Speaker- and group-specific information in formant dynamics: a forensic perspective

Vincent Hughes

Paul Foulkes

LabPhon 15 Satellite

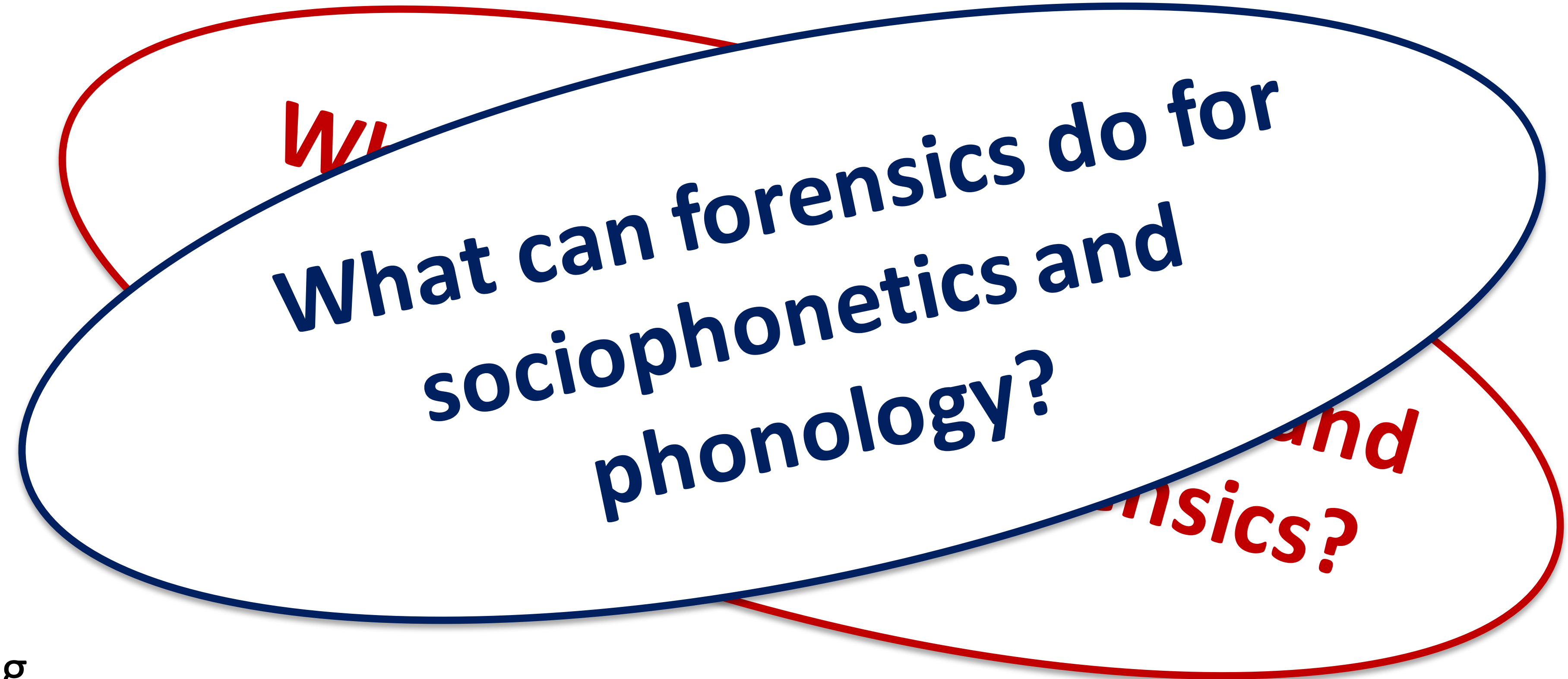
Speech dynamics, social meaning, and phonological categories

UNIVERSITY of York

13th July 2016

voice and identity

0101101ID1



e.g.

Roberts (2012)

Rhodes (2012)

Harrison (2013)

