



Te Kāhui Roro Reo - New Zealand Institute of Language Brain & Behaviour

UC  UNIVERSITY OF CANTERBURY  
22 RUAKO, HAMILTON & MOTUTU, CHRISTCHURCH NEW ZEALAND

**Voice and Identity:**  
exploring the contribution of voice quality

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

voice and identity  
0101101101



Te Kāhui Roro Reo - New Zealand Institute of Language Brain & Behaviour  
THE UNIVERSITY OF YORK

**overview**

1. forensic voice analysis
2. voice quality analysis
3. methods
4. results
5. voice quality in FVC experiment
6. summary

**Overview**

Te Kāhui Roro Reo - New Zealand Institute of Language Brain & Behaviour

- **Voice & Identity** project, 2015-2018
  - forensic identification of speakers
  - aim: to explore relationships between automatic speaker recognition (ASR) & forensic phonetics
  - compare results on same data
  - assess scope for complementary application

voice and identity  
0101101101



**Overview**

Te Kāhui Roro Reo - New Zealand Institute of Language Brain & Behaviour  
THE UNIVERSITY OF YORK

- this study: analysis of **voice quality/vocal setting**
  1. procedure & reliability
  2. contribution to overall forensic voice analysis
    - added as stage in voice classification experiment
- largely methodological issues
- work in progress, so comments especially welcome!

Te Kāhui Roro Reo - New Zealand Institute of Language Brain & Behaviour  
THE UNIVERSITY OF YORK

**overview**

1. **forensic voice analysis**
2. voice quality analysis
3. methods
4. results
5. voice quality in FVC experiment
6. summary

**1. Forensic voice analysis**

Te Kāhui Roro Reo - New Zealand Institute of Language Brain & Behaviour  
THE UNIVERSITY OF YORK

- context: voice as a biometric
  - voice is a marker of human identity
  - but it's an imperfect biometric
    - within-speaker variation; no feature permanent or unchanging (cf. DNA or fingerprints); technical effects; health & ageing...
    - **no voiceprint** (despite CSI-type claims)

## 1. Forensic voice analysis

Te Kaitiaki Raukawa - New Zealand Institute of Language Brain & Behaviour

The slide shows two cassette tapes. The left one is labeled 'unknown offender' and the right one is labeled 'A known suspect'. Below the tapes, there is a cartoon illustration of a man with a mask and a gun, and another cartoon illustration of a man in a suit sitting at a desk with a typewriter.

## 1. Forensic voice analysis

Te Kaitiaki Raukawa - New Zealand Institute of Language Brain & Behaviour

- three main methods of analysis

**linguistic-phonetic**

**automatic (ASR)**

**semi-automatic (S-ASR)**

## 1. Forensic voice analysis

Te Kaitiaki Raukawa - New Zealand Institute of Language Brain & Behaviour  
THE UNIVERSITY OF YORK

- largely separate fields of research & case practice
  - in UK/Europe/Australia – forensic phonetics
  - in US – ASR (automatic speaker recognition)
- developments of integration
  - NIST evaluations of ASR + human component
  - lab practices in UK, Germany, Sweden...
  - a few research papers in forensic phonetics

## 1. Forensic voice analysis

Te Kaitiaki Raukawa - New Zealand Institute of Language Brain & Behaviour  
THE UNIVERSITY OF YORK

- pros and cons

	phonetic	ASR
mapping to concrete entities	✓	(x)
explainable in court	✓	x
time & effort	x	✓
robust to channel	(✓)	xx
objectivity	(x)	✓
quantify strength of evidence	(x)	✓
quantify error rates	x	✓
works with limited/poor materials	(✓)	(x)

- improve forensic voice analysis via combination?

## overview

Te Kaitiaki Raukawa - New Zealand Institute of Language Brain & Behaviour  
THE UNIVERSITY OF YORK

- forensic voice analysis
- voice quality analysis**
- methods
- results
- voice quality in FVC experiment
- summary

## 2. Voice quality analysis

Te Kaitiaki Raukawa - New Zealand Institute of Language Brain & Behaviour  
THE UNIVERSITY OF YORK

- forensic phonetic analysis
  - componential
    - vowels, consonants, f0, lexis, syntax, VQ...
  - mainly standard analytic methods
    - formants, durations, f0 range & mean...

The diagram shows a waveform and a spectrogram. The waveform is labeled 'Hello world. Goodbye.' and the spectrogram is labeled 'Hello world. Goodbye.'.

