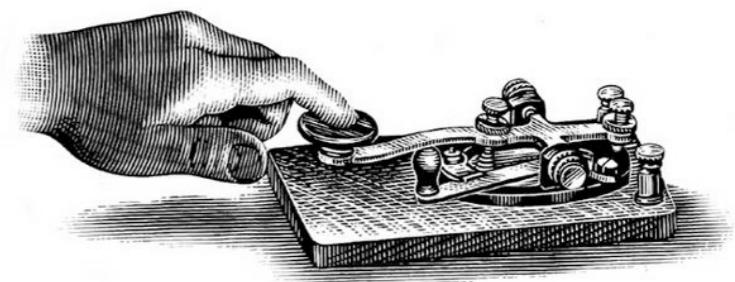


Speech Perception



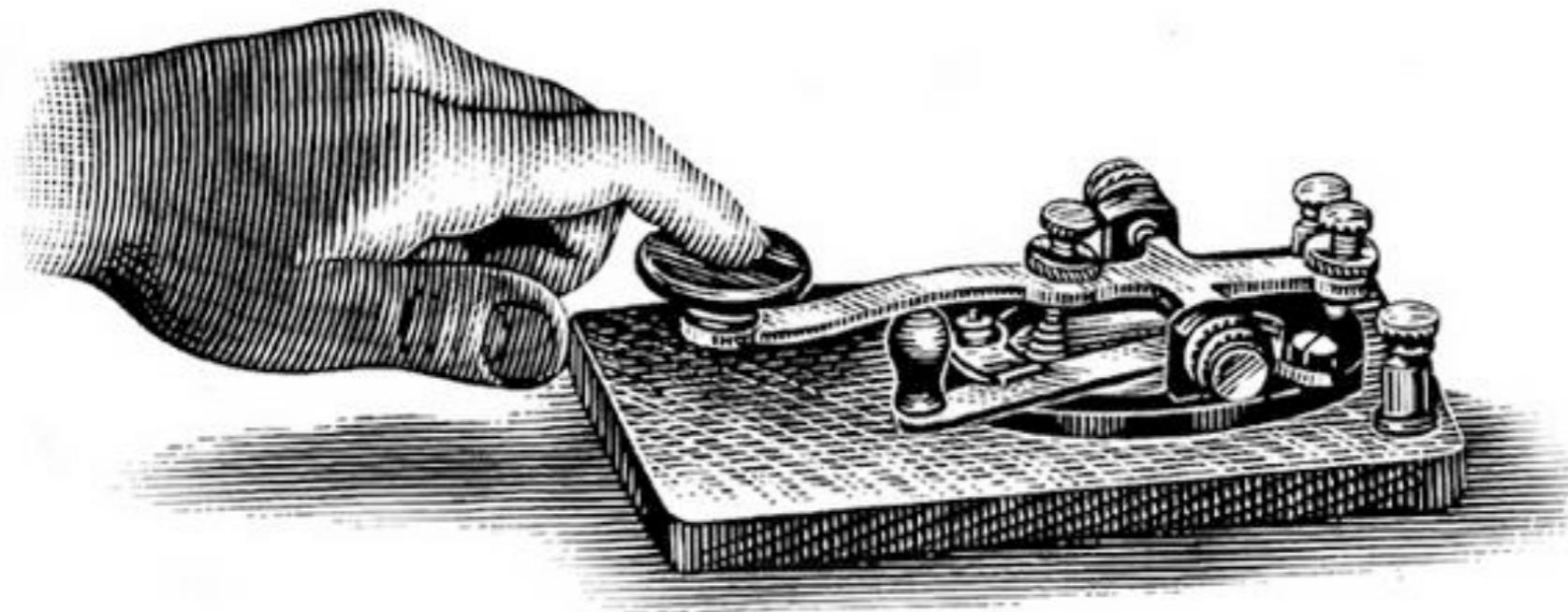
Perceiving social variation:
Perceptions of identity among native speakers & listeners

UK College of Arts
and Sciences
Department of Linguistics

Kevin's office hours

- Monday 7-14: 3:00-4:00
- Thursday 7-18: 12:00-1:00

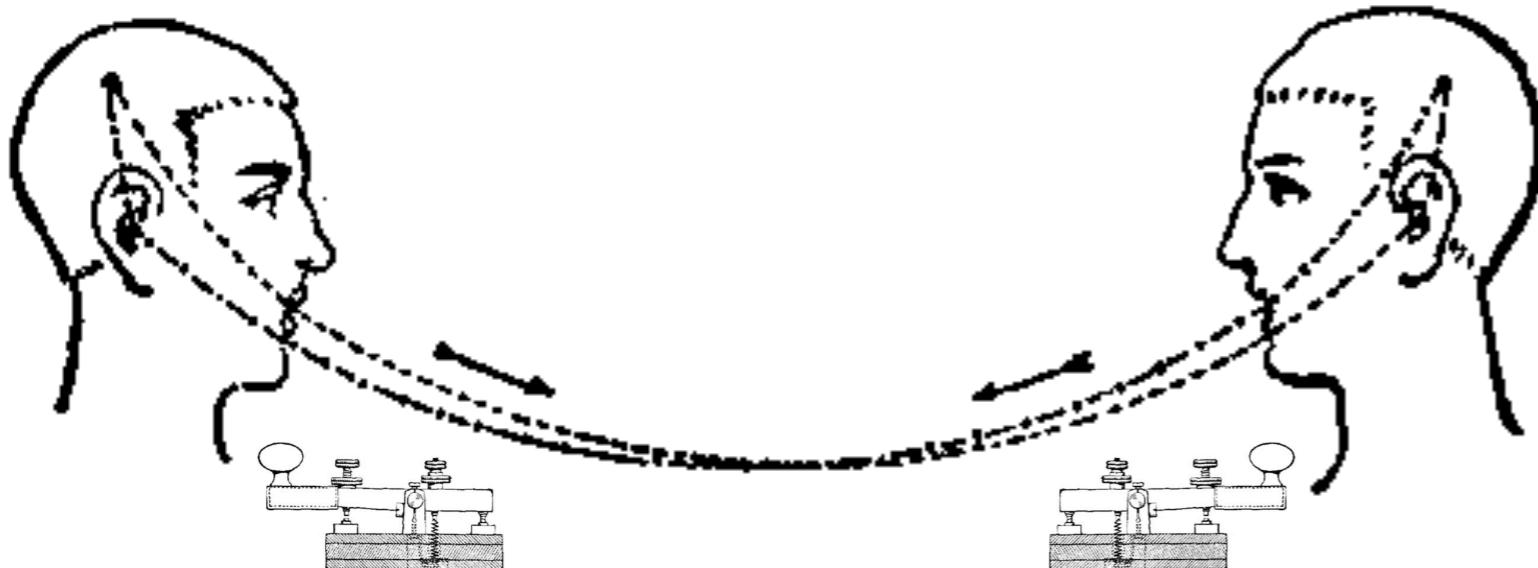
Morse Code



.... . -... .-.. --- -.-. .-- --. -.-. .-. -... -..
h e l l o , w o r l d

A telegrapher's “fist”

