Stabilizing Label Assignment for Speech Separation by Self-supervised Pre-training

Sung-Feng Huang[1], Shun-Po Chuang[1], Da-Rong Liu[1], Yi-Chen Chen[1],
Gene-Ping Yang[2], Hung-yi Lee[1]

[1]National Taiwan University, Taiwan
[2]University of Edinburgh, UK

{f06942045,f04942141,f07942148,f06942069}@ntu.edu.tw,
s2064029@ed.ac.uk, hungyilee@ntu.edu.tw

Abstract

Speech separation has been well developed, with the very successful permutation invariant training (PIT) approach, although the frequent label assignment switching happening during PIT training remains to be a problem when better convergence speed and achievable performance are desired. In this paper, we propose to perform self-supervised pre-training to stabilize the label assignment in training the speech separation model. Experiments over several types of self-supervised approaches, several typical speech separation models and two different datasets showed that very good improvements are achievable if a proper self-supervised approach is chosen.

Index Terms: Speech Enhancement, Self-supervised Pre-train, Speech Separation, Label Permutation Switch

1. Introduction

Supervised learning has been extremely successful in recent years in machine learning, except the huge quantity of labeled data needed causes the major problem. On the other hand, self-supervised learning tries to train the model using only unlabeled data, such as reconstructing the original data from some transformed representations or leveraging some parts of data to predict the other parts, therefore becomes highly attractive. In natural language processing (NLP) [1, 2, 3], BERT [1] learned powerful representations by self-supervised pre-training to encode contextual information. In computer vision (CV) [4, 5, 6, 7, 8], SimCLRv2 [5] outperformed the previous state-of-the-art on ImageNet by self-supervised pre-training. Examples in NLP and CV have shown self-supervised pre-trained models are more label-efficient than previous semi-supervised training methods. In the speech processing area, self-supervised learning also showed great advantages when labeled data are limited [6, 9, 10, 11, 12, 13, 14]. CPC [6] and APC [12] learned to extract useful representations for speech using a probabilistic contrastive loss to capture information for predicting future samples. Wav2vec [9] benefited from the idea of CPC and outperformed the state-of-the-art in character-based ASR with representations learned from 1000 hours of unlabeled speech. Wav2vec 2.0 [10] further showed that 10 minutes of labeled data were enough for training an ASR system with 53k hours of unlabeled data. TERA [14] pre-trained a Transformer model with a BERT-like objective. The learned representations were shown to be robust for a wide range of downstream tasks. The model could even outperform supervised learning when fine-tuned with only 0.1% of labeled data.

On the other hand, speech separation has long been a fundamental problem towards robust speech processing un-der the real-world acoustic environment, in which the considered speech signal is inevitably disturbed by some additional signals produced by other speakers. In general, deep learning techniques for single-channel speech separation can be divided into two categories: time-frequency (T-F) domain methods and end-to-end time-domain approaches. Based on T-F features obtained with short-time Fourier transform (STFT), T-F domain methods separate the T-F features for each source and then reconstruct the source waveforms by inverse STFT [15, 16, 17, 18, 19]. Time-domain approaches then directly process the mixture waveform using an encode-decoder framework, and this line of research has achieved significant progress in recent years [20, 21, 22, 23]. But both the T-F domain and time-domain approaches suffer from the label ambiguity problem when evaluating the reconstruction errors by matching the ground truths with the estimated signals. Permutation-invariant training (PIT) [24] has been very useful to handle this problem by dynamically choosing the best label assignment each time. However, the very unstable label assignment during the early training stage in PIT was shown to lead to slower convergence and lower performance [25].

In this paper, we made the following contributions:

- We point out the self-supervised pre-training is also extremely helpful to speech separation.

- We show the self-supervised pre-training can effectively stabilize the label assignment in PIT during training speech separation models, and the significantly reduced label assignment switching during training directly lead to faster convergence and improved performance.

- The proposed approach is shown to be equally useful to all different separation models over different datasets, because PIT has been widely used across almost all speech separation tasks.