## 2. Voice quality analysis



- voice quality/vocal setting
- definition: Laver (1994)



'the tendency underlying the production of the chain of segments in speech towards maintaining a particular configuration or state of the vocal apparatus'

## 2. Voice quality analysis



- apparently widely used (Gold & French 2011)
  - thus surprising how little published work!
- multiple frameworks for analysis
- but far from easy to operate

## 2. Voice quality analysis



- Laver framework
  - originally developed for clinical use
- modified at J P French Associates for forensic casework

Visual Profile Analysis Protocol

Judge: \_\_\_\_\_ Date of Analysis: \_\_\_\_\_ Speaker: \_\_\_\_\_ Sex: \_\_\_\_\_ Age: \_\_\_\_\_

1. VOCAL QUALITY FEATURES

Category	FIRST PASS		SECOND PASS	
	Neutral	Non-neutral	Neutral	Non-neutral
A. Supralaryngeal Features				
1. Labial				
2. Mandibular				
3. Lingual				
4. Lingual Body				
5. Velopharyngeal				
6. Pharyngeal				
7. Laryngeal				
8. Laryngeal Body				
9. Laryngeal Tip				
10. Laryngeal Base				
11. Laryngeal Tip/Blade				
12. Laryngeal Tip/Blade				
13. Laryngeal Tip/Blade				
14. Laryngeal Tip/Blade				
15. Laryngeal Tip/Blade				
16. Laryngeal Tip/Blade				
17. Laryngeal Tip/Blade				
18. Laryngeal Tip/Blade				
19. Laryngeal Tip/Blade				
20. Laryngeal Tip/Blade				
21. Laryngeal Tip/Blade				
22. Laryngeal Tip/Blade				
23. Laryngeal Tip/Blade				
24. Laryngeal Tip/Blade				
25. Laryngeal Tip/Blade				
26. Laryngeal Tip/Blade				
27. Laryngeal Tip/Blade				
28. Laryngeal Tip/Blade				
29. Laryngeal Tip/Blade				
30. Laryngeal Tip/Blade				
31. Laryngeal Tip/Blade				
32. Laryngeal Tip/Blade				
33. Laryngeal Tip/Blade				
34. Laryngeal Tip/Blade				

## 2. Voice quality analysis



CATEGORY	FIRST PASS			SECOND PASS						
	Neutral	Non-neutral	Abnormal	SETTING	Scalar Degrees					
					Normal		Abnormal			
					1	2	3	4	5	
A. Supralaryngeal Features										
1. Labial				Lip Rounding/Protrusion						
				Lip Spreading						
				Labiodentalisation						
2. Mandibular				Extensive Range						
				Minimised Range						
				Close Jaw						
				Open Jaw						
				Protruded Jaw						
				Extensive Range						
3. Lingual				Minimised Range						
				Advanced						
				Retracted						
4. Lingual Body				Fronted Body						
				Backed Body						
				Raised Body						
				Lowered Body						
				Extensive Range						
				Minimised Range						
5. Velopharyngeal				Nasal						
				Audible Nasal Escape						
				Denasal						

## 2. Voice quality analysis



- illustrative recordings (Laver 1980)

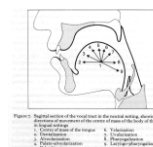
**learning to speak well is an important and fruitful task**

modal (normal)		creak	
raised larynx		falsestto	
nasal		denasal	

## 2. Voice quality analysis



- Laver VPA framework
- componential, organised ~ vocal tract
- 34 physical descriptors ('settings')
  - lip rounding, denasal, lowered tongue body, tense larynx
- but perceptual judgment
- orientation to 'neutral setting'
- annotate departures from neutral



## 2. Voice quality analysis



- main obstacles to widespread use in PVC (Nolan 2005)
  - (1) lack of training in use of the VPA
  - (2) high variability of VQ settings within a sample
    - largely unexplored intra-speaker variability
    - known effects for emotion, health etc
  - (3) VQ can be compromised
    - distorting effect of telephone or background noise, typical of forensic recordings

## 2. Voice quality analysis



- (some) further problems:
  - lack of simple acoustic correlates of physical settings or perceptual categories
  - lack of (reliable) software for automation
  - interaction of dimensions
  - very few published studies of inter-rater agreement
  - fairly poor results where published (Kreiman & Gerratt 1998, Webb et al 2004, Beck 2005)

## 2. Voice quality analysis



- hence this study...
- modified Laver scheme used by 3 analysts to test:
  - between-rater reliability
  - within-rater reliability
  - procedural issues in using the tool
  - potential contribution to overall analysis with ASR

## overview

1. forensic voice analysis
2. voice quality analysis
- 3. methods**
4. results
5. voice quality in FVC experiment
6. summary

## 3.1 Methods: corpus



- DyViS corpus (Dynamic Variability in Speech)
  - Nolan et al (2009)
  - 100 young RP men (Cambridge students)
  - simulated police interview & phone call with 'accomplice'
  - homogeneous!



