="Adult speech intensity*Match:Hearing Age")) %>%
  mutate_if(is.numeric, round, digits=2) %>%
  rename(Parameter=term,
         Estimate=estimate,
         S.E. = std.error,
         `z-statistic`=statistic,
         `p-value`=p.value) %>%
  mutate(`95% CI`=paste(conf.low,"-",conf.high)) %>%
  select(-conf.low,-conf.high)

knitr::kable(intens_m5_tbl,
             caption = 'Model predicting timing of contingent vocalizations',
             booktabs=T) %>%
  kable_styling() %>%
  landscape()
```

Table 3: (#tab:create model summary table)Model predicting timing of contingent vocalizations

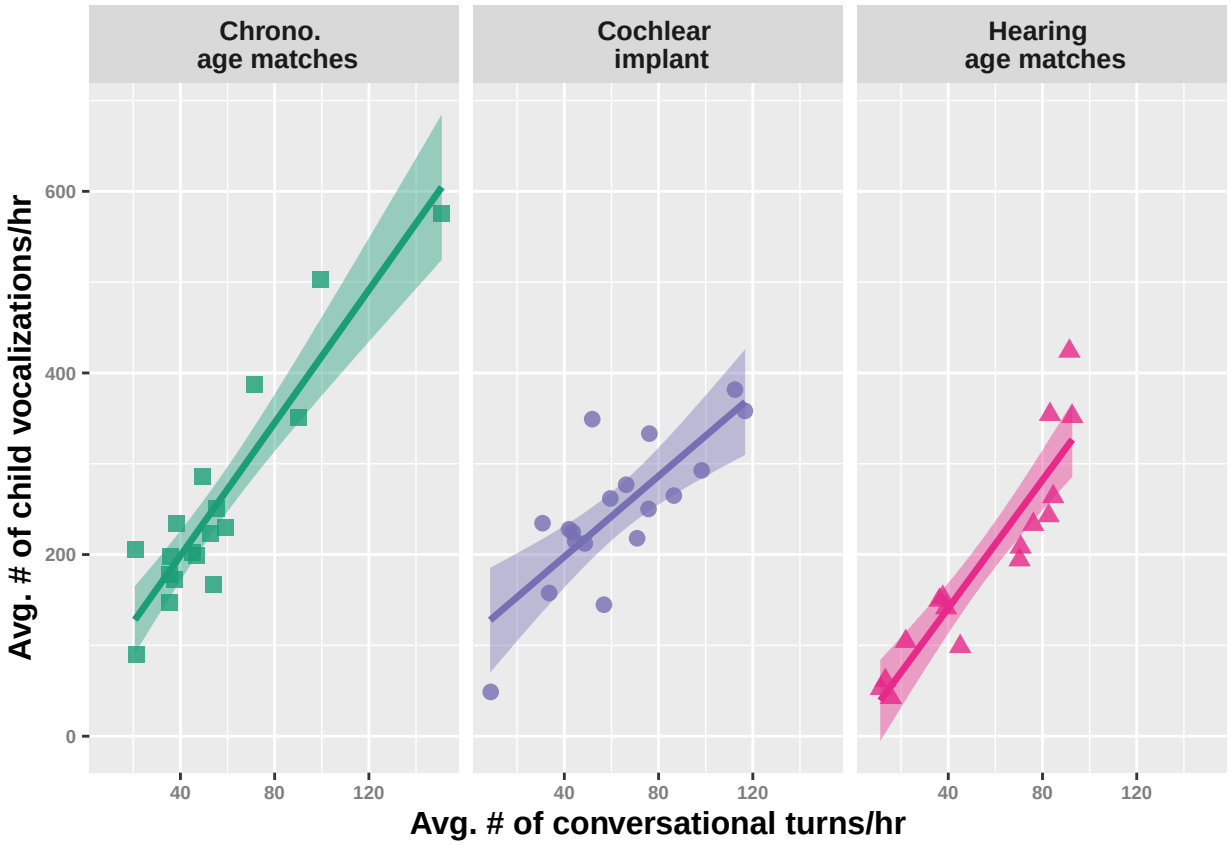
Parameter	Estimate	S.E.	z-statistic	df	p-value	95% CI
Intercept	1185.48	9.83	120.58	42.90	0.00	1165.65 – 1205.31
Adult speech intensity (dB)	-3.24	0.94	-3.45	17768.31	0.00	-5.07 – -1.4
Match:Chronological	-6.88	14.13	-0.49	46.25	0.63	-35.31 – 21.56
Match:Hearing Age	-20.09	14.59	-1.38	44.98	0.18	-49.47 – 9.29
Adult speech intensity*Match:Chronological	-3.37	1.36	-2.48	16826.20	0.01	-6.04 – -0.7
Adult speech intensity*Match:Hearing Age	-1.73	1.37	-1.26	18139.18	0.21	-4.42 – 0.96

7 Predicting vocal maturity from the speech environment

