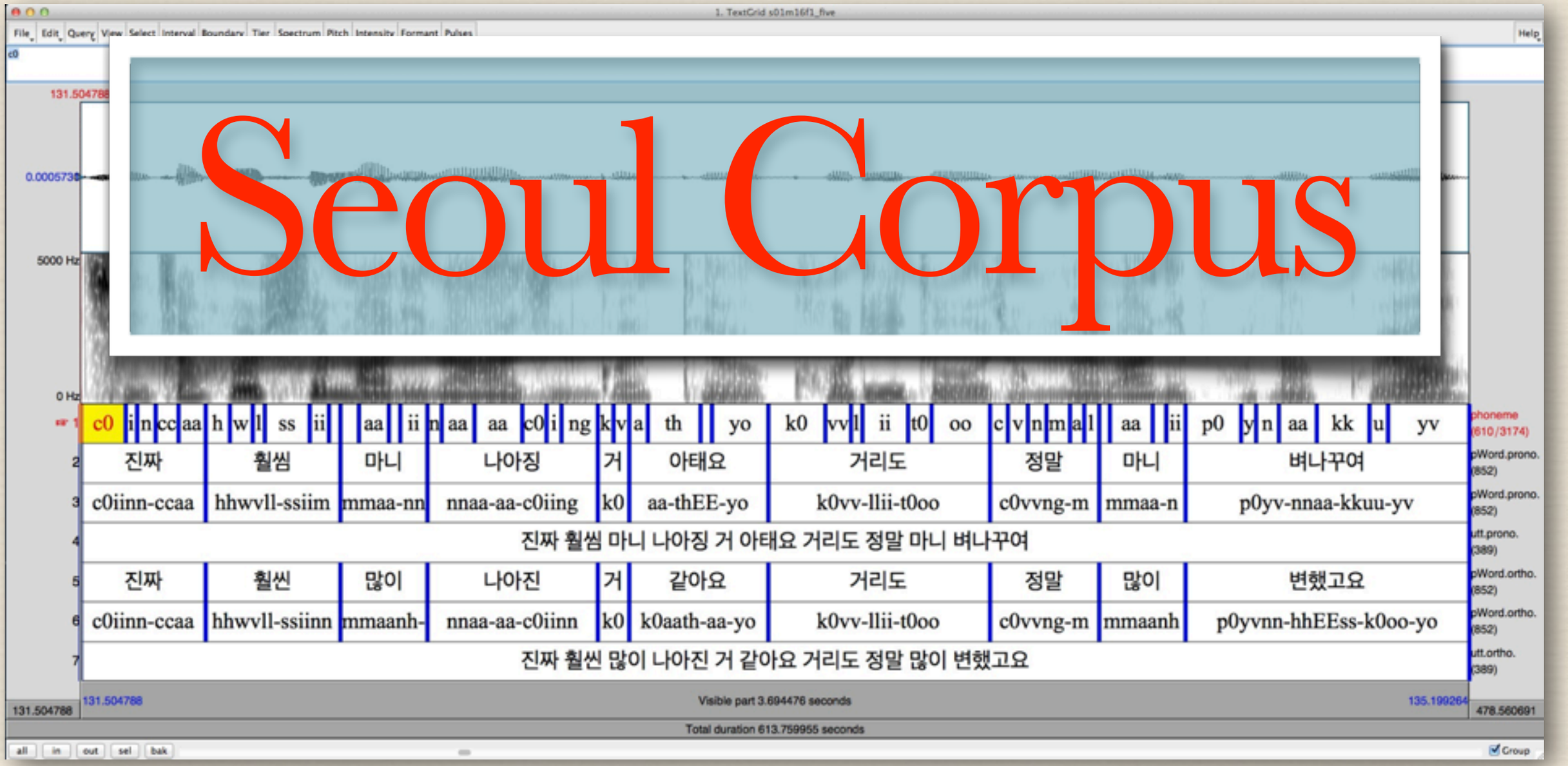
한국어 자연발화 음성코퍼스

The Korean Corpus of Spontaneous Speech

윤원희, 윤규철, 박선우, 이주희, 조성문, 강덕수

2015. 03. 21

This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2012S1A5A2A03034027).



한국어 자연발화 음성코퍼스

The Korean Corpus of Spontaneous Speech

한국어 자연발화 음성코퍼스

The Korean Corpus of Spontaneous Speech

Project Personnel 1

Inspired by the creators of the Buckeye corpus…

Corpus Recording 2

Corpus Transcription 3

Using Praat & HTK

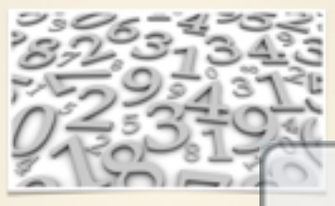
Agreement among Transcribers 4

We got nine of them!

CONTENTS

Applications 5

in pure and applied areas of study



Corpus Statistics 6

Corpus Tools 7

Corpus Availability 8

Since early 2015

Seoul Corpus

Corpus Summary

- 40 speakers of Seoul Korean interviewed.
- 10 speakers each (5 males and 5 females) in their teens, twenties, thirties, and forties.
- Recordings transcribed in Korean hangul orthography and with symbols for phonemes.
- Sampled at 44kHz with 16-bit quantisation in .wav files, each file having matching Praat TextGrid file.
- 40 hours, 231,632 phrasal word tokens (51,443 types) and 1,134,781 phonemes.
- Available for free to the research community from March 2015.

Project Summary

- Funded for two years by the Korea Research Foundation (2012.09.01 ~ 2014.08.31, NRF-2012S1A5A2A03034027)
- Project Title: “A Study for Building a Korean Corpus of Spontaneous Speech”
- Six professors from five universities, three full-time researchers and over ten research assistants
- Inspired by the creators of the Buckeye Corpus

Project Personnel

Inspired by the creators of the Buckeye corpus…

1 Personnel | 2 Recording | 3 Transcription | 4 Agreement | 5 Applications | 6 Statistics | 7 Tools | 8 Availability

PIs, researchers, and RAs

- Professor Yun, Weonhee
  Professor Yoon, Kyuchul
  Professor Park, Sunwoo
  Professor Lee, Juhee
  Professor Cho, Sungmoon
  Professor Kang, Ducksoo

- Post-doctoral researchers
  Hahn, Hyeseung
  Byun, Kunhyuck
  Kim, Jungsun

- Graduate research assistants
  Kim, Soonok
  Lee, Yuri
  Lee, Jinhee
  Chung, Hyejung
  Choi, Subin

Corpus Recording

Speakers of Seoul Korean

- Native speakers of Seoul Korean whose parents also born and raised in Seoul or Gyeonggi area.
  (some exceptions where one of the parents moved to the area before the graduation of their elementary school)
- Class was not strictly controlled.
- Recordings made from Sept. 2012 to end of 2013.

Recruitment of target speakers

- Target speakers recruited by
  1. advertisements
  2. referrals from target speakers
  3. referrals from project members

- Target speakers
  1. were told the purpose of this project
  2. signed the agreement sheet
  3. were rewarded financially

  (This procedure was approved by the Internal Review Board)

Speaker groups

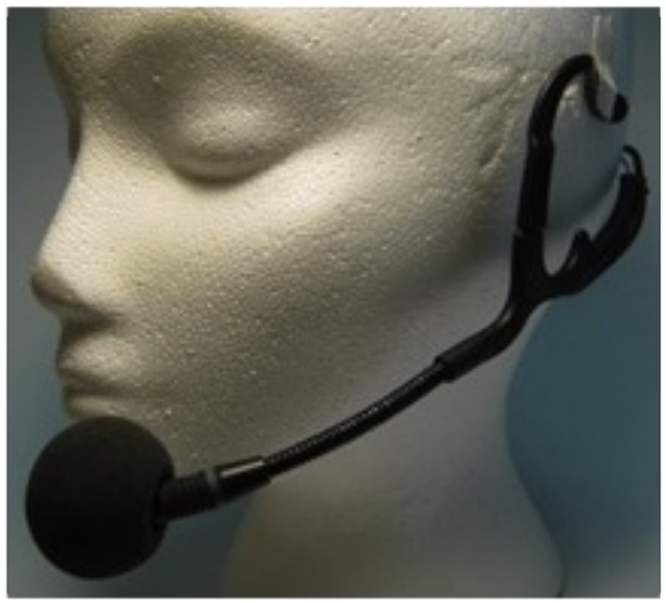
in their	male	female	total
10s	5	5	10
20s	5	5	10
30s	5	5	10
40s	5	5	10
	20	20	40

Speakers recorded

speaker No.	in their	age	gender	interviewer gender	height (cm)	weight (kg)	speaker No.	in their	age	gender	interviewer gender	height (cm)	weight (kg)
s01	10s	16	m	f	172	54	s21	30s	31	m	m	177	79
s02		16		f	175	62	s22		37		m	176	63
s03		15		m	166	50	s23		36		f	181	78
s04		15		m	175	84	s24		36		f	176	81
s05		16		f	179	70	s25		32		f	170	90
s06		18	f	m	163	49	s26		32	f	f	165	51
s07		16		m	167	50	s27		32		f	159	51
s08		16		m	167	51	s28		34		m	168	52
s09		17		f	171	55	s29		37		m	163	57
s10		18		f	169	59	s30		38		m	162	60
s11	20s	25	m	f	162	58	s31	40s	43	m	m	171	75
s12		23		f	183	70	s32		43		m	170	67
s13		26		m	182	92	s33		44		m	170	72
s14		23		m	177	85	s34		47		f	181	88
s15		22		m	179	64	s35		43		f	160	68
s16		22	f	m	158	49	s36		43	f	m	159	48
s17		24		f	159	52	s37		46		m	160	60
s18		27		m	162	48	s38		46		f	150	54
s19		24		f	160	53	s39		43		f	165	60
s20		24		f	160	47	s40		43		f	162	55

Recording equipments

- Recordings made in the recording studio of the Department of English Language & Literature, Hanyang University, South Korea
- TASCAM HD-P2 recorder
- AKG C420 headworn microphone
- .wav audio files sampled at 44kHz, 16-bit



Recording of target speakers

- Interviews were conducted in the recording room by the male/female interviewer.
- Only the speaker wore the head-worn microphone.
- Small talk for adjusting the recording level.

Topics covered in interviews

1. Tell us about
   yourself
   when and where you were born and related stories

2. Tell us about
   your family members
   their personality, what they do and related stories

3. Tell us about
   your place, type of residence and community,
   e.g. where you shop
   your neighbors and stories about them

4. Tell us about
   your school or workplace and study- or work-related stories
   your friends, teachers, colleagues or bosses at school or work
   what your friends talk about
   where you hang out with your friends and what you do

5. Tell us about
   your opinion on various political issues
   your thoughts on past or recent political elections
   your thoughts on expressing political views on the internet

6. Tell us about
   how much money you get or spend every week or month
   your thoughts on past or recent (inter)national economic crises
   your thoughts on your current financial situation or status
   your thoughts on the rich and the poor

7. Tell us about
   how you spend your leisure time, e.g. going to the movies/plays
   your domestic or international travel experiences
   your thoughts or experiences on multi-cultural families
   your favorite online or offline games
   your experiences on smartphones

Recording times & file formats

- Recording lasted for approximately
  1 hour for each speaker,
  40 hours in total.

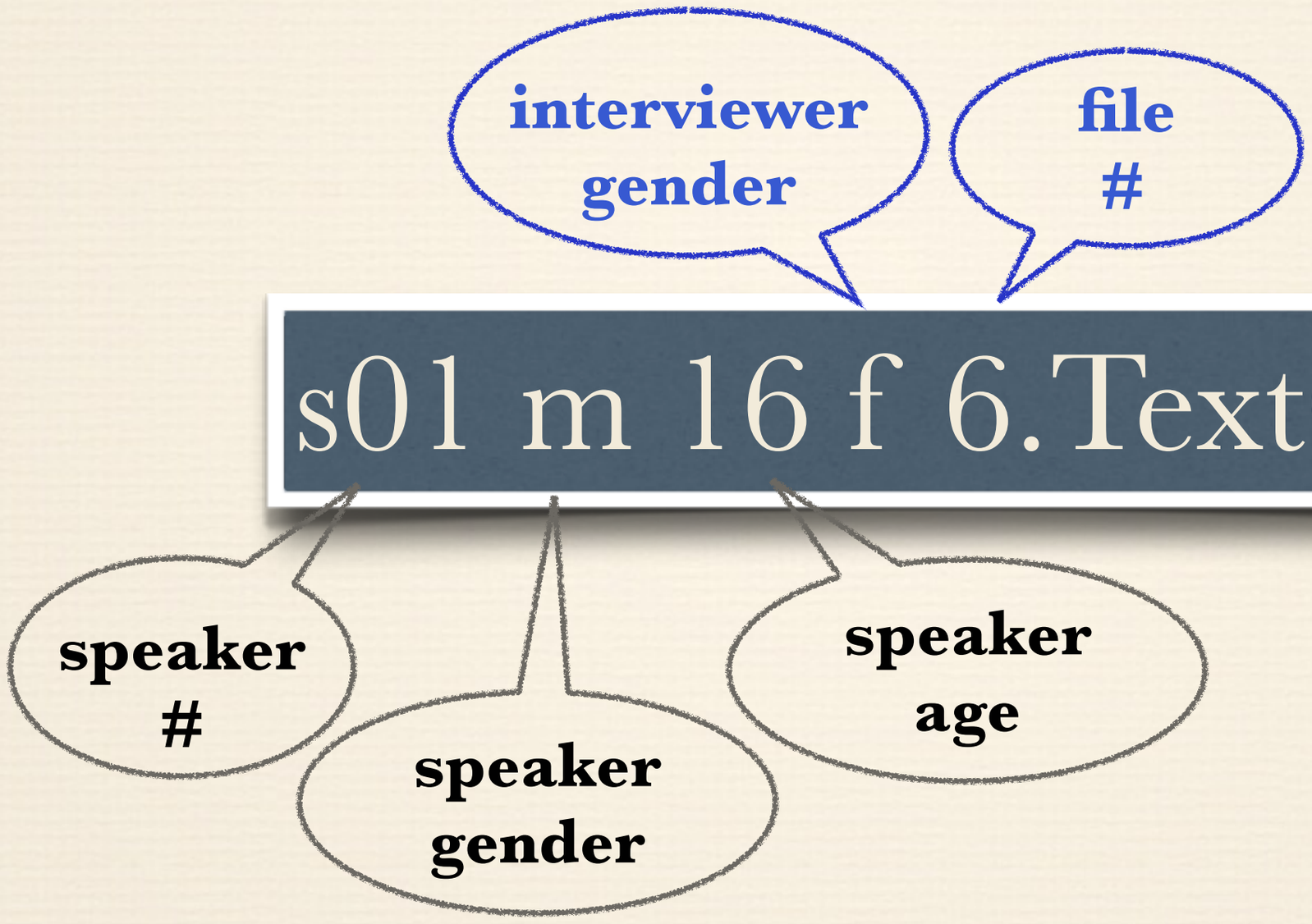
- Each one-hour recording divided into
  6 ten-min audio files, 240 files in total.
  (see Figure)

- file naming convention (for one speaker):

s01m16f1.wav		s01m16f1.TextGrid
s01m16f2.wav		s01m16f2.TextGrid
s01m16f3.wav		s01m16f3.TextGrid
s01m16f4.wav	&	s01m16f4.TextGrid
s01m16f5.wav		s01m16f5.TextGrid
s01m16f6.wav		s01m16f6.TextGrid

File naming

- File names contain information about recording



Corpus Transcription

Using Praat & HTK

Korean phoneme set

phoneme	IPA	hangul onset	hangul coda
p0	p	ㅂ	ㅂ
ph	p^h	ㅍ	
pp	p'	ㅃ	
t0	t	ㄷ	ㄷ
th	t^h	ㅌ	
tt	t'	ㄸ	
k0	k	ㄱ	ㄱ
kh	k^h	ㅋ	
kk	k'	ㄲ	

phoneme	IPA	hangul onset	hangul coda
s0	s	ㅅ	
ss	s'	ㅆ	
hh	h	ㅎ	
c0	tɕ	ㅈ	
ch	$tɕ^h$	ㅊ	
cc	tɕ'	ㅉ	
mm	m	ㅁ	ㅁ
nn	n	ㄴ	ㄴ
ng	ŋ		ㅇ
ll	l	ㄹ	ㄹ

CONSONANTS

phoneme	IPA	hangul nucleus
ii	i	ㅣ
ee	e	ㅔ, ㅐ
aa	a	ㅏ
xx	ɨ	ㅡ
vv	ə	ㅓ
uu	u	ㅜ
oo	o	ㅗ

phoneme	IPA	hangul nucleus
ye	je	ㅖ, ㅒ
ya	ja	ㅑ
yv	jə	ㅕ
yu	ju	ㅠ
yo	jo	ㅛ
wi	wi	ㅟ
we	we	ㅚ,ㅙ,ㅞ
wa	wa	ㅘ
wv	wə	ㅝ
xi	ɨi	ㅢ

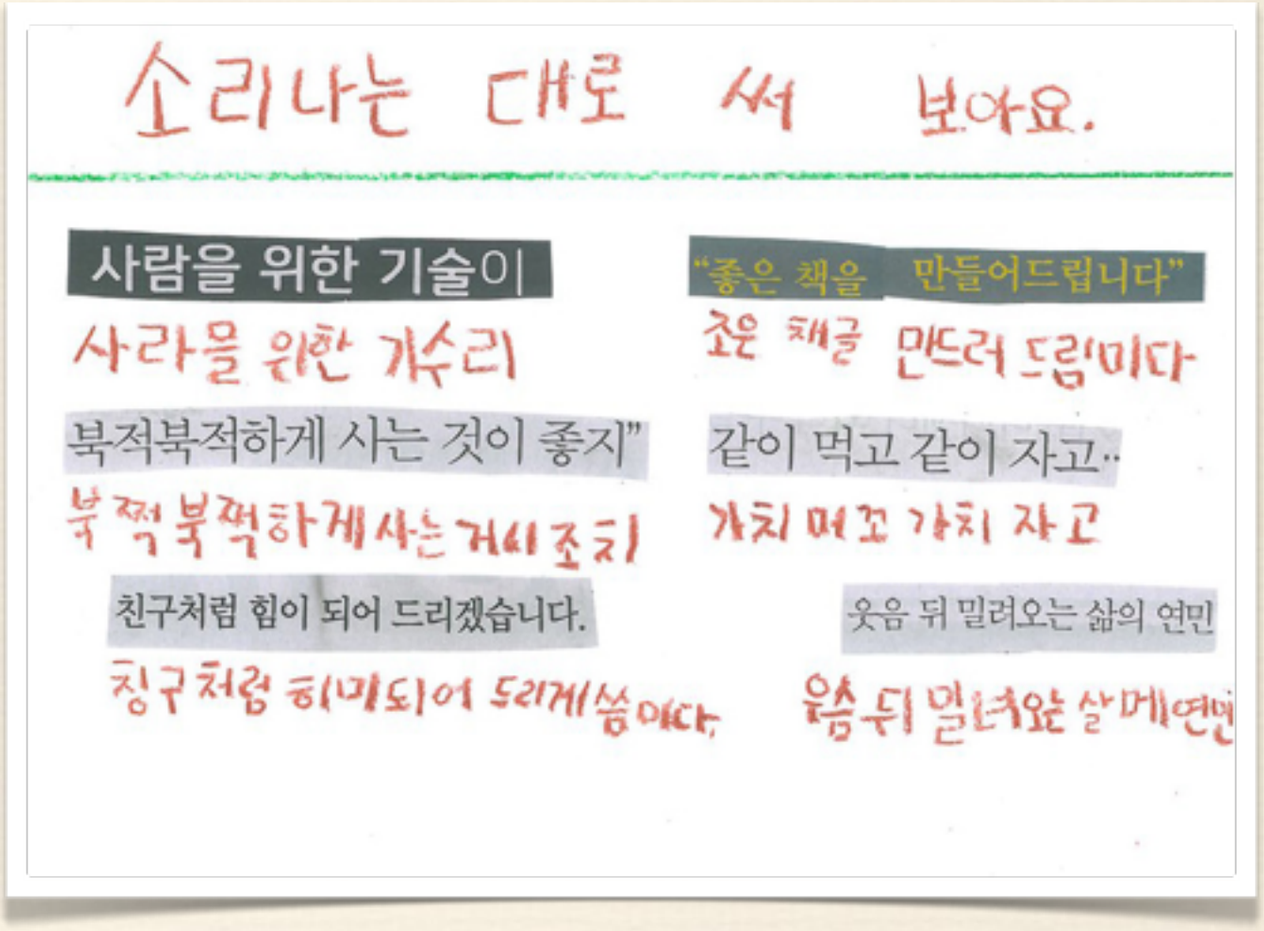
VOWELS

Non-speech labels:

<IVER>, <SIL>, <VOCNOISE>, <LAUGH>, <NOISE>, <UNKNOWN>, <PRIVATE.INFO>

Transcribed as pronounced

- Recordings were transcribed in Korean hangul orthography by RAs.
- Transcribed by the utterance.
- Native Korean RAs acted as the human speech recognizers for phoneme identification.



Phoneme boundaries placed by ASR

- Output from the human phoneme identification fed into the ASR (Yun, 2003) for segmentation, i.e. placing phoneme boundaries automatically.

