The vocal tract as a biometric: output measures, interrelationships, and efficacy

Peter French, Paul Foulkes, Philip Harrison, Vincent Hughes, Eugenia San Segundo & Louisa Stevens

University of York & J P French Associates

Discussant Session:
Forensic phonetics and speaker characteristics

voice and identity

ICPhS2015

Arts & Humanities Research Council

1. Introduction

1. Introduction

voice and identity

- forensic voice comparison (FVC)
  - 400-500 cases per year in UK
- **Voice and Identity: source, filter, biometric**
  - best way to discriminate between speakers
  - best variables
  - best method(s): phonetic, acoustic, ASR…
- **starting point:** vocal tract output (VTO) measures
  - vocal tract as a biometric

1. Introduction

VTO measures

- **vocal profile analysis** (VPA; Laver et al. 1981)
  - auditory analysis
  - 27 supralaryngeal features
- **long-term formant distributions** (LTFDs)
  - *global* analysis of formant distributions across a sample
  - information about vowel system and space
- **mel-frequency cepstral coefficients** (MFCCs)
  - *global* variables extracted from across a sample
  - developed in ASR

1. Introduction

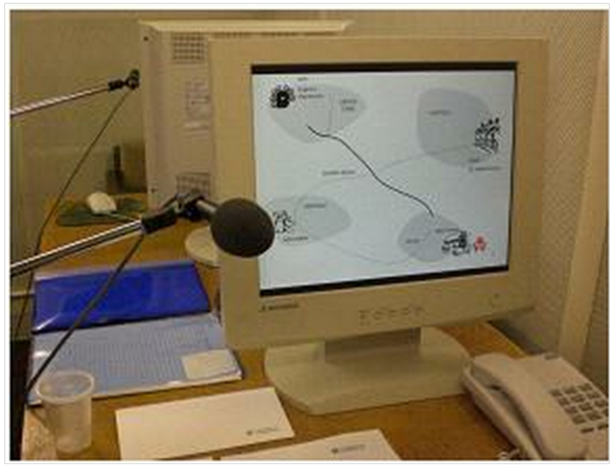
aims

- investigate the interrelationships between these supralaryngeal VTO measures

- investigate the relative discriminant power and limitations of the three methods

2. Data and Methods

2.1 Corpus



DyViS

- DyViS (Nolan et al. 2009)
  - 100 male speakers
  - Standard Southern British English (RP)
  - 18-25 years old
- Task 2 studio (near-end) recordings
  - information exchange task over telephone
  - 44.1kHz/ 16-bit depth audio
  - 10-15 minutes in duration
  - manually edited (silences removed, 4 min samples…)

2.2 Method

- extraction of data for the three measures

- for each measure:

(a) **distances** (degree of divergence) between each pair of voice samples

(a) **identification** (speaker discrimination) **score** for each pair of same speaker (SS) and different speaker (DS) samples

2.3 VPA analysis

- in-house version of VPA scheme
  - 7 scalar degrees (0 → 6)
  - 27 supralaryngeal features

VOCAL PROFILE ANALYSIS PROTOCOL

Speaker: .... Date of recording: .... Judge: .... Recording ID: ....

	FIRST PASS		SECOND PASS						
				moderate			extreme		
	Neutral	Non-neutral	SETTING	1	2	3	4	5	6
A. VOCAL TRACT FEATURES									
1. Labial			Lip rounding/protrusion						
			Lip spreading						
			Labiodentalization						
			Extensive range						
			Minimised range						
2. Mandibular			Close jaw						
			Open jaw						
			Protruded jaw						
			Extensive range						
			Minimised range						
3. Lingual tip/blade			Advanced tip/blade						
			Retracted tip/blade						
4. Lingual body			Fronted tongue body						
			Backed tongue body						
			Raised tongue body						
			Lowered tongue body						
			Extensive range						
			Minimised range						
5. Pharyngeal			Pharyngeal constriction						
			Pharyngeal expansion						
6. Velopharyngeal			Audible nasal escape						
			Nasal						
			Denasal						
7. Larynx height			Raised larynx						
			Lowered larynx						
B. OVERALL MUSCULAR TENSION									
8. Vocal tract tension			Tense vocal tract						
			Lax vocal tract						
9. Laryngeal tension			Tense larynx						
			Lax larynx						

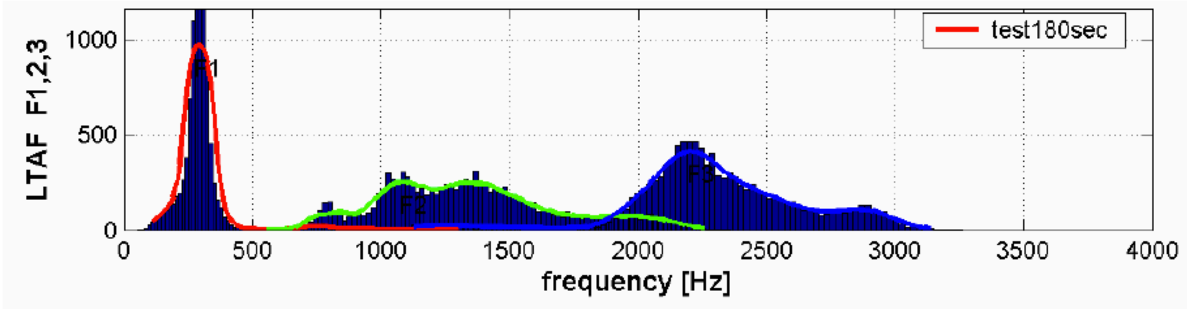
(a) speaker distances

- Euclidean distances between speaker pairs

(b) identification score

- currently one data set per speaker (i.e. no SS comparisons)
- **close match** = speakers with VPA profiles differing by ≤ 2 scalar degrees

