Lexical diffusion of vowel length merger in Seoul Korean: a corpus-based study

Yoonjung Kang[1], Tae-Jin Yoon[2], Sungwoo Han[3]
University of Toronto Scarborough[1],
Cheongju University[2], Inha University[3]

LabPhon 14
July 26, 2014

Overview

- Usage-based model of phonology:
  - Frequency interacts with sound change reflecting the mechanism behind the change
  - But, studies on the frequency effect on sound change that track the effect over time are rare.
- Shortening of long vowels in Seoul Korean, sound change in progress

I. Korean Vowel Length Contrast

Vowel length contrast in Seoul Korean

/nun/	‘eye’	/nu:n/	‘snow’
/pɛ/	‘pear’	/pɛ:/	‘double’
/pɑm/	‘night’	/pɑ:m/	‘chestnut’
/il/	‘one’	/i:l/	‘work’
/tsʌk-ta/	‘to write’	/tsʌ:k-ta/	‘plenty’
/mos/	‘nail’	/mo:s/	‘Neg.’

Word-initial position only

- Long vowels are restricted to word-initial position.

/nu:n-s'ʌlmɛ/	[nu:ns'ʌlmɛ]	'toboggan'
/tsʰʌs-nu:n/	[tsʰʌnnun]	'first snow'
/so:k-sɑtsʌŋ/	[so:ks'ɑdzʌŋ]	'behind story'
/mɑɨm-so:k/	[mɑɨms'ok]	'one's mind'
/pʌ:lli-ta/	[pʌ:llida]	'open, spread'
/t'ʌ-pʌ:llita/	[t'ʌbʌllida]	'brag'

Some interaction with morphology

- Some verbs/adjectives alternate:

C-initial suffix	V-initial suffix	
[ku:lm-ta](>[ku:mt'a])	[kulm-ʌ]	'to starve'
[nʌ:h-ta] (>[nʌ:tʰa])	[nʌh-ɨni] (>[nʌɨni])	'to insert'
[ki:l-ta]	[kil-ʌ]	'long'
[pu:s-ta] (> [pu:t'a])	[pu-ɨni]	'to pour'

- Not all verbs/adjectives do:

[ca:k-ta](>[ca:kt'a])	[ca:k-a]	'small'
[ʌ:ps-ta] (>[ʌ:pt'a])	[ʌ:ps-ʌ] (>[ʌps'ʌ])	'absent'

- Nouns don't:

C-initial suffix	V-initial suffix	
[sa:lm-to](>[sa:mdo])	[sa:lm-i]	'life'
[ka:m-kwa](>[ka:mgwa])	[ka:m-i]	'persimmon'

Previous studies

- Earlier studies describe vowel length contrast as phonemic, although some individual or lexical variation is noted already.

(Hur, Woong 1960, Lee, Soong Nyong 1955, Lee, Hi Sung 1956, Martin 1951, 1954, Han 1964)

- More recent studies observe generational change and variation in vowel length realization.

(Jung & Hwang 2000, Ko 1988, Lee & Zhi 1981, Zhi 1993, Kim 2003, Park 1985, Nam 2012).

- Lexical diffusion: certain words are more consistently long than others.

II. Frequency Effect on Sound Change

Frequency and Lexical Diffusion

- Sound change may affect low- and high-frequency words differently.

  (Schuchardt 1885, Fidelholtz 1975, Hooper 1976, Phillips 1980, 1983, 1984, 2001, 2006, Bybee and Slobin 1982, Bybee 1985, 1995, 2001, Rhodes 1996, Bybee and Hopper 2001, Corbett *et al.* 2001, Pierrehumbert 2001, Kang 2003, Schleef 2013, Sonderegger and Niyogi 2013)

High frequency words first

- Phonetically motivated changes arising from shortening and lenition affect high-frequency words first.
- If representation of a word gets updated with each use of the word, a reductive change will be more advanced in high frequency words than in low frequency words.

  e.g., Schwa deletion in English : *ev**e**ry* vs. *mamm**a**ry*

Low frequency words first

- Analogically motivated changes that involve (re)-analysis affect low-frequency words first.
- High-frequency words form a strong mental representation and resist change motivated by analogy to other forms.
- E.g., Regularization of English irregular past: *weeped (< wept)* vs. **keeped* (*<kept*)

Phillips (2001)

- Some changes that affect high-frequency words first are not lenition-based.
  - English stress shift in *–ate* verbs: e.g., *frústrate* > *frustráte*
- Some (seemingly) reductive changes affect low-frequency words first.
  - Middle English mid front vowel unrounding: ö > e
  - English yod dropping: tj > t

"Changes which require analysis—whether syntactic, morphological, or phonological—during their implementation affect the least frequent words first; others affect the most frequent words first."

