Voice quality similarity based on a simplified version of the Vocal Profile Analysis: A preliminary approach with Spanish speakers including identical twin pairs

Eugenia San Segundo[1]; Jose A. Mompean[2]

[1] Dept. of Language and Linguistic Science, University of York, UK
[2] English Department, University of Murcia, Spain

INTRODUCTION

- **Voice quality (VQ)** = characteristic timbre or quasi-permanent quality resulting from a combination of long-term laryngeal and supralaryngeal features which make a speaker's voice recognizably different from others [1].
- **Forensic Speaker Comparison (FSC)** = comparison of voice samples belonging to an offender and a suspect in order to assist courts in determining speaker identity [2].
  - Forensic phoneticians have shown an increasing interest in VQ [3].
  - The Vocal Profile Analysis (VPA) protocol [4] is the most common perceptual assessment scheme in forensic phonetic casework [5].
- Auditory-acoustic approaches to FSC are the most popular methods worldwide [5], only followed by human-supervised Automatic Speaker Recognition (ASR) systems. However analyses based on the perceptual skills of trained experts are sensitive to bias and errors [6]. This may call into question their reliability and validity.
  - Intra and inter-rater agreement have seldom been investigated in VQ and FSC.
  - Monozygotic (MZ) twins have proved suitable to test extreme similar-soundingness in ASR systems [7] and could also prove adequate for perceptual analyses.

OBJECTIVES

1. Design a simplified VPA (henceforth SVPA) for Standard Peninsular Spanish (SPS).
   - The original VPA protocol is very comprehensive and useful but generally thought to be too complex: its greater scope is at the expense of reliability [8] and easy quantification, especially in FSC [3].
   - This simplification aims to tackle typical problems in VQ perceptual assessment, e.g. voice multidimensionality and the difficulty in isolating dimensions [9].
2. Evaluate to which extent two trained phoneticians agree with each other (inter-rater reliability) and with themselves (intra-rater reliability) in VQ perceptual assessment.
3. Explore whether SVPA can be considered as a useful tool for forensic phonetics by testing voice quality similarity in a population of identical twins and unrelated speakers.

Simplified VPA (SVPA) protocol for Spanish: scheme and examples

	Major setting groups	Settings			
FEATURAL ANALYSIS	Vocal tract	Labial	Spreading	N	Rounding
		Mandibular	Close	N	Open
		Apical/alveolar	Retracted	N	Advanced
		Dorsal	Back & Lowered	N	Front & Raised
		Velopharyngeal	Denasal	N	Nasal
		Pharyngeal	Constricted	N	Expanded
		Laryngeal height	Lowered	N	Raised
	Overall muscular tension	Vocal tract tension	Lax	N	Tense
		Laryngeal tension	Lax	N	Tense
	Phonation	Voice type	Whisper/Breathy	N	Creaky/Harsh
HOLISTIC DESCRIPTION					

		Settings	Subject 19 (MML)			Subject 21 (PML)		
FEATURAL	Vocal Tract	Labial	Spreading	N x	Rounding	Spreading	N x	Rounding
		Mandibular	Close x	N	Open	Close	N x	Open
		Apical/alveolar	Retracted	N x	Advanced	Retracted	N	Advanced x
		Dorsal	B & L	N x	F & R	B & L	N	F & R
		Velopharyngeal	Denasal x	N	Nasal	Denasal	N x	Nasal
		Pharyngeal	Constricted x	N	Expanded	Constricted x	N	Expanded
		Laryngeal height	Lowered	N x	Raised	Lowered	N x	Raised
	Muscular tension	VT tension	Lax	N	Tense x	Lax x	N	Tense
		Laryngeal tension	Lax	N	Tense x	Lax x	N	Tense
	Phonation	Voice type	W/B	N x	C/H	W/B x	N	C/H
HOLISTIC			Denasal; fast rate; hypoarticulation (close jaw) but precise consonant articulation (tense vocal tract)			Sibilance/Lisp (advanced tongue tip), key segments "misma", "pista", "susto"; laxness; glottal air escape		

MATERIALS & METHOD

Participants: 24 non-pathological male speakers of Standard Peninsular Spanish (SPS), aged 20-36 (mean 26.83, SD 6.6). They make up 12 pairs of MZ twins.

Data: Spontaneous speech samples (ca. 130 sec) of researcher-speaker conversations [10].

Procedure: Blind VQ perceptual assessment of each speech sample (random order) carried out by two trained phoneticians. Two independent evaluations per rater (one week lapse).