```
num_convo <- convo %>%
  group_by(child_id) %>%
  summarize(normed_turns = sum(convo_count)/total_hrs) %>%
  distinct(child_id,.keep_all = T)

num_words <- speech %>%
  group_by(child_id) %>%
  summarize(normed_words = sum(wordCount)/total_hrs) %>%
  distinct(child_id,.keep_all = T)

num_vocs_only <- vocs %>%
  group_by(child_id, match) %>%
  summarize(normed_vocs = sum(childUttCnt)/total_hrs) %>%
  distinct(child_id, .keep_all = T)

num_vocs <- num_vocs_only %>%
  merge(., num_convo, by='child_id') %>%
  merge(., num_words, by=c('child_id')) %>%
  merge(., its_df, by=c('match','child_id'))

# get a dataframe that contains the avg # of seconds/hr of adult speech
seconds_adult_speech <- speech %>%
  group_by(child_id, total_hrs) %>%
  summarize(total_input = sum(duration)) %>% # total input, in seconds, over the whole recording
  mutate(adult_speech_per_hour = total_input/(total_hrs)) %>% # normalize by duration of recording
  merge(., num_vocs, by='child_id')

# get a dataframe that contains the avg num of words per hour that were relatively loud
num_loud_words <- speech %>%
  merge(., med_meas, by='child_id') %>% # merge with df containing median measurement of adult speech i
  group_by(child_id) %>%
  mutate(adult_loudness=if_else(avg_dB>med_dB, "loud", "soft")) %>%
  filter(adult_loudness=='loud') %>% # select only those words that were relatively loud
  summarize(normed_loud_words = sum(wordCount)/total_hrs) %>%
  distinct(child_id,.keep_all = T) %>%
  merge(., num_vocs_only, by='child_id')

# what's the correlation between seconds of input and words of input?
cor.test(seconds_adult_speech$adult_speech_per_hour,seconds_adult_speech$normed_words)
```

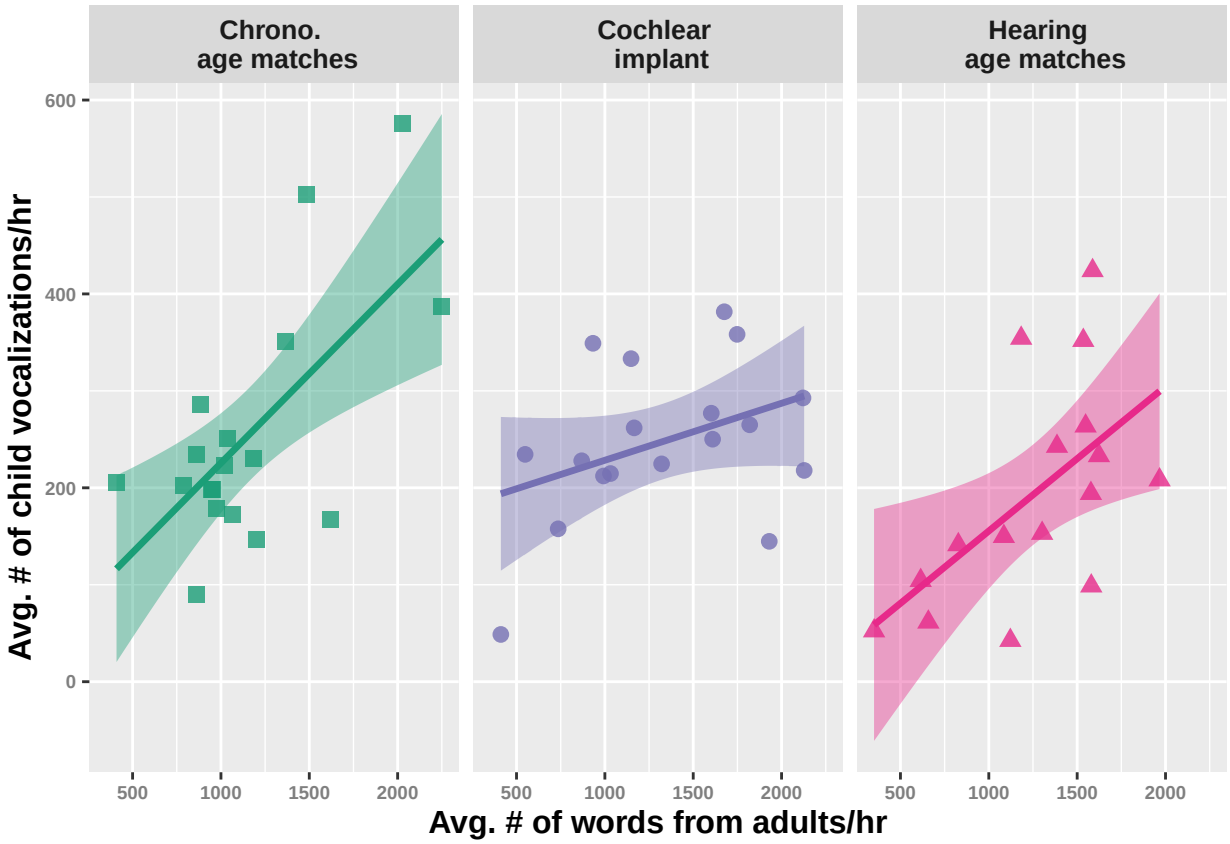
```
##
##  Pearson's product-moment correlation
##
## data:  seconds_adult_speech$adult_speech_per_hour and seconds_adult_speech$normed_words
## t = 55.635, df = 50, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.9860709 0.9954337
## sample estimates:
##       cor
## 0.9920197
```