Synchrony and Diachrony of frequency effect

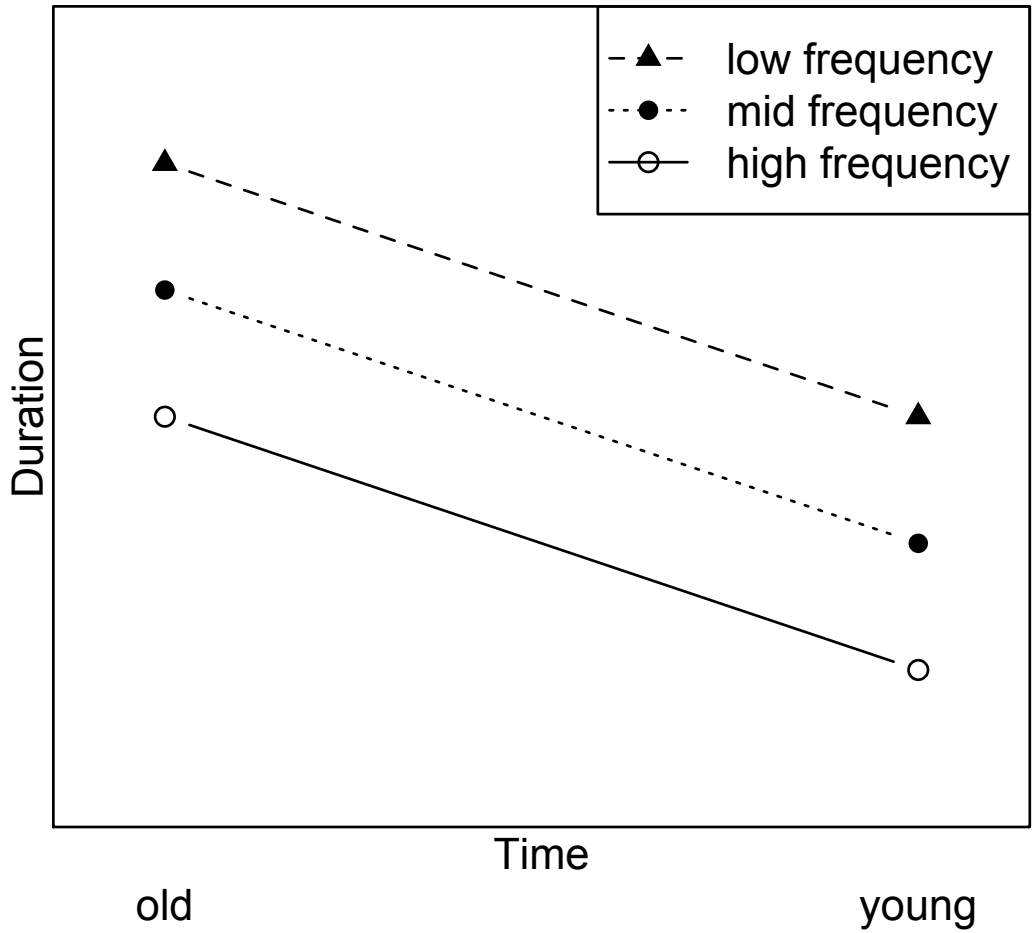
- Synchronic effect of frequency on segment duration: high frequency words are shorter and reduced.

  (Bell et al. 2003, 2009, Ernestus et al. 2006, Gahl 2008…)

- But, less clear is how this synchronic frequency effect interact with sound change over time?

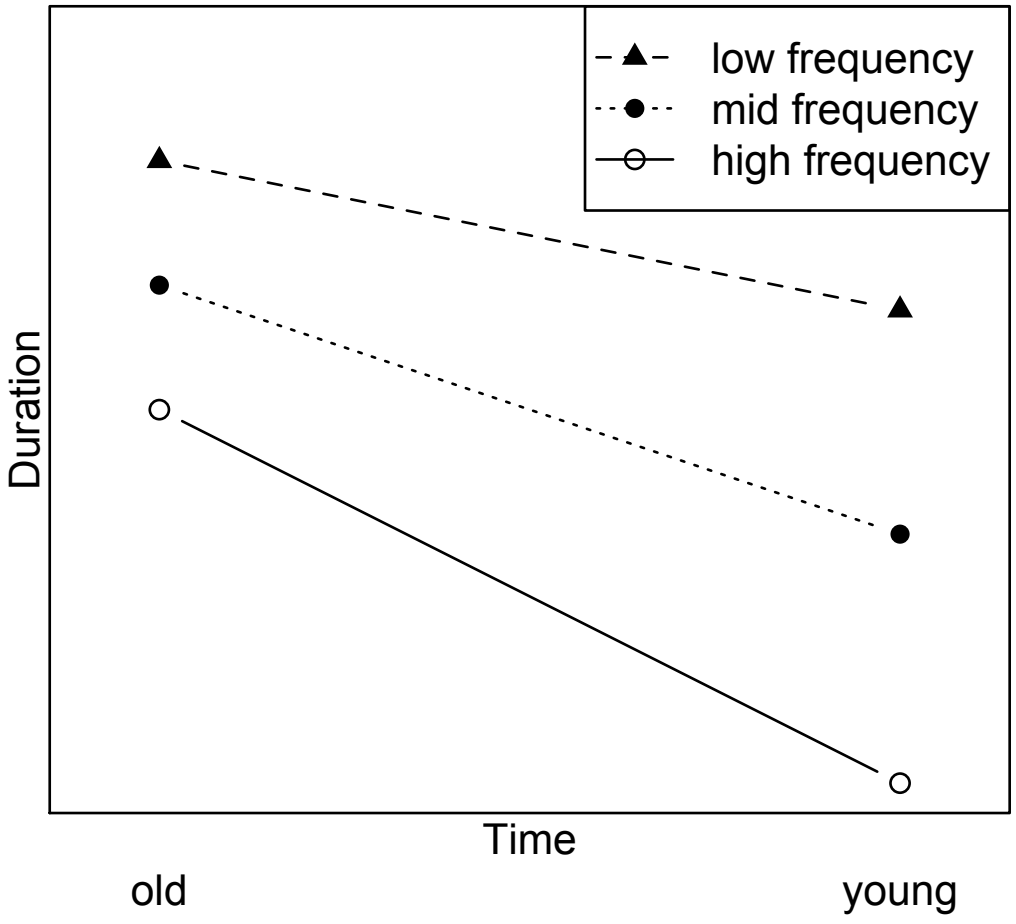
A) Diachrony mirrors synchronic frequency effect

