

Sociolinguistic and acoustic variability in filled pauses

Paul Foulkes

Gareth Carrol & Samantha Hughes

University of York &

JP French Associates, York, UK

pf11@york.ac.uk



R000234892



Res. grant 2003

Introduction and outline

- analysis of filled pauses (**FPs**) - *uh, um*
- structure
 - rationale for the study
 - methodology
 - results
 - implications for forensic speaker ID

Introduction

- forensic speaker identification
- comparison of speech samples:
 - criminal sample + reference sample from suspect
- identify similarities + differences
- conclude whether or not same speaker

Introduction

- **problems**
 - natural variability in speech
 - effects of emotion, alcohol, shouting, telephone...
- thus we need to know:
- what are the most/more robust diagnostics?

Introduction

- FPs may be useful:

‘Individuals tend to be quite **consistent** in using “their” respective personal variant of the hesitation sound, in particular with respect to the **optional** addition of a bilabial **nasal** consonant and the **colour** of the **vocalic** component’

(Künzel 1997)

Introduction

- FPs may present:
 - X-speaker variation in vowel quality
 - X-speaker variation in relative frequency of *uh/um*
- less variable than lexical vowels
 - longer duration than lexical vowels
 - less subject to coarticulation etc

Introduction

- FPs may present:
 - X-speaker variation in vowel quality
 - X-speaker variation in relative frequency of *uh/um*
- less variable than lexical vowels
 - longer duration than lexical vowels
 - less subject to coarticulation etc
- non-linguistic: less likely to be disguised

Introduction

- however...
- few empirical studies

Aims of our study

- analysis of FPs in large corpus
 - distribution of types: *uh* and *um*
 - phonetic qualities of vowel
 - assess acoustic variability in FPs
 - discriminatory power of FP data

Methodology: corpus

- corpus of 32 speakers from Newcastle
- *Phonological Variation & Change*
project (Milroy et al 1997, Docherty & Foulkes 1999...)

Methodology: corpus

- corpus of 32 speakers from Newcastle
- sample design:

'middle class'				'working class'			
16-25		45-65		16-25		45-65	
m	f	m	f	m	f	m	f
4	4	4	4	4	4	4	4

Methodology: recordings

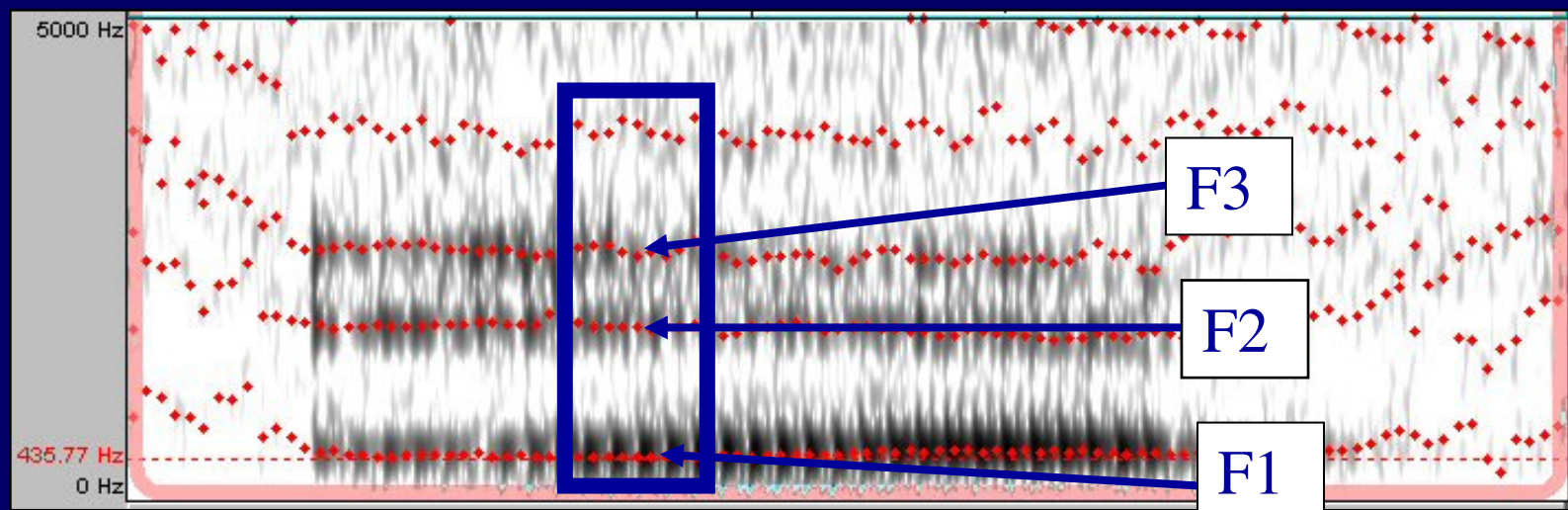
- recordings: 45 min. conversation
 - self-selected pairs
 - mainly single sex
 - spontaneous, natural data
- phonologically balanced word-lists
 - c. 200 items