## 3.1 Methods: corpus



- DyViS sample for VQ analysis
  - 99 young RP men
  - Task 2 – near end of phone call
  - various pre-processing steps taken in *Voice & Identity*
    - i.e. not original DyViS files
- examples

#067



#072

### 3.2 Methods: analysis



- 3 analysts (PF, JPF, ESS) used modified Laver VPA
- 10 voices selected as pilot
- calibration process before full analysis

### 3.3 Methods: modified VPA



- based on JP French version
- removed clinical scale points 4-6
- removed purely clinical settings
  - e.g. audible nasal escape, protruded jaw
- combined settings with no clear distinction
  - e.g. creak/creaky, whisper/whispery
- discarded 'intermittent' categorisation
  - by definition not long term feature

version used for  
analysis

(based on JP  
French version)

	SETTING	Scale Degree			Notes
		1	2	3	
		1	2	3	
A. VOCAL TRACT FEATURES					
Labial	Lip rounding/protrusion				
	Lip tension				
	Labiodentalization				
Alveolar	Alveolar spread				
	Alveolar nasal range				
	Alveolar nasal range				
Nasal	Chin up				
	Chin down				
	Protrusion				
Uvular	Uvular spread				
	Uvular nasal range				
	Uvular nasal range				
Uvular body	Uvular spread				
	Uvular nasal range				
	Uvular nasal range				
Pharynx	Pharyngeal spread				
	Pharyngeal nasal range				
	Pharyngeal nasal range				
Larynx	Laryngeal spread				
	Laryngeal nasal range				
	Laryngeal nasal range				
B. OVERALL MUSCULAR TENSION					
Vocal tract tension	Vocal tract tension				
	Vocal tract tension				
	Vocal tract tension				
Laryngeal tension	Laryngeal tension				
	Laryngeal tension				
	Laryngeal tension				
C. PHONATION FEATURES					
	SETTING	Scale Degree			Notes
		1	2	3	
Labial	Labial				
	Labial				
	Labial				
Alveolar	Alveolar				
	Alveolar				
	Alveolar				
Nasal	Nasal				
	Nasal				
	Nasal				
Uvular	Uvular				
	Uvular				
	Uvular				
Pharynx	Pharynx				
	Pharynx				
	Pharynx				
Larynx	Larynx				
	Larynx				
	Larynx				

### 3.3 Methods: modified VPA



- pilot & calibration
- 10 voices
- same hardware and software
- blind analysis
- as much time as needed

### 3.3 Methods: modified VPA



- findings and issues
- segments vs. settings
  - labiodentalisation ... on /r/ only?
- scale
  - what's marked, extreme...?
- judgment baseline
  - neutral setting, or dialect norm?

### 3.3 Methods: modified VPA



- findings and issues
- agreed distinctions, e.g.
  - breathy vs. whispery (based on degree of voicing)
- tense/lax vocal tract to cover overall laxity
  - but a characteristic of SSBE is combination of lenited unstressed vowels/syllables & precise consonants

### 3.3 Methods: modified VPA



- after pilot & calibration
- analysed all voices, inc. redoing the 10 pilot voices

### overview

- forensic voice analysis
- voice quality analysis
- methods
- results
- voice quality in FVC experiment
- summary

### 4. Results of VPA analysis



- 3 analysts' data compared side-by-side
- collective reanalysis where
  - differences in >1 scalar degree (e.g. ratings 3 – 1 – 1)
  - differences in presence/absence (e.g. 0 – 0 – 1)
- corrected any clear errors
- developed agreed final VPA version for comparison with other forms of analysis (ASR etc)

### 4. Results of VPA analysis



- assessed **inter-rater** agreement via
  - % agreement, **absolute** and **within 1 scalar degree**
  - Fleiss' **kappa** values (chance-corrected measure)
- assessed **intra-rater** agreement in separate experiment (Klug, IAFPA 2017)