Outline

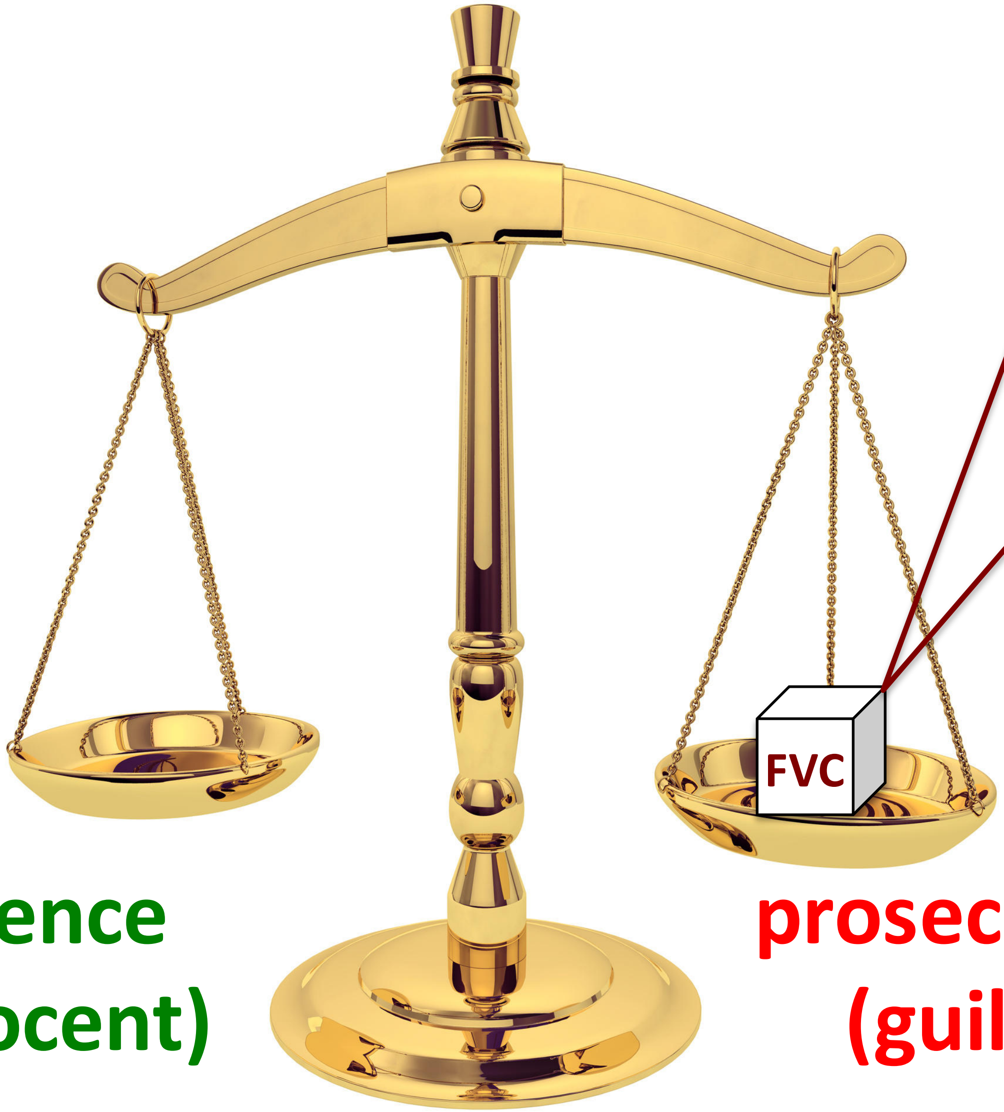
1. the forensic problem
2. formant dynamics in forensics
3. research questions
4. method
5. experiments:
   - speaker discrimination
   - group discrimination
6. discussion

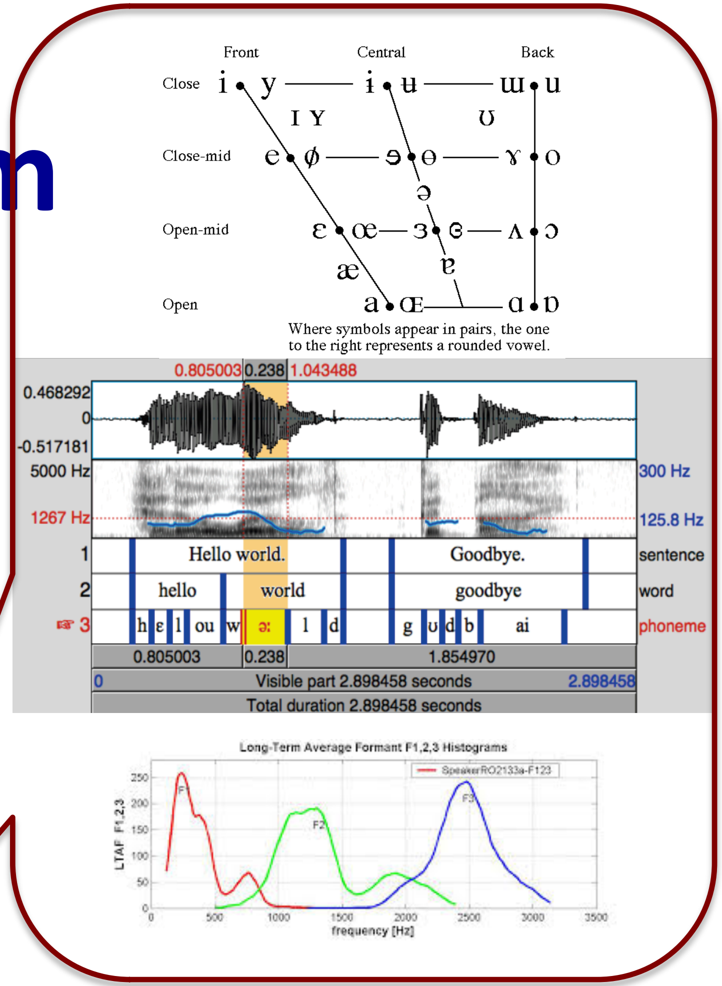
1. The forensic problem

- forensic voice comparison (FVC):

vs.

