Global TIMIT Thai and /aj/ raising

Jonathan D. Wright

University of Pennsylvania, Linguistic Data Consortium

jdwright@ldc.upenn.edu

NWAV AP 7 Dec. 16th, 2022

Original TIMIT

- Global TIMIT is an initiative at LDC to create TIMIT like corpora in a variety of languages
- The original TIMIT was an expensive undertaking that created a valuable speech sample of American English widely used in engineering
- TIMIT involved 630 speakers reading 10 sentences each, from a variety of dialect areas in the US
- Sentences were in three categories regarding how many different people read them: shared by all, shared by few, unique to one speaker
- Word and phone level transcriptions were produced

Global TIMIT Thai

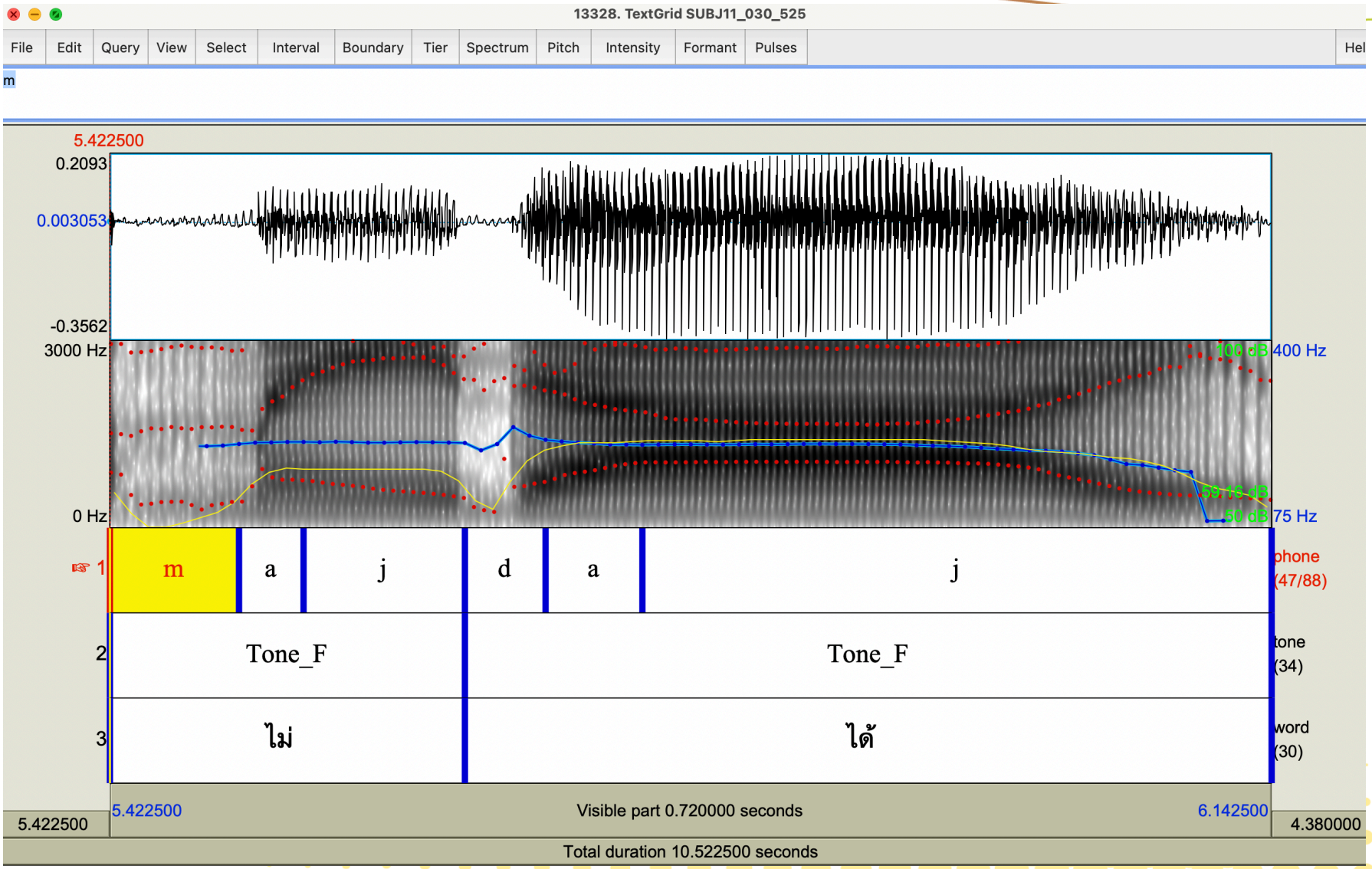
- The Global TIMIT Thai recordings were collected in Thailand in 2016 by Nattanun (Pleng) Chanchaochai for LDC, and is one of the December 2022 publications
- 50 speakers recorded 120 sentences each for 6000 total, producing roughly the same amount of speech as TIMIT, but fewer speakers for lower cost
- Like TIMIT, there are three categories of sentence: shared by all, shared by few, unique to one speaker
- Basic demographics are included like date of birth, sex, etc.
- Speakers are categorized into the 4 commonly used dialect areas of Central, North, Northeast, and South, with additional notes on linguistic background

