

coartic_midpts

Meg Cychosz

3/30/2019

An analysis of coarticulation across versus within morphemes in adult and child Quechua. Here we are analyzing adjacent coarticulation between [a] and [p] across morphemes as in “llama-pi” and within morphemes as in “api”.

Load data

```
data <- read.csv('all_files_mfccs_3tpt.csv', na.strings = c("NA", "NaN", " "))
```

Prepare and process data

```
new_morphDF <- Reduce(function(x,y) merge(x,y,all=TRUE) ,list(api2, imilla,  
    jukucha, llama, papa2, sunkha, tika, uhuta, waka, wallpa, wawa))  
new_morphDF$Word <- droplevels(new_morphDF$Word)
```

Convert structure of spectral measurements

```
new_morphDF$Spectrum_int_2 <- gsub( ']', '', new_morphDF$Spectrum_int_2)
new_morphDF$Spectrum_int_2 <- gsub( '[ ', '', new_morphDF$Spectrum_int_2, fixed = TRUE)

# make string, then numeric
new_morphDF$variable_sep_tp2 <- str_extract_all(new_morphDF$Spectrum_int_2, "[-0-9\\.]+")
new_morphDF$spec_vector_tp2 <- lapply(new_morphDF$variable_sep_tp2 , FUN = as.numeric)
new_morphDF <- as.data.frame(new_morphDF)
```

Calculate euc distance between vectors

```
# sanity check
if(any(grepl("package:plyr", search())) detach("package:plyr") else message("plyr not loaded"))

## plyr not loaded

# euclidean distance and raw difference
diff_morphdf <- new_morphDF %>%
  group_by(Word, Speaker, Note) %>%
  mutate(raw_diff = map2(spec_vector_tp2, lead(spec_vector_tp2), `~`)) %>% # sanity check
  mutate(euc_dist = map2(spec_vector_tp2, lead(spec_vector_tp2), function(x, y)
    sqrt(sum((x-y) ^ 2)))) %>%
  as.data.frame()

# sequence duration
diff_morphdf <- diff_morphdf %>%
  group_by(Word, Speaker, Note) %>%
  mutate(ap_duration = map2(Phone_duration, lead(Phone_duration), `+`)) %>%
  as.data.frame()

# remove NA rows where measurement was made upon but not stored
df.final <- subset(diff_morphdf, euc_dist != 'NA')
df.final$euc_dist <- as.numeric(df.final$euc_dist)
df.final$ap_duration <- as.numeric(df.final$ap_duration)
```

Descriptive stats

averages by age

```
## [1] "Average euc. distance between spectral vectors, all environments: adults 15.04"  
## [1] "Average euc. distance, all environments: children 16.46"  
## [1] "Average euc. distance, all environments: ten y/os 16.2"  
## [1] "Average euc. distance, all environments: nine y/os 15.34"  
## [1] "Average euc. distance, all environments: eight y/os 16.52"  
## [1] "Average euc. distance, all environments: seven y/os 16.25"  
## [1] "Average euc. distance, all environments: 5 & 6 y/os 18.56"
```

age*morphological status

```
## [1] "Average euc. distance across morphemes: adults 15.57"  
## [1] "Average euc. distance within morphemes: adults 13.31"  
## [1] "Average euc. distance across morphemes: children 16.4"  
## [1] "Average euc. distance within morphemes: children 16.59"
```

visualize some stuff

```
jpeg('age_eucdist.jpeg', width = 500, height=500)

ggplot(df.final, aes(Age, euc_dist, fill = Morph_status)) +
  geom_boxplot() + guides(fill = guide_legend(title = "Position")) +
  ylim(4, 30) +
  scale_fill_manual(values=c("gray39", "gray77")) +
  labs(title = "Spectral distance between \n middle thirds of [a] & [p]",
        y = "Euclidean distance", x = "Age") +
  theme(legend.position = c(0.6, 0.88)) +
  theme(legend.title=element_text(size=12,face="bold")) +
  theme(legend.text=element_text(size=12)) +
  theme(plot.title = element_text(size = 18, face = "bold")) +
  theme(
    axis.text = element_text(face = "bold", size = 22),
    axis.title = element_text(size=22,face="bold"))

dev.off()

## pdf
## 2
```