1. The forensic problem





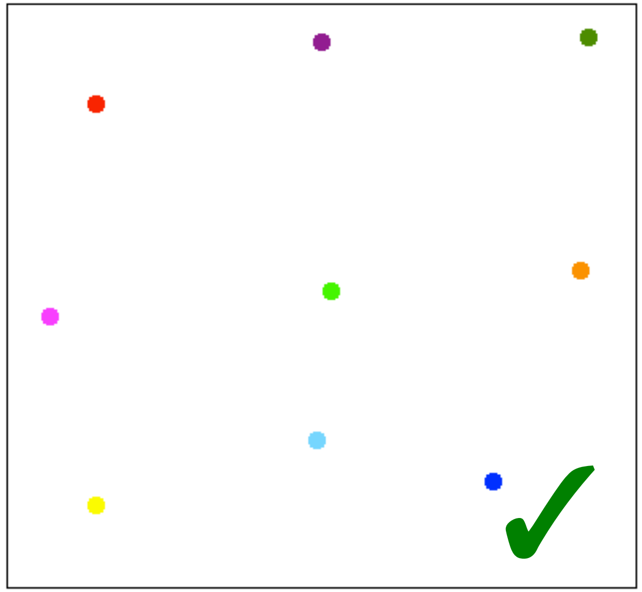
defence
(innocent)

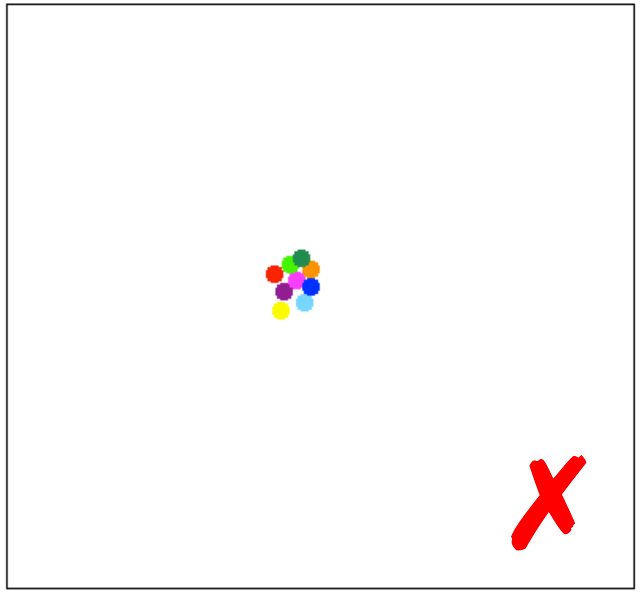
prosecution
(guilty)

1. The forensic problem

- properties of ideal features:
  - **high between-speaker variability**
  - low within-speaker variability

  - resistance to disguise
  - robustness in transmission
  - measurability
  - availability

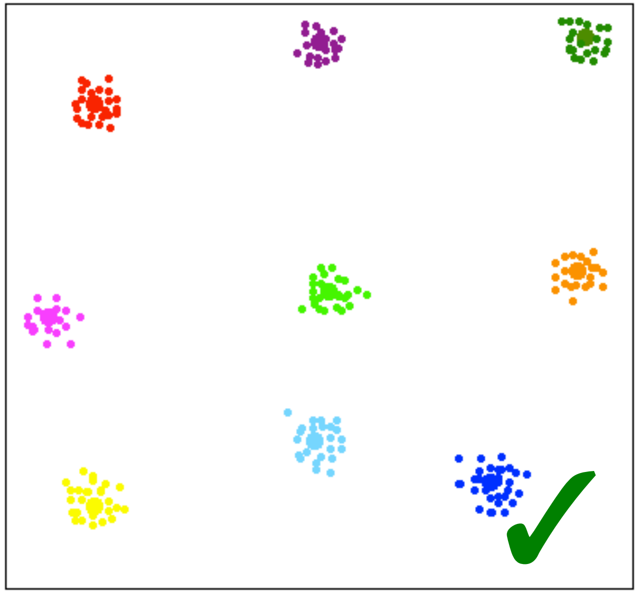


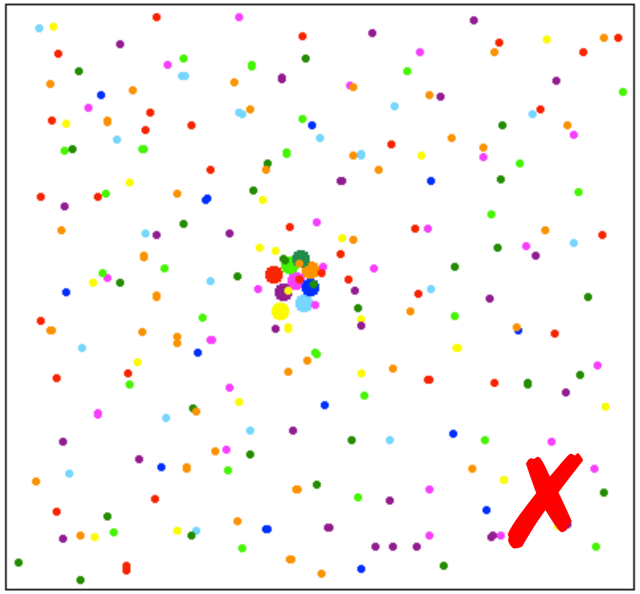


from Nolan (1983)