Global TIMIT Thai

- A forced aligner was developed shortly after collection by Jiahong Yuan to create timestamps for the words, phones, and tones.
- Some manual correction of timestamps and labels was done
- The timestamps are included in the corpus as plain text labels as well as Praat TextGrids

Goals

- Choose a problem of the right scope where Global TIMIT Thai might provide some insight
- Attempt automatic analysis to the extent possible
- I chose a specific pattern that I called here /aj/ raising:
  - In casual speech, /aj/ is often pronounced [e]
  - For example: ไม่ได้, “not able/ok”, /maj daj/, [me daj]
  - No mention (that I’ve seen) of this in popular or linguistic descriptions, so the pattern is unknown (to me)
  - What can we learn from the corpus?

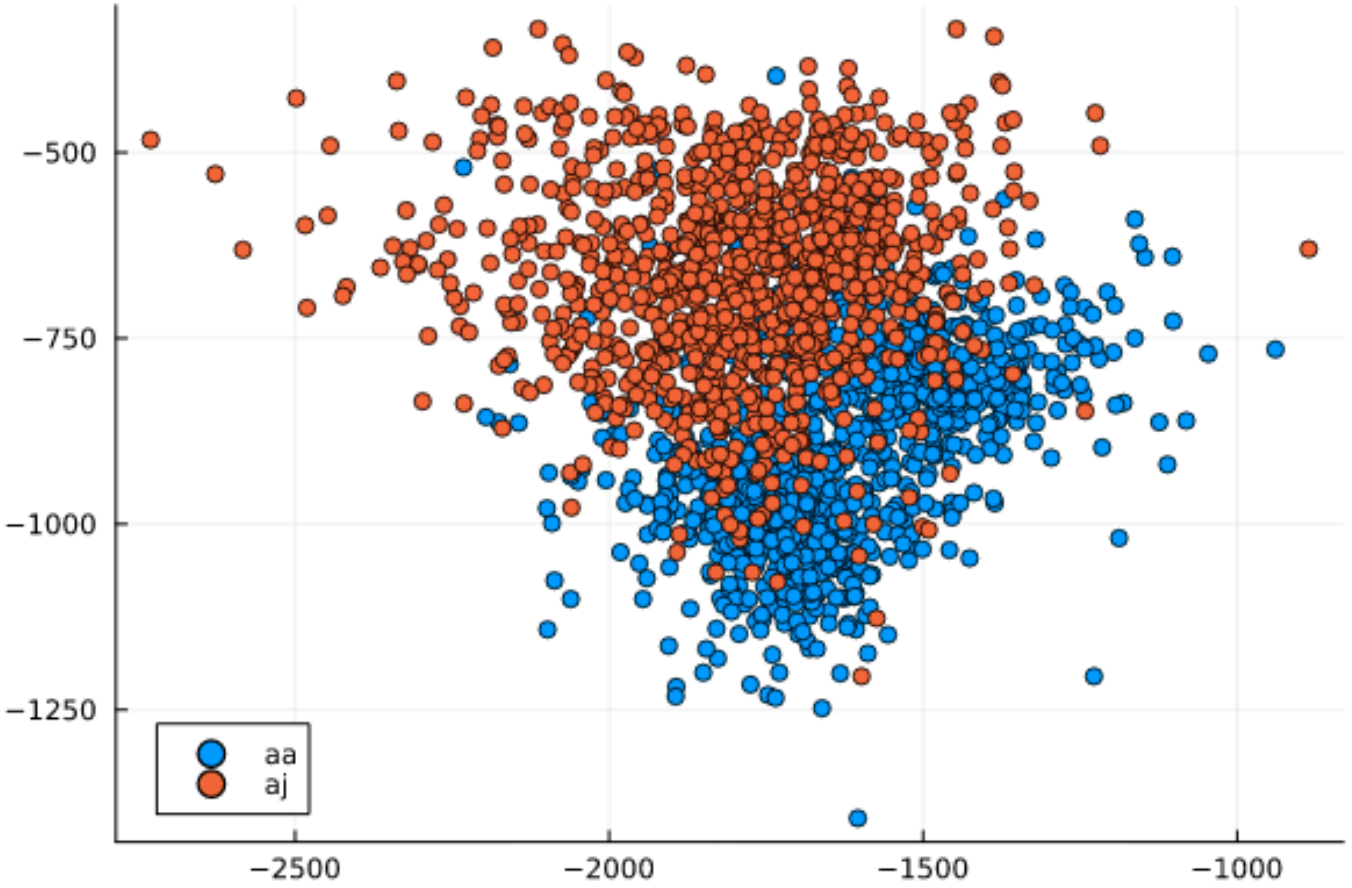

13328. TextGrid SUBJ11_030_525
File Edit Query View Select Interval Boundary Tier Spectrum Pitch Intensity Formant Pulses Hel
m
5.422500
0.2093
0.003053
-0.3562
3000 Hz
0 Hz
100 dB
400 Hz
59.15 dB
50 dB
75 Hz
1
m a j d a j
phone
(47/88)
2
Tone_F Tone_F
tone
(34)
3
ไม่ ได้
word
(30)
5.422500
Visible part 0.720000 seconds
6.142500
5.422500
4.380000
Total duration 10.522500 seconds