Synchrony=Diachrony



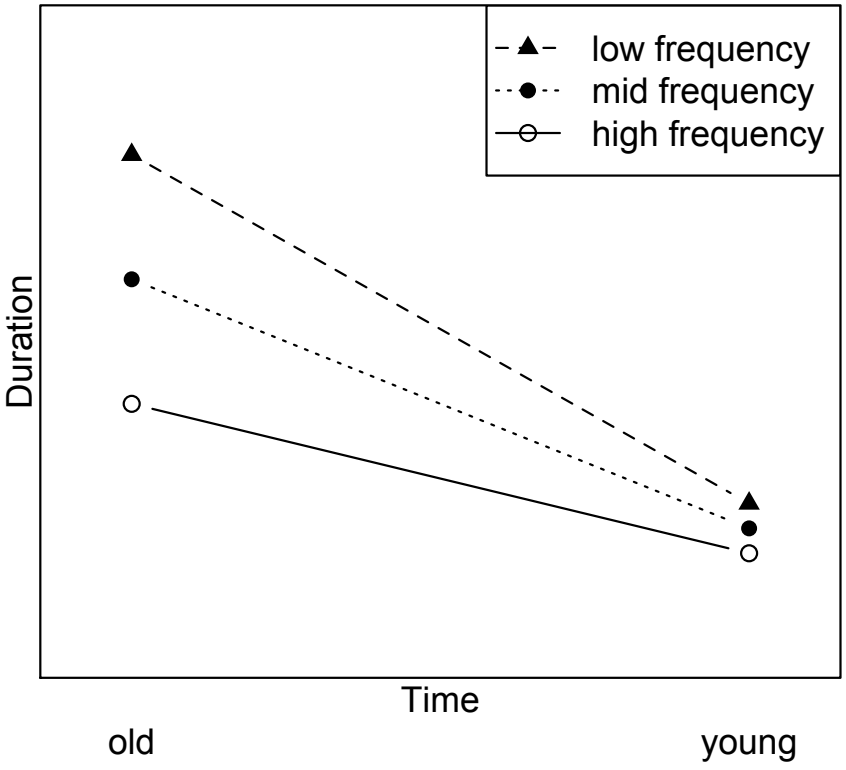
B) Diachrony **amplifies** synchronic frequency effect

