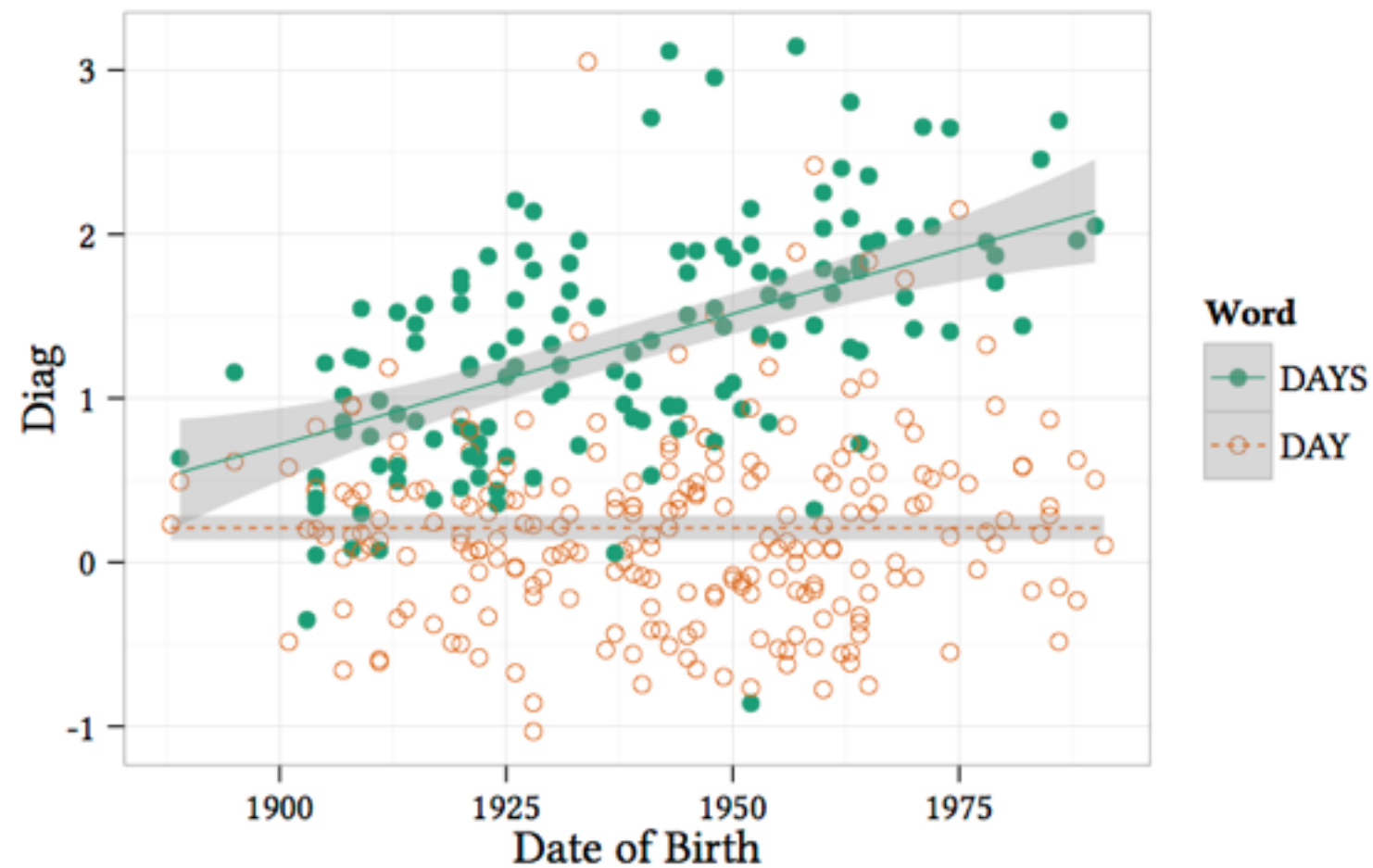


Corpus sociophonetics at home... without the mess!

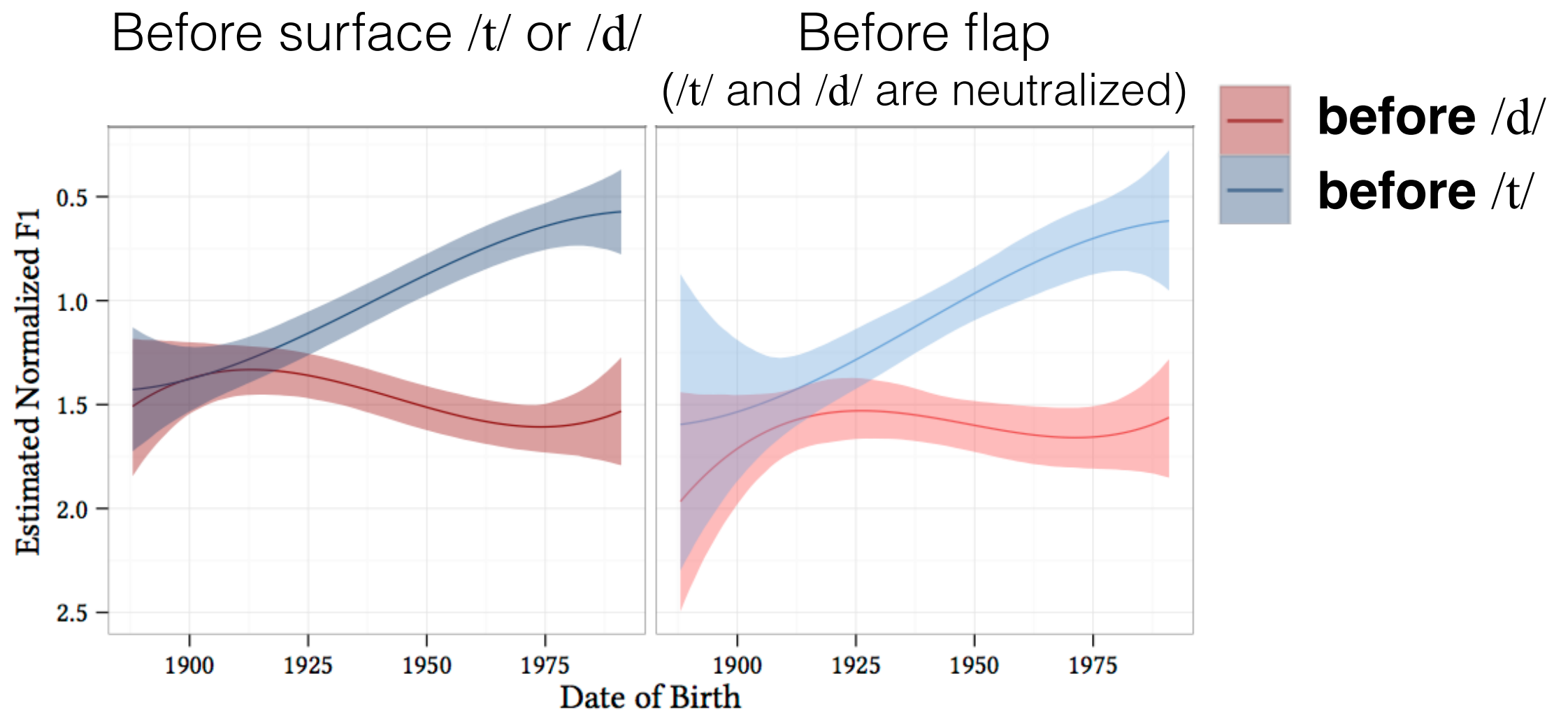
Ewan Dunbar
April 9, 2015

Phonetic variation and change



/ey/ raising in Philadelphia English (apparent time)

Phonetic variation and change



Incipient /ay/ raising in Philadelphia English (apparent time)

Corpus sociophonetics

The majority of the work on phonetic variation and change (ever) has been done without large corpora

Corpus sociophonetics

The majority of the work on phonetic variation and change (ever) has been done without large corpora

This sucks

Corpus sociophonetics

The majority of the work on phonetic variation and change (ever) has been done without large corpora

This sucks

But so do corpora

Corpus sociophonetics: the state of the art

Old school sociophonetics

Step 1. Administer a “sociolinguistic interview”

TABLE 2. *Sample of 35 participants, by age and sex*

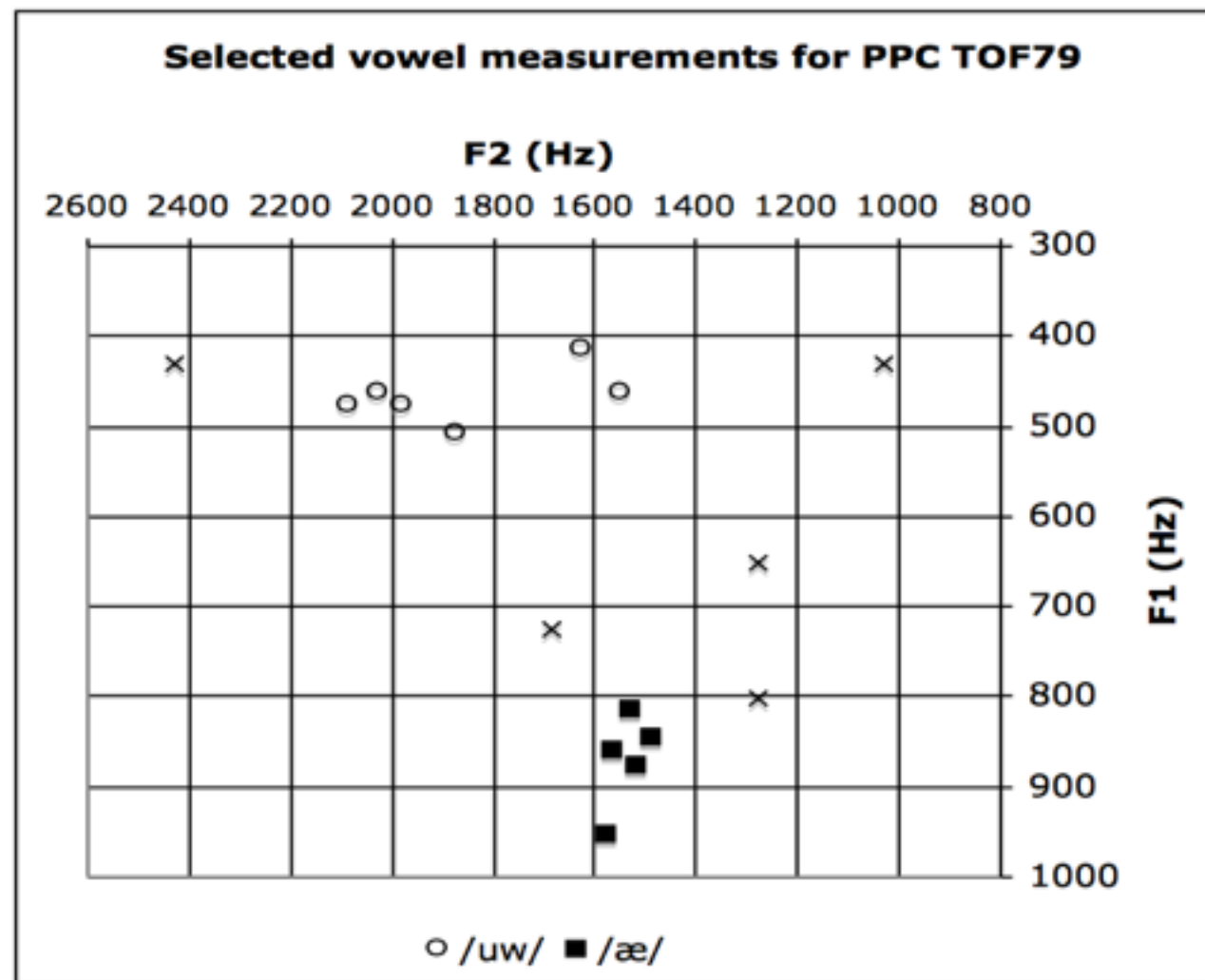
Sex	Birth year		
	<1946	1946–1965	>1965
Female	8	6	7
Male	5	5	4
Total	13	11	11

**PLEASE READ THE FOLLOWING WORDS AS YOU WOULD
NORMALLY SAY THEM**

bar	cow	tooth	girl	sad
sit	pair	seed	did	sore
file	proud	stayed	bang	cook
student	tide	carry	new	do
collar	star	step	state	spirit
stare	sod	mafia	code	tight
pajamas	Pakistani	boots	dull	turn
lasaona	laøer	writer	cool	ferrv

Old school sociophonetics

Step 2. Take some measure of interest



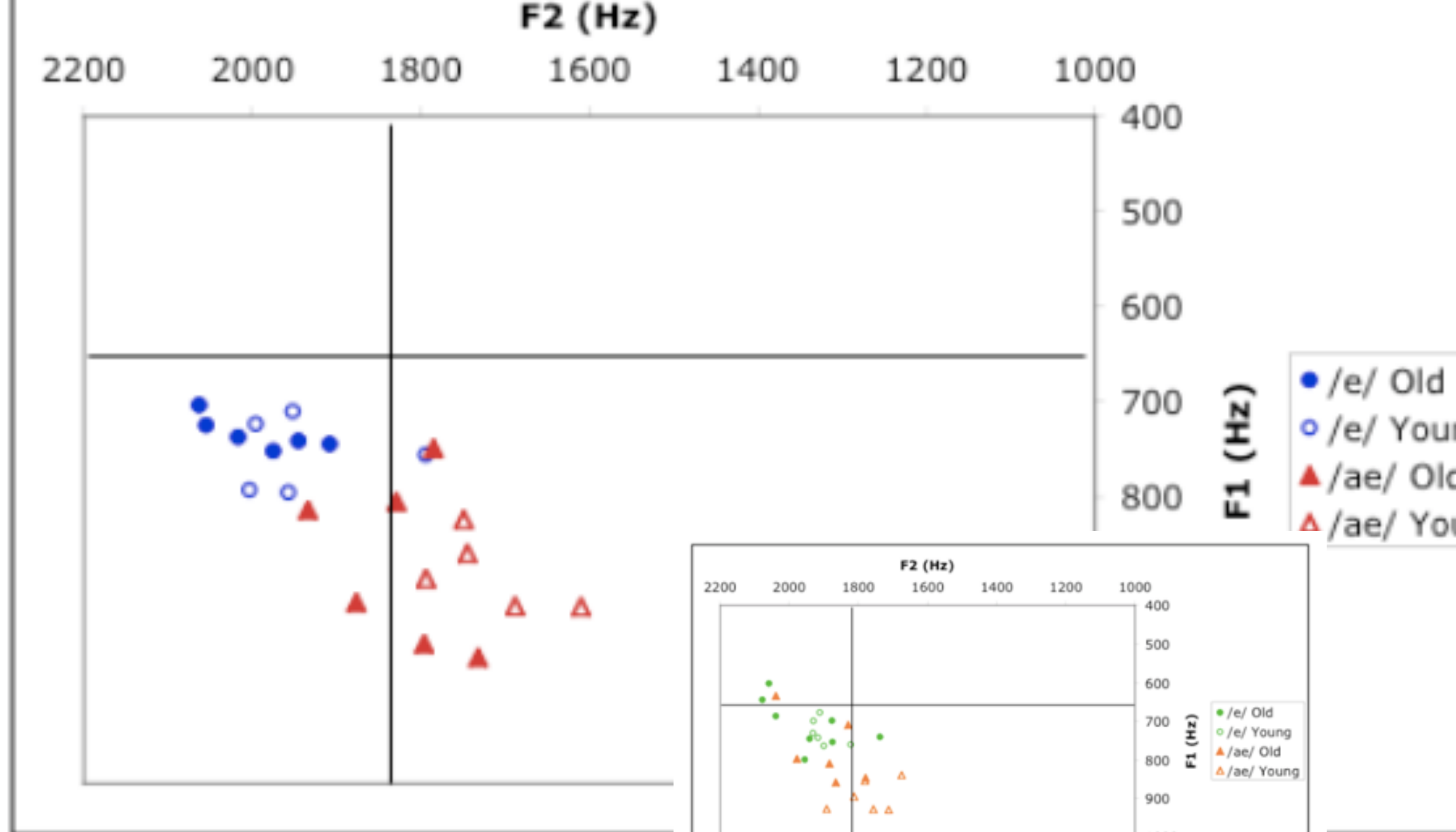


Fig 4. Individual means of /e/ and /ae/

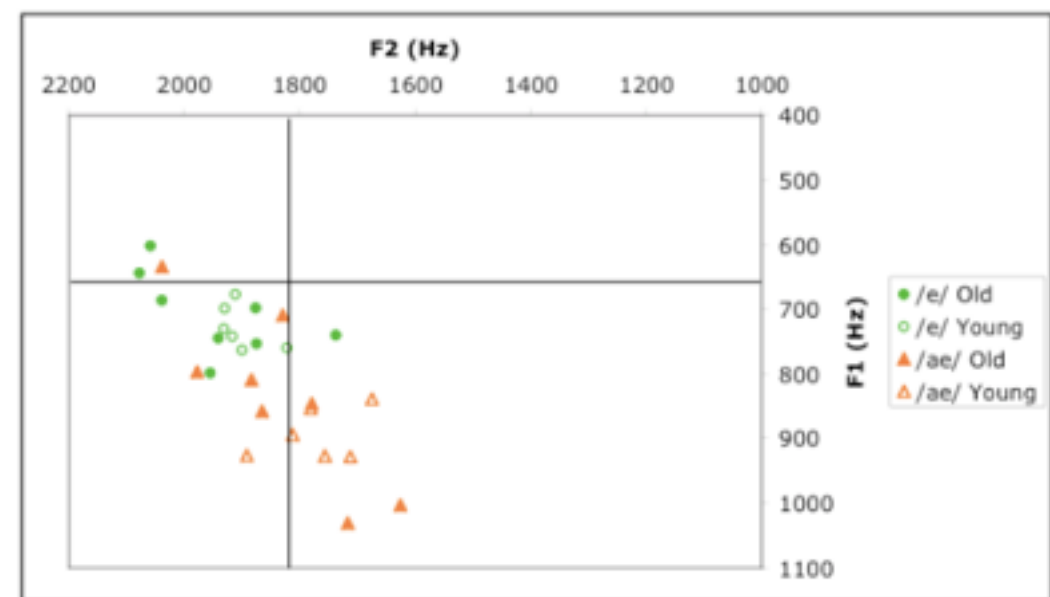
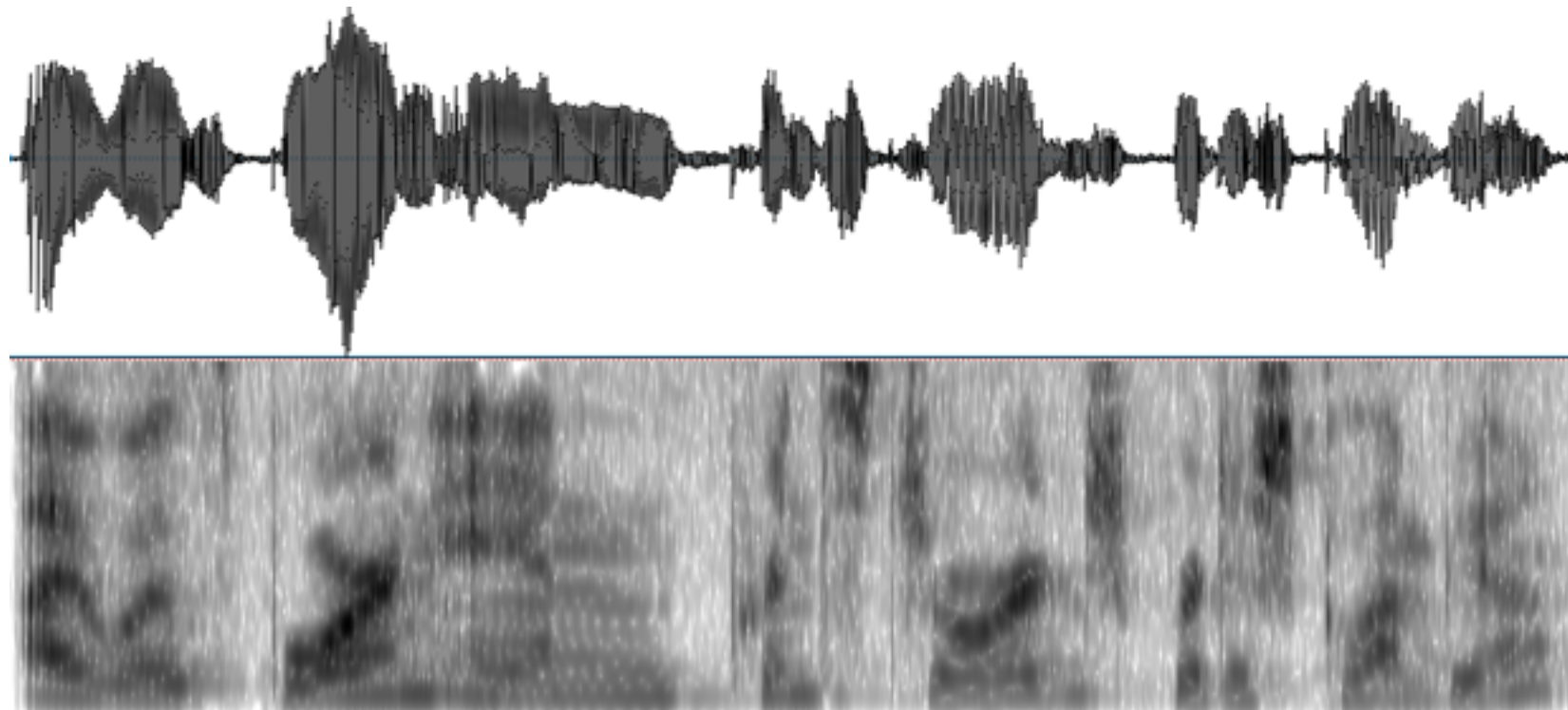


