



Large Scale Sampling of Spoken Language Across Space and Time with the Penn Sociolinguistic Archive

Bill Labov
University of Pennsylvania

LDC Workshop on Crowdsourcing
and Novel Incentives in Language
Resource Creation Oct 3, 2016

BACK

NOTE
PAD

NAVIG

HELP

MENU

GLOSS

EXIT

The reliability of monolingual eliciting and the efficiency and dispatch of bilingual elicitation can be obtained by combining **magnetic recording** with bilingual field work.

--C.F. Voegelin and Z. Harris. 1951. Methods for determining intelligibility among dialects of natural languages., *Proceedings of the American Philosophical Society* 95:3, p. 322.



NAGRA III Tape Recorder

- ABOUT
- MEMBERS
- COMMUNICATIONS
- LANGUAGE RESOURCES ▾
- Data ▾
- Obtaining Data
- Catalog
- By Year
- Top Ten Corpora
- Projects
- Search
- Memberships
- LDC Online
- Data Scholarships
- Tools >
- Papers >
- LR Wiki
- DATA MANAGEMENT
- COLLABORATIONS

[Home](#) > [Language Resources](#) > [Data](#)

SLX Corpus of Classic Sociolinguistic Interviews

<i>Item Name:</i>	SLX Corpus of Classic Sociolinguistic Interviews
<i>Author(s):</i>	Stephanie Strassel, Jeffrey Conn, Suzanne Evans, Christopher Cieri, William Kazuaki Maeda
<i>LDC Catalog No.:</i>	LDC2003T15
<i>ISBN:</i>	1-58563-273-2
<i>ISLRN:</i>	034-299-958-433-1
<i>Release Date:</i>	November 25, 2003
<i>Member Year(s):</i>	2003
<i>DCMI Type(s):</i>	Sound
<i>Sample Type:</i>	pcm
<i>Sample Rate:</i>	22050
<i>Data Source(s):</i>	field recordings

LDC Corpus of Classic Sociolinguistic Interviews by Bill Labov

Speaker	Age	Speech Community	Occupation	Ethnicity	Education
Adolphus H.	81	Near Hillsboro, NC	Farmer	African American	Very little
Bobbie A.	22	Ayr, Scotland	Saw Doctor	Scottish/Italian	Some technical college
Henry G.	60	E. Atlanta, GA (DeKalb Co.)	Railroad foreman	European American	High school graduate
Jerry T.	19	Near Leakey, Texas	Gas station attendant	European American	Some high school
Joe D. (interviewed with Eddie M.)	21	Liverpool, England	Docker	English	Some high school
Eddie M. (Interviewed with Joe D.)	19	Liverpool, England	Docker	English	Some high school
Kathy D.	15	Rochester, NY	Student	European American	In 11th grade
Louise A.	53	Knoxville, TN	Mother	European American	Unknown
Rose B.	43	New York, NY (Lower East Side)	Factory seamstress	Italian American	Sixth Grade

Penn Sociolinguistic Archive

N=5,391

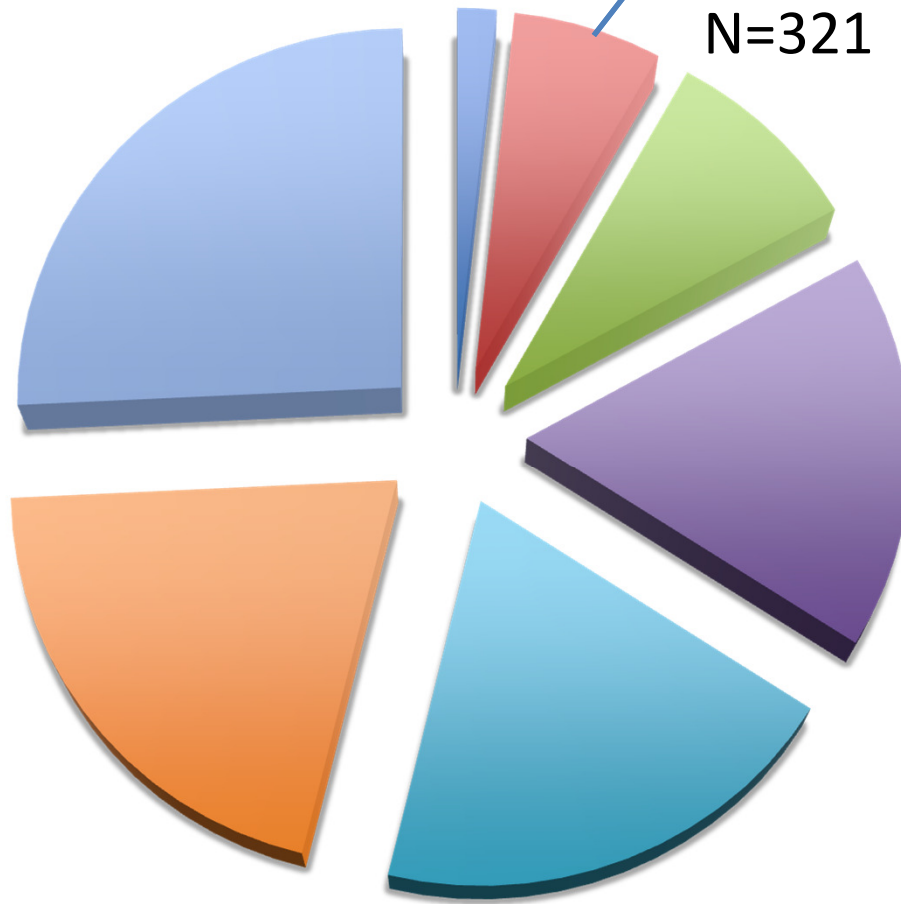
1961-2014

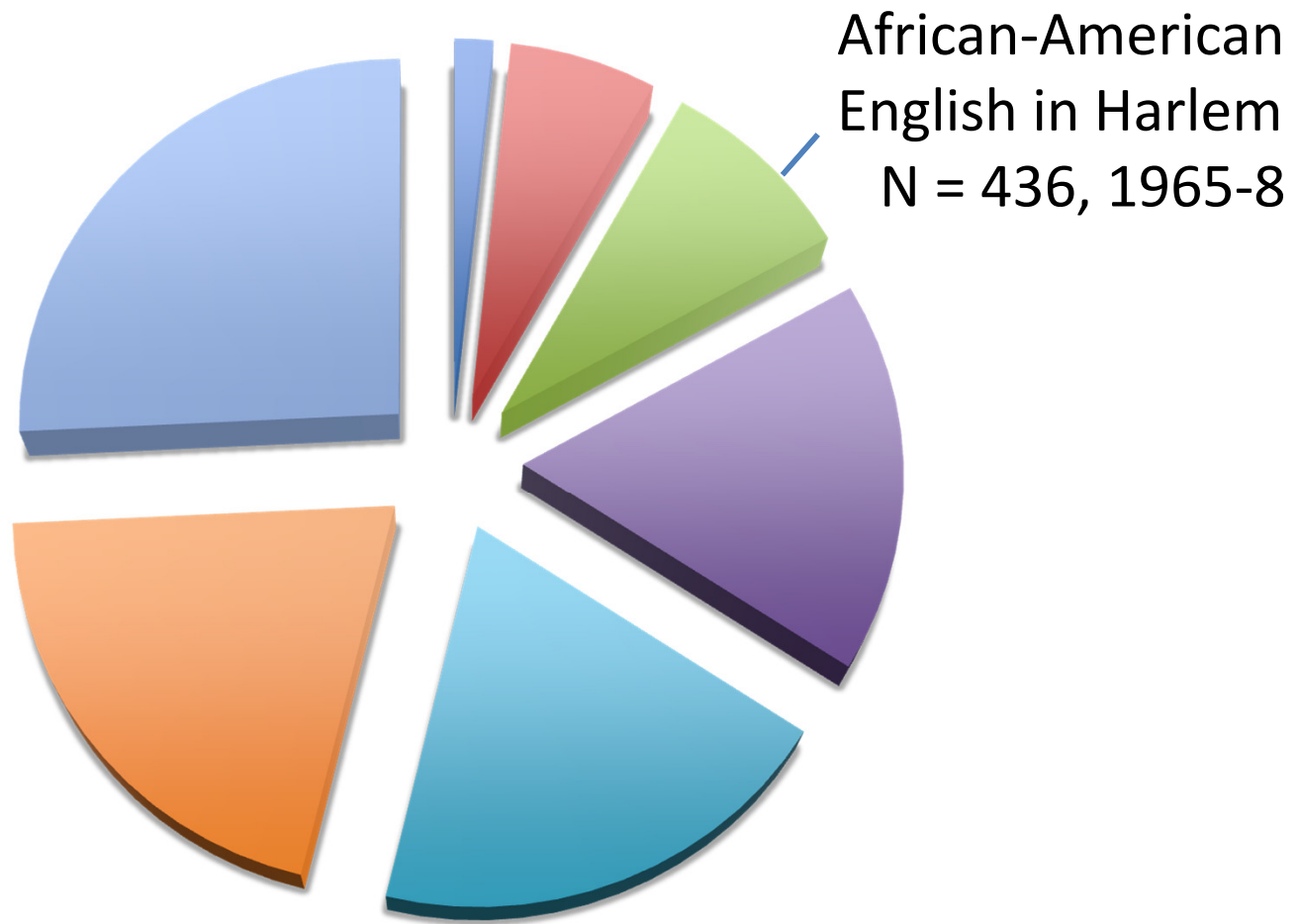


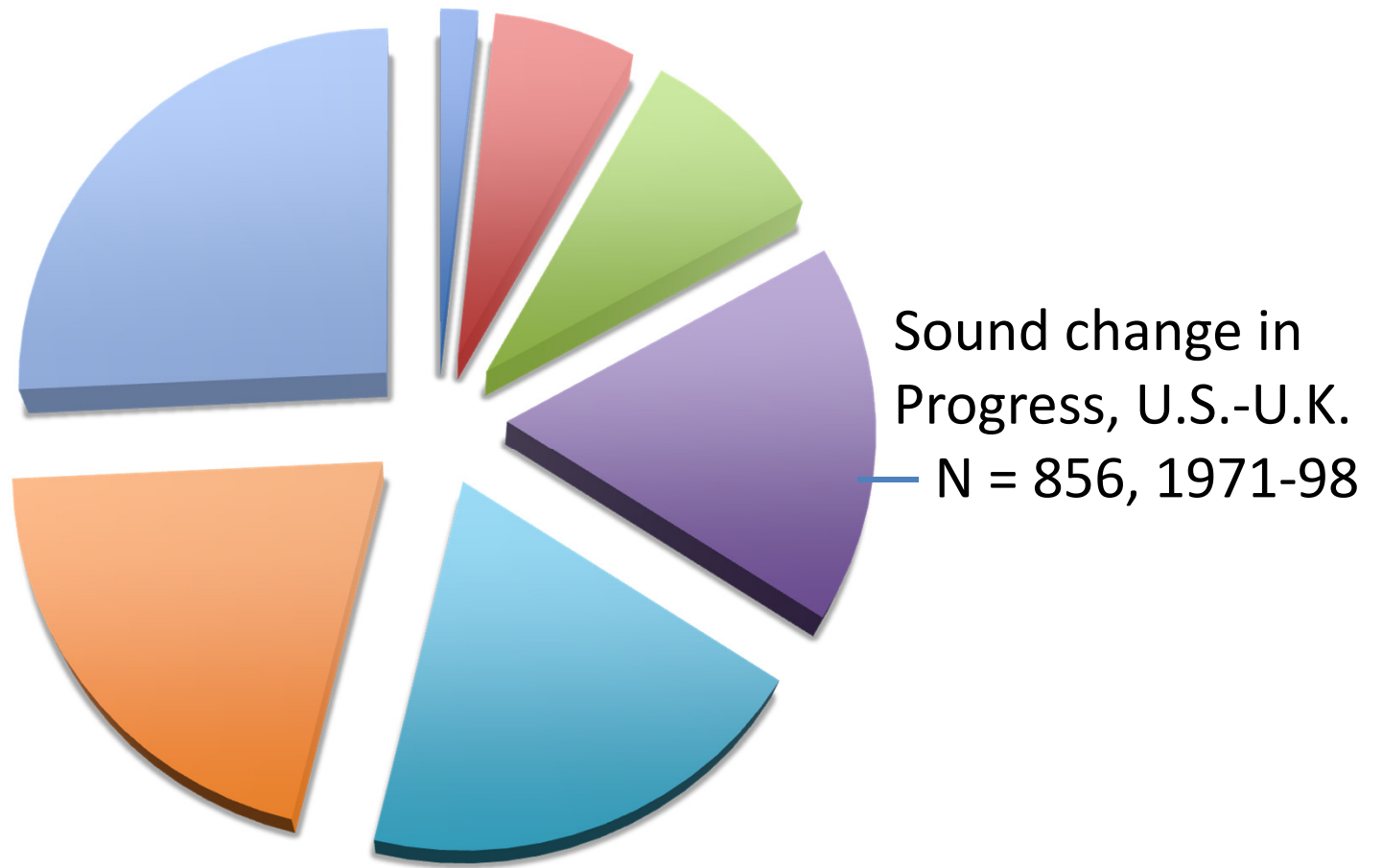
Social Motivation of a Sound Change Martha's Vineyard, N=84 1961-2