Synchrony²=Diachrony

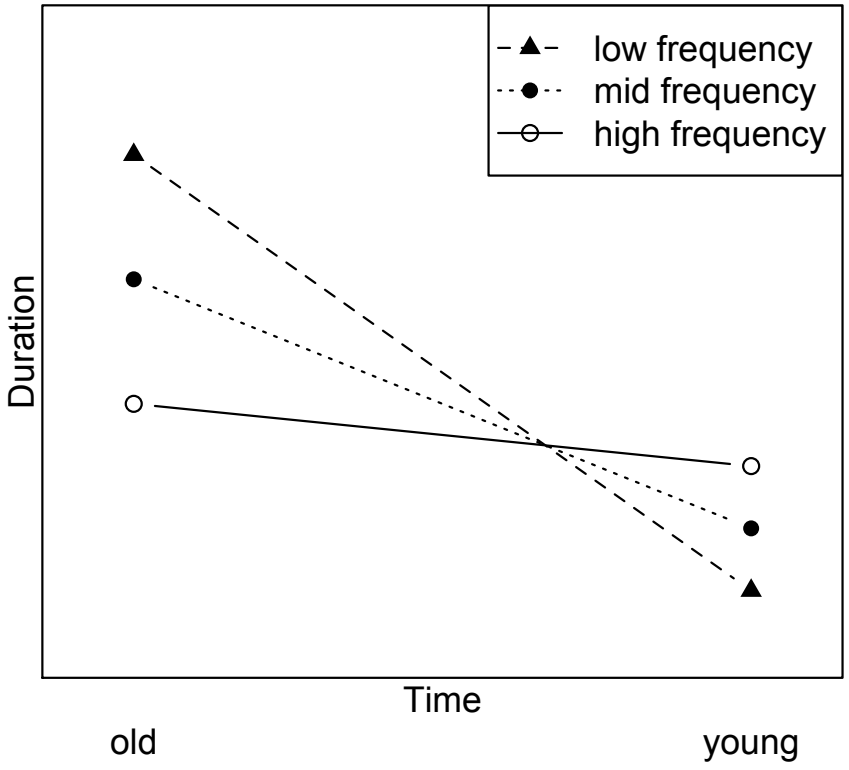


C) Diachrony introduces an "analytical" frequency effect, opposite of synchrony

Synchrony≠Diachrony

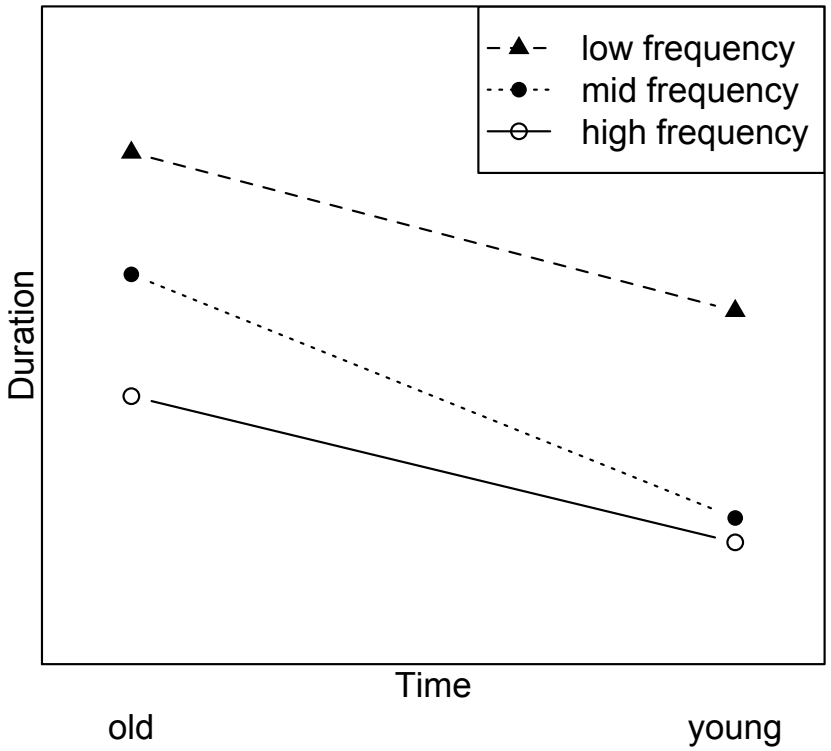


Synchrony≠Diachrony

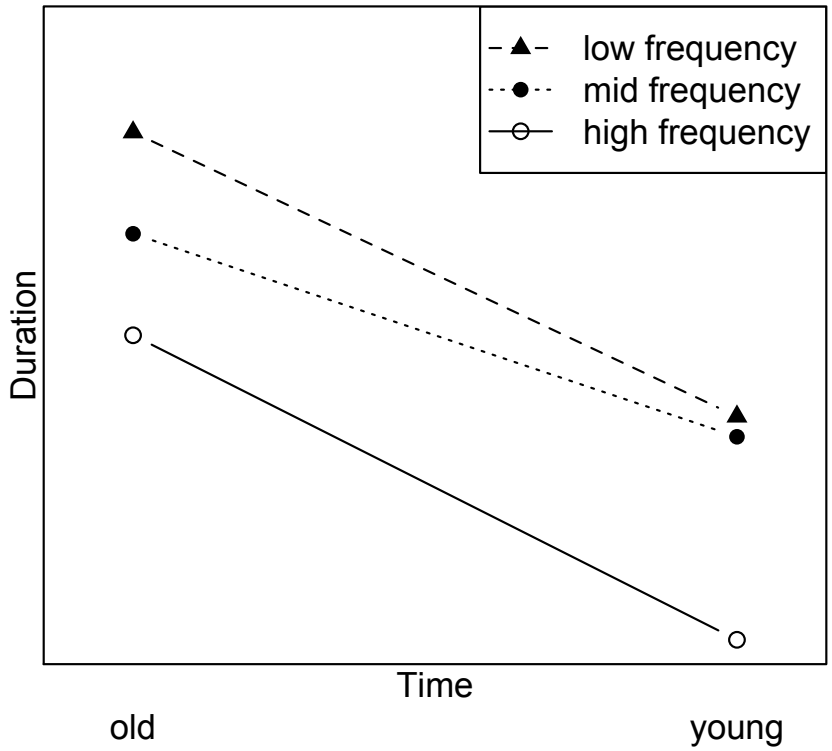


D) Mixed

High & Low Last



High & Low First



Which type of change is loss of length contrast in Korean?