Automatic Speech Recognition

recorded speech & transcribed text

pronunciation database

monophone HMM

ASR training

dictionary w/ pronunciation variation

Viterbi alignment

phones segmented & annotated

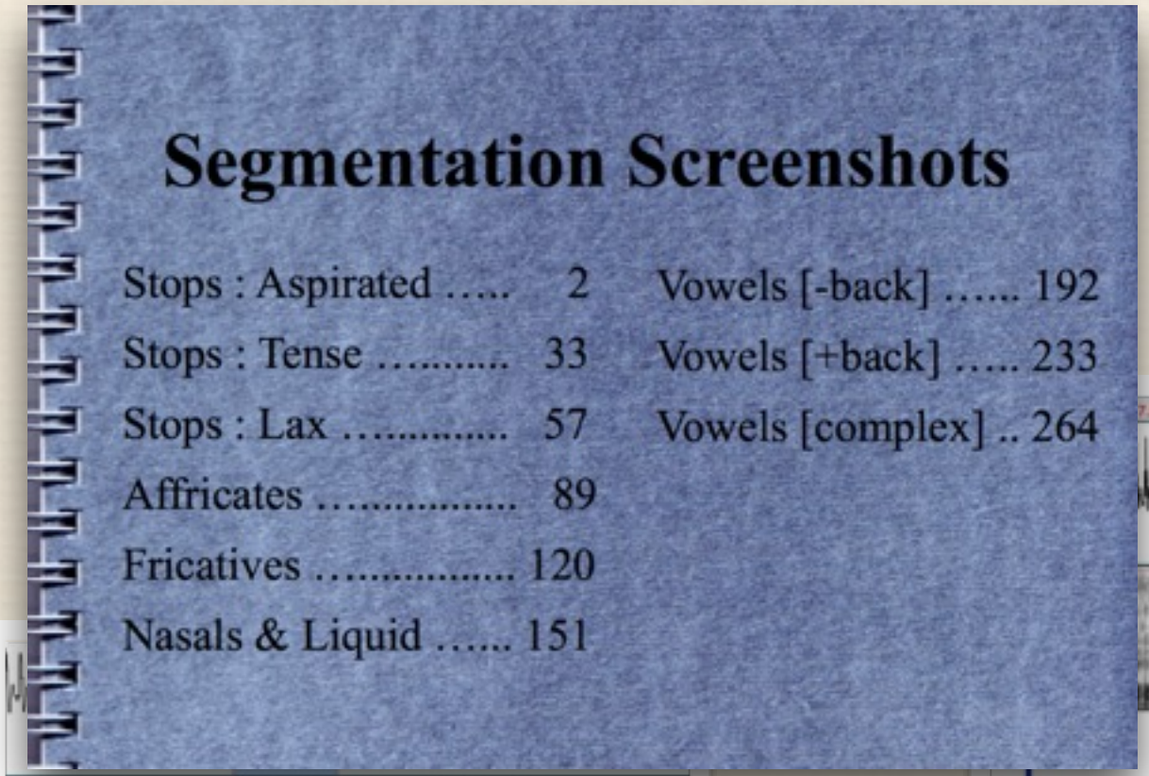
Hand realignment

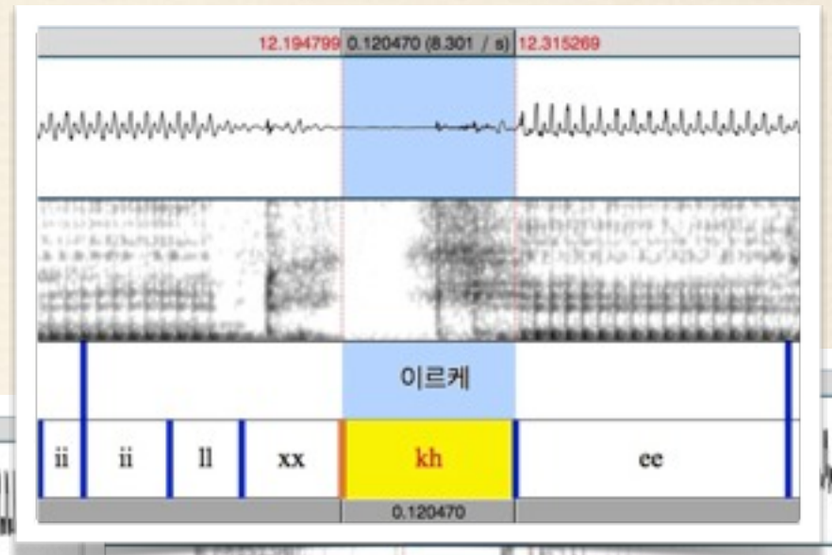
- Corrections and readjustments made for the phoneme identification and segmentation by the 9 transcribers.

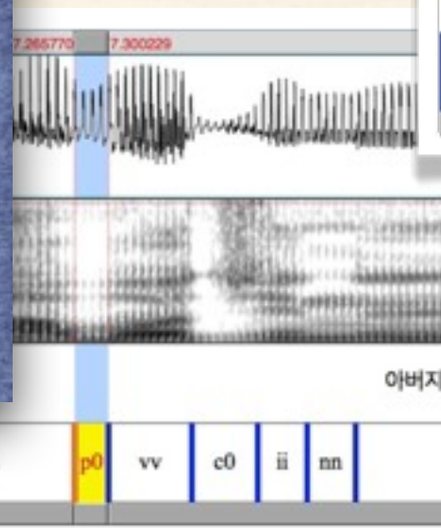
- Phoneme symbols replaced and boundaries for phonemes, phrasal words and utterances readjusted.

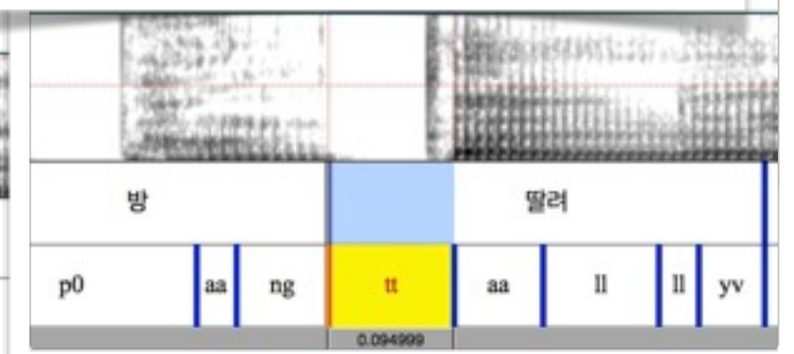
- Various errors were corrected by hand/scripts, e.g. boundary sync, word spacing, etc.

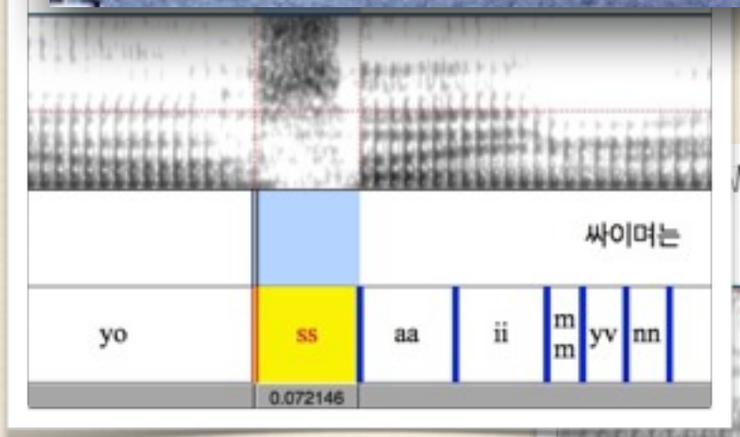
Segmentation guidelines

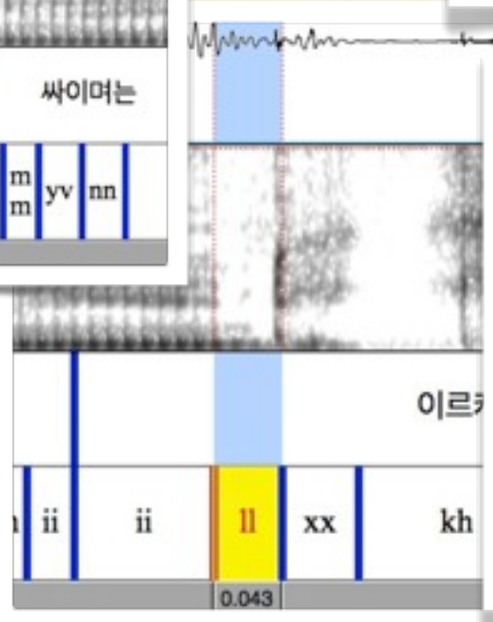


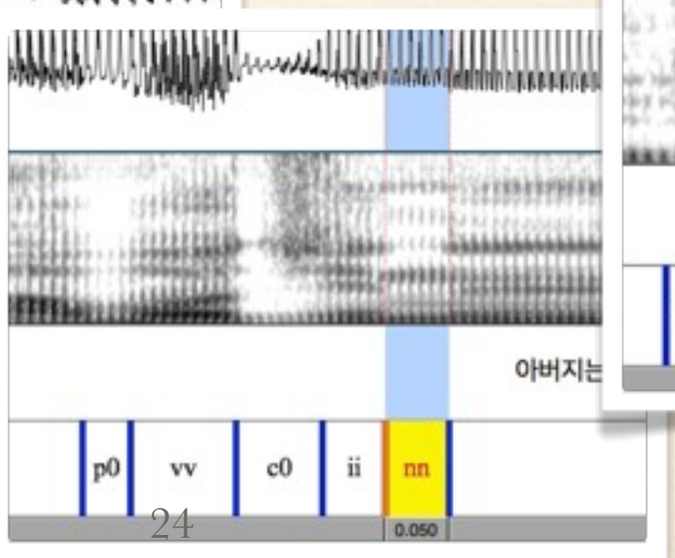


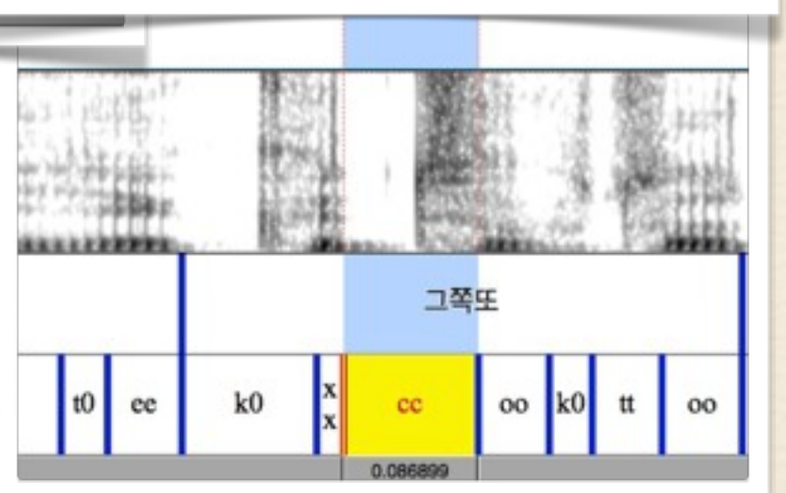




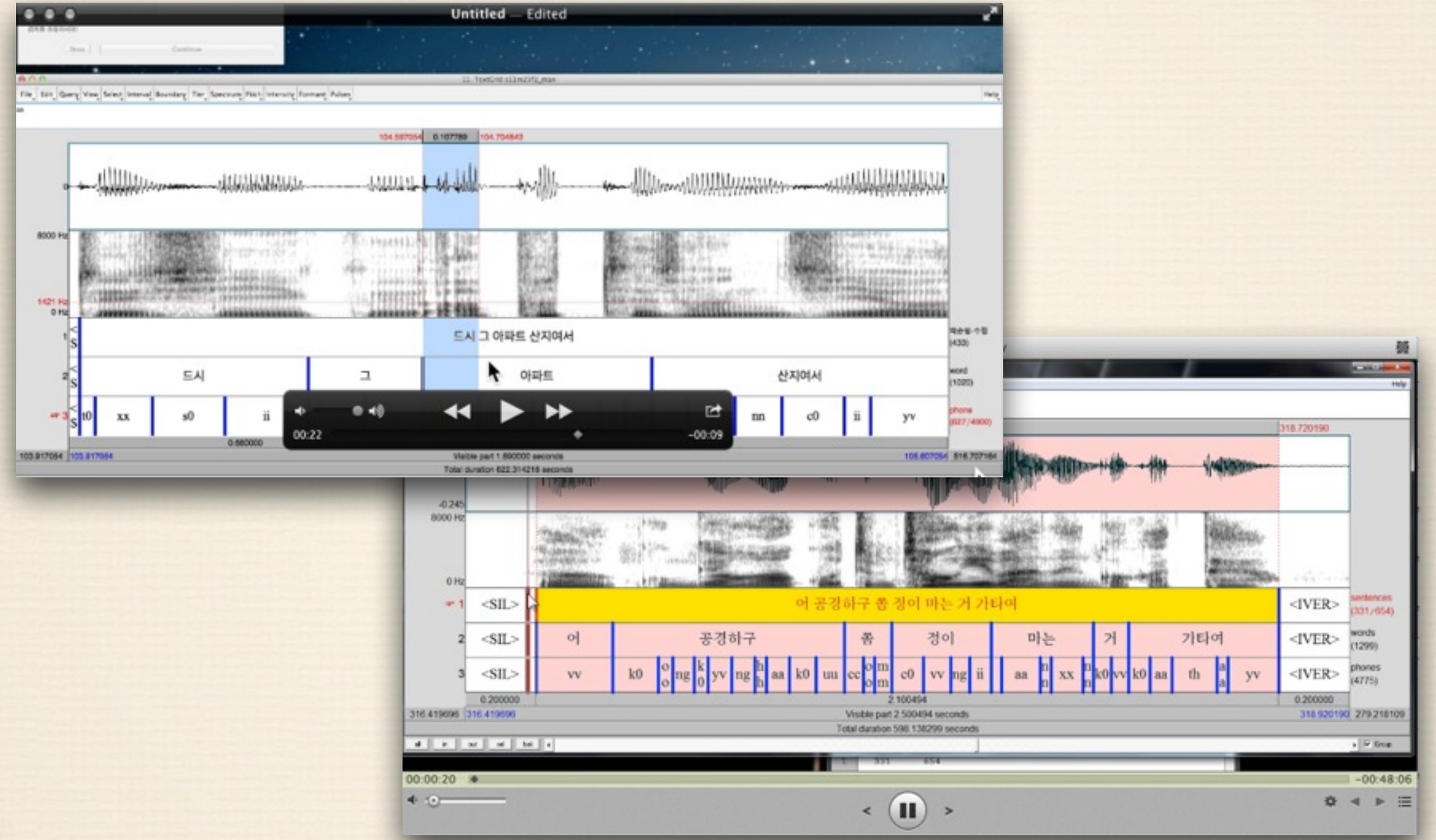








Segmentation video tutorials



Graduate course for segmentation

* one-semester hands-on class on hand realignment to train RAs

RAs

Romanization scheme used

roman	IPA	hangul onset	hangul coda
p0	p	ㅂ	ㅂ
ph	p^h	ㅍ	ㅍ
pp	p'	ㅃ	
t0	t	ㄷ	ㄷ
th	t^h	ㅌ	ㅌ
tt	t'	ㄸ	
k0	k	ㄱ	ㄱ
kh	k^h	ㅋ	ㅋ
kk	k'	ㄲ	ㄲ
ks	ks		ㄱㅅ
nc	ntɕ		ㄴㅈ
nh	nh		ㄴㅎ
lk	lk		ㄹㄱ
lm	lm		ㄹㅁ
lp	lp		ㄹㅂ
ls	ls		ㄹㅅ
lT	lt^h		ㄹㅌ
lP	lp^h		ㄹㅍ
lh	lh		ㄹㅎ
ps	ps		ㅂㅅ

roman	IPA	hangul onset	hangul coda
s0	s	ㅅ	ㅅ
ss	s'	ㅆ	ㅆ
hh	h	ㅎ	ㅎ
c0	tɕ	ㅈ	ㅈ
ch	$tɕ^h$	ㅊ	ㅊ
cc	tɕ'	ㅉ	
mm	m	ㅁ	ㅁ
nn	n	ㄴ	ㄴ
ng	ŋ		ㅇ
ll	l	ㄹ	ㄹ

CONSONANTS

roman	IPA	hangul nucleus
ii	i	ㅣ
ee,EE	e	ㅔ,ㅐ
aa	a	ㅏ
xx	ɨ	ㅡ
vv	ə	ㅓ
uu	u	ㅜ
oo	o	ㅗ

roman	IPA	hangul nucleus
ye,YE	je	ㅖ,ㅒ
ya	ja	ㅑ
yv	jə	ㅕ
yu	ju	ㅠ
yo	jo	ㅛ
wi	wi	ㅟ
wE,WE,we	we	ㅚ,ㅙ,ㅞ
wa	wa	ㅘ
wv	wə	ㅝ
xi	ɨi	ㅢ

VOWELS

- Information
  on syllabic boundaries
  in orthographic/pronounced
  forms
  preserved as hyphens.

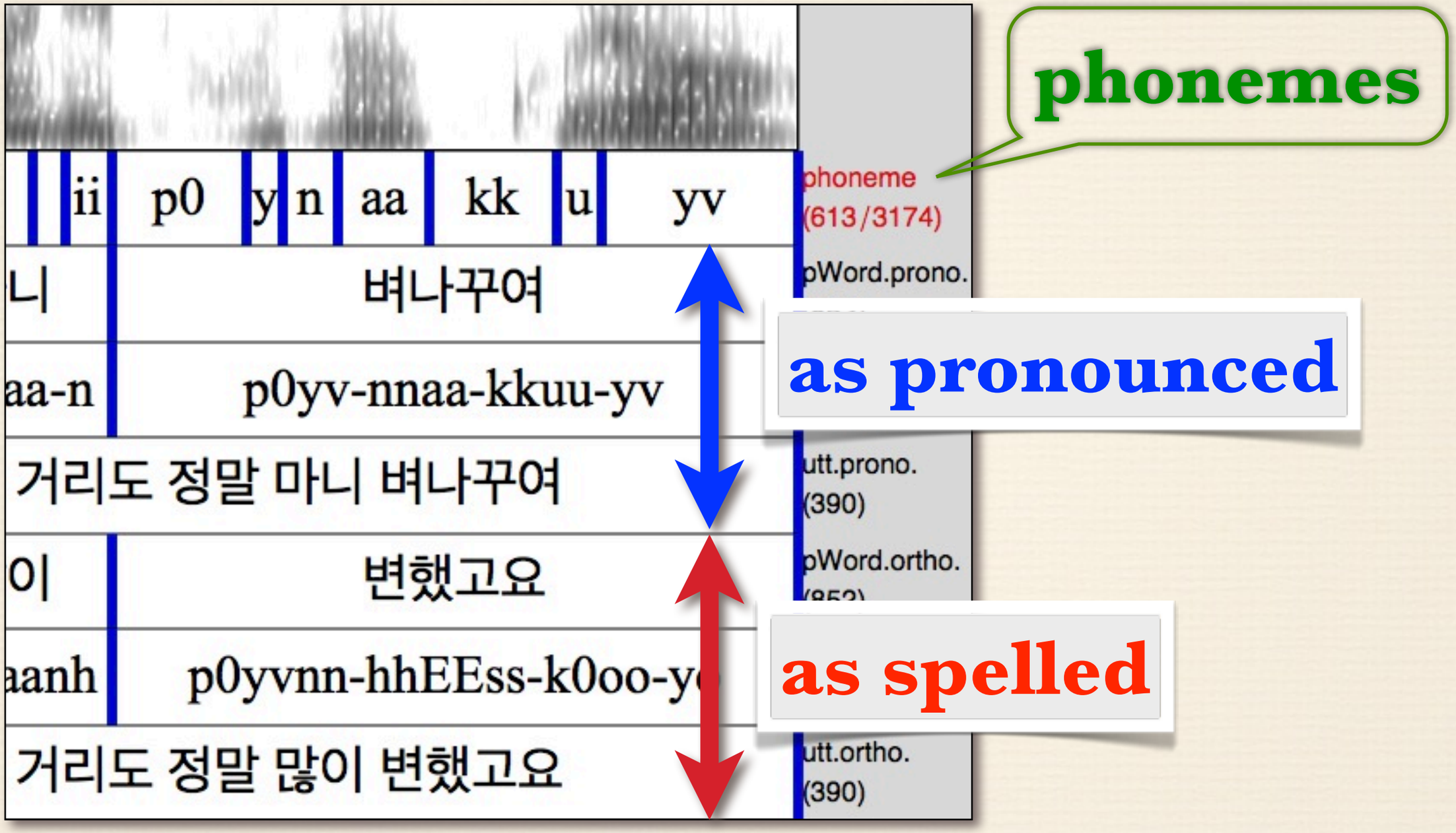
Transcription tiers

All boundaries in tiers were synced

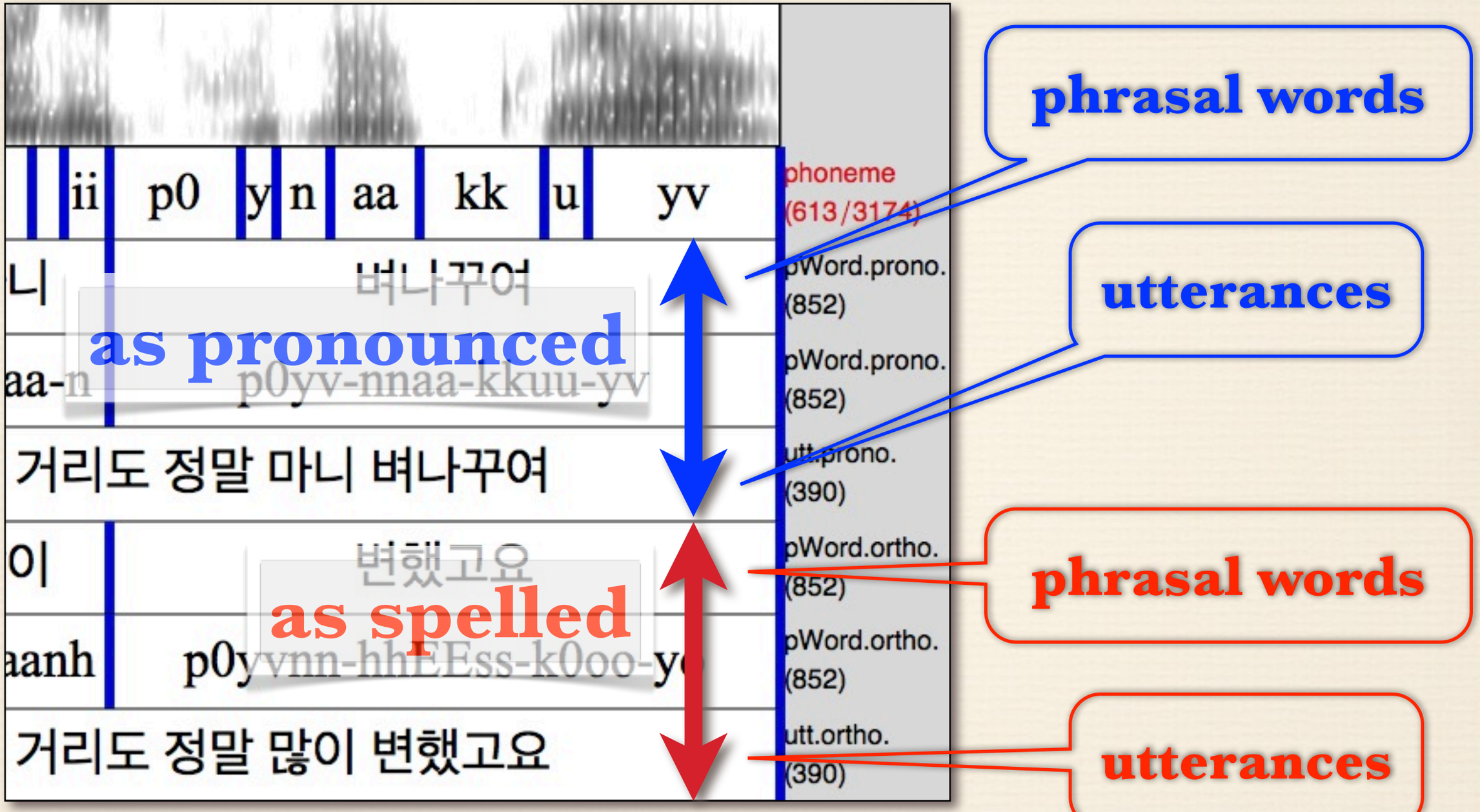
1. *phonemes*
2. *phrasal words*
3. *utterances*
4. *hangul*
5. *romanization*