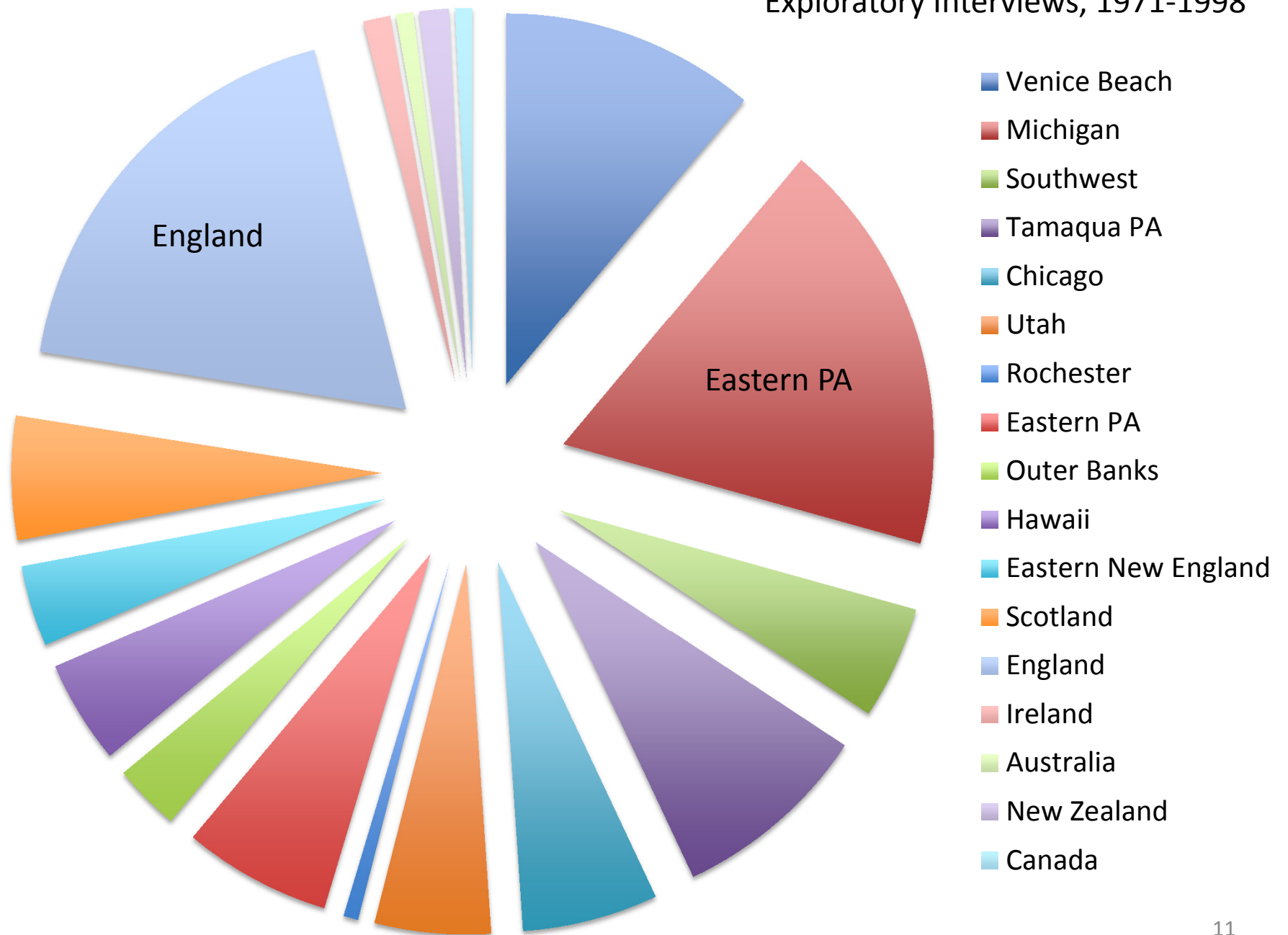
The Social Stratification of
English in New York City
N=321 1962-4