make new age groups

```
library('plyr')

## -----
## You have loaded plyr after dplyr - this is likely to cause problems.
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:
## library(plyr); library(dplyr)
## -----
##
## Attaching package: 'plyr'
##
## The following objects are masked from 'package:dplyr':
##
##   arrange, count, desc, failwith, id, mutate, rename, summarise,
##   summarize
##
## The following object is masked from 'package:purrr':
##
##   compact
df.final$Age_group <- mapvalues(df.final$Age_group, from =
                               c('adult', '9', '10', '8', '7', '5-6'),
                               to = c("adult", "9-10", "9-10", "7-8",
                                       "7-8", "5-6"))

df.final$Age_group <- factor(df.final$Age_group,
                             levels = c("5-6", "7-8", "9-10", "adult"))

jpeg('agegroup_eucdist.jpeg', width = 500, height=500)

ggplot(df.final, aes(x = Age_group, y = euc_dist, fill = Morph_status)) +
  geom_boxplot() + guides(fill = guide_legend(title = "Position")) +
  scale_fill_manual(values=c("gray39", "gray77")) +
  ylim(4, 30) +
  labs(title = "Spectral distance between \n middle thirds of [a] & [p]",
       y = "Euclidean distance", x = "Age (in years)") +
  theme(legend.position = c(0.69, 0.88)) +
  theme(legend.title=element_text(size=12,face="bold")) +
  theme(legend.text=element_text(size=12)) +
  theme(plot.title = element_text(size = 18, face = "bold")) +
  theme(
    axis.text = element_text(face = "bold", size = 22),
    axis.title = element_text(size=22,face="bold"))

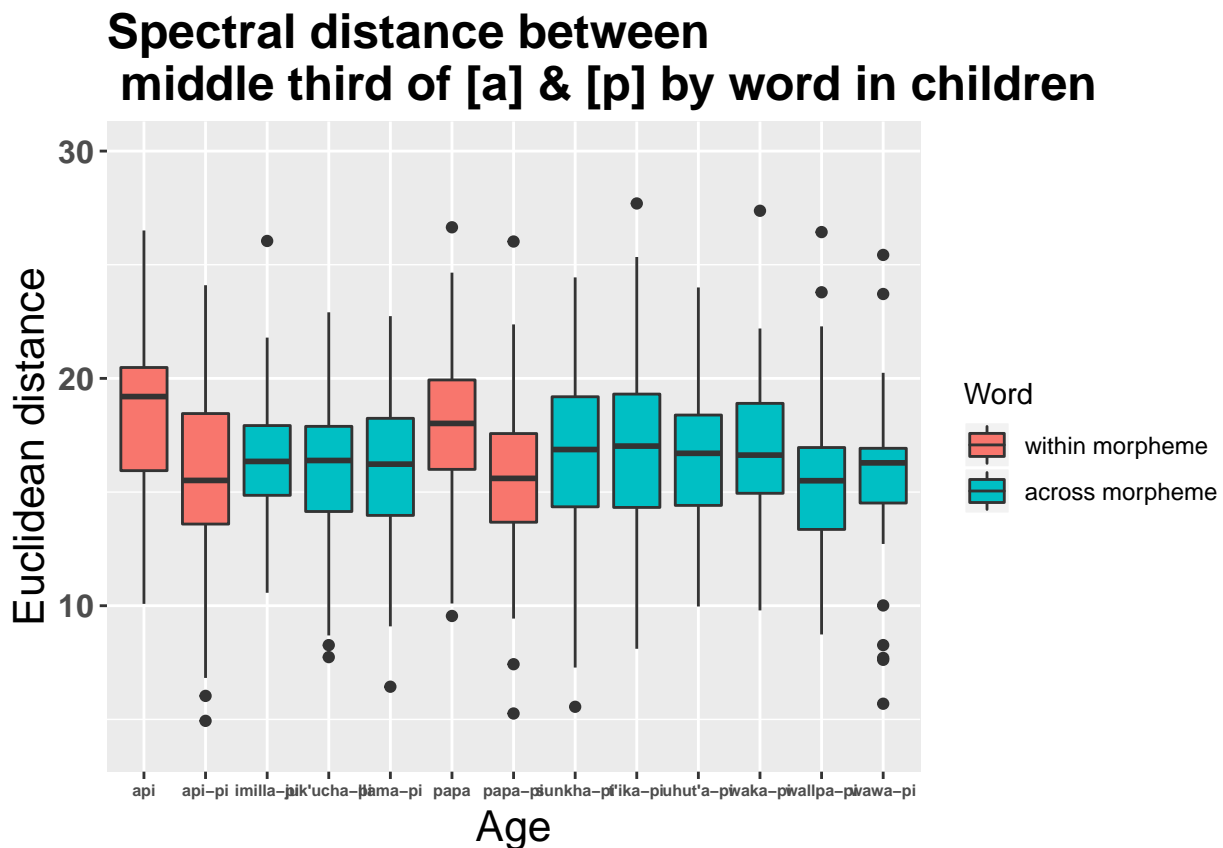
dev.off()

## pdf
## 2
```