Methodology: tokens

- for each speaker we extracted:
 - all FPs
 - 30 tokens of lexical vowels / ɪ ɛ a ə /
- N tokens:
 - 1,690 FPs
 - 3,958 lexical vowels

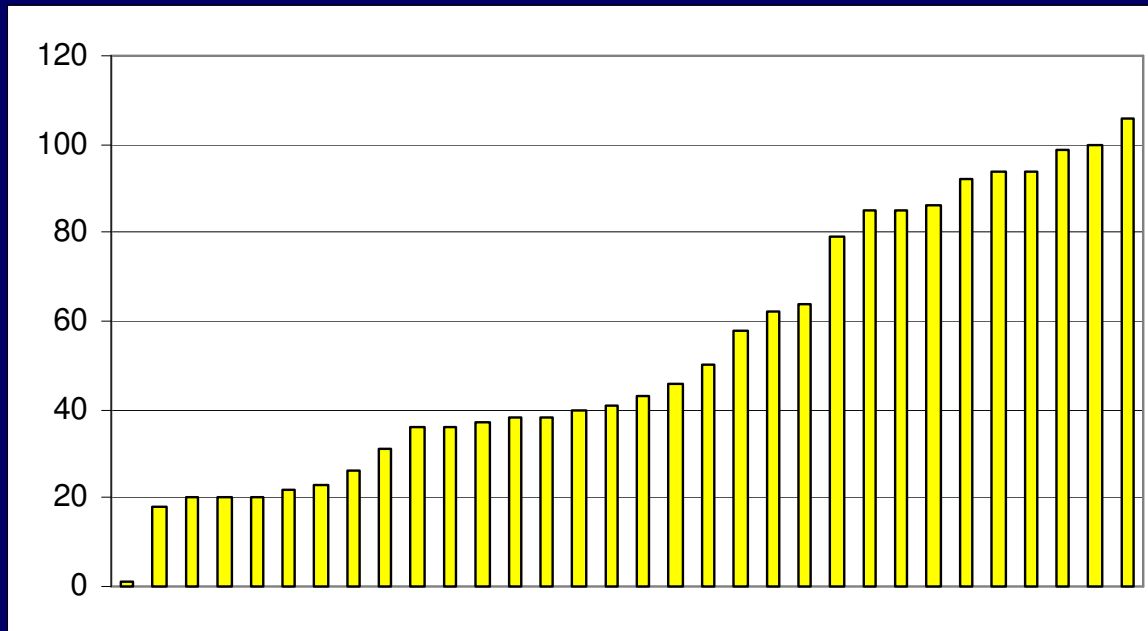
Methodology: analysis

- acoustic qualities of vocalic portions
 - acoustic analysis by *Praat*
 - formant extraction function
 - average over stable section at midpoint