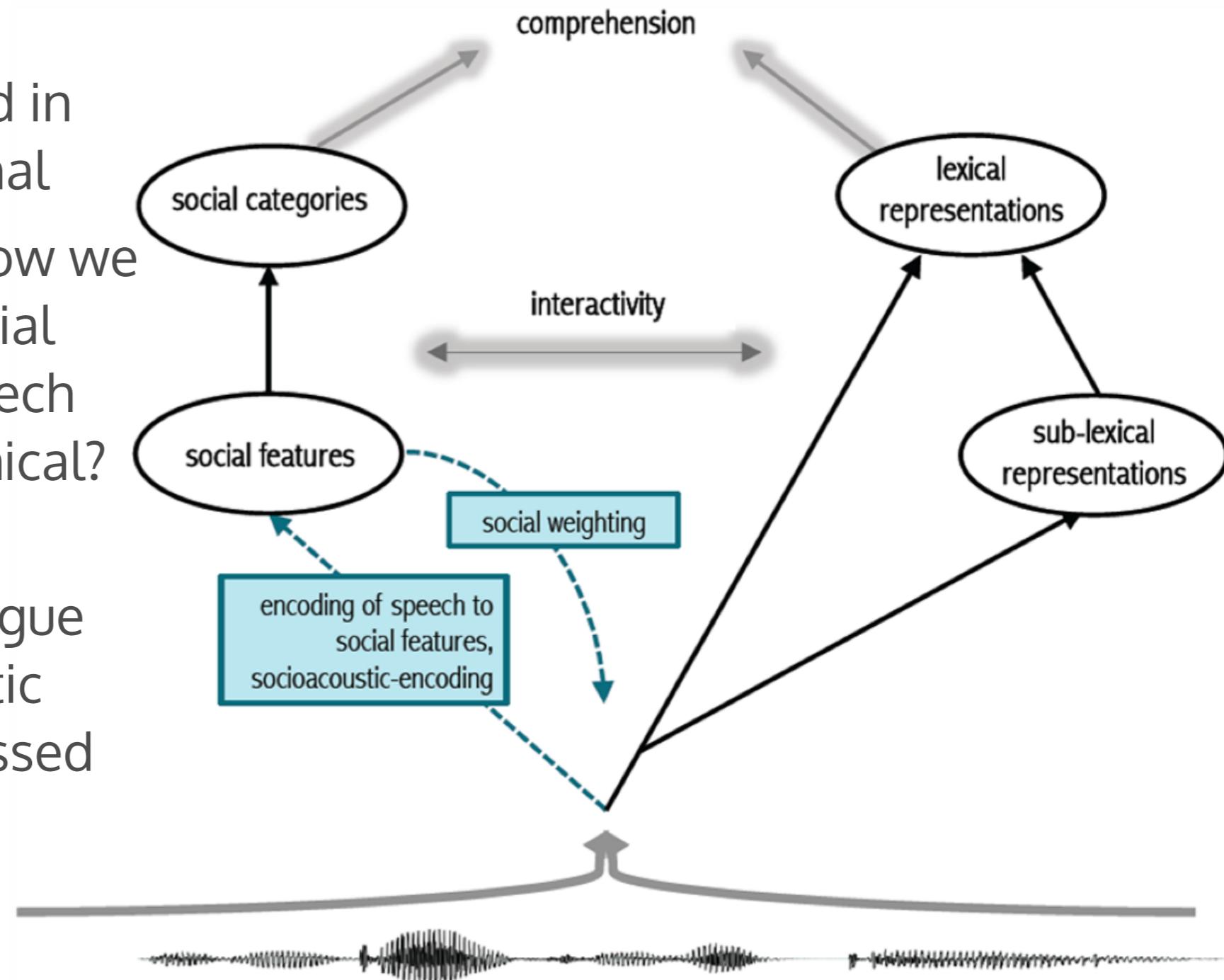
- “A telegrapher's Morse ... is as distinctive as his face, his tones, or his handwriting and as difficult to counterfeit as his voice or writing.”
- “A woman's Morse is as feminine as her voice or her handwriting.”
- “And it would be quite possible for an imaginative operator to build up a fairly accurate mental image of [an unknown sender], whether he ate with his knife, or wore his hat cocked on the side of his head, or talked loud in public places.”



– Hall, L.C. (1902). McClure's Magazine, January, p. 227-231.

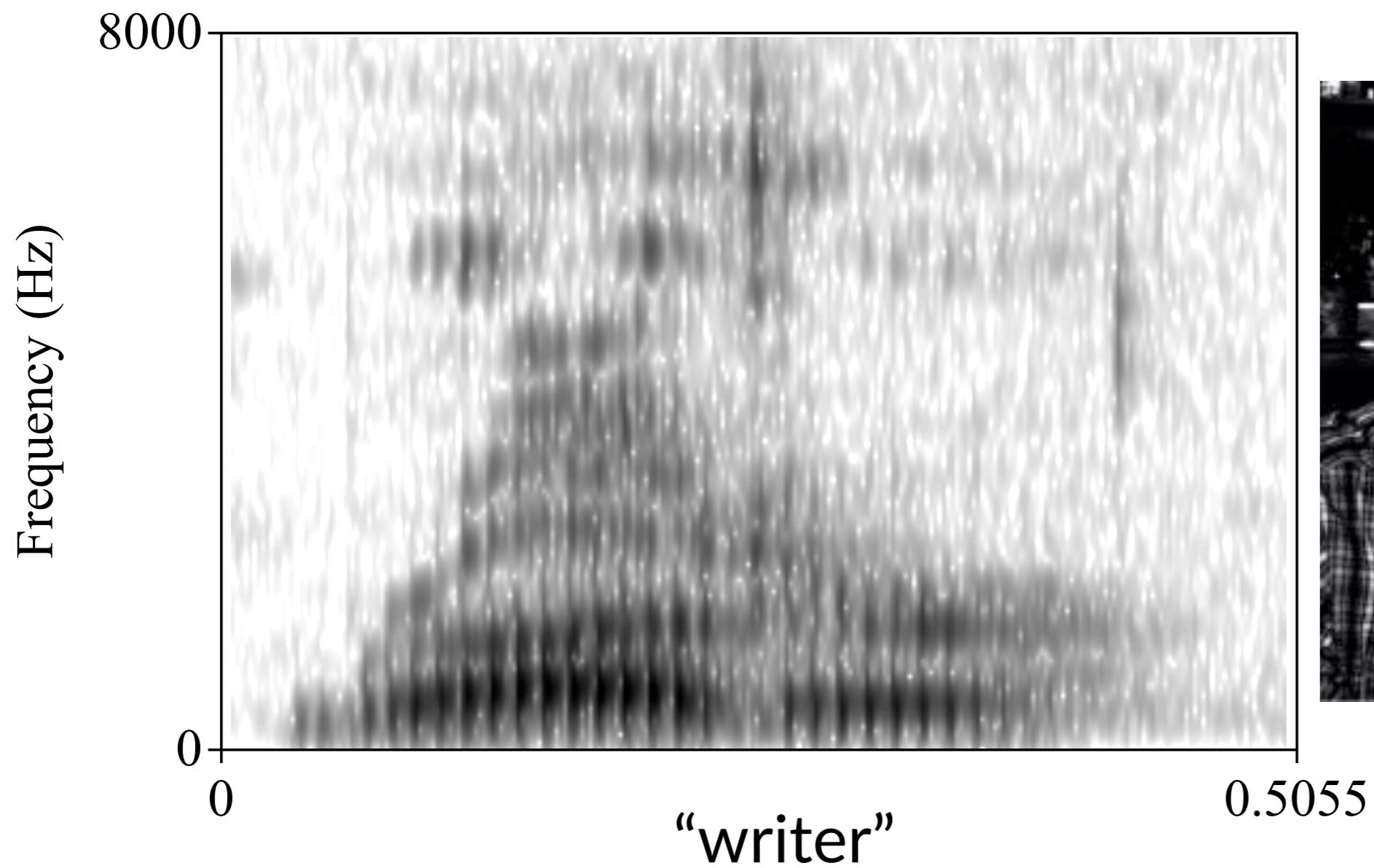
Speech is at least this informative!