Fig. 5. Individual means of /e/ and /ae/ for Halifax speakers

Corpus sociophonetics

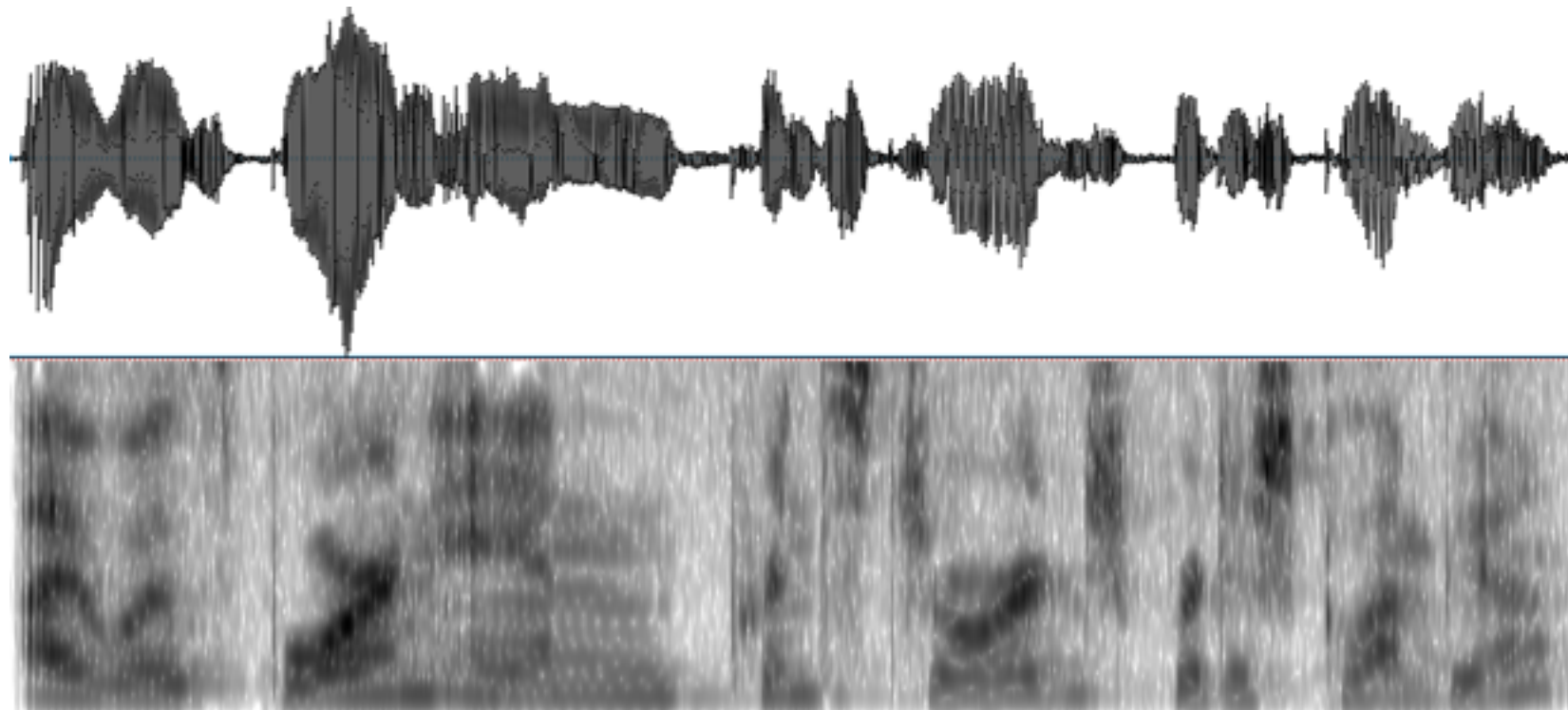
Step 1. Forced alignment



I WAS BORN IN PRINCE GEORGE BRITISH COLUMBIA

Corpus sociophonetics

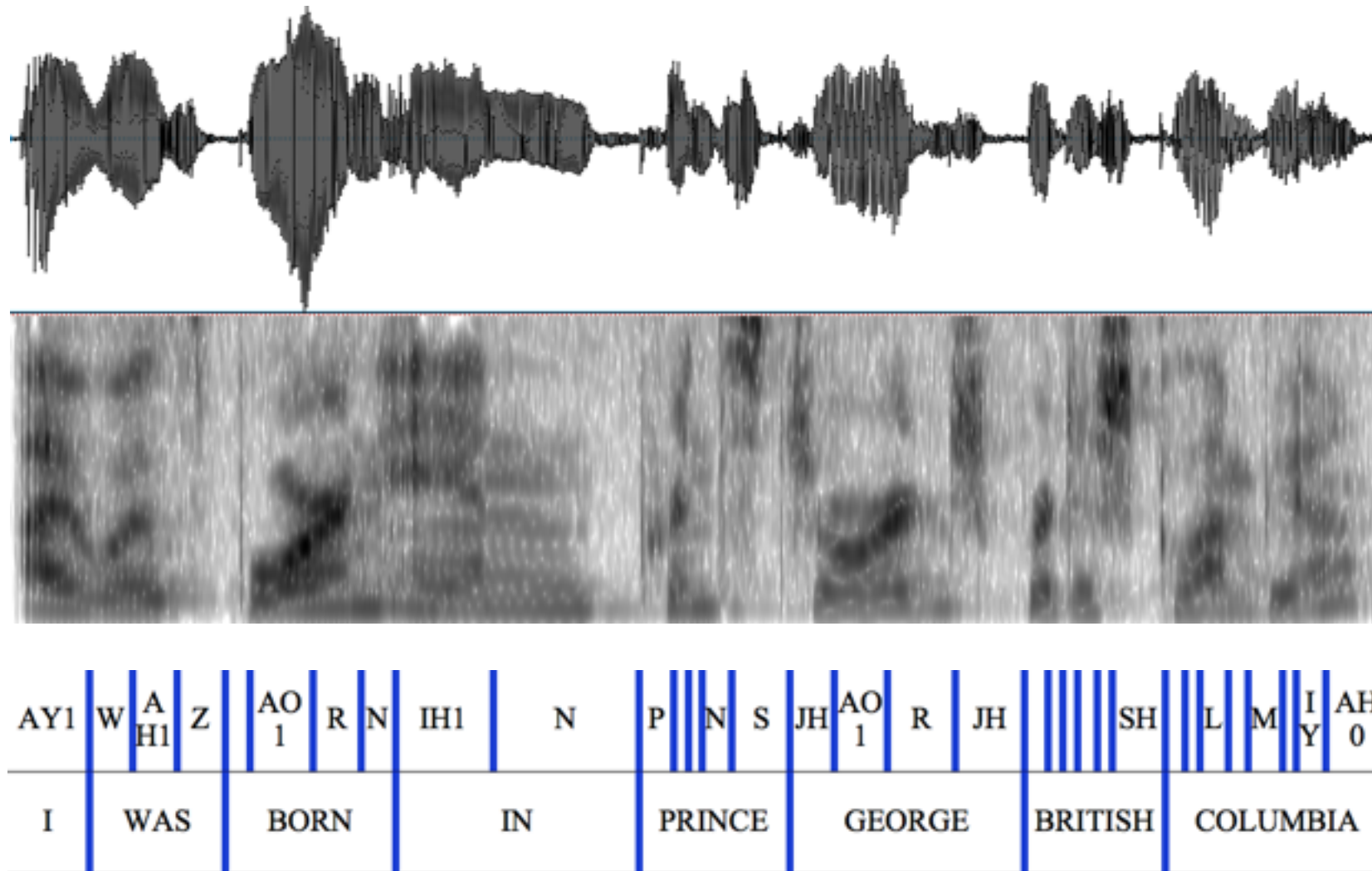
Step 1. Forced alignment



I WAS BORN IN PRINCE GEORGE BRITISH COLUMBIA

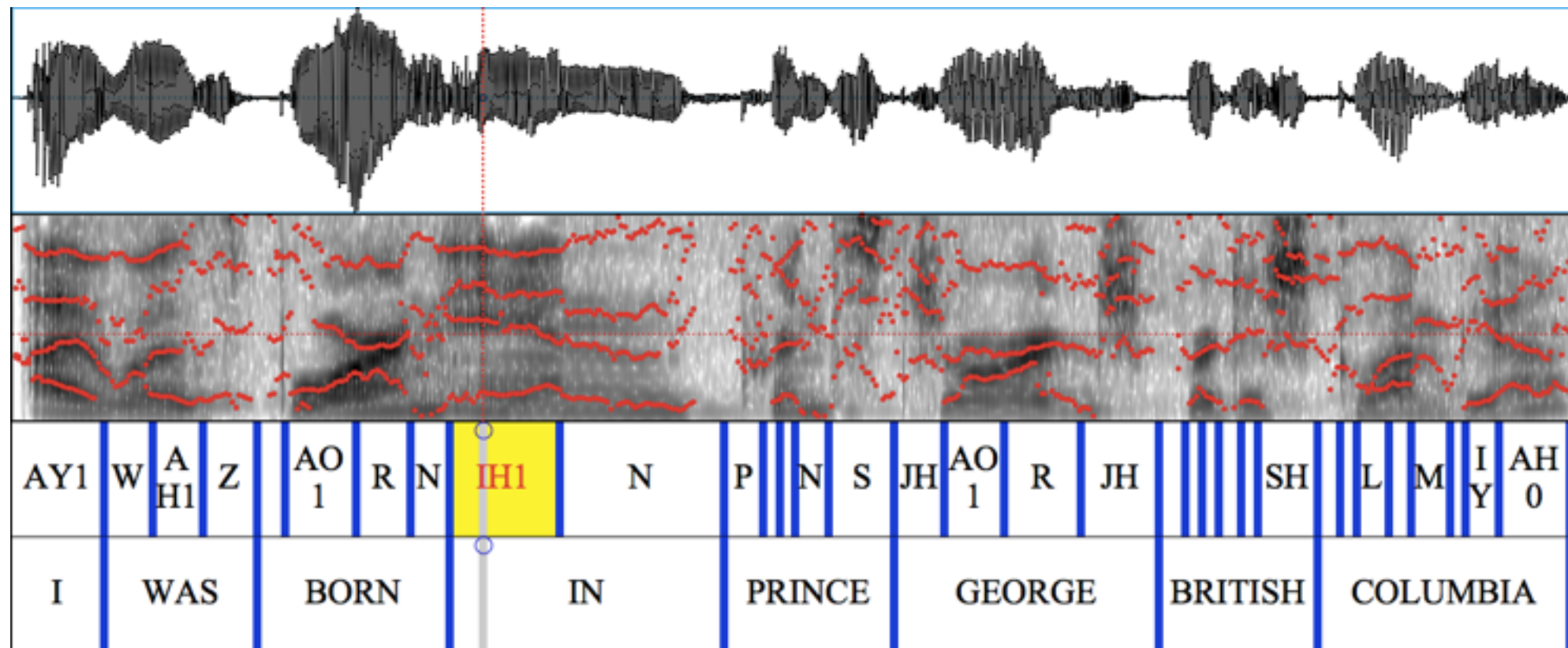
Corpus sociophonetics

Step 1. Forced alignment



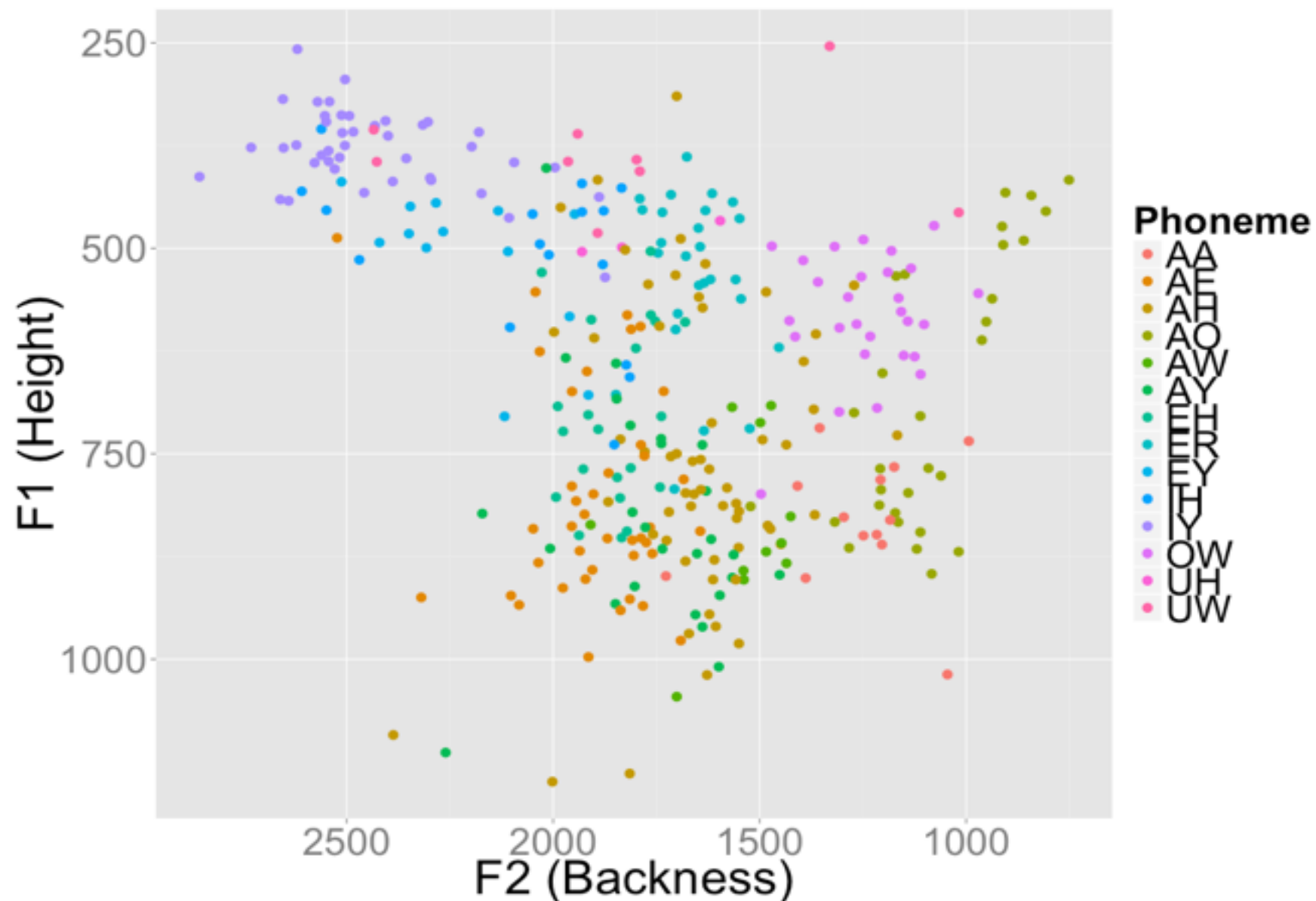
Corpus sociophonetics

Step 2. Automatic formant extraction



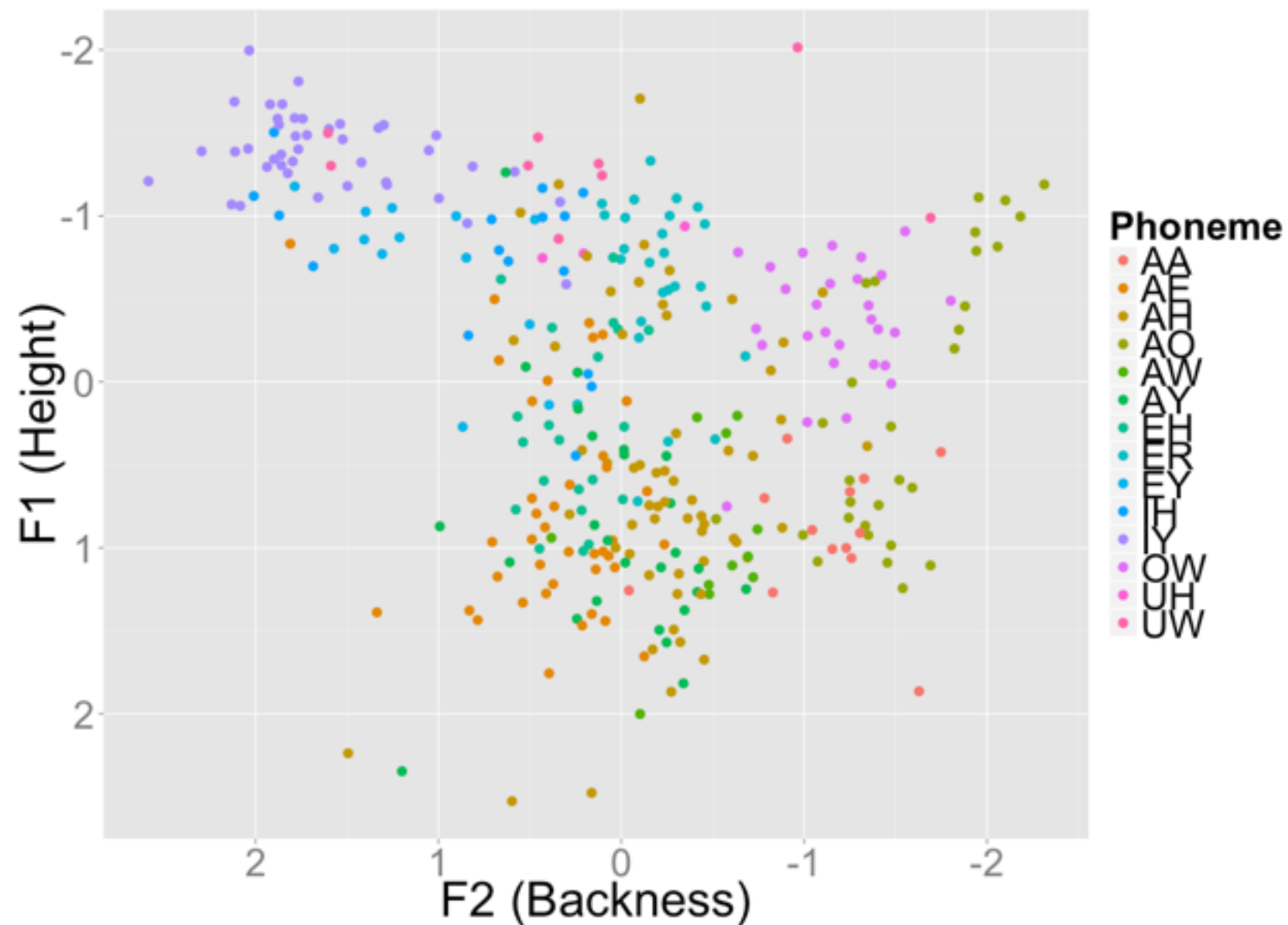
Corpus sociophonetics

Step 3. Normalization



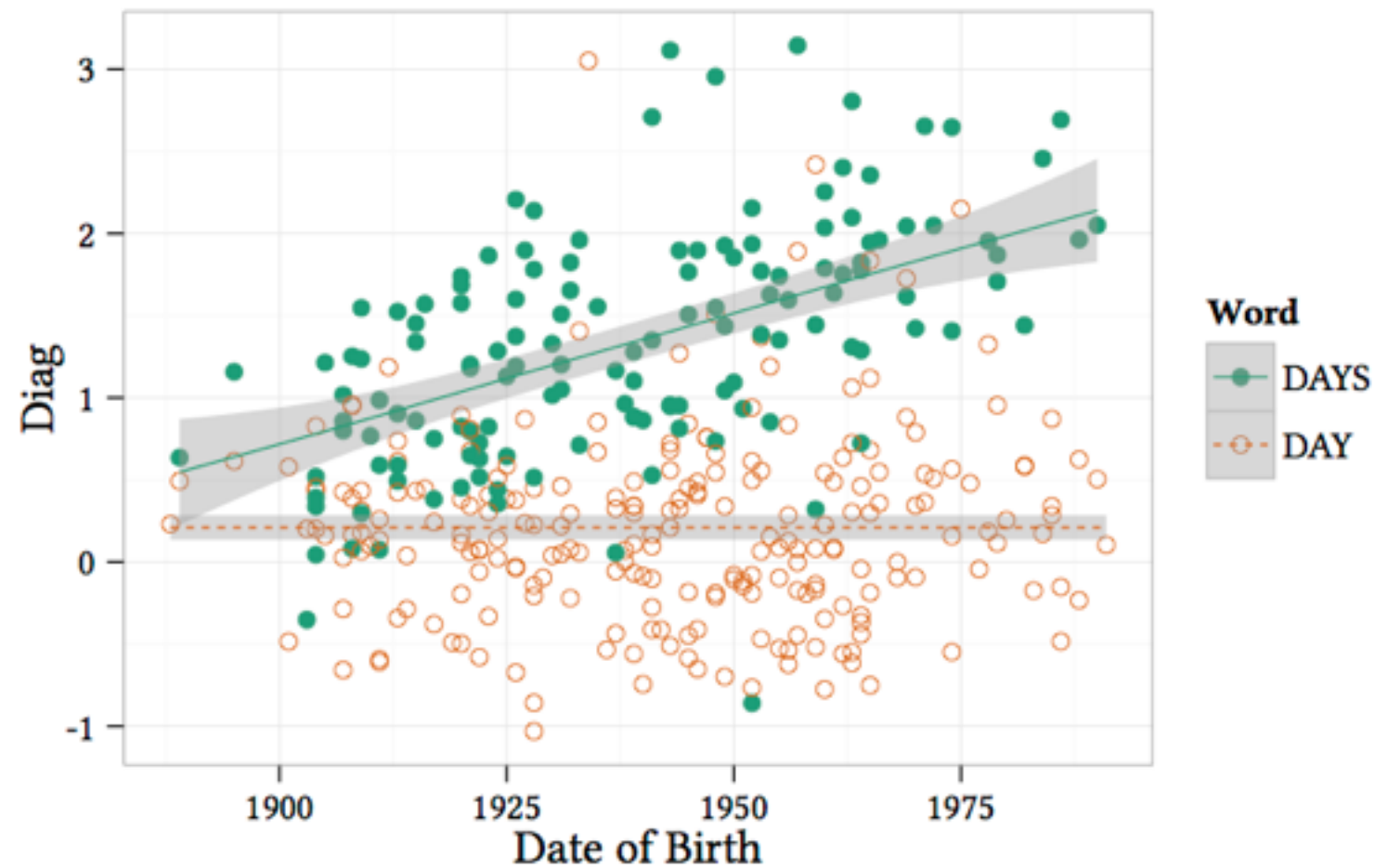
Corpus sociophonetics