1. The forensic problem

- properties of ideal features:
  - high between-speaker variability
  - **low within-speaker variability**
  - resistance to disguise
  - robustness in transmission
  - measurability
  - availability





from Nolan (1983)

2. Formant dynamics in forensics

- commonly used in forensics for last 20 years
  - – starting with… Greisbach et al. (1995)
  - – McDougall (2006)
    - value of parametric representations
    - polynomials better than raw Hz input
  - – Morrison (2009)
    - comparison of different parametric representations

2. Formant dynamics in forensics

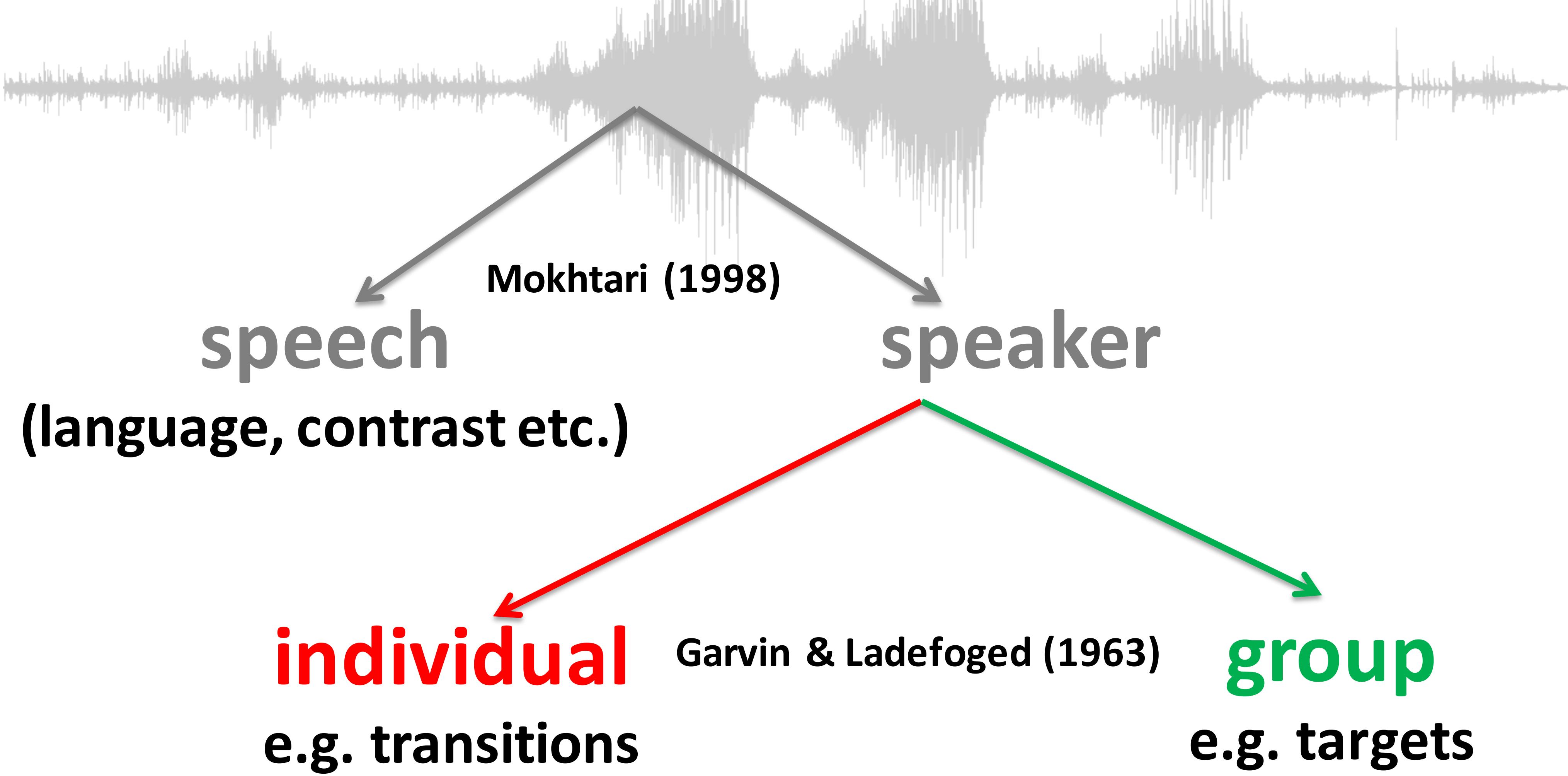
why dynamics?