Methods

- Praat script was used to extract formants for all files ahead of time, using the suggested settings for male and female
- Label files from GT Thai were used to search for words and phones of interest, using the timestamps to extract formant values from Praat output
- To normalize across speakers we took advantage of the fact that variation is along the front diagonal of the vowel space
- Using the 2D means of /i:/ and /a:/ for each speaker, a single dimension was used to evaluate raising
- In other words, what fraction of the front diagonal had the nucleus for /aj/ moved?
- 5th and 95th percentiles used to remove outliers
- Plots and statistics done in Julia

Overview

- As a first look, you can see the tokens of /aj/ only partially overlap with /aa/ (tokens are unnormalized)

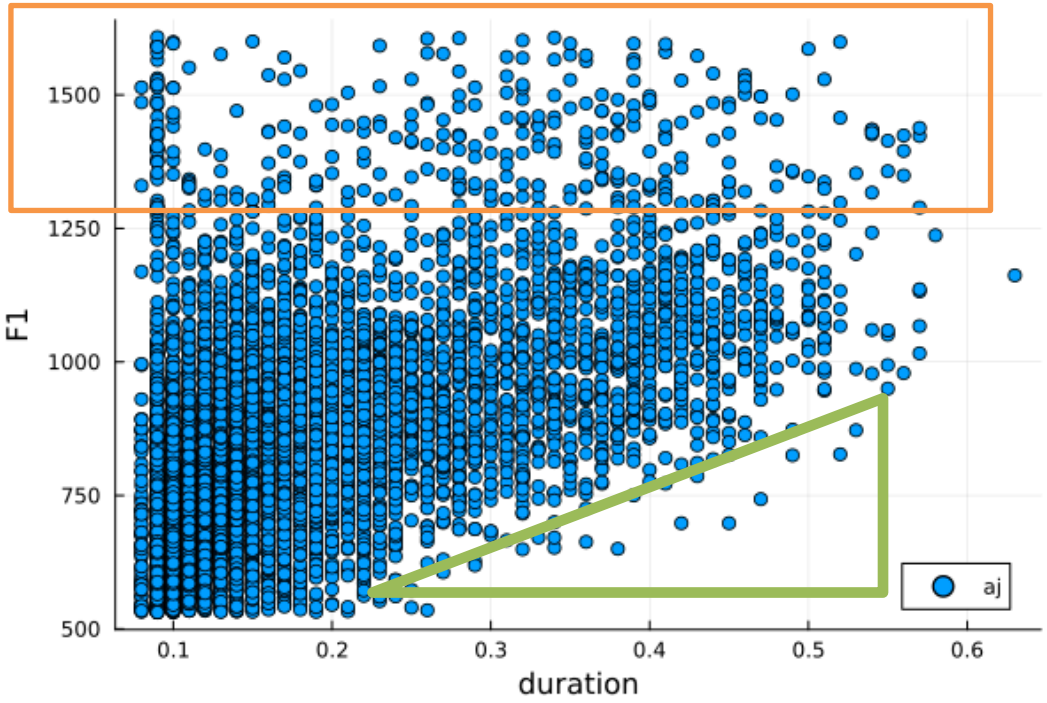


Overview

- Superficially (to the ear) there appears to be an alternation between [aj] and [e]
- Initial pooling of all data with no normalization didn't suggest a bimodal distribution looking at various plots
- Manual coding also revealed a lot of ambiguity
  - Does speech rate obscure the pattern?
- In short, it's unclear whether the variable is discrete or continuous

Duration

- Stress appears to be a strong factor; Thai is generally iambic, and [e] tends to appear in a weak position
- This can be demonstrated by looking at the length of the tokens