Step 3. Normalization



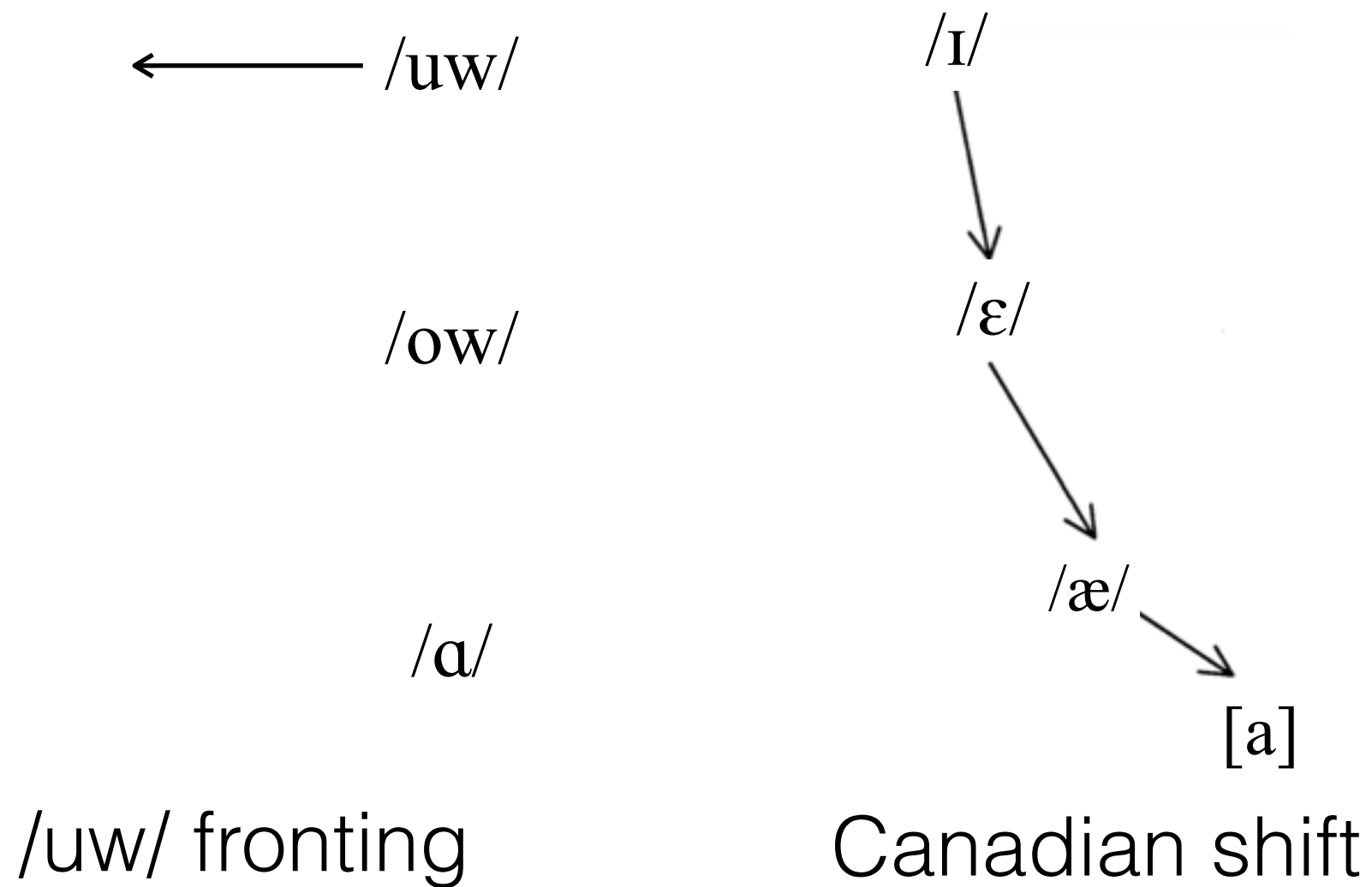
Corpus sociophonetics

Step 4. Analysis



Sounding board

Canadian English



Corpus

Canadian English

Two publicly available audio + transcript archives
63 speakers (31 male, 32 female), mainly from Ontario (45)
4.5 hours of speech; 56,781 vowel tokens in transcripts

For comparison, the Philadelphia Neighborhood Corpus

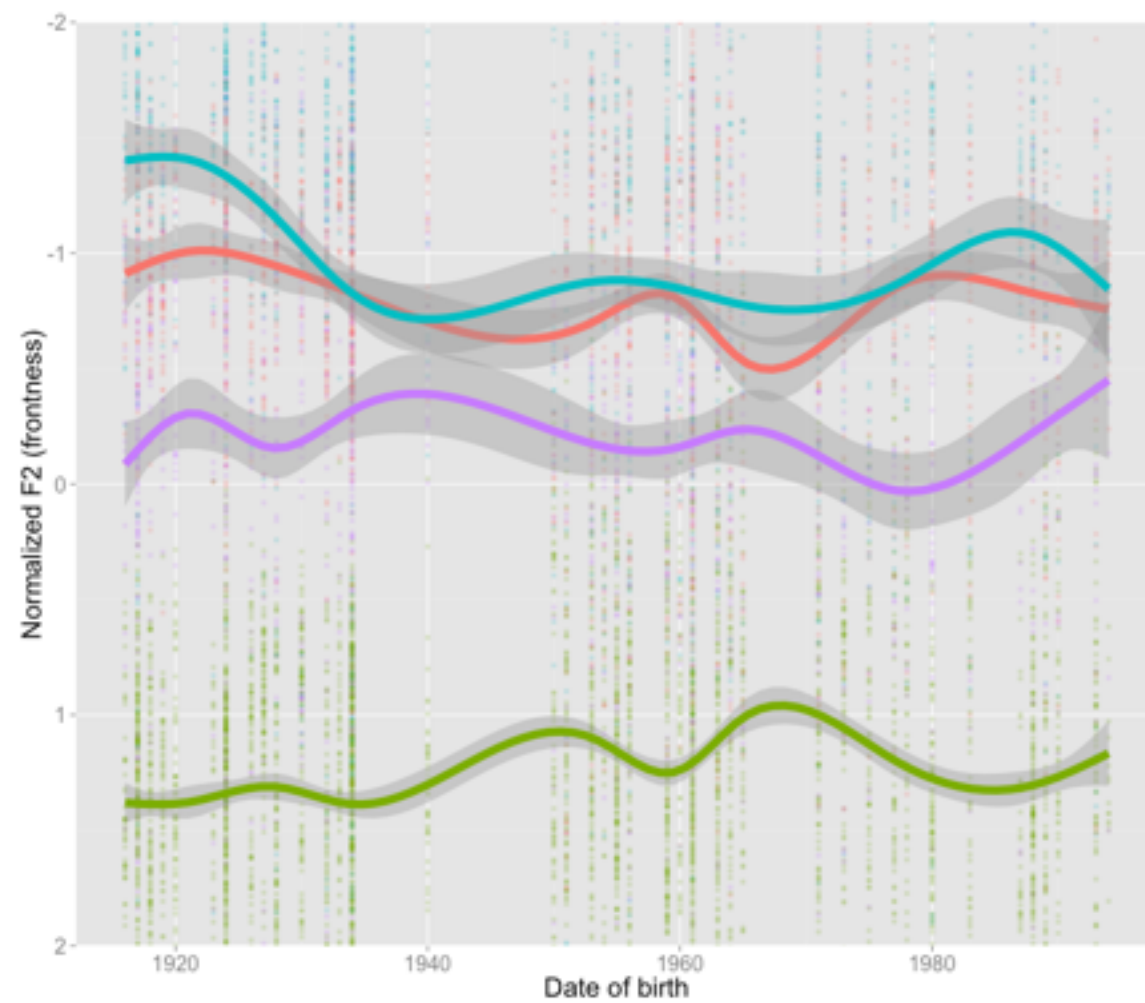
308 speakers, 735,408 vowel tokens

Treatment

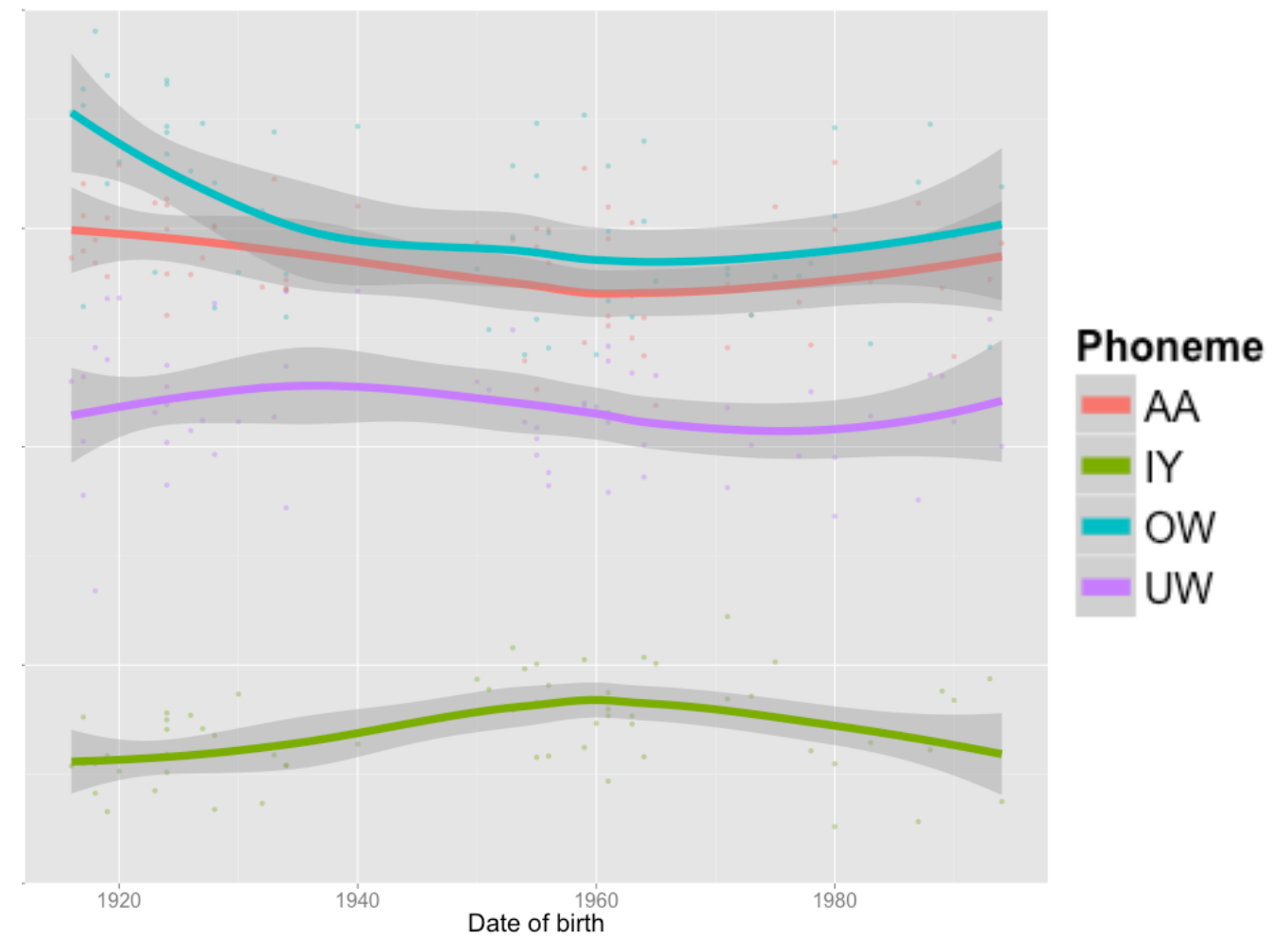
Forced alignment with Penn Forced Aligner (sloppy)
Formant extraction at first third of vowel using FAVE (sloppy)
Normalized formant values within speaker

