

This document discusses the value of the per-speaker input SNR for the generation of the reverberant LibriCHiME-5 dataset. It corresponds to the value of 'mu' in data_creation.md (section 'Create mixtures').

Table of content

[Distribution of the SNR estimated with Brouhaha](#)

[Input SI-SDR](#)

[Conclusion](#)

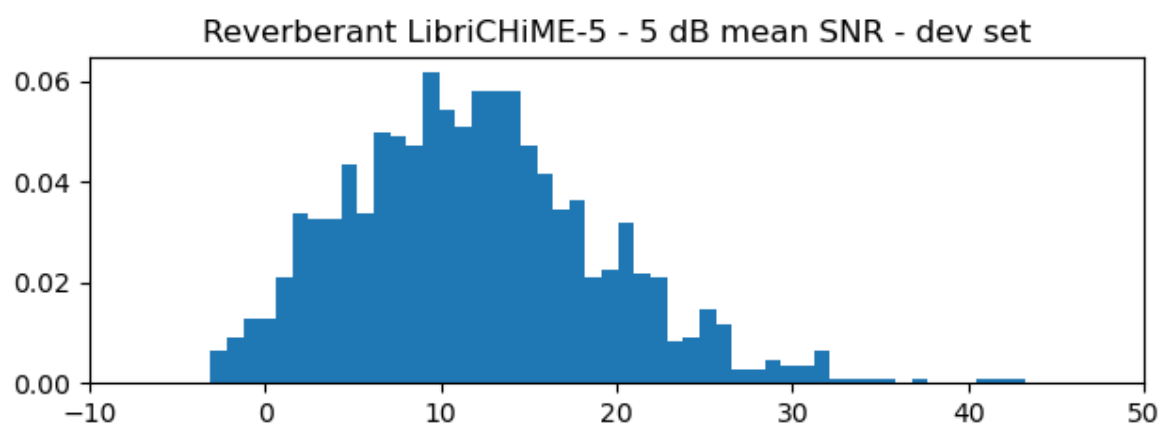
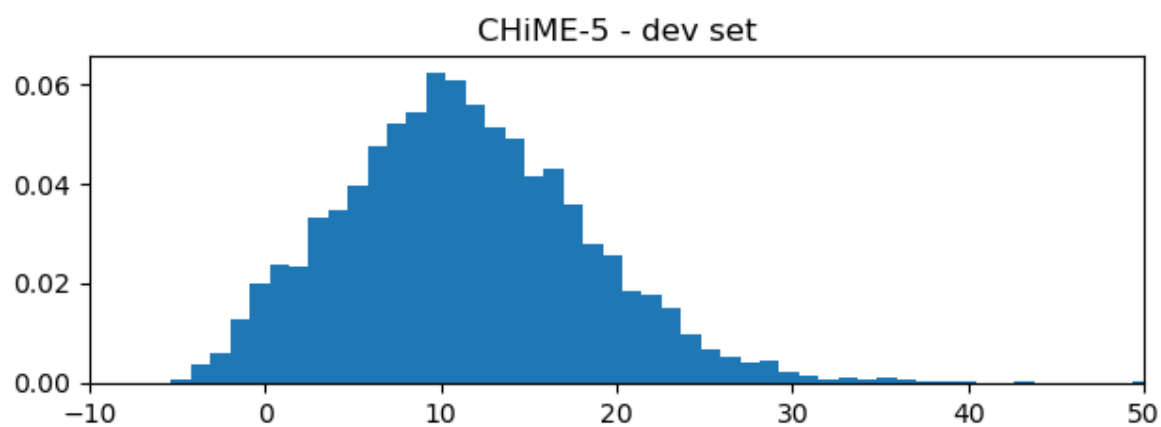
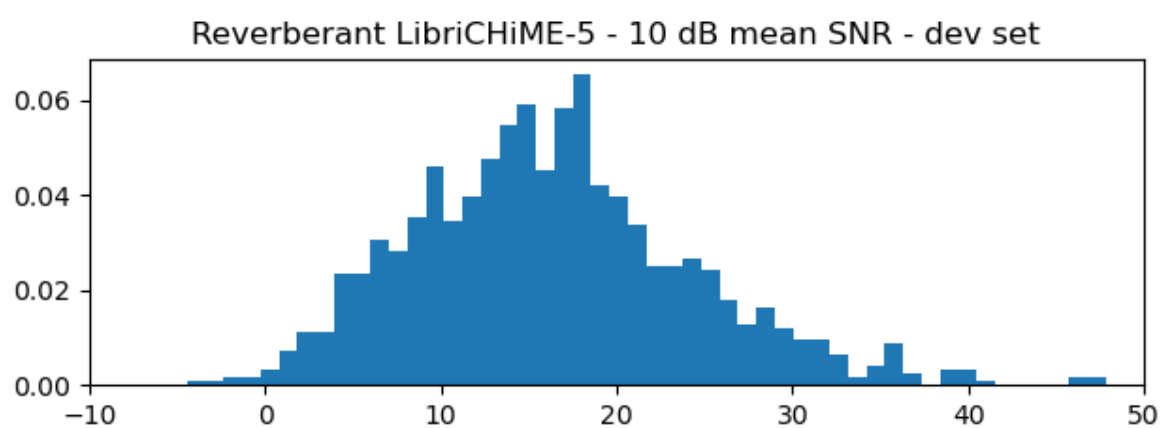
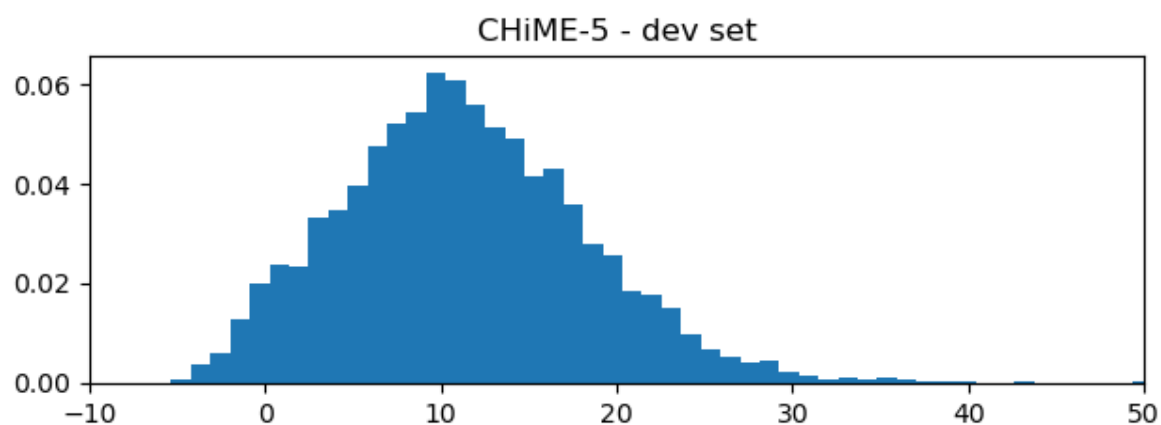
Distribution of the SNR estimated with Brouhaha