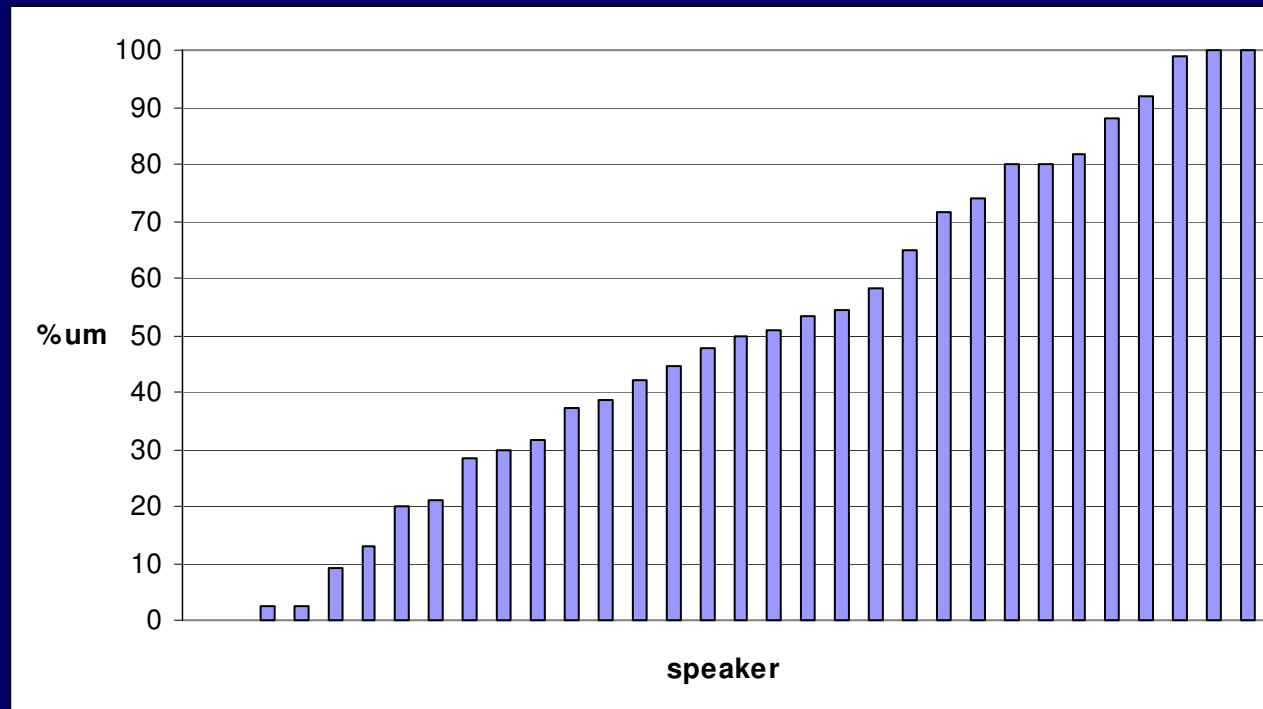
Results 1: pause types

- distribution of FPs
- N tokens per speaker, rank order



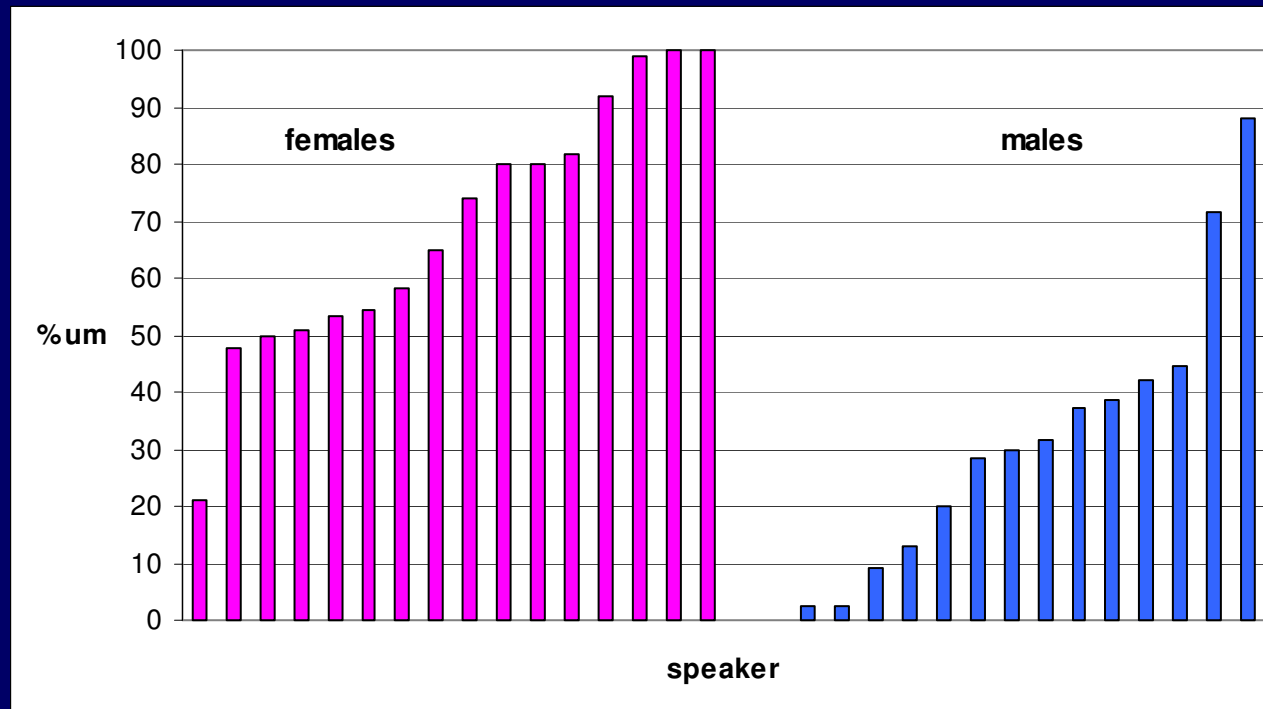
Results 1: pause types

- % *um*: all speakers, rank order