Pronouncing dictionary for Korean spontaneous speech constructed (51,443 phrasal word entries)

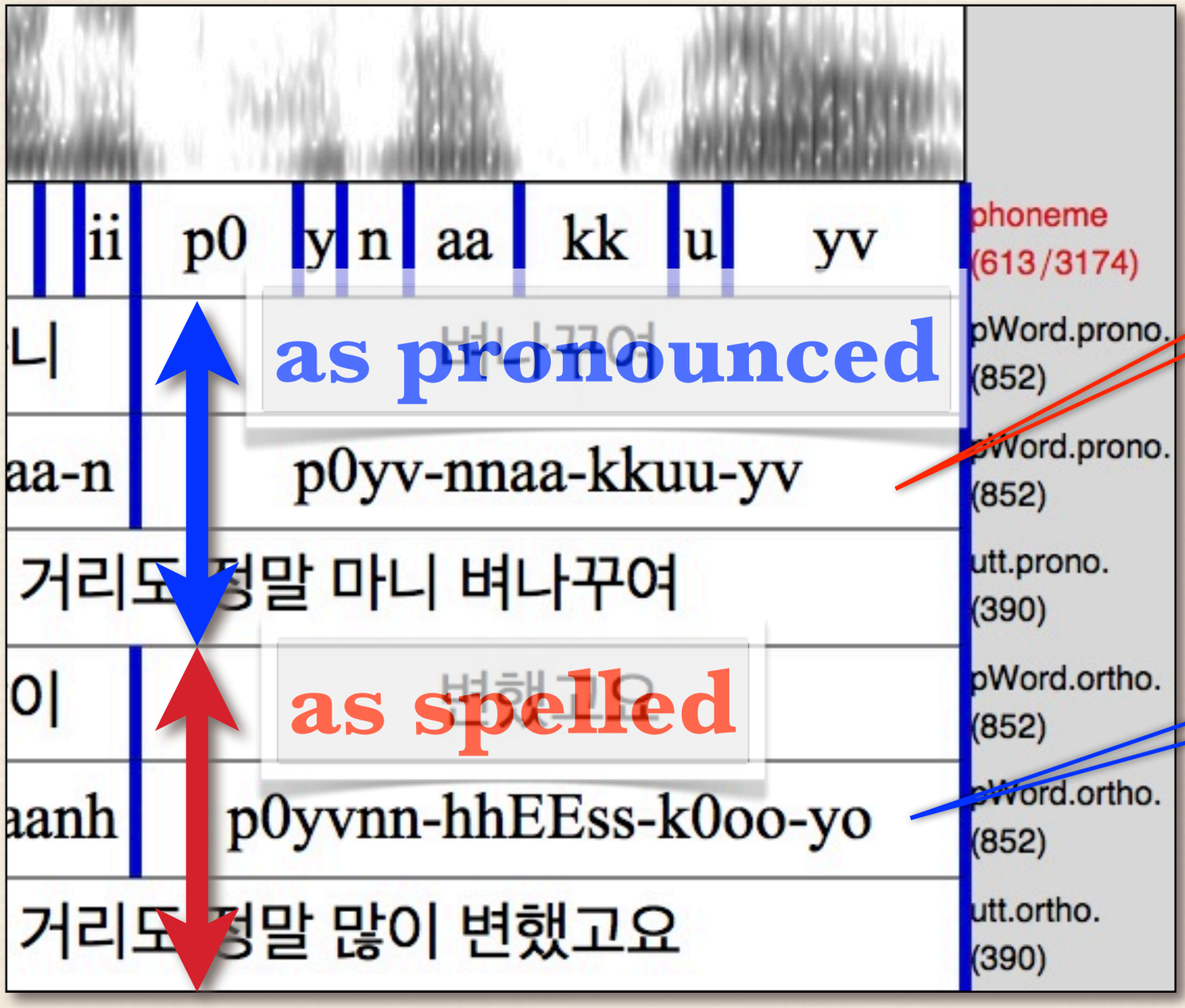
Transcription tiers explained



Transcription tiers explained



Transcription tiers explained



Agreement among Transcribers

We got nine of them!

Test materials extracted from 8 groups

10s - male	random 1 min.
10s - female	random 1 min.
20s - male	random 1 min.
20s - female	random 1 min.
30s - male	random 1 min.
30s - female	random 1 min.
40s - male	random 1 min.
40s - female	random 1 min.

Manual transcription by 9 transcribers

random 1 min.

random 1 min.

random 1 min.

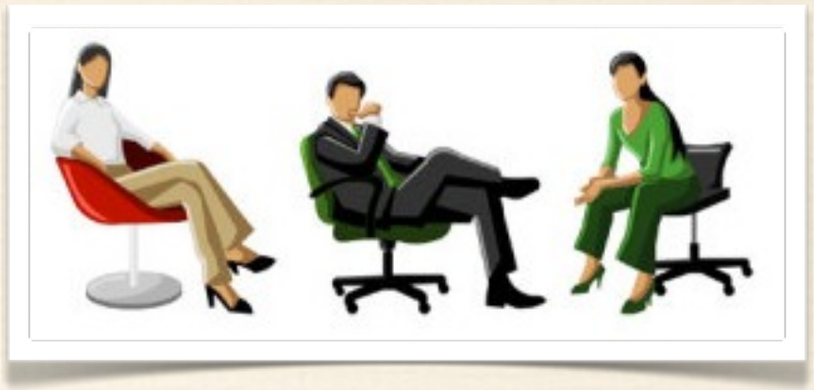
random 1 min.

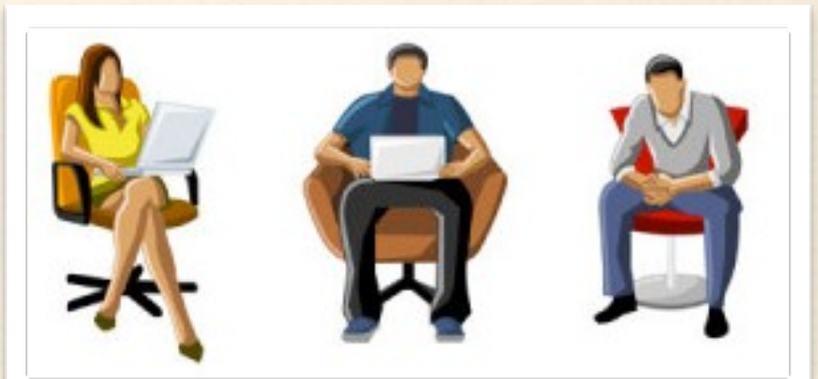
random 1 min.

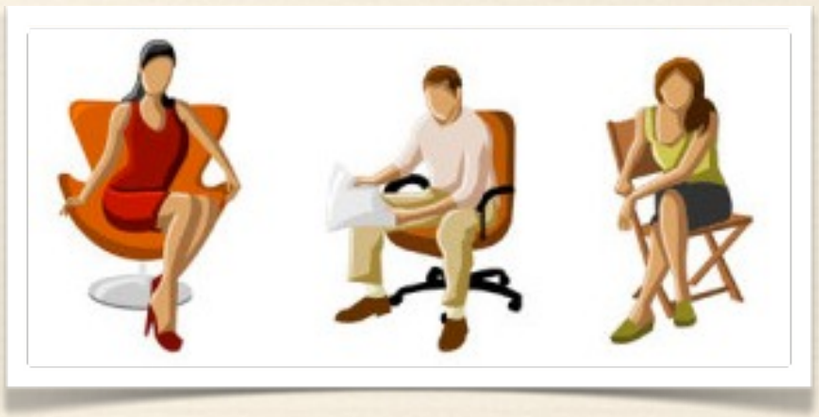
random 1 min.

random 1 min.

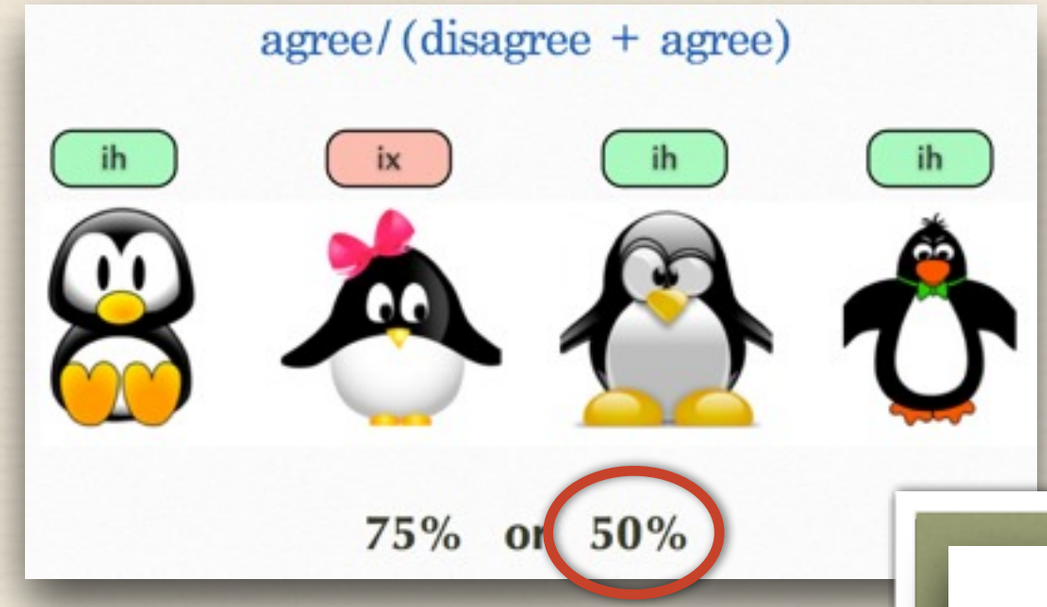
random 1 min.

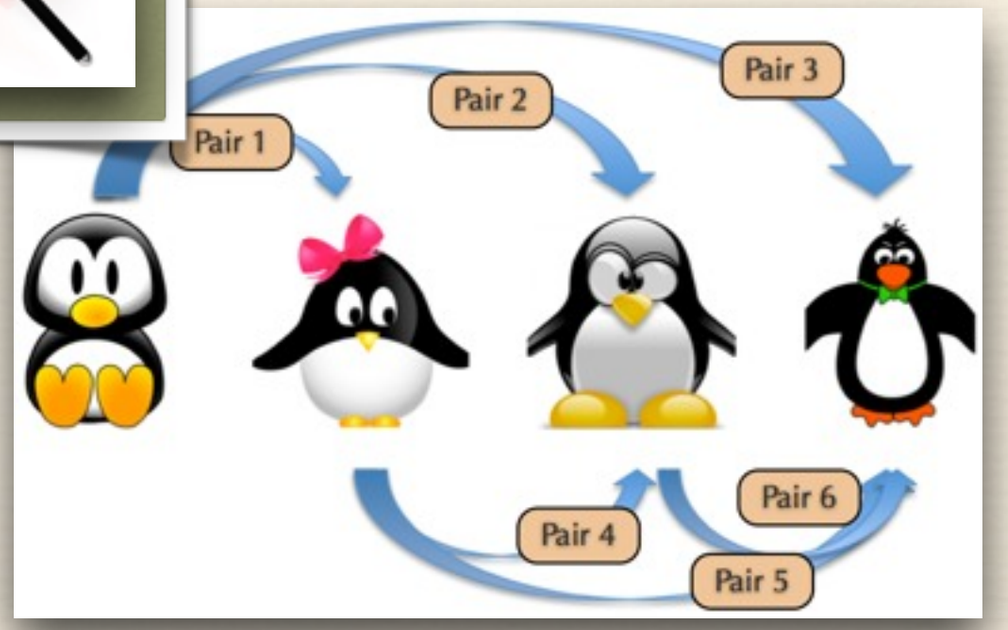






% transcriber pair agreement





Statistics on transcriber agreements

	N	% Agree	Kappa	Max (kappa)	% Unanimous
all	5,152	98.1	0.980	0.995	89.6
stops	1,013	99.1	0.996	1.001	89.8
fricatives	309	98.6	0.991	1.002	85.1
affricates	221	98.3	0.985	0.998	95.5
nasals	902	96.6	0.971	0.991	83.7
liquid	326	99.5	1.000	1.003	91.4
vowels	2,414	97.7	0.977	0.997	90.3

Percent agreement on phoneme identification

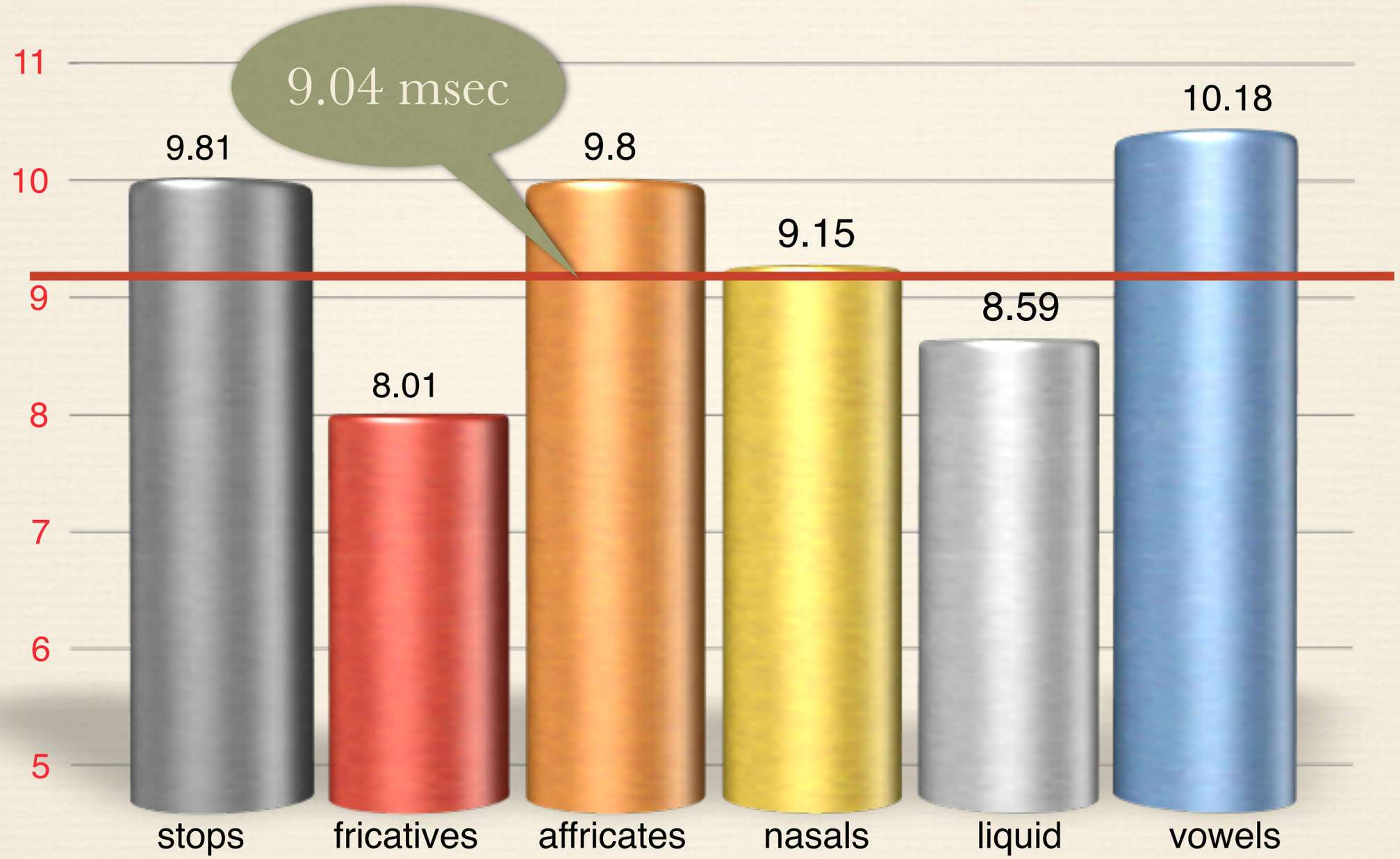
98.1 %

Mean deviation in phoneme segmentation

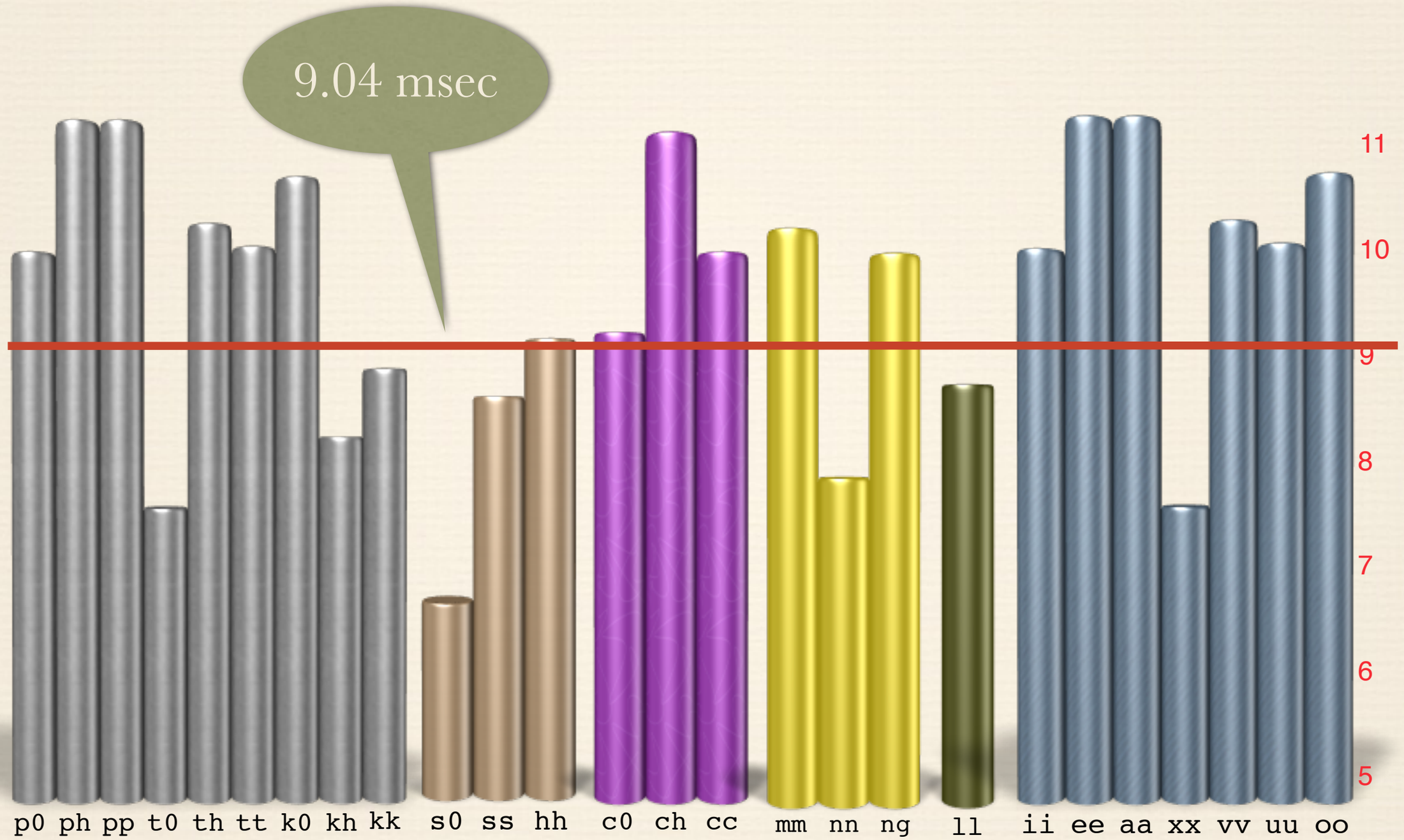
9.04 msec

Agreement does not necessarily mean **correctness**!

