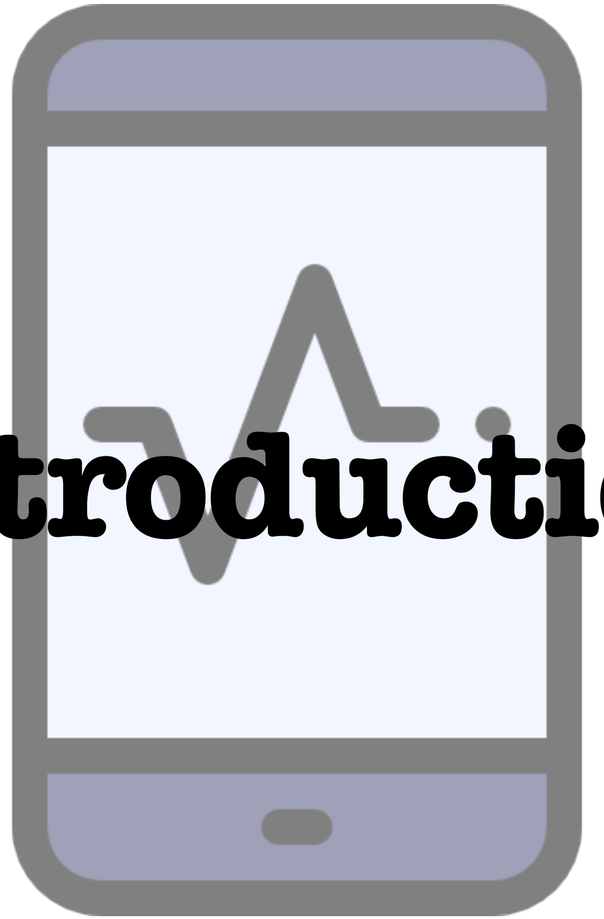
Voice onset time in English voiceless initial stops in read and spontaneous speech of Thai students with English as a second language

Chanakan Wittayasakpan

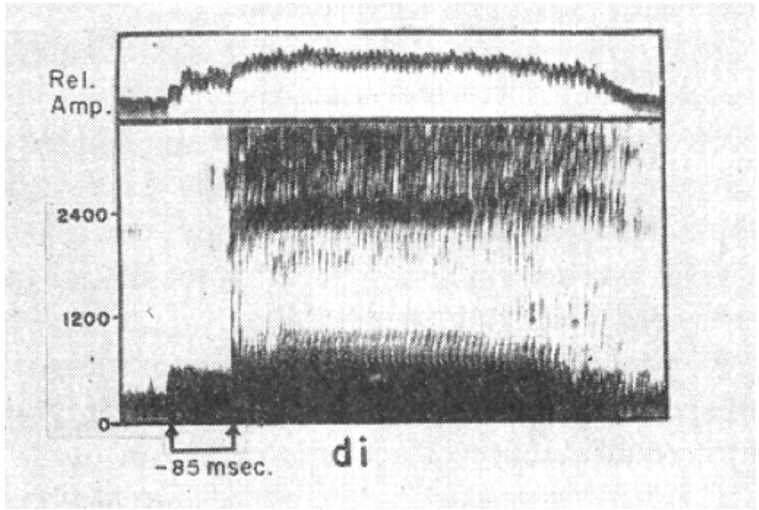


Introduction

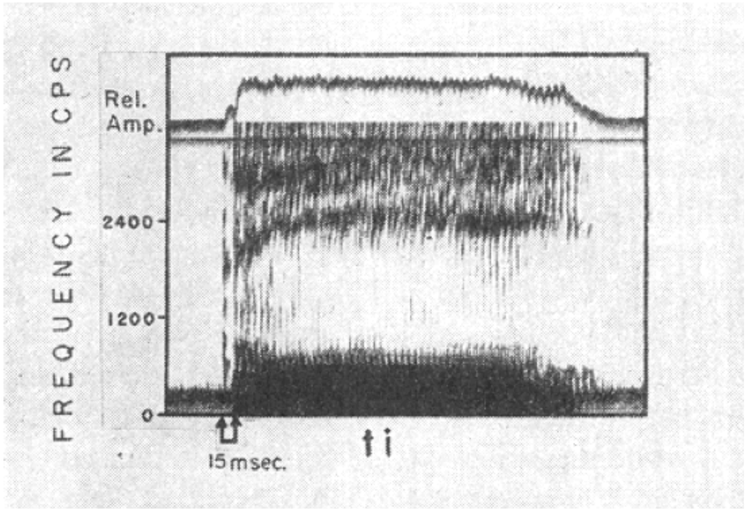
Voice onset time

- The time between the burst and the onset of vocal fold vibration
- Usually studied in word-initial positions