- Social and linguistic information are carried in the same acoustic signal
- A crucial question is how we want to model the social information in the speech signal: irrelevant? inimical? informative?
- Sumner et al. (2014) argue that social and linguistic information are processed simultaneously and in parallel

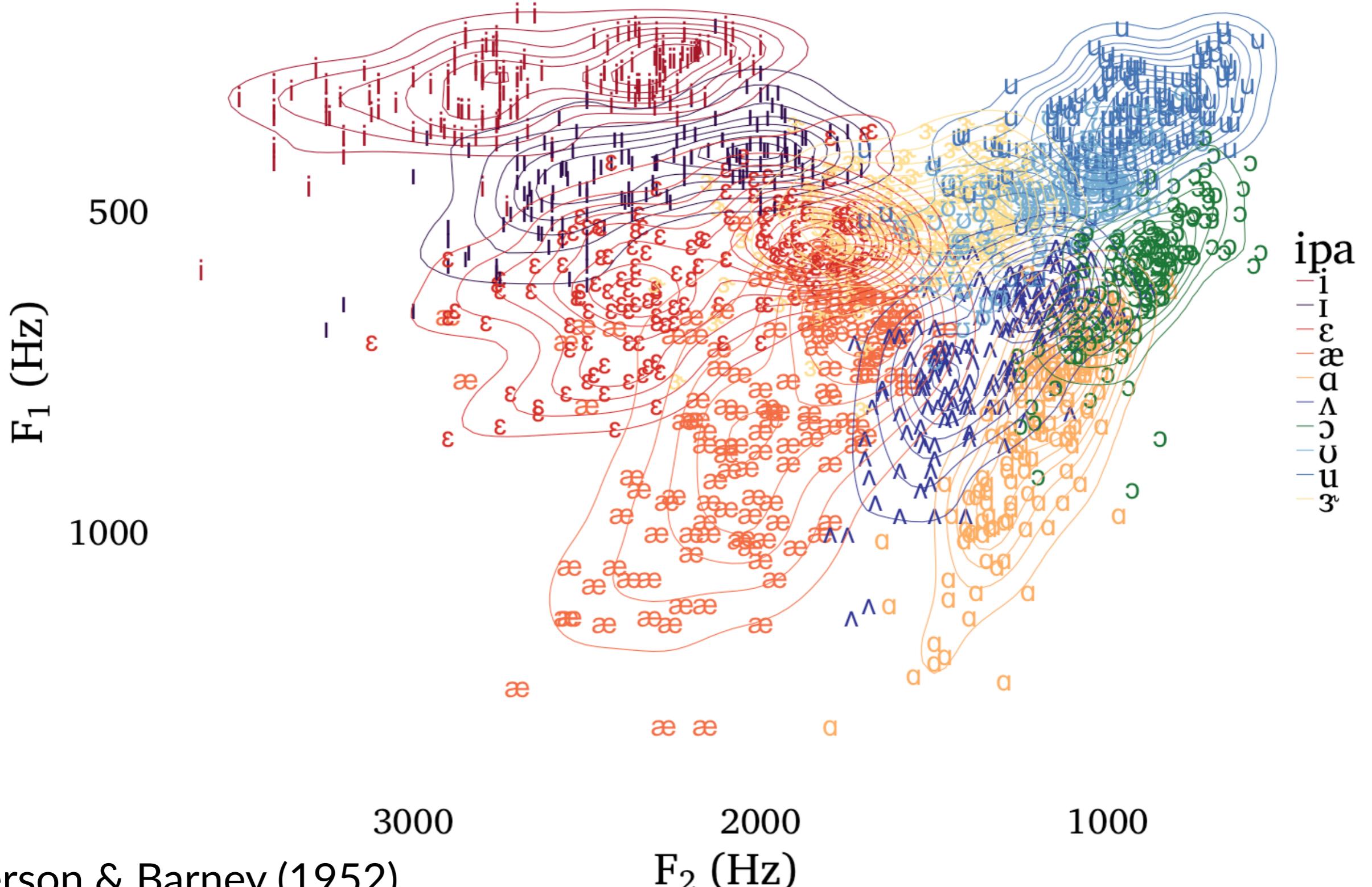


Sumner et al. (2014)

Variation cues social information



Social information cues variation



F_1 (Hz)

500

1000

3000

2000

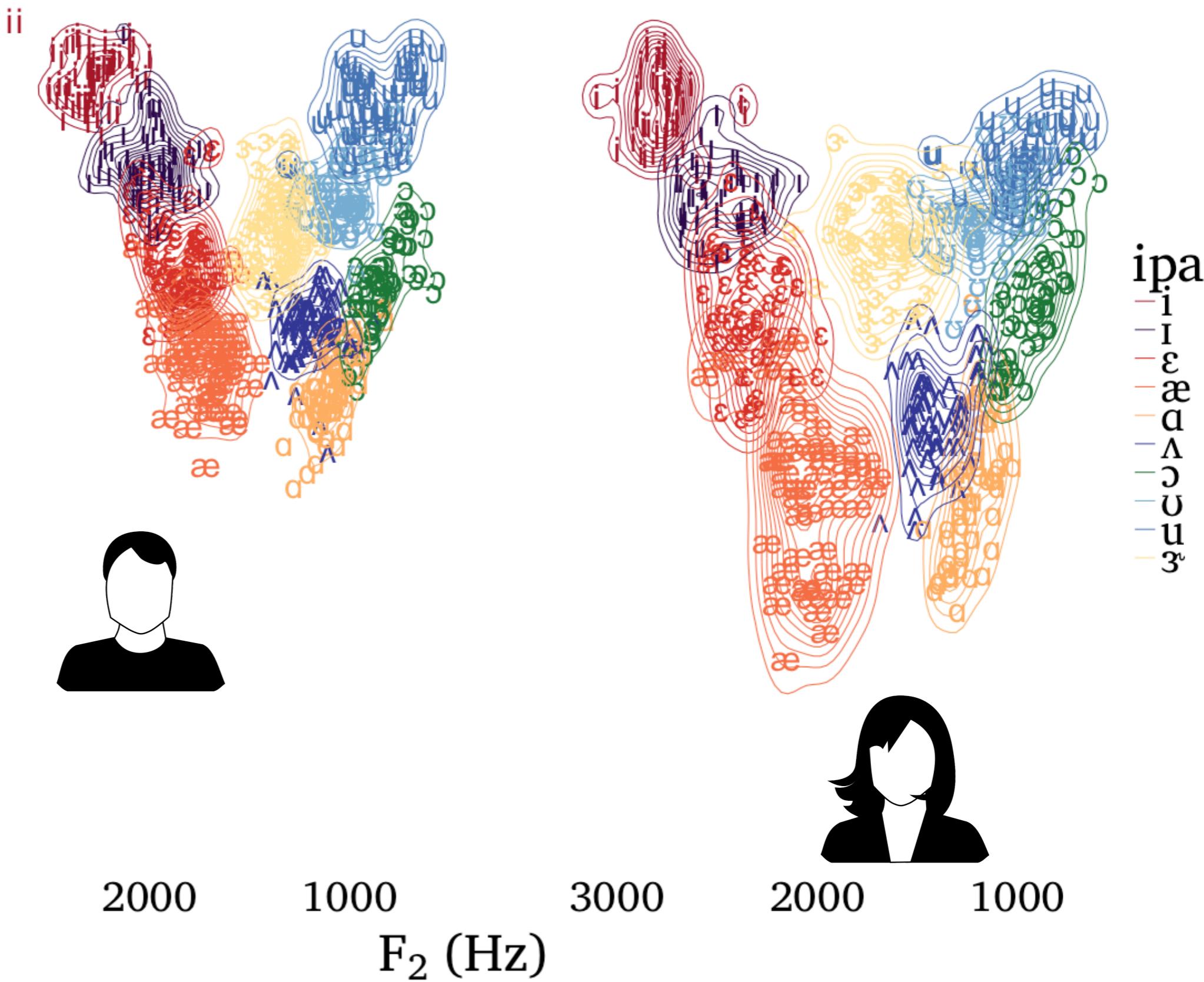
1000

3000

2000

1000

F_2 (Hz)