Exploratory Interviews, 1971-1998

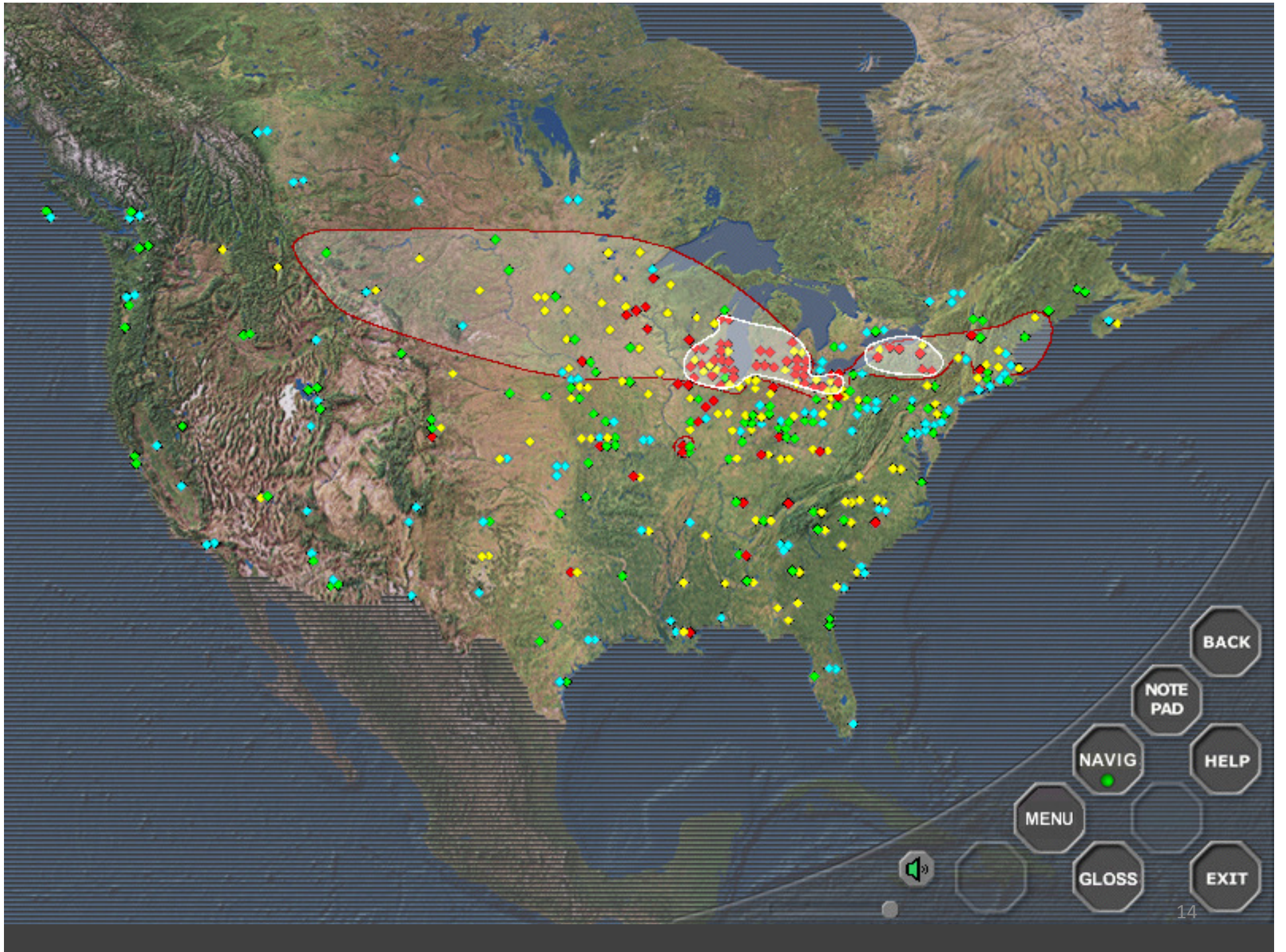




Linguistic change and
Variation in Philadelphia
N=976, 1972-79

Atlas of North
American English
N=1017, 1992-1999

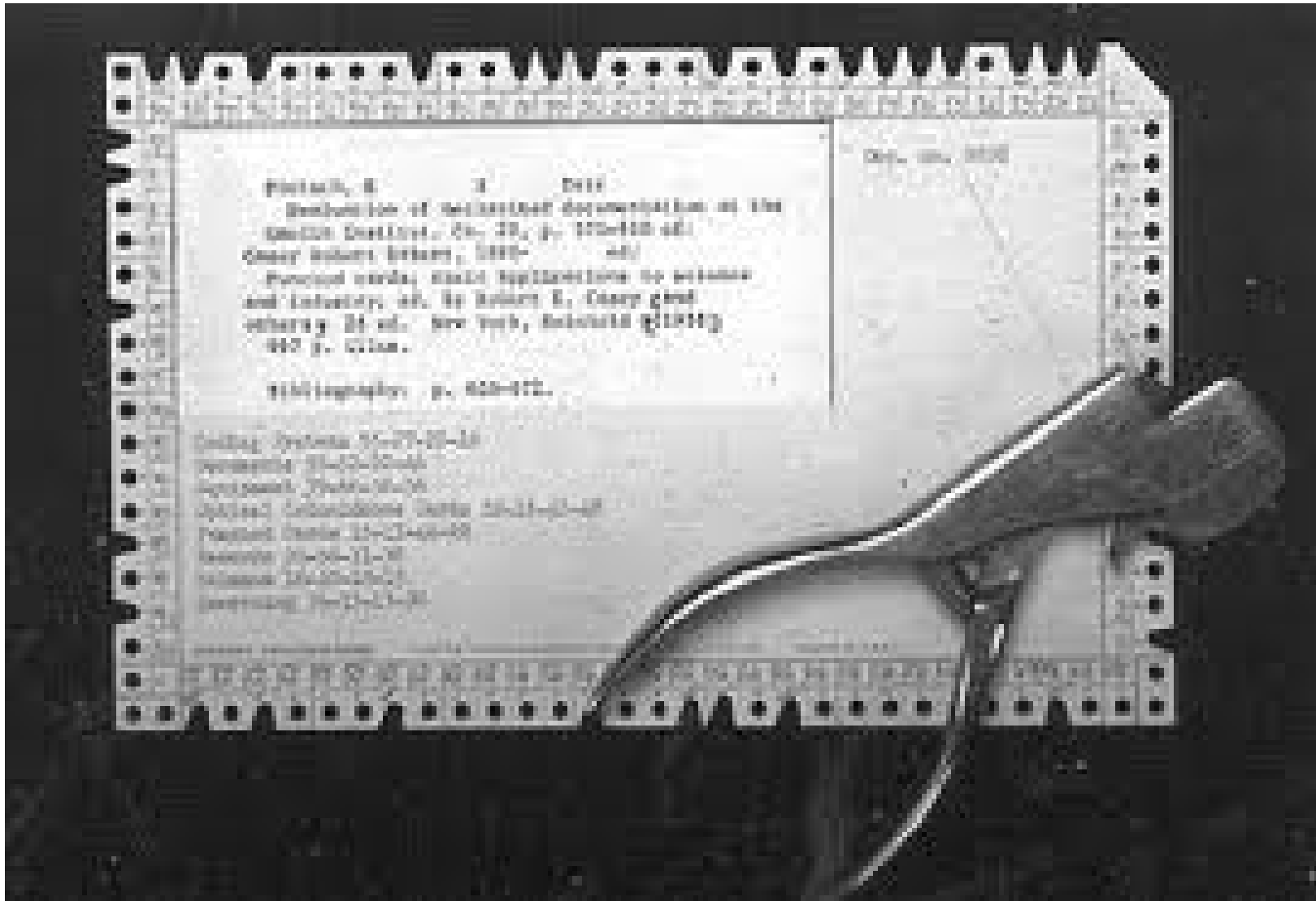




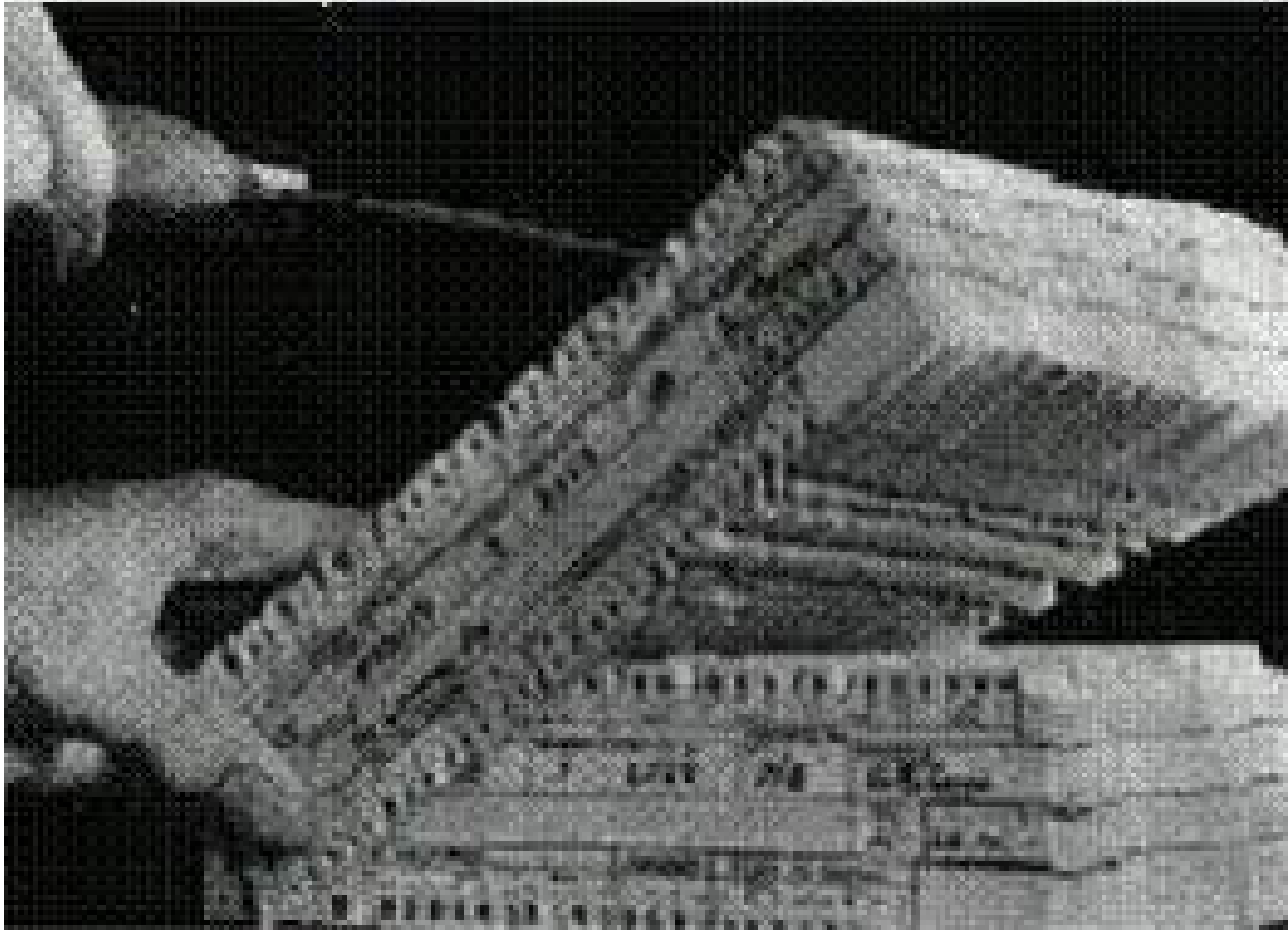
Philadelphia Neighborhood Corpus
N=1280, 1973-2012