2. Label ambiguity problem and permutation invariant training (PIT)

2.1. Label ambiguity problem

In single-channel speech separation, several speech signals are mixed: $y = \sum_{n=1}^{N} x_n$, where N is the number of sources; the goal is to extract all individual speech signals $\{x_n\}_{n=1}^{N}$ from the mixed signal y. For simplicity, we consider two sources only, $y = x_1 + x_2$, and employ a model with two outputs, o_1 and o_2. There exist two possible label assignments: (1) o_1 regresses to x_1 and o_2 regresses to x_2, or (2) o_1 regresses to x_2 and o_2 regresses to x_1. These two label assignments lead to two different loss functions to be used in model training. There are

$N!$ possible label assignments for $N \geq 2$. Incorrect label assignments naturally force the separation model to be updated to wrong direction, or even possibly destroy what has been learned before.

2.2. PIT and label assignment switching problem

Permutation invariant training (PIT) [24] was proposed to solve the above problem. Every time when the model parameters are to be updated, all possible label assignments as mentioned above are used to calculate the regression loss, and the one with minimum loss is chosen to update the model. Although such a dynamic label selection principle sounds reasonable, the selected labels can be very different for different training epochs giving a very rugged training path. A soft version of PIT was proposed to relax the label assignment switching problem between epochs [26], but restricted to those with $L2$-based objective functions only. A cascaded training strategy was then proposed [25], in which a good label assignment was first obtained with PIT, based on which the model parameters were better updated, to be used as a good initialization for the third stage of PIT training. This approach properly reduced the assignment switching during training, but made the training time several times longer compared to the original PIT.

3. Proposed training strategies

Considering the unstable label assignment problem during training as mentioned above, plus the fact that self-supervised pre-training was shown to be able to assist the model to learn structural information from large-scale unlabeled data and benefit in boosting the following training procedures [1, 2, 3, 4, 5, 6, 7, 8], we propose a self-supervised pre-training and fine-tuning framework as below.

3.1. Pre-train

In this work, we consider three different self-supervised approaches for pre-training here: speech enhancement (SE), Masked Acoustic Model with Alteration (MAMA) used in TERA [14], and continuous contrastive task (CC) used in wav2vec 2.0 [10]. Speech enhancement (SE) simply tries to reconstruct the original signal when noise is added to the input. MAMA is a masked reconstruction task, where the input audio is disturbed by noise with some parts randomly picked up and masked, and the model is required to reconstruct the clean audio of the masked parts. CC is a contrastive task; we mask the spans of the input audio features, and the model is trained to predict the masked spans of features correctly. Fig. 1(a) (colored part) shows the flowchart of pre-training, where the input signal is probably mixed with random noise and masked, and the model is to reconstruct the original clean source.

3.2. Fine-tune

After pre-training, the model is then fine-tuned with the normal separation training objective to produce the desired individual signals. All model parameters for fine-tuning are loaded from the pre-trained model as long as available, but the parameters used to generate specific output channels are re-initialized, as shown in Fig. 1(b). PIT is performed as usual.

3.3. Multi-task learning

To verify that whether the proposed framework really benefits from the "pre-train then fine-tune" procedure, jointly learning from the self-supervised training plus separation in a multi-task learning framework is also tested as a baseline for comparison, as in Fig. 1(c).

4. Experimental setup

4.1. Dataset

In this work, speech separation was trained and evaluated on WSJ0-2mix [18] and Libri2Mix [27] `train-100` set, and self-supervised approaches were trained using Libri1Mix [27] `train-360` set [27]. The WSJ0-2mix dataset was derived from the WSJ0 data corpus [28]. The training and validation data contained two-speaker mixtures generated by randomly selecting utterances from different speakers in the WSJ0 `si_tr_s` set, and the test set was similarly generated using utterances from unseen speakers in WSJ0 `si_dt_05` and `si_et_05` set. Libri2Mix is created based on the Librispeech dataset [29] with a similar generating procedure as WSJ0-2mix. Libri2Mix `train-100` set used speakers randomly selected from the `train-clean-100` set of Librispeech, while the dev and test set used the utterances from unseen speakers in the `dev` and `test` sets of Librispeech respectively. Libri1Mix `train-360` dataset was created with the same settings as Libri2Mix, while only one speaker was randomly selected from the `train-clean-360` set of Librispeech, and mixed with a random noise sampled from WHAM! [30]. The speaker groups of Libri1Mix `train-360` and Libri2Mix `train-100` set were disjoint.