Results 1: pause types

- % *um*: females and males



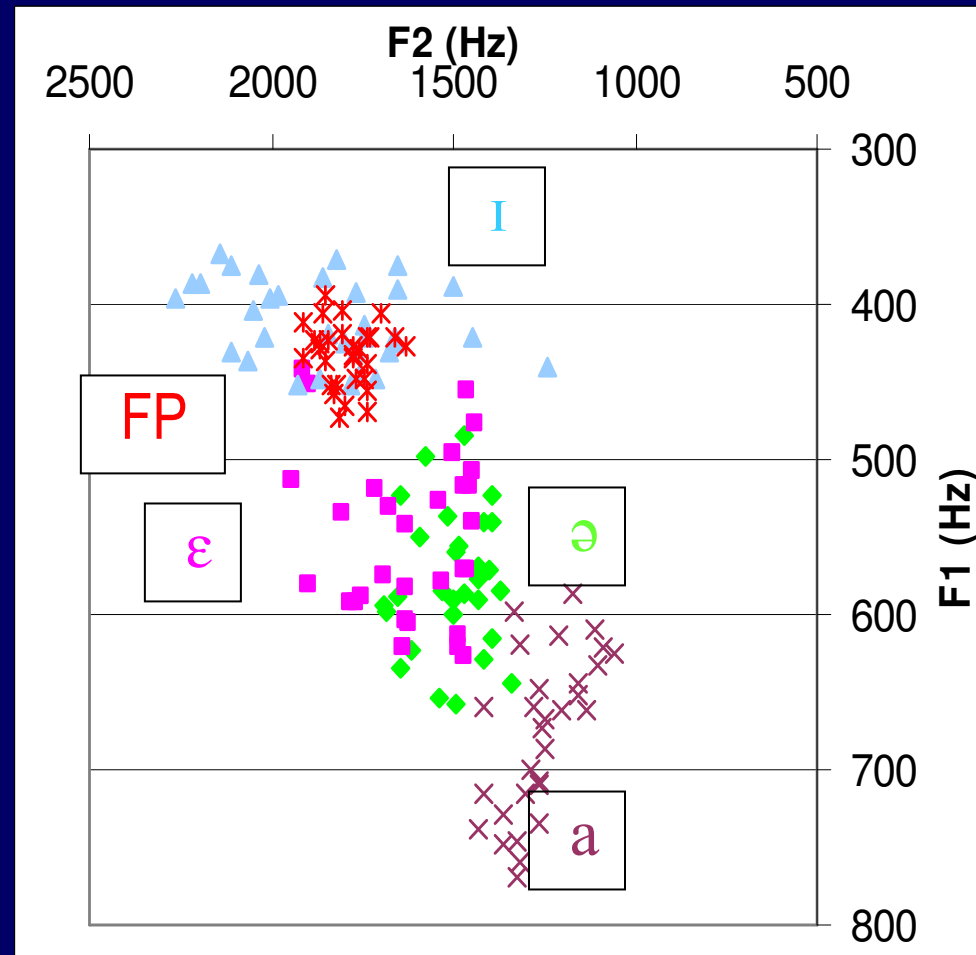
Results 1: pause types

- *um* sig. > *uh* for:
 - females
 - middle class
 - younger speakers

(Anova: $p < .001$ in each case)