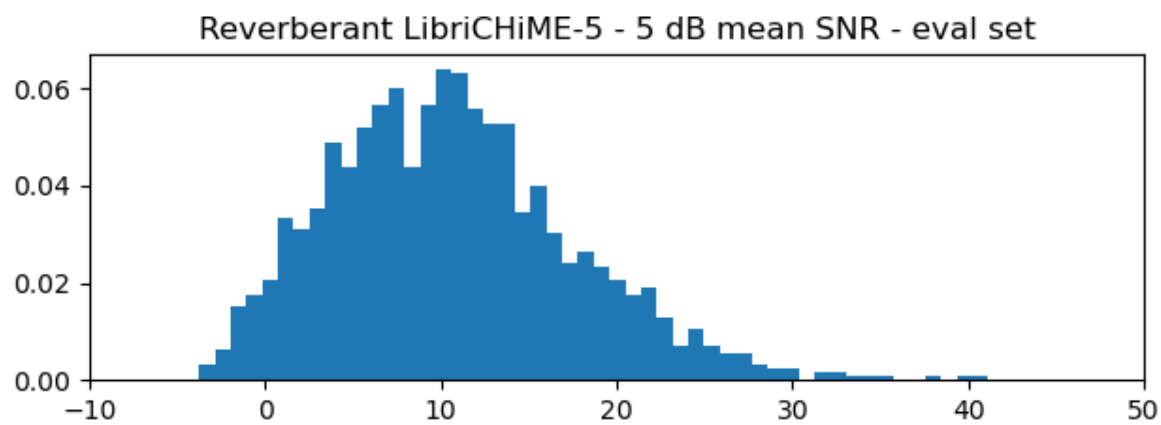
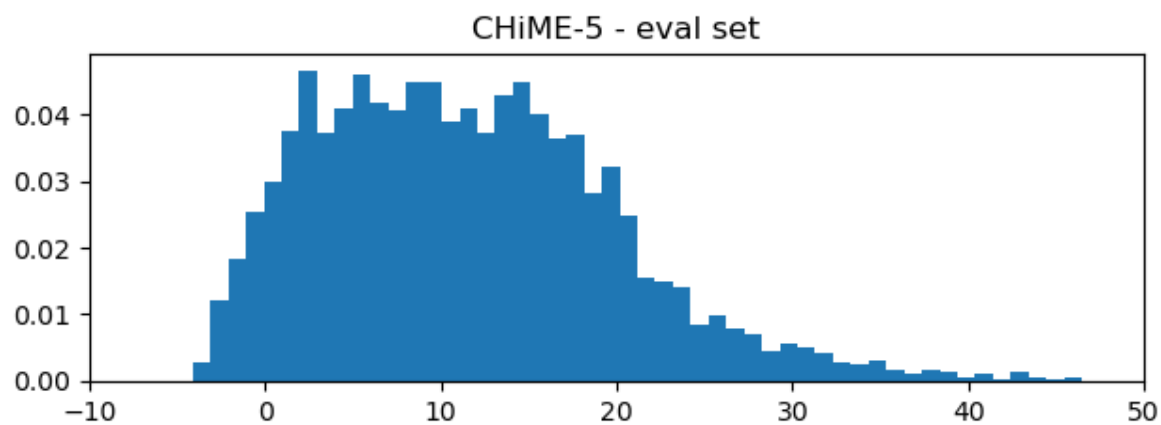
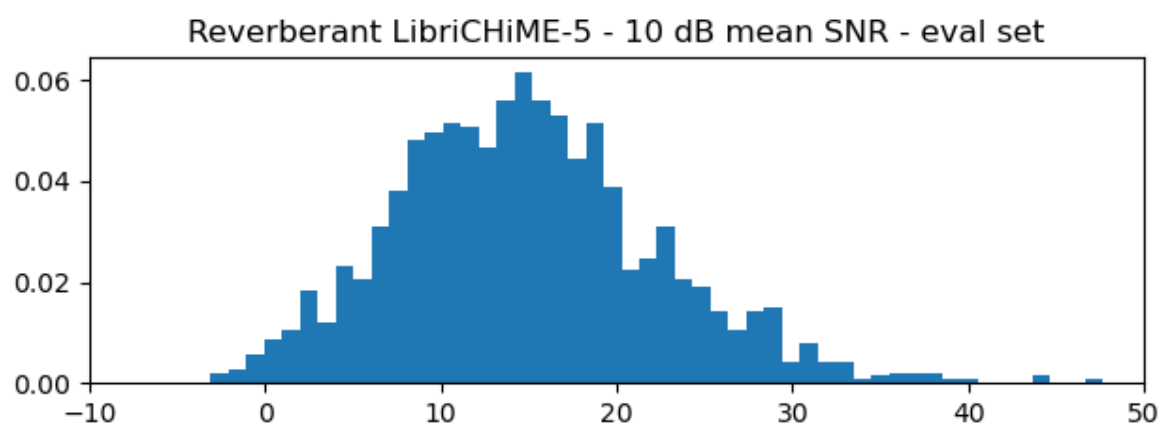
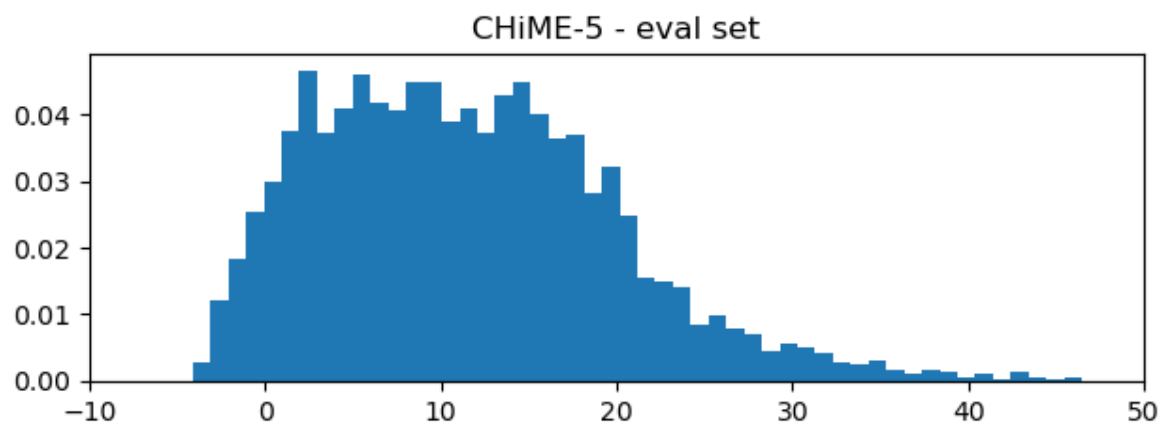
Below we plot the distribution (normalized histogram) of the SNR estimated using Brouhaha on the single-speaker noisy mixtures of (i) the CHiME-5 dataset (dev and eval sets) and (ii) the reverberant LibriCHiME-5 dataset (dev and eval sets), for two different values of mu: 10 dB and 5 dB.