Mean deviation in segmentation by phoneme



Mean deviation in segmentation by phoneme



Transcription matrix

		stops									fricatives			affricates			nasals			liquid	vowels																
		p0	ph	pp	t0	th	tt	k0	kh	kk	s0	ss	hh	c0	ch	cc	mm	nn	ng	ll	ii	ee	aa	xx	vv	uu	oo	ye	ya	yv	yu	yo	wi	we	wa	wv	xi
stops	p0	4,082																																			
	ph	26	997																																		
	pp	9		252																																	
	t0	1			7,078																																
	th				12	929																															
	tt	7		7	14	8	1,720																														
	k0	26						15,013																													
	kh							16	820																												
	kk									3,001																											
fricatives	s0										6,272																										
	ss										80	1,645																									
	hh		20					18	8		6		2,055																								
affricates	c0													5,527																							
	ch														1,078																						
	cc													96	20	1,148																					
nasals	mm							2					14				8,062																				
	nn				44				3						8		327	15,079																			
	ng							22	5								133	353	6,220																		
liq	ll				7													36		11,063																
vowels	ii							7												11,087																
	ee				16												8	16			45	10,337														
	aa																			8		50	17,836													
	xx							9									6	40			7	22	38	12,870													
	vv				2												2			7		29	61	145	10,926												
	uu							1									8		8					74	56	5,548											
	oo							5																22	100	333	6,878										
	ye																					193	6				404									
	ya																				8	19	38				1	1,038								
	yv																				14	30		66	48				3,910							
	yu																									16				144						
	yo																								18					860						
	wi																				58										215					
	we																					134											303			
	wa								8									8					31		5									238		
	wv																			1					40		27								242		total pairs
	xi																					8		8											28	178,223

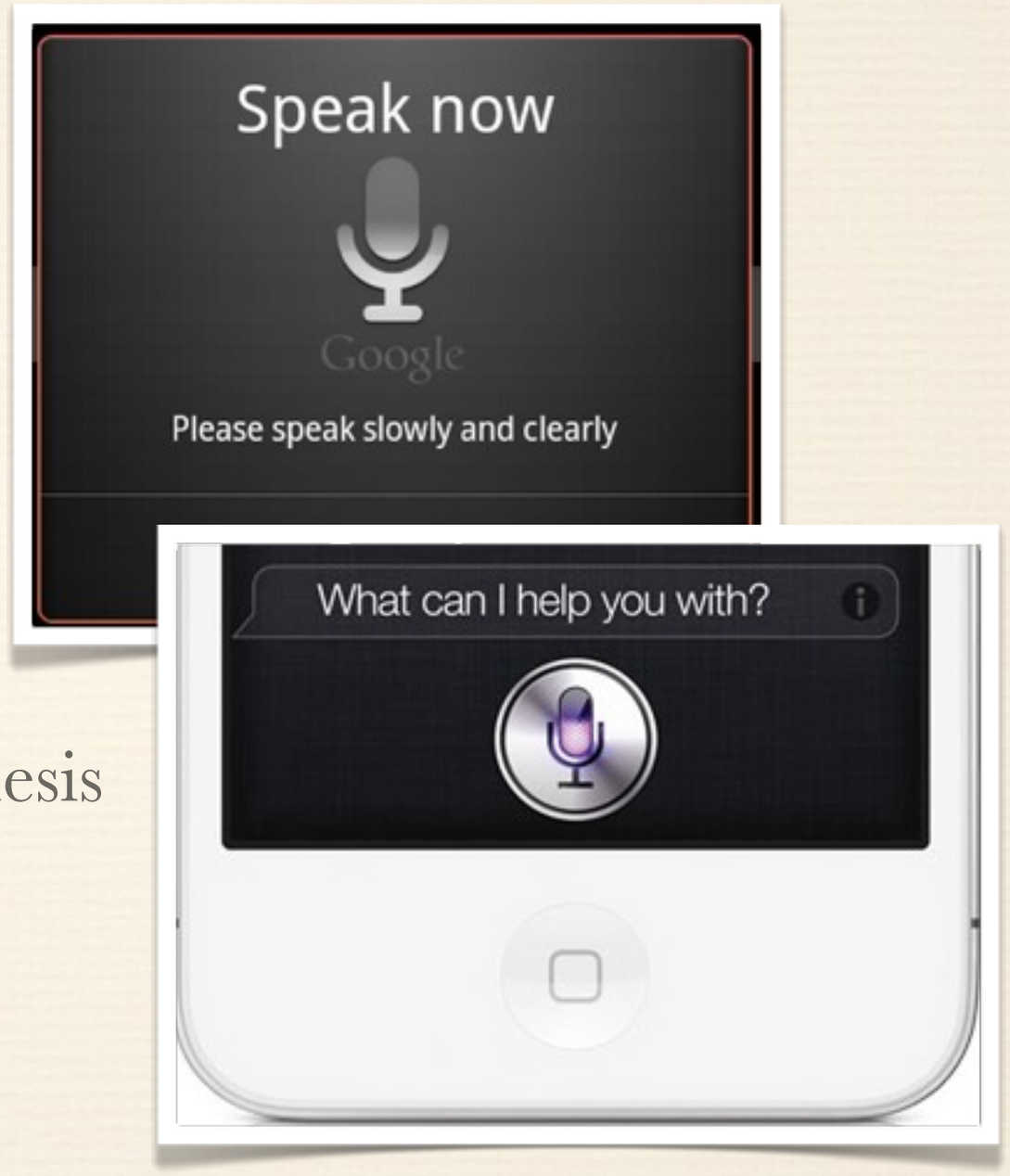
Applications

in pure and applied areas of study

Application

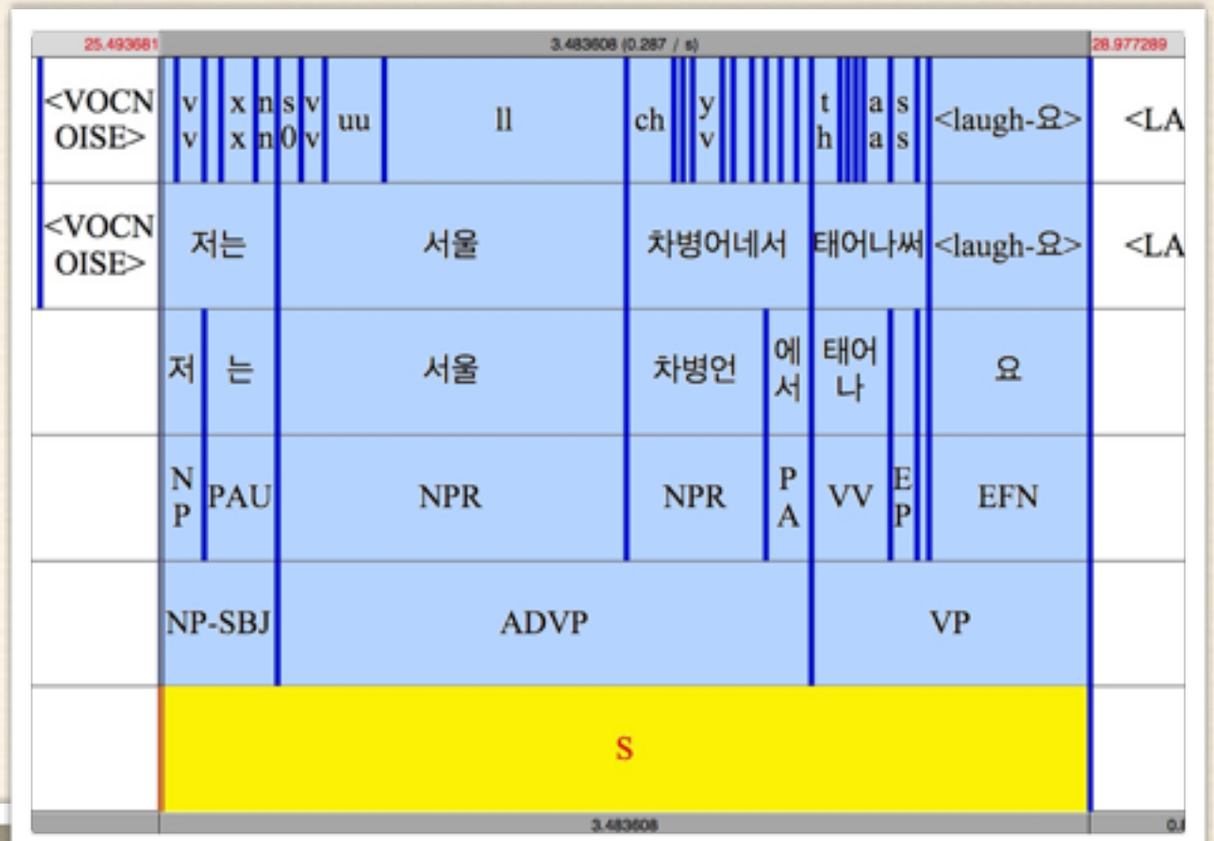
- In general linguistics
  1. Phonetics/Phonology
  2. Morphology/Syntax
  3. Pragmatics, etc.

- In applied linguistics
  1. Speech recognition/synthesis
  2. Education

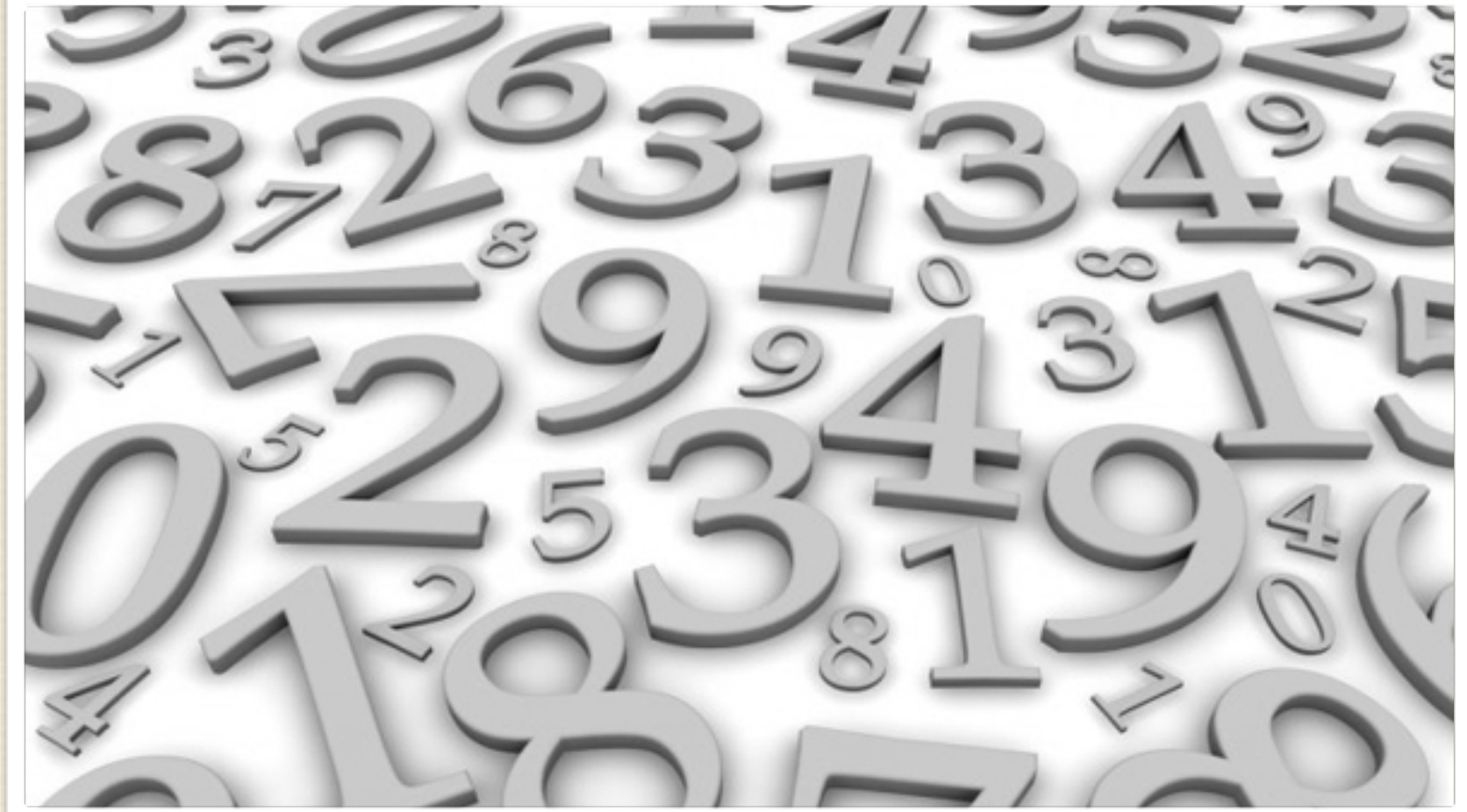


Future Plans

- Morphological/syntactic annotation
  (as in Penn Korean Treebank)



- Pronunciation & grammar database for teaching Korean to speakers of other languages



Corpus Statistics

Corpus sample

1. TextGrid s01m16f1_five

File Edit Query View Select Interval Boundary Tier Spectrum Pitch Intensity Formant Pulses Help

472.183040

0.0007736

5000 Hz

0 Hz

1	e	n	n	aa	ll	ee	xx	h	aa	nn	c0	aa	k0	ii	ll	ll	ii	aa	n	yv	nn	tt	e	n	xx	nn	cc	vv	c0	vv	l	aa	n	vv	mm	aa
2	엔나레는							한			주악			일리				앙년				때는					쩌		저랑					엄마		
3	eenn-nnaa-llee-nnxxnn							hhaann			c0uu-aak0			iill-llii				aang-nnyv				ttEE-nnxxnn					ccvv		c0vv-llaang					vvmm-mmaa		
4	엔나레는 한 주악 일리 앙년 때는 쩌 저랑 엄마																																			
5	옛날에는							한			중학			일이				학년				때는					쩌		저랑					엄마		
6	yes0-nnaall-ee-nnxxnn							hhaann			c0uung-hhaak0			iill-ii				hhaak0-nn				ttEE-nnxxnn					ccvv		c0vv-llaang					vvmm-mmaa		
7	옛날에는 한 중학 일이 학년 때는 쩌 저랑 엄마																																			

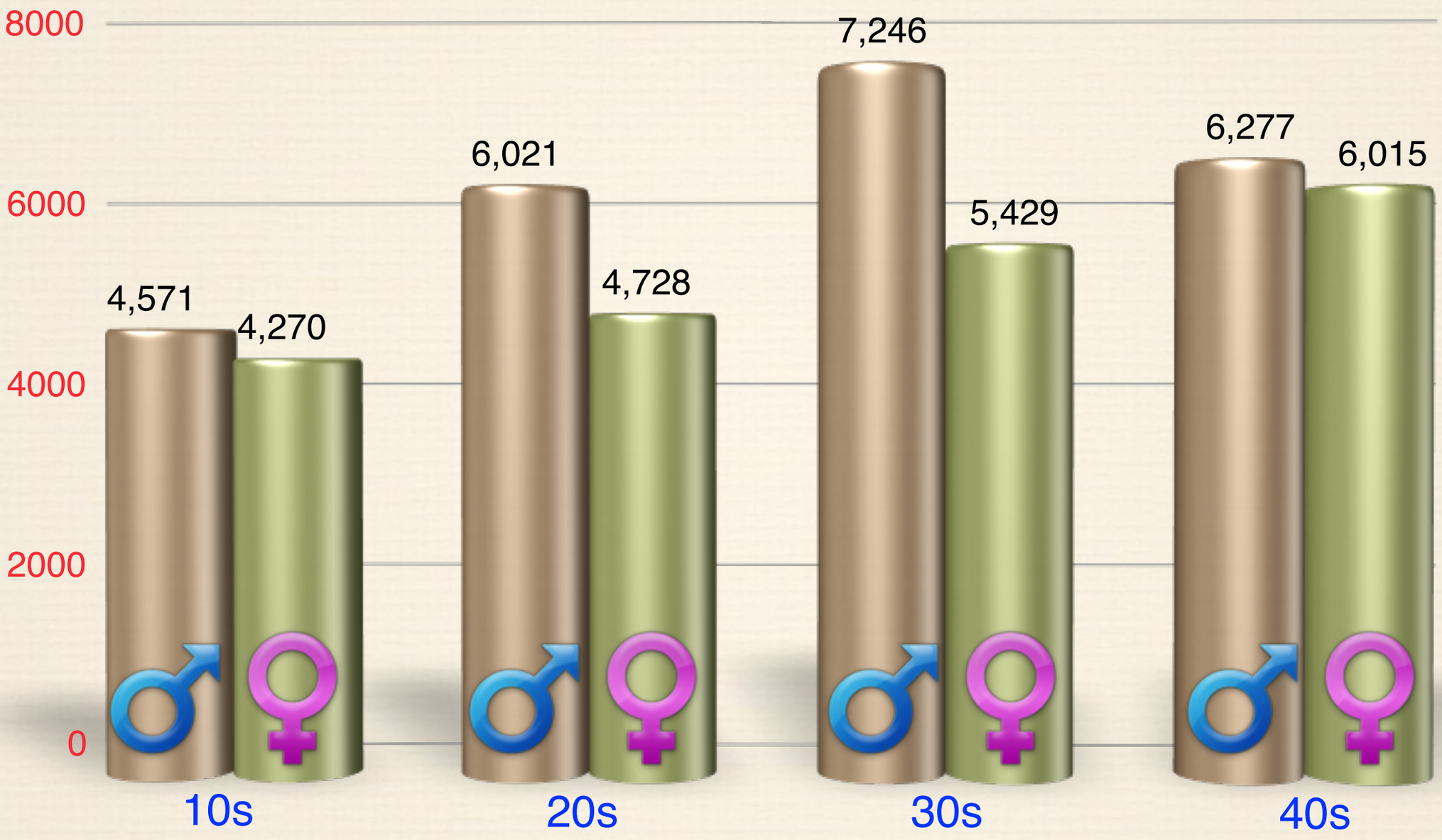
phoneme (2282/3174)
pWord.prono. (852)
pWord.prono. (852)
utt.prono. (389)
pWord.ortho. (852)
pWord.ortho. (852)
utt.ortho. (389)

472.183040 — 472.183040 — Visible part 3.762083 seconds — 475.945123 — 137.814832