visualize distance by word

```
adult <- subset(df.final, Age == 'adult')
child <- subset(df.final, Age == 'child')

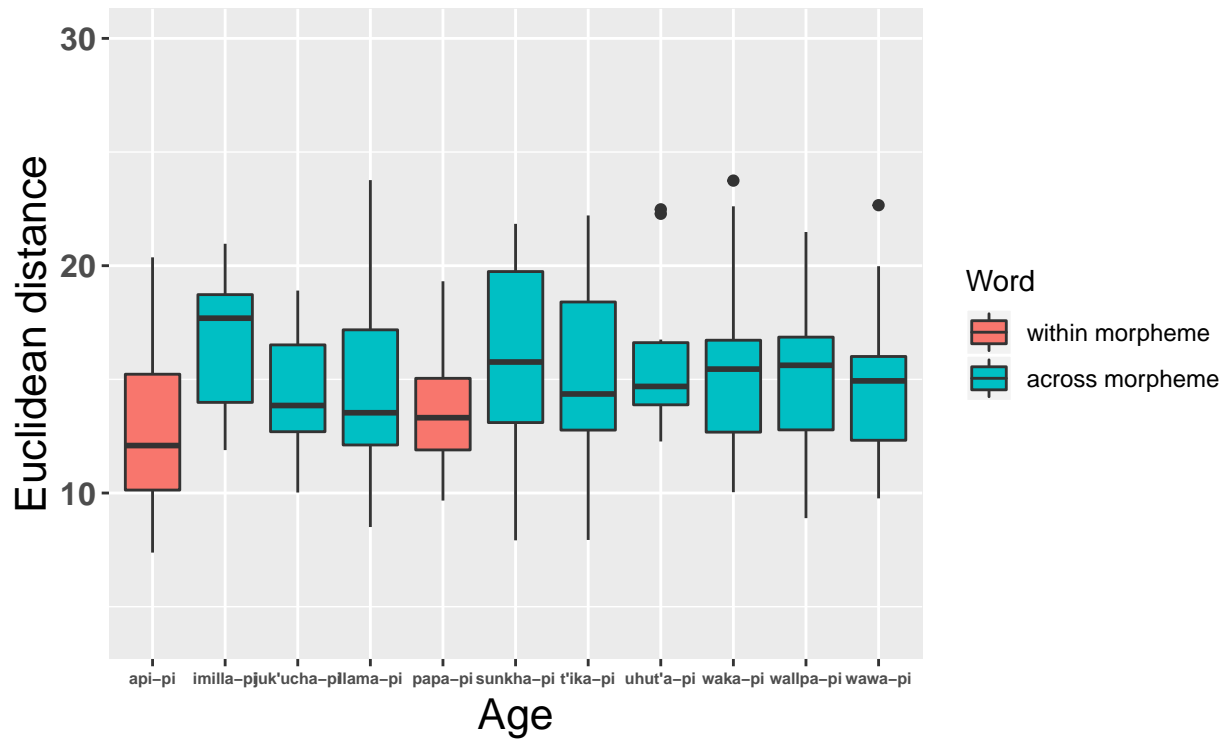
ggplot(child, aes(x = Word, y = euc_dist, fill = Morph_status)) +
  geom_boxplot() +
  guides(fill = guide_legend(title = "Word")) +
  ylim(4,30) +
  labs(title = "Spectral distance between \n middle third of [a] & [p] by word in children",
       y = "Euclidean distance", x = "Age") +
  theme(axis.title = element_text(size = 16)) +
  theme(plot.title = element_text(size = 18, face = "bold")) +
  theme(
    axis.text.x = element_text(face = "bold", size = 6),
    axis.text.y = element_text(face = "bold", size = 12)
  )
)
```



```
ggplot(adult, aes(x = Word, y = euc_dist, fill = Morph_status)) +
  geom_boxplot() + guides(fill = guide_legend(title = "Word")) +
  ylim(4,30) +
  labs(title = "Spectral distance between \n middle third of [a] & [p] by word in adults",
       y = "Euclidean distance", x = "Age") +
  theme(axis.title = element_text(size = 16)) +
  theme(plot.title = element_text(size = 18, face = "bold")) +
  theme(
```

```
axis.text.x = element_text(face = "bold", size = 6),
axis.text.y = element_text(face = 'bold', size = 12)
)
```