- High frequency words first
  - Long vowels become short vowels.
  - Shortening/Reduction
- Low frequency words first
  - Removal of a marked structure
  - Analogy to a dominant pattern (short vowels)?
    - Long vowels: restricted to word-initial position, estimated to be <20%

III. Data

The Reading-Style Speech Corpus of Standard Korean

- The National Institute of the Korean Language (2002)
- 60 male, 60 female, Seoul-Kyeoungki area
- Age distribution

Age	**20s**	**30s**	**40s**	**50s**	**60~70s**
Male	20	20 (19)		20	20
Female	20		20	20	20 (19)

- Apparent time study method (Bailey et al. 1991)

	Text	# of sentences	Duration (mv01)
1.	요람기	51	321.92 sec (5.37 min)
2.	소나기	87	240.12 sec (4.00 min)
3.	아름다움에 대하여	69	311.92 sec (5.20 min)
4.	방망이를 깎던 노인	62	275.50 sec (4.59 min)
5.	문학의 세계와 삶의 세계	47	321.01 sec (5.35 min)
6.	독서와 인생	54	294.97 sec (4.92 min)
7.	수난 이대	62	317.31 sec (5.29 min)
8.	메밀꽃 필 무렵	94	425.98 sec (7.10 min)
9.	**토끼와 자라**	**60**	**185.27 sec (3.09 min)**
10.	**선녀와 나무꾼**	**73**	**295.03 sec (4.92 min)**
11.	**호랑이와 곶감**	**42**	**159.92 sec (2.67 min)**
12.	**해님 달님**	**28**	**124.10 sec (2.07 min)**
13.	**그리운 시내가**	**36**	**276.49 sec (4.61 min)**
14.	**광화문 지하도 아주머니**	**27**	**148.40 sec (2.47 min)**
15.	**막 지은 밥**	**17**	**106.36 sec (1.77 min)**
16.	**눈 오던 날**	**35**	**147.77 sec (2.46 min)**
17.	**숭늉의 지혜**	**19**	**95.00 sec (1.58 min)**
18.	**까만 눈동자들 앞에서**	**27**	**205.92 sec (3.43 min)**
19.	**내 고향 개울 물에서**	**40**	**163.34 sec (2.72 min)**
	Total	930	4416.40 sec (73.61 min)

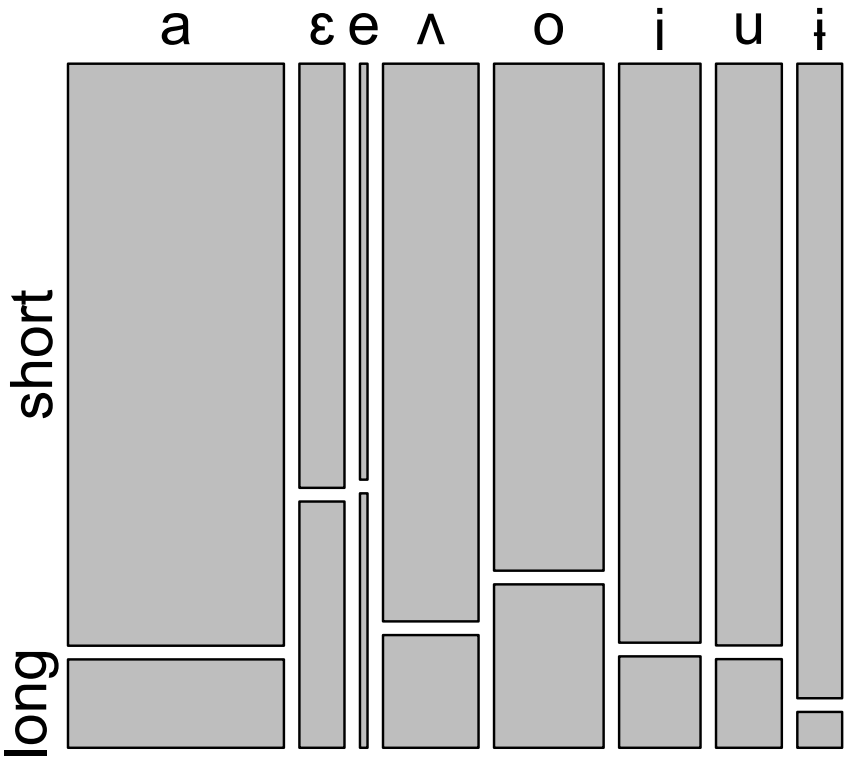
From Yoon (2013)

Seoul Korean vowels

i	ɨ	u
e	ʌ	o
(ɛ)	ɑ	

Vowels in word-initial syllables

- 2,250 word types
- 3,366 word tokens
- 1,251 lexemes
- 118 speakers
- <300 ms only, voiced vowels only
- 373,733 tokens of vowels in word-initial syllable