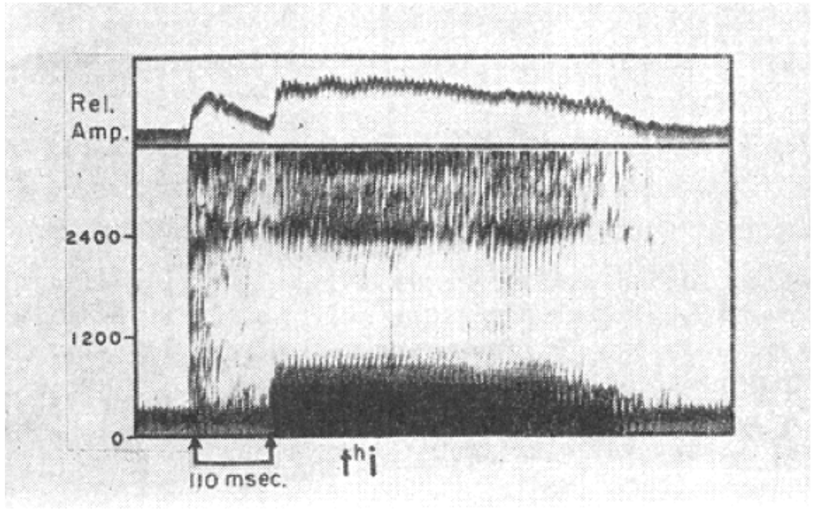
Lisker&Abramson, 1964



Voice lead

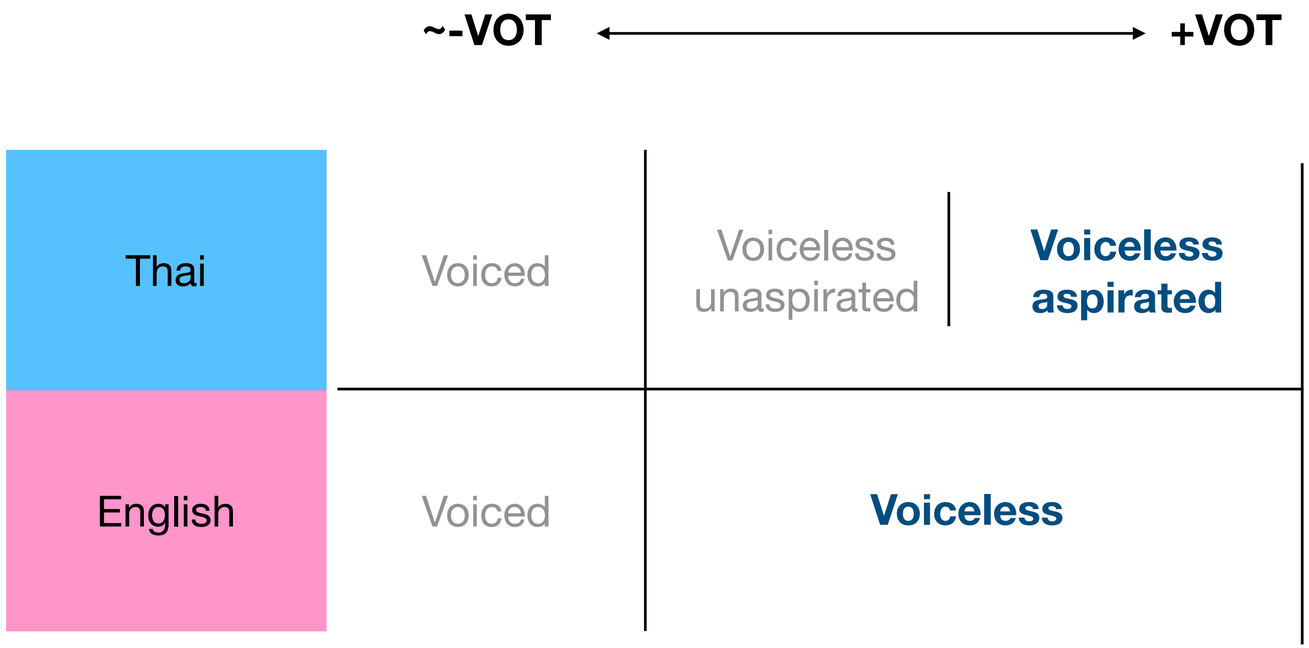


Short lag



Long lag

VOT in Thai and English



Lisker&Abramson, 1964

TABLE 6. Voice Onset Time in Msec: English (4 speakers)

	/p/	/t/	/k/
Av.	58	70	80
R.	20:120	30:105	50:135
N.	102	116	84

TABLE 8. Voice Onset Time in Msec: Thai (3 speakers)

	/pʰ/	/tʰ/	/kʰ/
Av.	64	65	100
R.	25:100	25:125	50:155
N.	33	33	38

L1 Transfer

Shimizu, 2011

"Voiceless stops are produced with VOT values which are close to those of aspirated stops in Thai," with Thai has higher average VOT

Table 3: Mean VOT values of Thai stops (ms) (N=24) (s.d. in parenthesis).

Voiced	vl.unaspirated	vl.aspirated
b -87(16.9)	p 11 (6.1)	p^h 87(20.8)
d -75(3.7)	t 10 (3.7)	t^h 91(27.6)
	k 21 (20.0)	k^h 113(19.5)

← Thai

Table 4: Mean VOT values of English stops by Thai subjects (ms) (N=12) (s.d. in parenthesis).

E-Voiced	E-Voiceless
b -100 (23.7)	p 74 (23.0)
d - 68(57.6)	t 91 (31.1)
g - 25 (24.9)	k 94 (21.8)

← English produced by Thai: Shorter VOT

Factors that affect VOT

Yao, 2009	Smith et al, 2015
Understanding VOT Variation in Spontaneous Speech	The private life of stops: VOT in a real-time corpus of spontaneous Glaswegian

VOT is sensitive to many factors, including following vowel height, speech rate, place of articulation, and speaker style.

Time

Balukas&Koops, 2015

The effect of code-switching levels off quickly after a point of time

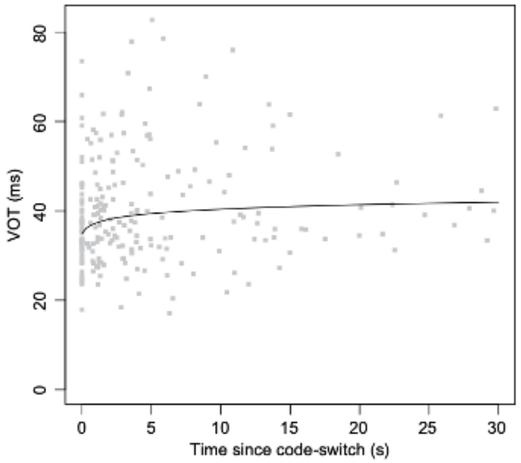


Figure 5. Regression line showing the log-linear increase in the English VOTs.

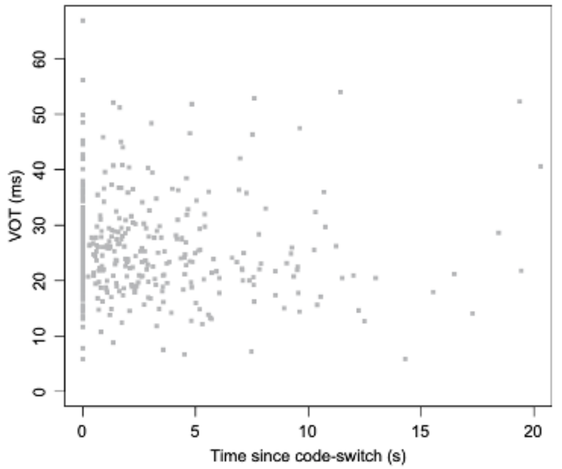


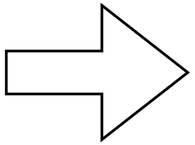
Figure 6. Distribution of the Spanish VOT values.

No study on direct correlation between speech time and VOT

Time

Timngam, 2020

Tense markers significantly decrease later in essays produced by Thai students



Morphosyntactic construction in L2 long production varies.

Writing is less spontaneous in speech.

Will phonetic production also change in long L2 production?

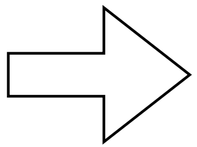
Read and spontaneous speech

Nakamura, Iwano & Furui S, 2008

Compared with read speech, the spectral distribution is reduced and phonemes vary more in spontaneous speech.

Baran et al 1977; Chodroff & Wilson 2017

VOT in read citation forms have shorter VOT duration than VOT in spontaneous speech



Will speech style affect VOT variation?

Research Questions

1. To test whether L1 transfer in terms of voice onset time intensifies with speech time, here defined as time elapsed from the first syllable

2. To test whether this variation is the same across speech styles

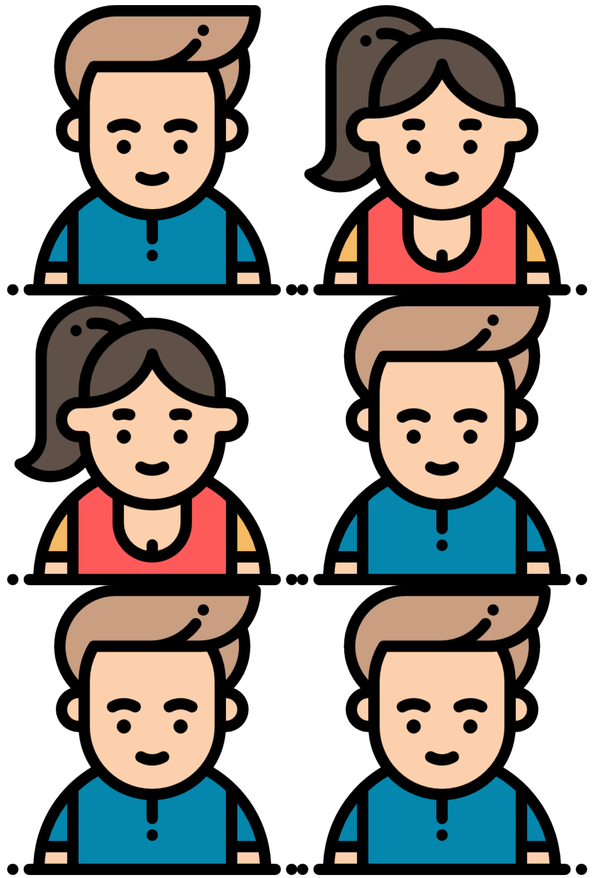
Hypothesis

1. VOT values should significantly increase as speech time elapses because their production gets similar to their L1 VOT, which has higher values.