### 4.1 inter-rater results



Setting	absolute agreement (%)				agreement within 1 scalar degree (%)			
	ES-PF	ES-JPF	JPF-PF	mean	ES-PF	ES-JPF	JPF-PF	mean
lip rounding	96	96	100	97	96	96	100	97
lip spreading	94	95	95	95	94	95	95	95
labio-dentalisation	98	100	98	99	98	100	98	99
extensive labial range	100	100	100	100	100	100	100	100
minimised labial range	100	100	100	100	100	100	100	100
close jaw	96	96	100	97	96	96	100	97
open jaw	100	100	100	100	100	100	100	100
ext. mandibular range	99	99	100	99	99	99	100	99
min. mandibular range	96	96	98	97	98	98	98	98
advanced tongue tip	55	56	66	59	60	73	78	73
retracted tongue tip	92	99	92	94	93	99	92	95
fronted tongue body	33	43	31	36	51	69	62	60
backed tongue body	97	97	100	98	97	97	100	98
ext. lingual range	98	99	99	99	100	100	100	100
min. lingual range	98	98	100	99	99	99	100	99
pharyngeal constriction	97	95	98	97	98	97	99	98
pharyngeal expansion	97	98	97	97	99	100	99	99

### 4.1 inter-rater results



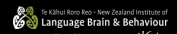
Setting	absolute agreement (%)				agreement within 1 scalar degree (%)			
	ES-PF	ES-JPF	JPF-PF	mean	ES-PF	ES-JPF	JPF-PF	mean
nasal	43	36	49	43	66	75	75	72
denasal	90	87	92	90	91	88	93	91
raised larynx	78	73	71	74	85	84	79	82
lowered larynx	62	70	71	67	72	79	79	76
tense vocal tract	53	55	59	55	75	65	66	68
lax vocal tract	66	55	58	59	76	65	71	70
tense larynx	69	66	68	67	74	80	74	76
lax larynx	66	69	51	62	71	85	58	71
falsehood	100	100	100	100	100	100	100	100
creaky	42	37	59	46	80	79	85	81
whispery	90	94	88	91	95	98	95	96
breathy	49	42	64	52	72	77	85	78
murmur	99	100	99	99	100	100	100	100
harsh	75	74	76	75	84	80	84	82
tremor	100	100	100	100	100	100	100	100
Overall rate				76				82

## 4.1 inter-rater results



- impressive enough...
- but many dimensions never/hardly used
  - hence 100% agreement = 100% avoidance!
- Fleiss kappa scores control for chance and categorical patterns in the data...

## 4.1 inter-rater results



Setting	Fleiss' Kappa	agreement class
murmur	0.83	'almost perfect'
ext. lingual range	0.77	
pharyngeal expansion	0.59	'moderate'
min. mand. range	0.53	
whispery	0.53	
min. lingual range	0.49	
pharyngeal constriction	0.49	
raised larynx	0.46	
harsh	0.43	
lowered larynx	0.41	

## 4.1 inter-rater results



Setting	Fleiss' Kappa	agreement class
advanced tongue tip	0.35	'fair'
tense larynx	0.34	
lax larynx	0.31	
creaky	0.31	
breathy	0.31	
lax vocal tract	0.29	
denasal	0.22	
tense vocal tract	0.22	
lip spreading	0.18	'slight'
nasal	0.13	
fronted tongue body	0.01	
others		undefined

## 4.2 VQ setting correlations



Correlated VPA settings		Spearman's r	Contingency Coefficient
raised larynx	tense larynx	.62	.58
harsh	tense larynx	.36	.57
lax larynx	lowered larynx	.57	.52
creaky	lax larynx	.46	.45
advanced tongue tip	fronted tongue body	.38	.41
creaky	lowered larynx	.35	.35
creaky	whispery	-.36	.37
lowered larynx	tense larynx	-.47	.46
creaky	raised larynx	-.43	.44
lax larynx	raised larynx	-.51	.47
lowered larynx	raised larynx	-.55	.51
lax larynx	tense larynx	-.66	.57
lax vocal tract	tense vocal tract	-.73	.61

## 4.2 VQ setting correlations



- aligns well with predictions made by e.g. Laver (1994), Esling (p.c.)
- e.g. larynx constriction is fundamental to speech production
  - constriction = raising, tensing, compressing vocal folds
  - thus predict larynx raising ~ tense larynx ~ harsh phonation