McBee Keysort Cards, 1961-64



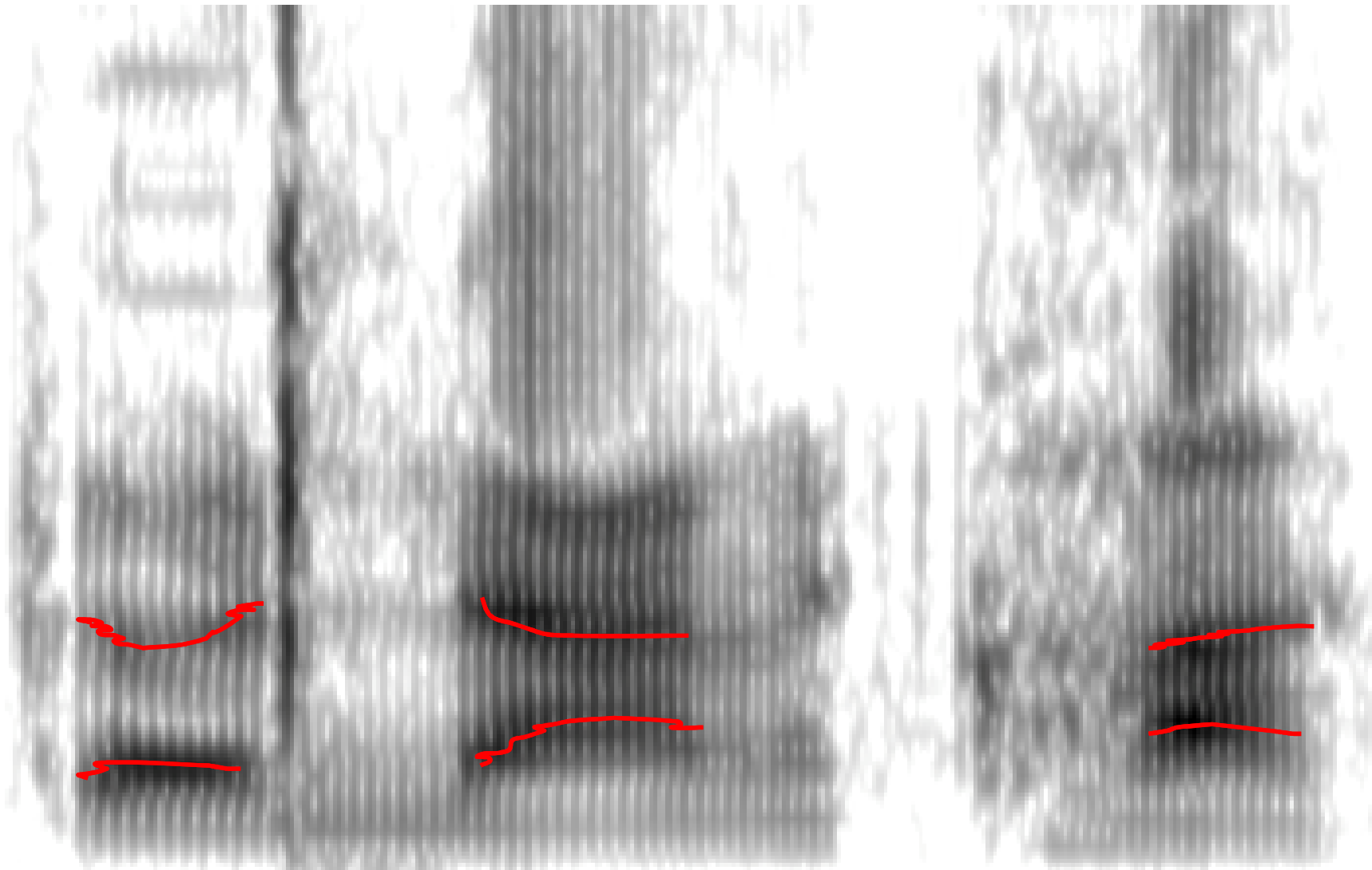
Sorting McBee Cards, 1961-64



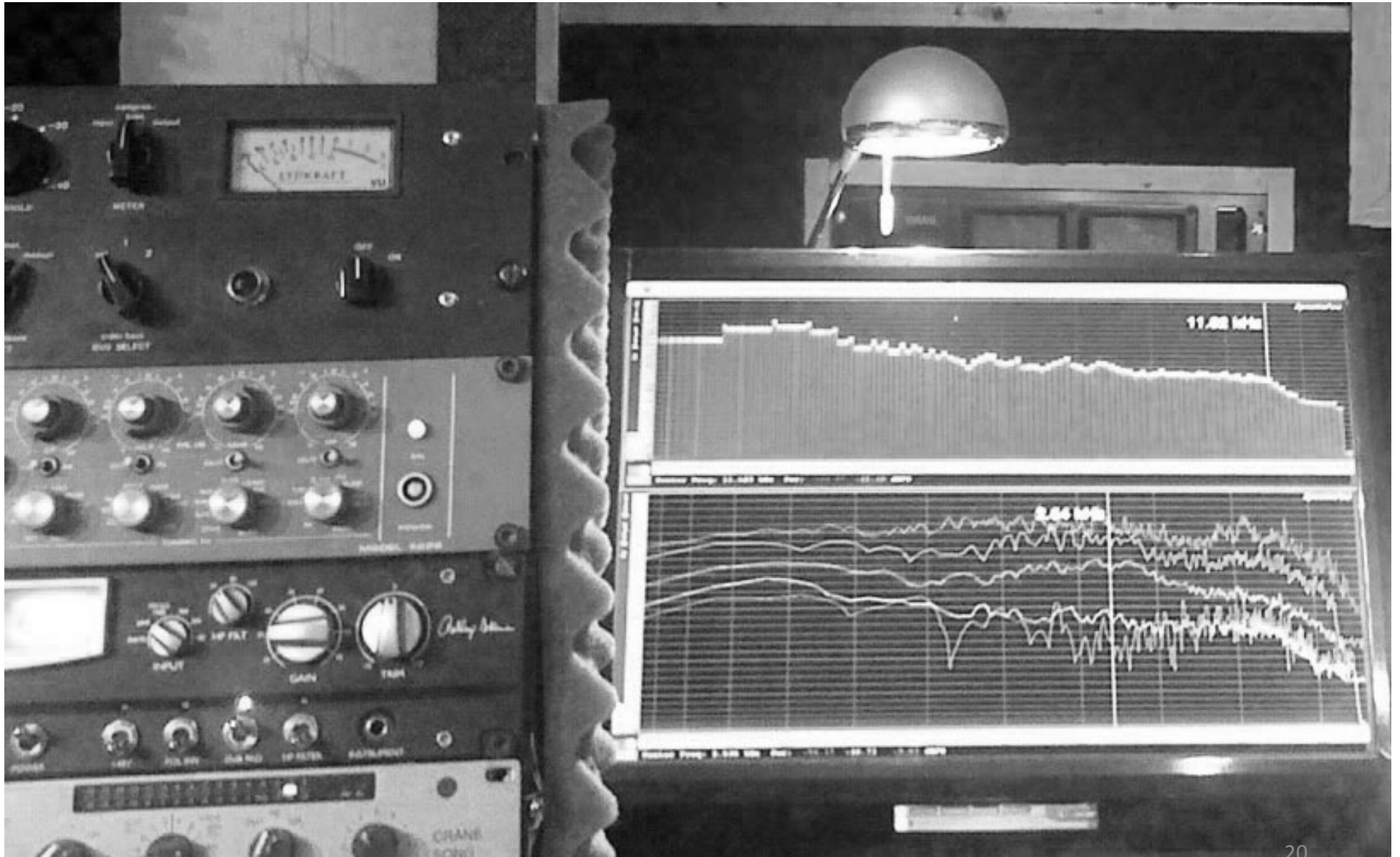
Kay Sonograph, 1969



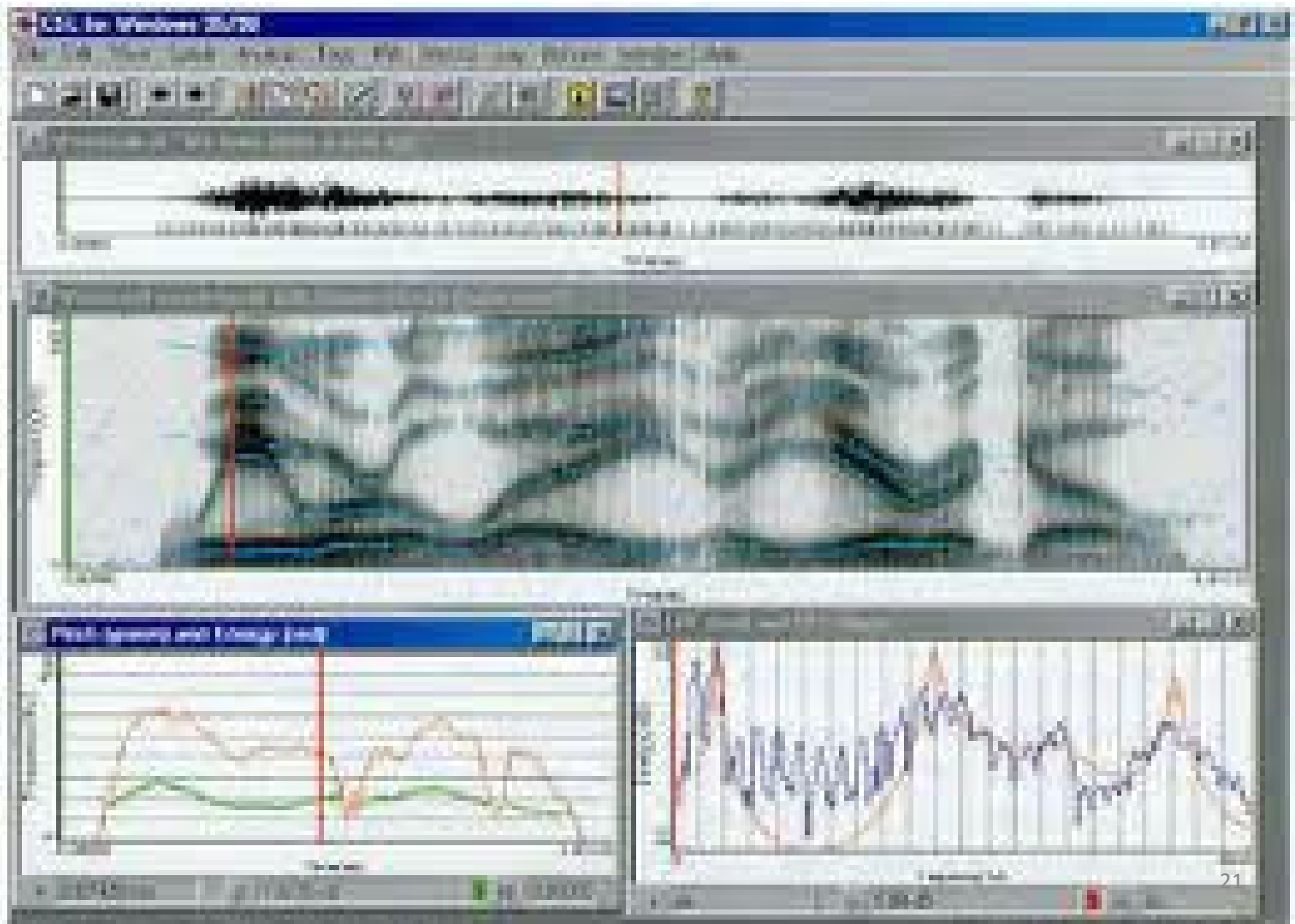
Drawing formant centers on sonograms



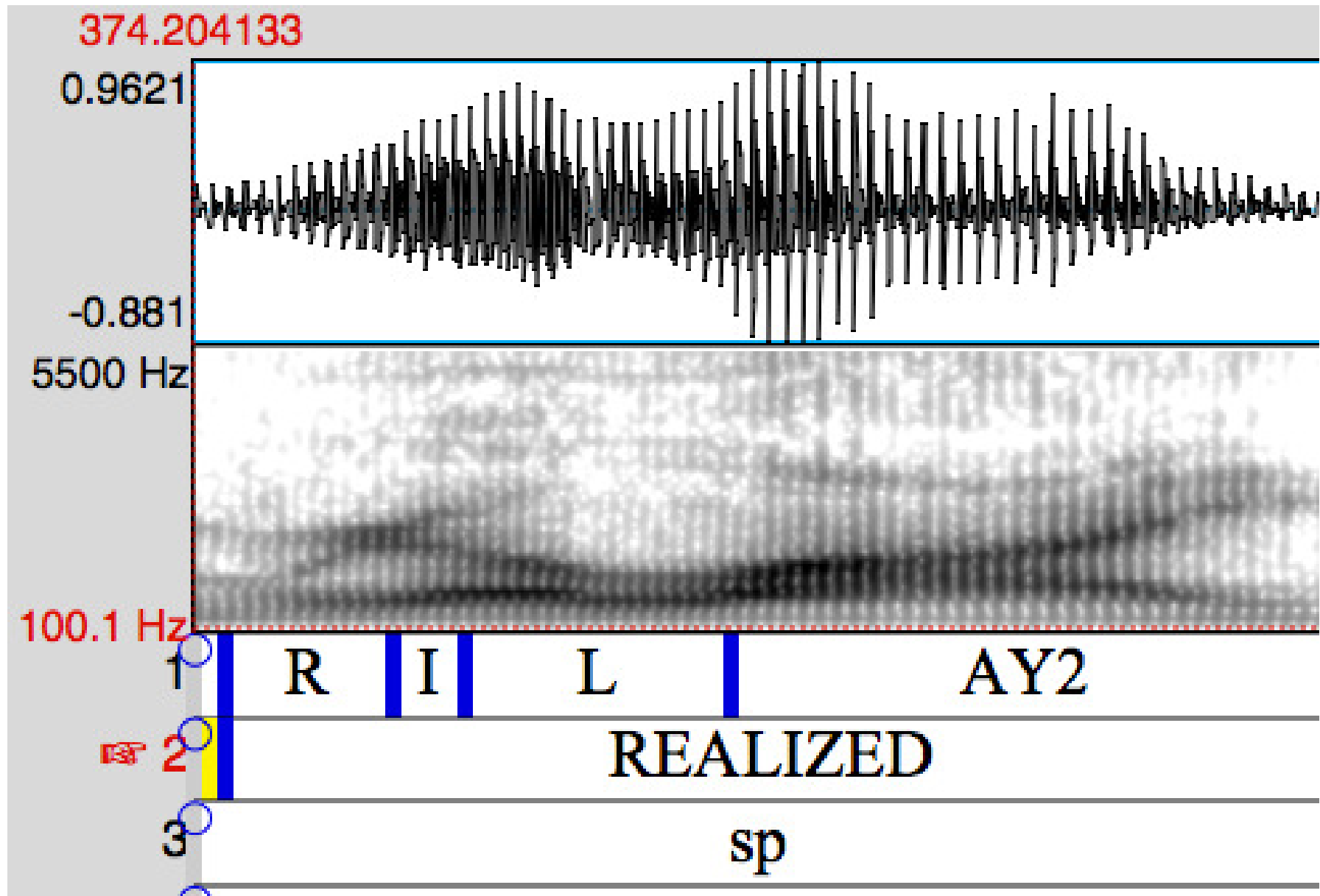
Hardware Spectrum Analyzer, 1975



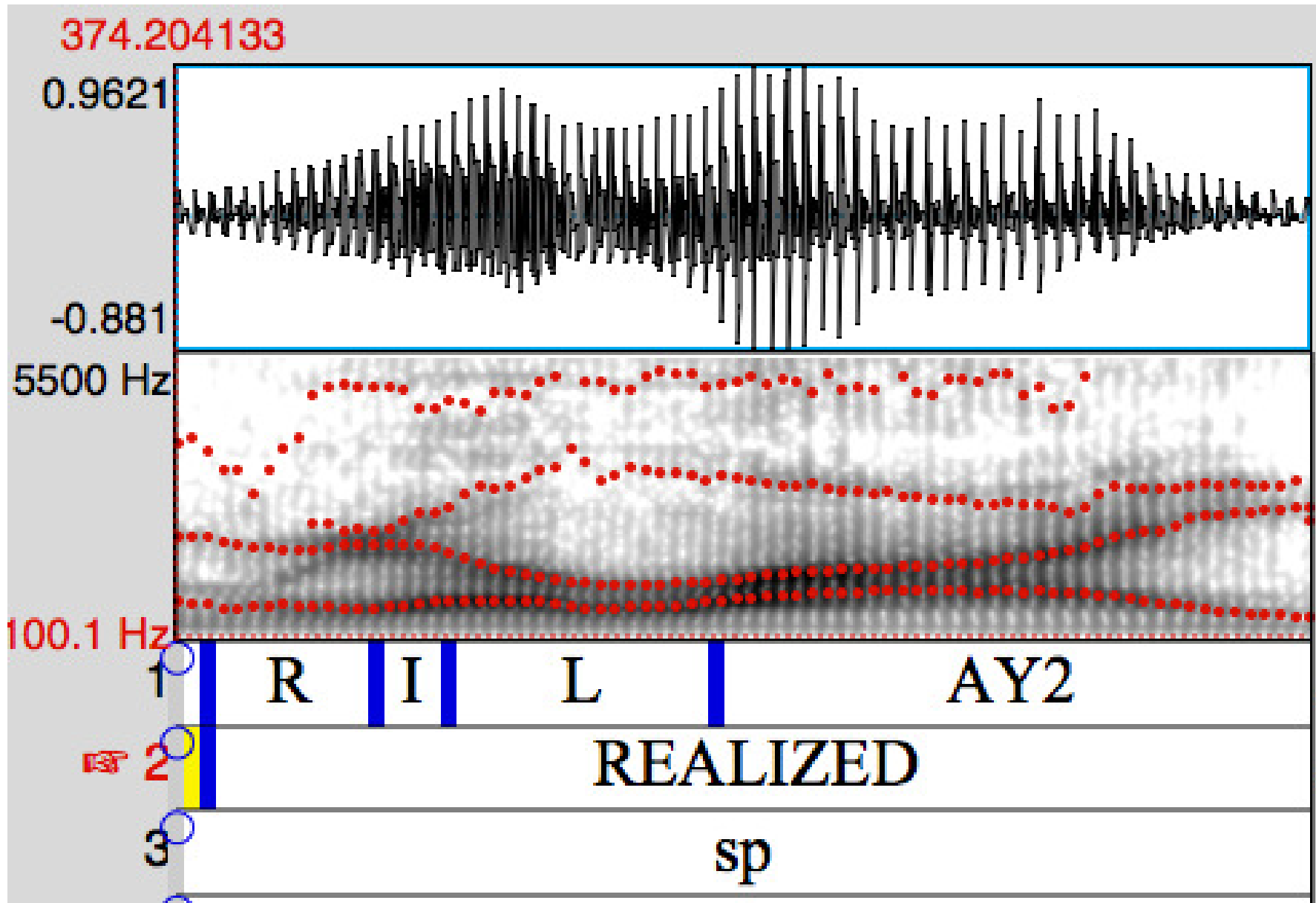
Kay Computer Speech Lab, 1995-2004



Kay wave form and spectrum



Linear predictive coding (B. S. Atal)



The FAVE web site fave.ling.upenn.edu

J. Yuan, M. Liberman, I. Rosenfelder, K. Evanini.