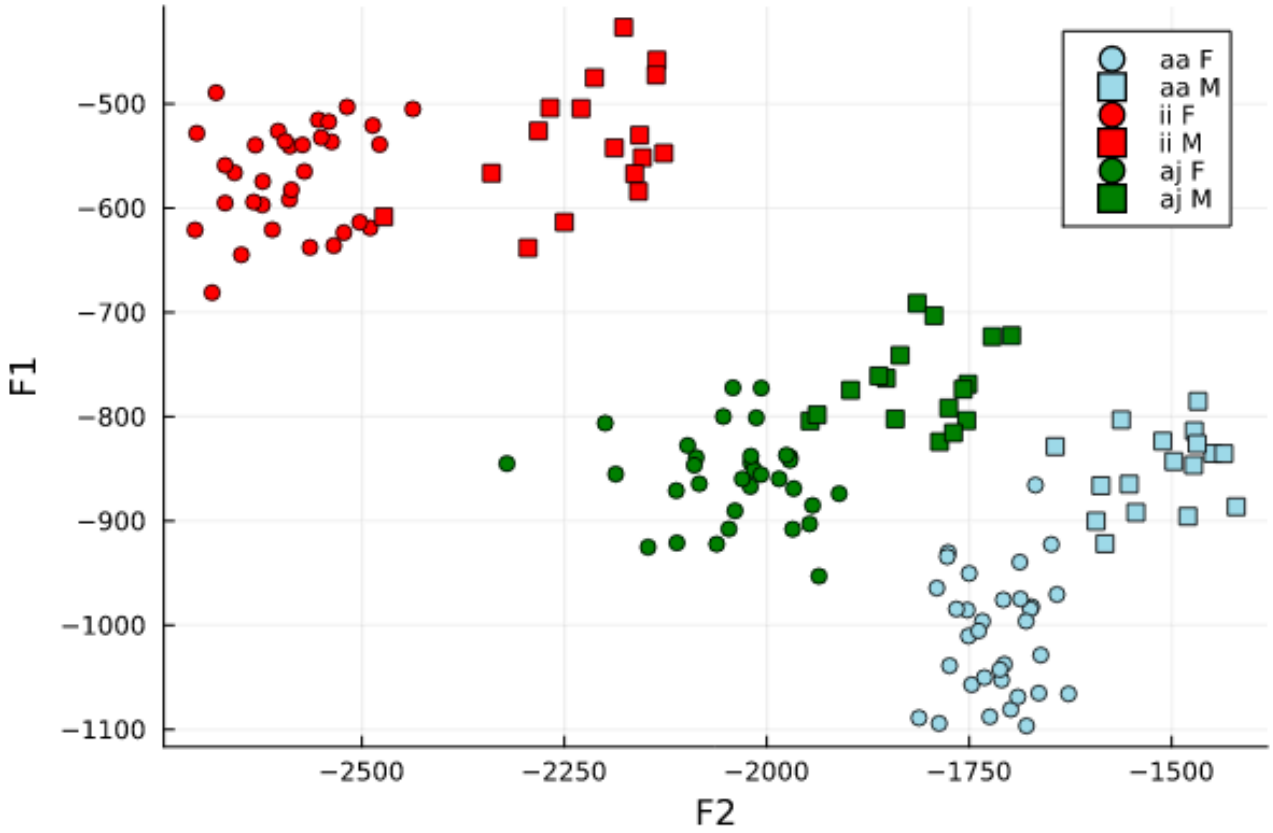


outliers, impossible F1

Gap where long, raised tokens would be

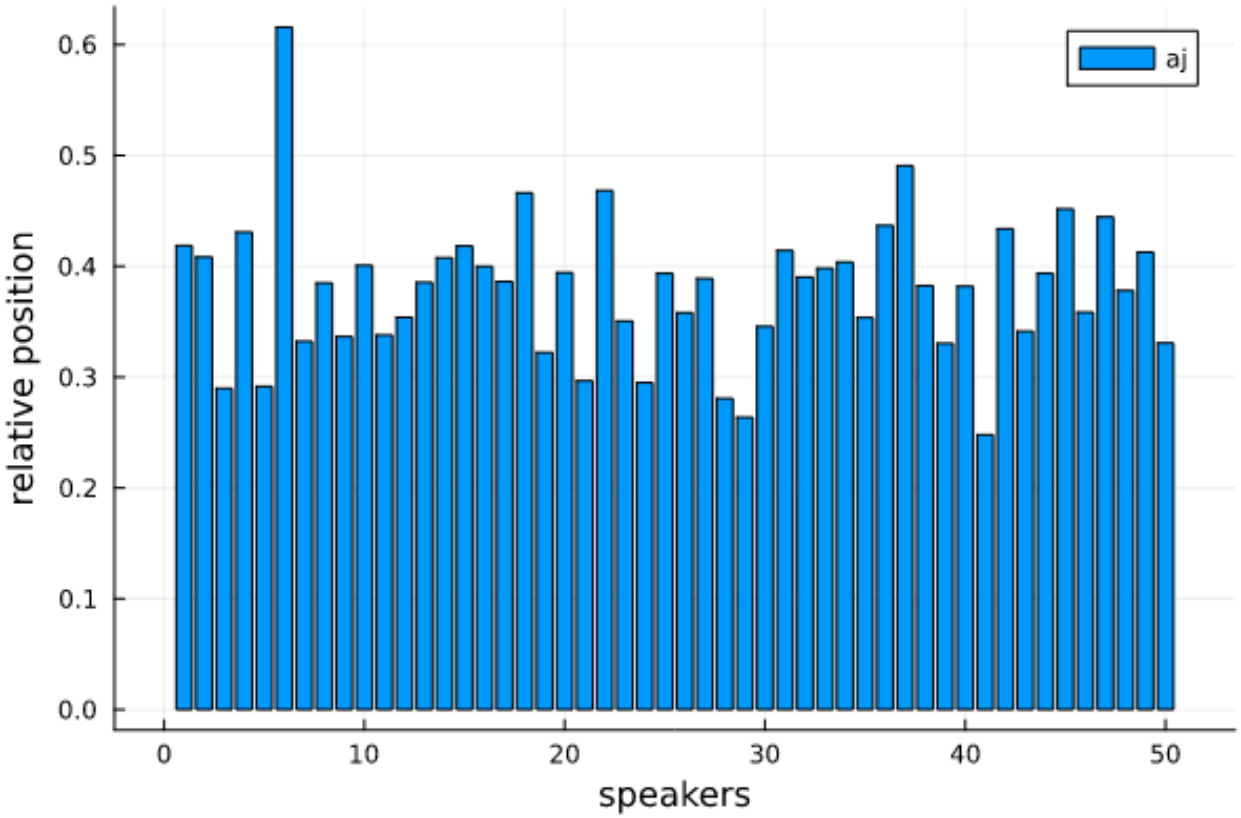
Speaker means

- Situate mean /aj/ values along the front diagonal (green)
- Tokens unnormalized, so sex separation is clear



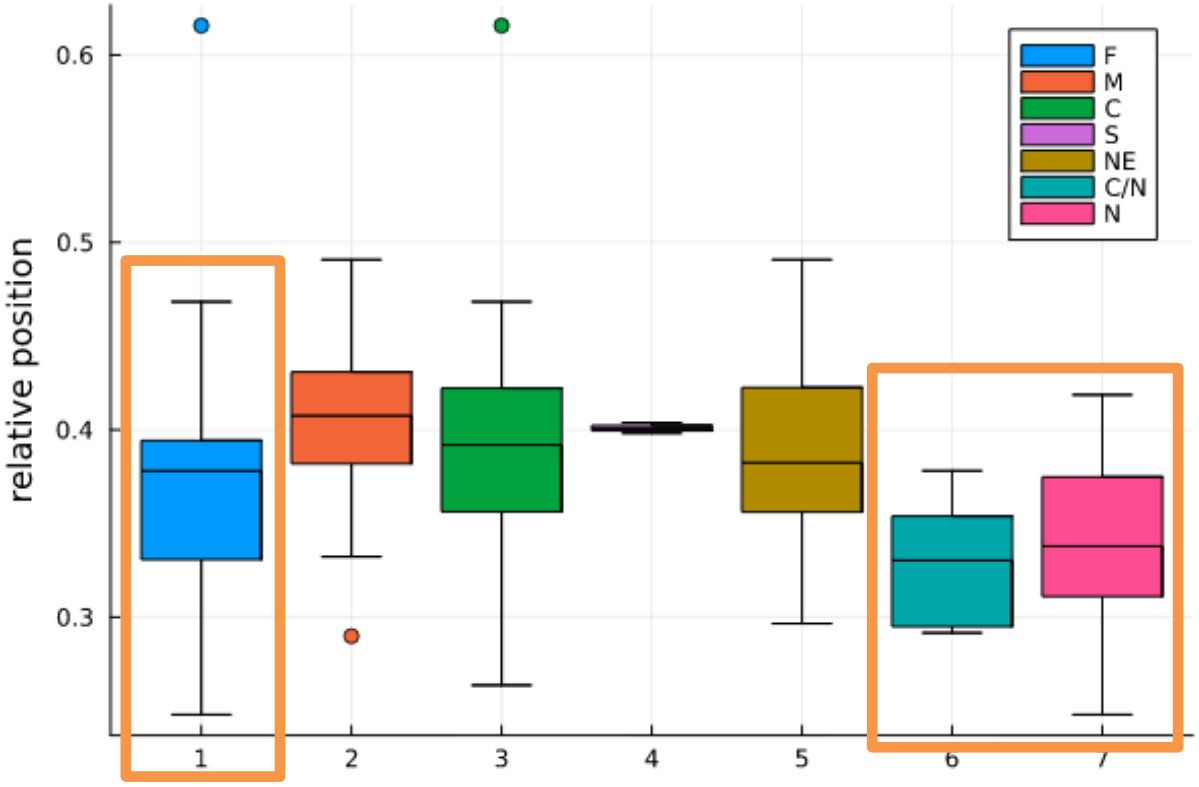
Normalized raising

- Normalized means, from 0 to 1 (/aa/ to /ii/)



Grouped means

- Normalized means, by sex and region
- sex = F and region = N are possibly conservative



Multiple regression

- Multiple regression suggests significance at p < .05 or better for both sex and region…

```
ajdistc ~ 1 + region + sex

Coefficients:
```