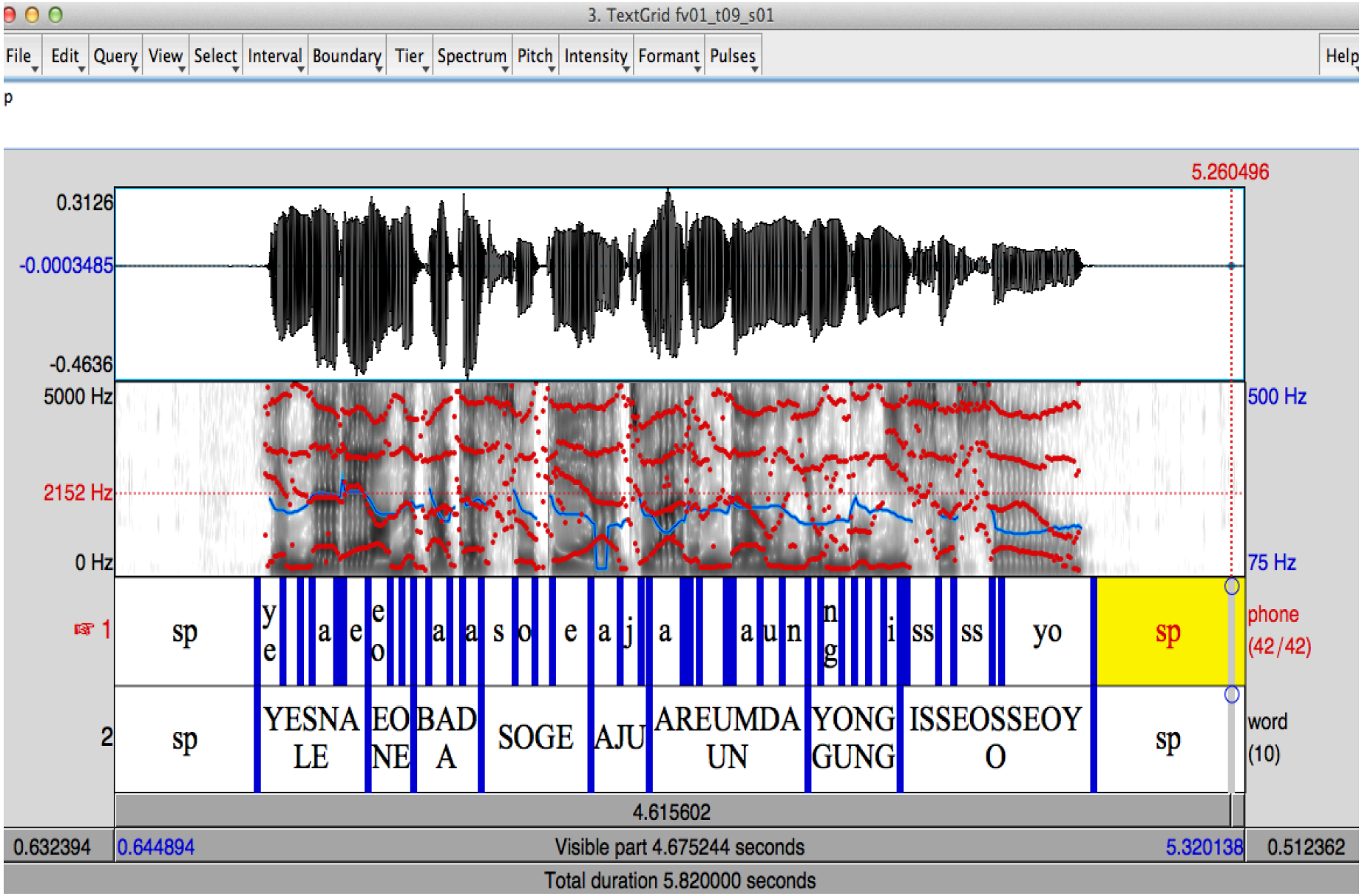


Korean phone-aligner

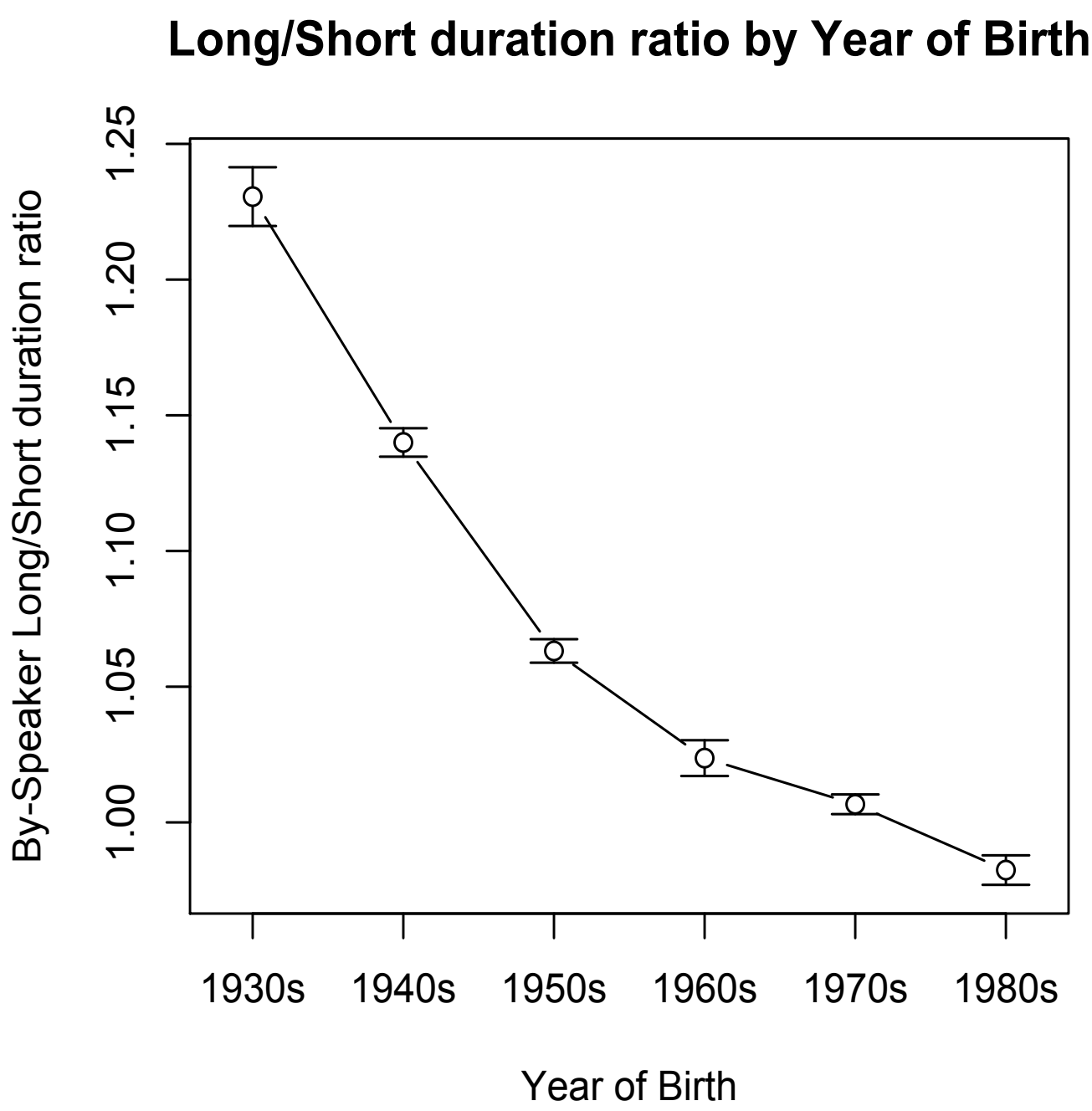
- Developed by Tae-Jin Yoon (cf. Yoon and Kang 2012, Yoon 2013)