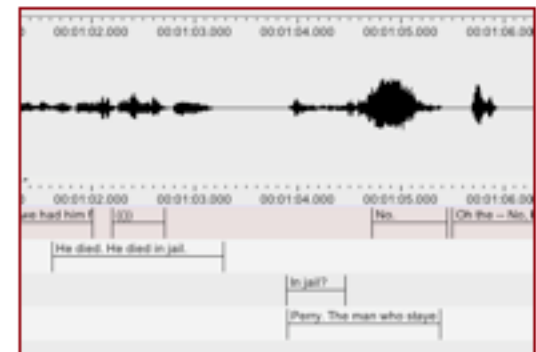
Forced Alignment & Vowel Extraction (FAVE)

An online suite for automatic vowel analysis

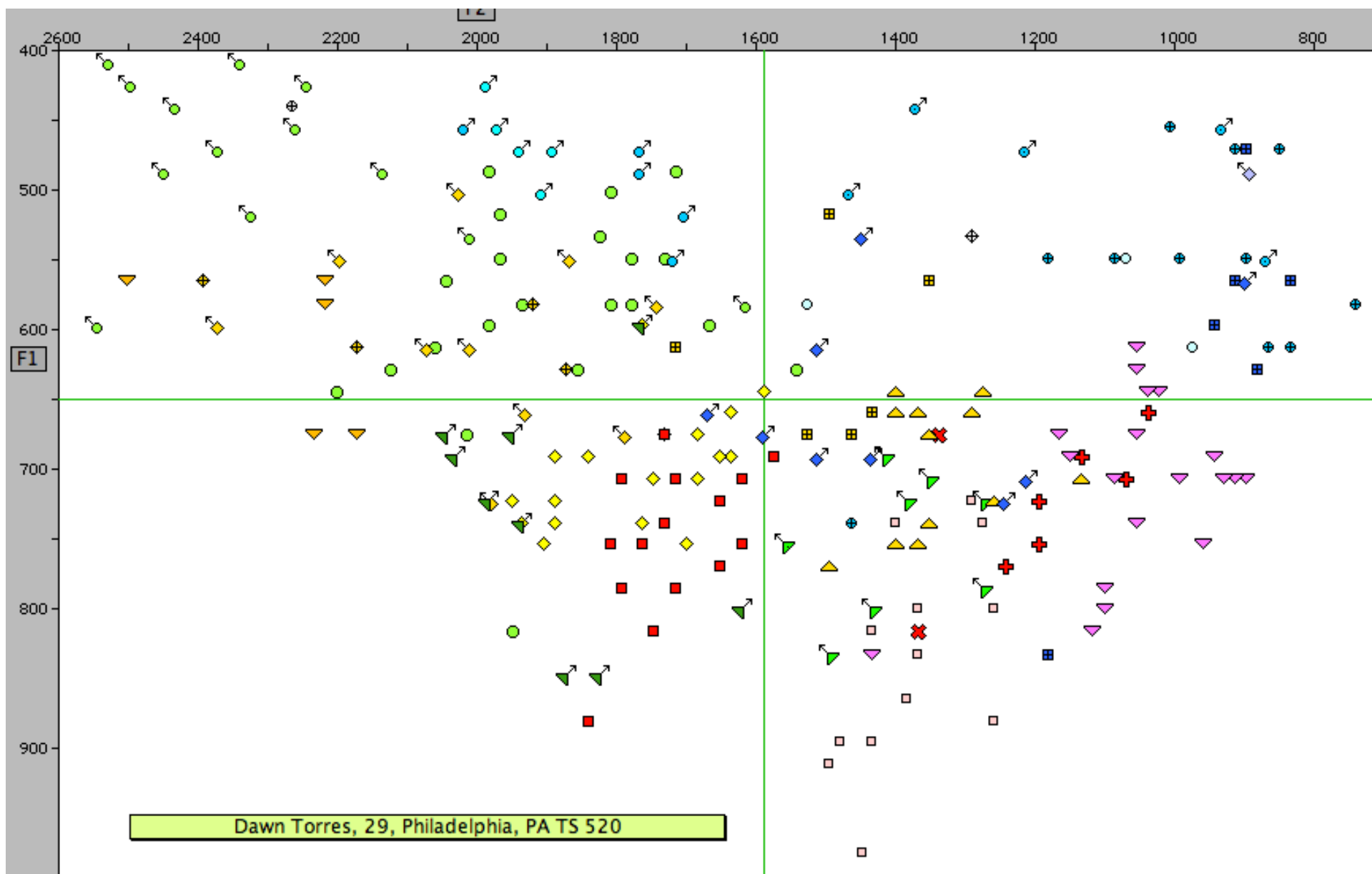
What is FAVE?

The FAVE (Forced Alignment & Vowel Extraction) program suite allows you to automatically align and extract large quantities of vowel formant measurements from sociolinguistic interviews or other bodies of orthographically transcribed data.

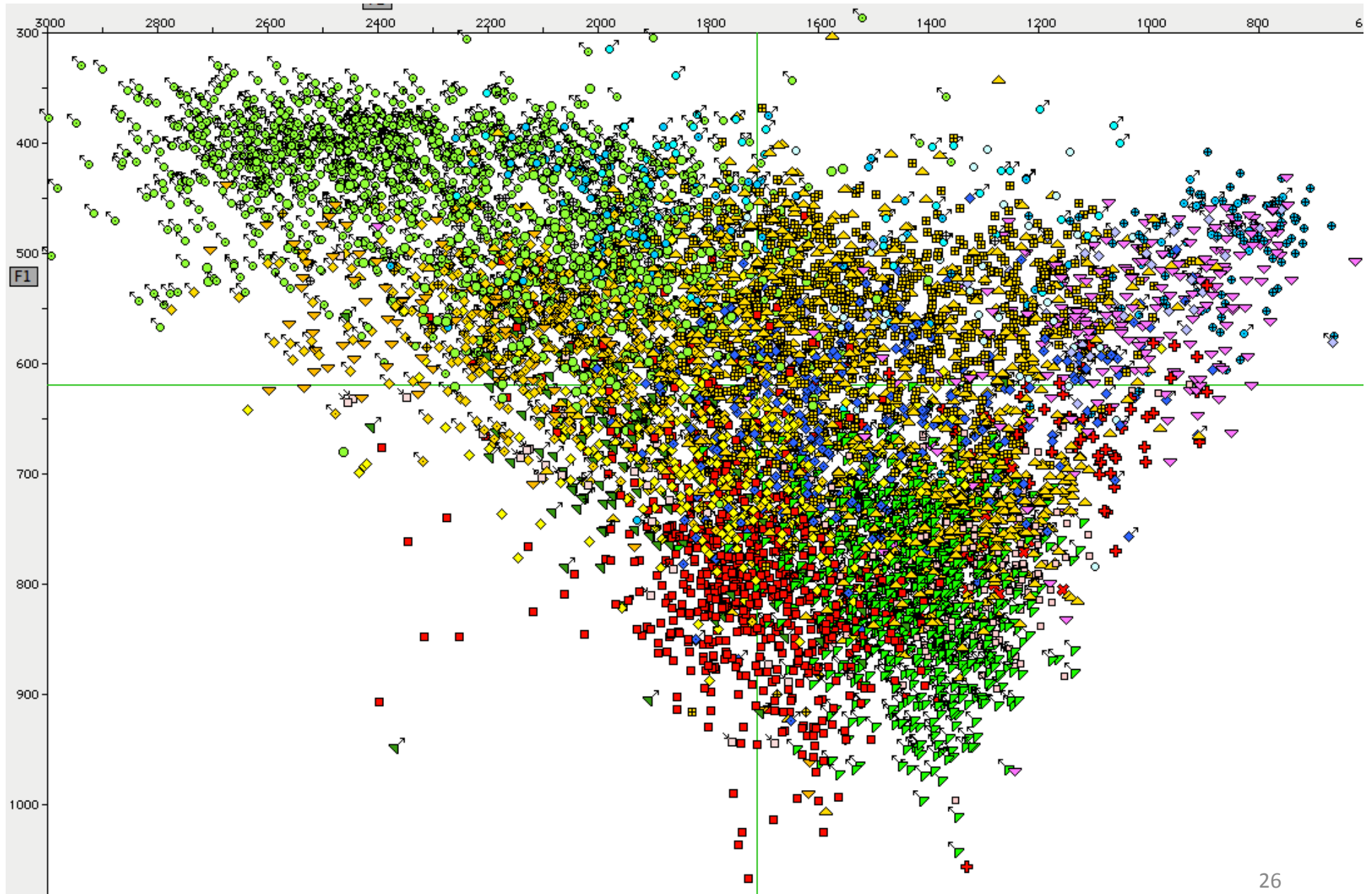
FAVE is a set of two programs: FAVE-align and FAVE-extract.