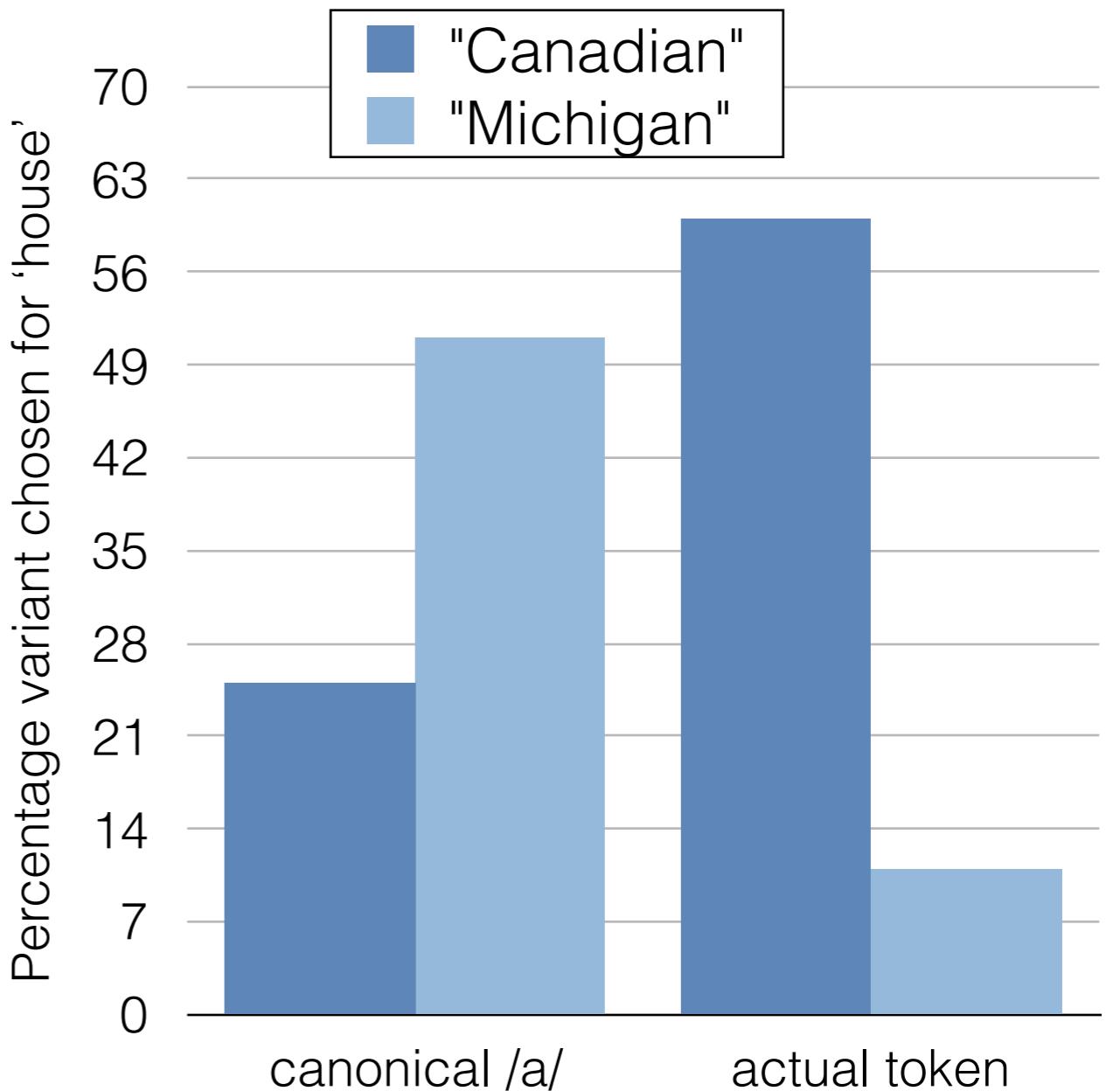


Social effects in speech perception

Top down social information can shift perception of phonetic detail

- Niedzielski (1999), Strand (1999), Hay, Warren, and Drager (2006), Staum Casasanto (2009), Walker & Hay (2011), Drager (2010) and Docherty & Foulkes (2014), etc.

Similar results are often explained in terms of exemplar models (e.g. Johnson, 2006; Hay, Nolan, and Drager, 2006, etc.)



Niedzielski (1999)

But phonetics, as a field, didn't suddenly discover social information in 1997.

Ladefoged and Broadbent (1957)

- “Most speech sounds may be said to convey three kinds of information:
 - linguistic information which enables the listener to identify the words that are being used;
 - socio-linguistic information, which enables him (sic) to appreciate something about the background of the speaker;
 - and personal information which helps to identify the speaker.”

Ladefoged and Broadbent (1957)

- Ladefoged and Broadbent tested their four test words on listeners and then compared results by the language background of the listener

TABLE IV. Identifications of test word *D* in association with version one of the introductory sentence by different groups of subjects.

Number in group	Character of group	Identified as:	
		<i>bat</i>	<i>but</i>
7	Scots	3	4
19	English influenced		
	Scots	4	15
19	English (R.P.)	1	18

So what happened to social information between 1957 and 1997?

Social information never went away

- 1972: Willis, C (JSHR) finds that listeners' vowel perceptions change with their native dialect (Ontarians without Northern Cities perceive set/sat differently from Buffalonians who have it)
- 1983: Schulman, R. (PERILUS) The same Swedish/English bilinguals can hear the set/sat distinction when given instructions in English, but not when given instructions in Swedish.
- 1991: Green, Kuhl, Metzof, Stevens (P&P) Changing gender of face does not break McGurk
although...

Affordances of theory

- Theories of speech perception make **predictions** and have **assumptions**, but they also have *affordances*.
- Gibson (1966) coined the term *affordance* for (visual) perception to describe the actions offered to a particular animal by a particular environment.
- If the animals are speech perception researchers then the theory (or theories) they adopt make some insights easier or more difficult than if one adopts a different theory

It matters how we frame the problem: MT

- What does Motor Theory have to say about social information?
- Motor Theory asserted that speech was special.
- If speech is processed in its own module in the mind. Social information must be extraneous.
- What did Motor Theorists think about McGurk again?

It matters how we frame the problem: MT

- Motor Theory claims that our mental representations (and our memories) of speech are organized by gesture rather than by raw acoustic targets.
- This eventually lead researchers to the surprising, maybe counterintuitive, fact that coarticulation (which can mask or obliterate acoustic targets) is *useful* for listeners.
- We can restate that usefulness in terms of Claude Shannon's (1948) Information Theory, but how does one overcome the assumption that coarticulation is damaging in the first place without the affordances of MT?