/uw/ fronting

Raw values

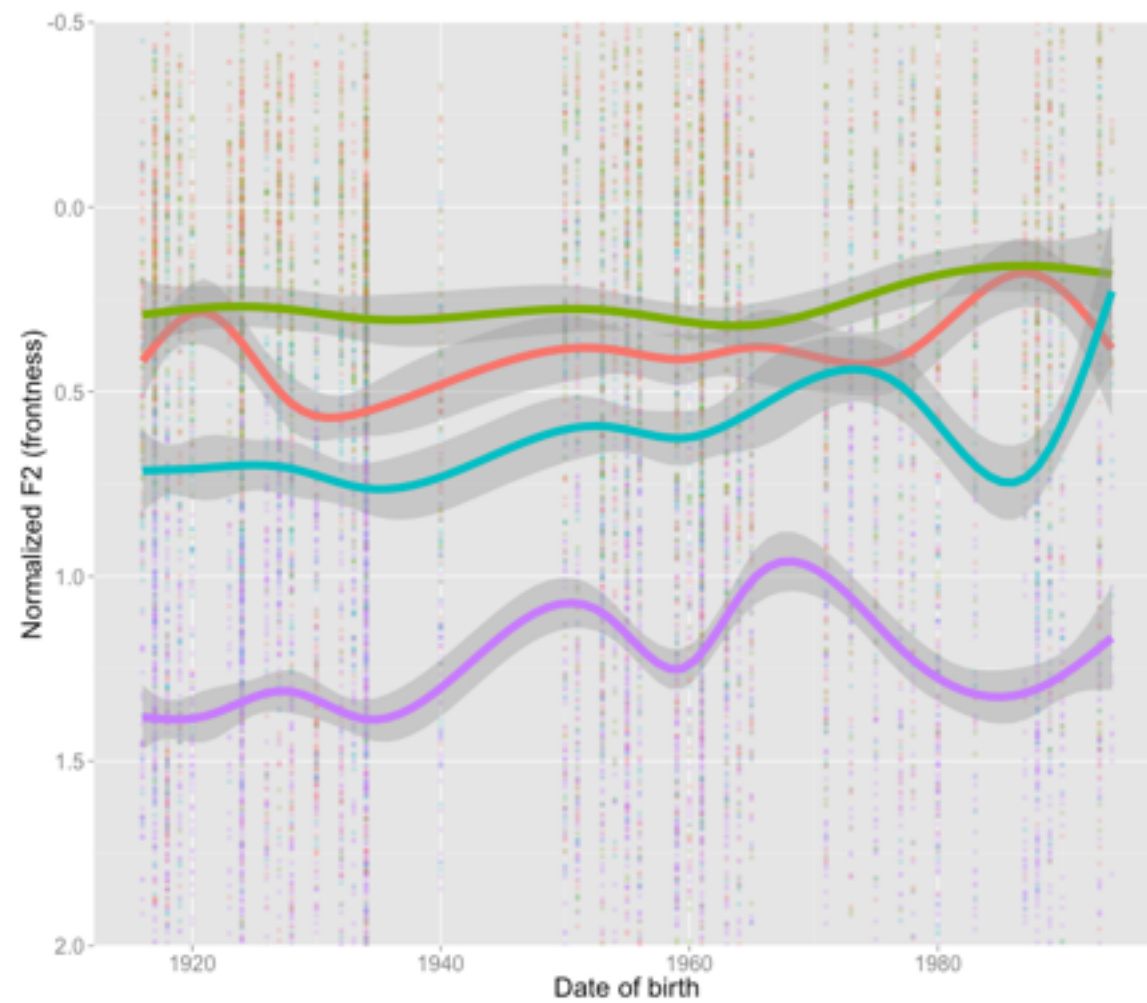


Subject means

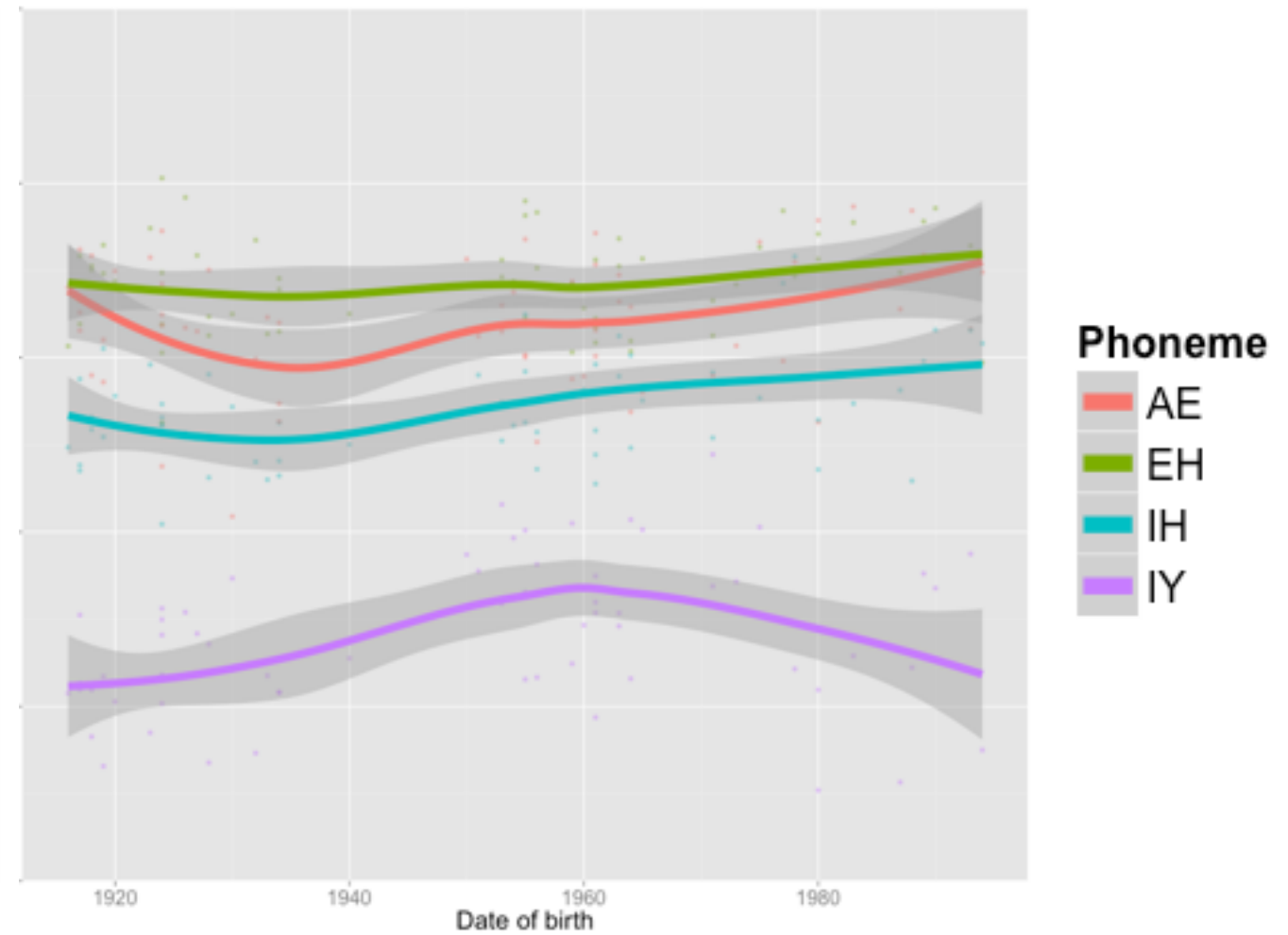


Canadian shift

Raw values

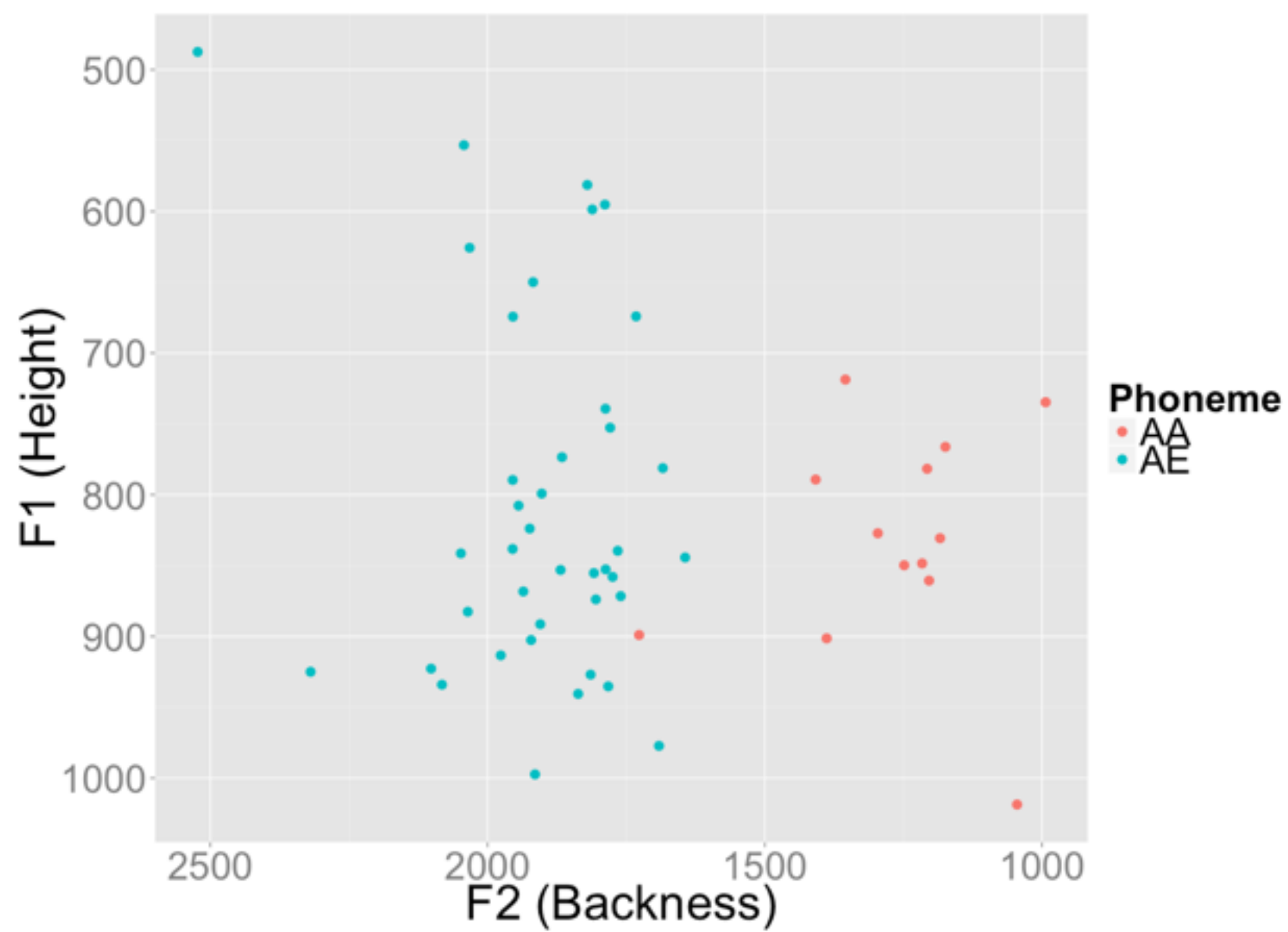


Subject means

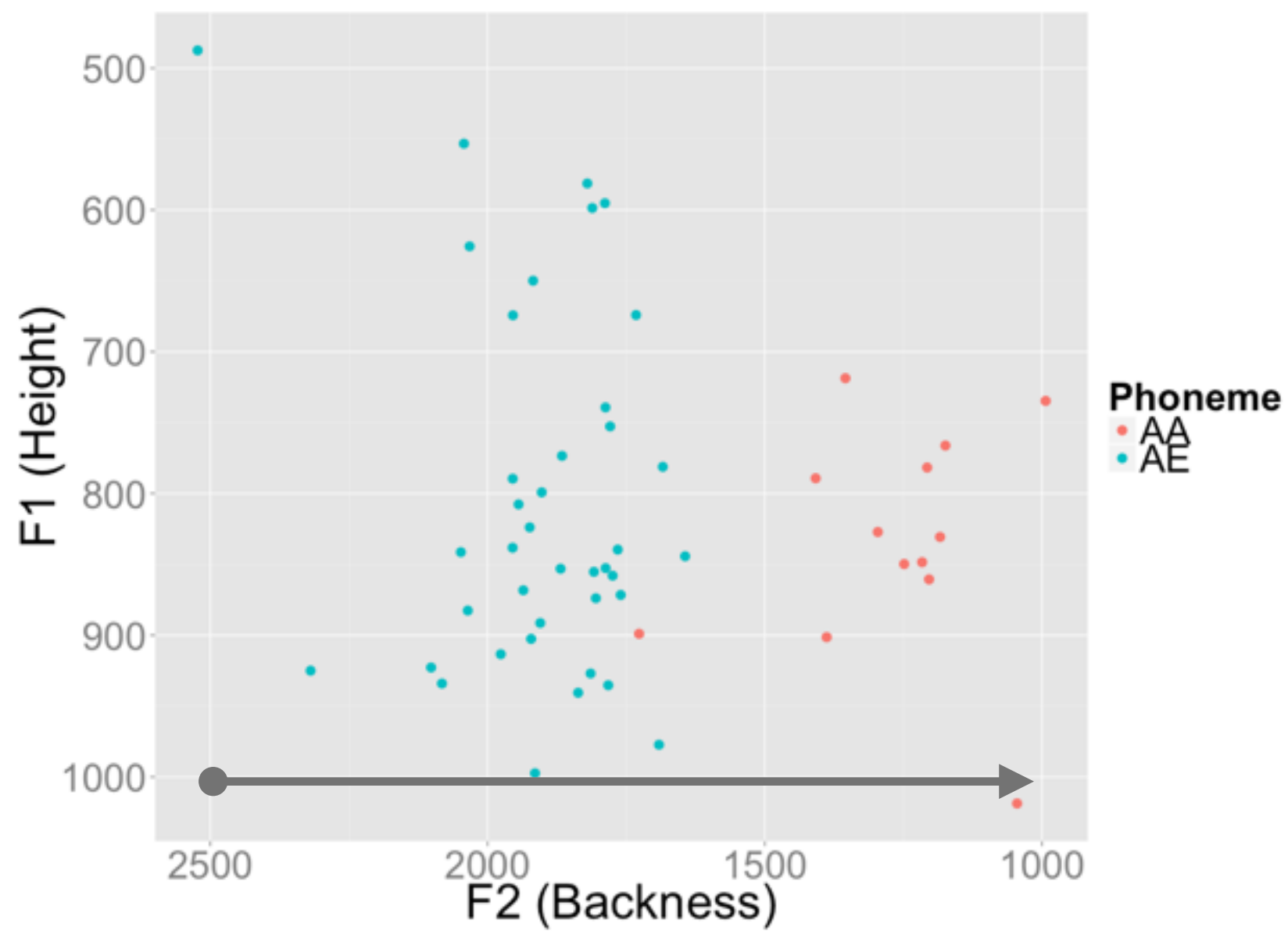


A new approach

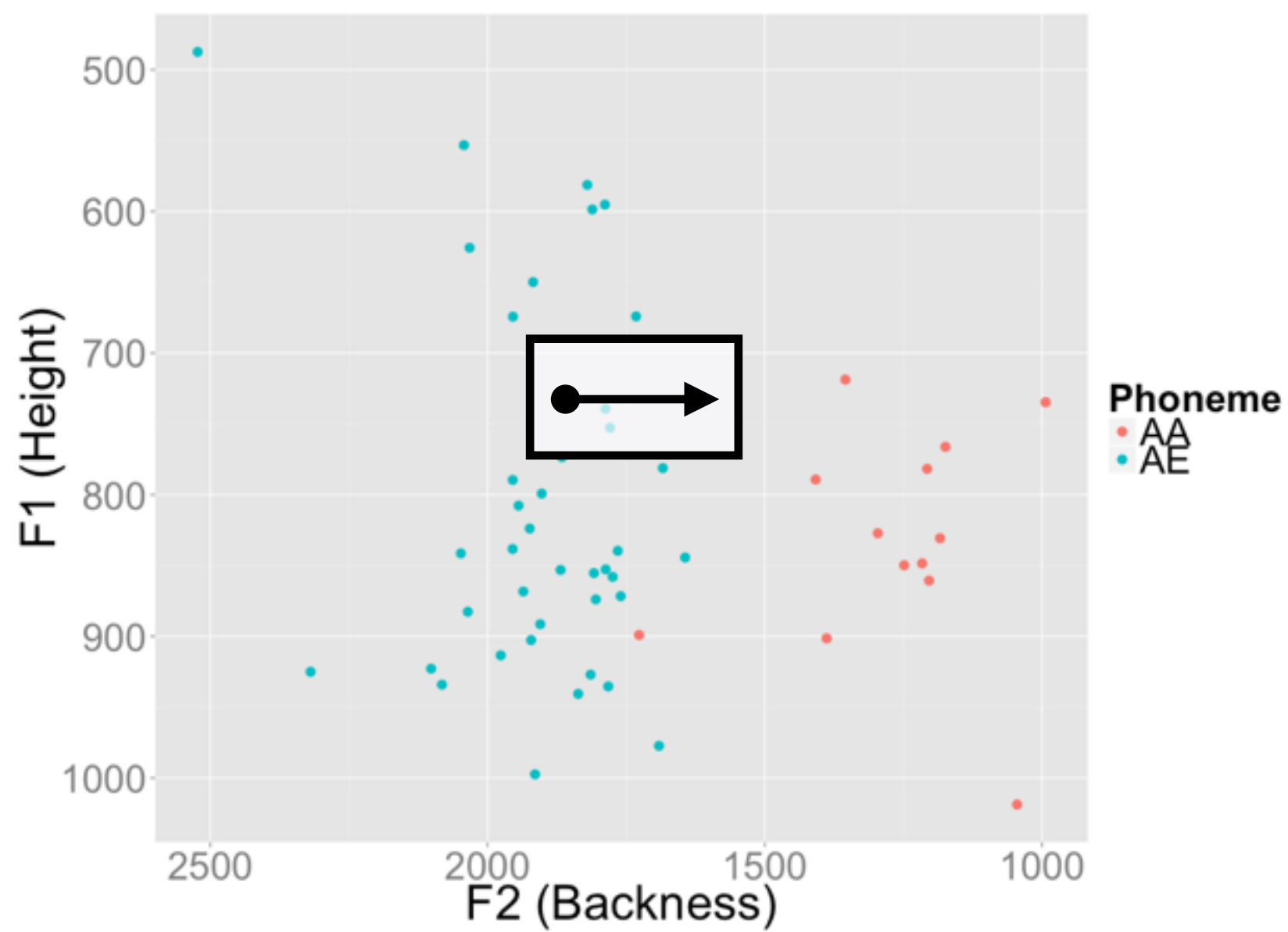
Idea



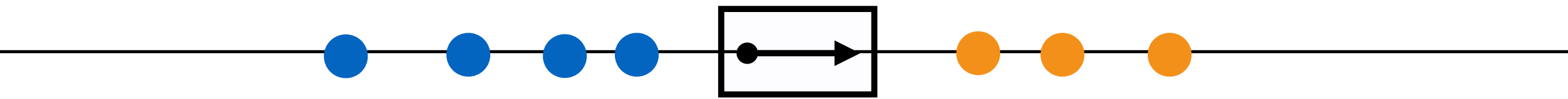
Idea



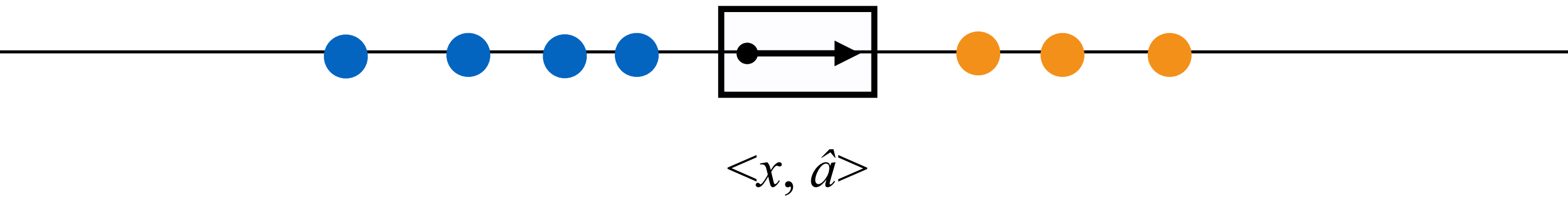
Idea



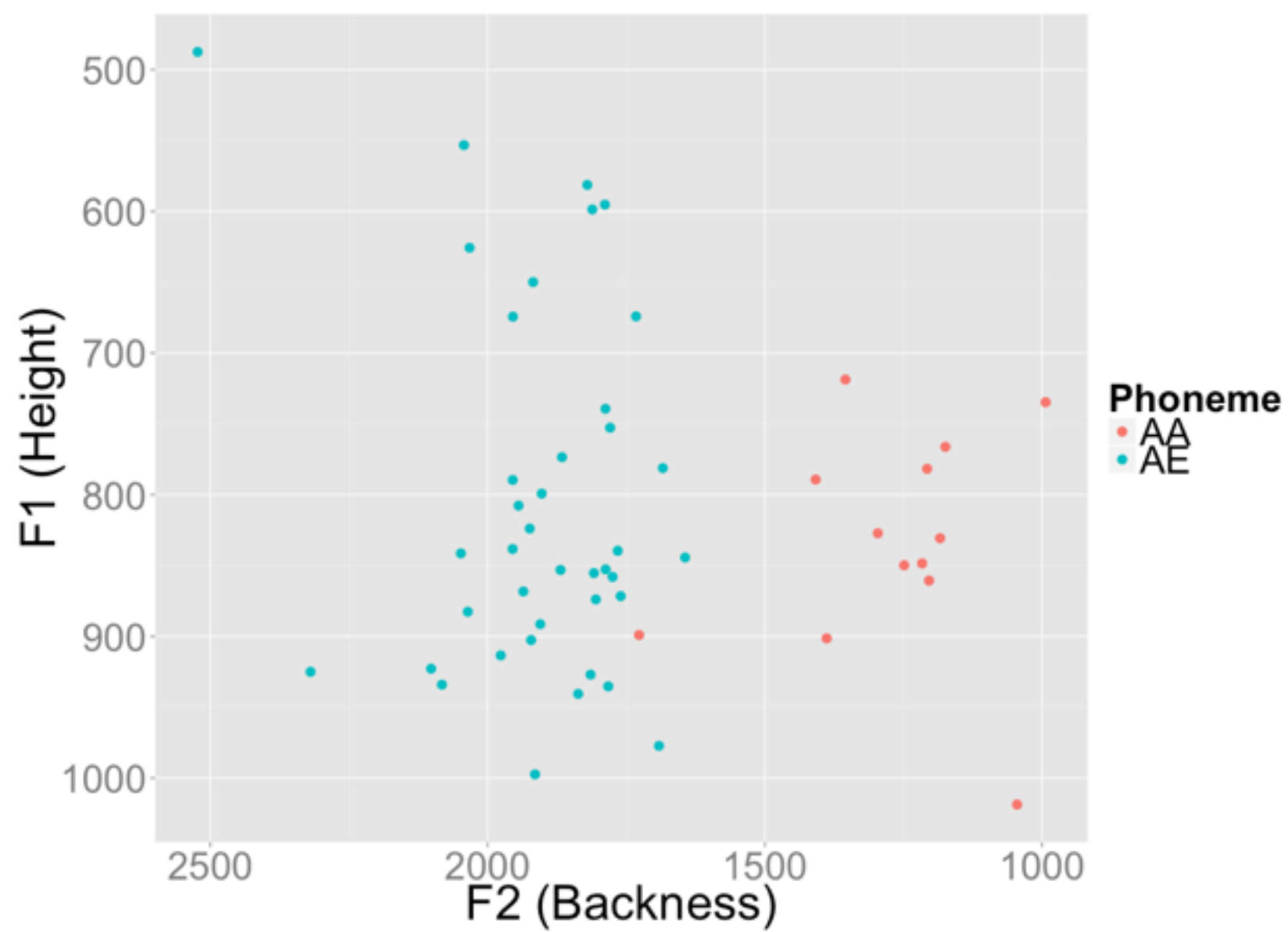
Idea



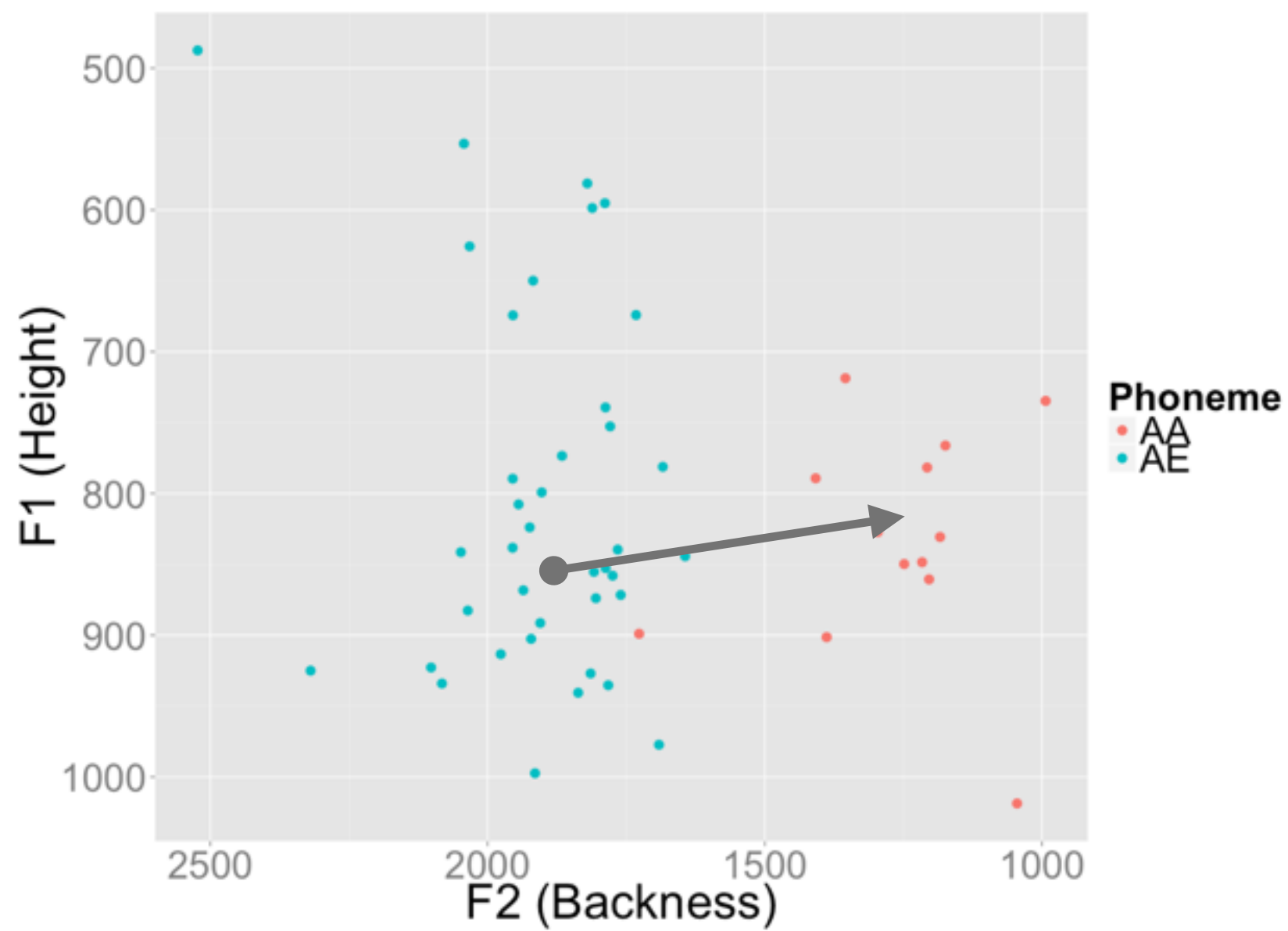
Idea



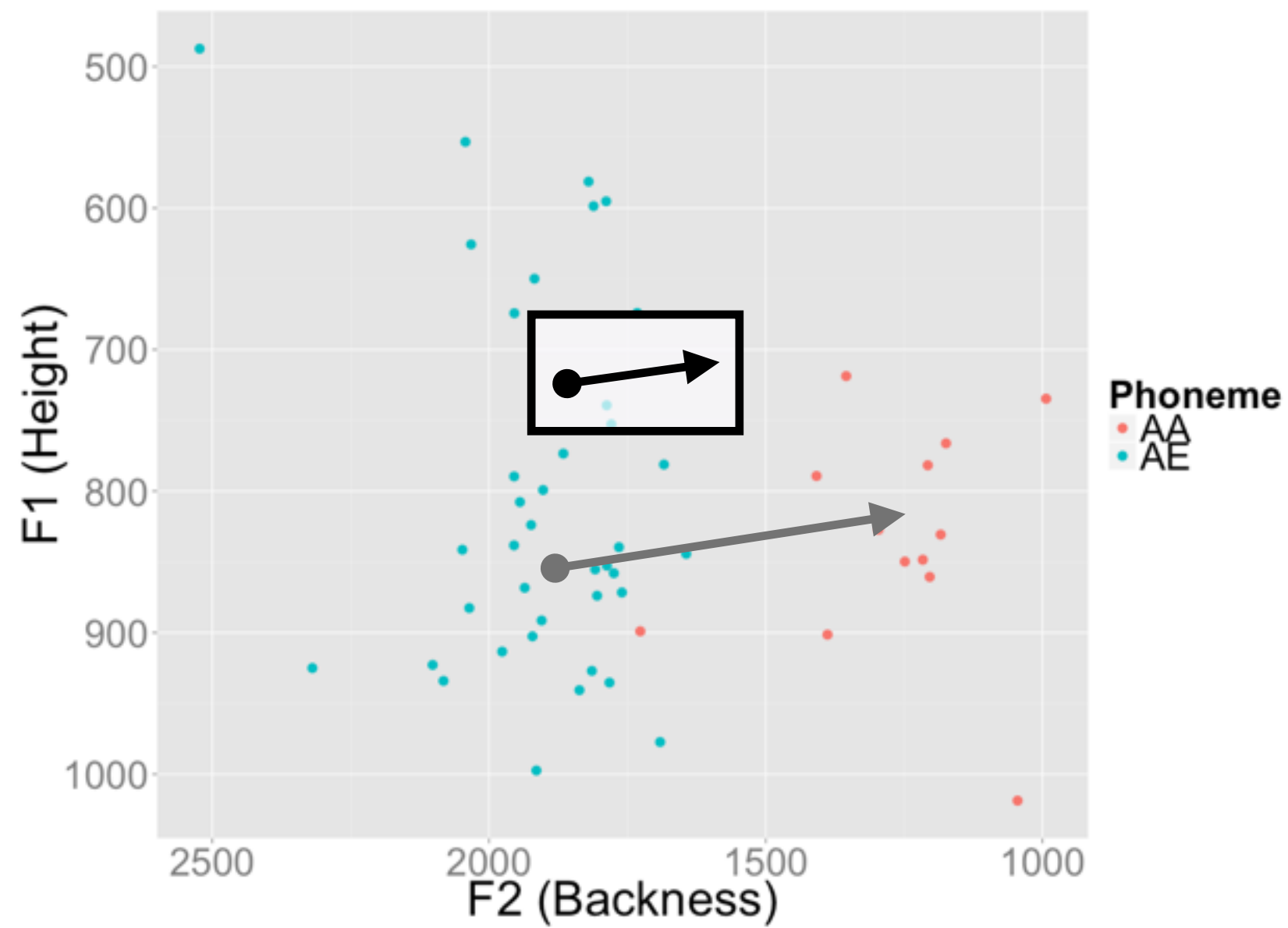