Analysis method:

1. **Perceptual analysis:** Simplified Vocal Profile Analysis (SVPA)

- Definition of neutral setting based on phonological descriptions of SPS [11].
- Main characteristics of the simplification:
  - ✓ Reduction from 36 to 22 features
  - ✓ No use of scalar degrees; use instead of a binary (neutral/non-neutral) system
  - ✓ Use of only two possible settings for most deviations from neutrality
  - ✓ Possibility of including holistic descriptions

2. Statistical analyses

- For inter and intra rater analyses we used *Percent Agreement* (rough estimate of reliability) and three chance-corrected measures, suitable for two raters and nominal variables: *Cohen's Kappa* and its variants *Scott's Pi* and *Krippendorf's Alpha.*
- For analysis of VQ similarity we used *Similarity Matching Coefficients* (SMC), a common distance measure for categorical data where the number of matches for each variable is divided by the number of variables. SMC range from 0 (very different) to 1 (very similar).

RESULTS

Intra-rater agreement

Setting	Percent Agreement		Cohen's Kappa		Krippendorff's Alpha		N disagreements (speaker)	
	R1	R2	R1	R2	R1	R2	R1	R2
Labial	95.83	100	0.91	1.00	0.91	1.00	$1_{(18)}$	0
Mandibular	95.83	91.67	0.93	0.80	0.93	0.80	$1_{(7)}$	$2_{(2,16)}$
Apical	100	100	1.00	1.00	1.00	1.00	0	0
Dorsal	91.67	100	0.81	1.00	0.81	1.00	$2_{(10,11)}$	0
Velopharyngeal	83.33	91.67	0.74	0.87	0.74	0.88	$4_{(3,7,10,14)}$	$2_{(3,22)}$
Pharyngeal	91.67	95.83	0.87	0.93	0.88	0.94	$2_{(2,4)}$	$1_{(4)}$
Larynx height	87.50	87.50	0.81	0.81	0.81	0.81	$3_{(8,11,18)}$	$3_{(3,15,22)}$
VT tension	87.50	95.83	0.81	0.93	0.82	0.94	$3_{(4,18,22)}$	$1_{(24)}$
Larynx tension	91.67	87.50	0.84	0.71	0.84	0.72	$2_{(12,11)}$	$3_{(8,12,19)}$
Voice type	83.33	87.50	0.70	0.78	0.71	0.78	$4_{(3,4,12,18)}$	$3_{(10,12,18)}$

Inter-rater agreement

Setting	Percent Agreement	Cohen's Kappa	Krippendorff's Alpha	N disagreements
Labial	**75.00**	**0.55**	**0.55**	**6**
Mandibular	50.00	0.06	0.04	12
Apical	54.17	0.11	0.09	11
Dorsal	**91.67**	**0.78**	**0.79**	**2**
Velopharyngeal	**70.83**	**0.55**	**0.56**	**7**
Pharyngeal	37.50	0.11	0.02	15
Larynx height	**66.67**	**0.50**	**0.50**	**8**
VT tension	41.67	0.13	0.13	14
Larynx tension	**66.67**	**0.30**	**0.30**	**8**
Voice type	**66.67**	**0.42**	**0.43**	**8**

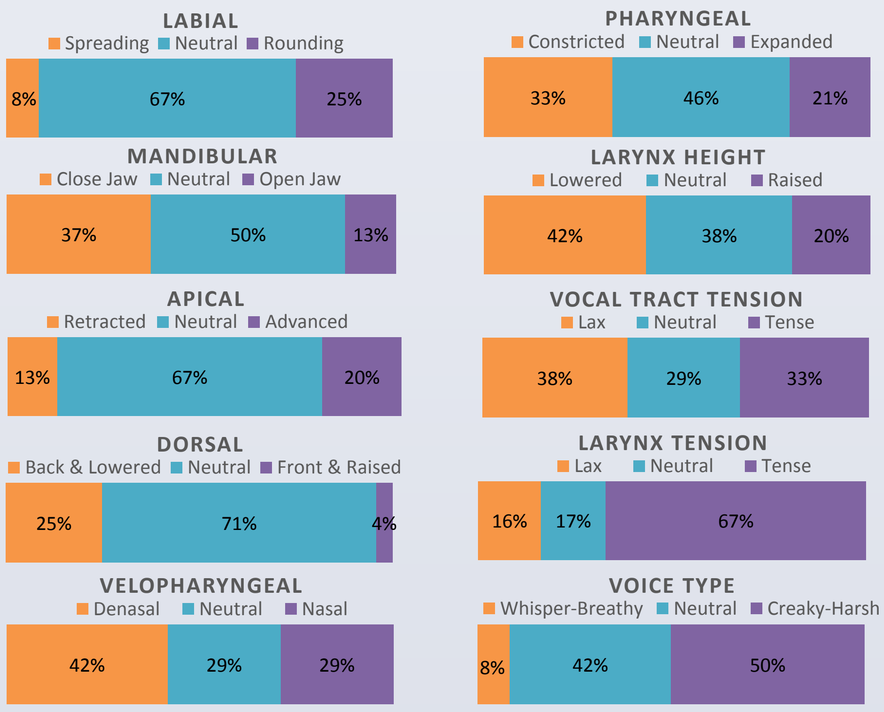
Similarity Matching Coefficients

SMC

Twin pair	SMC
1-AGF	
24-SGF	0.8
2-AGP	
8-CGP	0.7
3-AMG	
13-EMG	0.8
4-APJ	
22-RPJ	0.4
5-ARJ	
17-JRJ	0.5
6-ASM	
23-RSM	0.7
7-CAS	
20-PAS	0.6
9-CSD	
12-DSD	0.8
10-DCT	
15-JCT	0.5
11-DSA	
14-ISA	0.6
16-JHB	
18-MHB	1
19-MML	
21-PML	0.3