```
num_vocs %>%
  mutate(match=factor(match, levels=c("chrono", "CI", "HA"))) %>%
  mutate(match=recode(match,
                      chrono='Chrono. \n age matches',
                      CI='Cochlear \n implant',
                      HA='Hearing \n age matches')) %>%
ggplot(., aes(normed_turns, normed_vocs)) +
  geom_jitter(aes(color=match, fill=match, shape=match),size=2.6,alpha=.8,width = .3) +
  geom_smooth(aes(fill=match, color=match), method = "lm",size=1.2) +
  facet_grid(~match) +
  #ylim(0, 600) +
  scale_color_manual(values=c("#1B9E77", "#7570B3", "#E7298A"))+
  scale_fill_manual(values=c("#1B9E77", "#7570B3", "#E7298A"))+
  scale_shape_manual(values=c(15,16,17)) +
  xlab("Avg. # of conversational turns/hr") +
  ylab("Avg. # of child vocalizations/hr") +
    theme(axis.title = element_text(face ="bold", size=12),
        legend.position = "none",
        axis.text = element_text(face="bold", color='gray50', size=7),
        strip.text=element_text(face='bold', size=10))
```




```
num_vocs %>%
  mutate(match=factor(match, levels=c("chrono", "CI", "HA"))) %>%
  mutate(match=recode(match,
                      chrono='Chrono. \n age matches',
                      CI='Cochlear \n implant',
```

```
                      HA='Hearing \n age matches')) %>%
ggplot(., aes(normed_words, normed_vocs)) +
  geom_jitter(aes(color=match, fill=match, shape=match),size=2.6,alpha=.8,width = .3) +
  geom_smooth(aes(fill=match, color=match), method = "lm",size=1.2) +
  facet_grid(~match) +
  #ylim(0, 600) +
    scale_color_manual(values=c("#1B9E77", "#7570B3", "#E7298A"))+
  scale_fill_manual(values=c("#1B9E77", "#7570B3", "#E7298A"))+
  scale_shape_manual(values=c(15,16,17)) +
  xlab("Avg. # of words from adults/hr") +
  ylab("Avg. # of child vocalizations/hr") +
    theme(axis.title = element_text(face ="bold", size=12),
        legend.position = "none",
        axis.text = element_text(face="bold", color='gray50', size=7),
        strip.text=element_text(face='bold', size=10))
```




```
num_vocs <- num_vocs %>%
  mutate(age_mos_centered = center_scale(age_mos),
         normed_turns_centered = center_scale(normed_turns),
         normed_words_centered = center_scale(normed_words))

# which measure of input fits the outcome measure better?
num_vocs$match <- relevel(factor(num_vocs$match), ref = "CI")
in_out_m0 <- lm(normed_vocs~normed_turns_centered+age_mos_centered, data=num_vocs)
in_out_m1 <- lm(normed_vocs~normed_words_centered+age_mos_centered, data=num_vocs)
# turns results in a better model fit
```

```
AIC(in_out_m0) #560.5966
```

```
## [1] 560.5966
```

```
AIC(in_out_m1) #611.2076
```

```
## [1] 611.2076
```

```
# match improves
in_out_m2 <- lm(normed_vocs~match, data=num_vocs)
in_out_m3 <- lm(normed_vocs~normed_turns_centered+match, data=num_vocs)
anova(in_out_m2,in_out_m3)
```