Idea



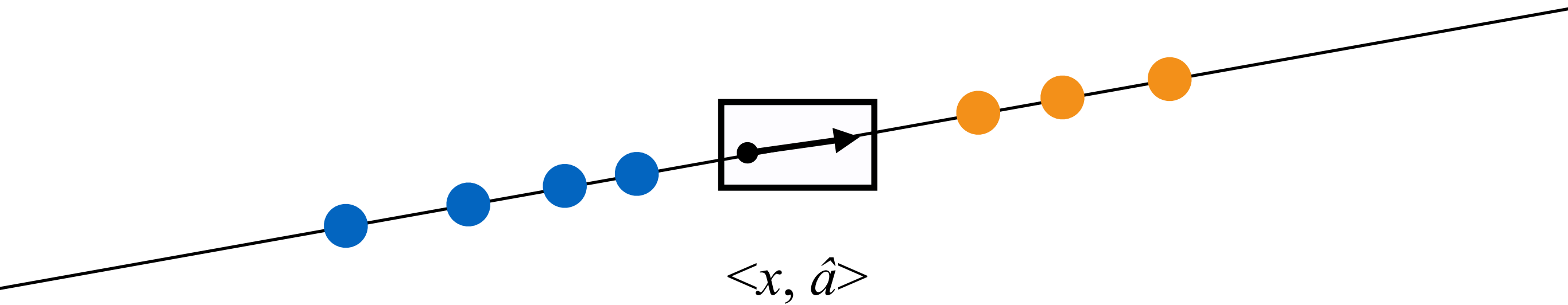
Idea



Idea

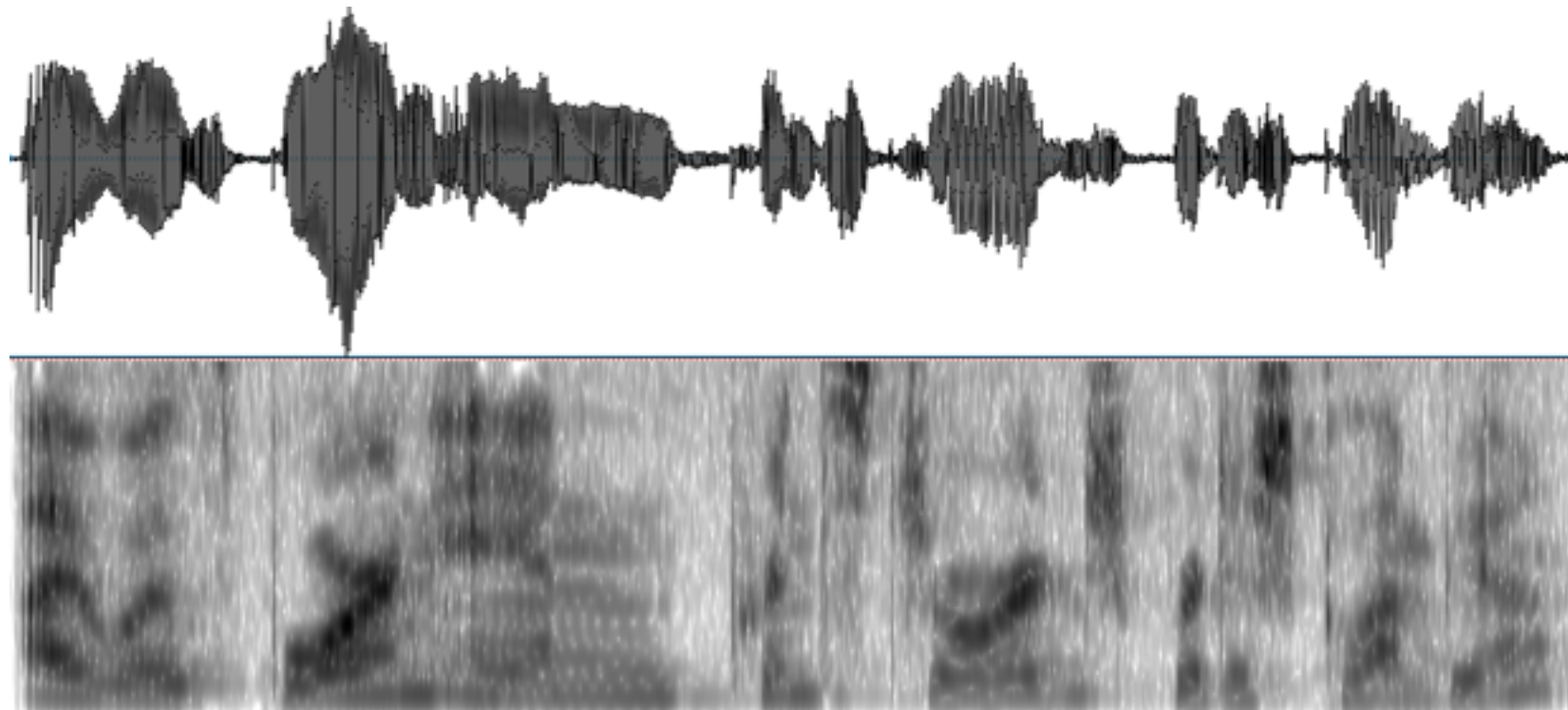


Idea



A new approach

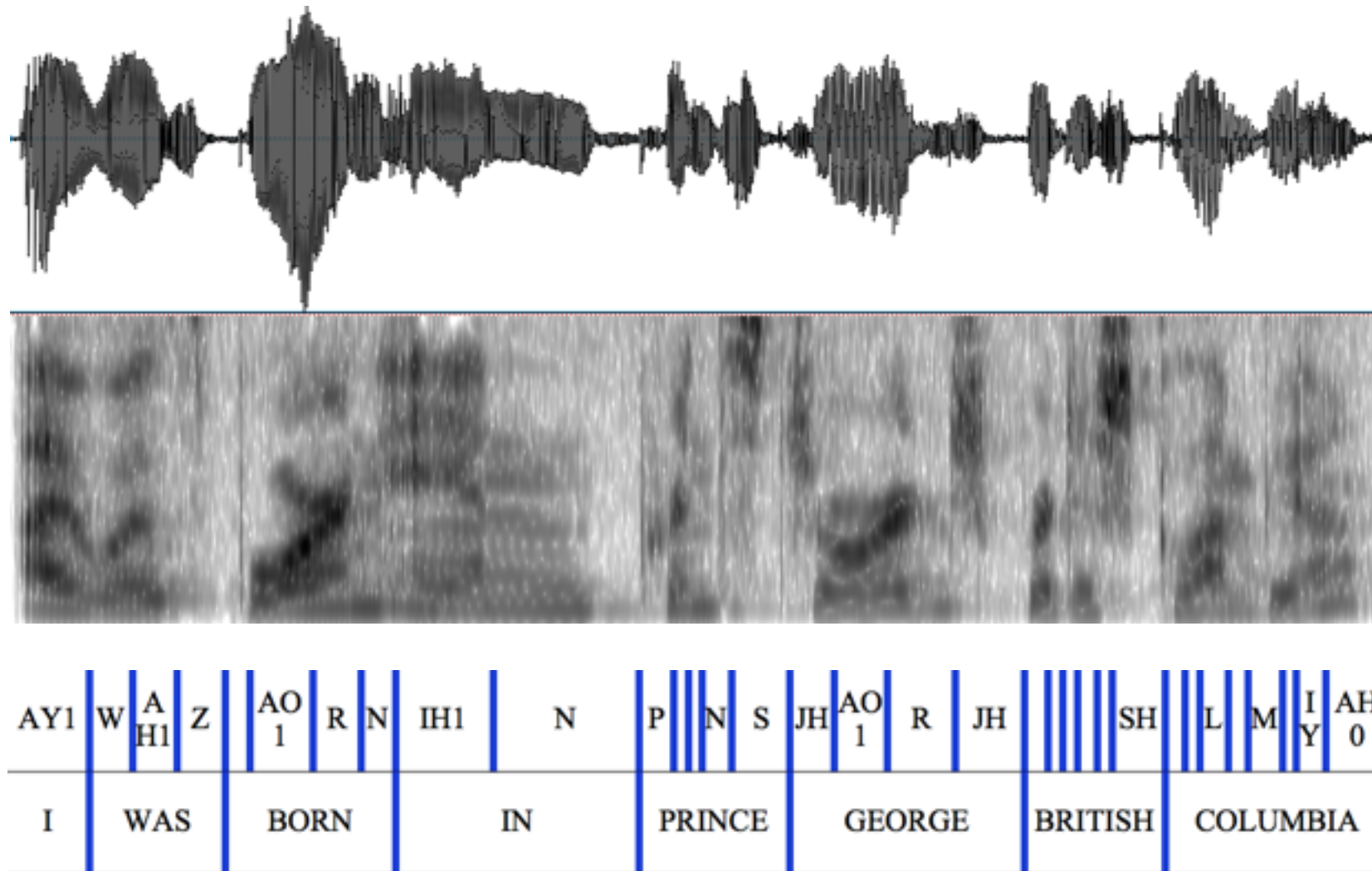
Step 1. Forced alignment



I WAS BORN IN PRINCE GEORGE BRITISH COLUMBIA

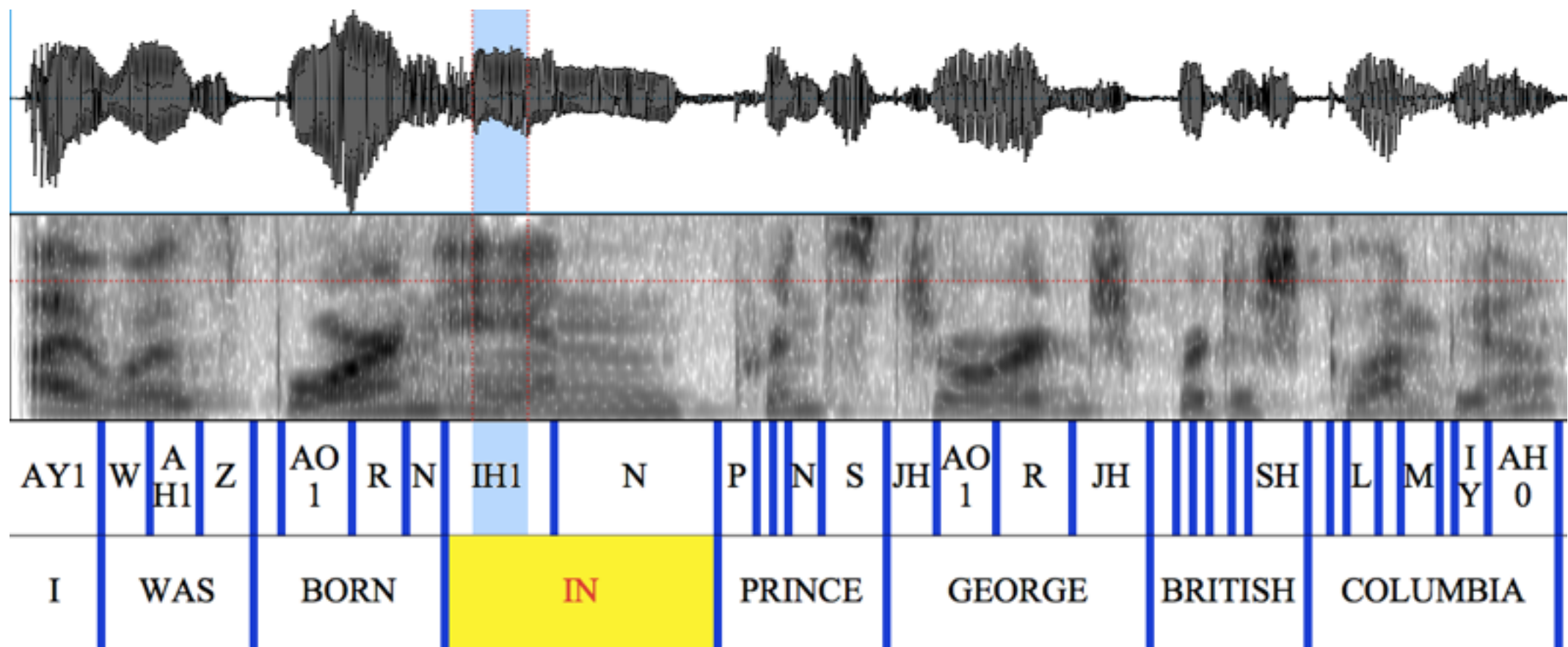
A new approach

Step 1. Forced alignment



A new approach

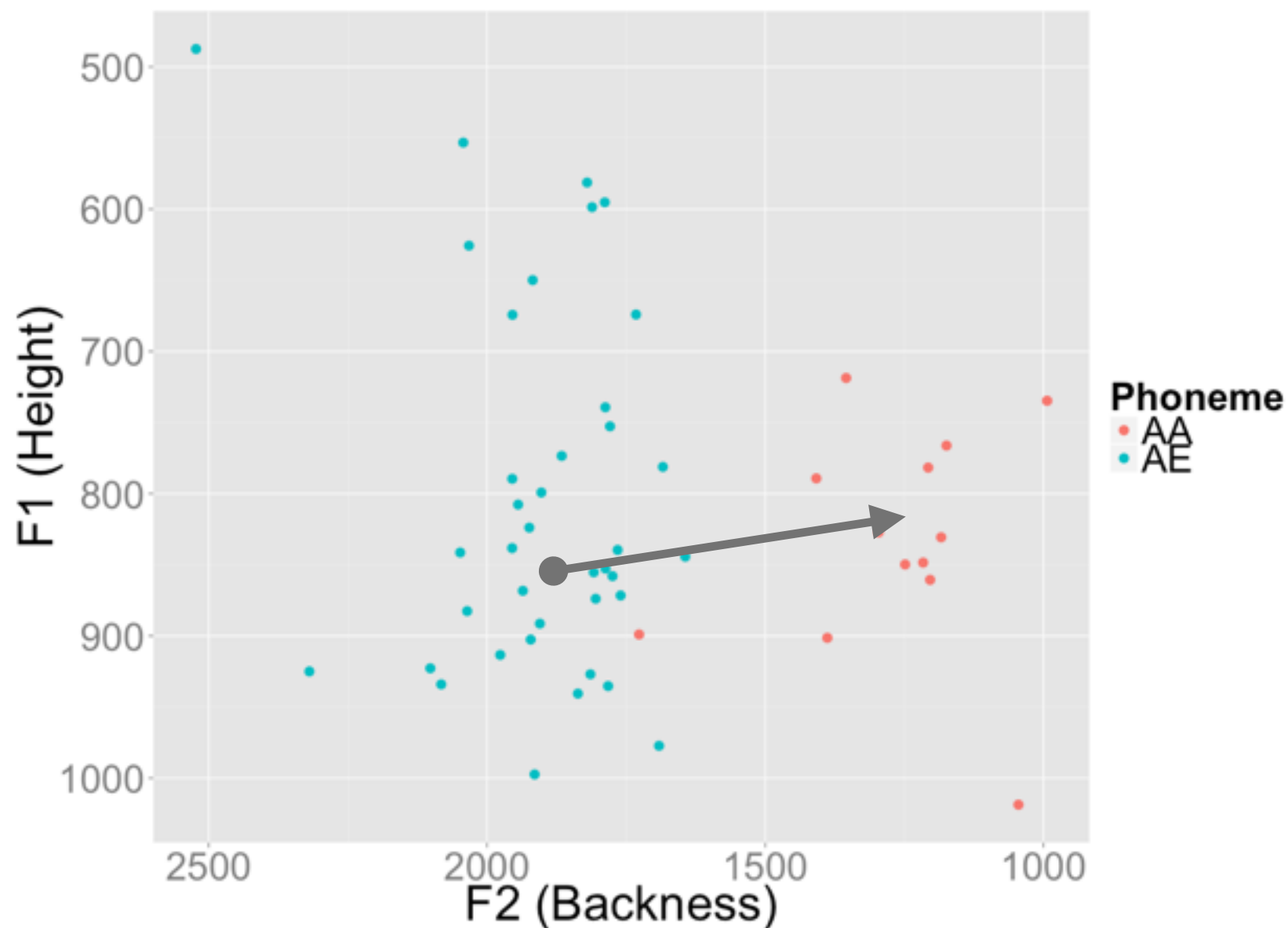
Step 2. Extract your favorite acoustic features



40 dimensional Mel-scaled filterbanks x 5 frames centered at midpoint
(200 dimensions)