4.2. Implementation details

The proposed self-supervised pre-training can be used with any separation model, and the study here was mainly focused on the effectiveness of pre-training. In this work, we choose Conv-TasNet [20] as our main baseline model, DPRNN [21] and DPT-Net [22] were also used in later experiments. All experiments were implemented with Asteroid [31], and the detailed training configurations are in the repository[1].

The model was trained with three different strategies for comparison in our experiments: from scratch, pre-trained then fine-tune (PT-FT), and multi-task training. We purposely let the three strategies have the same number of update steps in training the separation task for fairness. Separation performance was evaluated in scale-invariant signal-to-noise ratio improvement (SI-SNRi) [32] and signal-to-distortion ratio improvement (SDRi) [33].

5. Experimental results

5.1. Comparison between the self-supervised pre-training tasks

We first wished to find out which self-supervise pre-training approach was more helpful to the separation task. Speech enhancement (SE), Masked Acoustic Model with Alteration (MAMA) and continuous contrastive task (CC) as described in Section 3.1 were tested. We used Conv-TasNet as our separation model. After pre-trained with SE, MAMA and CC respectively, we fine-tuned the obtained models for 100 epochs for the speech separation task. The pre-training tasks were all trained on Libri1Mix `train-360` set, and the fine-tuning task was trained on WSJ0-2mix. The results listed in Table 1 showed

---

[1]`https://github.com/SungFeng-Huang/`
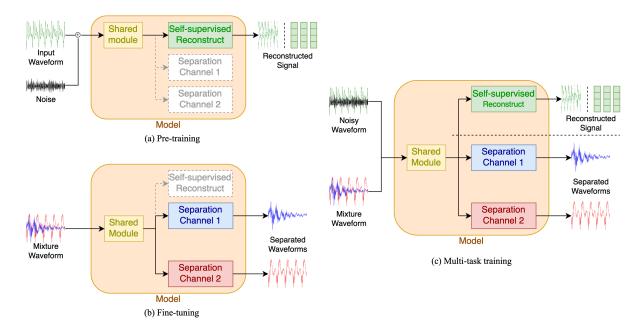`SSL-pretraining-separation/tree/main/local`

Figure 1: *The flowchart for the proposed training framework: (a) pre-training, (b) fine-tuning after pre-training, and (c) multi-task training for comparison. Gray blocks indicate the corresponding parts are not used during training.*

Table 1: *Comparison between different self-supervised pre-training approaches when fine-tuned with Conv-TasNet in SI-SNRi and SDRi. The first row is for training from scratch.*

Pre-training task	SI-SNRi (dB)	SDRi (dB)
–	15.6	15.8
SE	**16.3**	**16.5**
MAMA [14]	16.2	**16.5**
CC [10]	15.5	15.8

Table 2: *Comparison between different training strategies for Conv-TasNet on two datasets (WSJ0-2mix and Libri2Mix) in SI-SNRi (dB) and SDRi (dB).* **PT-FT**: *pre-trained then fine-tune.* **SE**: *with speech separation for self-supervised. Number in the parentheses are the improvements over "from scratch".*