2. VOT in spontaneous speech should be shorter and varies more than in read speech, thus resulting in a steeper slope.

Methods

Methods: Participants

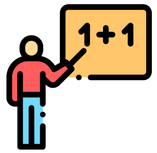


6 participants

- Thai as the first language and English as their second language
- Are all university students
- Studied in international schools/English program
- Have been trained to give 7-min debate speech
- Have never stayed abroad for longer than 6 months

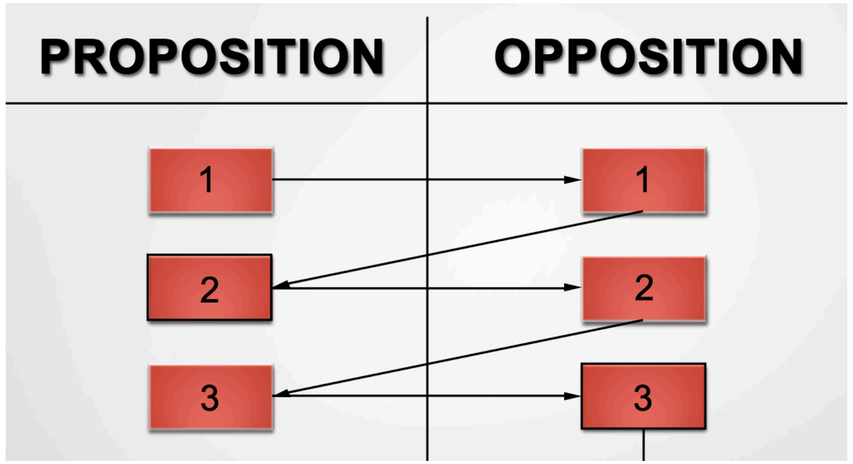
Methods: Elicitation and Recording

Debate Task to obtain spontaneous speech



- Participants engage in an online impromptu parliamentary debate

  “This House would punish natural or legal persons who are accused of cultural appropriation.”



- 30-min prep time, 7-min speech without interruption

- Two simultaneous means of recording: mixidea online recorder and participants’ own phones

Methods: Elicitation and Recording

Reading task 

- Article 'Kept Women’ from Aeon.com
    - rearranged to contain targeted stops
    - expected to last ~7 minutes

- Participants were allowed to practice as they wished before recording

- Self-record using participants’ phones

Methods: Analysis

- For spontaneous speech, use sounds with clearer signals
- manually transcribe
- auto-segment using Bavarian Archive for Speech Signals (BAS)'s service
- manually segment VOT in Praat only in word-initial position
- Code place of articulation, following vowel duration, following vowel height, participants, and the words containing the stops

Spontaneous Speech

Approximately 903 tokens

Average duration 7:21 minutes

Read Speech

498 Expected tokens

Average duration 7:36 minutes

Method: Data Selection

Exclusion Criteria

- Technical problems

(eg. noise, no clear burst signal)

- Stop deletion
- Devoiced following vowel
- Affrication and frication
- Voicing

Spontaneous Speech

523 tokens

Read Speech

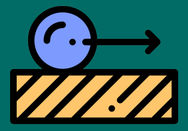
367 tokens

(DEL=9, DV=15, Af/F=102 ,V=5)

Method: Normalization

Linear Mixed-Effects Model

- is adapted from Balukas&Koops (2015)
- uses lme4 package and nlme package in R
- allows for variables to be controlled
- uses random intercepts



Fixed Effects

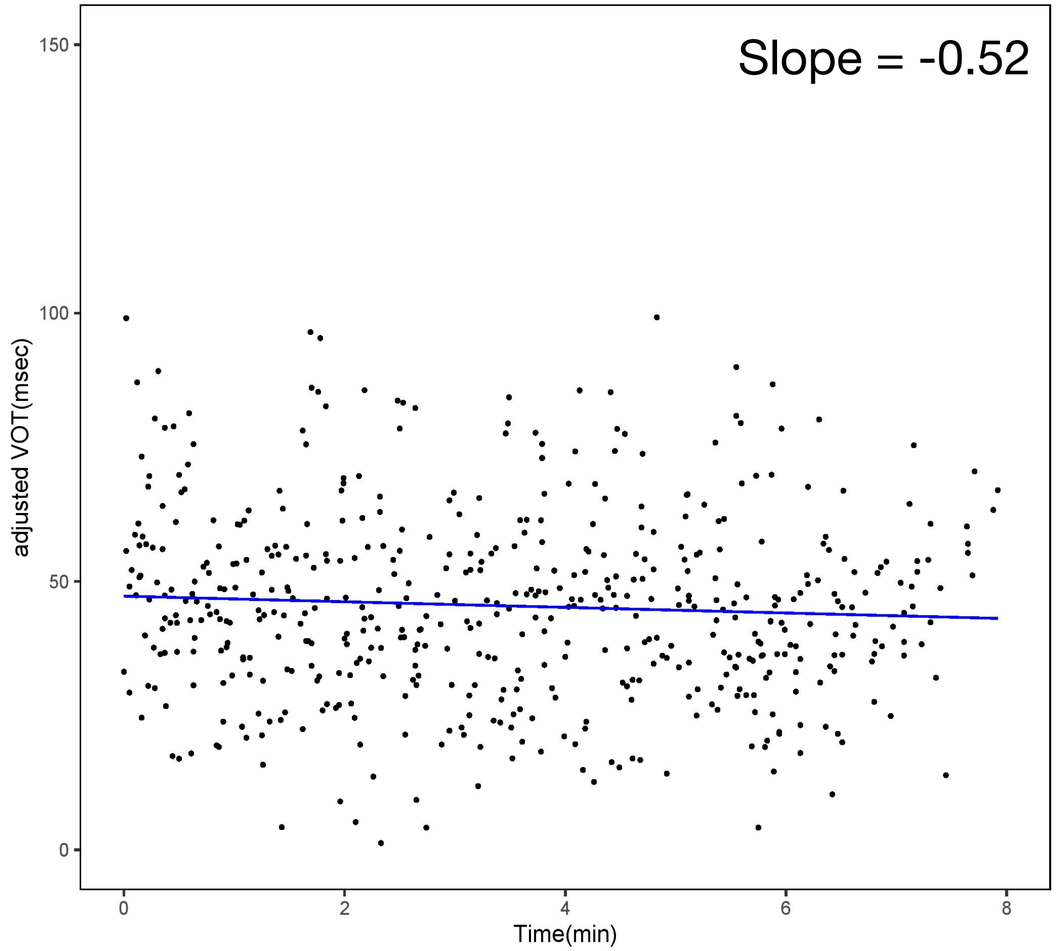
Time from the onset of the first syllable
Vowel duration (= speech rate)