It matters how we frame the problem: DR & GA

- What do our two domain general theories have to say about social information?
- For both DR and GA, social information contributes to the lack of invariance problem
- So under both theories you get to normalization, but the focus is still on recovering the invariant gestures or the sounds themselves — not on the source of the variation

It matters how we frame the problem: Exemplars

- Early exemplar models were *not* explicitly sociophonetic or sociolinguistic in design or intention.
- They are models of categorization.
- The assumptions include:
 - people can and do learn.
 - variation does not need to be normalized, it is stored.
 - experiences are then labeled with abstract category labels.

It matters how we frame the problem: Exemplars

- The abstract category labels quickly became a focus of theoretical investigations themselves (e.g. Strand, Johnson, D'Imperio 1999; Pierrehumbert 2001)
- Exemplar models afford not only asking social questions about speech perception but then provide a framework for making sense of the results

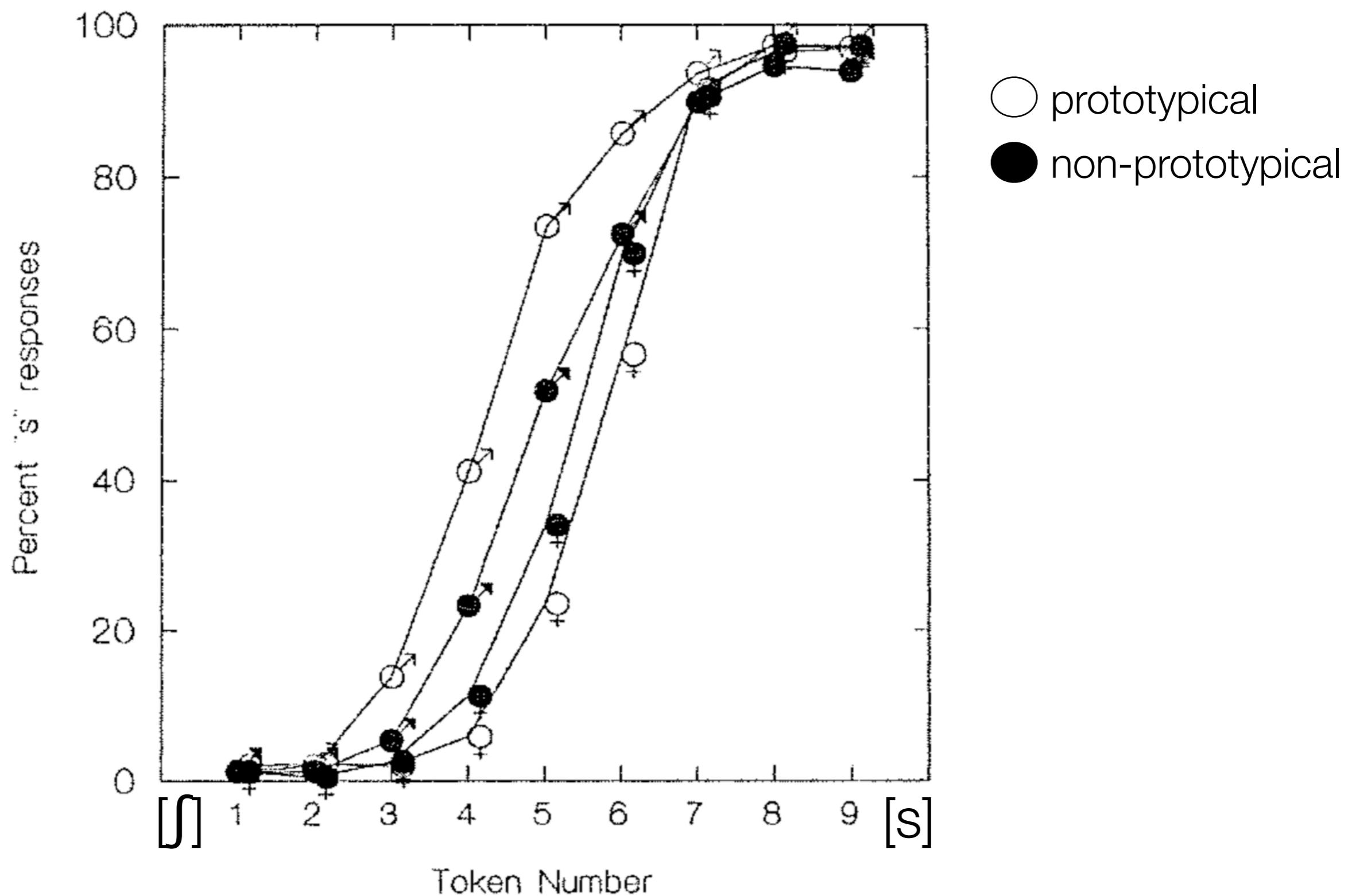
Perception of [s] and [ʃ]

- Strand and Johnson (1996) conducted three experiments to better understand the role of gender in fricative place perception.
- Assumptions from prior work include:
 - Gender differences are purely physical/biological and have to do primarily with size.
 - Listeners normalize these differences to arrive at invariance.

Perception of [s] and [ʃ]

- Stimuli: 9 step synthesized fricative continuum from most [ʃ] like to most [s] like. (Klatt and Klatt 1990)
- Spliced with [ad] and [u] from 4 talkers:
 - non-prototypical female and male
 - prototypical female and male
- (Typicality determined by a 58 person norming study conducted prior to experiment 1)

