Corrections to "Speech Enhancement Based on Bayesian Low-Rank and Sparse Decomposition of Multichannel Magnitude Spectrograms"

Yoshiaki Bando, *Student Member, IEEE,* Katsutoshi Itoyama, *Member, IEEE,*
Masashi Konyo, *Member, IEEE,* Satoshi Tadokoro, *Fellow, IEEE,* Kazuhiro Nakadai, *Senior Member, IEEE,*
Kazuyoshi Yoshii, *Member, IEEE,* Tatsuya Kawahara, *Fellow, IEEE,* Hiroshi G. Okuno, *Fellow, IEEE*

We describe two corrections to the paper entitled "Speech Enhancement Based on Bayesian Low-Rank and Sparse Decomposition of Multichannel Magnitude Spectrograms" [1].

A. Correction to Figs. 9(b) and 14(b)

The noisy signals used in Figs. 9(b) and 14(b) were described in the second paragraph of Section V-B in [1] as follows:

> Fig. 9(a) shows the performances for all the noisy signals. Fig. 9(b) shows those for the subset of Fig. 9(a) by excluding the ego-noise and speech signals used for the parameter optimization.

In fact, however, these figures were obtained in a wrong setting where the ego-noise used for the parameter optimization was included in the test data. To show the generalization capability to new noise signals, we used additional 40-second ego-noise signals for the Open conditions and Door-4ch condition, and the description is revised as follows:

> Fig. 9(a) shows the performances for all the noisy signals. Fig. 9(b) shows those for a different set of noisy signals whose ego-noise and speech signals were not used for the parameter optimization.

Here, the new versions of Figs. 9(b) and 14(b) are shown in the next page.

B. Correction to Fig. 17

Fig. 17 in the original paper has the results of VB-SRNTF-$g^{(n)}_{mft}$ that was incorrectly implemented. The correct version of Fig. 17 in [1] is shown in the next page. The discussion in the third paragraph of Section V-D in [1] is revised as follows:

> On the other hand, the SDR performance was degraded in the Door-2ch condition by VB-SRNTF-$g^{(n)}_{mft}$ whose gain has both frequency and time dependencies.

Y. Bando, K. Itoyama, K. Yoshii, and T. Kawahara are with Graduate School of Informatics, Kyoto University, Kyoto, 606-8501, Japan, email: yoshiaki@kuis.kyoto-u.ac.jp.

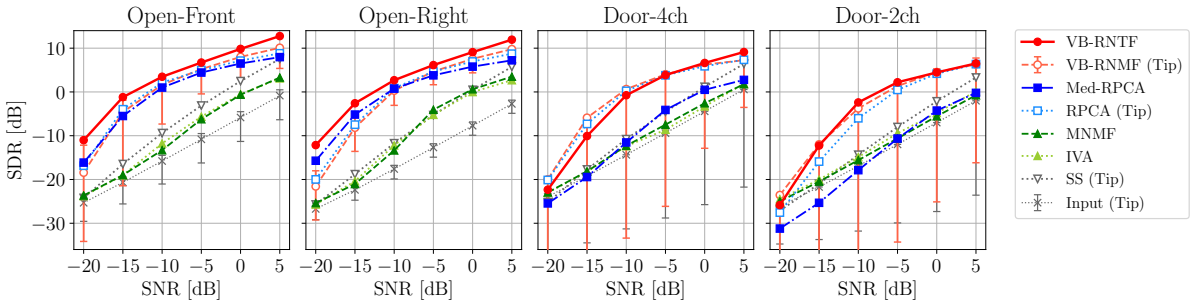
M. Konyo and S. Tadokoro are with Graduate School of Information Science, Tohoku University, Miyagi, 980-8579, Japan.

K. Nakadai is with Tokyo Institute of Technology, Saitama, 351-0114, Japan / Honda Research Institute Japan Co., Ltd.

H. G. Okuno is with Graduate Program for Enbodiment Informatics, Waseda University, Tokyo, 169-0072, Japan.

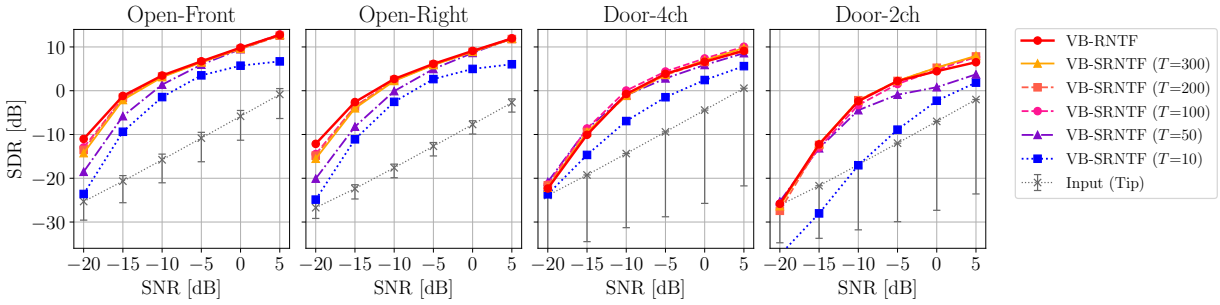
References

[1] Y. Bando, K. Itoyama, M. Konyo, S. Tadokoro, K. Nakadai, K. Yoshii, T. Kawahara, and H. G. Okuno, "Speech enhancement based on Bayesian low-rank and sparse decomposition of multichannel magnitude spectrograms," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 2, pp. 215–230, 2018.



(b)

Fig. 9. Speech enhancement performances of VB-RNTF, VB-RNMF, and existing methods. Each line indicates average SDR at the specified condition. Error bars for VB-RNMF and the input signal span the maximum and minimum SDRs in all the microphones. (a) Average SDRs for all the noisy test signals. (b) Average SDRs for the different set of noisy signals that were not used for the parameter optimization.



(b)

Fig. 14. SDR performances of VB-SRNTFs with different mini-batch sizes. (a) Average SDRs for all the noisy test signals. (b) Average SDRs for the different set of noisy signals that were not used for the parameter optimization.

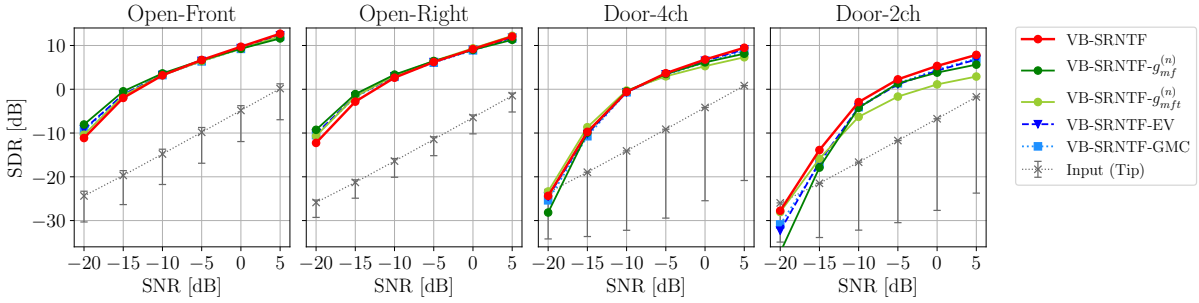


Fig. 17. Comparison of VB-SRNTF (T=200) proposed in Section III-C and the variants of VB-SRNTF with the frequency-dependent gains and the temporal-continuous gains.