Random Effects

Place of articulation
Vowel height
Words containing the items
Participants

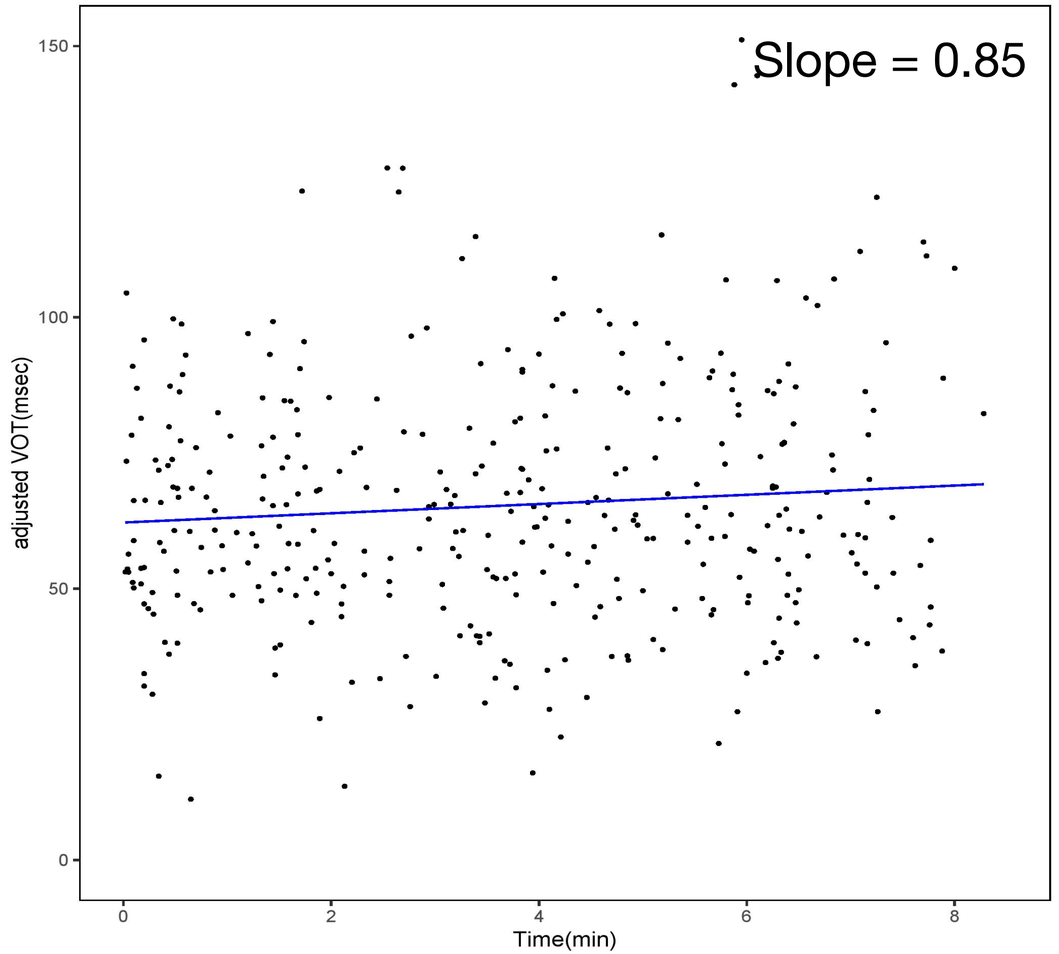
Results

Spontaneous Speech



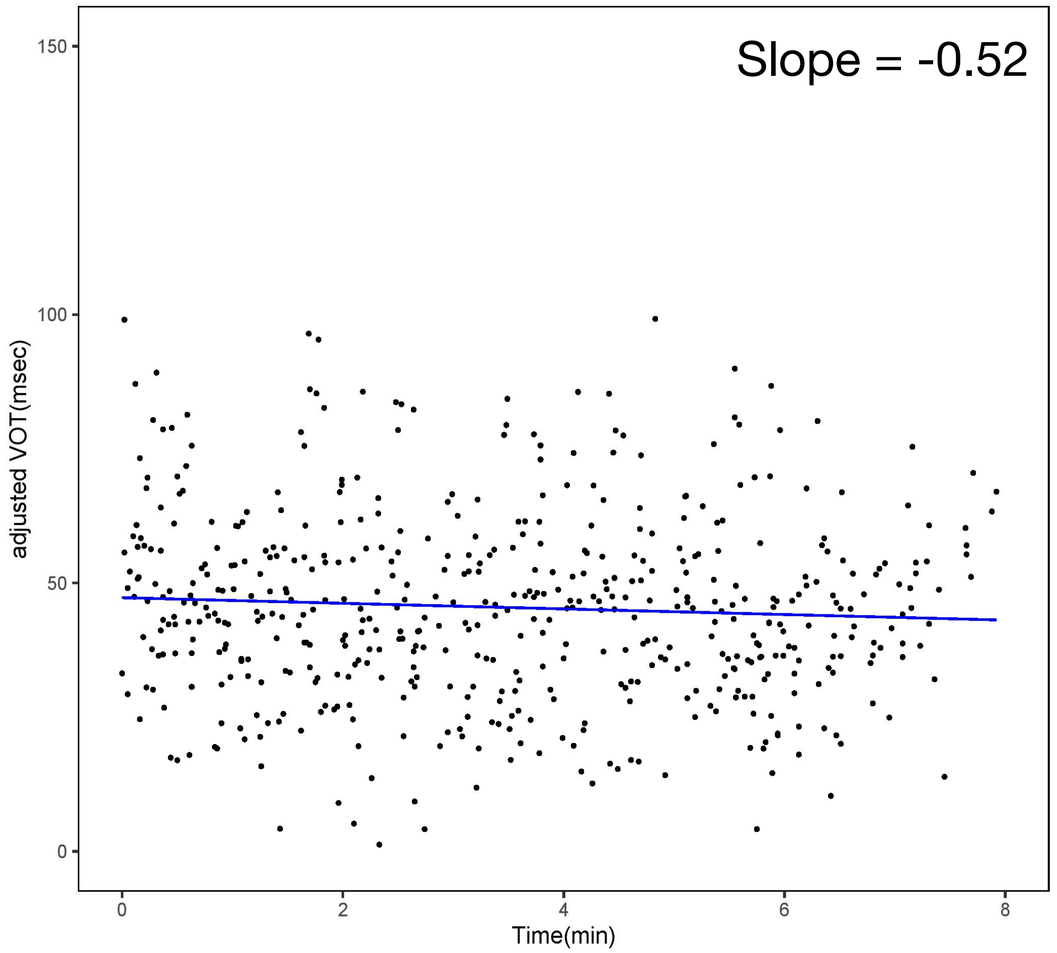
- All dots are intercept-adjusted from the model. The model shows linearity.
- In spontaneous speech, VOT slightly drops as the time goes.

