Evidential value of long-term laryngeal voice quality acoustics

Ricky K.W. Chan[1] *and Bruce Xiao Wang*[2]
[1]*Speech, Language and Cognition Laboratory, School of English, University of Hong Kong, HK*
rickykwc@hku.hk
[2]*Department of Chinese and Bilingual Studies, Hong Kong Polytechnic University, HK*
bruce.wang@alumni.york.ac.uk
a 'work in progress' poster

Introduction. One of the main goals in forensic voice comparison (FVC) is to identity speech features that are useful for distinguishing voices under forensically relevant conditions. Voice quality (VQ) was reported to be useful features for FVC (e.g. Gold & French, 2011), but empirical studies that test this claim are surprisingly limited. This contribution focuses on the acoustics of laryngeal voice quality, and tests how the use of non-contemporaneous recordings affect their evidential value.

Methods. 75 male speakers aged 18-45 were selected from a forensically-oriented database of 552 Australian English speakers (Morrison et al., 2015). Speakers were recorded on more than one occasion under different speech styles, i.e. the casual telephone conversation (CNV) and pseudo police interview (INT) (i.e. CNV1, CNV2, INT1, INT2). Around 33 seconds of vocalic material per recording was analyzed. The VQ parameters (Tables 1 and 2) reported in Hughes et al. (2019) were extracted using VoiceSauce (Shue et al., 2011) and served as input for score generation and LR computation. The fvclrr package (Lo, 2018) was used to implement the multivariate kernel-density (MVKD; Aitken & Lucy, 2004) formula for same-speaker and different-speaker comparisons. Calibrations were conducted using logistic regression (Brümmer et al., 2007). The 75 speakers were randomly assigned to training, test and reference set respectively (25 speakers in each set). The procedure above was replicated 100 times with different speakers in the training, test, and reference sets, as it has been demonstrated that the reliability of system performance hinges on the speaker samples involved (Wang et al., 2019). This contribution reports two comparisons: CNV1 vs. INT1 (contemporaneous recordings) and CNV1 vs. INT2 (non-contemporaneous recordings).

Results and discussion. Overall, all the input parameters yielded a small standard deviation in C_{llr} (less than 0.1) and EER (mostly less than 5%) values across the 100 replications, suggesting that system performance using these parameters as input were stable. Individual VQ parameters performed rather poorly, with mean C_{llr} values close to 1 and mean EER value mostly greater than 40%. This suggests that individual VQ parameters carry little speaker-discriminatory information. System performances improved considerably when combining all the spectral tilt parameters or all the additive noise parameters, but the results are still less promising than those reported in Hughes et al. (2019). Surprisingly, using all spectral tilt and additive noise parameters as input led to worse performance, suggesting that these two types of measures provide overlapping or conflicting information for distinguishing speakers. More comprehensive analysis, theoretical and forensic implications, and suggestions for future research will be presented in the conference.

CNV1 vs. INT1

	Cllr				EER			
VQ parameter	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**
H1 - H2	0.97	1.09	1.00	0.02	36.00	64.42	48.26	4.55
H2 - H4	0.89	1.41	1.00	0.06	32.75	56.42	44.11	4.83
H1 - A1	0.99	1.03	1.00	0.00	40.08	60.00	49.30	3.45
H1 - A2	0.99	1.04	1.00	0.01	39.50	60.33	49.15	3.72
H1 - A3	0.92	1.11	0.98	0.03	32.83	52.00	42.51	3.73
Spectral tilt	0.91	1.04	0.96	0.02	34.67	51.42	41.11	3.61
CPP	0.93	1.12	0.98	0.03	32.00	52.00	42.59	3.87
HNR05	0.91	1.15	0.96	0.03	36.00	53.00	44.84	4.07
HNR15	0.88	1.13	0.97	0.05	28.25	48.75	41.32	4.27