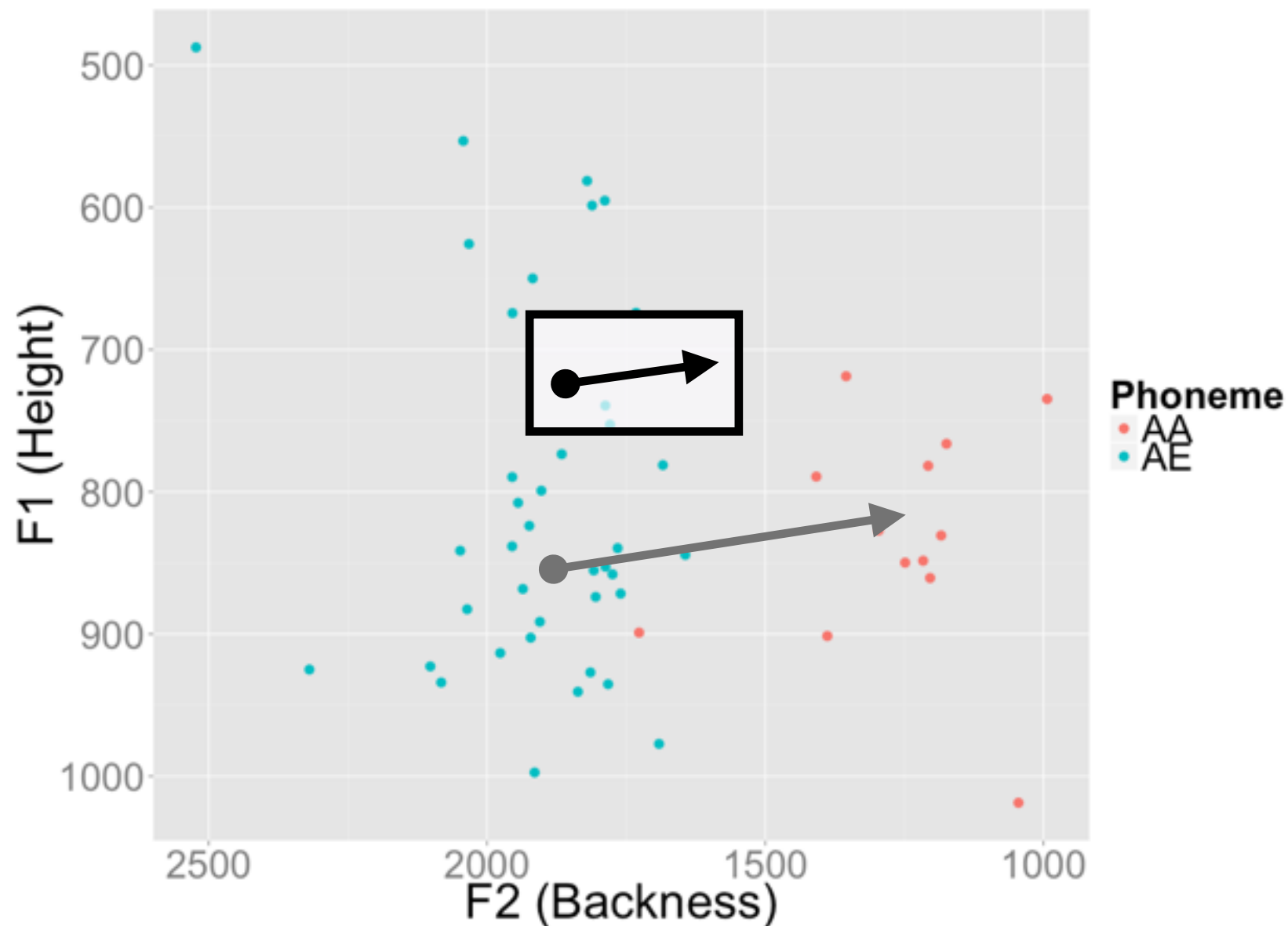
A new approach

Step 3. Construct a contrast vector



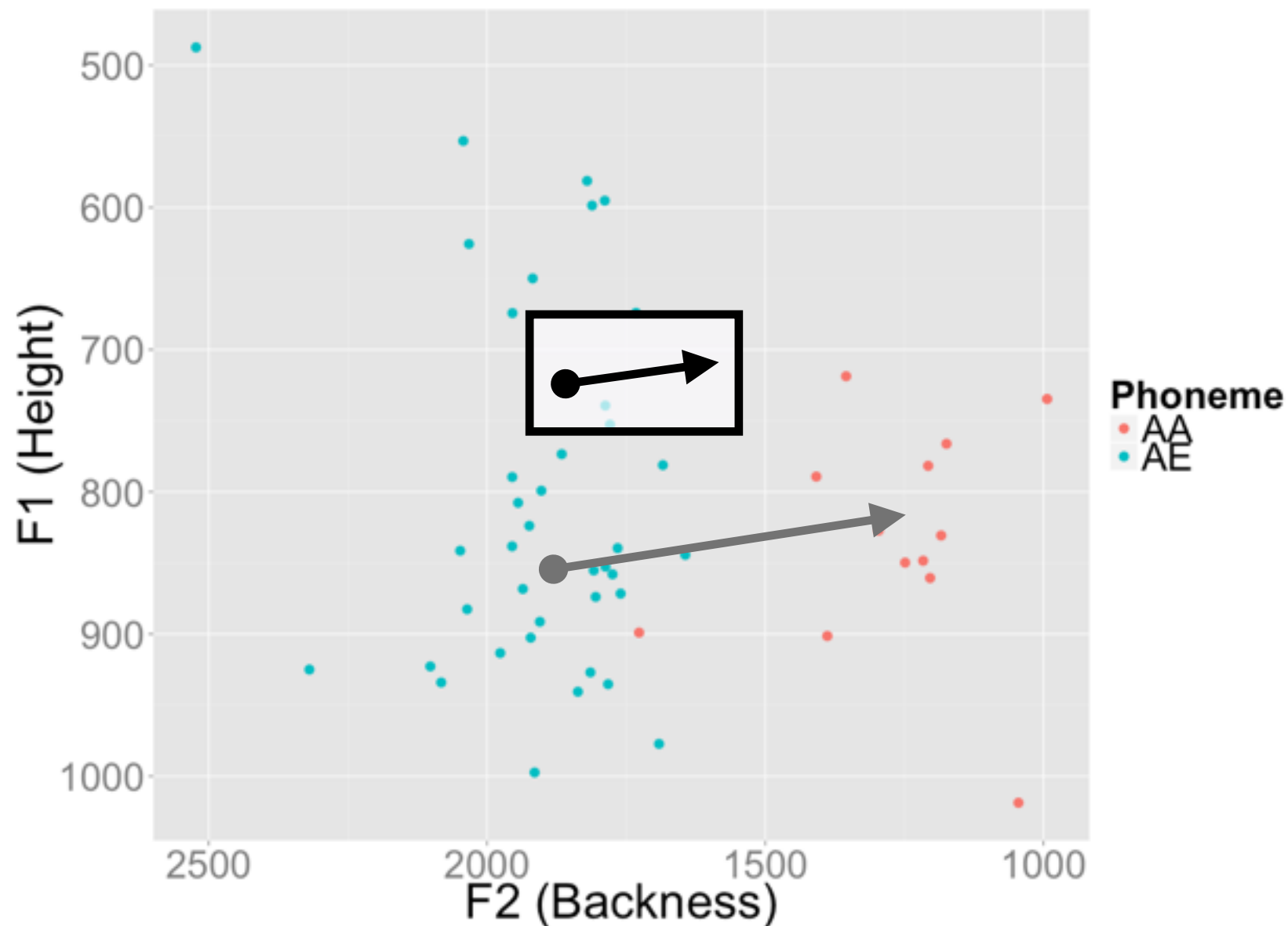
A new approach

Step 3. Construct a contrast vector



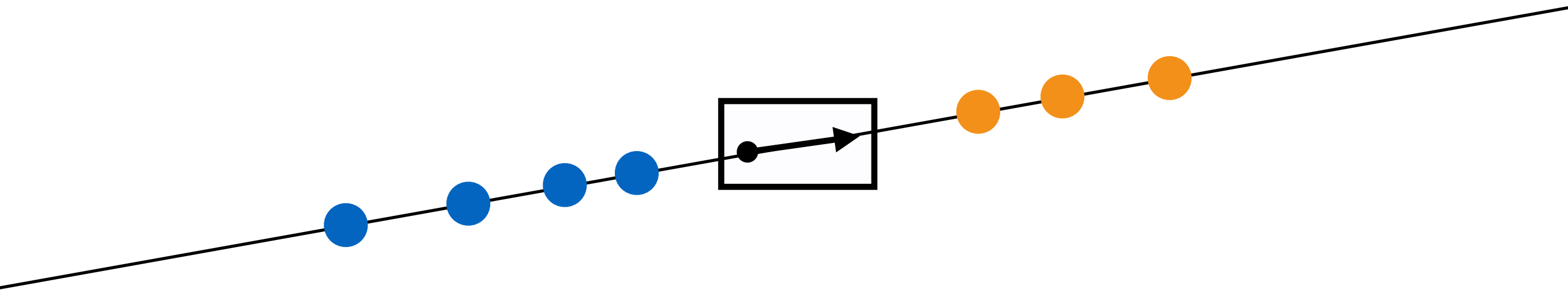
A new approach

Step 4. Project everything onto the contrast vector and normalize within speaker



A new approach

Step 4. Project everything onto the contrast vector and normalize within speaker