- targets = learned by speech community
- transitions = “acquired ... by trial and error”
- “speakers' 'vocal signatures' lie in the rapid, transitional movements of the speech organs between sounds”

from Nolan (1997)/ McDougall (2004)

2. Formant dynamics in forensics

- so... phonology is all about targets?



2. Formant dynamics in forensics

- but… inconsistent with e.g. usage-based models?
  - any element of phonetic/ phonological structure can be learned & represented cognitively
  - thus potential for transitions to carry ‘group’ information

- formant dynamics increasingly used to explore group-patterns in sociophonetics

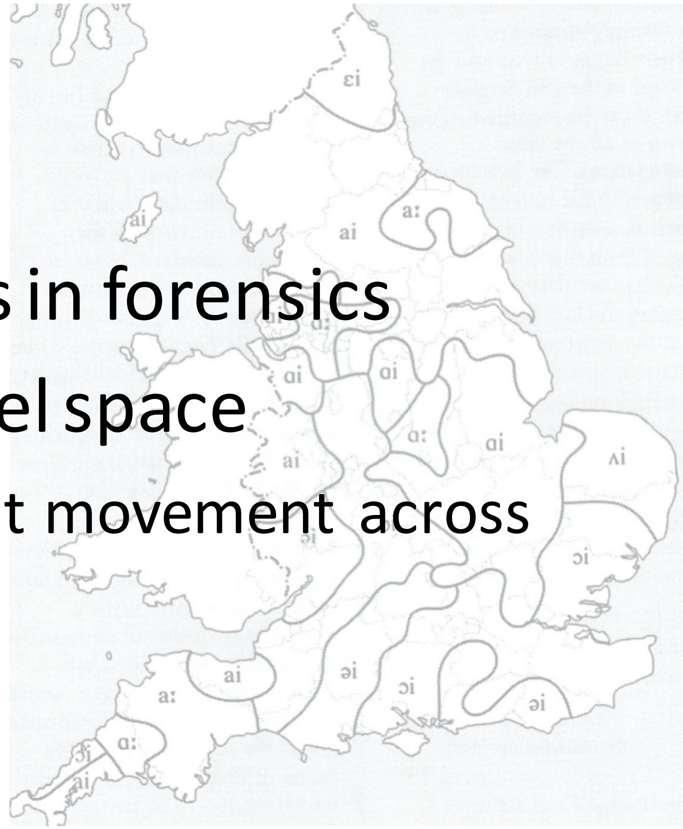
3. Research questions

- to what extent is **speaker**- and **group**-specific information encoded in the dynamics of formant trajectories?
  - implications for models of phonology
  - value of the forensic perspective

4. Method

variable

- PRICE /aɪ/
  - subject of considerable analysis in forensics
  - covers a wide range of the vowel space
    - potential for considerable formant movement across the duration of the vowel



4. Method: datasets

(1) Standard Southern British English (SSBE)

- DyViS corpus (Nolan et al. 2009)
- 97 male speakers
- 18-25 years
- mock police interview (map task)

