Voice and Identity: applications and limitations of the voice as a biometric

Colleen Kavanagh, Paul Foulkes, Peter French, Vincent Hughes,
Philip Harrison, & Eugenia San Segundo

University of York,
J P French Associates,
& Royal Canadian Mounted Police

2 October 2017

Outline

1. voice comparison

2. automatic speaker recognition (ASR)
   - principles
   - theoretical limitations
   - practical limitations

3. phonetics and linguistics

4. ongoing research at York

1. Voice comparison

- comparison of voice in criminal recording with voice in recording of known suspect

- assist court with determining identity/non-identity of suspect and criminal

1.1 ‘Cooper’ case example

- case referred to as ‘Cooper’ in Foulkes & French (2012: 564-5)
- theft at care home
- 4 seconds of speech recorded via intercom system
- suspect read version of text

1.1 ‘Cooper’ case example

example: QS (4 seconds)

I’ve come to see the lady at number two

av k^{h}ʊm tsiː ʔ ɬɛɪd jəʔ nʊmbə tsuwːː

(I’m fro)m the Home Care I’ve come to collect her sheet

mʔ oʊm k^{h}ɛːɹ av k^{h}ʊm tə k^{h}əlɛkt ə ʃiːːʔ

2. Automatic speaker recognition (ASR)

- voice of **questioned (criminal) speaker** is put into system
- voice of **suspect** also put into system
- database of voices: **reference population** is put into system

2.1 ASR: principles

- system mathematically reduces speech samples to statistical models reflecting **vocal tract geometry**
- 1 model for suspect, 1 model for criminal sample
- models for reference population
- criminal model is compared with suspect model and reference population models

http://parole.loria.fr

2.1 ASR: principles

- ‘distance’ between criminal and suspect models → **similarity**
- ‘distance’ between criminal and reference population models → **typicality**

- **high similarity** with suspect and **low typicality** → more likely to be **same speaker**
- **low similarity** with suspect and **high typicality** → more likely to be **different speakers**

2.2 ASR limitations: theoretical and practical

2.2.1 theoretical problems with ASR

- inherent limitations of underlying assumption of ASR: **the vocal tract as a biometric**

- vocal tract is probably unique to each speaker

- but limited by:
  - **small differences between speakers**
  - **plasticity**
  - **indirectness**
  - **exogenous influences**

2.3 practical problems with ASR

- recordings are usually degraded in quality
  - channel mismatch (e.g. phone versus direct)
  - recording media (mobile phone, CCTV, poor technology...)
  - acoustic environments (traffic, noise, distance from mic...)
  - UK: Regulation of Investigatory Powers Act (2000) prohibits use of intercepts (phone taps) as evidence

- example: French & Harrison (2010)
  - 767 trials using **real forensic case data** with known outcomes
  - EER = equal error rate (classifies SS as DS; DS as SS)

adequacy rating (given by Batvox system)	**# trials**	**EER**
OK	171/767 (22%)	5.4 %
OK + Warning	369/767 (48%)	15.1 %
All	767/767 (100%)	24.2 %

- to achieve EER of 5%, 78% of cases must be rejected as unsuitable for analysis

2.4 summary

- ASR has great advantages
  - speed, replicability, very good performance in experiments...
- but inherent limitations of vocal tract acoustic output as a biometric
  - relative lack of variability across individuals
  - high variability within individuals
  - these factors yield **greater overlap between speakers**
- does vocal tract output alone really have the potential to discriminate a population of e.g. 16m adult Caucasian males in England?

2.4 summary

- voice is different from most other biometrics: much less fixed, much more subject to within-individual variation

- but sources of variation in speech and language are relatively well understood
  - patterns are usually **principled** not random

- we can exploit knowledge of these patterns to assist in forensic voice comparison

3. Contributions to forensic speaker comparison from phonetics and linguistics

3. Contributions from phonetics

- ASR as one component in a broader approach
  - avoid dependence on a single type of metric
  - seek alternative features for analysis to circumvent inherent problems in ASR
  - stronger evidence where multiple lines of enquiry (independent features) yield consistent conclusions

- incorporate **componential phonetic-linguistic** analysis

3.1 componential phon-ling analysis

- application of standard, largely uncontroversial, analytic techniques from phonetics & linguistics

- views speech signal as complex & divisible, composed of (semi-)independent elements, vs. holistic approach of automatic systems