```
## Analysis of Variance Table
##
## Model 1: normed_vocs ~ match
## Model 2: normed_vocs ~ normed_turns_centered + match
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     49 575123
## 2     48 131913  1    443210 161.27 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# interaction improves
in_out_m4 <- lm(normed_vocs~normed_turns_centered*match, data=num_vocs)
anova(in_out_m3,in_out_m4)
```

```
## Analysis of Variance Table
##
## Model 1: normed_vocs ~ normed_turns_centered + match
## Model 2: normed_vocs ~ normed_turns_centered * match
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     48 131913
## 2     46 113563  2     18350 3.7165 0.0319 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# repeated measures - not using
#predict_df <- num_vocs %>%
#  merge(., vocs, by=c('child_id', 'match'))


#in_out_m2 <- lmer(log(childUttLen)~normed_turns+match+(1|child_id),data=predict_df)
#in_out_m3 <- lmer(log(childUttLen)~normed_words*match+(1|child_id),data=predict_df)
#in_out_m4 <- lmer(log(childUttLen)~normed_turns*match+(1|child_id),data=predict_df)
```

8 Create speech measures tables

8.1 Standard measures

```
# recording duration
rec_dur_tbl <- its_df %>%
  group_by(match) %>%
  summarize(recording_duration = mean(total_hrs),
            recording_duration_sd = sd(total_hrs),
            recording_duration_min = min(total_hrs),
            recording_duration_max = max(total_hrs)) %>%
  mutate_if(is.numeric, round, 2) %>%
  mutate(recording_stats = paste0(recording_duration,"(",recording_duration_sd,")",",",recording_duratio
  select(match, recording_stats) %>%
  spread(match, recording_stats) %>%
  rownames_to_column(.,var = 'measure')

# Input
speech_quantity_tbl
```

```
## # A tibble: 3 x 4
##   measure        chrono                          CI                          HA
##   <chr>          <chr>                           <chr>                       <chr>
## 1 intensity      68.12(5.98),47.1-84.22          68.82(5.87),45.46-84.2      68.7~
## 2 num_words_hr   1162.83(448.41),408.96-2250.39 1321.66(528.56),411.36-21~ 1246~
## 3 mins_speech_hr 291.29(111.26),112.59-545.81    327(125.54),100.7-514.19    314.~
```

```
speech_consis_tbl
```

```
## # A tibble: 1 x 4
##   measure chrono               CI                  HA
##   <chr>   <chr>                <chr>               <chr>
## 1 1       0.53(0.12),0.35-0.78 0.59(0.12),0.25-0.73 0.53(0.11),0.31-0.66
```

```
# Output
voc_tbl
```

```
## # A tibble: 3 x 4
##   measure     chrono                       CI                          HA
##   <chr>       <chr>                        <chr>                       <chr>
## 1 intensity   76.78(4.39),47.51-84.77      76.95(4.35),43.16-85.79     77.15(4.95~
## 2 num_vocs_hr 308.03(142.81),90.12-575.81 271.75(69.23),48.75-381.62 254.5(108.~
## 3 voc_dur     1004.46(662.3),80-10940      937.93(569.76),80-13270     966.59(627~
```

```
voc_consis_tbl
```

```
## # A tibble: 1 x 4
##   measure chrono               CI                  HA
##   <chr>   <chr>                <chr>               <chr>
## 1 1       0.55(0.15),0.34-0.84 0.58(0.13),0.17-0.72 0.49(0.14),0.22-0.69
```

```
# Interaction
convo_consis_tbl
```

```
## # A tibble: 1 x 4
##   measure chrono               CI                  HA
##   <chr>   <chr>                <chr>               <chr>
## 1 1       0.58(0.14),0.38-0.84 0.64(0.13),0.22-0.77 0.56(0.12),0.36-0.74
```

```
turn_quantity_tbl
```

```
## # A tibble: 1 x 4
##   measure chrono                    CI                      HA
##   <chr>   <chr>                     <chr>                   <chr>
## 1 1       61.71(32.78),20.69-150.94 68.17(26.47),8.5-116.75 65.13(25.47),11.12-~
```

```
measure_tbl <- rec_dur_tbl %>%
  rbind(., speech_quantity_tbl) %>%
  rbind(., speech_consis_tbl) %>%
  rbind(., voc_tbl) %>%
  rbind(., voc_consis_tbl) %>%
  rbind(., turn_quantity_tbl) %>%
  rbind(., convo_consis_tbl)

meas_mat <- measure_tbl %>% select(-measure) %>% as.matrix(.)