Key Speech Lab measurements of a Philadelphia ANAE subject, N= 214



FAVE analysis of Jean H., 60, Interviewed by J. Fruehwald in 2006 [N=7,101]



A natural misunderstanding (/ey/ -> /iy/)

Rosemarie D.: All right, come to dinner! [carrying out the food on a tray]

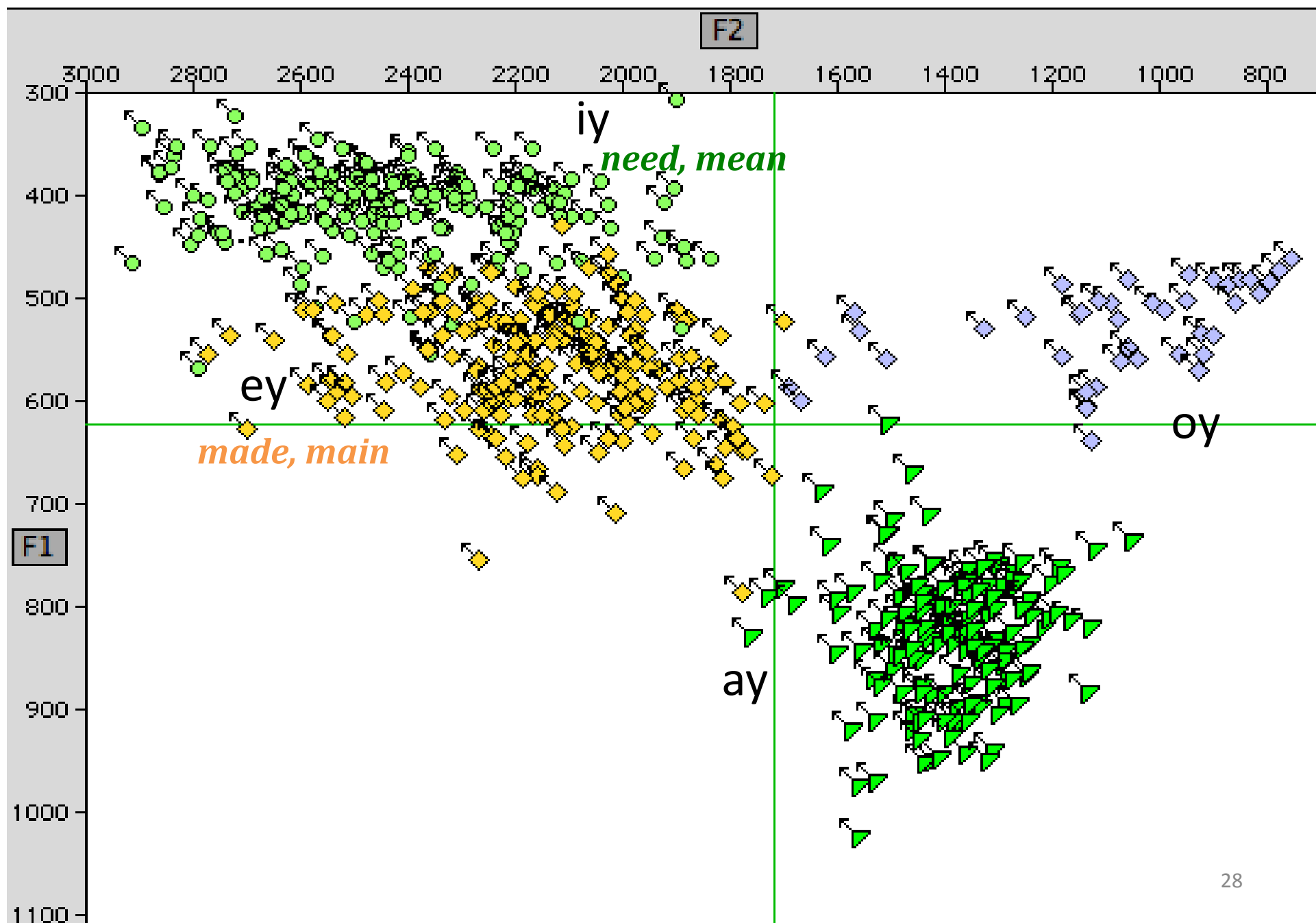
WL: You run a tight ship.

Tom D. [husband]: She makes us slave. >> “makes us leave”

Rosemarie: Why would I want you to leave?

Tom D.: One day, we'll explain it all to Rosemarie

Raising of /ey/ for Jean M., 60, Philadelphia [2006]



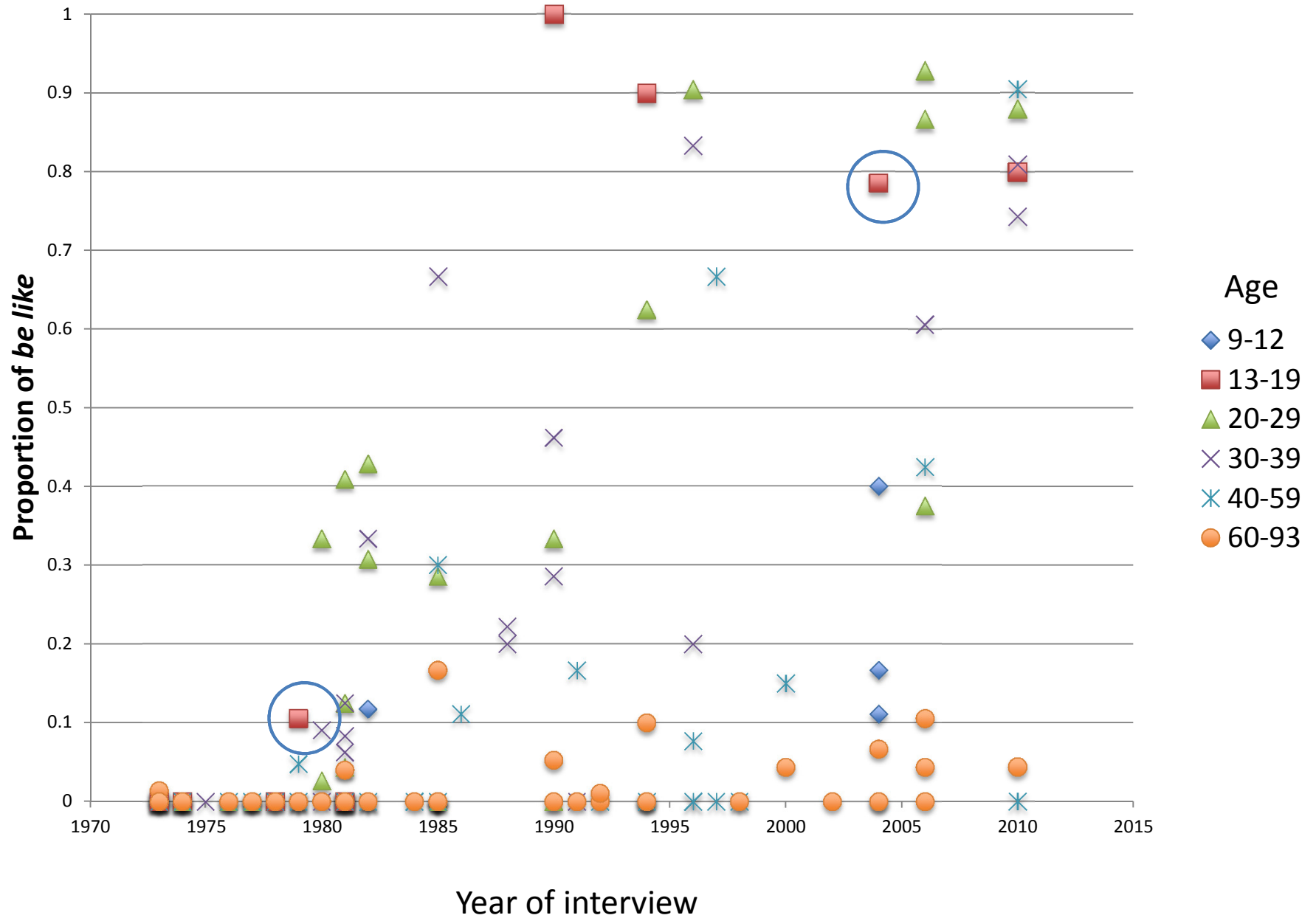
The raising of short-a in Philadelphia

- Anne Bower interviewing Celeste Sullivan on Daley St., 1973
[æ] → [i:ə] in *bad, dance*



But I used to go *dance* at the *canteen*. (The canteen?) Yeah they had a *canteen* that— it was all service men. And you couldn't get in if you didn't have a date. And you'd have to be a service *man*-- to take you in. But once you got in you could leave this guy, you know, and *dance* with everybody else. And that's what we used to do, *stand* on the corner. And wait for the fellas, and they'd-- I'd say to one of the sailors, "Are you going in there?" And he'd say "Yeah", "Would you take me in?" "Sure!" And they would take you--and nothing *bad* would happen in there. And— the Salvation Army *ran* a beautiful thing. It was only coffee and doughnuts, there was no drinking, and soda, and music. All night you'd *dance, dance, dance*.

Philadelphia development of new verb of quotation *be like* by year and age



--Brigitte Cash, PH94-1-6, 16, 1994



This number kept on coming up on his pager when he was in this-- he had to drop off papers, in Center City so I was like checking his beeper.

And **I'm like** "OK, it came up once." **I'm like**, "Alright."

Came up again two seconds later and **I'm like** . . . "Alright!!"

And it came up five times total within ten minutes.

So I wrote the number down and called it.

And **I said** you know, "How do you know [X]?"

And **she's like**, "Because . . uh . . he's my brother-in-law."

And **I was like**— and **she was like**, "How do you know him?"

You one of his sluts?"

And **I was like**, "Oh my god!" I hung up the phone, I called him.

He's like, "Why did you call?" whatever her name was.

And **I was like**, "I'm sorry."

He was like, "She told my wife."