Spectral distance between middle third of [a] & [p] by word in adults



Fit some models

baseline and baseline + ap_duration

Less coarticulation in temporally longer segments, as expected.

```
summary(m <- lmer(euc_dist ~ + (1 | Speaker) + (1 | Word), data = df.final))

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: euc_dist ~ +(1 | Speaker) + (1 | Word)
## Data: df.final
##
## REML criterion at convergence: 4387.6
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.5288 -0.6006  0.0419  0.5391  6.6525
##
## Random effects:
##  Groups   Name      Variance Std.Dev.
## Speaker  (Intercept) 6.3026   2.5105
## Word     (Intercept) 0.4827   0.6948
## Residual                    8.5258   2.9199
## Number of obs: 855, groups: Speaker, 40; Word, 13
##
## Fixed effects:
##              Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)  16.3270     0.4538 46.4107   35.98  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

summary(mdur <- lmer(euc_dist ~ ap_duration + (1 | Speaker) + (1 | Word), data = df.final))

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: euc_dist ~ ap_duration + (1 | Speaker) + (1 | Word)
## Data: df.final
##
## REML criterion at convergence: 4372.2
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.2502 -0.5858  0.0456  0.5462  6.8238
##
## Random effects:
##  Groups   Name      Variance Std.Dev.
## Speaker  (Intercept) 6.1583   2.4816
## Word     (Intercept) 0.3908   0.6251
## Residual                    8.4481   2.9066
## Number of obs: 855, groups: Speaker, 40; Word, 13
##
## Fixed effects:
##              Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)  13.7853     0.8726 324.4931   15.799  < 2e-16 ***
```



```
## ap_duration 10.9481      3.2480 700.9051    3.371 0.000791 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr)
## ap_duration -0.862
anova(m, mdur) # improves

## Data: df.final
## Models:
## m: euc_dist ~ +(1 | Speaker) + (1 | Word)
## mdur: euc_dist ~ ap_duration + (1 | Speaker) + (1 | Word)
##      Df    AIC    BIC logLik deviance Chisq Chi Df Pr(>Chisq)
## m      4 4395.8 4414.8 -2193.9  4387.8
## mdur   5 4386.6 4410.4 -2188.3  4376.6 11.211      1 0.0008132 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Morphological position + sequence_duration

```
summary(m1.a <- lmer(euc_dist ~ Morph_status + ap_duration + (1 | Speaker) + (1 | Word), data = df.final)

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: euc_dist ~ Morph_status + ap_duration + (1 | Speaker) + (1 |
##      Word)
##      Data: df.final
##
## REML criterion at convergence: 4370.8
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.2730 -0.5848  0.0424  0.5396  6.8214
##
## Random effects:
##      Groups      Name              Variance Std.Dev.
##      Speaker  (Intercept)  6.1874      2.4874
##      Word     (Intercept)  0.3591      0.5992
##      Residual                    8.4512      2.9071
## Number of obs: 855, groups:  Speaker, 40; Word, 13
##
## Fixed effects:
##
##              Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)      13.3512    0.9470 180.3725  14.099 < 2e-16
## Morph_statusacross morpheme    0.4864    0.4297  10.2679   1.132 0.283426
## ap_duration       11.3666    3.2589  680.1911   3.488 0.000518
##
## (Intercept)          ***
## Morph_statusacross morpheme
## ap_duration          ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Correlation of Fixed Effects:
##           (Intr) Mrph_m
## Mrph_sttscm -0.394
## ap_duration -0.829  0.102
anova(mdur, m1.a) # doesn't improve

## Data: df.final
## Models:
## mdur: euc_dist ~ ap_duration + (1 | Speaker) + (1 | Word)
## m1.a: euc_dist ~ Morph_status + ap_duration + (1 | Speaker) + (1 |
## m1.a:      Word)
##      Df    AIC    BIC logLik deviance Chisq Chi Df Pr(>Chisq)
## mdur  5 4386.6 4410.4 -2188.3  4376.6
## m1.a  6 4387.3 4415.8 -2187.6  4375.3 1.3423    1    0.2466
```

Age[child v adult] + sequence_duration + Morph_status

```
summary(m1 <- lmer(euc_dist ~ Age + ap_duration + Morph_status + (1 | Speaker) + (1 | Word), data = df.

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: euc_dist ~ Age + ap_duration + Morph_status + (1 | Speaker) +
##      (1 | Word)
##      Data: df.final
##
## REML criterion at convergence: 4368.1
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.2800 -0.5887  0.0433  0.5360  6.8296
##
## Random effects:
##   Groups   Name                Variance Std.Dev.
##   Speaker  (Intercept)  6.1783     2.4856
##   Word     (Intercept)  0.3547     0.5955
##   Residual                    8.4524     2.9073
## Number of obs: 855, groups:  Speaker, 40; Word, 13
##
## Fixed effects:
##              Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)      13.6197    0.9827 184.2677  13.859 < 2e-16
## Ageadult         -0.9565    0.9371  37.4222  -1.021 0.313952
## ap_duration       11.2215    3.2616 679.0725   3.441 0.000616
## Morph_statusacross morpheme  0.4946    0.4279  10.2630   1.156 0.273964
##
## (Intercept)          ***
## Ageadult
## ap_duration          ***
## Morph_statusacross morpheme
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) Agedlt ap_drt
## Ageadult      -0.271
## ap_duration   -0.810  0.046
## Mrph_sttscm  -0.374 -0.017  0.102