Read Speech

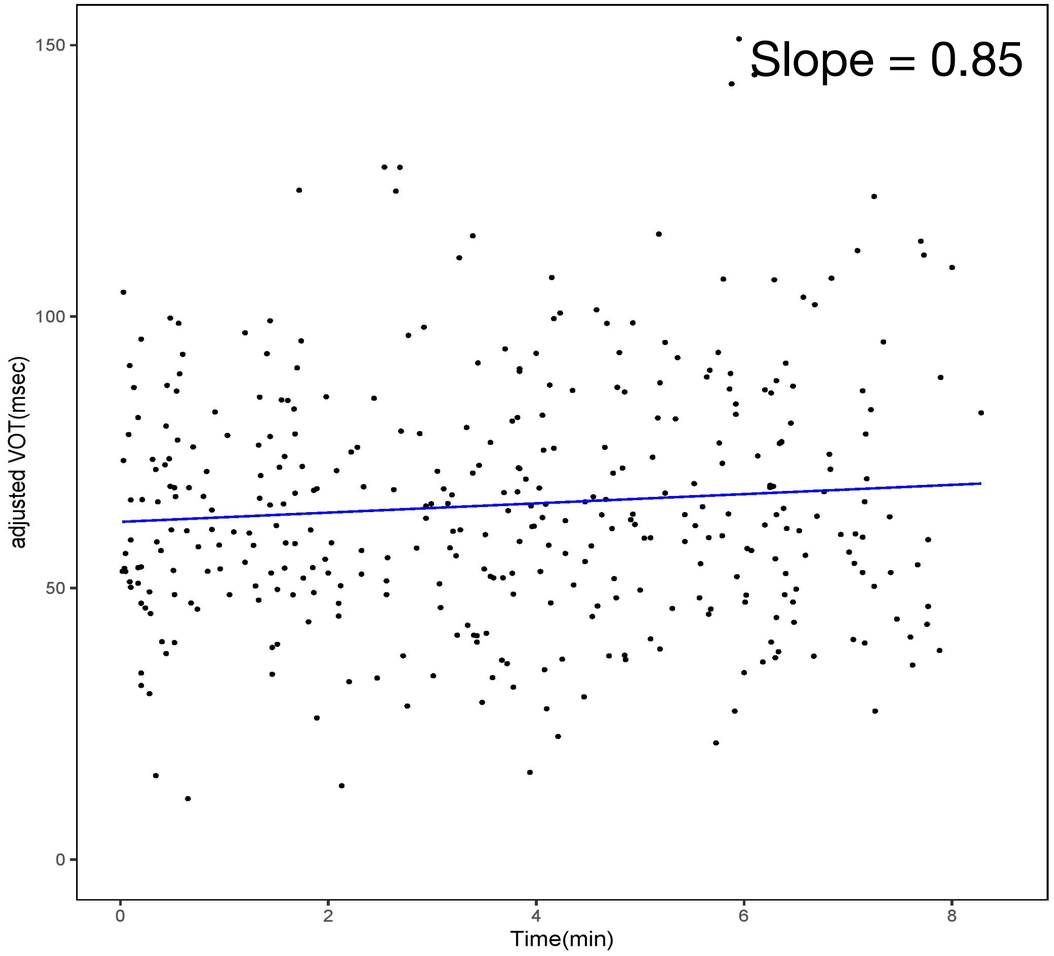


- All dots are intercept-adjusted from the model. The model shows linearity.
- Unlike spontaneous speech, in read speech, VOT increases slightly but with greater slope than in spontaneous speech.

Spontaneous Speech



Read Speech



- Not significant, given that the fact that there is much room for variation (~50 milliseconds)
- F-test via Kenward-Roger approximation : p-value = 0.20 in spontaneous speech, 0.24 in read speech

Discussion

Hypothesis fail.

1. VOT does not significantly increase as speech time elapses.

2. VOT in spontaneous speech is shorter, but does not vary significantly.

Confirmation or contradiction?

- May align with Shimizu's findings that L2 VOT is affected by L1 VOT

Shimizu, 2011

Table 3: Mean VOT values of Thai stops (ms) (N=24) (s.d. in parenthesis).

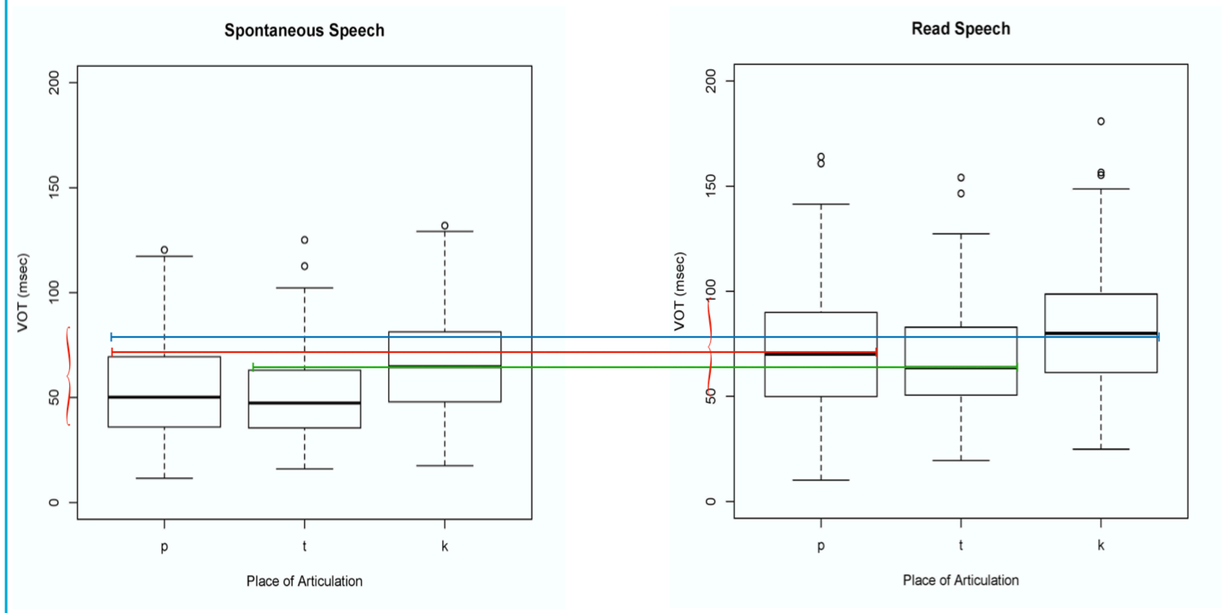
Voiced	vl.unaspirated	vl.aspirated
b -87(16.9)	p 11 (6.1)	p^h 87(20.8)
d -75(3.7)	t 10 (3.7)	t^h 91(27.6)
	k 21 (20.0)	k^h 113(19.5)

Thai

Table 4: Mean VOT values of English stops by Thai subjects (ms) (N=12) (s.d. in parenthesis).

E-Voiced	E-Voiceless
b -100 (23.7)	p 74 (23.0)
d - 68(57.6)	t 91 (31.1)
g - 25 (24.9)	k 94 (21.8)

English produced by Thai: Shorter VOT



Why are VOT values different in two types of speech?

- In line with Baran et al 1977; Chodroff & Wilson 2017

- Stylistic variation - attention and awareness affect stop articulation.
  - Participants tend to be aware of their speech more when they are reading.
  - In spontaneous speech, participants have to pay attention to content.
- Linguistic and pragmatic constraints, e.g. redundancy and less information
  - This allows a greater degree of imprecision to be tolerated.

Why does VOT not significantly vary with time?

Possible explanation:

- Despite the effects, participants still separate VOT values for each language in order to maintain the contrast between languages.
- More data are still needed to establish the conclusion.

Why do different speech styles have different trends?

Possible explanation for downward trend in spontaneous speech:

- Vocal fatigue (Caraty & Montacié, 2010)

Possible explanation for upward trend in read speech:

- Leaning towards L1, but minimized by category separation
- Overshooting to compensate fatigue

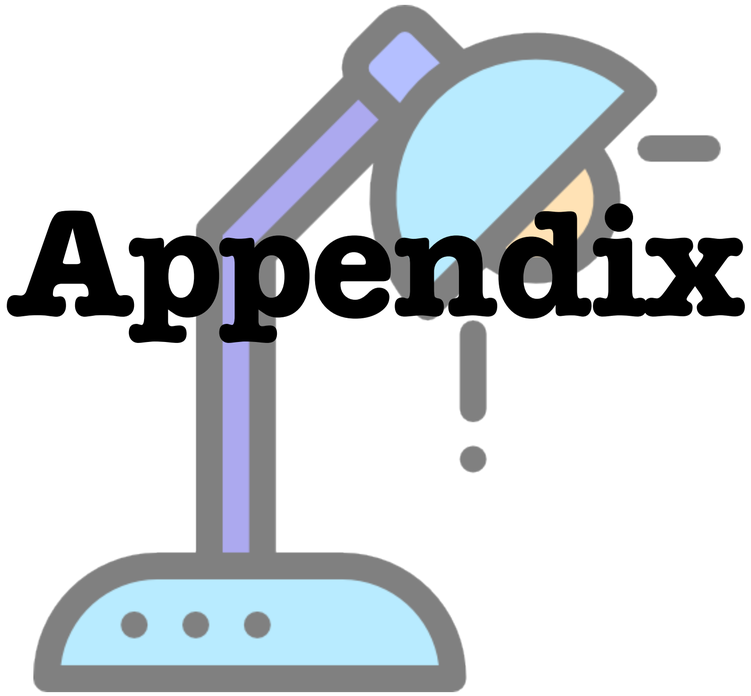
Conclusion - Limits and suggestions

- VOT in tasks without topics for spontaneous speech, e.g. picture description, should also be examined to reaffirm the conclusion.