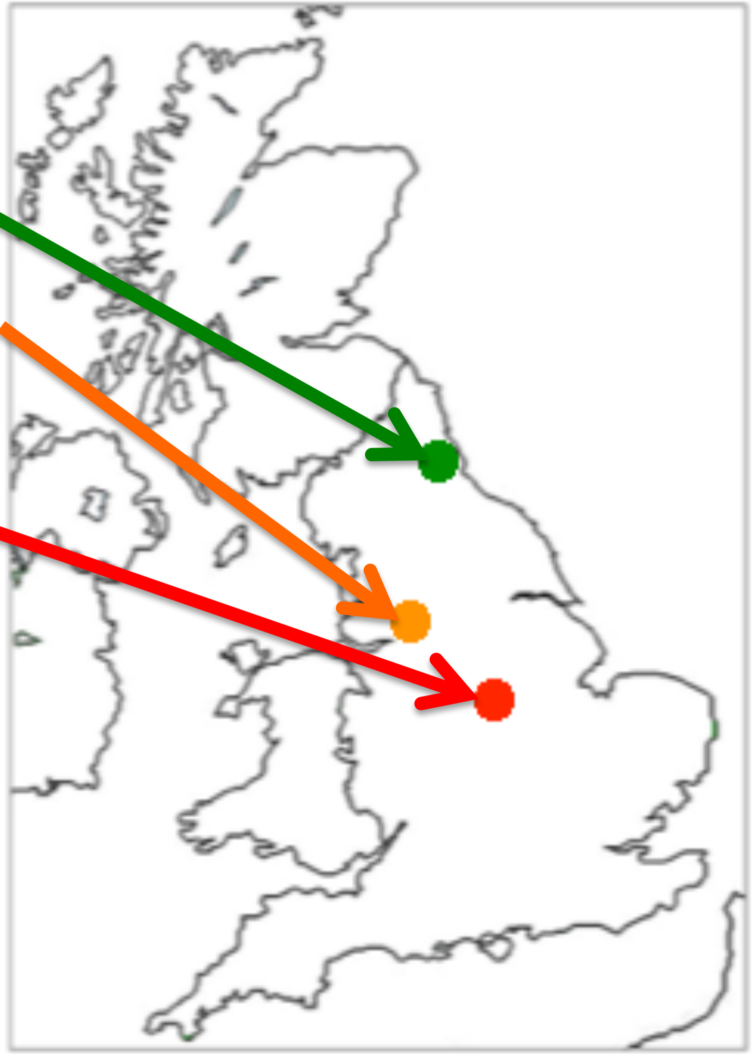
4. Method: datasets

(2) Newcastle (Milroy et al. 1994-97)

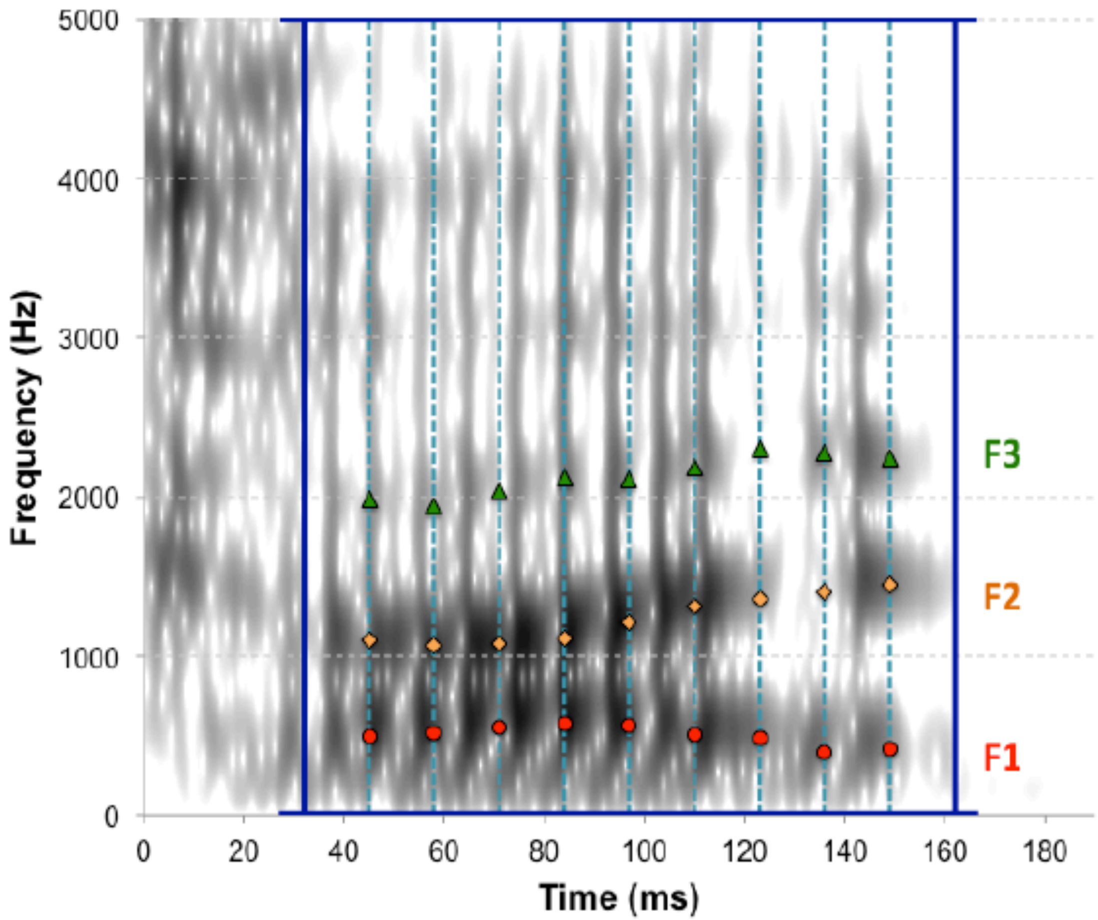
(3) Manchester (Haddican et al. 2013)

(4) Derby (Milroy et al. 1994-97)