Corpus	Training strategy	SI-SNRi	SDRi
Libri2Mix	from scratch	13.2	13.6
	PT-FT (SE)	**14.1 (0.9)**	**14.6 (1.0)**
	multi-task (SE)	13.7 (0.5)	14.1 (0.5)
WSJ0-2mix	from scratch	15.6	15.8
	PT-FT (SE)	**16.3 (0.7)**	**16.5 (0.7)**
	multi-task (SE)	15.7 (0.1)	16.0 (0.2)

that both SE and MAMA led to significant improvement, but not CC. Note that both SE and MAMA had input speech disturbed by noise, while the model was to reconstruct the whole utterance (SE) or only the masked parts (MAMA), as mentioned in Section 3.1. So we may conclude that approaches trying to reconstruct the clean input speech from the noisy and/or masked one are probably more effective for pre-training speech separation tasks. A possible explanation may be here SE and MAMA already learned to extract from disturbed signals the information about each individual speaker, so all the following Conv-TasNet model needed to learn is to separate the extracted information into two channels, therefore the learning process was more stable and efficient. This is why in the following tests we only used speech enhancement (SE) for self-supervised pre-training.

5.2. Effectiveness of pre-training and fine-tuning (PT-FT)

Table 2 shows the results of three different training strategies: from scratch, pre-training and fine-tuning (PT-FT) and multi-task, with the latter two using speech enhancement (SE) for self-supervised learning, all trained with Conv-TasNet as the main separation model. As shown, both pre-training and multi-task learning improved the separation model on both WSJ0-2mix and Libri2Mix, while pre-training improved more significantly (0.7 - 1.0 dB improvements for PT-FT (SE) compared to "from scratch" v.s. 0.1 - 0.5 dB for multi-task (SE)). A good explana-

tion for this is that, as mentioned above, the pre-trained model already learned to extract from disturbed signals the information about the individual target speakers, so the following separation model could focus on the construction of the two masks, for the two sources. In contrast, for multi-task learning, the two different tasks of speech enhancement and speech separation were learned jointly, while sharing the knowledge learned for the two very different tasks may not be easy. This further showed the effectiveness of learning the two different tasks sequentially instead of jointly (self-supervised for enhancement then fine-tuning for separation). Also noted that since the corpus used to train speech enhancement was Libri1Mix `train-360` set, which was closer to Libri2Mix `train-100` set but farther from WSJ0-2mix, which may be the simple reason why the results in Table 2 on Libri2Mix (upper half) showed more improvements than those on WSJ0-2mix (lower half).

Validation SI-SNR results during training are reported in Figure 2(a)(b) for Libri2Mix and WSJ0-2mix respectively. As shown in the figure, improvements for multi-task learning gradually decreased while training on Libri2Mix and were nearly hard to see on WSJ0-2mix. The proposed pre-trained model
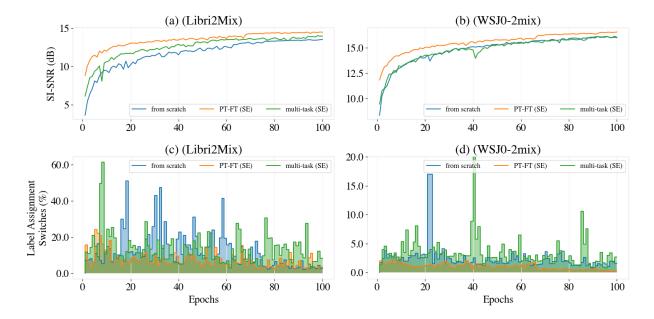
Figure 2: *(a)(b) validation SI-SNR (dB) and (c)(d) percentage of label assignment switches in total training data (%) at each epoch on two datasets Libri2Mix and WSJ0-2mix respectively. In (d), the green bars reach 89% at around epoch 40.*

led the baselines all the way and achieved the final result of the baselines in only 37 epochs for Libri2Mix and 66 epochs for WSJ0-2mix, which are about one-third to two-thirds of the baseline training epochs. Figure 2(c)(d) show the percentage of label assignment switches in total training data. Here we can see only the proposed pre-training with speech enhancement (orange bars) significantly reduced label assignment switches, while multi-task learning (green bars) not only failed to reduce the label assignment switches but sometimes increased them. Moreover, training from scratch (blue bars) and multi-task learning (green bars) sometimes got very high switching percentages (e.g., roughly 15% - 35 % of the label assignments were often switched over for both training from scratch and multi-task training in Figure 2(c), and most label assignments were switched at epoch 40 for multi-task training in Figure 2(d).