- Participants with different levels of English proficiency and Thai VOT in long speech could also be compared to see clearer contrast.

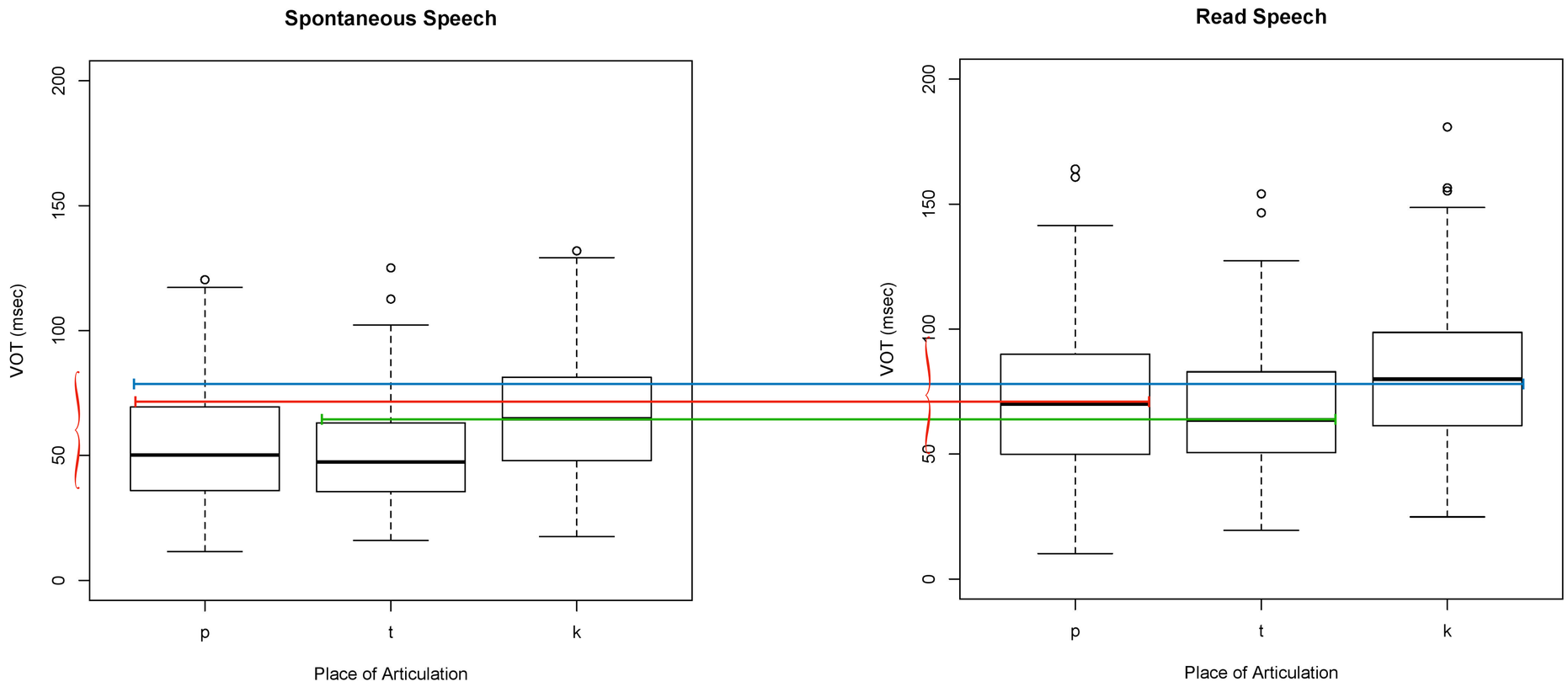
Thank you so much

Appendix



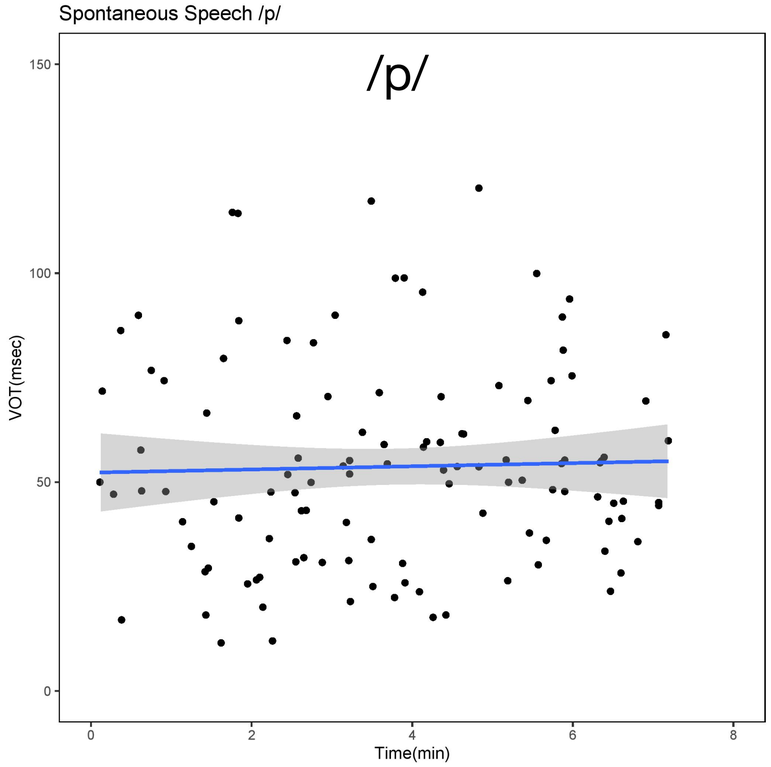
- Raw VOT plots
- Number of used tokens
- Speech time
- LMMs and residual plots