- 8 male speakers
- 18-31 years
- sociolinguistic interviews in peer-group pairs



4. Method



dynamics

- c. 10 tokens/ sp
- measurements at +10% steps
- pre-testing for optimal fit
  - cubic polynomials
- 4 coefficients/ formant

statics

- +20% & +80% Hz values/ formant

5. Results: speaker discrimination

- SSBE speakers:
  - 20 test speakers
  - 57 reference speakers
- **same-** (**SS**) & **different**-speaker (**DS**) comps
- likelihood ratios (LRs) used for discrimination

$$\frac{p(E \mid H_p)}{p(E \mid H_d)}$$

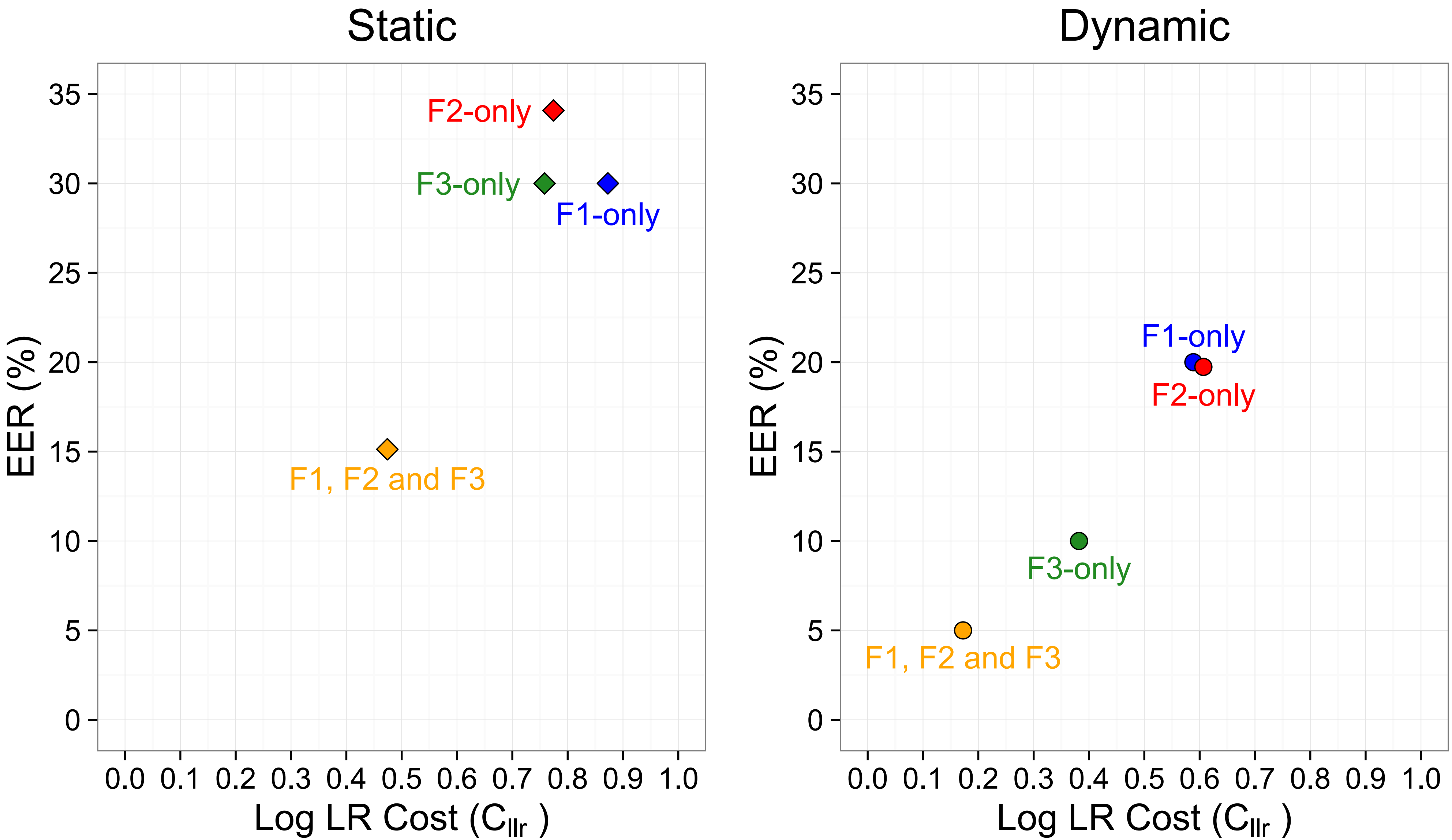
p = probability
E = evidence
| = 'given'
H_p = prosecution hyp
H_d = defence hyp

5. Results: speaker discrimination

- output = $\log_{10}$ LRs:
  - centered on 0 (no evidence)
  - > 0 = support for prosecution
  - < 0 = support for defence
- error metrics:
  - equal error rate (EER)
  - log LR cost function (C_{llr})