	Coef.	Std. Error	t	Pr(>|t|)	Lower 95%	Upper 95%
(Intercept)	0.385194	0.0118193	32.59	<1e-31	0.361373	0.409014
region: C/N	-0.0639856	0.0285179	-2.24	0.0299	-0.12146	-0.00651152
region: N	-0.0702835	0.0257265	-2.73	0.0090	-0.122132	-0.0184352
region: NE	-0.0201651	0.0257265	-0.78	0.4373	-0.0720133	0.0316832
region: S	-0.0129308	0.0366969	-0.35	0.7262	-0.0868885	0.0610268
sex: M	0.0429808	0.0189476	2.27	0.0283	0.00479439	0.0811671

T-test, sex

- … but T-test suggests sex isn't quite significant

```
Two sample t-test (equal variance)
----------------------------------
Population details:
    parameter of interest:   Mean difference
    value under h_0:         0
    point estimate:          -0.0327839
    95% confidence interval: (-0.07048, 0.004917)

Test summary:
    outcome with 95% confidence: fail to reject h_0
    two-sided p-value:           0.0868

Details:
    number of observations:   [33,17]
    t-statistic:              -1.7484081465632126
    degrees of freedom:       48
    empirical standard error: 0.018750705381951043
```

```
EqualVarianceTTest(jjjc.ajdistc[jjjc.sex .== "F"], jjjc.ajdistc[jjjc.sex .== "M"])
```

T-test, region

- Northern (N + C/N) does appear significant at p < .01

```
Two sample t-test (equal variance)
----------------------------------
Population details:
    parameter of interest:   Mean difference
    value under h_0:         0
    point estimate:          -0.0585881
    95% confidence interval: (-0.09822, -0.01895)

Test summary:
    outcome with 95% confidence: reject h_0
    two-sided p-value:           0.0046

Details:
    number of observations:   [12,38]
    t-statistic:              -2.9721320224730285
    degrees of freedom:       48
    empirical standard error: 0.019712491034005 84
```

```
EqualVarianceTTest(jjjc.ajdistc[(jjjc.region .== "N") .| (jjjc.region .== "C/N")],jjjc.ajdistc[(jjjc.region .!= "N") .& (jjjc.region .!= "C/N")])
```

Manual coding

- The drawback to the above approach is that means obscure what might be a discrete alternation
- Next step was manual coding of ~ 10K tokens of /aj/
- Tokens categorized as [aj] or [e]
- There were ambiguous or unintelligible tokens, but this generally worked out
- For tokens that truly seemed intermediate, I chose [aj], since the null hypothesis is no raising
- The previous approach allowed for continuous or discrete, but this approach really assumes a discrete alternation

Chi Square, sex

- Chi Square using sex, no significance

```
2×2 Named Matrix{Int64}
     sex ╲ v │ String3("aj")  String3("ej")
─────────────┼──────────────────────────────
String1("F") │          5942           1099
String1("M") │          2997            597
```

```
Pearson's Chi-square Test
-------------------------
Population details:
    parameter of interest:   Multinomial Probabilities
    value under h_0:         [0.556478, 0.284048, 0.105
    point estimate:          [0.558721, 0.281805, 0.103
    95% confidence interval: [(0.5467, 0.5707), (0.271,

Test summary:
    outcome with 95% confidence: fail to reject h_0
    one-sided p-value:           0.1817
```

Chi Square, region

- Chi Square using region, significant p <<< .001
- Less frequent in the North and East