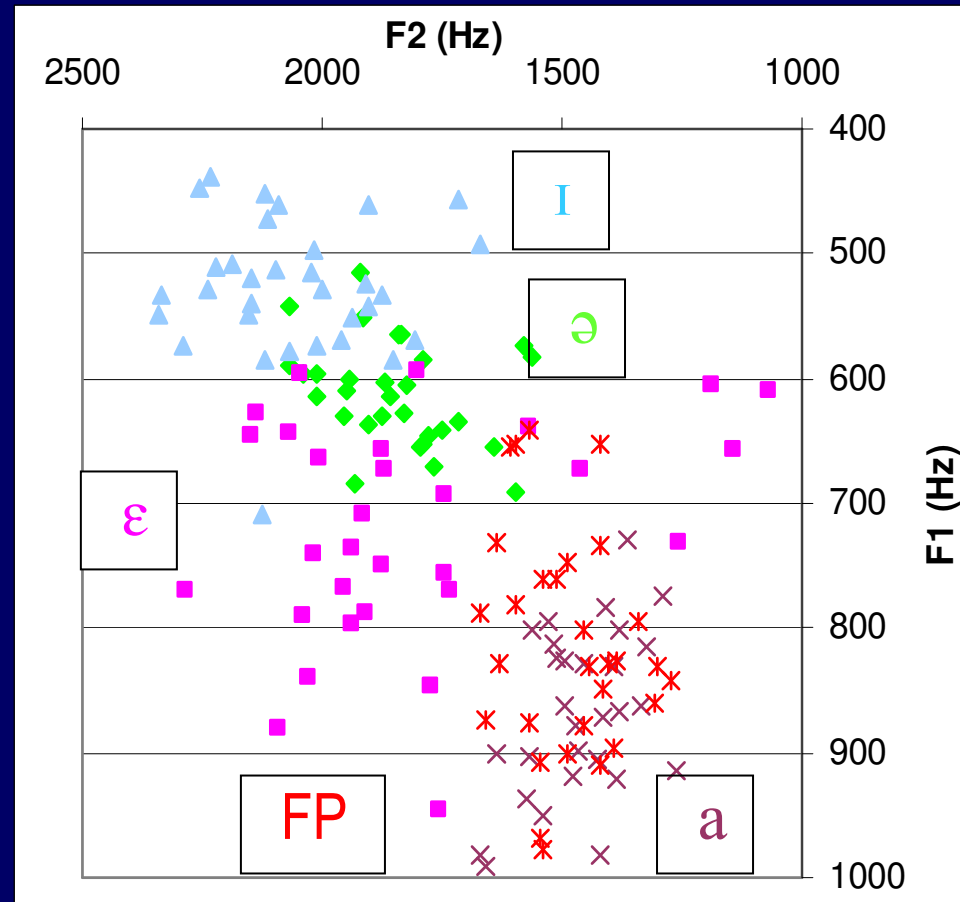
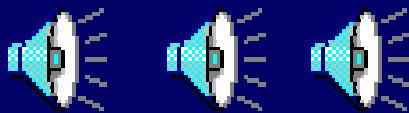
Results 2: vowel quality