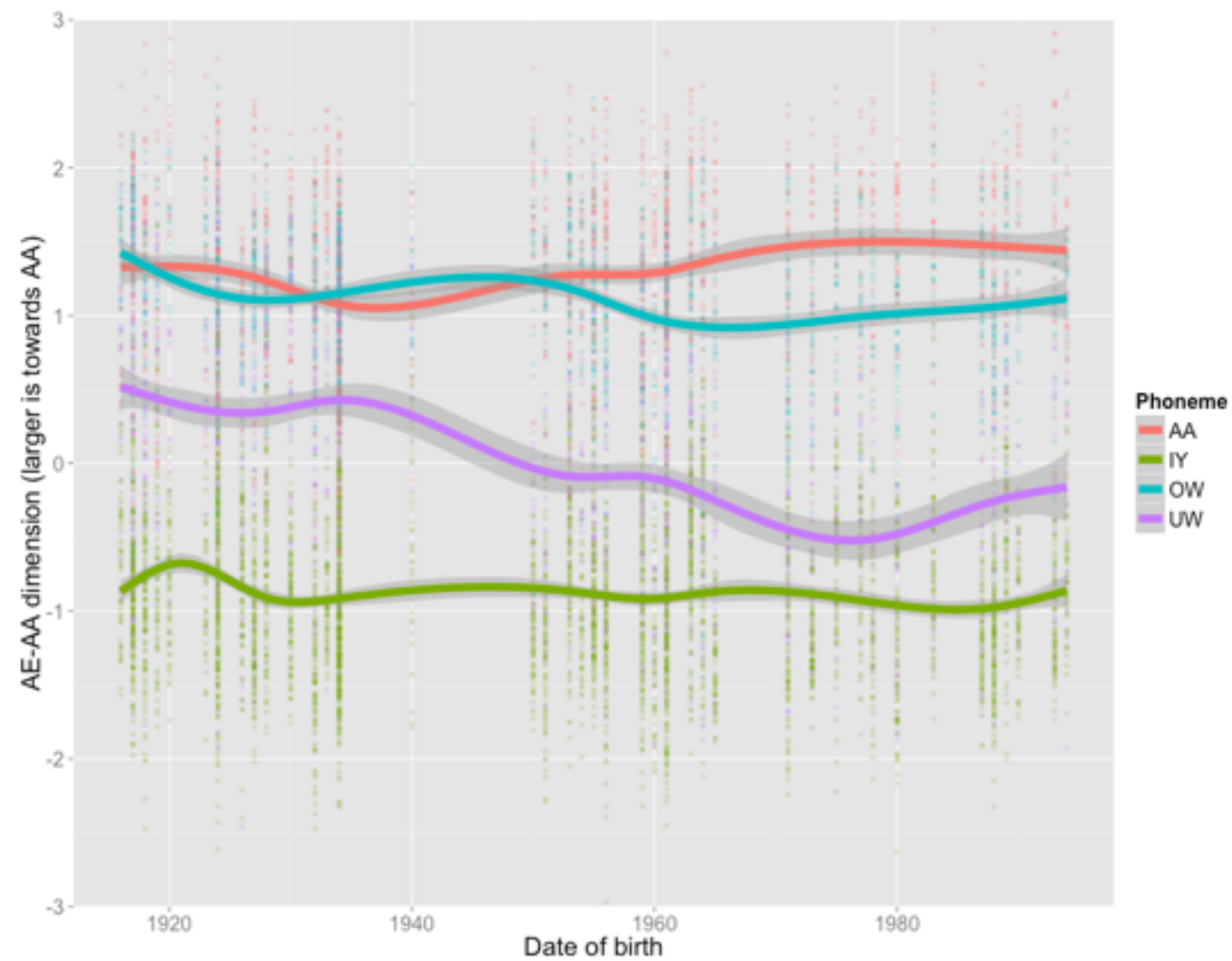


/uw/ fronting

AE-AA dimension

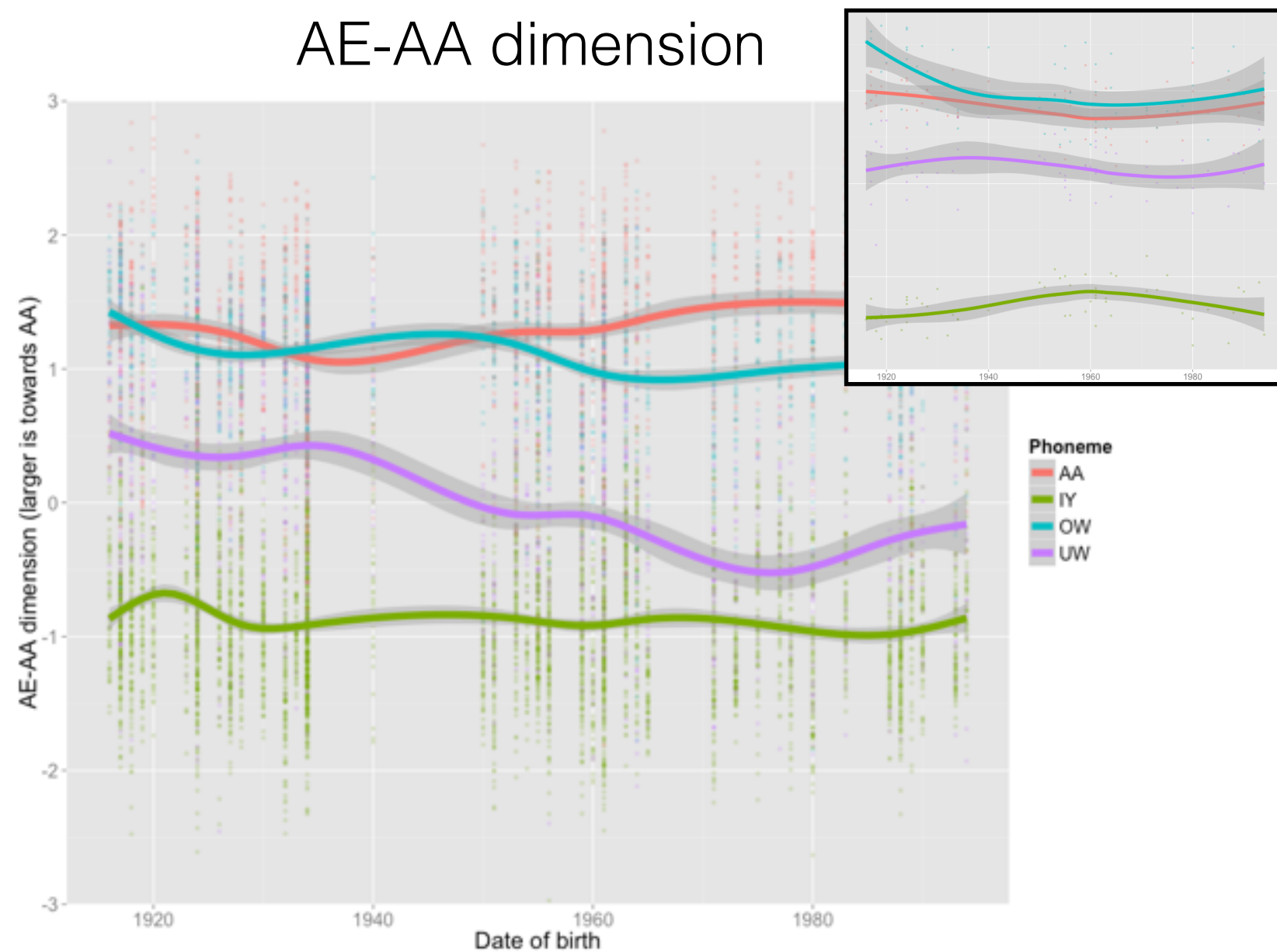
/uw/ fronting

AE-AA dimension



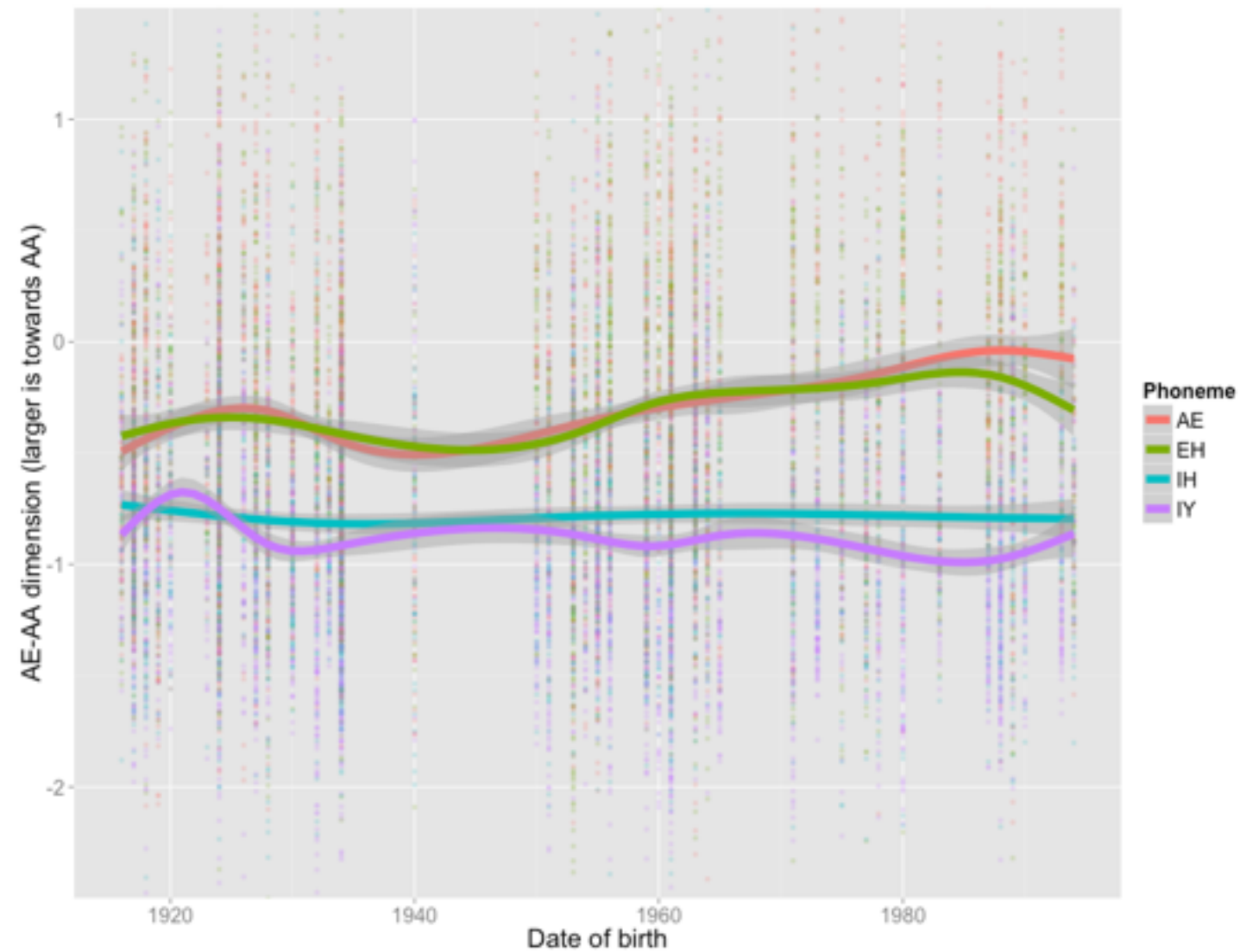
/uw/ fronting

AE-AA dimension



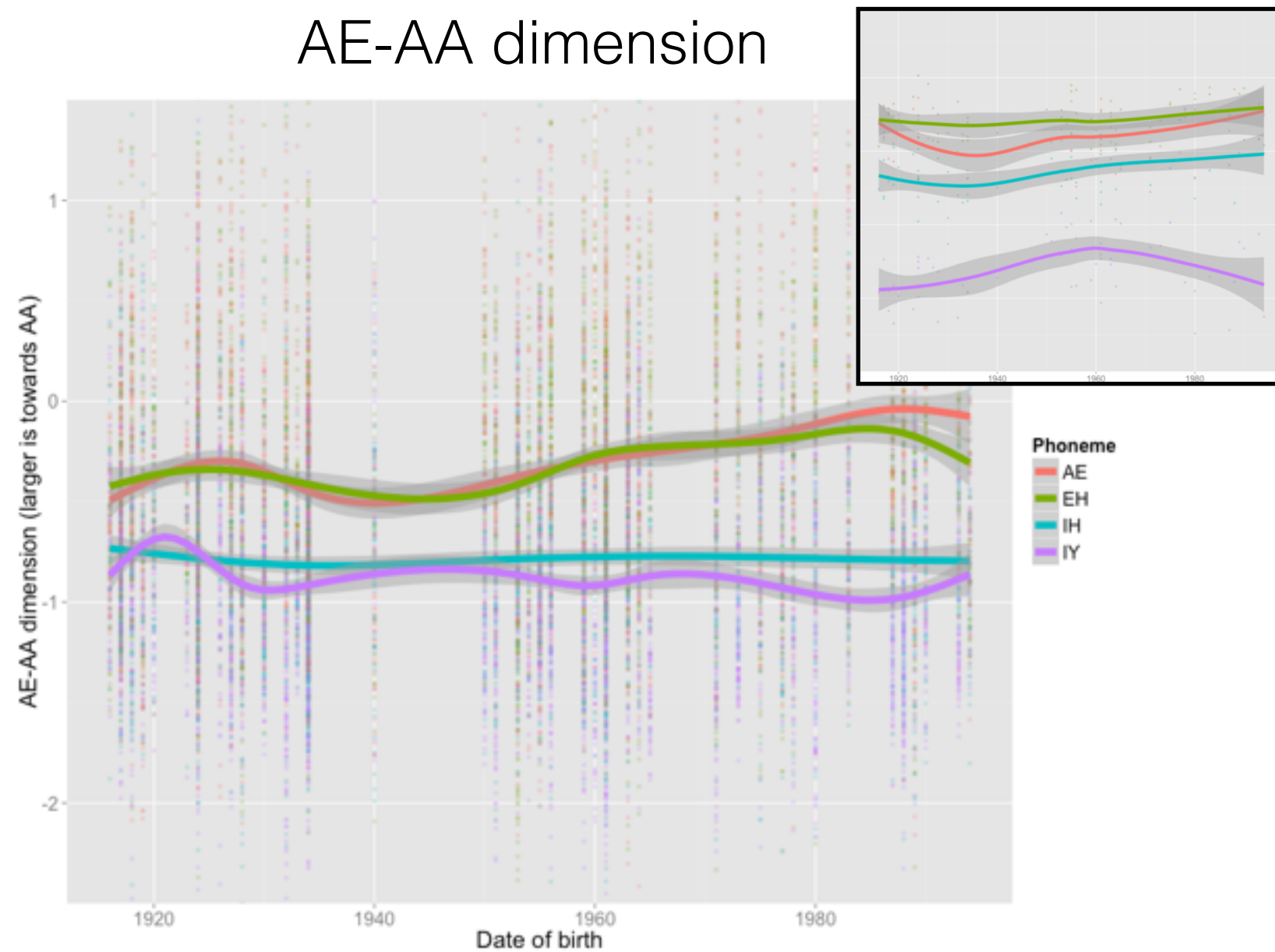
Canadian shift

AE-AA dimension



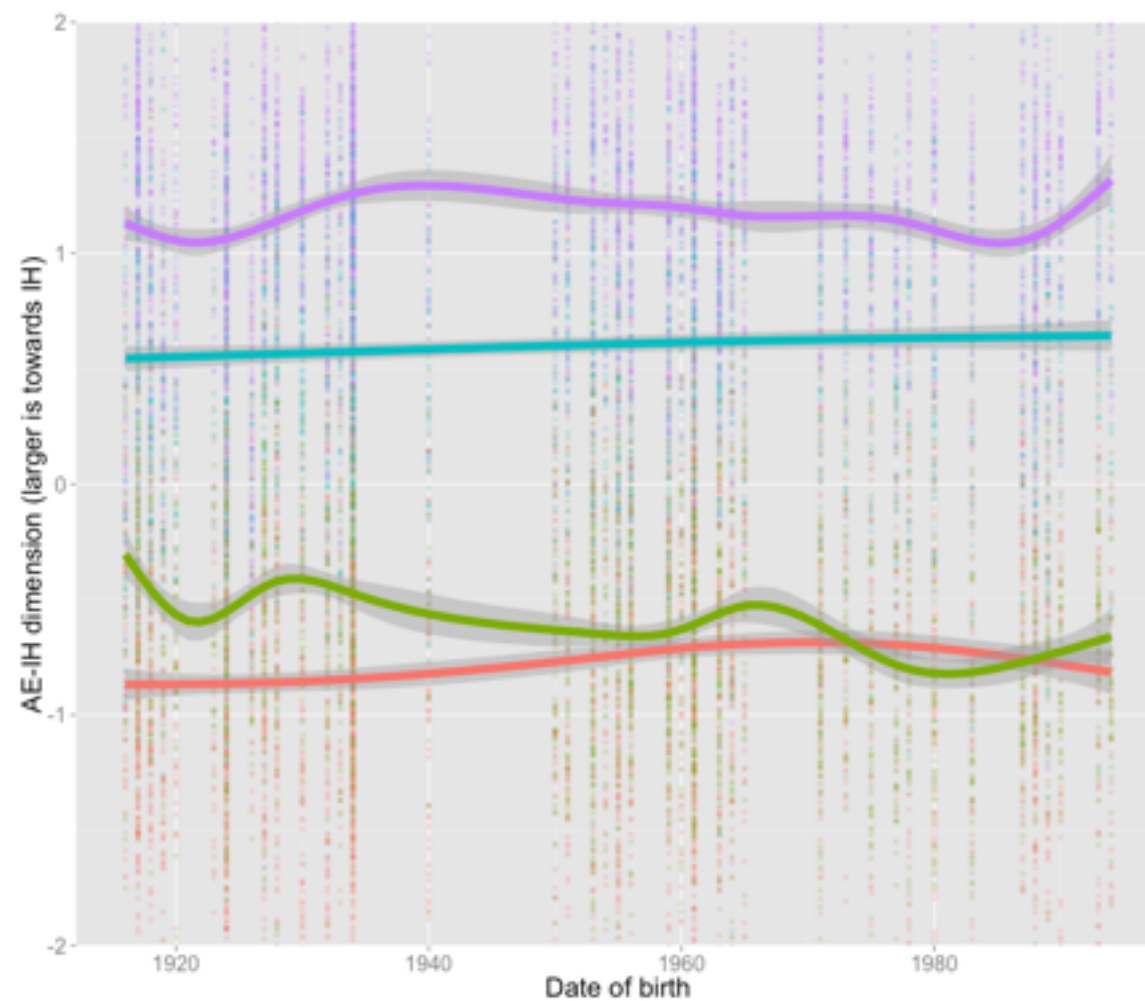
Canadian shift

AE-AA dimension

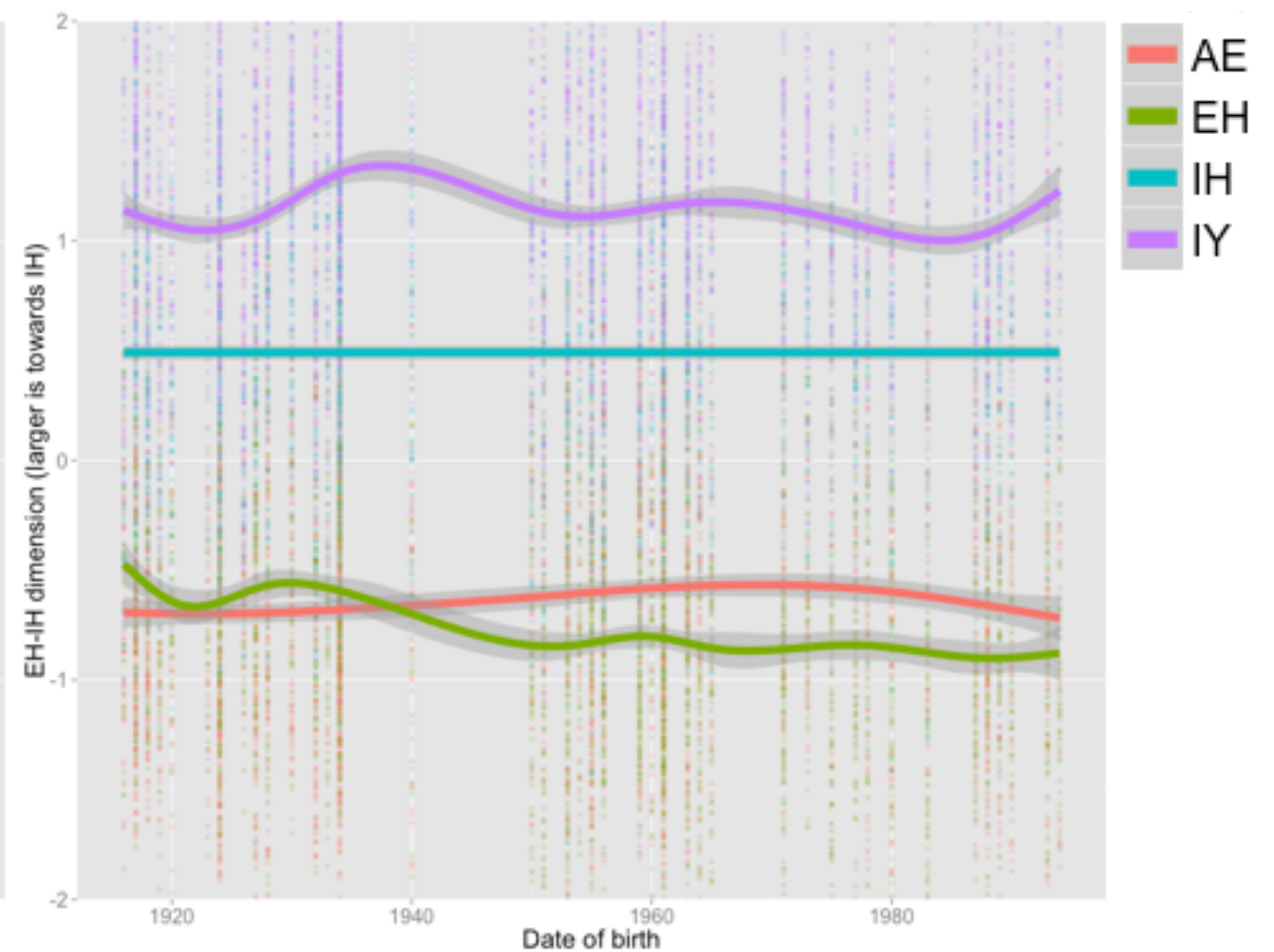


Canadian shift

AE-IH (“height”)