PRAAT: Boersma and Weenink

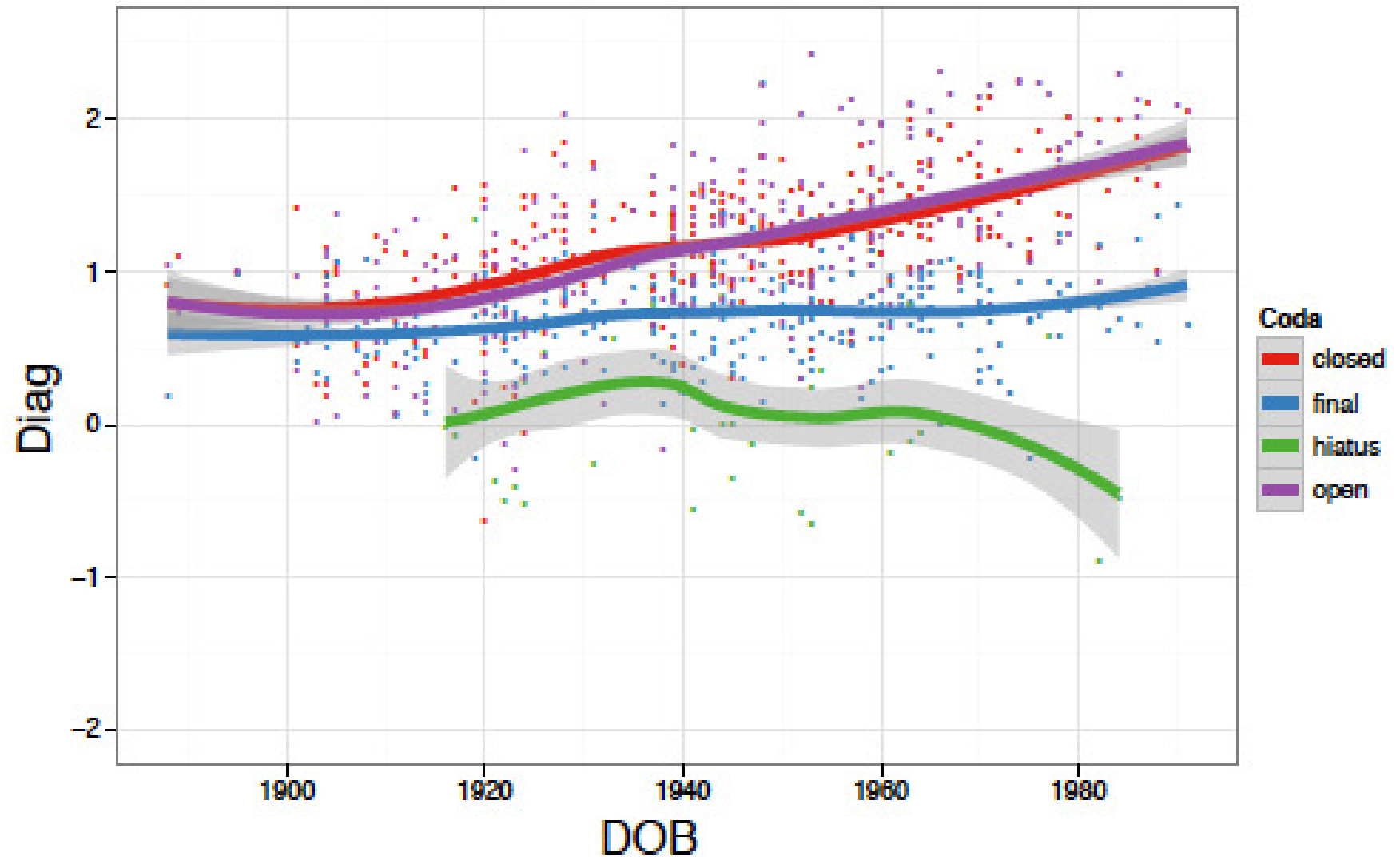
The screenshot displays the Praat software interface. On the left, the 'Objects' list shows: 2. LongSound PH06-2-3-Amanda, 3. TextGrid PH06-2-3-Amanda, and 8. Sound untitled. Below this are two portrait photos of men. The main window shows the 'Formant | Pulses' view for the file '3. TextGrid PH06-2-3-Amanda'. It features a menu bar with options like File, Edit, Query, View, Select, Interval, Boundary, Tier, Spectrum, Pitch, Intensity, and Fo. The main display area is divided into three horizontal panels: a waveform at the top with amplitude values from -0.8468 to 0.9578, a spectrogram in the middle with frequency markers at 0 Hz and 5500 Hz, and a text grid at the bottom. The text grid has four tiers. Tier 1 contains phonetic labels: AH1, HH, W, EH1, R, I, T, S. Tier 2 contains: UM, WHERE, IT'S. Tier 3 contains: sp. Tier 4 contains: sp. A small number '34' is visible in the bottom right corner of the text grid area.

1	AH1	HH	W	EH1	R	I	T	S
2	UM	WHERE			IT'S			
3						sp		
4						sp		

Defining the variable (eyC)

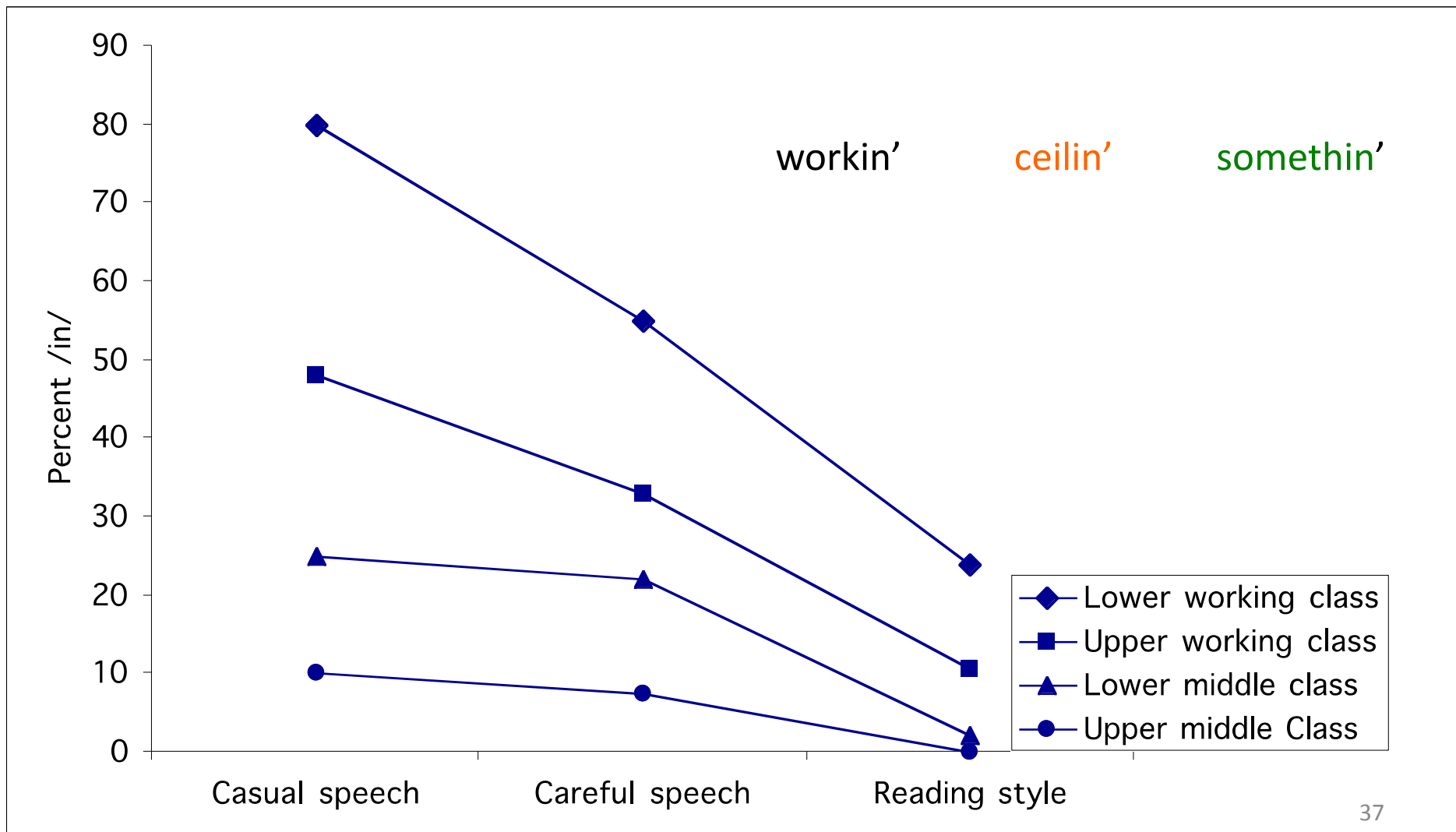
babe ←[?] *baby* →[?] *bay*
CVC CVCV CV

Syllable context for the raising of /eyC/: **open syllable** (*baby*), **closed syllable** (*babe*), **final** (*bay*) & **hiatus** (*mayor*)--from Fruehwald 2012



The discovery of orderly heterogeneity

Social and stylistic stratification of (ing) in the random sample of the Lower East Side of New York City adults [N=81/321]

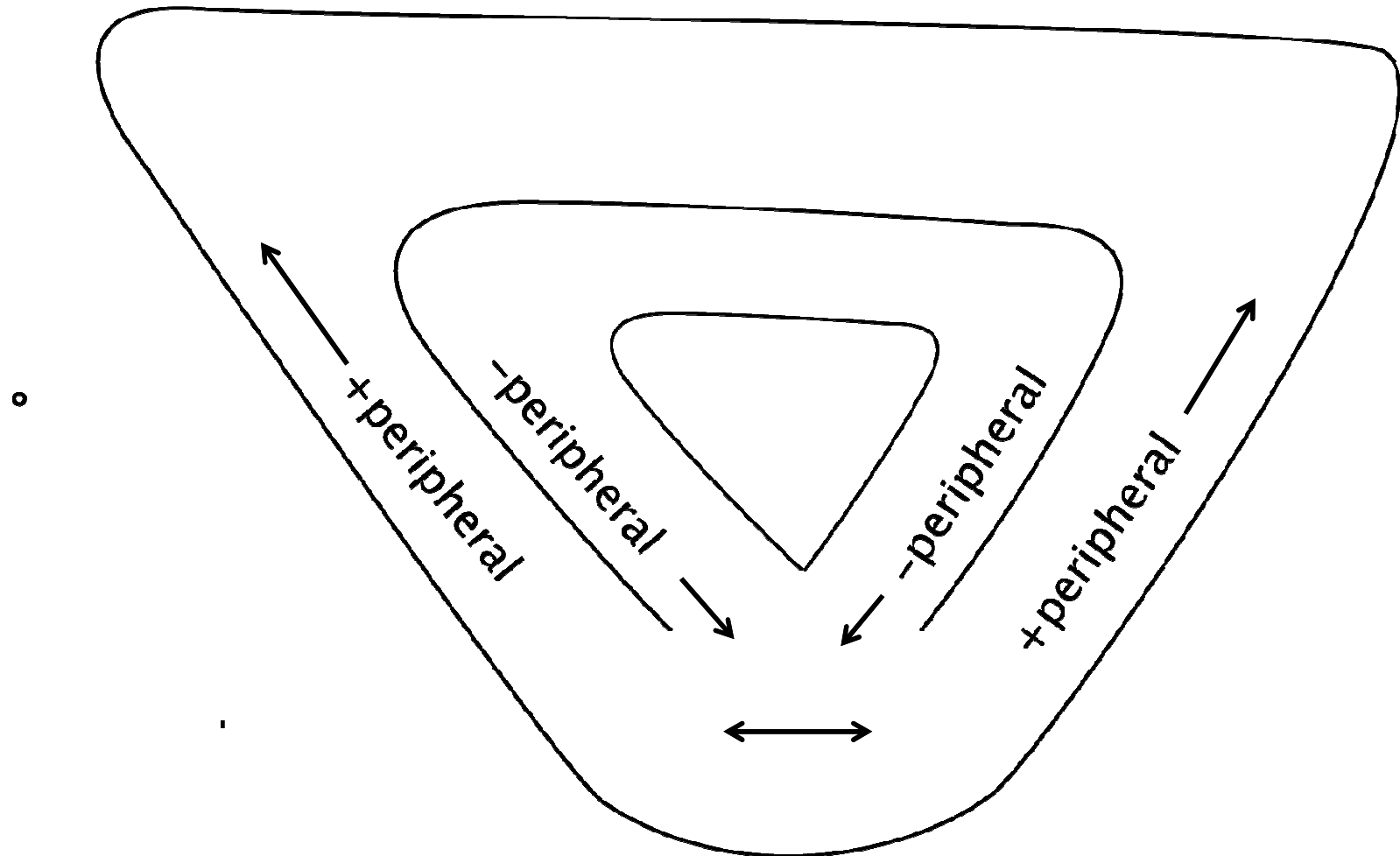


General principles of chain shifting

In chain shifts,

- I. Tense nuclei rise along a peripheral track
- II. Lax nuclei fall along a non-peripheral track