- typical pattern
- e.g.:
older male
- close-mid front
- distinct from /ə/



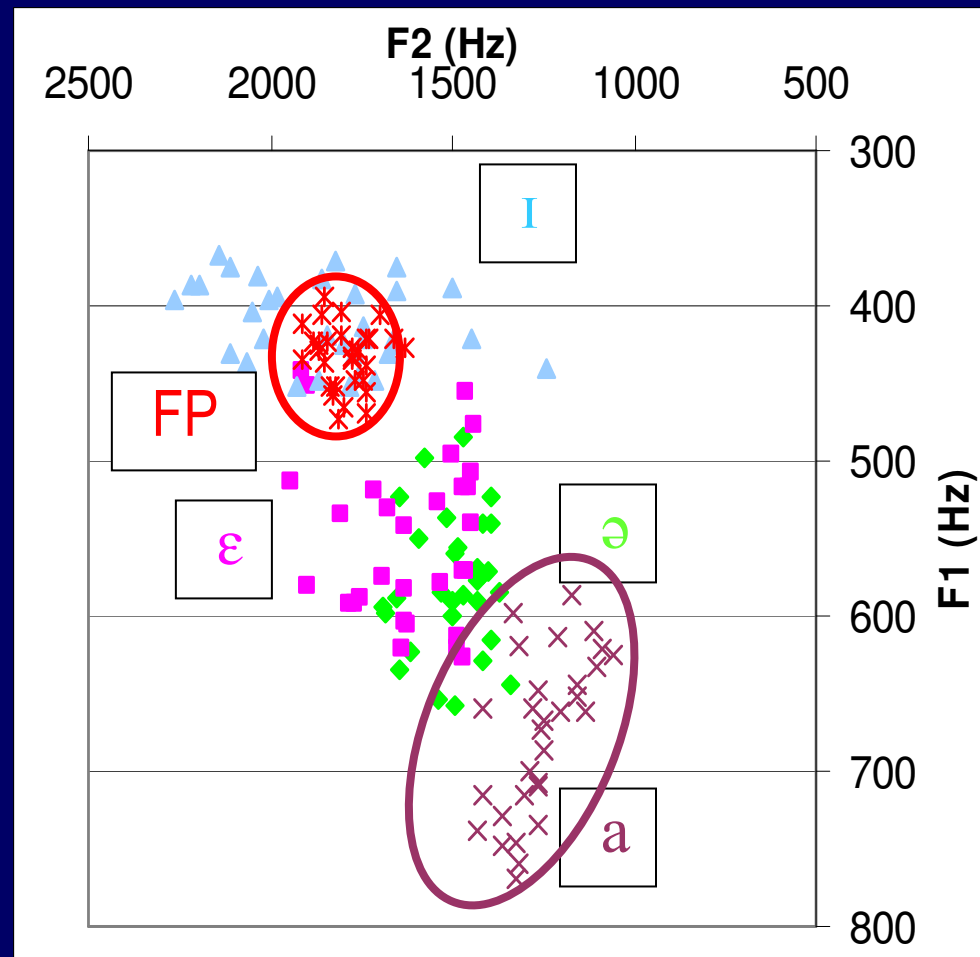
Results 2: vowel quality

- one major exception
- older WC female



Results 3: variance

- are FPs less variable than lexical vowels?
- spread of formant measures
- FPs appear more compact in F1-F2 space



Results 3: variance

- statistical analysis of formant measures
- data for F1, F2, F3 treated separately
- F-ratio to compare variance in s.d.

s.d. of F_n in FP

versus

s.d. of F_n in lexical vowel

Results 3: variance

- if $F < 1$, FP is **less** variable
- if $F > 1$, FP is **more** variable
- illustration: older MC male

	F1		F2		F3	
	F	<i>p</i>	F	<i>p</i>	F	<i>p</i>
θ	.95	ns	1.01	ns	.69	ns
ϵ	1.40	ns	.35	.03	.68	ns
a	.75	ns	3.34	.005	.30	.005

Results 3: variance

- all F1 comparisons ($p < .05$)

<u>% of comparisons</u>	<u>um</u>	<u>uh</u>
FP less variable	18	18
no sig. difference	62	66
FP more variable	20	15

Results 3: variance

- all F2 comparisons ($p < .05$)

<u>% of comparisons</u>	<u>um</u>	<u>uh</u>
FP less variable	22	24
no sig. difference	63	60
FP more variable	15	16

Results 3: variance

- all F3 comparisons ($p < .05$)

<u>% of comparisons</u>	<u>um</u>	<u>uh</u>
FP less variable	36	45
no sig. difference	55	48
FP more variable	10	7

- trend: $F < 1$ in 73% of cases

Results 3: variance

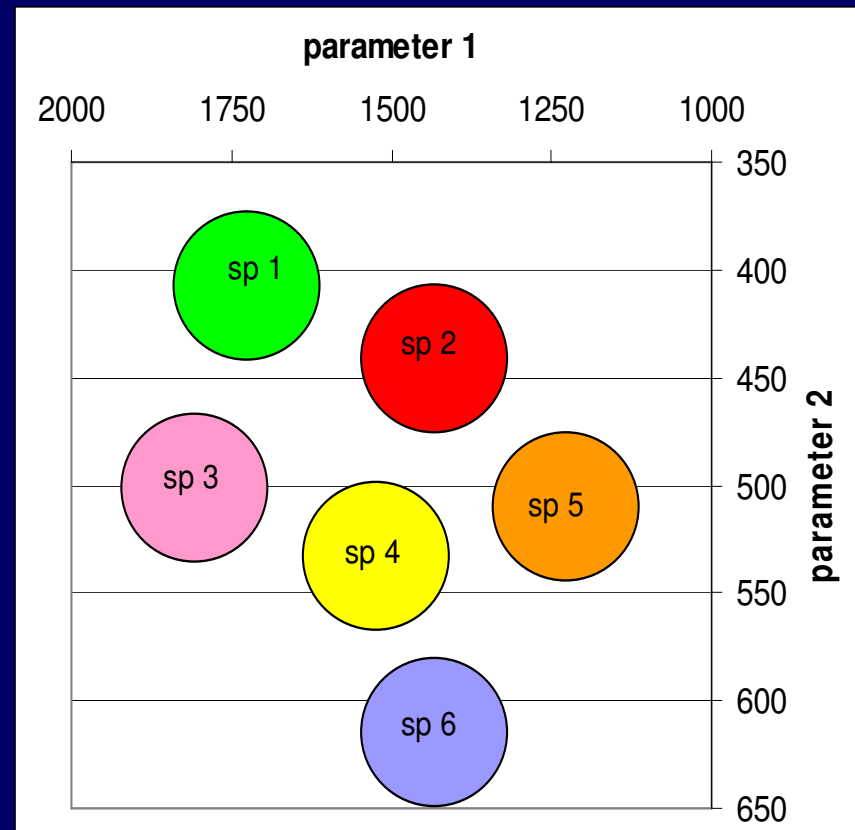
- summary:
 - F3 generally less variable in FPs
 - F1 and F2 are similar to lexical Vs in variance