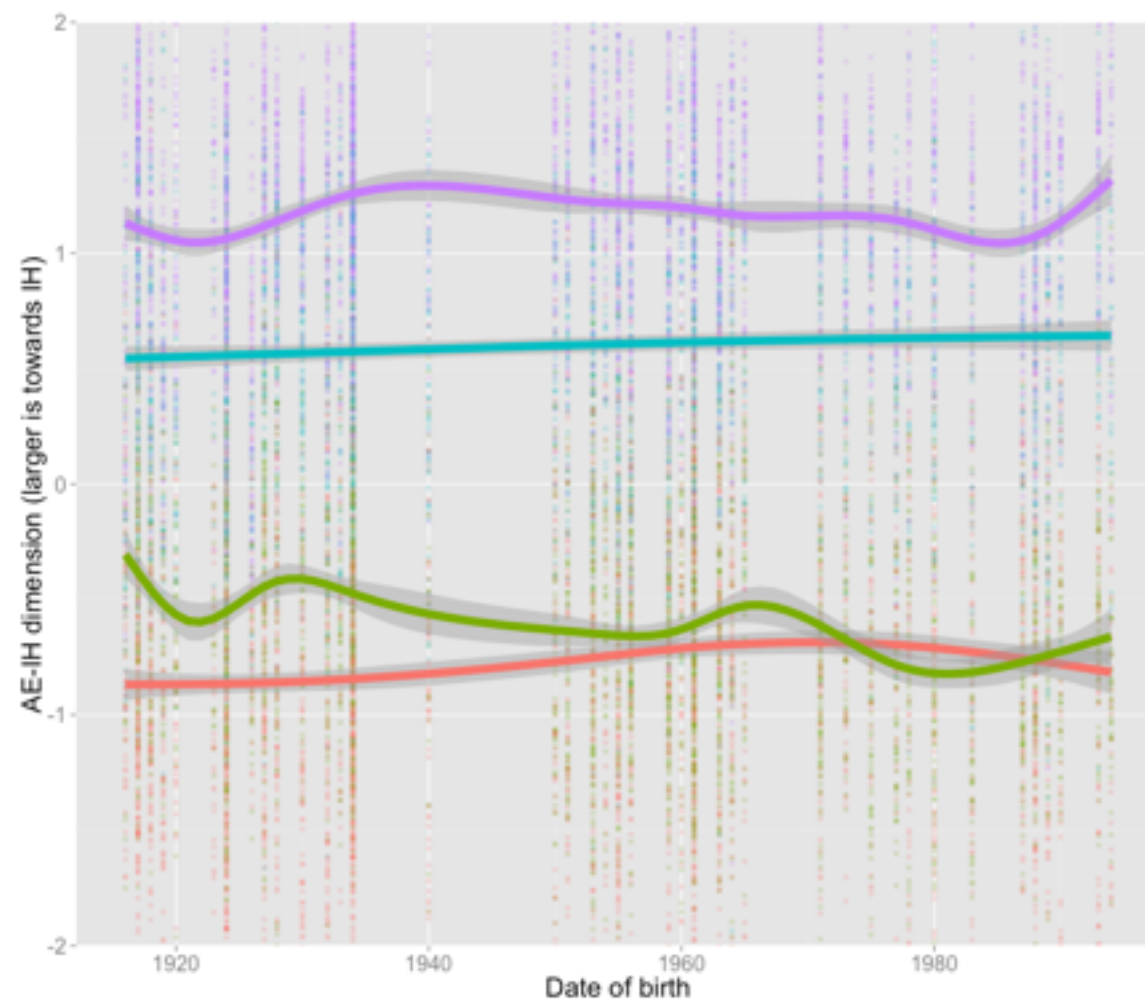


EH-IH (“height”)

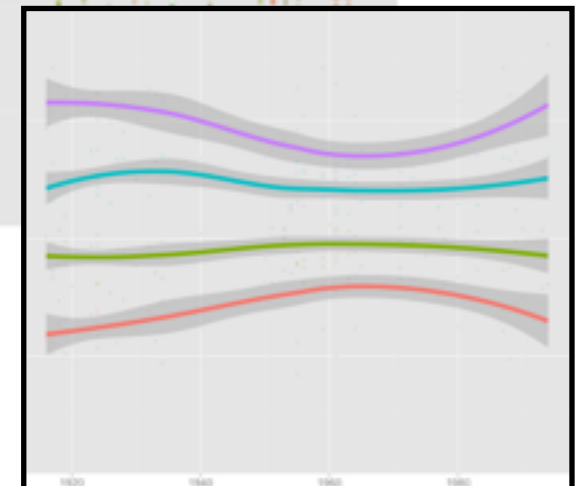
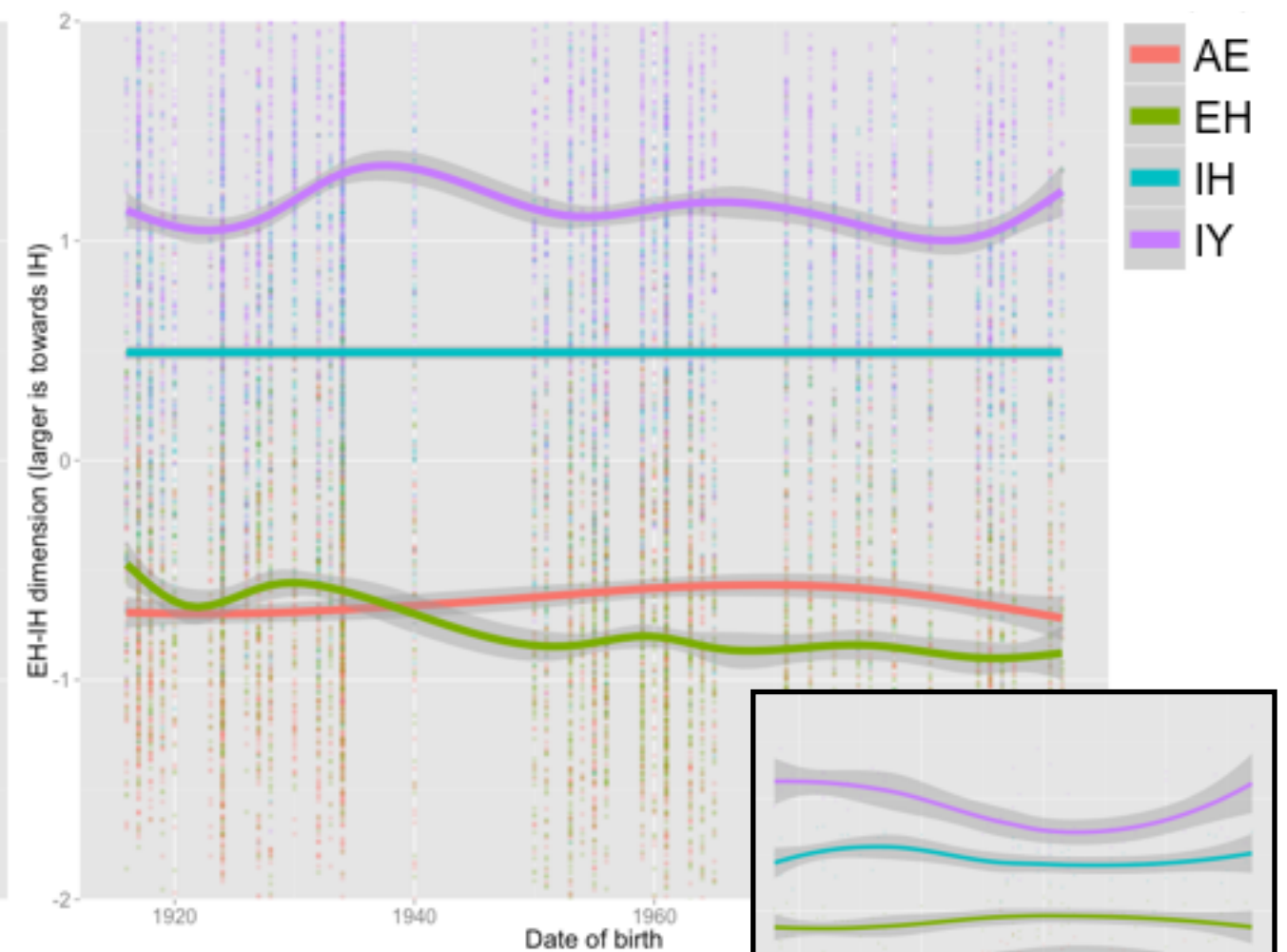


Canadian shift

AE-IH (“height”)

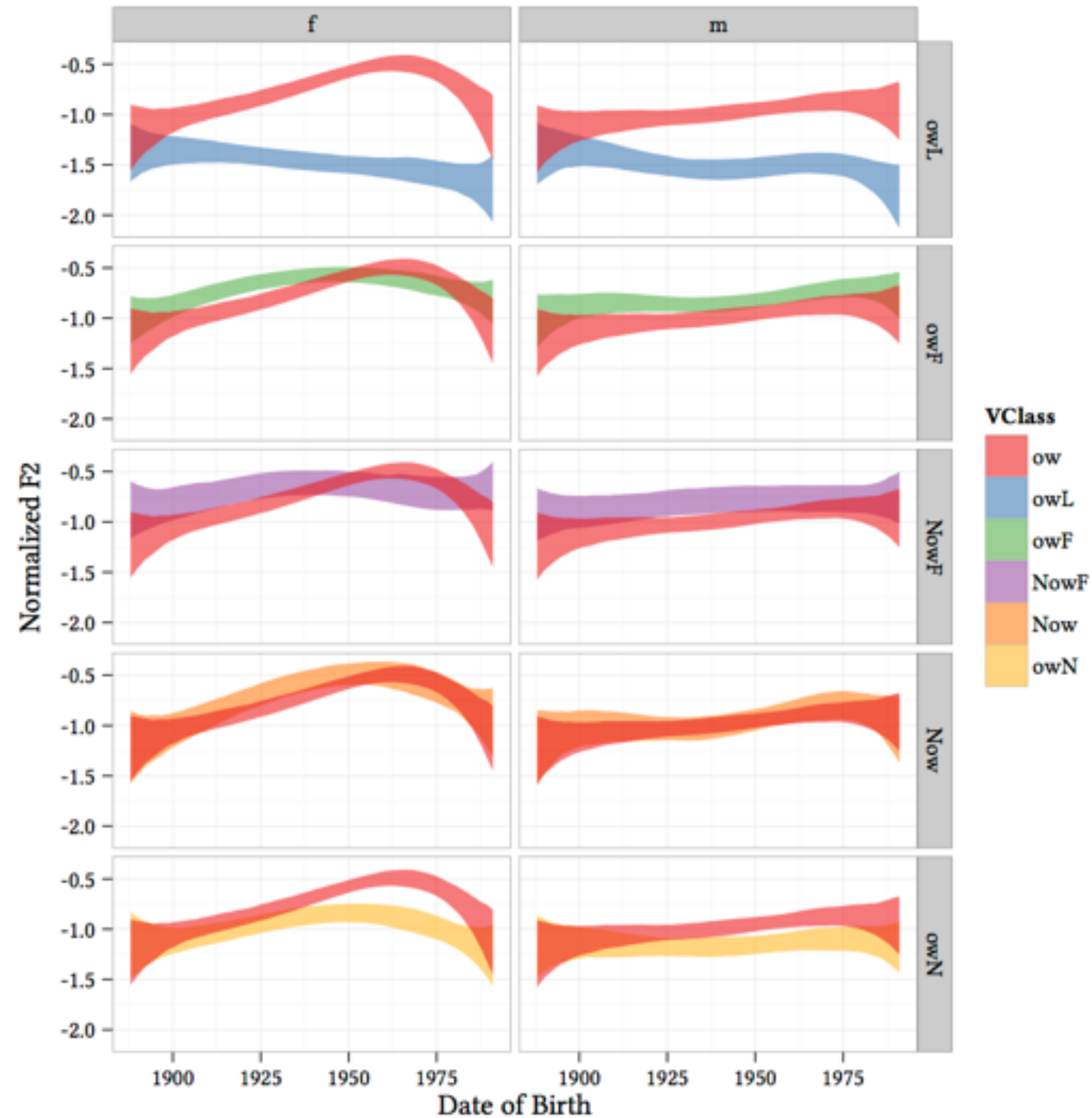


EH-IH (“height”)



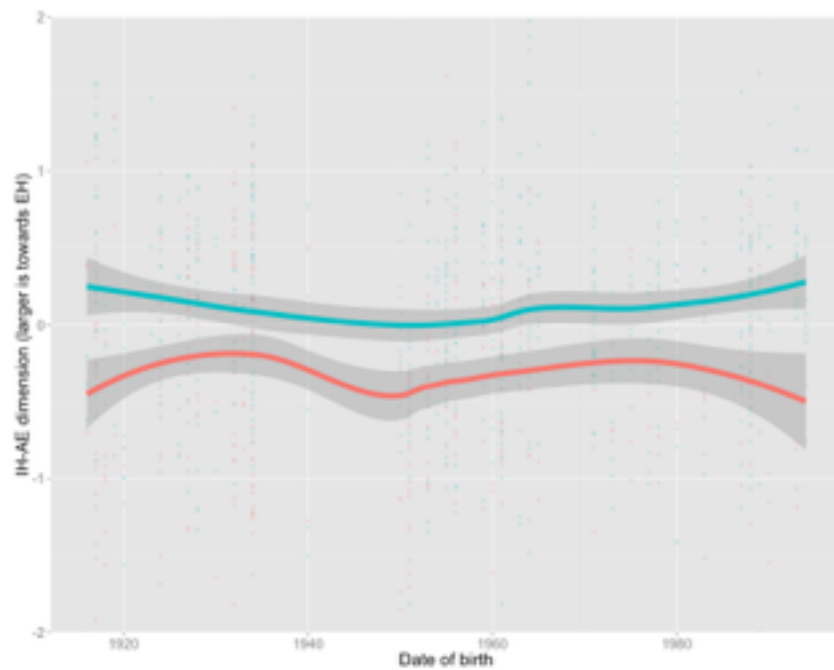
More interesting
examples

Diverging allophones

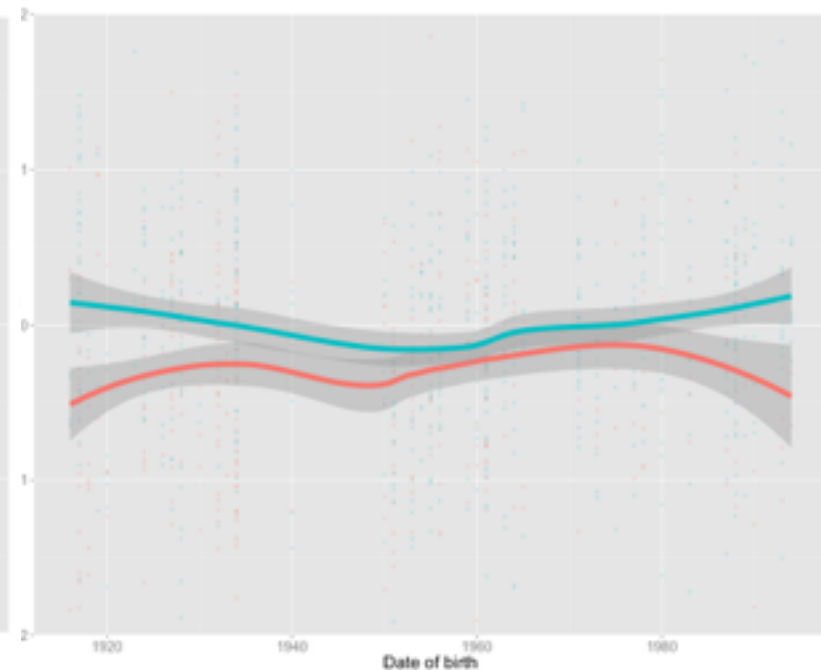


Ambiguity

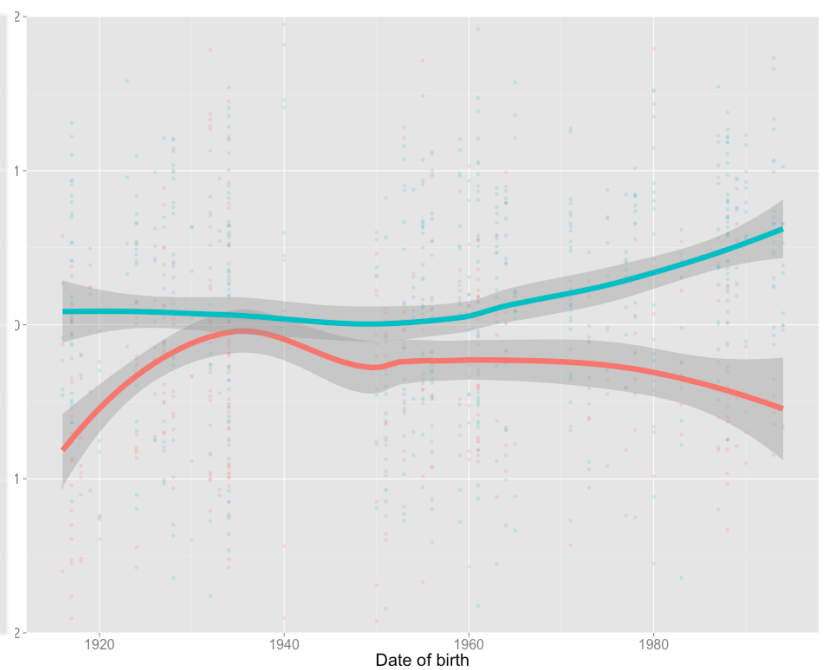
AE-IH (“height”)



EH-IH (“height”)



AA-AH (“centralization”)



Phoneme