2.4 LTFDs



- automatic separation into C and V (StkCV)

  → vowel-only samples:

  - 25ms Gaussian window shifted at 5ms
  - F1→F4 extracted from each frame using iCAbS tracker (Harrison & Clermont 2012)

(a) speaker distances

- LTFDs modelled as GMM (8 Gaussians)
- Kullback-Leibler (KL) divergence: distance between models

(b) identification score

- GMM-UBM: SS (100) & DS (4900) log LRs

2.5 MFCCs

- data extraction and analysis: BATVOX (v4)
  - 20ms hamming window shifted at 10ms intervals
  - 20 MFCCs/ deltas/ delta-deltas per frame

(a) speaker distances

- MFCCs modelled as GMM (1024 Gaussians)
- KL divergence: distance between models

(b) identification score

- BATVOX identification mode: SS (100) & DS (4900) log LRs

3. Results

3.1 Correlations: global

- correlations between VTO distance scores:

Comparison	*r*	*p*
LTFDs vs. MFCCs	0.49	<0.01
LTFDs vs. VPA	0.12	<0.01
MFCCs vs. VPA	0.17	<0.01

  – **but…** global scores might conceal stronger correlations between sub-components

3.1 Correlations:
formants vs. MFCCs/VPA distances

Comparison	MFCC:		VPA:	
	r	***p***	***r***	***p***
F1+F2+F3+F4	**0.49**	<0.01	0.12	<0.01
F1	**0.27**	<0.01	0.03	<0.05
F2	**0.30**	<0.01	0.07	<0.01
F3	**0.44**	<0.01	0.06	<0.01
F4	0.13	<0.01	0.13	<0.01

3.1 Correlations:
formants vs. VPA features

- by-speaker means calculated for LTFD1→4

- Spearman correlation matrix generated for LTFDs and raw VPA scores

3.1 Correlations:
formants vs. VPA features

LTFD 1

- backed tongue body rho = 0.200 $p = 0.045$ *
- pharyngeal constriction rho = 0.298 $p = 0.003$ **
- pharyngeal expansion rho = -0.213 $p = 0.034$ *
- raised larynx rho = 0.397 $p < 0.0001$ ***
- lowered larynx rho = -0.248 $p = 0.013$ *

LTFD 2

- fronted tongue body rho = 0.239 $p = 0.016$ *
- lowered larynx rho = -0.257 $p = 0.0097$ **
- lax vocal tract rho = -0.197 $p = 0.049$ *

LTFD 3

- tense vocal tract rho = 0.242 $p = 0.041$ *

LTFD 4

- pharyngeal constriction rho = -0.220 $p = 0.028$ *
- raised larynx rho = -0.385 $p < 0.0001$ ***

3.2 Speaker discrimination

Speaker discrimination performance (%)

	MFCC	LTFD	VPA (exact)	VPA (close)
True rejection	97.1	97.4	99.5	87.9
True acceptance	100	94	-	-
False acceptance	2.9	2.6	0.5	12.1
False rejection	0	6	-	-

3.2 Speaker discrimination

Speaker discrimination performance (%)

	MFCC	LTFD	VPA (exact)	VPA (close)
True rejection	97.1	97.4	99.5	87.9
True acceptance	100	94	-	-
False acceptance	**2.9**	**2.6**	0.5	12.1
False rejection	0	6	-	-

3.2 Speaker discrimination

Speaker discrimination performance (%)

	MFCC	LTFD	VPA (exact)	VPA (close)
True rejection	97.1	97.4	99.5	87.9
True acceptance	100	94	-	-
False acceptance	2.9	2.6	**0.5**	12.1
False rejection	0	6	-	-

3.2 Speaker discrimination

Speaker discrimination performance (%)

	MFCC	LTFD	VPA (exact)	VPA (close)
True rejection	97.1	97.4	99.5	87.9
True acceptance	100	94	-	-
False acceptance	2.9	2.6	0.5	12.1
False rejection	0	6	-	-

3.2 Speaker discrimination

Speaker discrimination performance (%)

	MFCC	LTFD	VPA (exact)	VPA (close)
True rejection	97.1	97.4	99.5	87.9
True acceptance	100	94	-	-
False acceptance	2.9	2.6	0.5	12.1
False rejection	**0**	**6**	-	-

4. Discussion and conclusion

4.1 Discussion

- strong correlations between acoustic VTO measures (LTFDs & MFCCs)
  - strongest correlation with F3
  - weakest correlation with F4

- weaker correlations between LTFDs/MFCCs and VPA
  - but some strong correlations between individual formants and individual VPA settings
  - different representations of VTO

4.1 Discussion

- speaker discrimination performance of all VTO measures = very good
  – although inevitably all yield errors

- given correlations between LTFDs & MFCCs no reason to expect different errors

- **but…** VPA different representation of VTO?
  – potential improvement in performance of LTFDs/ MFCCs with the inclusion of auditory VPA

4.2 Conclusion

- no perfect VTO measure given limitations of the supralaryngeal vocal tract as a biometric

- further limitations introduced in casework
  - channel mismatch/ background noise/ telephone transmission
  - benefit of using auditory measures which are more robust to some of these limitations

- future work: inclusion of laryngeal features

Thanks!
Questions?

voice and identity

0101101ID1

Arts & Humanities Research Council

THE UNIVERSITY of York

J P French Associates
Forensic speech and acoustics laboratory