row.names(meas_mat) <- c("Recording duration (hrs.)",
                         "Adult speech intensity (dB)",
                         "Adult speech/hr (words)",
                         "Adult speech/hr (s)",
                         "Adult word consistency",
                         "Voc. intensity (dB)",
                         "Child voc. quantity",
                         "Voc. duration (ms)",
                         "Child voc. consistency",
                         #"Input:output", # maybe include this?
                         #"% contingent vocs.",
                         "Convo. turn quantity",
                         "Convo turn consistency")


kable(meas_mat,
             caption= "Measures of the naturalistic speech environment, by hearing group",
             col.names = c("Chrono. age matches",
                           "CI",
                           "Hearing age matches"),
      escape=FALSE) %>%
  kable_styling(.) %>%
  pack_rows(., "Recording", 1, 1) %>%
  pack_rows(., "Input", 2,5) %>%
  pack_rows(., "Output", 6,9) %>%
  pack_rows(., "Interaction", 10,11)
```

8.2 Growth in measures table

```
# we create a separate table here because comparing four "matches" (to include growth by hearing age)
growth_tbl <- input_growth_tbl  %>%
  rbind(., voc_growth_tbl) %>%
  rbind(., voc_dur_growth_tbl) %>%
  rbind(., ctc_growth_tbl) %>%
```

Table 4: (#tab:create standard measures table)Measures of the naturalistic speech environment, by hearing group

	Chrono. age matches	CI	Hearing age matches
Recording			
Recording duration (hrs.)	15.82(0.75),12.83-16	16(0),16-16	16(0),16-16
Input			
Adult speech intensity (dB)	68.12(5.98),47.1-84.22	68.82(5.87),45.46-84.2	68.73(6.19),44.92-88.4
Adult speech/hr (words)	1162.83(448.41),408.96-2250.39	1321.66(528.56),411.36-2127.7	1246.53(444.48),350.59-
Adult speech/hr (s)	291.29(111.26),112.59-545.81	327(125.54),100.7-514.19	314.5(104.37),107.19-47
Adult word consistency	0.53(0.12),0.35-0.78	0.59(0.12),0.25-0.73	0.53(0.11),0.31-0.66
Output			
Voc. intensity (dB)	76.78(4.39),47.51-84.77	76.95(4.35),43.16-85.79	77.15(4.95),44.79-90.31
Child voc. quantity	308.03(142.81),90.12-575.81	271.75(69.23),48.75-381.62	254.5(108.83),42.5-424
Voc. duration (ms)	1004.46(662.3),80-10940	937.93(569.76),80-13270	966.59(627.6),80-19730
Child voc. consistency	0.55(0.15),0.34-0.84	0.58(0.13),0.17-0.72	0.49(0.14),0.22-0.69
Interaction			
Convo. turn quantity	61.71(32.78),20.69-150.94	68.17(26.47),8.5-116.75	65.13(25.47),11.12-92.62
Convo turn consistency	0.58(0.14),0.38-0.84	0.64(0.13),0.22-0.77	0.56(0.12),0.36-0.74

```
  select(measure, chrono, CI, CI_by_hearing_age, HA)


kable(growth_tbl, booktabs=T,
              caption= "Growth in measures of the naturalistic speech environment, by hearing group",
              col.names = c(" ",
                            "Chrono. age matches",
                           "CI by chrono. age",
                           "CI by hearing age",
                           "Hearing age matches")) %>%
  kable_styling() %>%
  kable_styling(latex_options = "hold_position")
```

Table 5: (#tab:create growth table)Growth in measures of the naturalistic speech environment, by hearing group

	Chrono. age matches	CI by chrono. age	CI by hearing age	H
Adult word growth	B=14.1,p=0.17,r=0.34	B=1.03,p=0.94,r=0.02	B=-2.03,p=0.86,r=-0.05	B=
Child voc. quantity growth	B=3.98,p=0.16,r=0.34	B=2.71,p=0.19,r=0.33	B=1.06,p=0.53,r=0.17	B=
Child voc. duration growth (ms)	B=564.77,p=0.79,r=0.07	B=3168.23,p=0.05,r=0.47	B=1850.89,p=0.18,r=0.35	B=
Convo. turn growth	B=0.83,p=0.25,r=0.29	B=0.01,p=0.99,r=0	B=-0.33,p=0.56,r=-0.16	B=

```
# note that in the ICPhS paper I reported the p-value for CI hearing age that included the two kids
# with HA < 12
```