Results 4: discrimination

- how well do FP measures discriminate between speakers?

Results 4: discrimination

- ideal world:
 - each speaker occupies own area in phonetic space
 - data yield **100% discriminatory power**

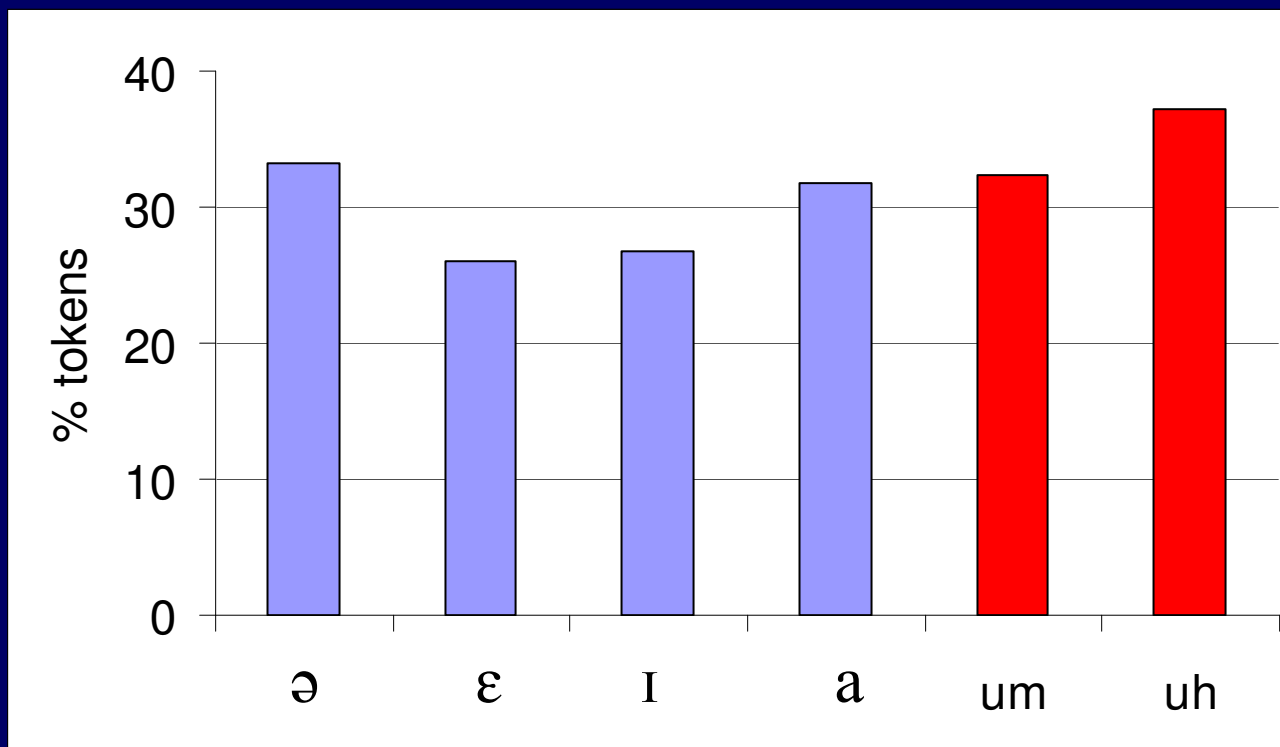


Results 4: discrimination

- discriminant analysis using SPSS
- males and females analysed separately
- F1, F2 and F3 combined
- /ɪ ɛ a ə/, *um*, *uh*