anova(mdur, m1) # doesn't improve

## Data: df.final
## Models:
## mdur: euc_dist ~ ap_duration + (1 | Speaker) + (1 | Word)
## m1: euc_dist ~ Age + ap_duration + Morph_status + (1 | Speaker) +
## m1:      (1 | Word)
##      Df      AIC      BIC logLik deviance Chisq Chi Df Pr(>Chisq)
## mdur  5 4386.6 4410.4 -2188.3  4376.6
## m1    7 4388.2 4421.4 -2187.1  4374.2 2.4327      2    0.2963
```

Age*Morphological status + sequence duration

The interaction of age and morphological status improves model fit. Adults differentiate between morphological environments, but children do not. Specifically, adults coarticulate less within morphemes than across.

```
summary(m2 <- lmer(euc_dist ~ Age*Morph_status + ap_duration + (1 | Speaker) + (1 | Word), data = df.fi
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: euc_dist ~ Age * Morph_status + ap_duration + (1 | Speaker) +
##      (1 | Word)
##      Data: df.final
##
## REML criterion at convergence: 4361.5
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.3168 -0.5950  0.0427  0.5597  6.7841
##
## Random effects:
##   Groups   Name                Variance Std.Dev.
##   Speaker  (Intercept)  6.1252     2.4749
##   Word     (Intercept)  0.2849     0.5338
##   Residual                    8.4205     2.9018
## Number of obs: 855, groups:  Speaker, 40; Word, 13
##
## Fixed effects:
##
##              Estimate Std. Error      df t value
## (Intercept)      13.7091    0.9687 195.3621  14.152
## Ageadult         -1.9369    1.0126  51.7836  -1.913
## Morph_statusacross morpheme      0.2352    0.4116  12.0485   0.571
## ap_duration      11.5086    3.2381 622.0922   3.554
## Ageadult:Morph_statusacross morpheme  1.2937    0.5222 773.6214   2.477
##
##              Pr(>|t|)
## (Intercept)      < 2e-16 ***
## Ageadult         0.061303 .
## Morph_statusacross morpheme  0.578213
## ap_duration      0.000408 ***
## Ageadult:Morph_statusacross morpheme 0.013449 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) Agedlt Mrph_m ap_drt
## Ageadult      -0.275
## Mrph_sttscom -0.361  0.088
## ap_duration -0.814  0.037  0.097
## Agdlt:Mrp_m  0.058 -0.388 -0.269  0.016
```

```
anova(mdur, m2) # improves
```

```
## Data: df.final
## Models:
## mdur: euc_dist ~ ap_duration + (1 | Speaker) + (1 | Word)
## m2: euc_dist ~ Age * Morph_status + ap_duration + (1 | Speaker) +
```

```
## m2:      (1 | Word)
##      Df      AIC      BIC logLik deviance Chisq Chi Df Pr(>Chisq)
## mdur  5 4386.6 4410.4 -2188.3  4376.6
## m2    8 4383.9 4422.0 -2184.0  4367.9 8.6575      3  0.03421 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Morphological status*sequence_duration

No reliable interaction of morphological status and sequence duration.

```
summary(m4 <- lmer(euc_dist ~ Morph_status*ap_duration + (1 | Speaker) + (1 | Word), data = df.final))
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: euc_dist ~ Morph_status * ap_duration + (1 | Speaker) + (1 |
##      Word)
##      Data: df.final
##
## REML criterion at convergence: 4365
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.2730 -0.5898  0.0440  0.5392  6.8025
##
## Random effects:
##      Groups      Name      Variance Std.Dev.
##      Speaker (Intercept) 6.2070   2.4914
##      Word    (Intercept) 0.3143   0.5606
##      Residual              8.4652   2.9095
## Number of obs: 855, groups: Speaker, 40; Word, 13
##
## Fixed effects:
##
##              Estimate Std. Error    df
## (Intercept)      12.651      1.265 140.284
## Morph_statusacross morpheme      1.529      1.376 170.888
## ap_duration      14.241      4.797 235.964
## Morph_statusacross morpheme:ap_duration    -4.392      5.584 415.647
##
##              t value Pr(>|t|)
## (Intercept)      9.999 <2e-16 ***
## Morph_statusacross morpheme      1.111   0.2682
## ap_duration      2.969   0.0033 **
## Morph_statusacross morpheme:ap_duration   -0.786   0.4320
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) Mrph_m ap_drt
## Mrph_sttscm -0.725
## ap_duration -0.912  0.723
## Mrph_mrph:_  0.670 -0.955 -0.735
```