HNR25	0.89	1.21	0.98	0.05	29.00	52.00	41.40	4.13
HNR35	0.92	1.35	0.98	0.07	32.00	48.42	40.89	3.29
Additive noise	0.84	1.05	0.92	0.04	28.00	47.83	36.77	4.25
Spectral tilt + additive noise	0.85	1.02	0.93	0.04	25.58	44.67	35.31	4.00

CNV1 vs. INT2

	Cllr				**EER**			
VQ parameter	**Min**	**Max**	**Mean**	**SD**	**Min**	**Max**	**Mean**	**SD**
H1 - H2	0.93	1.33	1.01	0.06	32.00	64.08	43.73	5.37
H2 - H4	0.97	1.07	1.00	0.02	37.00	59.08	48.49	3.66
H1 - A1	0.99	1.11	1.00	0.01	40.67	56.92	50.10	3.36
H1 - A2	0.95	1.08	0.99	0.02	40.00	56.00	49.27	3.36
H1 - A3	0.95	1.08	0.99	0.02	40.00	56.00	47.88	3.26
Spectral tilt	0.91	1.05	0.97	0.03	32.00	48.00	39.76	3.41
CPP	0.93	1.12	0.98	0.03	32.00	52.00	42.59	3.87
HNR05	0.91	1.15	0.96	0.03	36.00	53.00	44.84	4.07
HNR15	0.88	1.13	0.97	0.05	28.25	48.75	41.32	4.27
HNR25	0.89	1.21	0.98	0.05	29.00	52.00	41.40	4.13
HNR35	0.92	1.35	0.98	0.07	32.00	48.42	40.89	3.29
Additive noise	0.76	1.01	0.88	0.05	23.17	40.00	30.75	3.71
Spectral tilt + additive noise	0.87	1.02	0.93	0.03	27.33	44.08	35.84	3.61

Tables 1 and 2: statistics of Cllr and EER values across 100 replications with VQ parameters as input in CNV1 vs. INT1, and CNV1 vs. INT2 respectively. Spectral tilt: combination of H1-H2, H2-H4, H1-A1, H1-A2, H1-A3; Additive noise: combination of CPP and HNR05-35.

References

Aitken, C. G. G., & Lucy, D. (2004). Evaluation of trace evidence in the form of multivariate data. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 53*(1), 109–122. https://doi.org/10.1046/j.0035-9254.2003.05271.x

Brümmer, N., Burget, L., Cernocky, J., Glembek, O., Grezl, F., Karafiat, M., van Leeuwen, D. A., Matejka, P., Schwarz, P., & Strasheim, A. (2007). Fusion of Heterogeneous Speaker Recognition Systems in the STBU Submission for the NIST Speaker Recognition Evaluation 2006. Proceedings of IEEE Transactions on Audio, Speech, and Language, 15(7), 2072–2084. https://doi.org/10.1109/TASL.2007.902870

Gold, E., & French, P. (2011). International practices in forensic speaker comparison. *International Journal of Speech, Language & the Law*, *18*(2).

Hughes, V., Cardoso, A., Foulkes, P., French, J. P., Harrison, P. and Gully, A. (2019) Forensic voice comparison using long-term acoustic measures of voice quality. *Proceedings of the 19th International Congress of Phonetic Sciences (ICPhS)*. Melbourne, Australia. pp. 1455-1459.

Lo, J. (2018). FVClrr: likelihood ratio calculation and testing in forensic voice comparison (2.0.1) [Computer software]. https://github.com/justinjhlo/fvclrr.

Morrison G.S., Zhang C., Enzinger E., Ochoa F., Bleach D., Johnson M., Folkes B.K., De Souza S., Cummins N., Chow D., Szczekulska A. (2021). *Forensic database of voice recordings of 500+ Australian English speakers (AusEng 500+)*. [Available: http://databases.forensic-voice-comparison.net/]

Shue, Y.-L., P. Keating , C. Vicenik, K. Yu (2011) VoiceSauce: A program for voice analysis. *Proceedings of the ICPhS XVII*, 1846-1849.