Experiment 2: voices only



Experiment 2: non-prototypical male voice w/ prototypical male and female faces

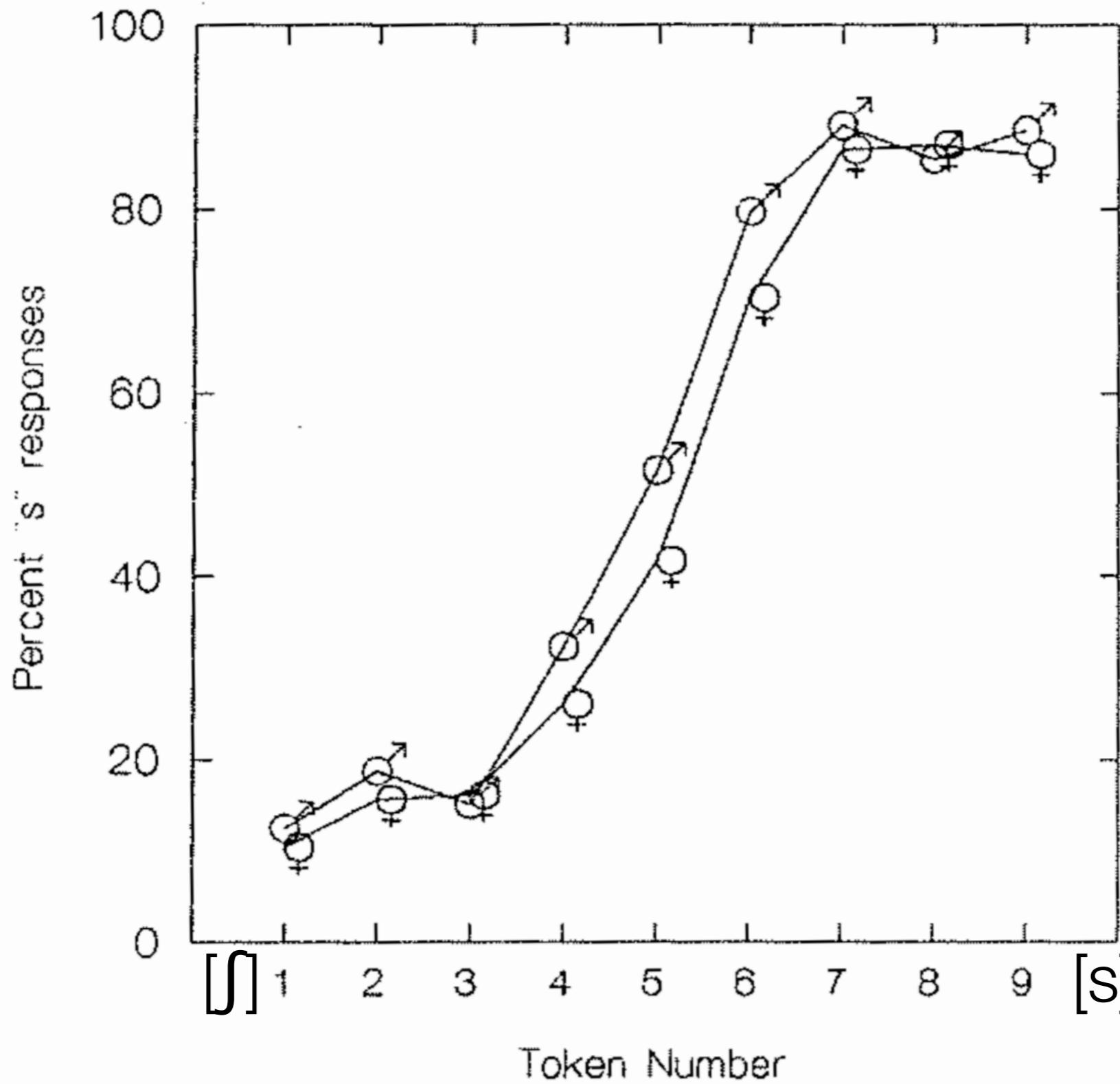


Figure 2.2: non-
prototypical male
voice paired with
prototypical male
and female faces

Experiment 3: faces w/voices

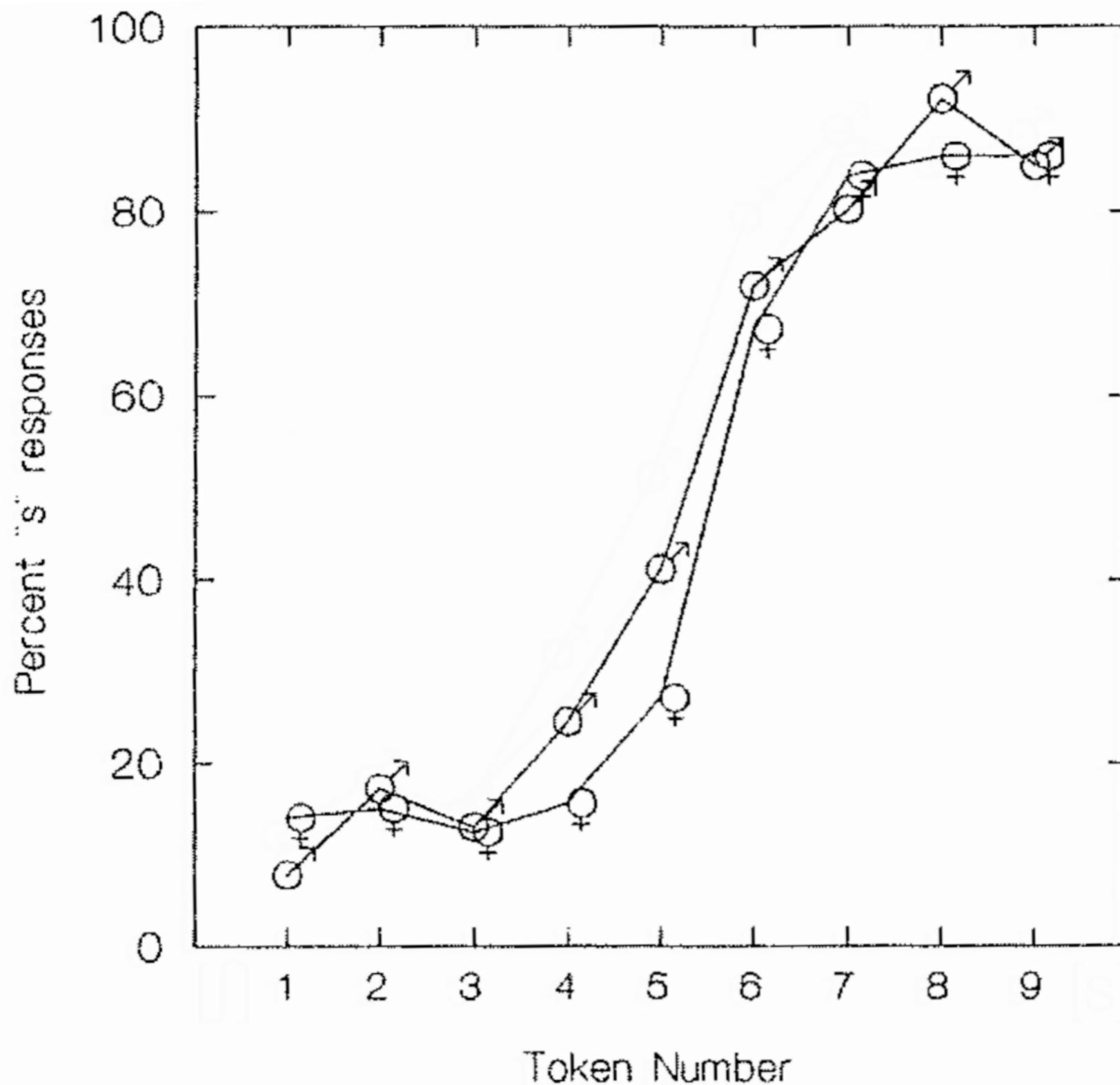
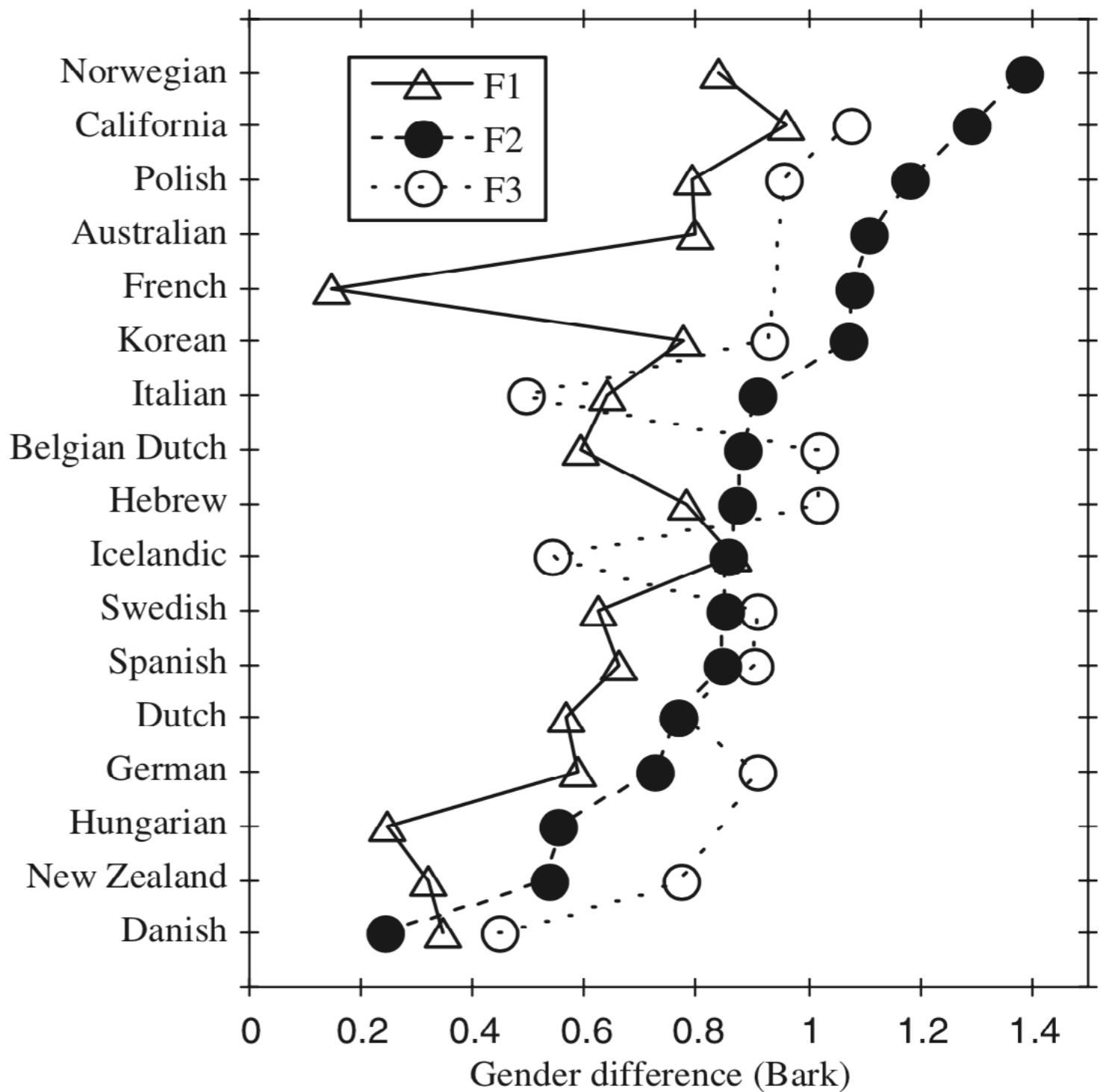