3.1 componential phon-ling analysis

- syntax/grammar
  - e.g. *I did it ~ I done it*
- morphology (word-structure)
  - e.g. *twenty-five pounds ~ twenty-five pound*
- lexical choices
  - e.g. *twenty-five pounds ~ twenty-five quid ~ pony*
- **phonology** (sound system)
  - e.g. distinction of *look/luck*, *which/witch*
- **phonetics/acoustics** (pronunciation)
  - e.g. /t/ variation: *ge**t** off* with [t – d – ʔ – r]

- numerous components for analysis
  - French et al (2011), Foulkes & French (2012)

feature	notes
Vowels	**English: 24 Vs**; different patterns for specific phonological environments; acoustic features (formant centre frequencies, densities, bandwidths), sociolinguistic variables...
Consonants	**English: 20 Cs**; different patterns for specific phonological environments; energy loci of fricatives and stop bursts; segment durations inc. VOT; sociolinguistic variables...
Vocal setting	Laver VPA scheme: 38 separate elements
Intonation	contours constrained by phonology & discourse
Pitch	mean, range, s.d. ...

- numerous components for analysis
  - French et al (2011), Foulkes & French (2012)

feature	notes
Articulation rate	speed of speech
Rhythm	
Tone	for languages with contrastive tone
Connected speech processes	assimilation, elision...
Discourse/ Pragmatics	discourse markers, turn-taking, telephone openings, code switching...
Non-linguistic	audible breathing, throat-clearing, tongue clicking, filled and silent hesitation phenomena...

‘Cooper’ case example

example: QS (4 seconds)

I’ve come to see the lady at number two

av k^{h}ʊm tsiː ʔ ɬɛɪd jəʔ nʊmbə tsuwːː

(I’m fro)m the Home Care I’ve come to collect her sheet

mʔ oʊm k^{h}ɛːɹ av k^{h}ʊm tə k^{h}əlɛkt ə ʃiːːʔ

‘Cooper’ case example

(some) observable features:

- general Yorkshire accent
- PRICE reduced to monophthong [a] (in both instances of *I’ve)*
- STRUT = typical northern English /ʊ/ (*come, number)*
- schwa fully elided (*to, collect)*
- /t/ = glottal stop in word-final position (*at, sheet)*
- /l/ is ‘dark’ in syllable-onset positions (*lady, collect)*
- despite Yorks accent, FACE & GOAT = diphthongs (*lady, Home)*
- GOOSE and FLEECE are not monophthongal (*two, sheet)*
- final syllable in each speaking turn markedly elongated
- definite article = local northern form [ʔ]
- /h/ is deleted (*Home, her)*
- the speaker is not rhotic *but* uses linking /r/ (*Care I’ve)*

3.1 componential phon-ling analysis

- some general advantages
  - many components robust to channel mismatch
  - can derive rich information from short samples
  - concrete reference: easily expressed in court
  - independence of features: increases depth of analysis, multiple evidence types in combination

4. York research

4.1 Voice and Identity: source, filter, biometric

- ASR seen as useful addition to other components within componential approach
  - **test & strengthen all components** in broader approach

1. what are best speaker-discriminating components?
2. robustness of components to exogenous factors
3. to show for the first time what the relationship is between what the ASR (black box) measures and features linguistic phoneticians examine

voice and identity 0101101ID1

Arts & Humanities Research Council

4.1 Voice and Identity: source, filter, biometric

- Comparison of 3 methods of analysis of vocal tract output using DyViS data (Nolan et al., 2009):