The title of the figures indicates the dataset (CHiME-5 or reverberant LibriCHiME-5), the subset (dev or eval), the value of mu (10 or 5 dB).

We can draw the following conclusions:

- With $\mu = 5$ dB we have a much better match between the two datasets than with $\mu = 10$ dB.
- Brouhaha overestimates the SNR (we know the ground-truth mean SNR for the reverberant LibriCHiME-5).
- On the dev set, we could probably lower the value of mu a bit (like 1 or 2 dB).
- On the eval set of the CHiME-5 dataset, the distribution seems multimodal, which is probably because we have multiple scenarios in the CHiME-5 data (cooking, dining, chatting). We could have taken this information into account to create the reverberant LibriCHiME-5 dataset, e.g. by drawing the input SNR according to a mixture of Gaussians.





Input SI-SDR

Below we compute the input SI-SDR on the dev and eval sets of the reverberant LibriCHiME-5 dataset, by taking the noisy mixture as the speech estimate:

With a per-speaker mean SNR of **10 dB** we obtain:

si-sdr on dev (in dB)

- 1 speaker(s): 10.27
- 2 speaker(s): 13.86
- 3 speaker(s): 15.61

si-sdr on eval (in dB)

- 1 speaker(s): 10.65
- 2 speaker(s): 13.74
- 3 speaker(s): 15.61

With a per-speaker mean SNR of **5 dB** we obtain:

si-sdr on dev (in dB)

- 1 speaker(s): 5.27
- 2 speaker(s): 8.86
- 3 speaker(s): 10.61

si-sdr on eval (in dB)

- 1 speaker(s): 5.64
- 2 speaker(s): 8.73
- 3 speaker(s): 10.61

Conclusion

It seems reasonable to adjust the value of the input SNR in the generation of the reverberant LibriCHiME-5 dataset so that the SNR distribution estimated with Brouhaha on this dataset matches the SNR distribution estimated with Brouhaha on the original CHiME-5 dataset. Indeed, Brouhaha's SNR estimate is biased but with this procedure (matching the SNR distributions) the bias is not taken into account in the generation of the reverberant LibriCHiME-5 dataset. Also, the input SI-SDR values obtained with the 5 dB mean SNR version of the reverberant LibriCHiME-5 dataset are more reasonable.