Stability of short-term voice quality parameters in GSM

Jitka Vaňková, Tomáš Bořil, and Radek Skarnitzl
Institute of Phonetics, Faculty of Arts, Charles University in Prague, Czech Republic
jitka.vanka@gmail.com, {tomas.boril|radek.skarnitzl}@ff.cuni.cz

Voice quality parameters have not been investigated to a great extent in technical speaker identification tasks, in spite of the fact that forensic phoneticians appear to make rather frequent use of voice quality in their casework (Nolan, 2005; Gold & French, 2011). The main reason for the lack of acoustic investigations – one of the few exceptions being Jessen (1997) – appears to be the fact that the presence of especially laryngeal voice quality features is compromised in telephone speech (Nolan, 2005). In addition, the plasticity of our voice production mechanism allows for great stylistically conditioned variability, and it is mostly voice quality which is affected.
Yet we believe that there still is space for acoustic examinations of speaker specificity of voice quality, especially of its short-term correlates which reflect spectral slope by comparing the amplitudes of various events in the acoustic spectrum (Hanson et al., 2001). The motivation for using the parameters H1*-H2*, H1*-A1*, H1*-A2*, H1*-A3* and H2*-H4* is twofold: first, it appears that some of them yield favourable rates of intra-speaker stability and inter-speaker variability (Vaňková and Skarnitzl, 2014); second, low frequencies relevant for H1 are actually not filtered out by the Adaptive Multi-Rate (AMR) codec, which is the current standard in mobile telephony (Guillemin and Watson, 2008; Vaňková and Bořil, submitted).

We analyzed recordings of 5 female and 5 male speakers which were passed through the AMR codec, using the lowest and highest bit rate of both its narrowband (NB) and wideband (WB) version (3GPP, 2012). 15 vowel items of each of the short Czech monophthongs /ɪ ɛ a o u/ were used, yielding 750 vowel tokens. F0 and formants were extracted from the central part of each token, and voice quality parameters computed using VoiceSauce (Shue, 2013) five times – from the original studio recordings and from the four types of GSM compression.

Table 1 displays mean differences between parameter values (in dB) in studio recordings and the four codec conditions (positive values signal higher studio values, negative ones higher codec values). Differences in values depend on individual parameters (H1*-A2* and H2*-H4* appearing most robust; note that H1*-H2* was identified by Jessen, 1997 as carrying the most speaker-specific information) and they also vary across codec conditions (WB performing overall better than NB). In terms of other sources of variability, the impact of the codec on the parameters was likewise found to differ for the two genders and individual vowel qualities. These results will also be included in the presentation.

Table 1. Mean differences between parameter values (in dB) in studio recordings and the four codec conditions. Positive values signal higher studio values; negative higher codec values.

	studio-NB 4.75	***studio-NB 12.20***	***studio-WB 6.60***	***studio-WB 19.85***	**Total (abs)**
H1*-H2*	1.46	0.70	0.97	0.19	**3.32**
H2*-H4*	0.55	0.13	0.59	0.43	**1.70**
H1*-A1*	2.52	1.35	1.86	0.64	**6.37**
H1*-A2*	-1.03	-0.27	-0.17	0.17	**1.64**
H1*-A3*	-1.76	-1.21	-0.43	0.42	**3.82**
Total (abs)	**7.33**	**3.65**	**4.02**	**1.84**	

References

3GPP (2012). TS 26.071 AMR speech CODEC; General description. Downloaded on February 2, 2014 from http://www.3gpp.org/ftp/specs/-archive/26_series/26.071/.

Gold, E. and P French. (2011). International practices in forensic speaker comparison. *The International Journal of Speech, Language and the Law,* **18**, 293–307.

Guillemin, B. J. and C Watson (2008). Impact of the GSM mobile phone network on the speech signal: some preliminary findings. *International Journal of Speech, Language and the Law*, **15**, 193–218.

Hanson, H. M., K N Stevens, H-K J Kuo, M Y Chen and J Slifka (2001). Towards models of phonation. *Journal of Phonetics*, **29**, 451–480.

Jessen, M. (1997). Speaker-specific information in voice quality parameters. *Forensic Linguistics*, **4**, 84–103.

Nolan, F. (2005). Forensic speaker identification and the phonetic description of voice quality". In W. Hardcastle & J Beck (Eds,), *A Figure of Speech*, 385–411. Mahwah, New Jersey: Erlbaum.

Shue, Y. (2013). VoiceSauce: A program for voice analysis (V1.14). Downloaded on October 7, 2013 from http://www.seas.ucla.edu/spapl/voicesauce/

Vaňková, J. and T Bořil (submitted). Impact of the GSM AMR codec on automatic vowel formant measurement in Praat and VoiceSauce (Snack). EUSIPCO 2014.

Vaňková, J. and R Skarnitzl (2014). Within- and between-speaker variability of parameters expressing short-term voice quality. *Proceedings of Speech Prosody 2014*, 1081–1085. Dublin.