Unrelated speakers	SMC
1-AGF	
2-AGP	0.3
3-AMG	
4-APJ	0.4
5-ARJ	
6-ASM	0.2
7-CAS	
8-CGP	0.4
9-CSD	
10-DCT	0.3
11-DSA	
12-DSD	0.4
13-EMG	
14-ISA	0.2
15-JCT	
16-JHB	0.3
17-JRJ	
18-MHB	0.2
19-MML	
20-PAS	0.5
21-PML	
22-RPJ	0.3
23-RSM	
24-SGF	0.5

Setting frequency of occurrence



DISCUSSION

- **Intra-rater agreement**: almost perfect (κ values between 0.81–1) regardless of rater.
- **Inter-rater agreement**: high overall but setting-dependent: *slight* agreement (mandibular, apical, pharyngeal and vocal tract tension); *fair* (laryngeal tension); *moderate* (voice type, larynx height, labial and velopharyngeal); and *substantial* (dorsal). ***Note***: Kappa is a conservative measure; better results with weighted Kappa. Previous voice studies show that reliability is usually not better than moderate.
- **SMC:** higher values in twin pairs (mean=0.64) than in unrelated speakers (mean=0.35), indicating more similarity among the former. ***Twin pairs:*** values tend to be pair-dependant: JHB & MHB very similar (1) while MML & PML very different (0.3). ***Unrelated speakers:*** no "vocal twins" as described in [4] and distinguished on average by more than 7 settings. Matches are based of sharing accent features/coincidence on neutral.

CONCLUSION

- We have proposed a **simplified VPA (SVPA) protocol** adapted to SPS, which has proved reliable from the point of view of inter-rater agreement and intra-rater consistency.
- SVPA is **useful for FSC**: it yields high SMC for very similar-sounding speakers (MZ twins) and low SMC for unrelated speakers. Percentages of setting distribution can be used to estimate reference population statistics to account for typicality.
- Other applications of our SVPA include L2 phonology, dialectal and sociolinguistic studies.

References

[1] Laver, J. (1980). *The phonetic description of voice quality*. Cambridge: CUP.

[2] Foulkes, P., & French, P. (2012). Forensic speaker comparison: a linguistic-acoustic perspective. In L. Solan & P. Tiersma (eds.) *The Oxford Handbook of Language and Law* (pp. 418–421). Oxford: OUP.

[3] San Segundo, E., French, P., Foulkes, P., Hughes, V. & Harrison, P. (2016). Developing the vocal profile analysis scheme for forensic voice comparison, *BAAP (British Association of Academic Phoneticians) Colloquium*, Lancaster University, 30 March – 1 April 2016.

[4] Beck, J. (2005). Perceptual analysis of voice quality: the place of Vocal Profile Analysis. In W.J. Hardcastle & J. Mackenzie Beck (Eds.) *A Figure of Speech: a Festschrift for John Laver* (pp. 285–322). London/Mahwah, NJJ: Laurence Erlbaum Associates.

[5] Gold, E. & French, P. (2011). International practices in forensic speaker comparison. *Int. Journal Speech Language & Law 18*, 293–307.

[6] Kent, R. D. (1997). Hearing and Believing: Some Limits to the Auditory-Perceptual Assessment of Speech and Voice Disorders. *American Journal of Speech-Language Pathology* 5, 7–23.

[7] Künzel, H. (2010). Automatic speaker recognition of identical twins. *Int. Journal of Speech, Language and the Law 17*, 251-277.

[8] Webb, A. L., et *al*. (2004). The reliability of three perceptual evaluation scales for dysphonia. *Eur Arch Otorhinolaryngol 261*, 429–434.

[9] Kreiman, J., Gerrat, B.R. & Ito, M. (2007). When and why listeners disagree in voice quality assessment tasks. *JASA* 122, 2354–2364.

[10] San Segundo, E. (2014). *Forensic speaker comparison of Spanish twins and non-twin siblings: A phonetic-acoustic analysis of formant trajectories in vocalic sequences, glottal source parameters and cepstral characteristics*, PhD. Diss., CSIC & UIMP.

[11] Gil, J. (2005). Implicaciones fonológicas de la base de articulación. In *Filología y Lingüística. Estudios ofrecidos a Antonio Quilis.* Vol 1. (pp. 219–252). Madrid: CSIC, UNED, Universidad de Valladolid.