Peripherality is defined for mid and high vowels only



Apparent time changes in the vowel system of Philadelphia

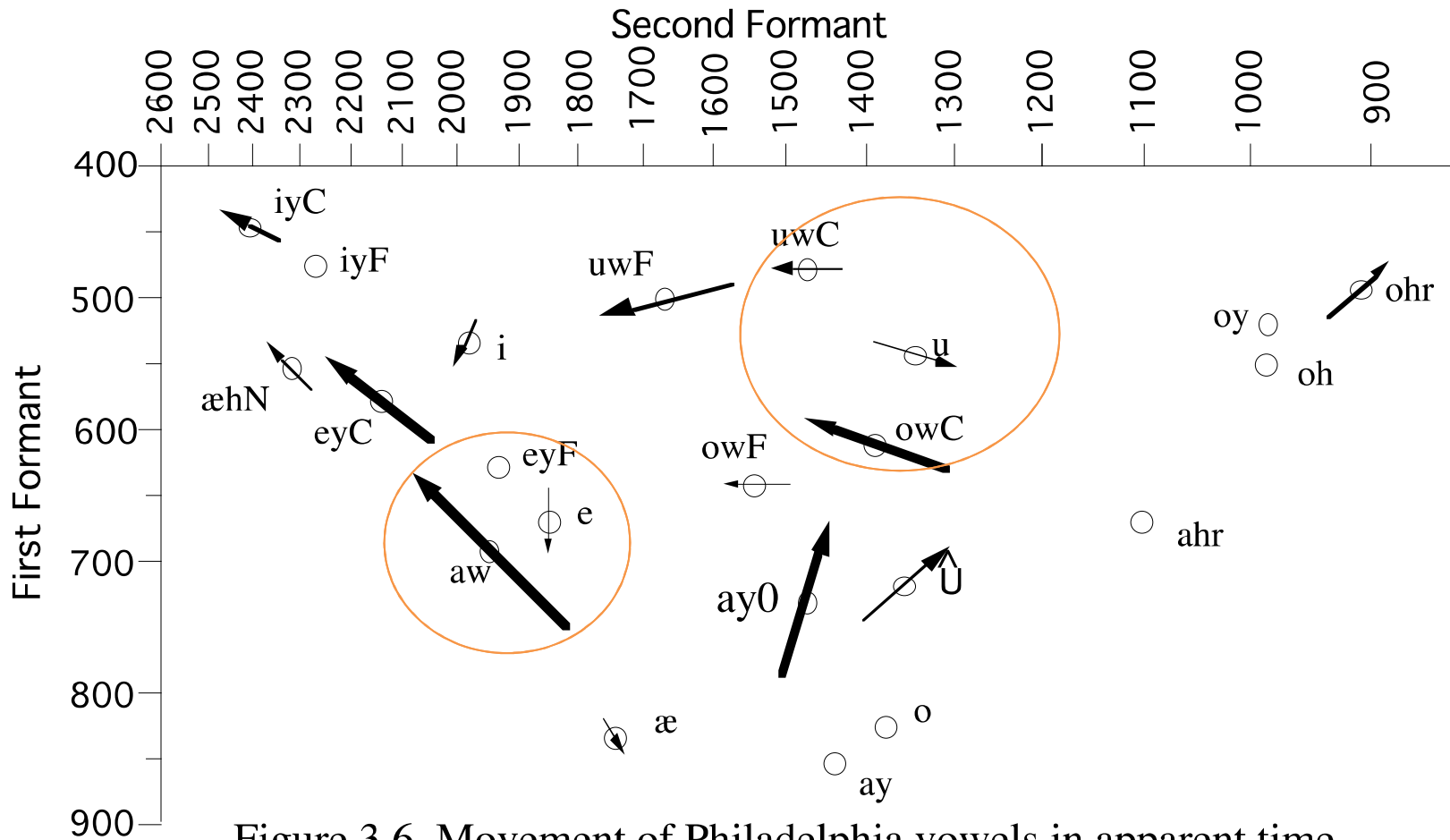
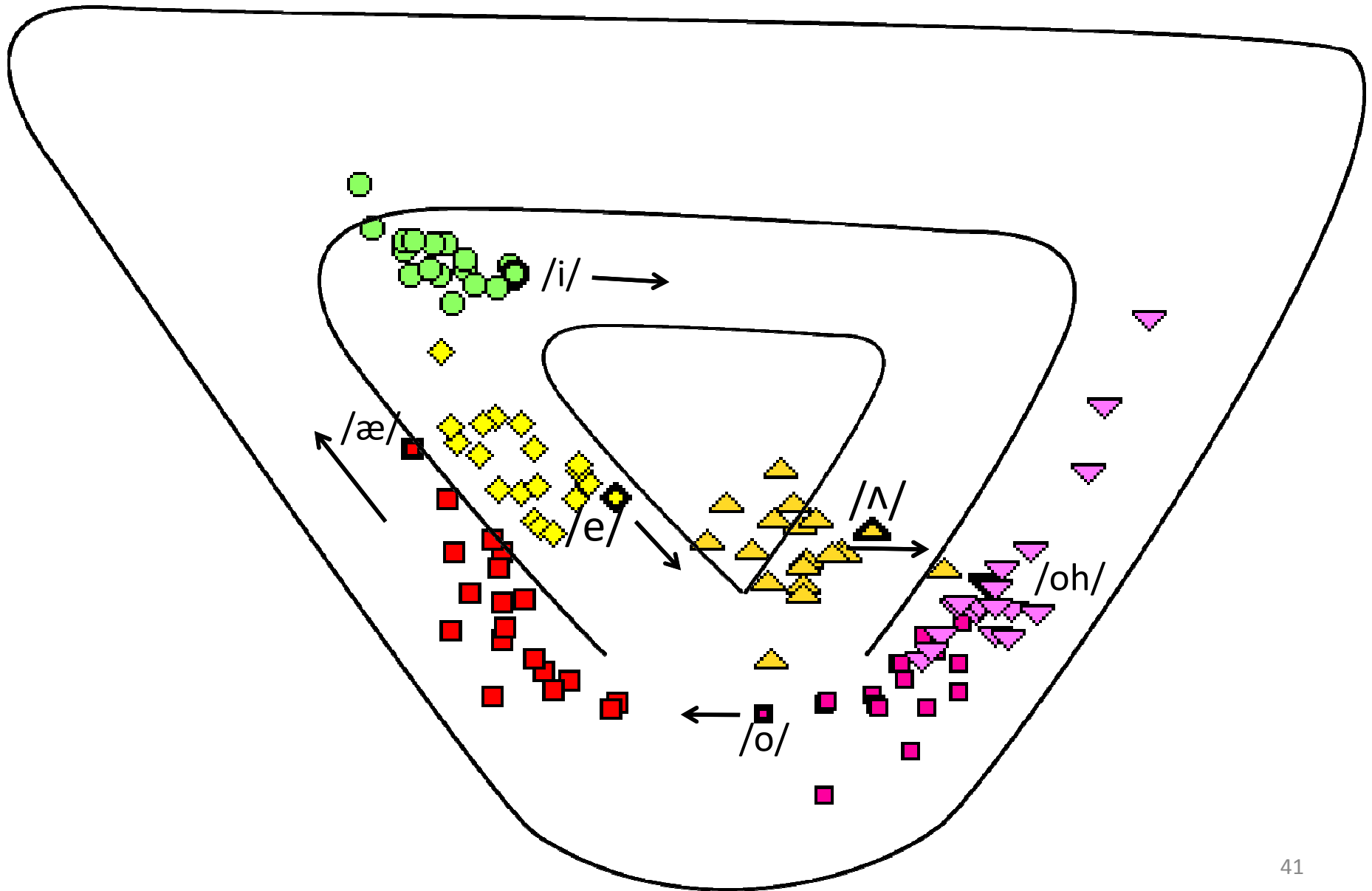


Figure 3.6. Movement of Philadelphia vowels in apparent time. Circles: mean values for 116 speakers in the Neighborhood Study. Vectors connect values for groups 25 yrs older & younger than mean. __F = free vowel; __C = checked vowel. __0 = before voiceless finals.



The Northern Cities Shift in phonological space.
Means for Inland North dialect highlighted)



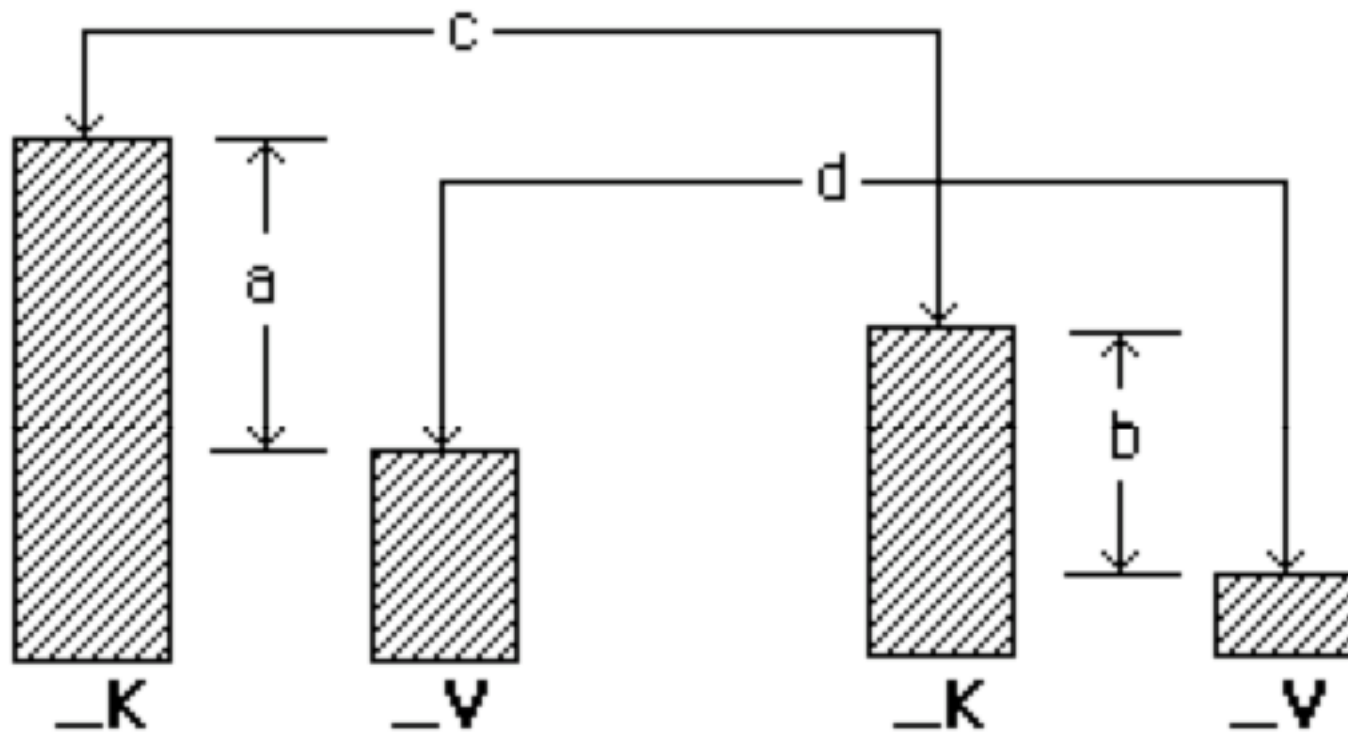
What can be learned from 7,101 vowel measurements in 2013 that could not be learned from 214 in 1995?

What can be learned from 344 speakers in 2013 that could not be learned from 112 speakers in 1975?

m

- More rapid and more accurate means of measurement
- New questions raised by previous finding

Lexically regular variation: coronal stop deletion in English. Two basic constraints on the simplification of /t,d/ clusters in English: the effect of a following consonant vs. a following vowel (a vs. b) and the effect of grammatical boundary (c vs. d)



Monomorphemic

hold

held?

Past tense

holed

DEC PDP-11, 1970s



28K Words

Distribution by age and year of interview for 164 Philadelphia Neighborhood Corpus speakers analyzed for verbs of quotation (N > 4)