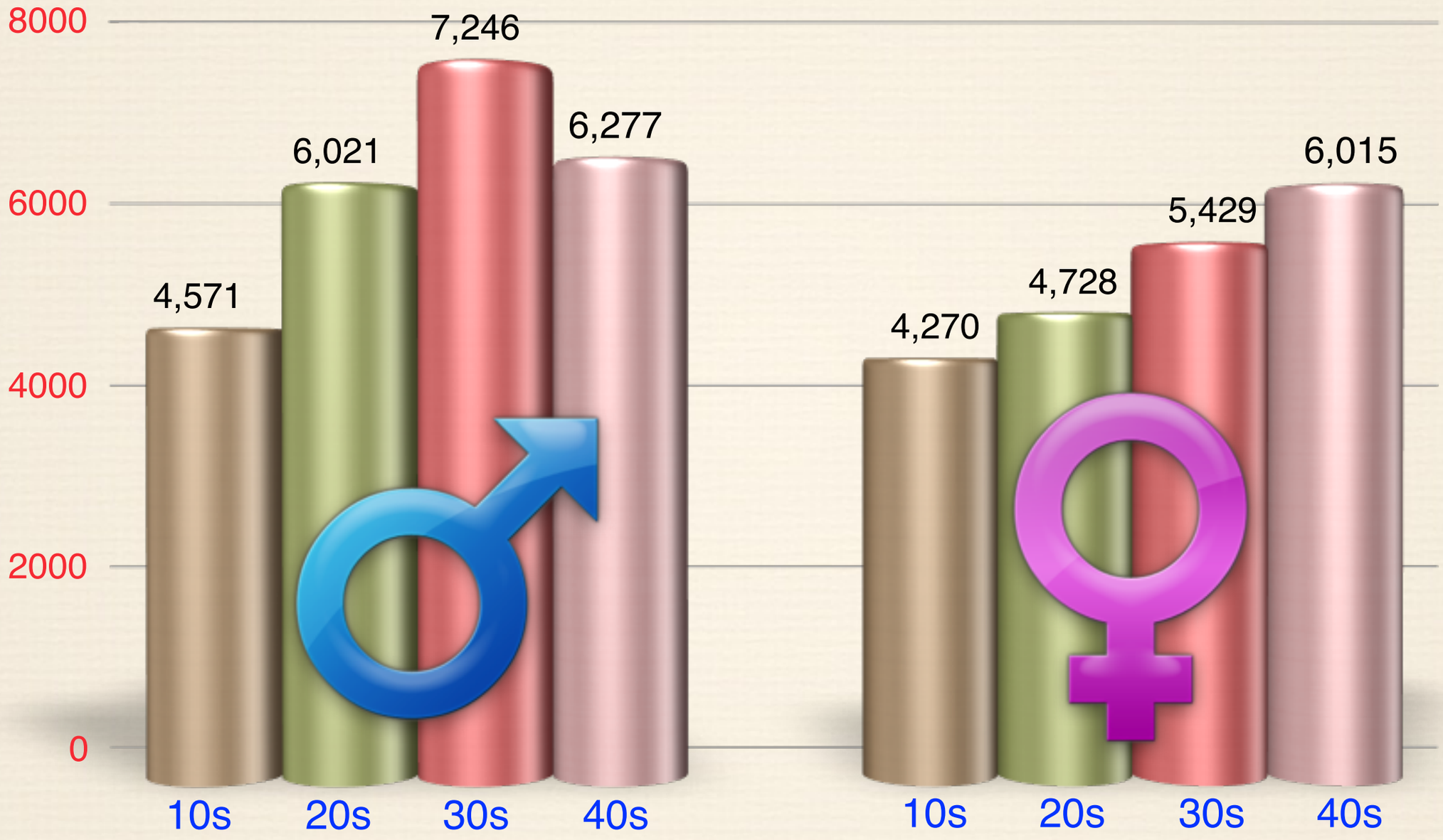
Total duration 613.759955 seconds

all in out sel bak Group

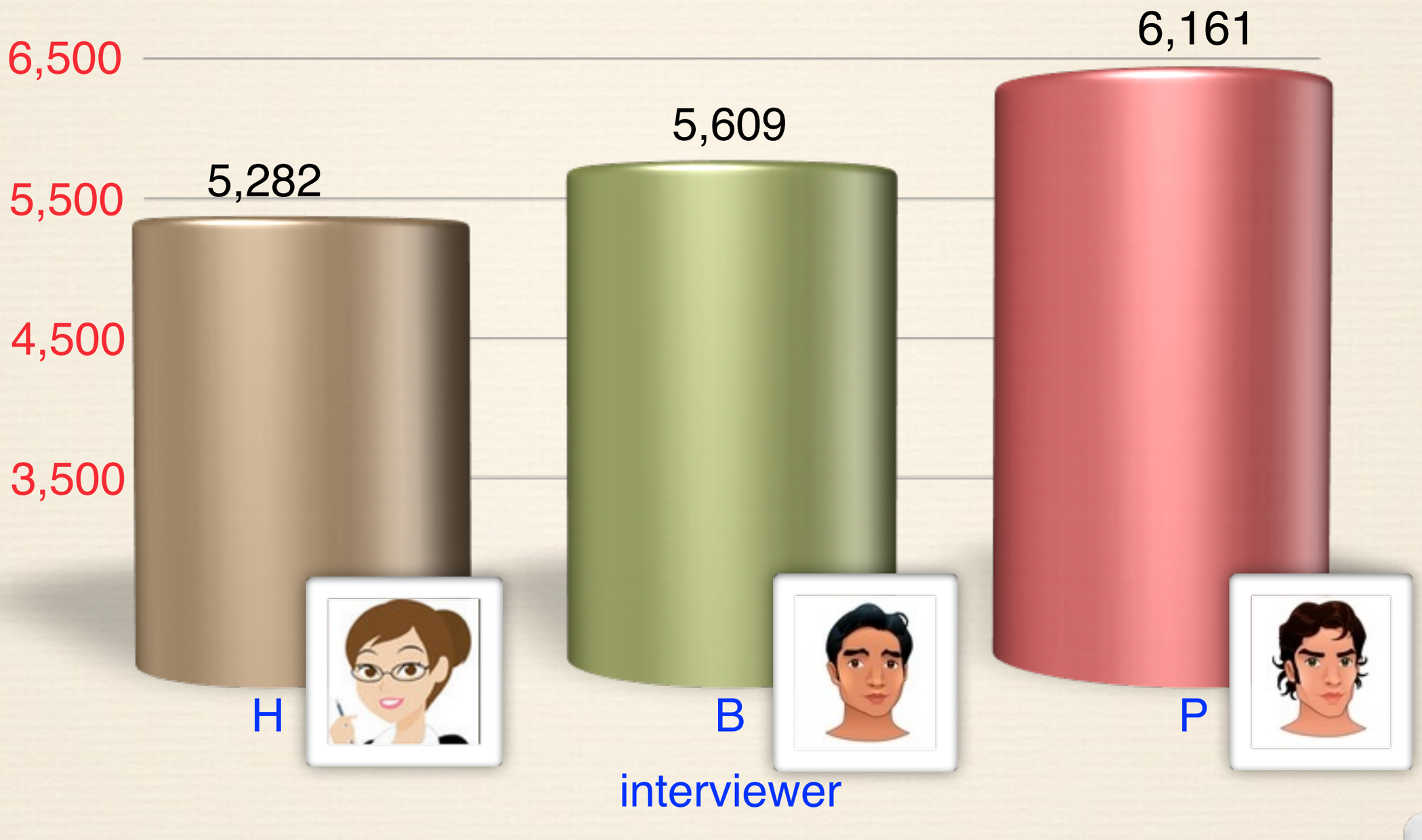
Mean number of pWords by age & gender



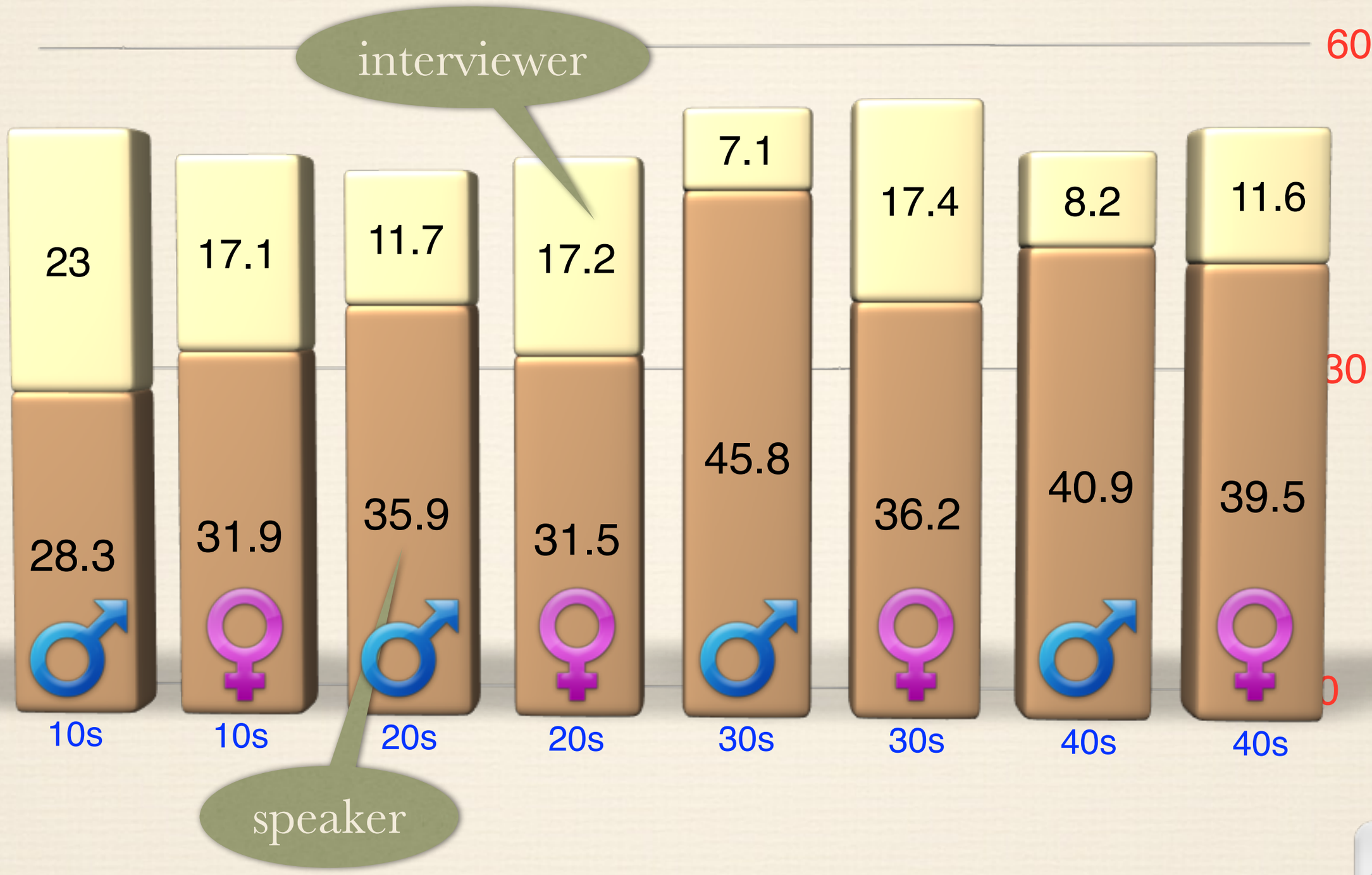
Mean number of pWords by age & gender



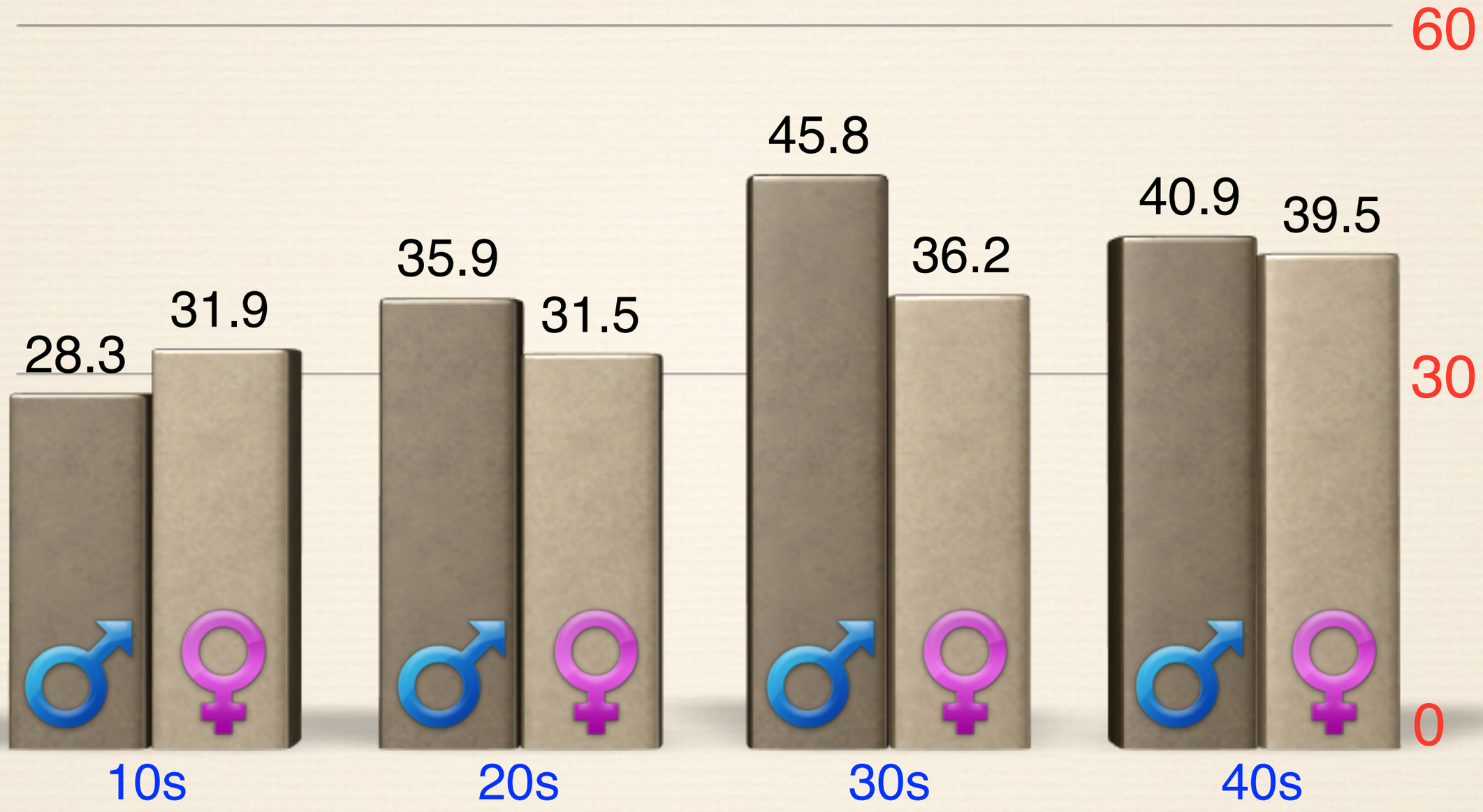
Mean number of pWords by interviewer



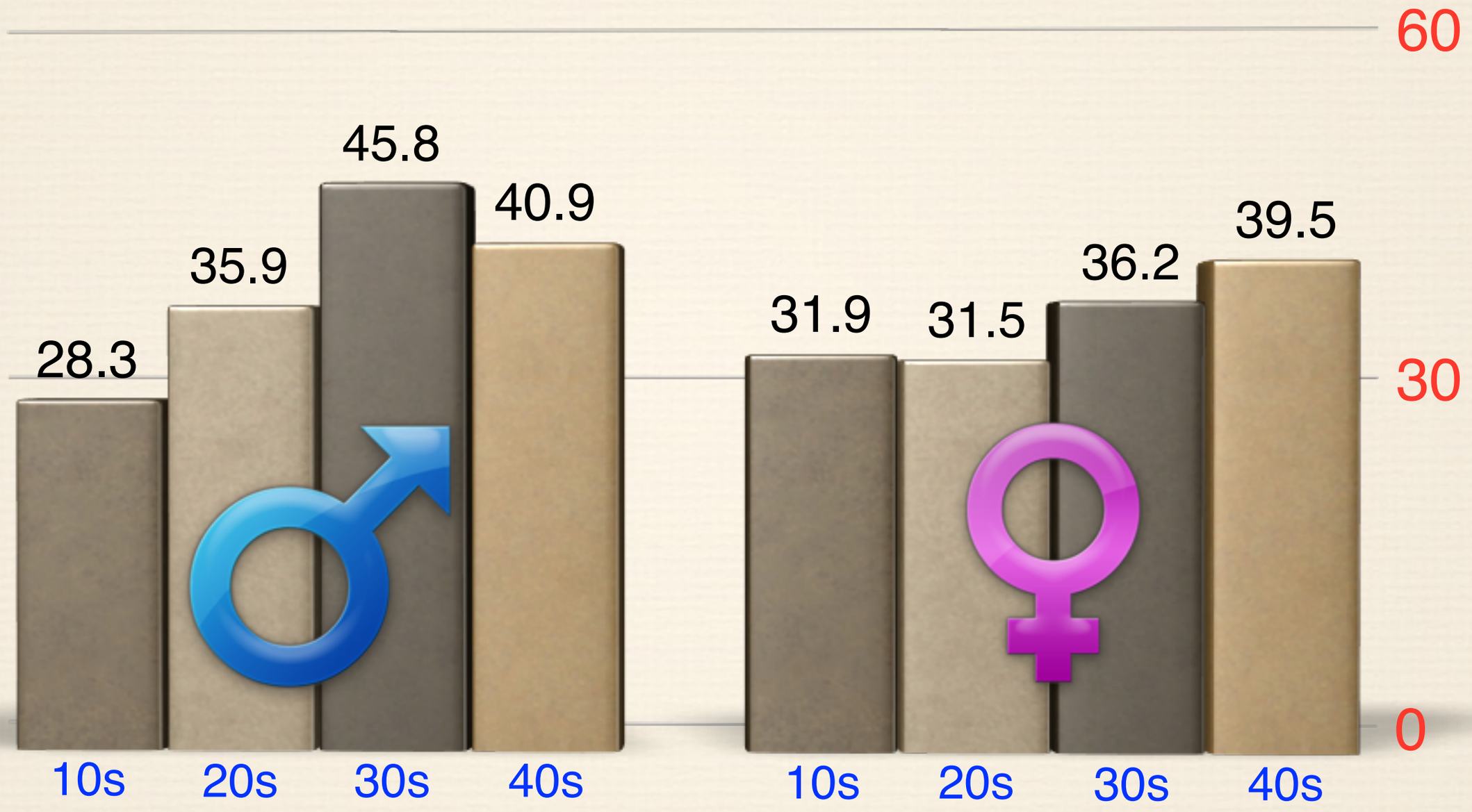
Mean minutes of speaker vs. interviewer



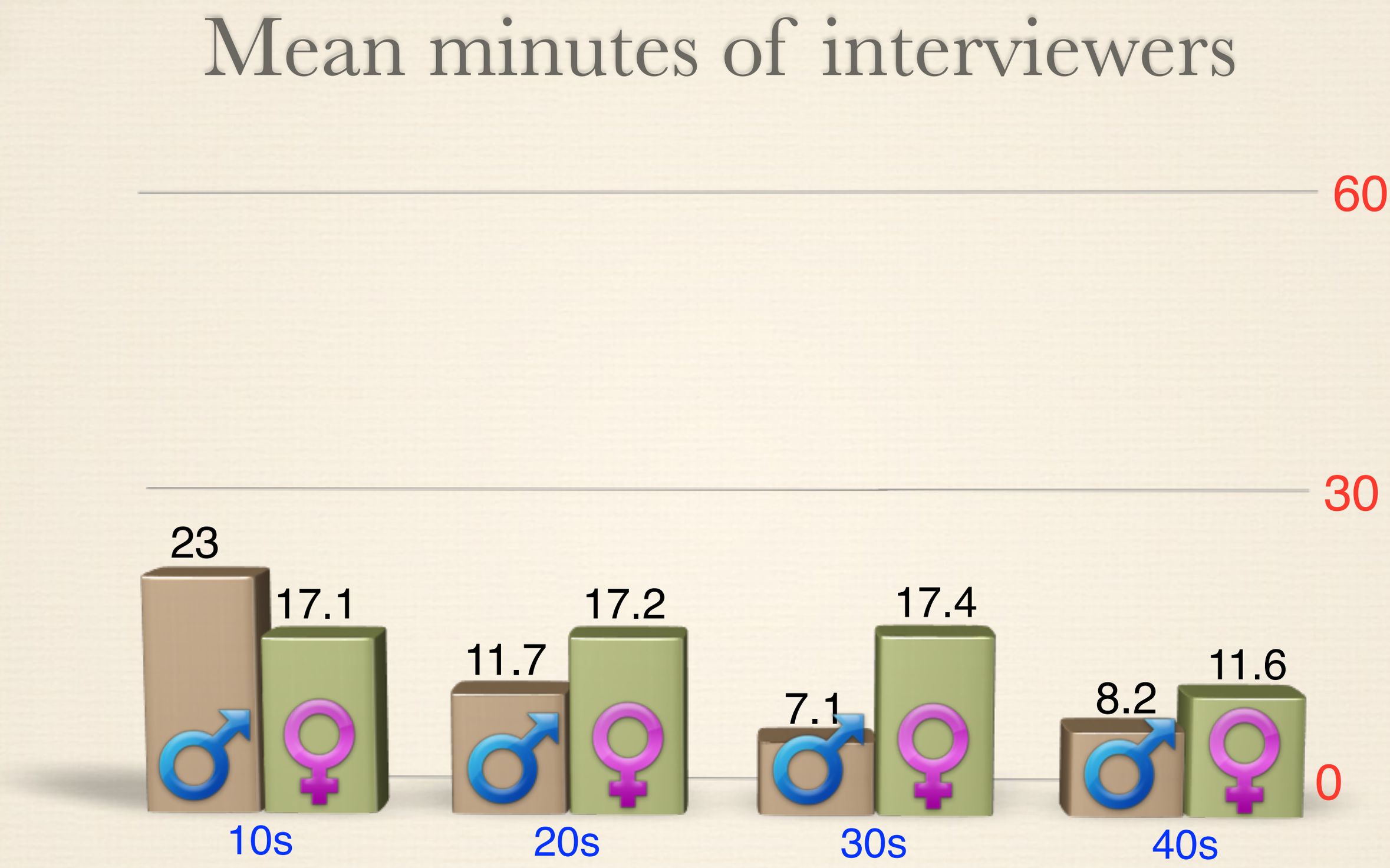
Mean minutes of speakers



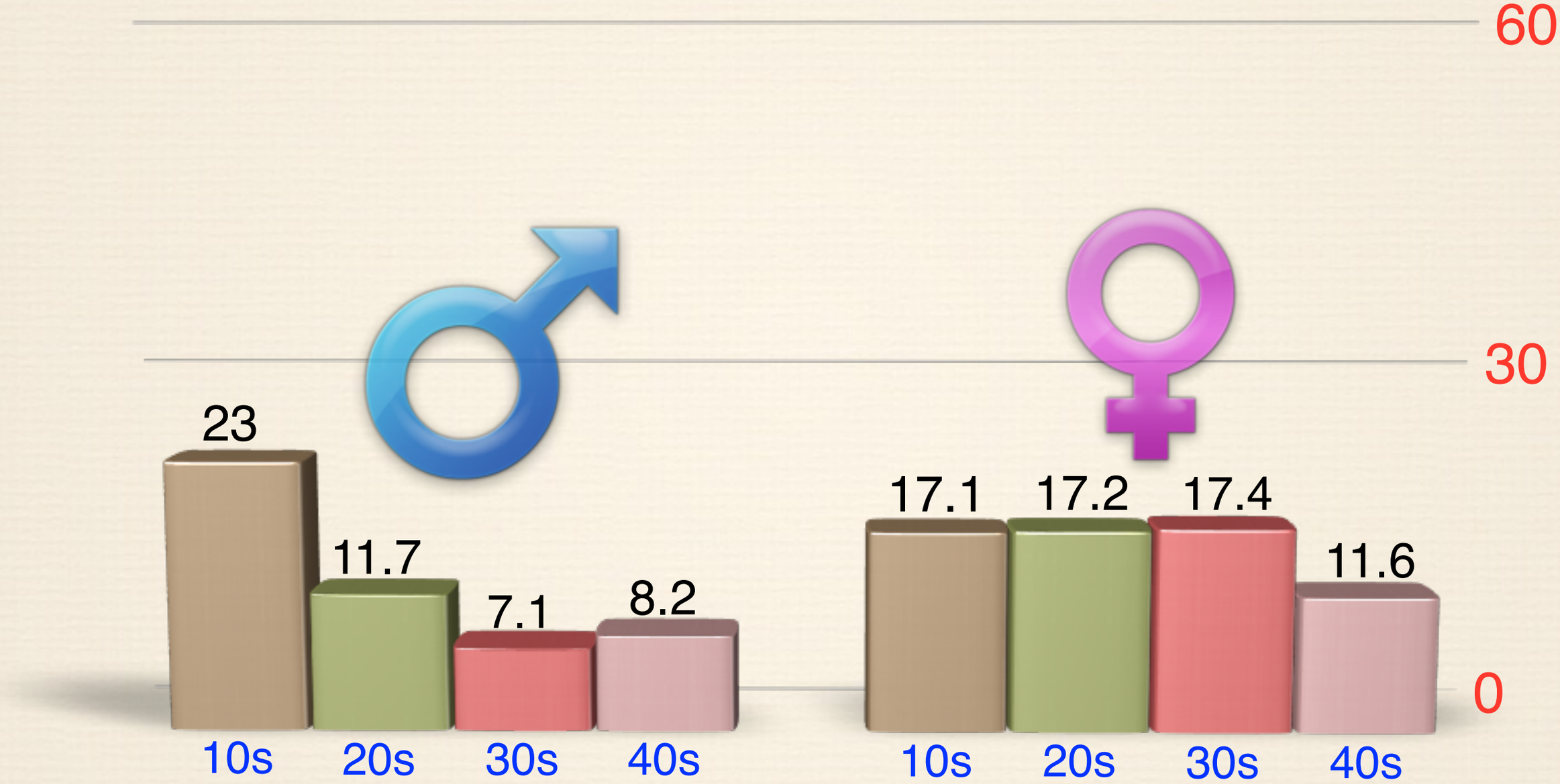
Mean minutes of speakers



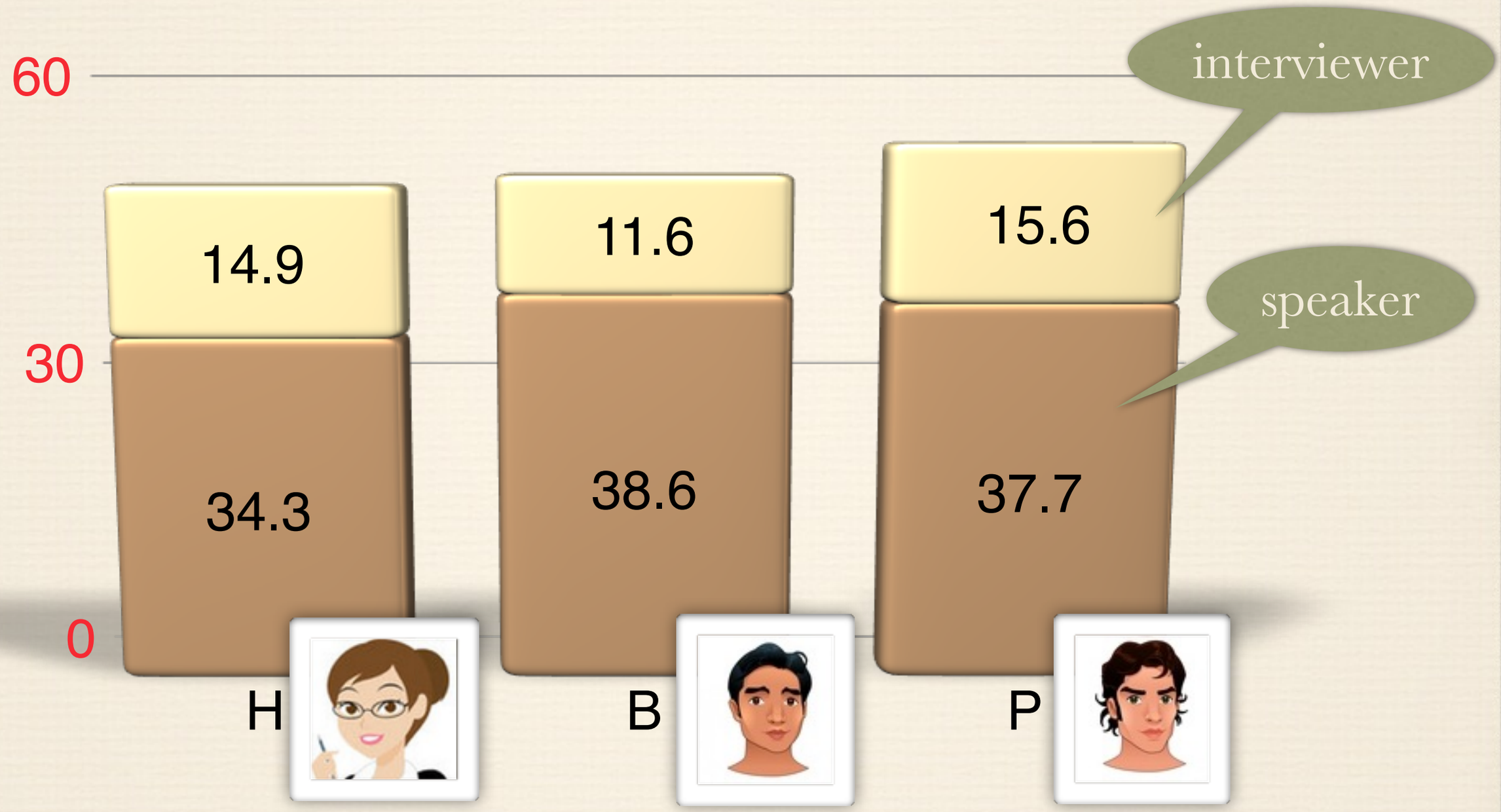
Mean minutes of interviewers



Mean minutes of interviewers



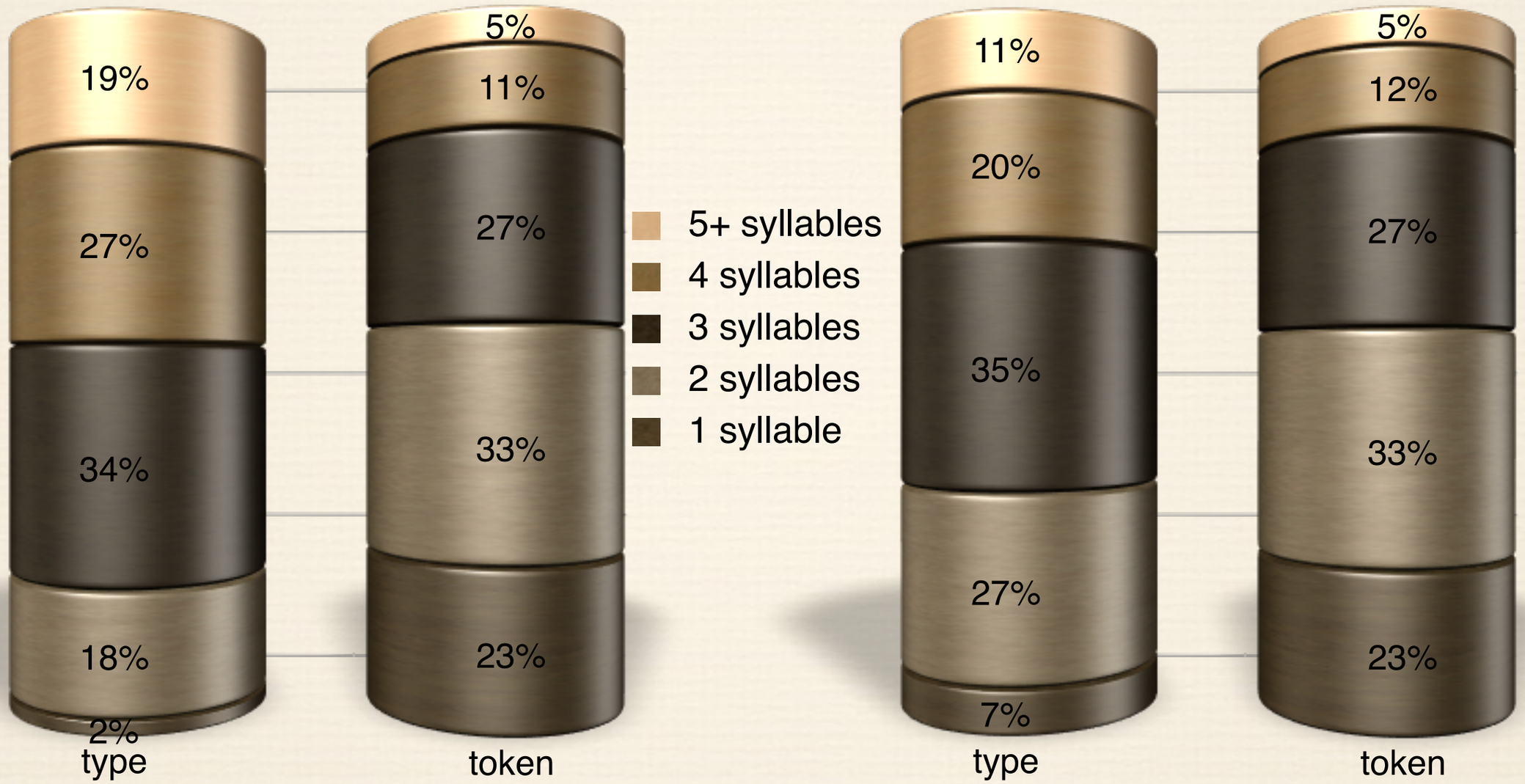
Mean minutes of speaker vs. interviewer



pWord frequency by number of syllables