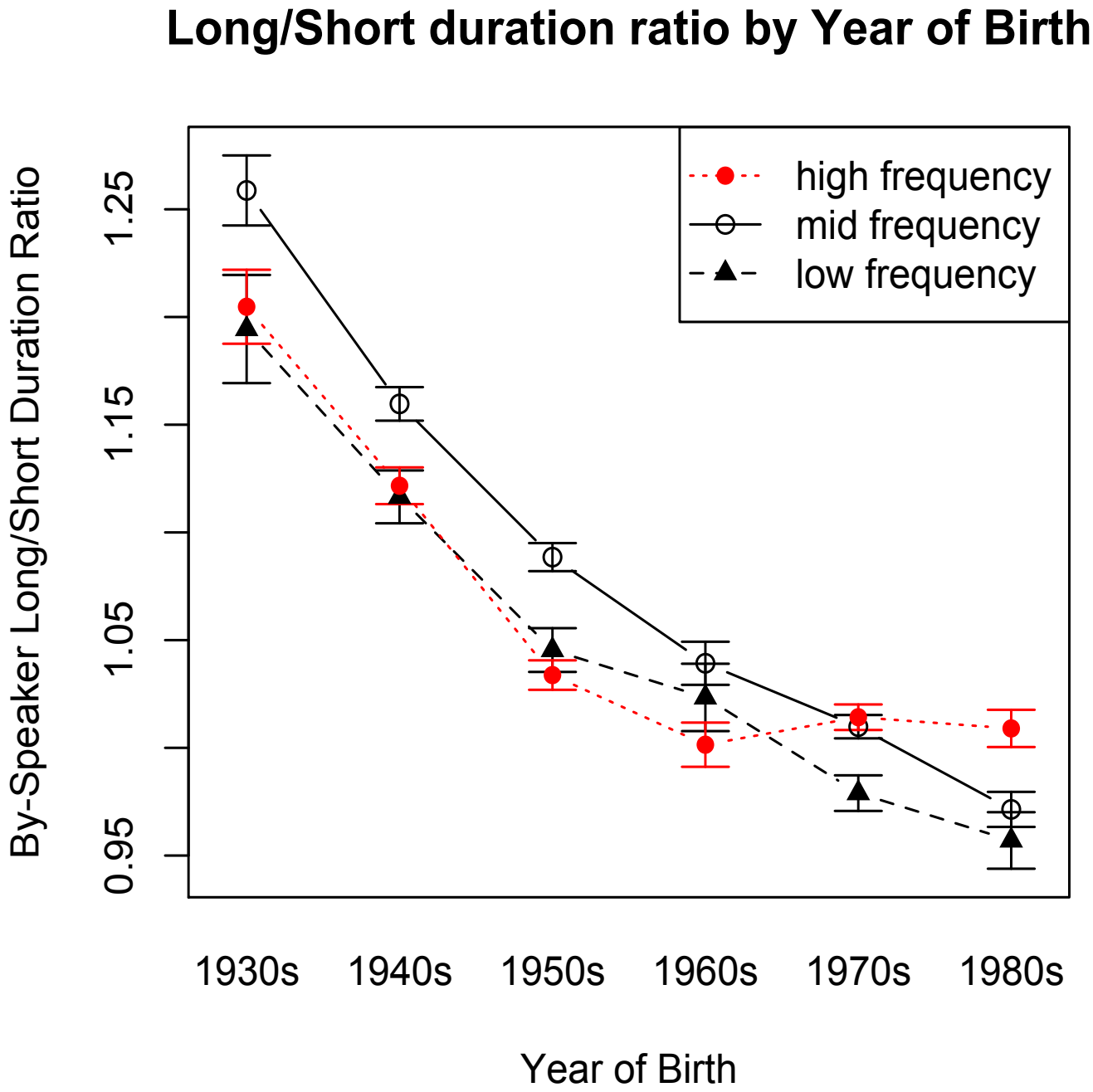
  http://korean.utsc.utoronto.ca/kpa/

- Given a sound file and its Korean transcription, the aligner provides a segmentation of the file into component phones.


3. TextGrid fv01_t09_s01
File Edit Query View Select Interval Boundary Tier Spectrum Pitch Intensity Formant Pulses Help
p
5.260496
0.3126
-0.0003485
-0.4636
5000 Hz
2152 Hz
0 Hz
500 Hz
75 Hz
phone
(42/42)
word
(10)
1
2
sp
y e a e e o a a s o e a j a a u n n g i ss ss yo
sp
sp
YESNA LE
EO NE
BAD A
SOGE
AJU
AREUMDA UN
YONG GUNG
ISSEOSSEOY O
sp
4.615602
0.632394
0.644894
Visible part 4.675244 seconds
5.320138
0.512362
Total duration 5.820000 seconds

IV. Results


Long/Short duration ratio by Year of Birth
By-Speaker Long/Short duration ratio
1.00
1.05
1.10
1.15
1.20
1.25
1930s
1940s
1950s
1960s
1970s
1980s
Year of Birth


Long/Short duration ratio by Year of Birth
By-Speaker Long/Short Duration Ratio
0.95
1.05
1.15
1.25
high frequency
mid frequency
low frequency
1930s
1940s
1950s
1960s
1970s
1980s
Year of Birth

Linear Mixed Effects Model

- Dependent variable: Vowel duration (log)
- Factors of main interest
  - YOB + YOB^2 (Year of Birth, quadratic)
  - Vowel length
  - Frequency (low, mid, high)
  - Vowel Length * [YOB + YOB^2]
  - Vowel Length * [YOB + YOB^2]* Frequency
- Control factors
  - Preceding consonants
  - Vowel height
  - Syllable structure
  - Position in prosodic phrase
  - Word length
  - Gender
- Random effects
  - Word: intercept and slope adjustment to [YOB + YOB^2]
  - Speaker: intercept
- 59.8% of variance explained by the model
- All fixed effects and random effects came out significant.

Results

- Vowel Length * [YOB+YOB^2]
  - As YOB increases, vowel length contrast reduces.
  - The reduction rate slows down toward the youngest group.
- Frequency
  - Higher the frequency, shorter the duration.
  - Duration: Low freq. > Mid freq. > High freq.
- Vowel Length * [YOB+YOB^2]* Frequency
  - Reduction slows down/flattens for high frequency words, compared to mid-low frequency words.
  - Difference between mid and low frequency words, not siginificant

Conclusion

- High frequency words resist the change, opposite of the synchronic frequency effect on duration.
- But, this difference is found mostly in the younger group, where change is already fairly advanced.
- Maybe this “analytical” frequency effect shows up only as change is close to being completed.
- Highlights the need to examine the frequency effect in change as it unfolds over time, rather than the final outcome or the pattern at a static point in time.

Thank you!

Thanks to:

- *The Social Sciences and Humanities Research Council of Canada—Partnership Development Grant* (# 890-2012-25)
- *The National Institute of the Korean Language*
- *The University of Toronto Scarborough Library Digital Scholarship Unit*
- Research Assistants: Yaruna Cooblal, Sohyun Hong, Roobika Karunananthan, Julianna So, Shawna-Kaye Tucker, and Cindy Yee
- *LVC group* and *Phonology/Phonetics group* at *the University of Toronto*