## 4.3 intra-rater agreement



- preliminary data from Klug (2017)
- assembled samples (61 voices) from DyViS where our agreed VPAs scored at 0 or +3/+2
- 3 analysts asked to judge
 

*Is the VQ feature x as a long term feature present to a marked/extreme degree in this voice?*





Te Kaitiaki Raukawa - New Zealand Institute of Language Brain & Behaviour  
THE UNIVERSITY of York

## overview

1. forensic voice analysis
2. voice quality analysis
3. methods
4. results
5. voice quality in FVC experiment
6. summary

Te Kaitiaki Raukawa - New Zealand Institute of Language Brain & Behaviour  
THE UNIVERSITY of York

## 5. VQ in FVC experiment

- can we use VQ to improve overall success of FVC?
- brief experiment (Hughes et al, *Interspeech* 2017)

Te Kaitiaki Raukawa - New Zealand Institute of Language Brain & Behaviour  
THE UNIVERSITY of York

## 5. VQ in FVC experiment

- ASR systems tested on long-term suprasegmental properties
- variables for comparison =
  - automatic: MFCCs**
    - Mel scale = captures human perceptual system
    - cepstrum = inverse of log power spectrum
    - in theory, **decouples source and filter** information, leaves only supralaryngeal information
  - semi-automatic: long-term formants (F1-F4)**

Te Kaitiaki Raukawa - New Zealand Institute of Language Brain & Behaviour  
THE UNIVERSITY of York

## 5. VQ in FVC experiment

- paired sample tests (Same/Different speaker)
  - system trained on subset of data
  - correct/incorrect scores calculated for new data

'suspect'	'offender'	classification
1	1	SS
1	2	DS
1	$n$	DS
...	...	
2	2	SS
2	3	DS
$n$	$n/n'$	

Te Kaitiaki Raukawa - New Zealand Institute of Language Brain & Behaviour  
THE UNIVERSITY of York

## 5. VQ in FVC experiment

- best system overall – excellent, in fact...
  - MFCCs+ $\Delta$ s+ $\Delta\Delta$ s** fused with **LTFDs**
  - EER (equal error rate) = 3.23%
  - $C_{llr}$  (log LR cost) = 0.137
- 14 errors, 13 false hits (DS categorised as SS)
- 9 involved same speakers, #067 & #072

} low is good!!

Te Kaitiaki Raukawa - New Zealand Institute of Language Brain & Behaviour  
THE UNIVERSITY of York

## 5. VQ in FVC experiment

- speakers **#067** and **#072** mystery speaker
- our agreed VPAs

		S 067	S 072
supra-laryngeal	advanced tongue tip	0	1
	fronted tongue body	1	1
	nasal	1	1
	tense vocal tract	1	0
laryngeal	<b>lax larynx</b>	<b>1</b>	<b>1</b>
	<b>creaky</b>	<b>1</b>	<b>2</b>
	<b>breathy</b>	<b>1</b>	<b>2</b>
	<b>harsh</b>	<b>1</b>	<b>0</b>

## 5. VQ in FVC experiment



- speakers **#067** and **#072**
  - fairly typical supralaryngeal VQ profiles
  - non-neutral for:
    - advanced tongue tip, fronted tongue body, nasality**
  - easily confused with other speakers?
- speakers with highly typical supralaryngeal VQ profiles should be those that the automatic system has difficulty separating from other speakers

## 5. VQ in FVC experiment



- error pairs presented to two forensic caseworkers for blind auditory analysis
- both classified all pairs correctly
- focused on **phonation** properties
- in theory phonation/source info deleted from ASR coefficients
- suggests VQ/phonation info of vital importance in integrated approach to FVC

## overview

1. forensic voice analysis
2. voice quality analysis
3. methods
4. results
5. voice quality in FVC experiment
6. summary

## 6. Summary



- increasing moves toward integration of ASR and forensic phonetics
- contribution of VQ still work in progress
  - more representative data sample of real value
- but some promise in VQ as reliable and consistent method of analysis, esp. for extreme properties
- experimental evidence shows improvement via combination of VQ and ASR/LTFD analysis
- VPA protocol to be understood as one tool in the forensic phonetician's toolbox



thanks, kia ora, ta

questions?

voice and identity

0101101101



paul.foulkes@york.ac.uk

voice and identity

0101101101



## 1.2 ASR

- signal divided into frames (e.g. 10 ms)
  - extract features from each frame
- standard approach: MFCC
  - Mel freq. cepstral coefficients