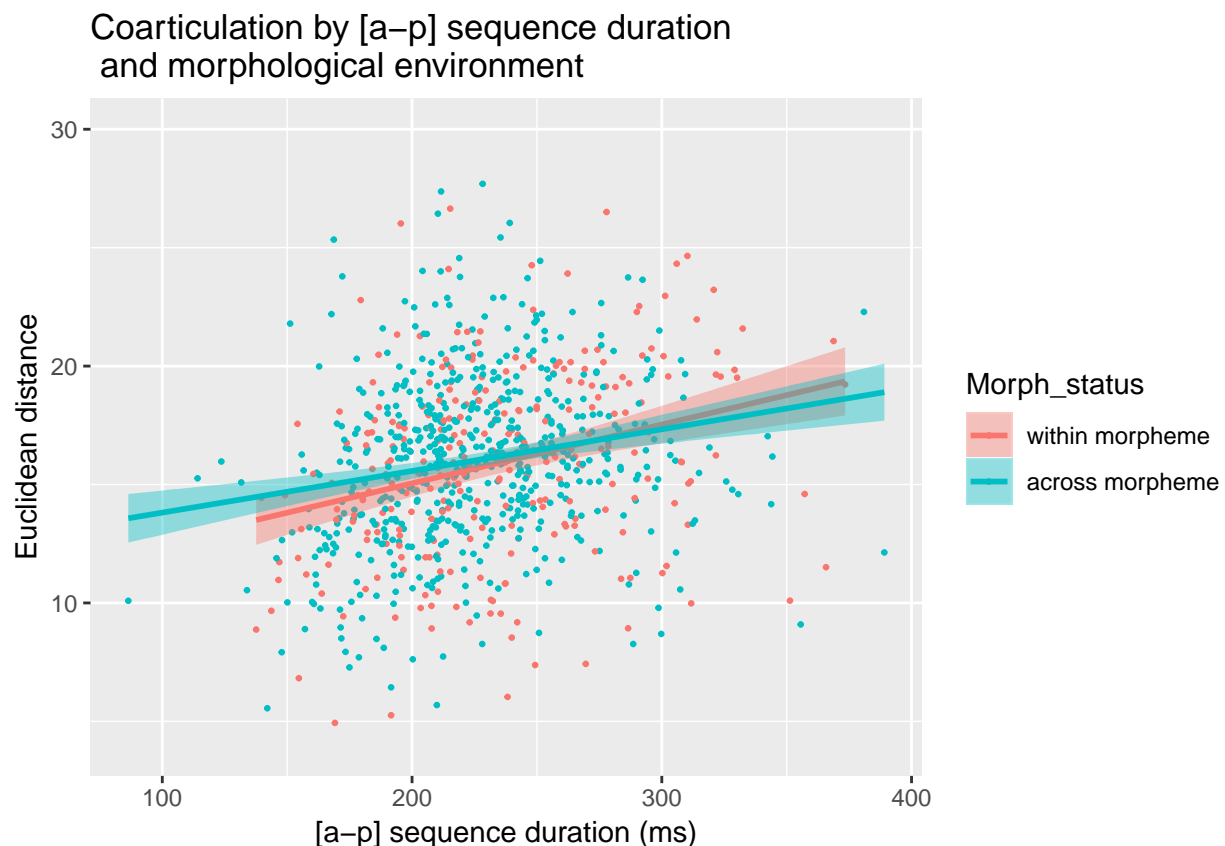
```
anova(mdur, m4) # doesn't improve
```

```
## Data: df.final
```

```
## Models:
## mdur: euc_dist ~ ap_duration + (1 | Speaker) + (1 | Word)
## m4: euc_dist ~ Morph_status * ap_duration + (1 | Speaker) + (1 |
## m4: Word)
##      Df      AIC      BIC logLik deviance Chisq Chi Df Pr(>Chisq)
## mdur  5 4386.6 4410.4 -2188.3  4376.6
## m4     7 4388.6 4421.8 -2187.3  4374.6 2.0519      2      0.3585
```

There is no relationship between segment duration and morphological environment - just less coarticulation in longer segments in all environments.

```
ggplot(df.final, aes(ap_duration*1000, euc_dist, color=Morph_status, fill=Morph_status)) +
  geom_point(size=.5) +
  geom_smooth(method = "lm") +
  labs(x = "[a-p] sequence duration (ms)", y = "Euclidean distance") +
  ggtitle("Coarticulation by [a-p] sequence duration \n and morphological environment") +
  ylim(4, 30)
```