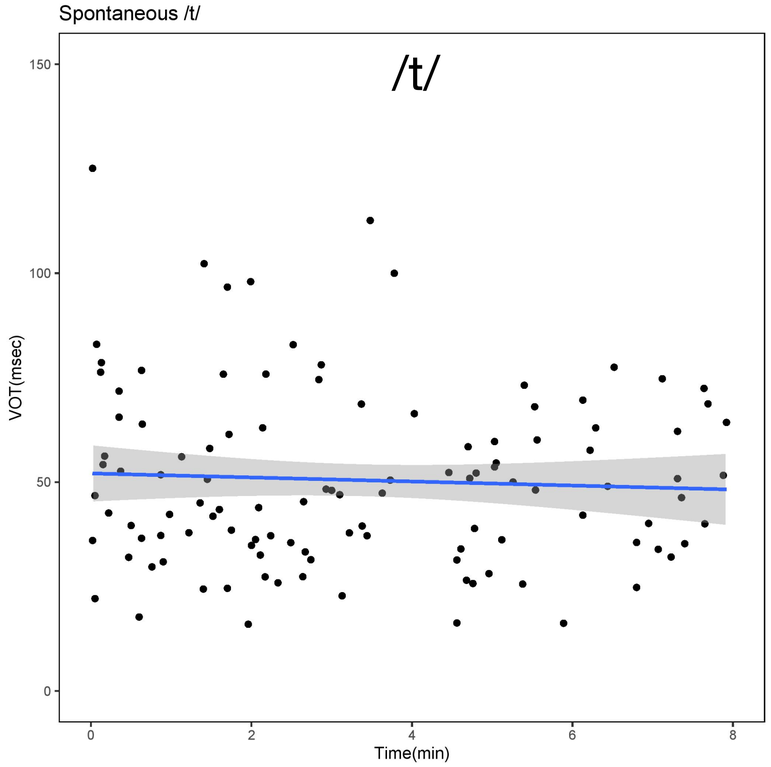
Mean Duration

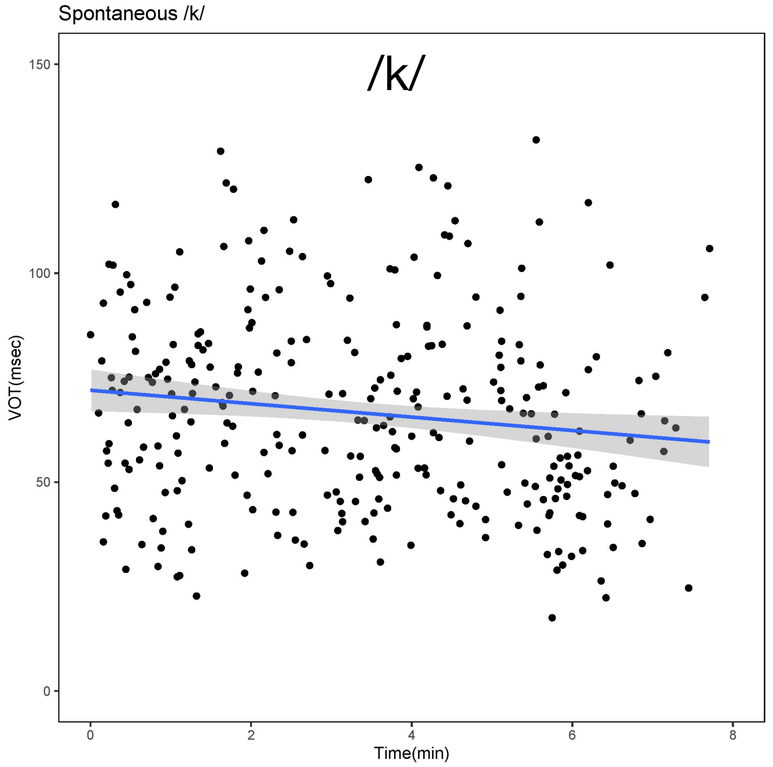


- All Mean VOT values in read speech is slightly higher than in spontaneous speech.
  - (This is no surprise as reading an article is similar to reading citation forms, resulting in the participants being more aware and careful.)
- Range in spontaneous speech: ~40-80 VS Range in read speech: ~50-100