orthographic forms

❖ 231,632 pWord tokens

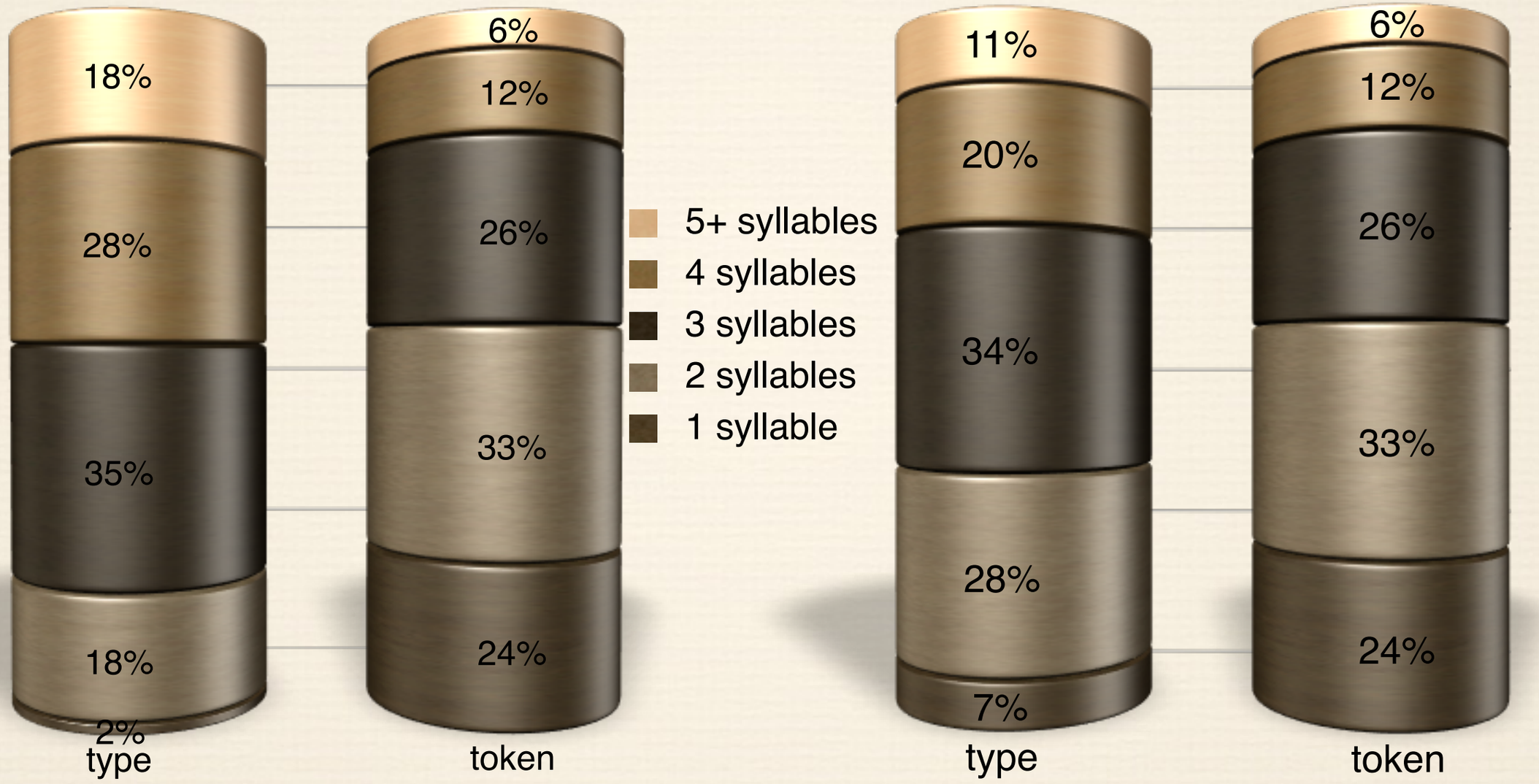


All Talkers

Average Talkers

pWord frequency by number of syllables

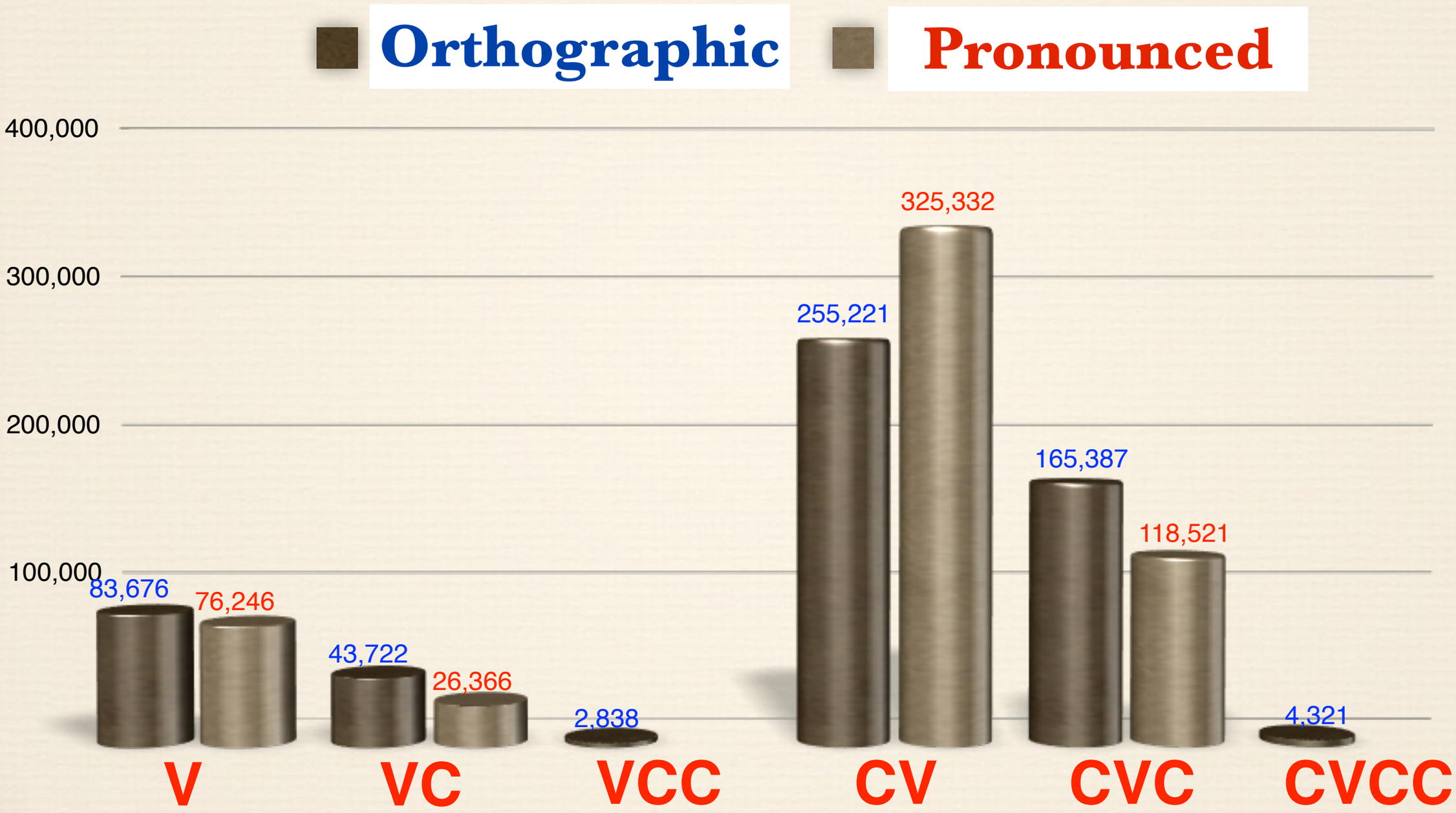
pronounced forms



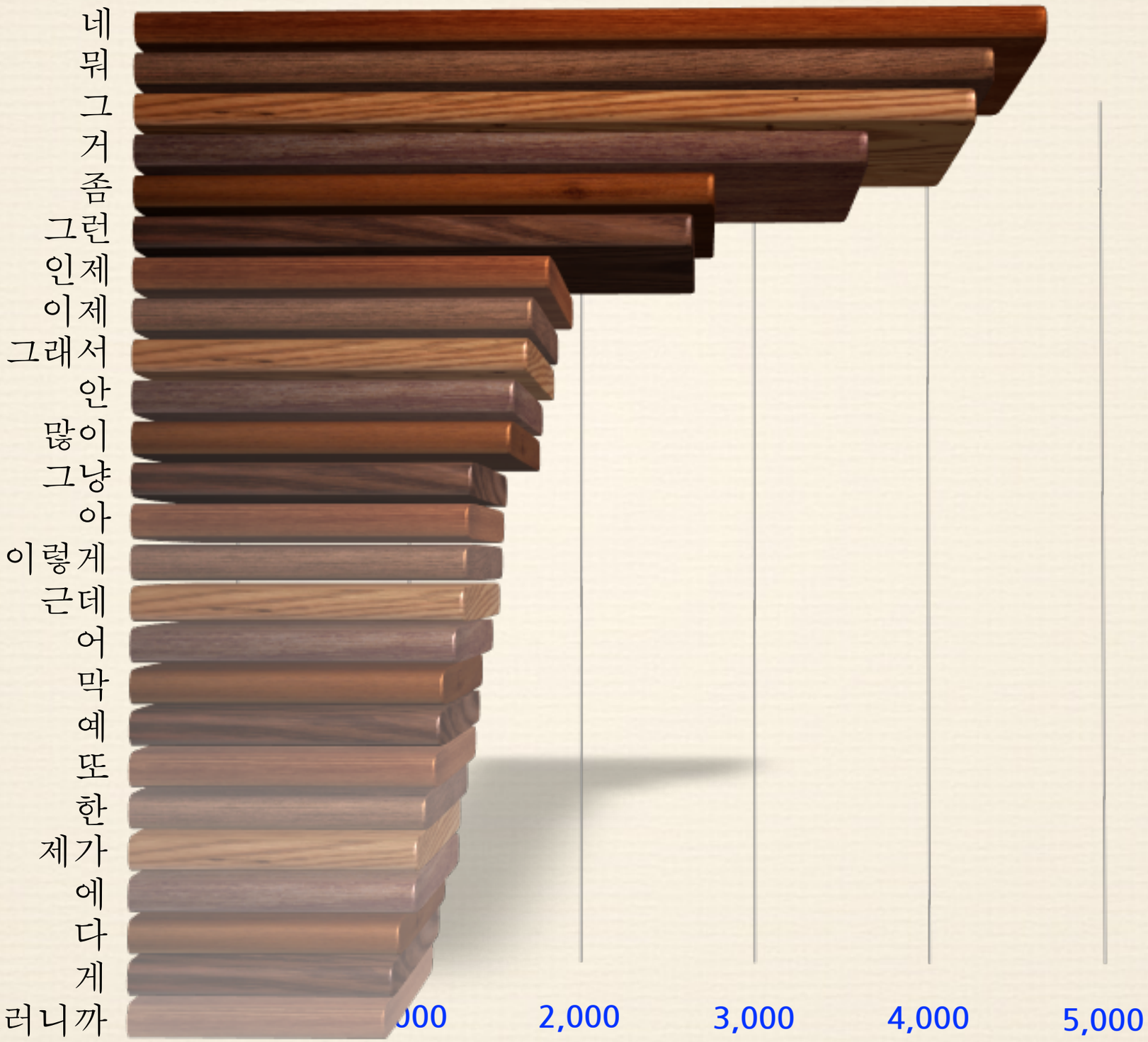
All Talkers

Average Talkers

Types of orthographic vs. pronounced syllables



pWord frequency overall



pWord frequency by age/gender

네 뭐 거 그 아 안 좀 냥 런 데 막 한 어 음 이 서 잘 다 가 이 까 게 는 게 때 그 그 근 많 그래 제 니 러 저 되

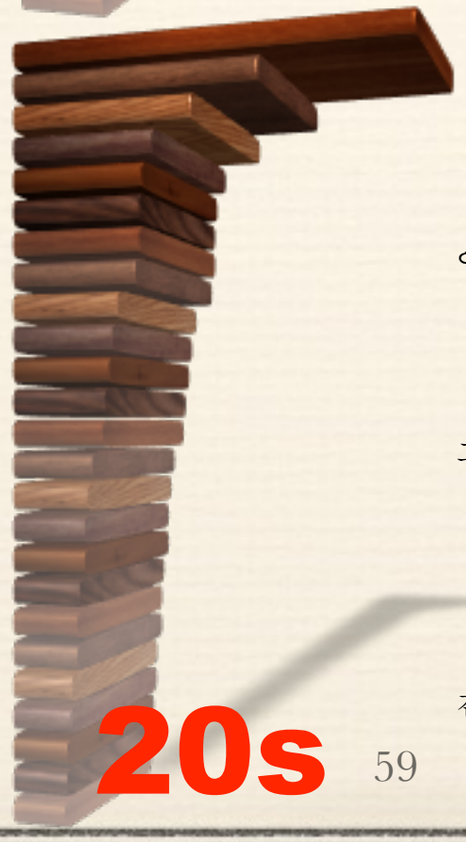
10s

네 제 거 런 그 좀 뭐 막 서 게 이 냥 게 안 아 어 런 는 게 가 간 요 더 다 잘 이 그 래 되 많 그 렇 이 저 제 약 아 같

10s

그 뭐 제 거 가 예 좀 이 런 제 네 또 게 에 아 서 막 고 때 다 한 안 는 까 금 인 제 많 그 이 그래 가 지 저 니 지

20s



네 거 그 게 런 서 좀 이 막 뭐 안 냥 어 제 다 데 아 는 게 음 런 한 가 요 잘 되 그래 많 그 이 근 저 이 제 애 같

20s

뭐 그 좀 런 거 제 게 제 예 한 네 냥 서 이 데 까 또 안 게 다 가 금 어 게 런 이 렇 이 그 래 많 근 니 러 지 이 제 렇 이

