Errata for the paper "An improved non-intrusive intelligibility metric for noisy and reverberant speech"

Correction for Equation 1

We have recently found out that the SRMRnorm implementation we distribute on Github is slightly different from what is presented in our paper titled "An improved non-intrusive intelligibility metric for noisy and reverberant speech", which was presented at IWAENC 2014. The difference is in how the normalization step is performed. In eq. 1 of the aforementioned paper, we show that the peak energy is computed as the maximum value of the average energy across **frames**; however, in our implementation, it is computed as the maximum value of the average energy across **acoustic channels**.

The equation should be corrected to the following:

$$\bar{E}_{peak} = \max_{m,f_b}(\frac{1}{J}\sum_{j=1}^{J} E_j(m, f_b))$$

where J is the number of acoustic channels (23).

We have also tested changing the implementation to correspond to the equation originally shown in the paper. However, that leads to higher variability and slightly lower Spearman and sigmoidal correlations. The evaluation metrics are, following the notation on the paper, $\rho_p = 0.78, \rho_{sp} = 0.91, \rho_{sig} = 0.91, RSD\% = 0.15, RMSE = 10.90$.

We thank Sebastian Braun for pointing out this issue.

Correction for the STOI scores on Table 1

Speech files for one of the reverberant conditions (T60 = 1.4 s) were not aligned to the corresponding clean files before the computation of STOI, which led to a discrepancy in the computed scores. The correct evaluation metrics, following the notation on the paper, are $\rho_p = 0.79, \rho_{sp} = 0.88, \rho_{sig} = 0.88, RSD\% = 0.06, RMSE = 11.93$.

We thank Steven Van Kuyk for pointing out this issue.