Closer to 0, the better the performance

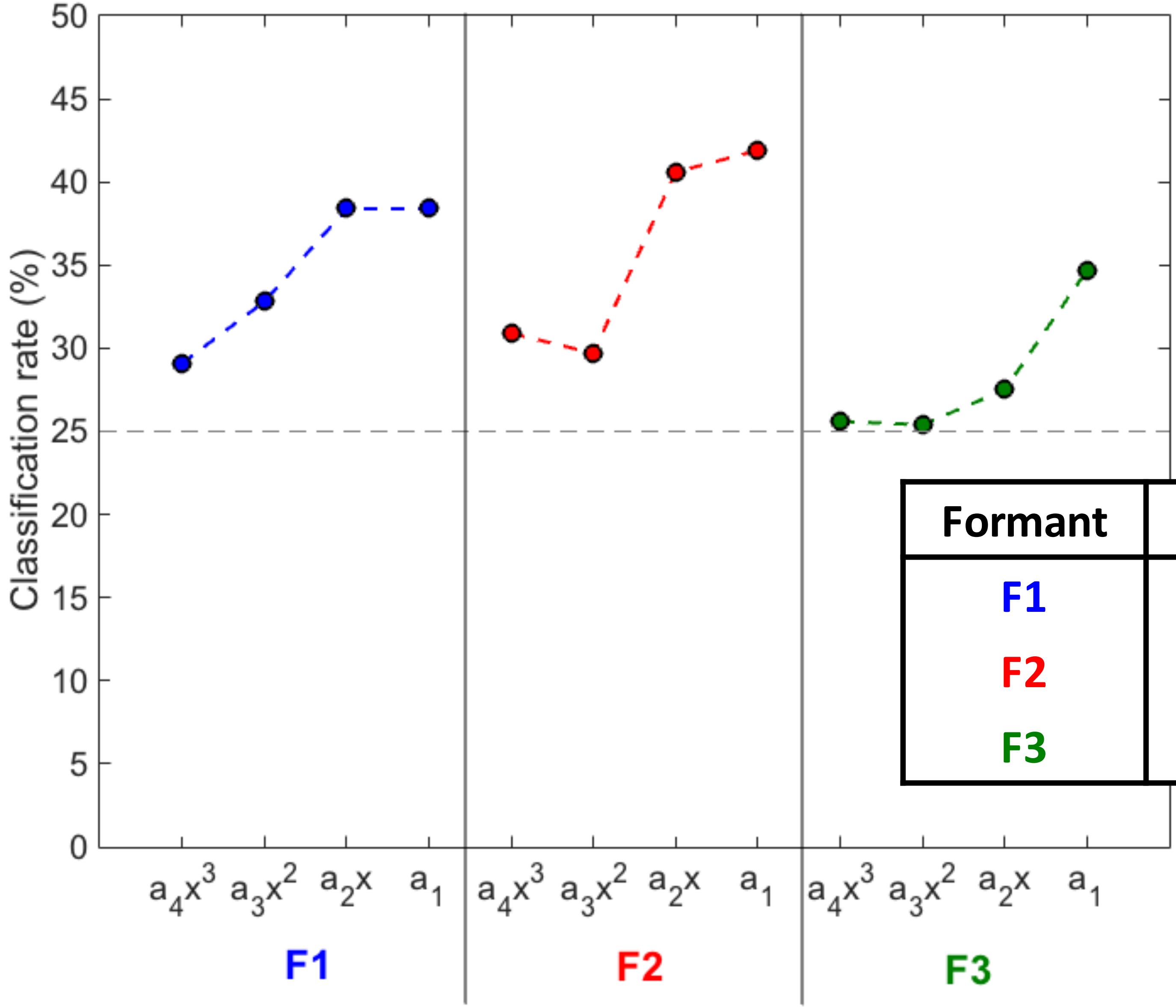
5. Results: speaker discrimination



5. Results: group discrimination

- predicting regional background
- cross-validated discriminant analysis:
  – each token assigned to 1 of 4 regional groups
  – models built on all data excluding target token
- generates classification rate based on posterior probability
- chance = 25% (1/4)

5. Results: group discrimination



Formant	Classification rate
F1	63.8%
F2	64.7%
F3	40.6%

6. Discussion

speaker discrimination

- formant dynamics contain considerable speaker-specific information:
  - better performance than static values
- higher formants = greater speaker-discriminatory power
  - speech-speaker dichotomy (Mokhtari 1998)
  - F1~F2 responsible for contrast

6. Discussion

group discrimination

- group-specific information isn't all about targets
  - individual cubic coefficients capable of predicting regional background above chance
  - all coefficients in combination outperform any one in isolation
- so… fine-grained phonetics clearly shared across speech communities

6. Discussion

- results challenge underlying phonological model for formant dynamics
  - groups = not all about targets
  - Individuals = not all about transitions
- need to rethink the dichotomies:
  - speech-speaker (Mokhtari 1998)
  - group-individual (Garvin & Ladefoged 1963)
  - maybe it's about continua?

7. Conclusion

- formant dynamics capable of encoding **both** speaker- and group-information
  - consistent with usage-based approaches?
- focus on the individual may help us better understand acquisition of variation
  - therefore a role for forensics (methodological and theoretical) in understanding phonology

Thanks!

Questions?

LabPhon 15 Satellite

Speech dynamics, social meaning, and phonological categories

UNIVERSITY of York

13th July 2016

voice and identity

0101101ID1