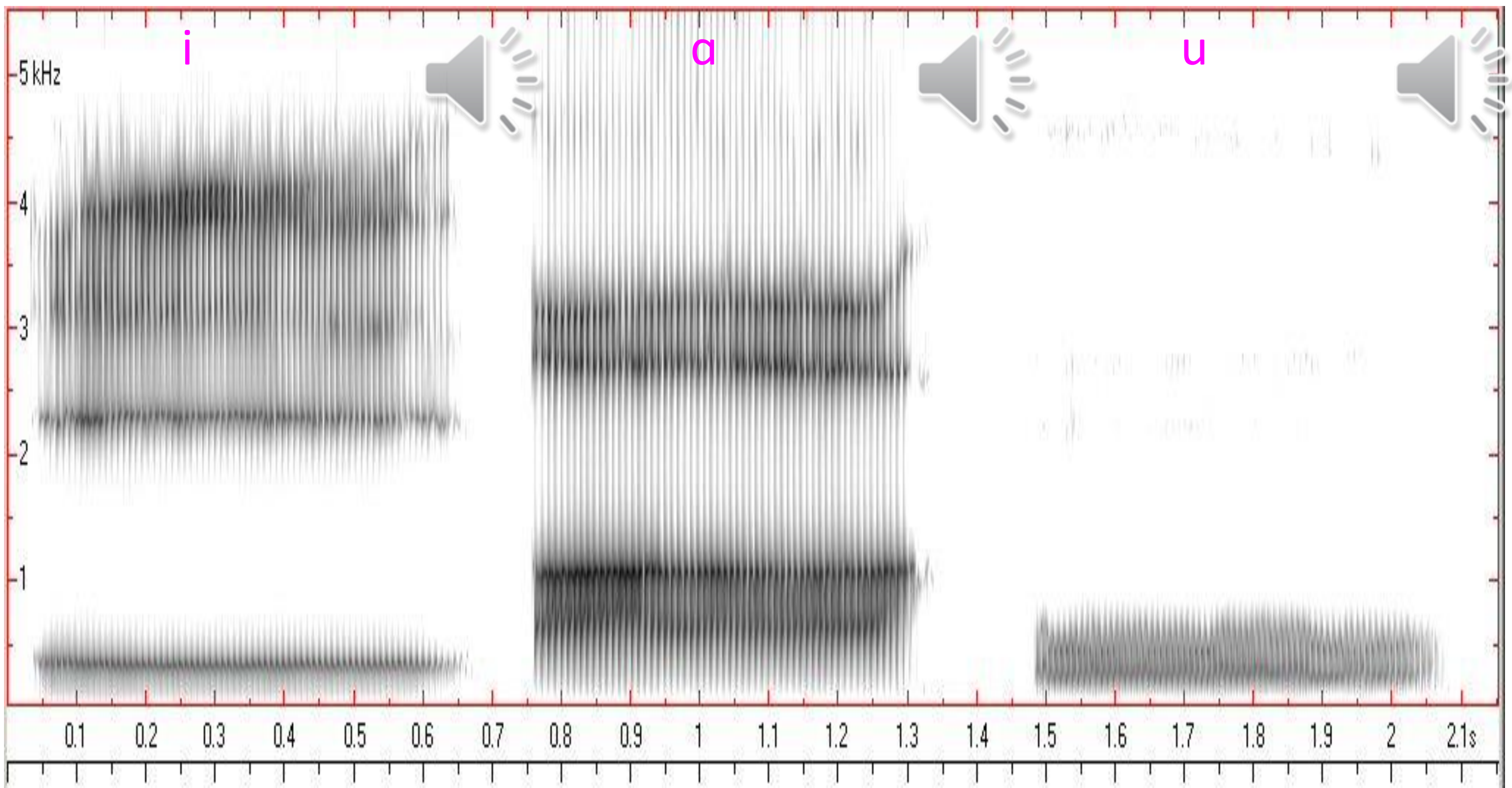
- **Automatic: MFCCs (Batvox)**

- **Semi-automatic: LTFDs**
  - acoustic phonetic

- **Phonetic: voice quality (VPA)**
  - auditory phonetic

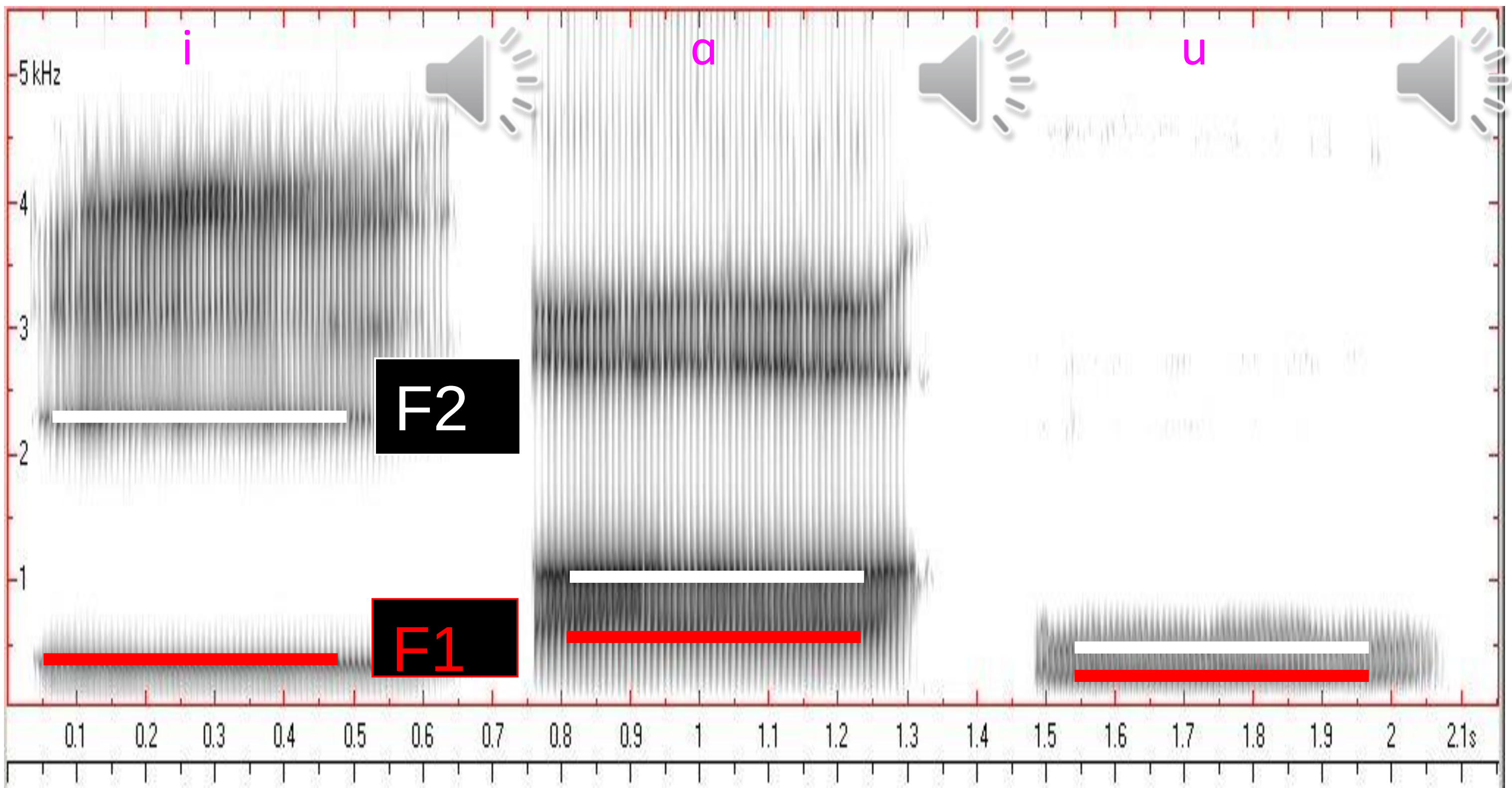
LTFD: acoustic analysis of vowels

- **vowels characterised by configuration of constituent resonances – formants**

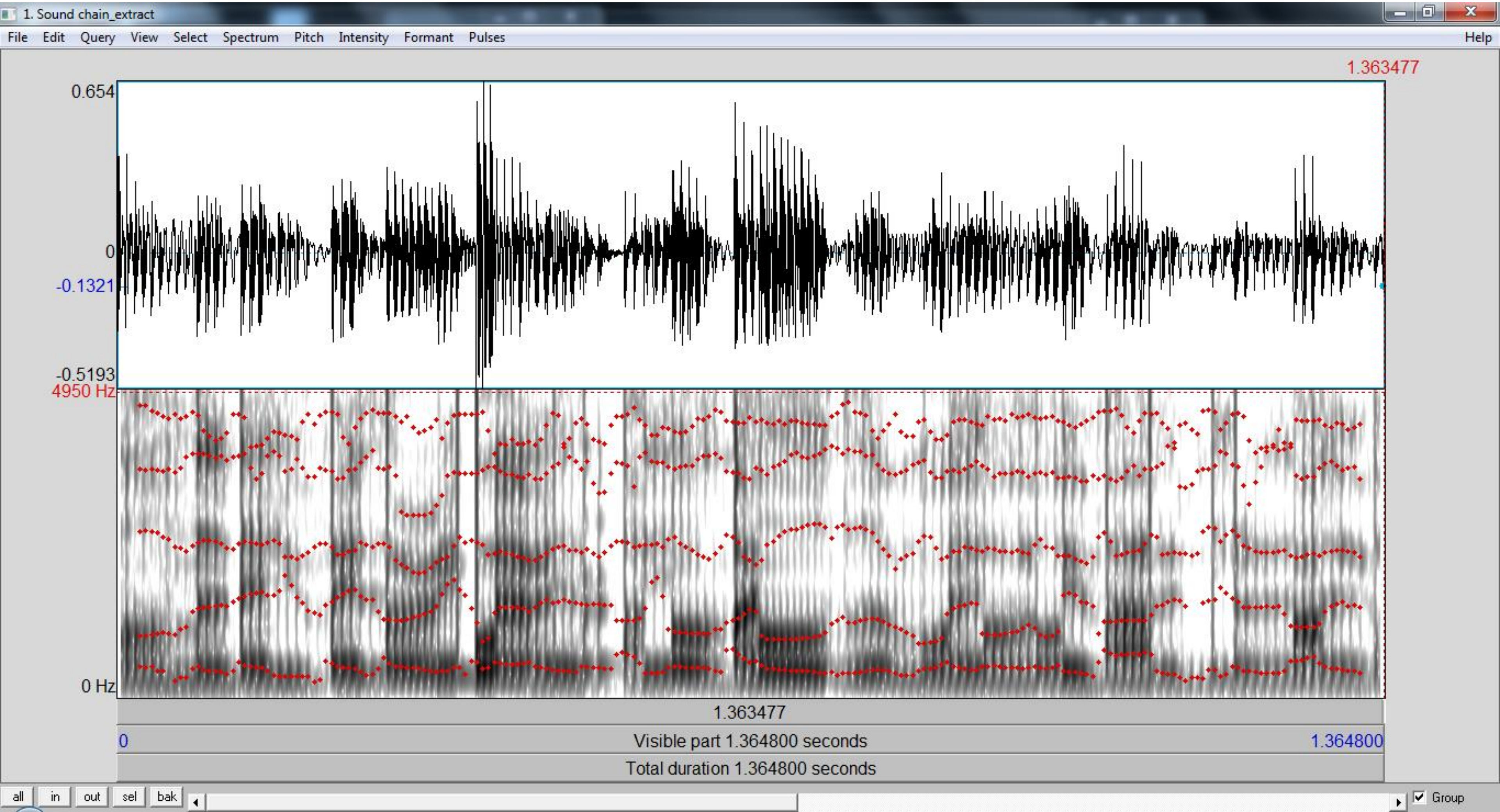


LTFD: acoustic analysis of vowels

- **vowels characterised by configuration of constituent resonances – formants**



- vowels extracted from running speech
- LTFD: long-term formant distribution



Voice quality: Vocal Profile Analysis (VPA)

- Modified VPA protocol (Laver et al 1981; Stevens & French 2012)

Denasal

Nasal

Fronted tongue

Pharyngeal constriction

	SETTING	SCALAR DEGREE			
		Slight	Marked	Extreme	