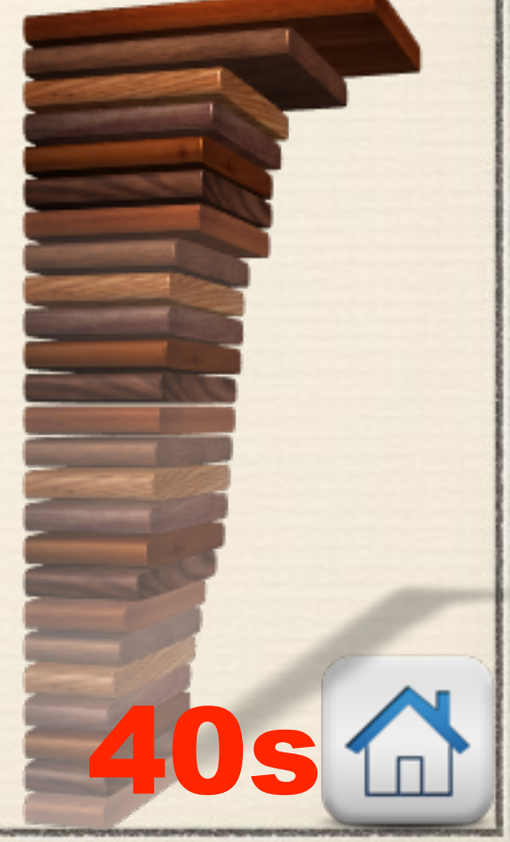
30s

네 거 뭐 런 그 에 좀 게 이 아 데 또 음 서 냥 안 제 런 제 가 다 요 어 는 한 이 렇 많 근 그 래 그 인 이 이 제 같 아 저

30s

그 뭐 제 거 어 서 좀 안 런 에 또 한 아 이 고 예 금 데 이 까 게 게 다 는 런 이 그 래 그 많 하 지 근 니 러 이 렇 이 있 이

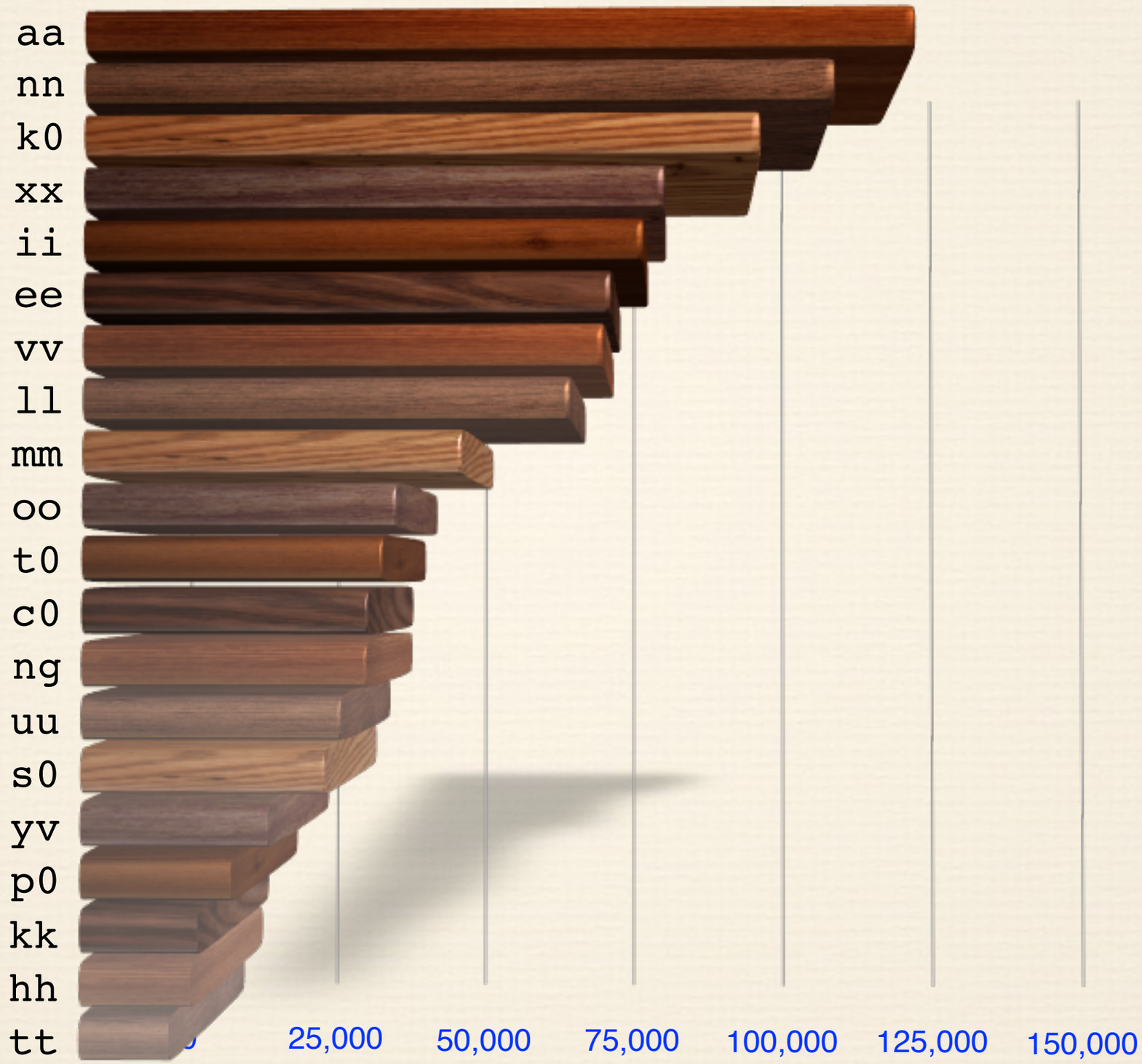
40s



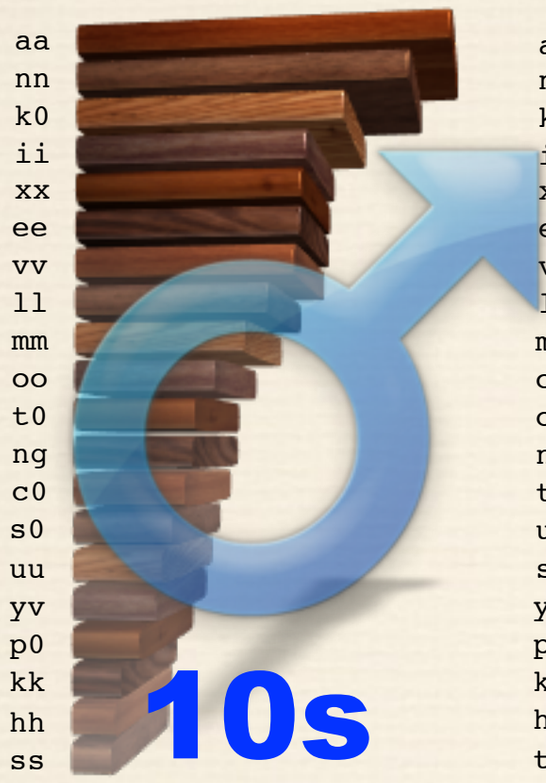
뭐 그 예 제 막 데 안 거 또 까 게 제 어 런 좀 에 네 냥 서 다 아 리 한 말 무 인 근 니 러 이 렇 이 그 래 우 정 너

40s

Individual phoneme frequency

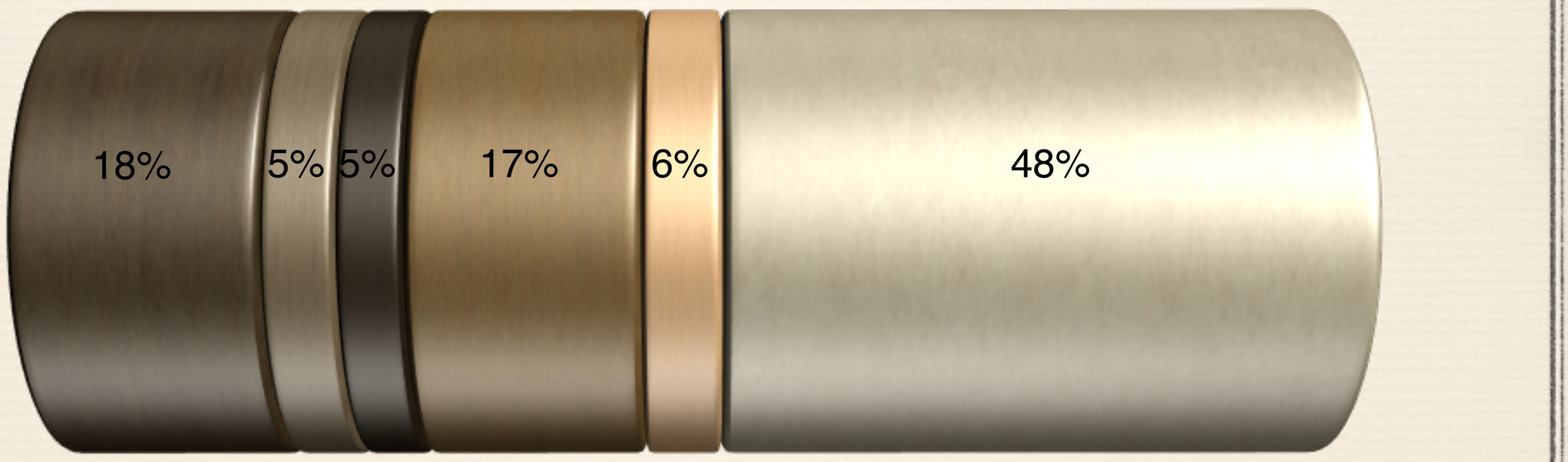


Phoneme frequency by age/gender



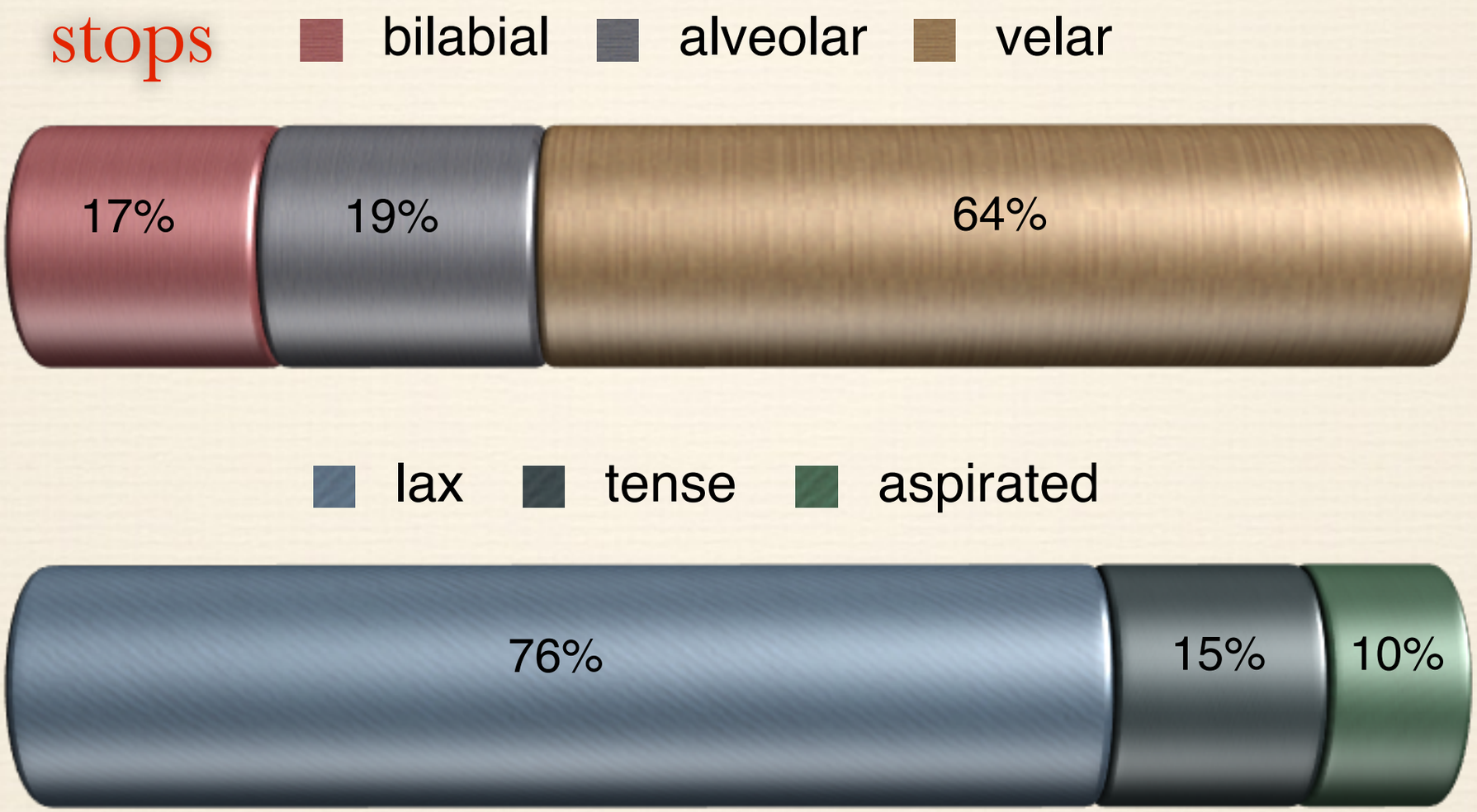
Phoneme frequency overall

- 1,134,781 phoneme tokens in the corpus



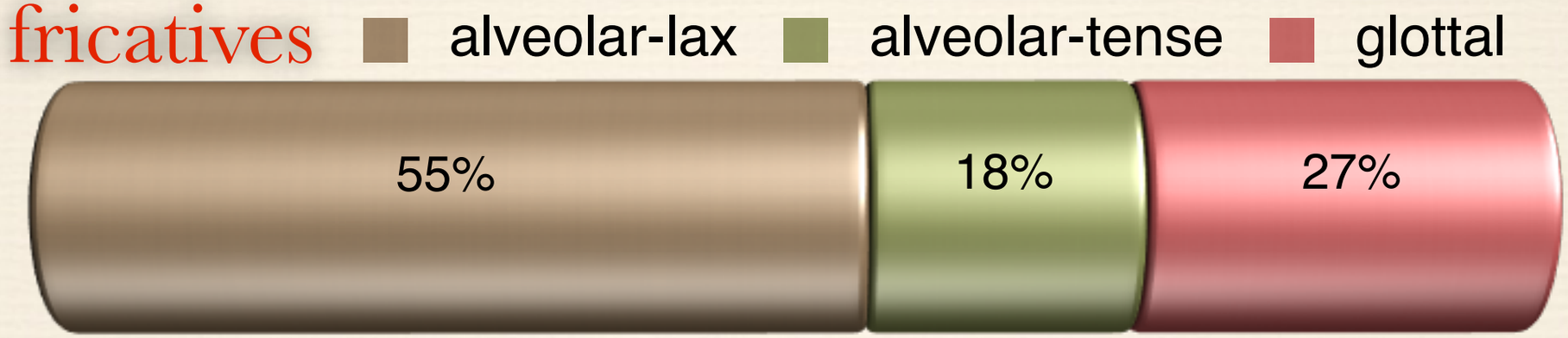
Phoneme frequency by categories

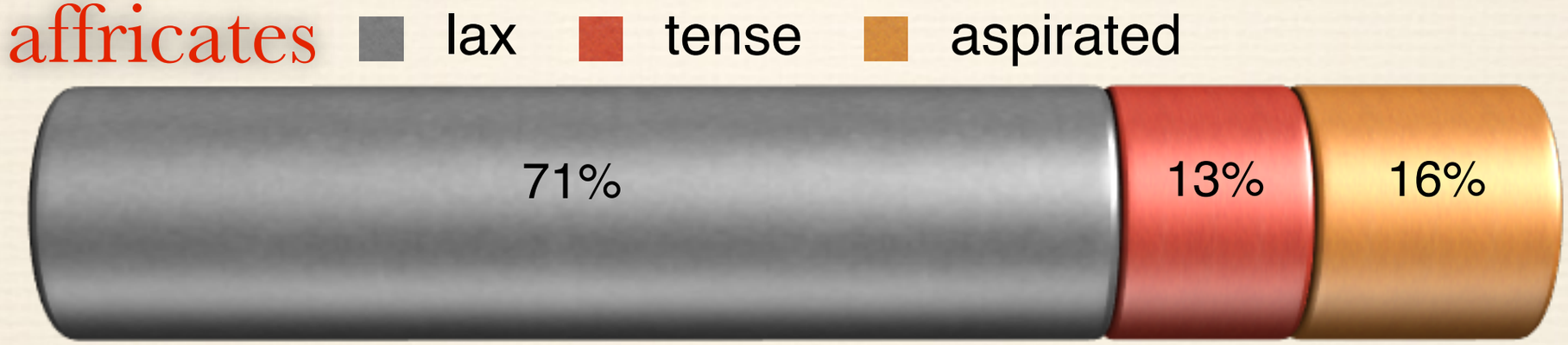
- stops by place & manner of articulation