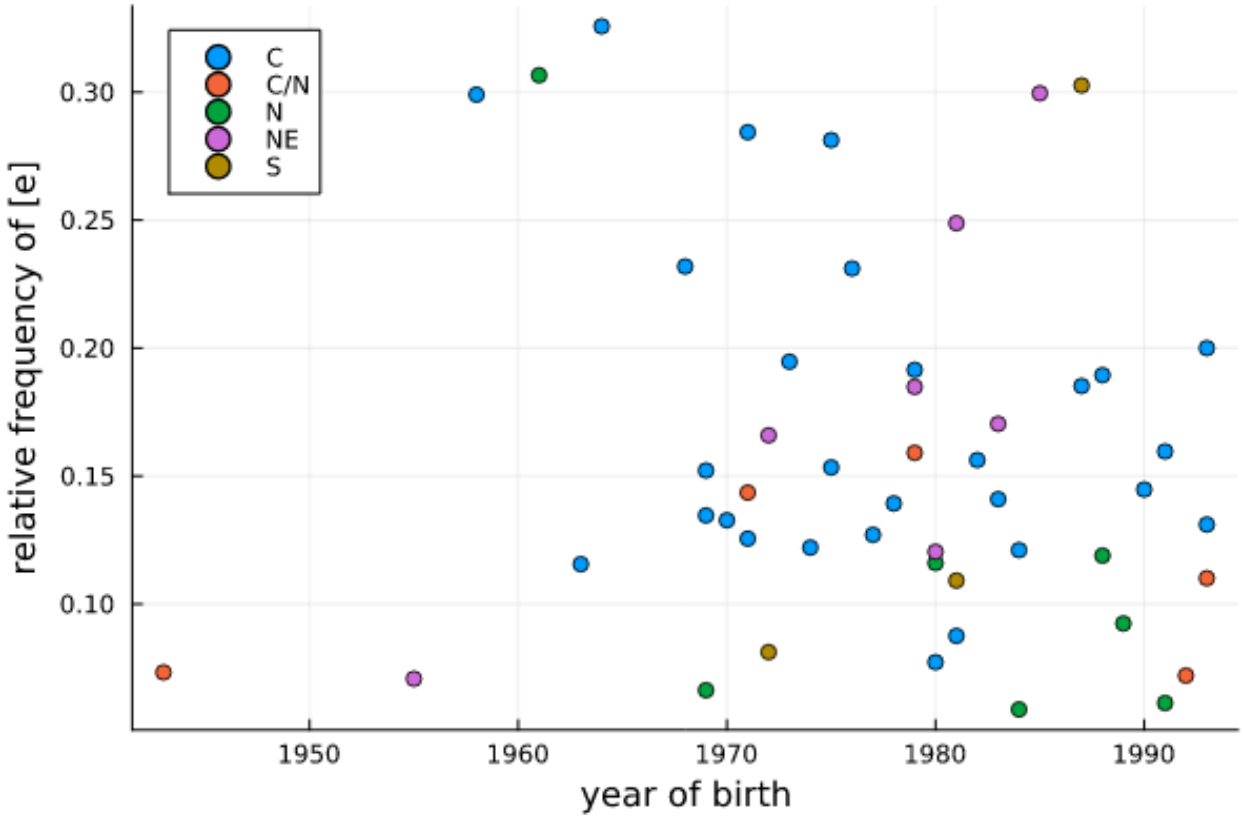
```
5×2 Named Matrix{Int64}
    region ╲ v │ String3("aj")  String3("ej")
───────────────┼──────────────────────────────
String3("C")   │          4977           1029
String3("C/N") │           954            120
String3("N")   │          1227            164
String3("NE")  │          1237            273
String3("S")   │           544            110
```

```
Pearson's Chi-square Test
-------------------------
Population details:
    parameter of interest:   Multinomial Probabilities
    value under h_0:         [0.474678, 0.0848825, 0.109
    point estimate:          [0.467983, 0.0897038, 0.115
    95% confidence interval: [(0.4544, 0.4816), (0.08223

Test summary:
    outcome with 95% confidence: reject h_0
    one-sided p-value:           <1e-09
```

Rel. Freq. by speaker

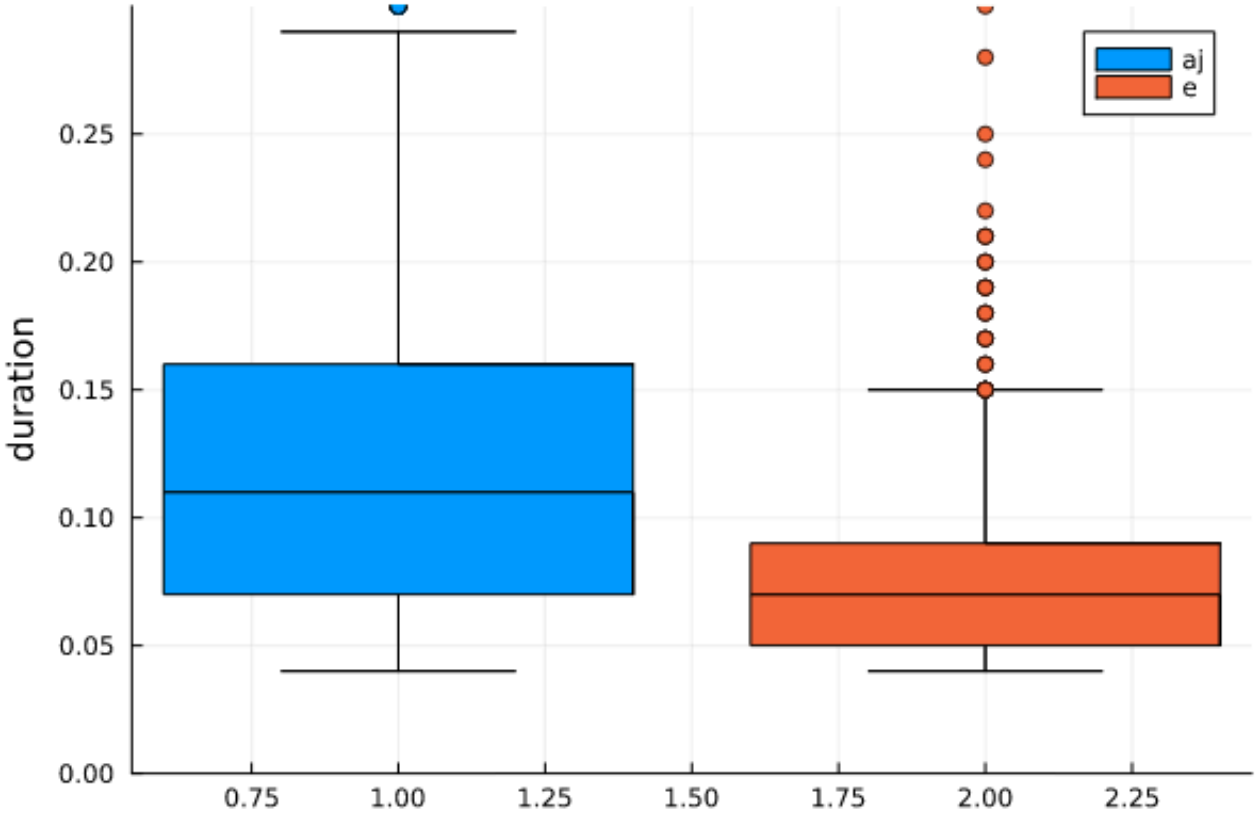
- Bars show relative frequency of raising, max is 1.0
- No apparent age grading / change in progress