Duration by morphological status by age

This model shows that there is a three-part interaction between morphological position, segment duration, and age.

```
summary(m5 <- lmer(euc_dist ~ Morph_status*ap_duration*Age + (1 | Speaker) + (1 | Word), data = df.fina
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: euc_dist ~ Morph_status * ap_duration * Age + (1 | Speaker) +
##      (1 | Word)
##      Data: df.final
##
## REML criterion at convergence: 4334.8
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -3.3662 -0.5902  0.0402  0.5498  6.8533
##
## Random effects:
##      Groups      Name              Variance Std.Dev.
## Speaker  (Intercept)  6.1877      2.4875
## Word     (Intercept)  0.2101      0.4583
## Residual                    8.3784      2.8945
## Number of obs: 855, groups: Speaker, 40; Word, 13
##
## Fixed effects:
##
##              Estimate Std. Error
## (Intercept)      11.655      1.391
## Morph_statusacross morpheme      2.862      1.570
## ap_duration      19.869      5.251
## Ageadult         4.713      2.596
## Morph_statusacross morpheme:ap_duration    -10.852      6.403
## Morph_statusacross morpheme:Ageadult       -6.346      2.808
## ap_duration:Ageadult    -31.082     11.113
## Morph_statusacross morpheme:ap_duration:Ageadult  35.487     12.805
##
##              df t value Pr(>|t|)
## (Intercept)    147.652   8.381 3.79e-14
## Morph_statusacross morpheme    240.771   1.823 0.069497
## ap_duration    184.106   3.784 0.000209
## Ageadult       669.169   1.815 0.069914
## Morph_statusacross morpheme:ap_duration    418.976  -1.695 0.090842
## Morph_statusacross morpheme:Ageadult       801.061  -2.260 0.024078
## ap_duration:Ageadult       709.658  -2.797 0.005299
## Morph_statusacross morpheme:ap_duration:Ageadult  781.865   2.771 0.005714
##
## (Intercept)          ***
## Morph_statusacross morpheme      .
## ap_duration          ***
## Ageadult              .
## Morph_statusacross morpheme:ap_duration      .
## Morph_statusacross morpheme:Ageadult          *
## ap_duration:Ageadult          **
## Morph_statusacross morpheme:ap_duration:Ageadult **
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##          (Intr) Mrph_m ap_drt Agedlt Mr_m:_ Mr_m:A ap_d:A
## Mrph_sttscm -0.685
## ap_duration -0.919  0.699
## Ageadult    -0.472  0.317  0.438
## Mrph_mrph:_  0.642 -0.971 -0.698 -0.306
## Mrph_mrph:A  0.327 -0.507 -0.344 -0.715  0.507
## ap_drtn:Agd  0.413 -0.315 -0.446 -0.920  0.313  0.756
## Mrph_mr:_:A -0.305  0.472  0.329  0.704 -0.484 -0.982 -0.775
anova(mdur, m5) # improves

## Data: df.final
## Models:
## mdur: euc_dist ~ ap_duration + (1 | Speaker) + (1 | Word)
## m5: euc_dist ~ Morph_status * ap_duration * Age + (1 | Speaker) +
## m5:      (1 | Word)
##      Df    AIC    BIC logLik deviance Chisq Chi Df Pr(>Chisq)
## mdur  5 4386.6 4410.4 -2188.3  4376.6
## m5    11 4380.9 4433.1 -2179.4  4358.9 17.738      6  0.006922 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```


What does this three part interaction look like?

Once again, adults distinguish between the two environments and children do not. Specifically, adults coarticulate less in longer duration segments across a morpheme boundary. But adults don't really change their coarticulatory behavior by segment duration within a morpheme. This could be some sort of planning effect.

You plan morphologically-complex words online, right? So you see a “hyper” planning effect in words of longer duration.

The words are longer maybe because they took longer to plan. They took longer to plan because the inflection of the word+suffix is less frequent/less used/less familiar. The inflection is less familiar and so speakers coarticulate less between the word and the suffix. It actually kind of makes sense that there isn't a relationship between in the kids - they don't have the experience.

I think in an ideal world, I would have frequency statistics for these inflections. But I don't and I have yet to hear a good suggestion for how to get them (adult ratings? - don't reflect children's experience. lexical statistics from the bible? - biased and unnatural language). Maybe one day once the child-directed speech in the Quechua corpus is annotated.