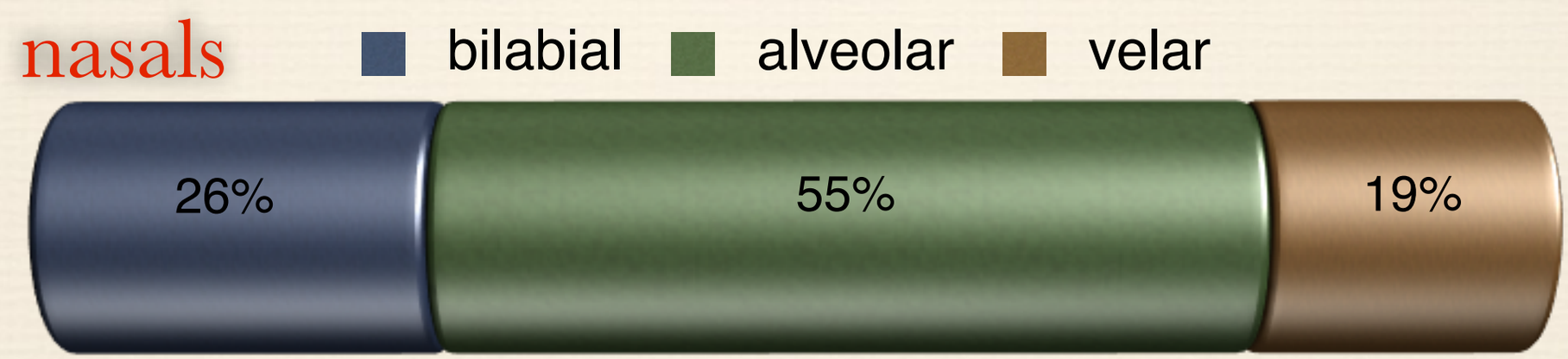


Phoneme frequency by categories

- fricatives, affricates & nasals

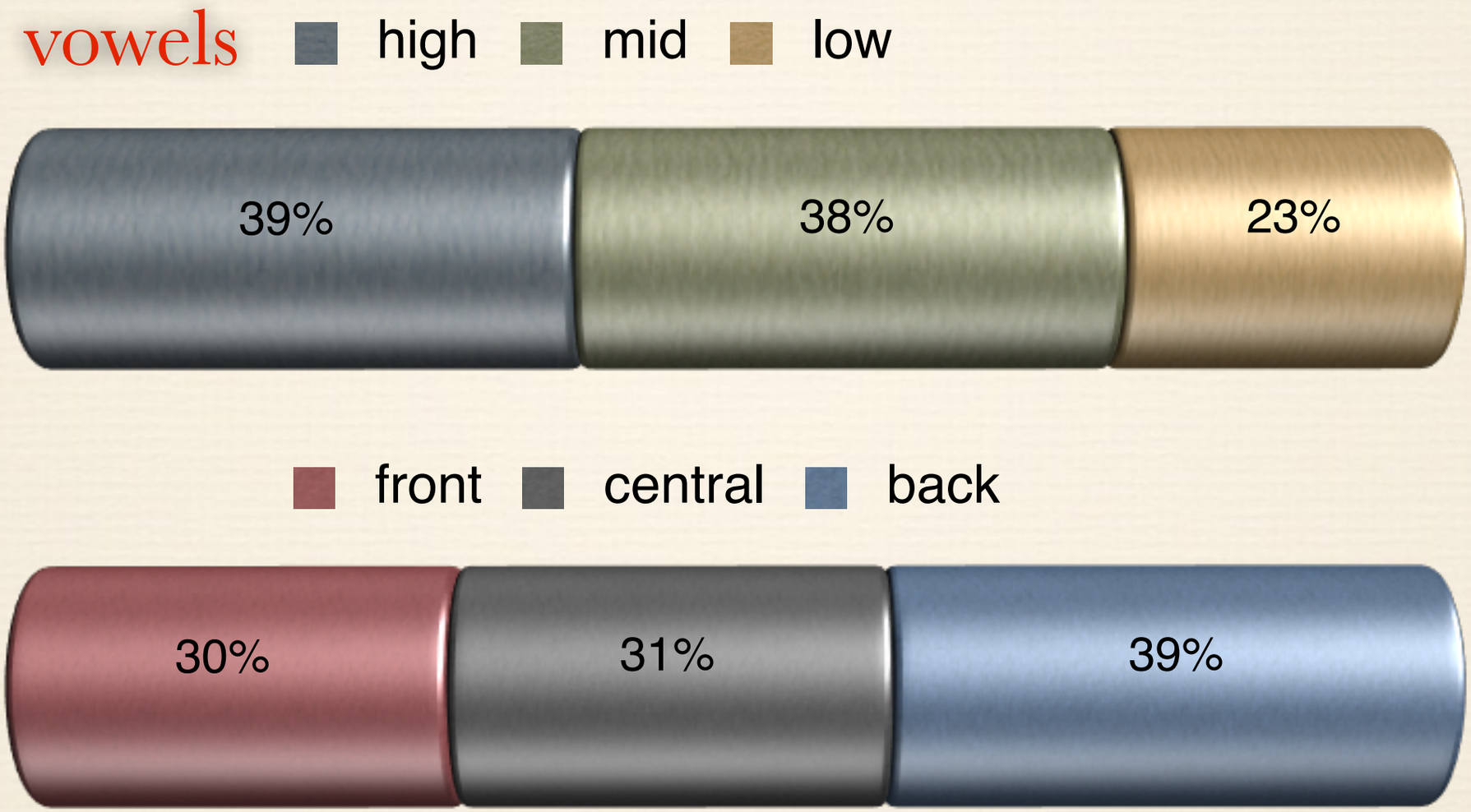






Phoneme frequency by categories

- vowels by tongue height and frontedness

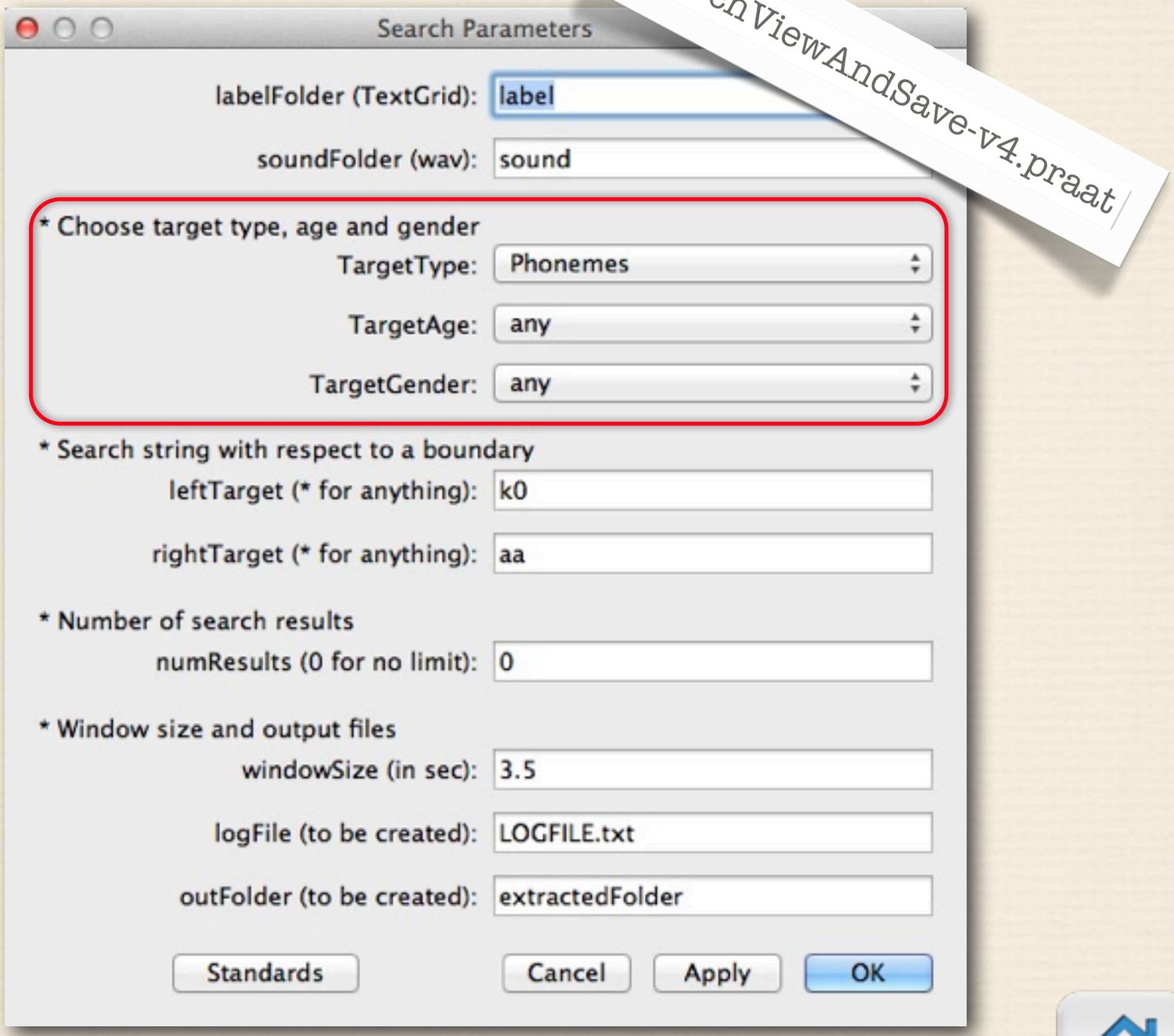


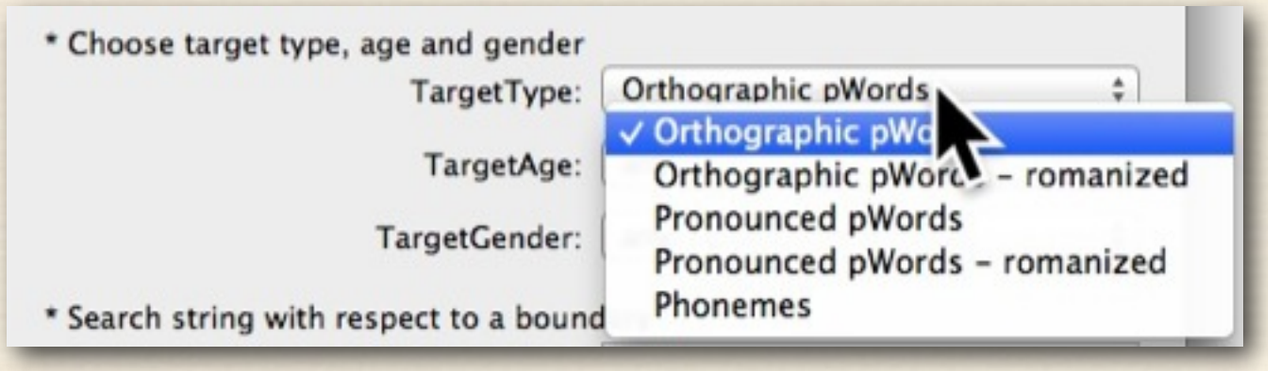
Corpus Tools

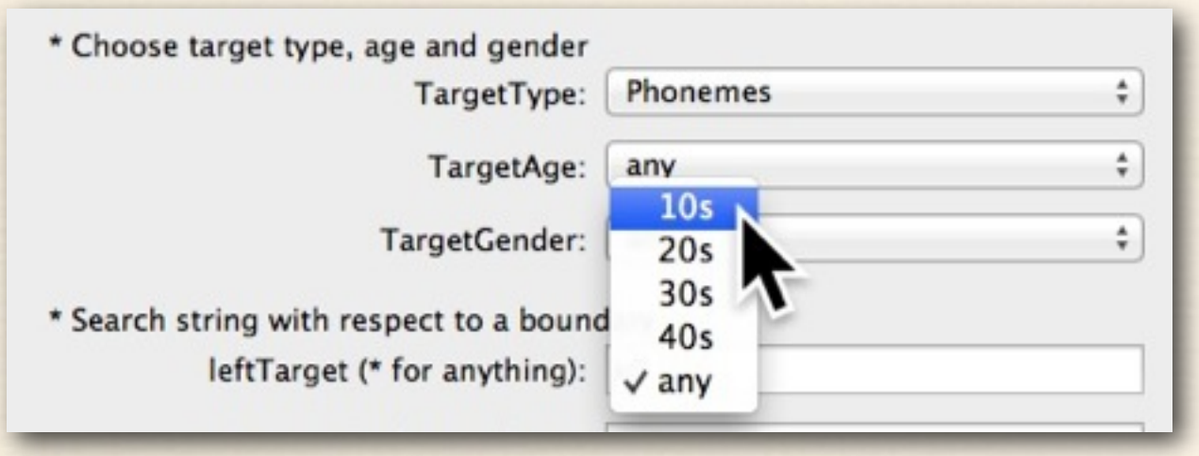
Search script in Praat

searchViewAndSave-v4.praat

- Search for phonemes and pWords in any age and gender







- Search for phonemes and pWords in any age and gender

Search Parameters

labelFolder (TextGrid): label

soundFolder (wav): sound

* Choose target type, age and gender

TargetType: Orthographic words

TargetAge: 30s

TargetGender: female

* Search string with respect to a boundary

leftTarget (* for anything): 그러니까

rightTarget (* for anything): *

* Number of search results

numResults (0 for no limit): 10

* Window size and output files

windowSize (in sec): 3.5

logFile (to be created): LOGFILE.txt

outFolder (to be created): extractedFolder

Standards Cancel Apply OK

Search Parameters

abelFolder (TextGrid): label

soundFolder (wav): sound

type, age and gender

TargetType: Pronounced words

TargetAge: 20s

TargetGender: male

with respect to a boundary

arget (* for anything): *

arget (* for anything): 고드아꼬

rch results

esults (0 for no limit): 15

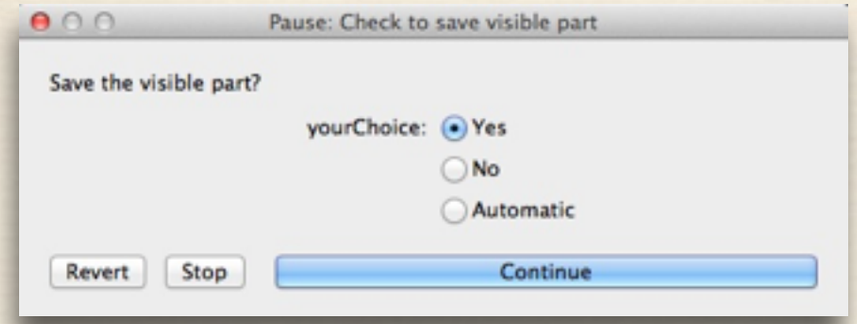
nd output files

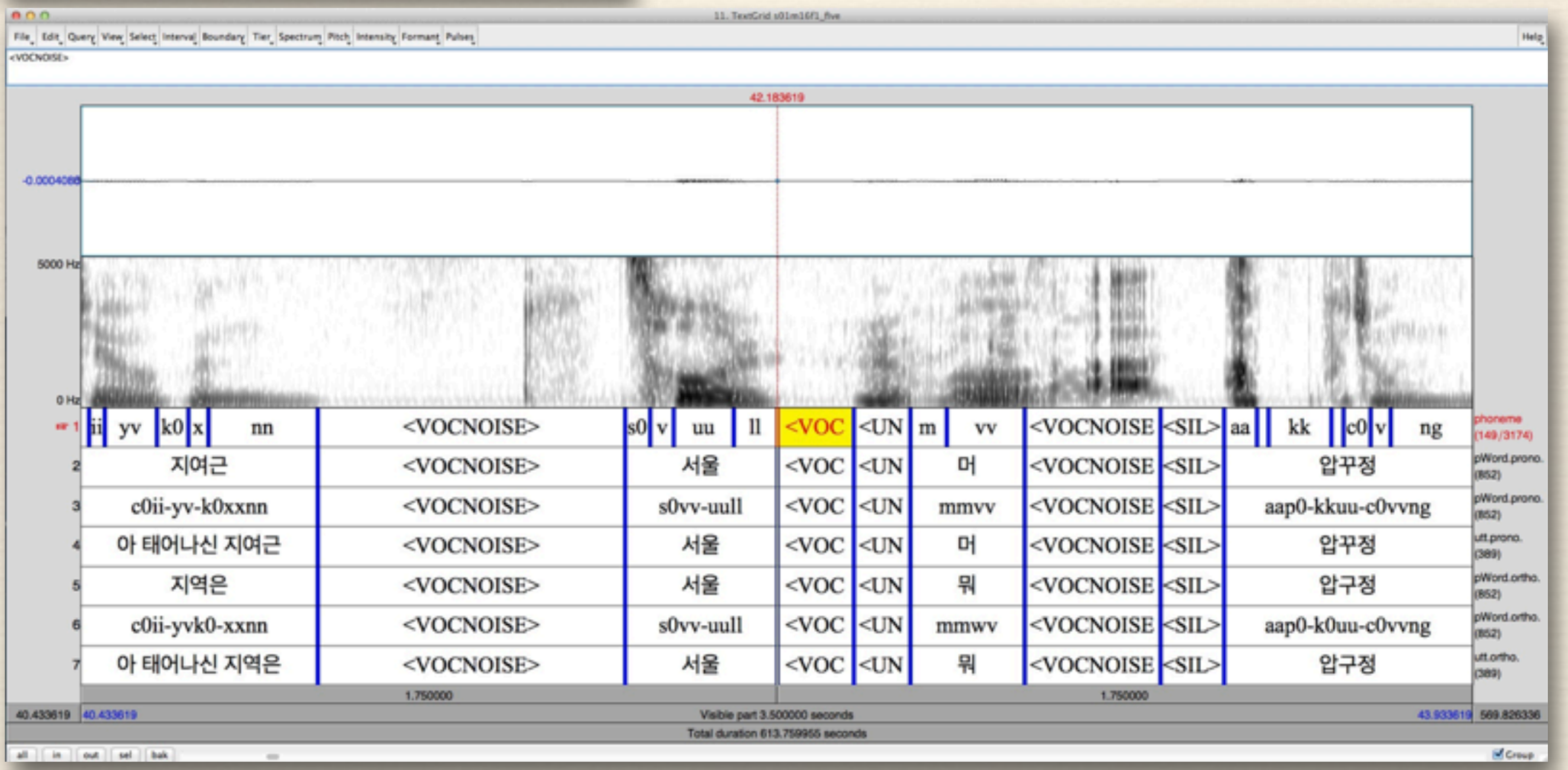
windowSize (in sec): 3.5

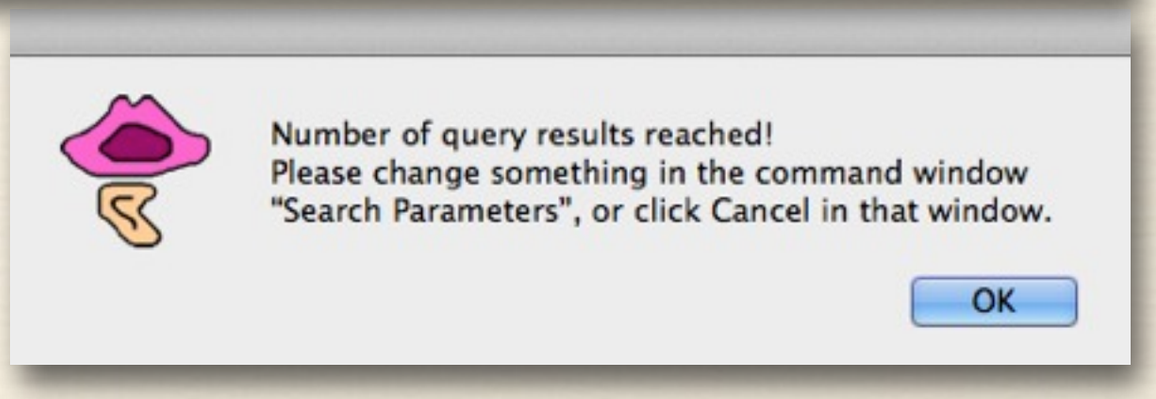
gFile (to be created): LOGFILE.txt

outFolder (to be created): extractedFolder

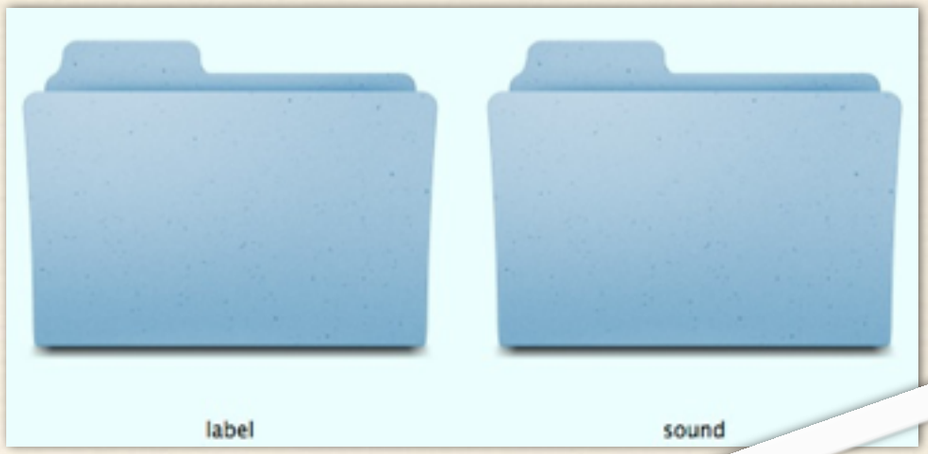
Standards Cancel Apply OK


Pause: Check to save visible part
Save the visible part?
yourChoice:
Yes
No
Automatic
Revert
Stop
Continue


11. TextGrid s01m16f1_five
File Edit Query View Select Interval Boundary Tier Spectrum Pitch Intensity Formant Pulses Help
<VOCNOISE>
42.183619
-0.0004088
5000 Hz
0 Hz
1 ii yv k0 x nn <VOCNOISE> s0 v uu ll <VOC <UN m vv <VOCNOISE <SIL> aa kk c0 v ng phoneme (149/3174)
2 지여근 <VOCNOISE> 서울 <VOC <UN 머 <VOCNOISE <SIL> 압꾸정 pWord.prono. (852)
3 c0ii-yv-k0xxnn <VOCNOISE> s0vv-uull <VOC <UN mmvv <VOCNOISE <SIL> aap0-kkuu-c0vvng pWord.prono. (852)
4 아 태어나신 지여근 <VOCNOISE> 서울 <VOC <UN 머 <VOCNOISE <SIL> 압꾸정 utt.prono. (389)
5 지역은 <VOCNOISE> 서울 <VOC <UN 뭐 <VOCNOISE <SIL> 압구정 pWord.ortho. (852)
6 c0ii-yvk0-xxnn <VOCNOISE> s0vv-uull <VOC <UN mmwv <VOCNOISE <SIL> aap0-k0uu-c0vvng pWord.ortho. (852)
7 아 태어나신 지역은 <VOCNOISE> 서울 <VOC <UN 뭐 <VOCNOISE <SIL> 압구정 utt.ortho. (389)
1.750000
1.750000
40.433619
40.433619
Visible part 3.500000 seconds
43.933619
569.826336
Total duration 613.759955 seconds
all in out sel bak
Group

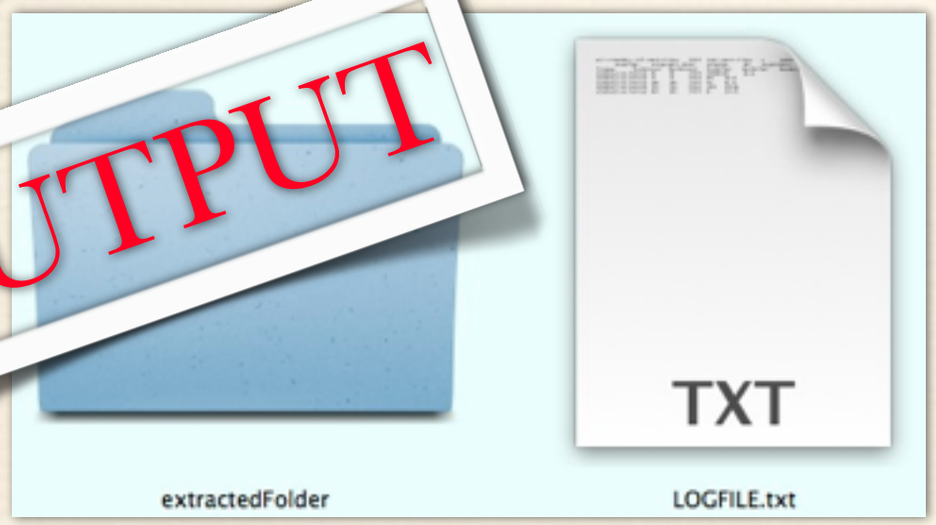

Number of query results reached!
Please change something in the command window
"Search Parameters", or click Cancel in that window.
OK

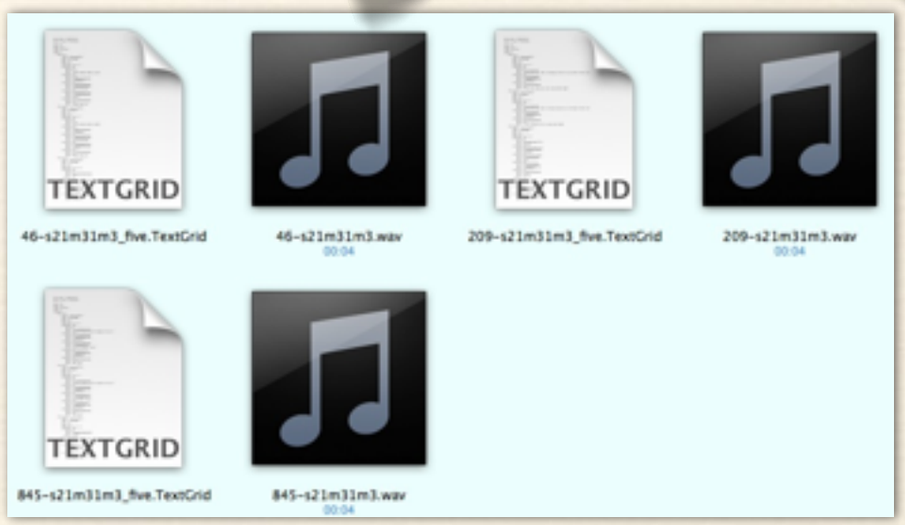
searchViewAndSave-v2.praat



INPUT

OUTPUT






```
wrt/ a boundary, left search string:    그러니까  right search string:    *       number of query results:       5
        targetType:     orthographic iWords     targetAge:      30s     targetGender:   male
filename        PrevInterval    NextInterval    PrevIntText     NextIntText     BoundaryTime
s21m31m3_five.TextGrid  45      46      그러니까  <VOCNOISE>      20.51
s21m31m3_five.TextGrid  71      72      그러니까  전에      28.51
s21m31m3_five.TextGrid  208     209     그러니까  먹는      83.31
s21m31m3_five.TextGrid  725     726     그러니까  기차역    270.20
s21m31m3_five.TextGrid  844     845     그러니까  잘        314.72
```

Corpus Availability

Since March 2015

Available for free

- Only to research community from March 2015
- For commercial use, please contact us

whyun@kmu.ac.kr
kyoon@ynu.ac.kr

Thank you