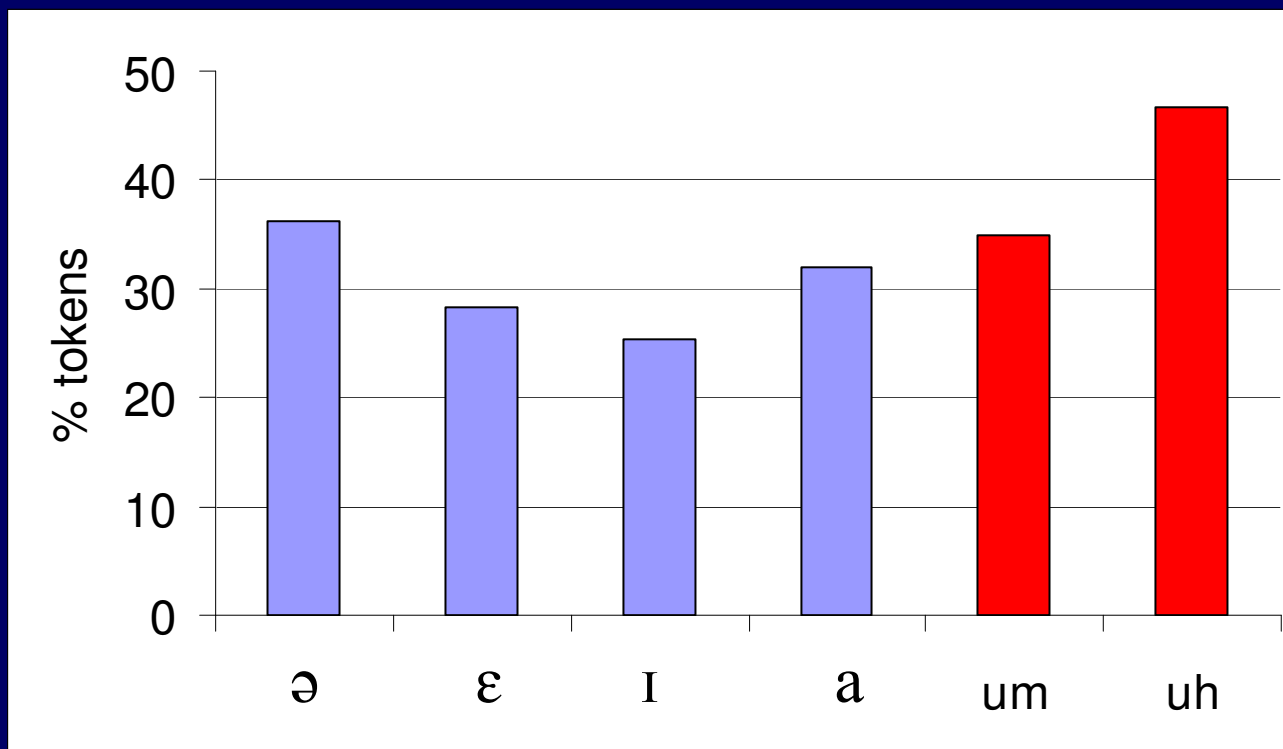
Results 4: discrimination

- males: % tokens correctly discriminated



Results 4: discrimination

- females: % tokens correctly discriminated



Results 4: discrimination

- summary:
 - FP data perform marginally better than lexical vowels

Discussion 1

- pause types: *um* - *uh*
 - individual variability in frequency
 - sig. variation by gender, class, age
 - fewer *um* in male speech

Discussion 1

- reflects different dynamics of discourse
 - *um*: prefaces long pause (Clark & Fox Tree 2001)
 - *uh*: prefaces short pause
 - males tend to be more competitive in turn-taking/control the floor more (Shriberg 2001)
 - males perhaps less tolerant of long pauses
 - requires further examination

Discussion 2

- vowel qualities:
 - generally close-mid front V
- of potential forensic significance:
 - marked individual departure from group norm by one female speaker

Discussion 3

- variability in FPs:
 - qualified support for Künzel et al
 - F3: FPs less variable in 40% of cases
 - discriminant power: at least as good as lexical Vs
- thus FPs moderately more reliable than lexical Vs for forensic speaker ID

Conclusion

- further study:
 - distribution of FP types:
 - different type of discourse?
 - different floor-holding strategies?
 - syntactic position of FPs?
 - other phonetic parameters of FPs:
 - F0
 - duration

fin

Acknowledgements:

IAFPA for Research Grant (2003)
Kirsty McDougall for statistical advice