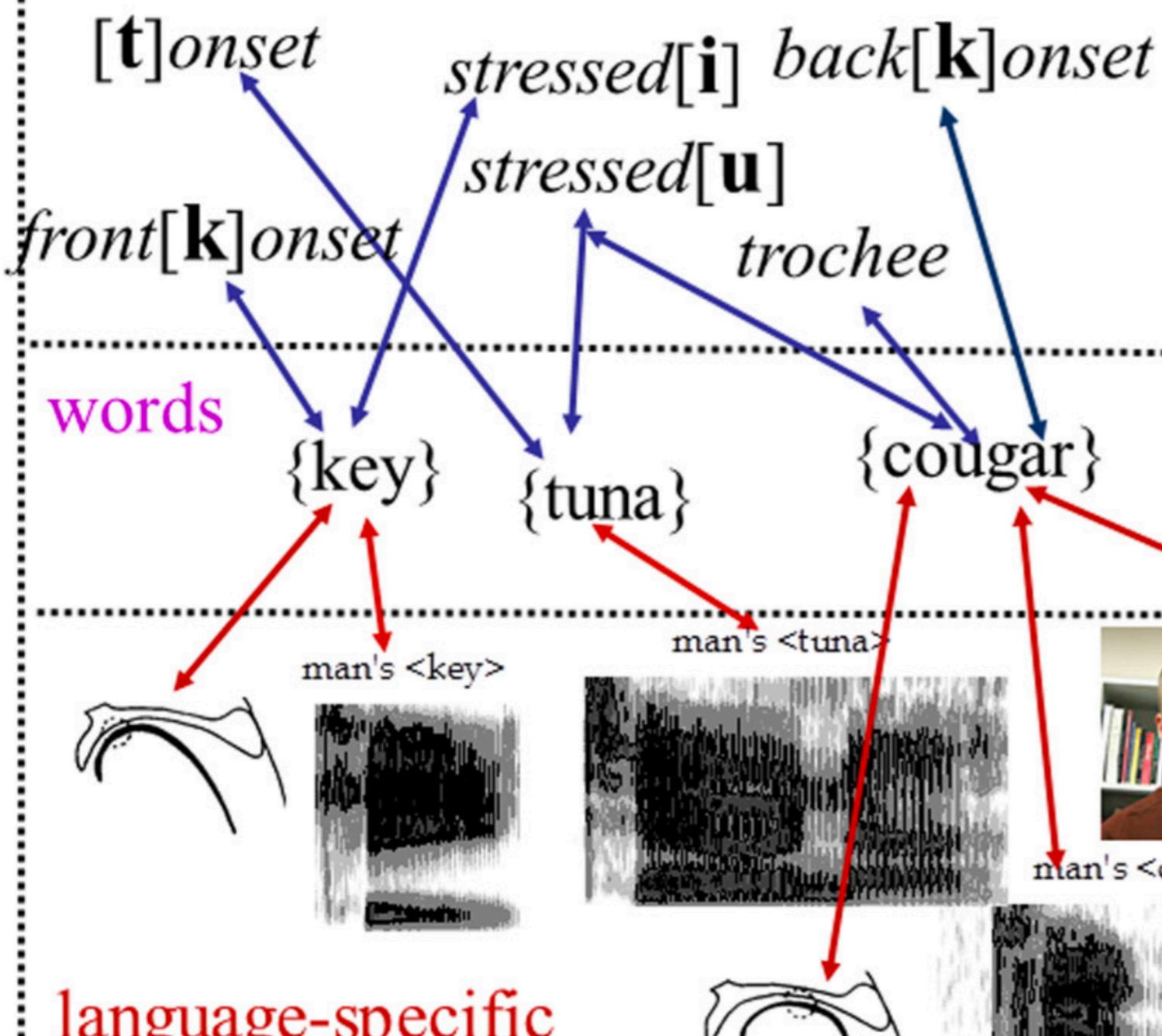
Figure 2.3: non-prototypical female voice paired with prototypical male and female faces

Making gender the question instead of the answer

- Johnson (2006) collects evidence from a number of studies to reveal systematic differences in the way gender is communicated through vowel formants cross-linguistically