		1	2	3	NOTES
	A. VOCAL TRACT				
Labial	Lip rounding/protrusion				
	Lip spreading				
	Labiodentalisation				
	Extensive labial range				
	Minimised labial range				
Mandibular	Close jaw				
	Open jaw				
	Extensive mandibular range				
	Minimised mandibular range				
Lingual tip/blade	Advanced tongue tip/blade				
	Retracted tongue tip/blade				
Lingual body	Fronted and raised tongue body				
	Backed and lowered tongue body				
	Extensive lingual range				
	Minimised lingual range				
Pharynx	Pharyngeal constriction				
	Pharyngeal expansion				
Velopharyngeal	Nasal				
	Denasal				
Larynx height	Raised larynx				
	Lowered larynx				
	B. OVERALL MUSCULAR TENSION				
Vocal tract tension	Tense vocal tract				
	Lax vocal tract				
Laryngeal tension	Tense larynx				
	Lax larynx				
	C. PHONATION				
	Falsetto				
	Creaky				
	Whispery				
	Breathy				
	Murmur				
	Harsh				
	Tremor				

4.1 Voice and Identity: source, filter, biometric

- no method perfect, but methods make **different** mistakes

#067 #072

- strong argument for combining methods

4.2 summary

- voice is not like other biometrics (e.g. fingerprints, DNA), but its problems are not insurmountable
- voice has considerable potential as evidence
- overarching aim to move towards unified theoretical position on speaker characterisation
- best means forward is tried and tested phonetic/acoustic methods in combination with ASR
  - Legally problematic?

thank you

https://sites.google.com/a/york.ac.uk/voice-and-identity/

https://sites.google.com/site/yorkfss/home

colleen.kavanagh@york.ac.uk

paul.foulkes@york.ac.uk