Inter-speakers variability of intensity levels across syllables

Lei He and Volker Dellwo
Phonetics Laboratory, University of Zurich, Zurich, Switzerland
{lei.he|volker.dellwo}@uzh.ch

People can easily be identified by their voice. Apart from information about the speaker's linguistic and socioeconomic background, the anatomical individualities of the speech organs and vocal tract, as well as the idiosyncratic control over the muscular movements of speech organs are fundamental to speaker idiosyncrasy in the speech signal (Dellwo et al., 2007). Such individual differences could result in speaker-specific pulmonic and sub-glottal pressure fluctuations. As a result, energy distributions in the acoustic signal could be idiosyncratic as well. We hypothesize that this speaker-specific energy distribution in the speech signal could be captured by measuring the intensity level variability across syllables in the utterances.

Based on the influential speech rhythm metrics (Ramus et al., 1999; Grabe & Low, 2002; Dellwo, 2006), which also yielded fair success in forensic voice comparison research (Dellwo et al., 2012; Leemann et al., 2014), we developed two sets of intensity variability metrics:

- The global measures:
  - stdevM: the standard deviation of average syllable intensity levels;
  - stdevP: the standard deviation of syllable peak intensity levels;
  - varcoM: the variation coefficient of average syllable intensity levels (normalized stdevM);
  - varcoP: the variation coefficient of syllable peak intensity levels (normalized stdevP).

- The local measures:
  - rPVIm: the raw pairwise variability of adjacent mean syllable intensity levels;
  - rPVIp: the raw pairwise variability of adjacent syllable peak intensity levels;
  - nPVIm: the normalized pairwise variability of adjacent mean syllable intensity levels;
  - nPVIp: the normalized pairwise variability of adjacent syllable peak intensity levels.

We applied these metrics to the TEVOID corpus built by Dellwo et al. (2012), which currently contains 16 gender-balanced Zurich German speakers, each producing 256 read sentences and 16 spontaneously uttered sentences. An initial visualization of the raw metrics scores (please see the box plots in Figure 1) suggests that a significant factor of the speakers is very likely to be found after transforming the data into more normally distributed ones. For further research, we would like to test these metrics on degraded speech as well, and work on possible optimizations of the metrics to make them more useful in forensic applications.

References

Dellwo, V. (2006). Rhythm and speech rate: a variation coefficient for deltaC. In P. Karnowski & I. Szigeti (Eds.), *Language and Language Processing*, 231-241, Frankfurt: Peter Lang.

Dellwo, V., Huckvale, M. and Ashby, M. (2007). How is individuality expressed in voice? An introduction to speech production and description for speaker classification. In C. Müller (Ed.), *Speaker Identification I*, 1-20, Berlin: Springer Verlag.

Dellwo, V., Leemann, A. and Kolly, M.-J. (2012). Speaker idiosyncratic rhythm features in the speech signal. In *Interspeech*, Portland, USA.

Grabe, E. and Low, E. L. (2002). Durational variability in speech and rhythm class hypothesis. In N. Warner & C. Gussenhoven (eds.), *Papers in Laboratory Phonology 7*, 515-543, Berlin and New York: Mouton de Gruyter.

Leemann, A., Kolly, M.-J., and Dellwo, V. (2014). Speech-individuality in suprasegmental temporal features: implications for forensic voice comparison. *Forensic Sci. Int.*, **238**, 59-67.

Ramus, F., Nespor, M. and Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, **73**, 265-292.

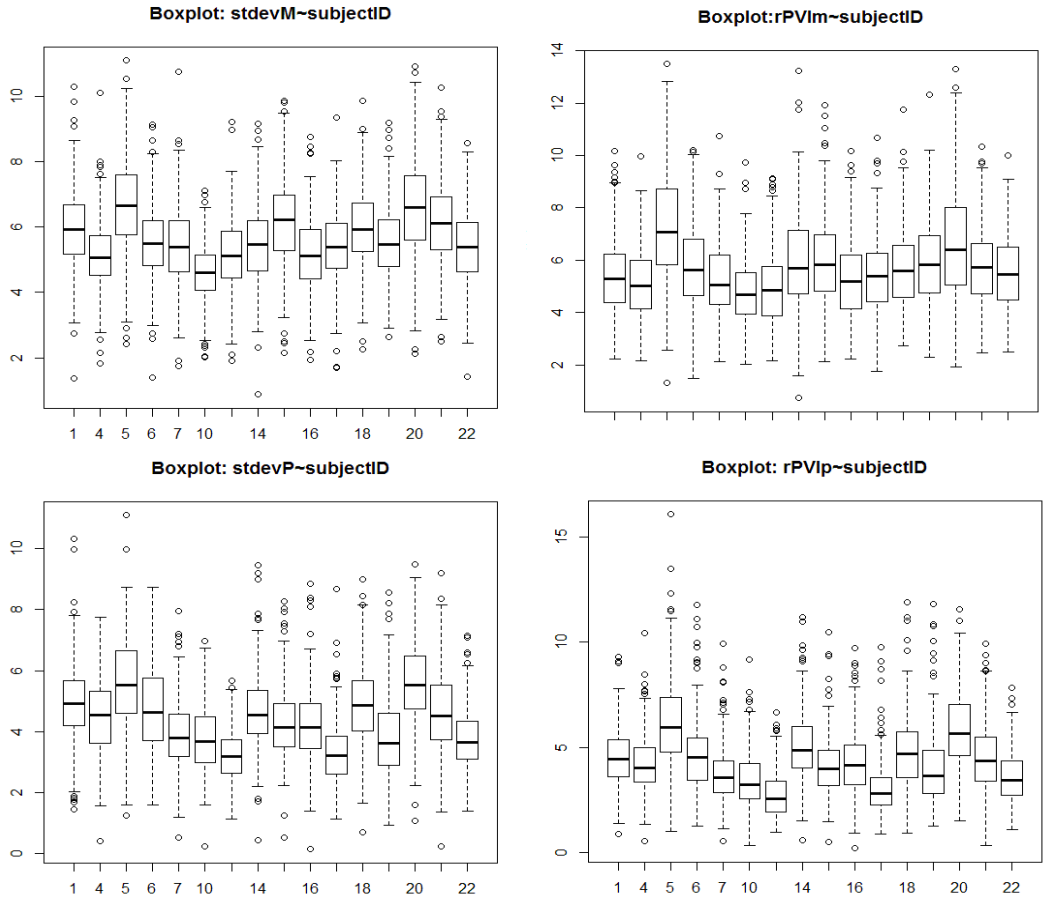

Figure 1 Box plots of the stdevM, rPVIm, stdevP and rPVIp scores of the read sentences in the TEVOID corpus.