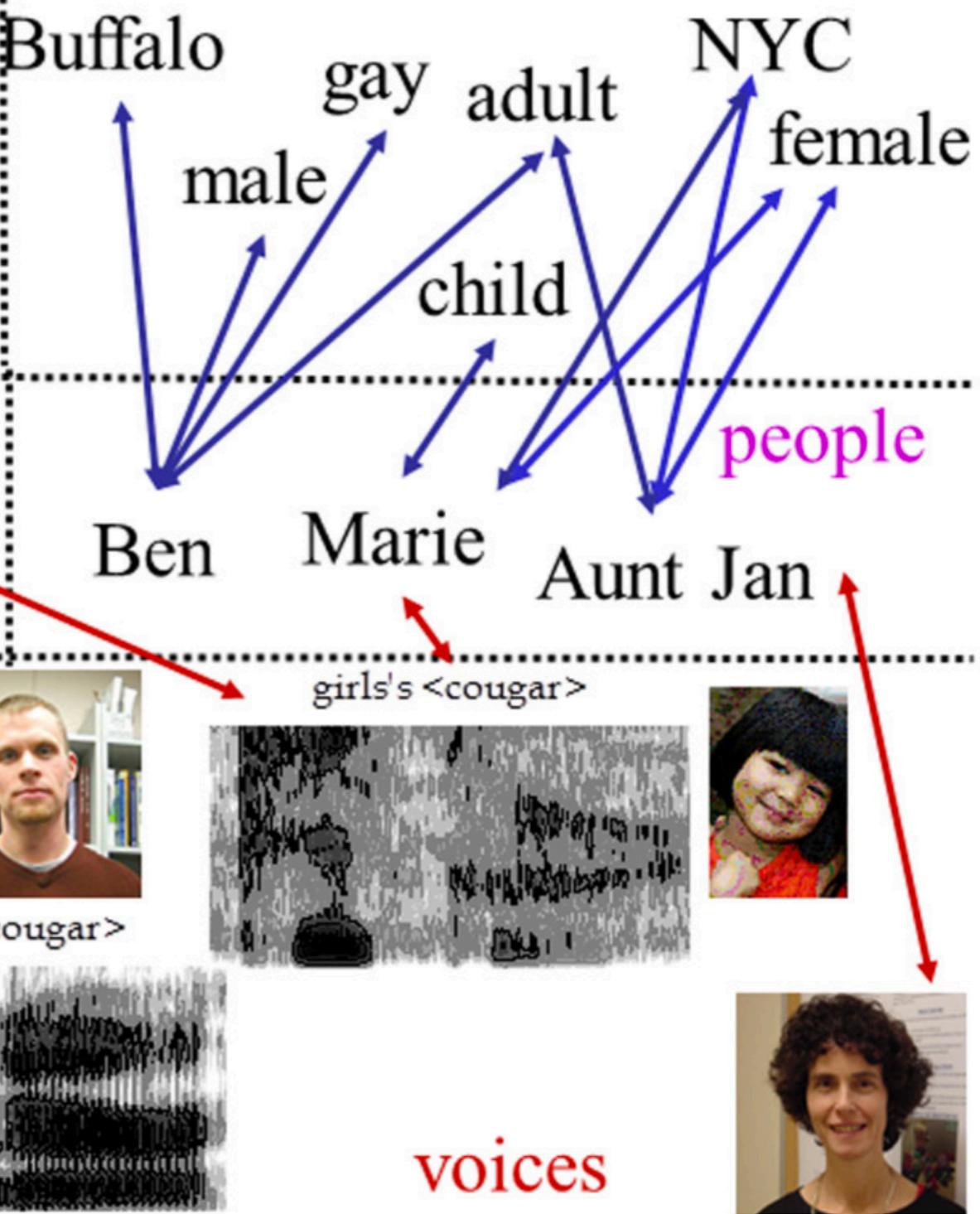


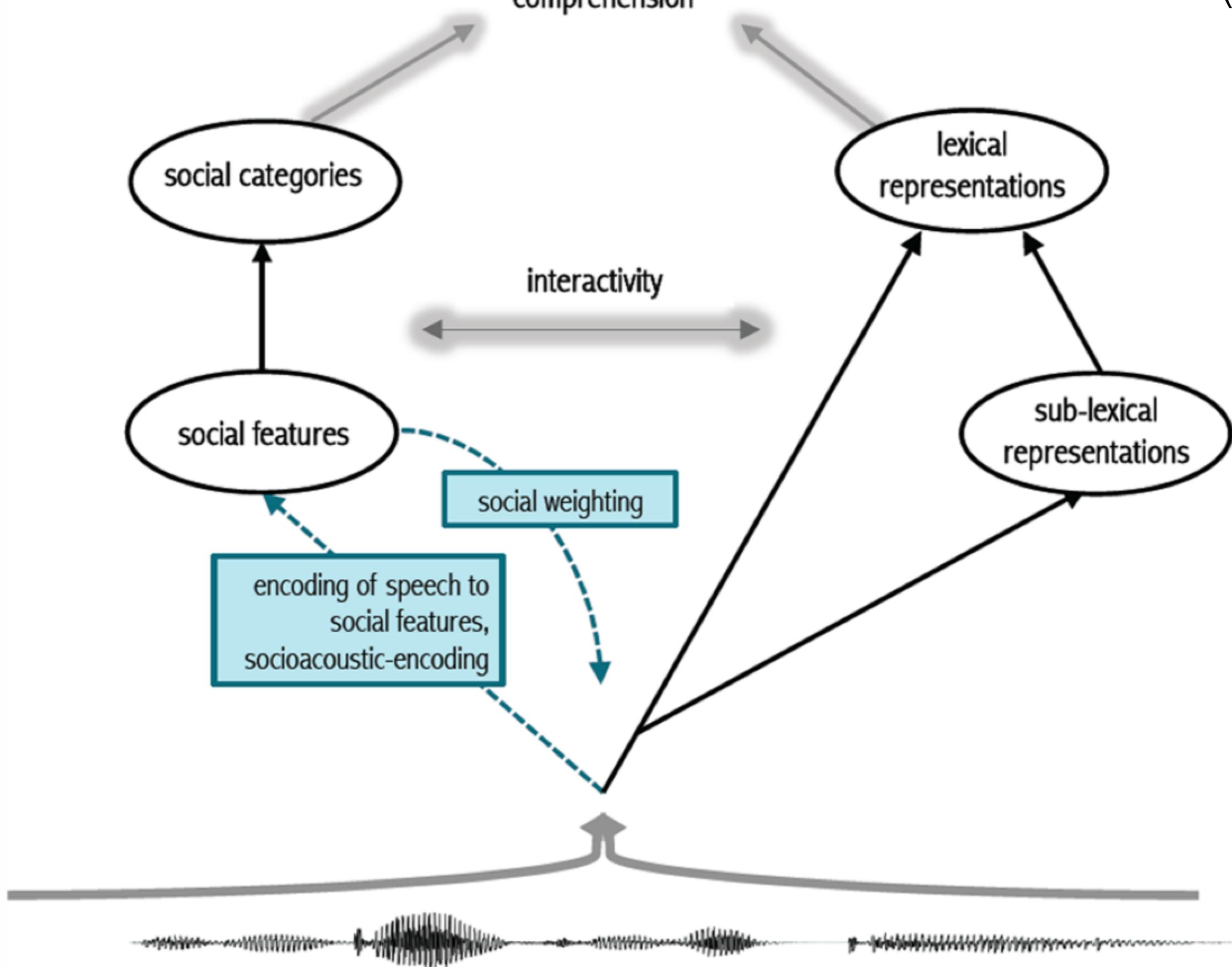
categorical phonological knowledge



language-specific
phonetic detail

socio-indexical knowledge





What is the mental lexicon made of?

- If Sumner, Kim, King, and McGowan are right, then we have to rethink a lot of assumptions about speech perception.
- How do we differentiate between so-called linguistic and non-linguistic cues in speech? What does a linguistic theory look like that can do this?
- How lexical is the mental lexicon (Pufahl & Samuel, 2014)?

Pufahl & Samuel (2014)

- 6 Experiments manipulating changes in both voice and ambient speech sounds and found that changes in either affected accuracy. They conclude:
- “...representations stored in the mental lexicon are not limited to linguistic information, nor are they limited to the addition of information from highly related sources like voices.”