some speakers use [e] very infrequently

Durations

- Returning to the beginning, is stress a factor?
- Durations of [e] are shorter

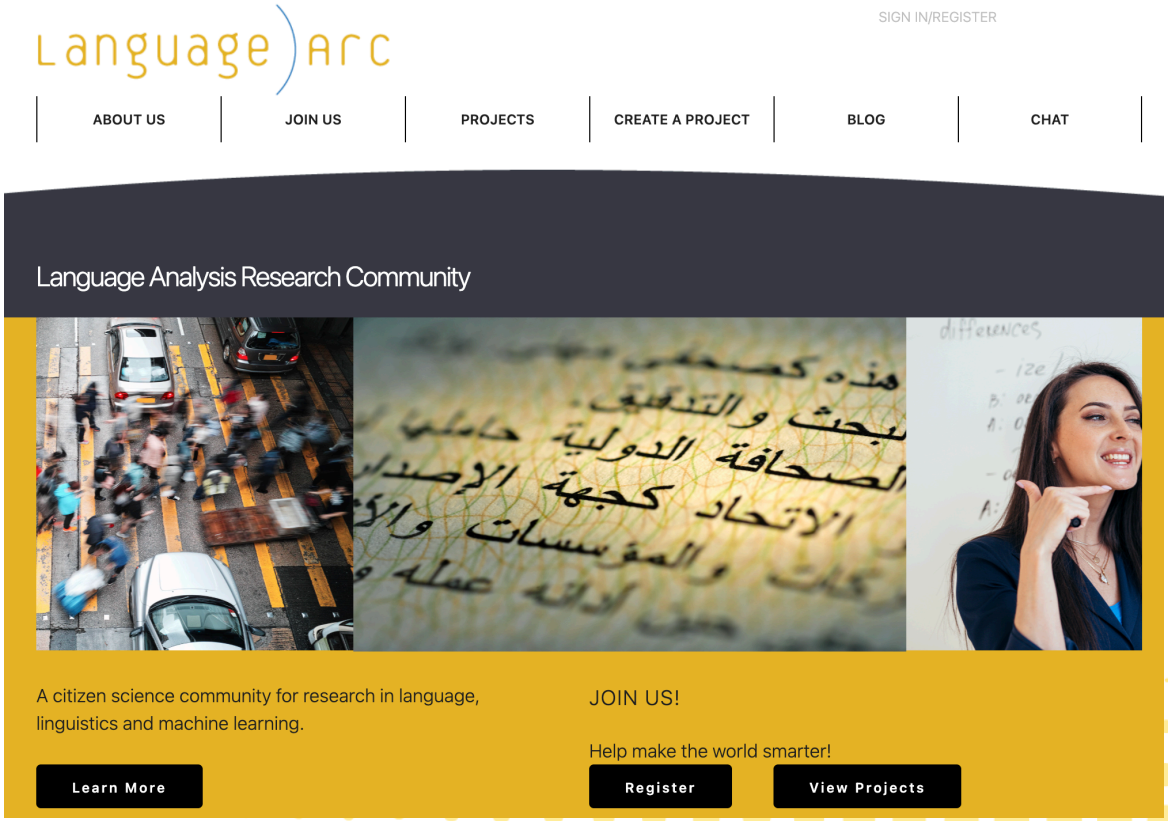


Discussion

- Slight evidence for sex and regional differences, but not exactly compelling
- Possible reasons for apparent patterns
  - By product of internal factors like stress/prosody
  - By product of lexical accident; for example, ไม่, /maj/, “not”, for syntactic reasons, is not only frequent, but frequently in the unstressed position
  - The sample is not well balanced regionally: e.g. 28 speakers are Central, and 3 are South
- On the other hand, maybe the corpus isn’t adequate for this variable
  - Given that some speakers rarely use [e], the relevant condition may not have been identified yet
  - Read speech may not be adequate (rather than spontaneous speech)
- For next steps, I’ll be looking more at stress/prosody as a condition

LanguageARC

- LDC has launched LanguageARC, a citizen linguist portal
- Further Thai data will be collected here



Linguistic Data Consortium

Thank You!
ขอบคุณครับ