```
#jpeg("interaction.jpg", width = 500, height = 500)

ggplot(df.final, aes(ap_duration*1000, euc_dist, color=Morph_status, fill=Morph_status)) +
  geom_point(size=.5) +
  geom_smooth(method = "lm") +
  labs(x = "[a-p] sequence Duration (ms)", y = "Euclidean distance") +
  facet_grid(~Age) +
  ggtitle("[a]-[p] coarticulation by sequence duration, \n morphological environment, and age") +
  ylim(4, 30)
```

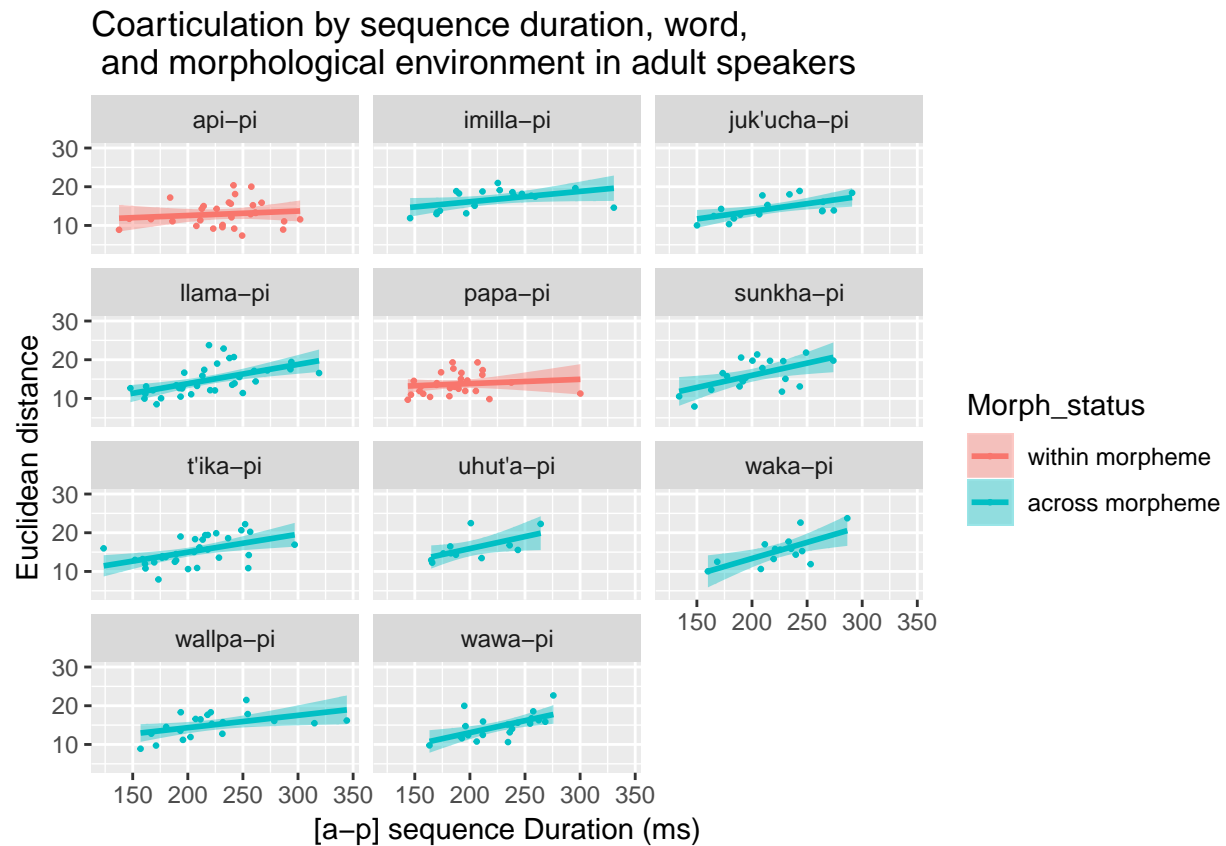


```
#dev.off()
```

I guess this could be conflated with word position. Quechua only has suffixes so all the tokens in the across-morpheme condition are word-final and all the tokens in the within-morpheme condition are word-initial. But word-initial segments are more emphasized and phonetically salient (less reduction, etc.) so if anything you would think you would see longer duration in the within-morpheme segments. Also stress was consistent between the two conditions.

```
adultdf <- df.final %>%
  filter(Age=='adult')
#jpeg("byword.jpg", width = 750, height = 750)

ggplot(adultdf, aes(ap_duration*1000, euc_dist, color=Morph_status, fill=Morph_status)) +
  geom_point(size=.5) +
  geom_smooth(method = "lm") +
  labs(x = "[a-p] sequence Duration (ms)", y = "Euclidean distance") +
  facet_wrap(~Word, nrow=4) +
  ggtitle("Coarticulation by sequence duration, word, \n and morphological environment in adult speakers")
ylim(4, 30)
```



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