Raw VOT in Spontaneous Speech

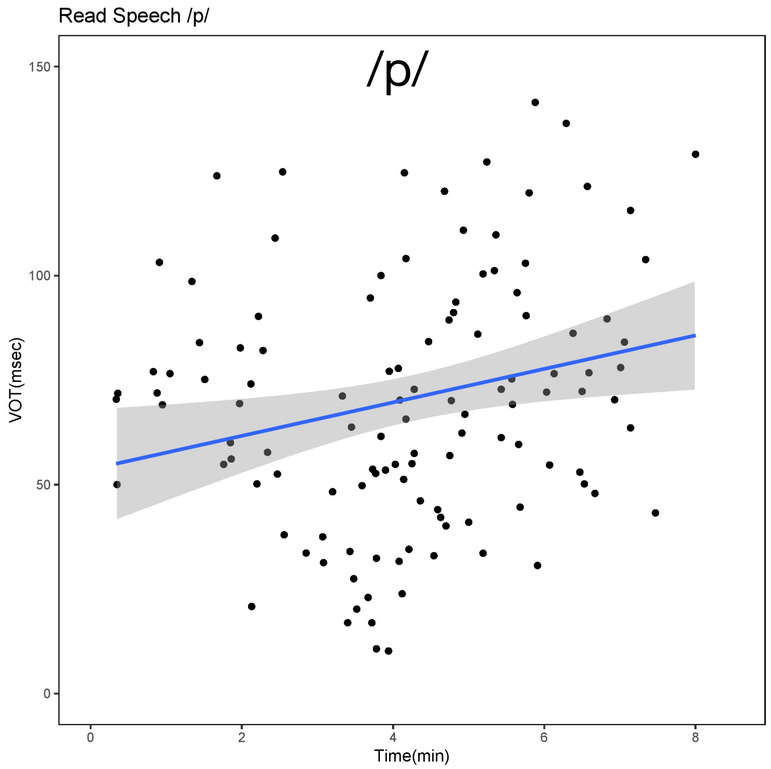


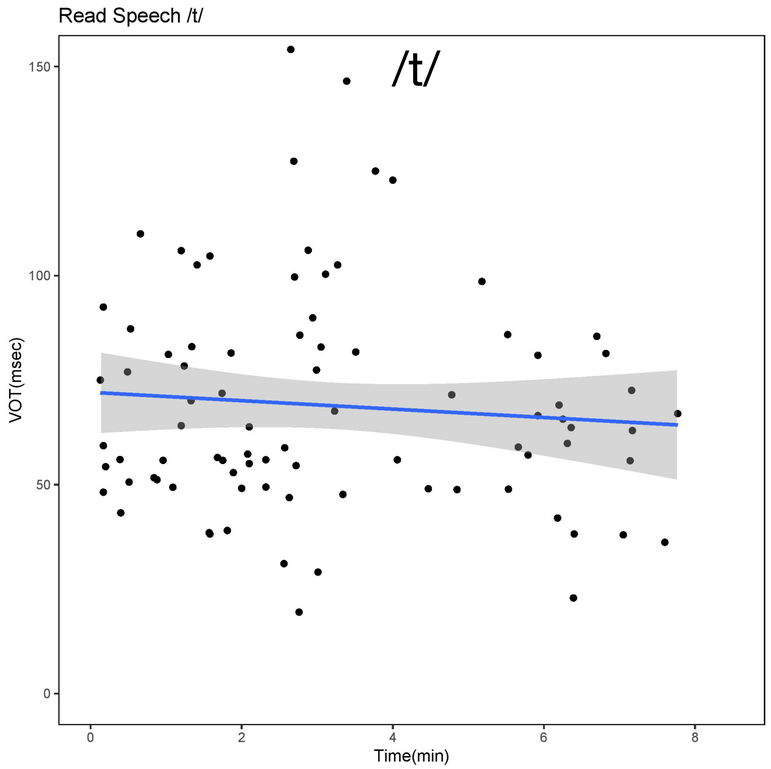


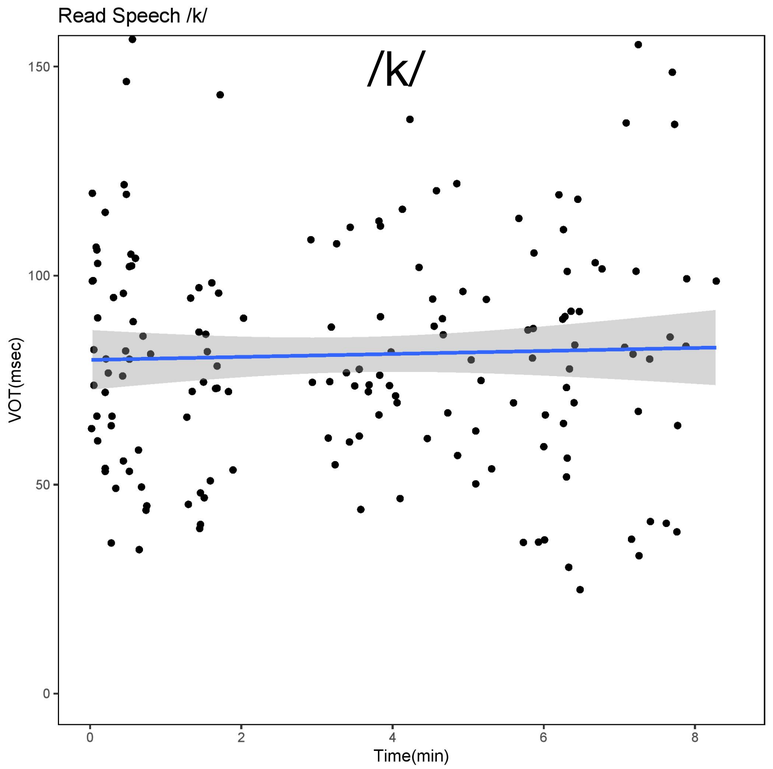


- Linear regression line added to see the trend, using lm function from ggplot2 package in R
- If there was linguistic convergence enhancement with time, we would expect VOT to increase or decrease significantly.
- Only one stop type /k/ shows decrease in VOT, whereas /p/ /t/, though their VOT do vary, do not have significant changes.

Raw VOT in Read Speech







- Linear regression line added to see the trend, using lm function from ggplot2 package in R
- Unlike spontaneous speech, /k/ seems to be the most stable stop. /p/ increases sharply and /t/ decreases.
- No systematic trend in neither between spontaneous and read speech nor among stop types.

Literature Review

	Fixed effect	Random Effect
Sonderegger (2015)	1.Clip 2.Day	1.Place of articulation 2.~~Following phone type~~ 3.Vowel height 4.Frenquency 5.Syllable stress 6.Position in phrase 7.Speaking rate
Balukas&Koops (2015)	1.Time elapsed 2.Vowel duration 3.~~The number of syllables~~	1.The speaker 2.The lexical item 3.Stop type 4.Vowel height 5.~~Presence of approximant~~

Number of used tokens

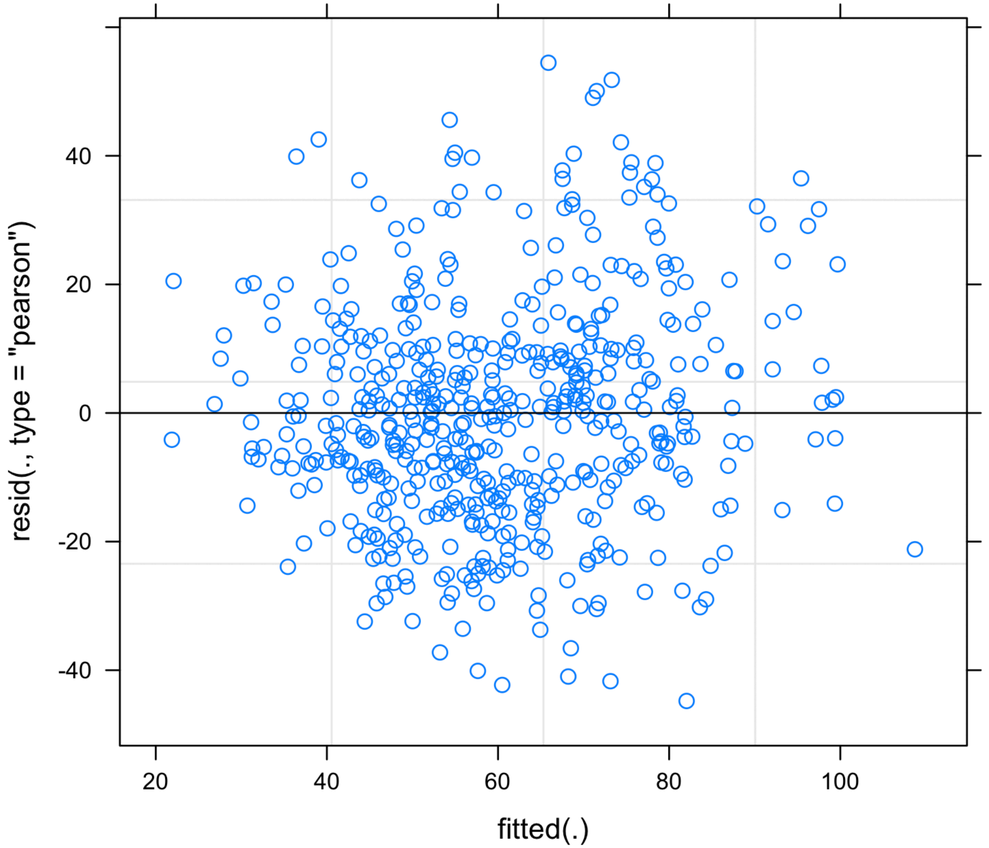
	Spontaneous	Read
/p/	113	119
/t/	116	85
/k/	294	163
Total	523	367

Time

	Spontaneous (Min)	Read (Min)
P1	6.57	7.50
P2	7.24	8.22
P3	6.34	6.44
P4	7.24	7.41
P5	7.56	6.23
P6	7.01	6.50
Average	7.21	7.36

Linear Mixed Model for Spontaneous Speech

```
> summary(mds)
Linear mixed model fit by REML ['lmerMod']
Formula: VOT ~ T + VD + (1 | POA) + (1 | VH) + (1 | Word) + (1 | PN)
   Data: spon

REML criterion at convergence: 4592.3

Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.44430 -0.61047 -0.03309  0.53784  2.97269

Random effects:
 Groups   Name        Variance Std.Dev.
 Word     (Intercept)  78.91    8.883
 PN       (Intercept) 144.03   12.001
 VH       (Intercept)  22.31    4.723
 POA      (Intercept)  36.03    6.003
 Residual             335.77   18.324
Number of obs: 523, groups:  Word, 111; PN, 6; VH, 3; POA, 3

Fixed effects:
            Estimate Std. Error t value
(Intercept)  47.2591     7.3081   6.467
T            -0.5226     0.4034  -1.296
VD          179.8961    27.8044   6.470

Correlation of Fixed Effects:
   (Intr) T
T  -0.201
VD -0.299  0.036
```



Linear Mixed Model for Read Speech

```
> summary(mdr)
Linear mixed model fit by REML ['lmerMod']
Formula: VOT ~ T + VD + (1 | POA) + (1 | VH) + (1 | Word) + (1 | PN)
   Data: read

REML criterion at convergence: 3416.3

Scaled residuals:
    Min      1Q  Median      3Q     Max
-2.1697 -0.6203 -0.1285  0.6071  3.5307

Random effects:
 Groups   Name        Variance Std.Dev.
 Word     (Intercept) 141.37   11.890
 PN       (Intercept) 109.67   10.473
 VH       (Intercept)  40.63    6.374
 POA      (Intercept)  16.22    4.028
 Residual             564.31   23.755
Number of obs: 367, groups:  Word, 53; PN, 6; VH, 3; POA, 3

Fixed effects:
            Estimate Std. Error t value
(Intercept)  62.1940     7.5672   8.219
T             0.8516     0.7147   1.192
VD          111.2686    33.5296   3.319

Correlation of Fixed Effects:
   (Intr) T
T  -0.312
VD -0.376 -0.082
convergence code: 0
Model failed to converge with max|grad| = 0.00417902 (tol = 0.002, component 1)
```