5.3. More Separation models tested on WSJ0-2mix

More test results on different separation models with different batch sizes (BS), utterance length (L), with pre-training (PT) or from scratch are listed in Table 3, all trained and evaluated on WSJ0-2mix. The first rows (a)(d)(g)(l) for each model are those reported in their original papers. For speeding up the experiments, Conv-TasNet and DPRNN (Sec. (I)(II) in Table 3) were trained with shorter utterance length (3 or 2 sec) and a larger batch size (24) with 200 epochs, which caused the slightly worse DPRNN results than those previously reported [21]. In addition to those for Conv-TasNet discussed previously, the pre-trained DPRNN (Sec. (II)) was improved significantly, even achieving comparable performance as the reported one (rows (f) v.s. (d)(e)), although with worse performance from scratch due to the hyper-parameters. DPTNet (Sec. (III)) was trained with batch size 1 and 4 with 100 epochs to speed up the training process. Setting batch size 4 instead of 1 gave 0.3 dB worse performance (rows (h) v.s. (j)). Nevertheless, the pre-trained DPTNet made up the gap, even doing slightly better (rows (i) v.s. (j)). Compared to the current state-of-the-art (Sandglasset [23]), the pre-trained DPTNet with a batch size 1 actually

Table 3: *Different separation models on WSJ0-2mix in SI-SNRi (dB) and SDRi (dB).* **BS***: batch size.* **L***: utterance length (sec).* **PT***: pre-training, "–" means training from scratch. The first rows (a)(d)(g)(l) for each model are the reported results from original papers. The blank indicates unknown. *Row (g) are actually SI-SNR and SDR.*

	Model	BS	L	PT	SI-SNRi	SDRi
(a)	(I) Conv-TasNet		4	–	15.3	15.6
(b)		24	3	–	15.6	15.8
(c)		24	3	SE	**16.3**	**16.5**
(d)	(II) DPRNN		4	–	18.8	19.0
(e)		24	2	–	17.0	17.3
(f)		24	2	SE	**18.6**	**18.9**
(g)	(III)DPTNet		4	–	20.2*	20.6*
(h)		4	4	–	20.4	20.6
(i)		4	4	SE	**20.8**	**21.0**
(j)		1	4	–	20.7	20.9
(k)		1	4	SE	**21.3**	**21.5**
(l)	(IV) Sandglasset		4	–	21.0	21.2

achieved the new state-of-the-art ((k) v.s. (l)).

6. Conclusion

In this paper, we propose to use self-supervised pre-training to stabilize the label assignment for speech separation. We show that pre-training with speech enhancement offers better training and consistently improves the separation performance across all different separation model architectures over two different datasets.

7. Acknowledgements

providing computational and storage resources.

8. References

[1] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[2] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep contextualized word representations," *arXiv preprint arXiv:1802.05365*, 2018.

[3] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding by generative pre-training," 2018.

[4] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," *arXiv preprint arXiv:2002.05709*, 2020.

[5] T. Chen, S. Kornblith, K. Swersky, M. Norouzi, and G. Hinton, "Big self-supervised models are strong semi-supervised learners," *arXiv preprint arXiv:2006.10029*, 2020.

[6] A. v. d. Oord, Y. Li, and O. Vinyals, "Representation learning with contrastive predictive coding," *arXiv preprint arXiv:1807.03748*, 2018.

[7] P. Bachman, R. D. Hjelm, and W. Buchwalter, "Learning representations by maximizing mutual information across views," in *Advances in Neural Information Processing Systems*, 2019, pp. 15 535–15 545.

[8] O. J. Hénaff, A. Srinivas, J. De Fauw, A. Razavi, C. Doersch, S. Eslami, and A. v. d. Oord, "Data-efficient image recognition with contrastive predictive coding," *arXiv preprint arXiv:1905.09272*, 2019.

[9] S. Schneider, A. Baevski, R. Collobert, and M. Auli, "wav2vec: Unsupervised pre-training for speech recognition," *arXiv preprint arXiv:1904.05862*, 2019.

[10] A. Baevski, H. Zhou, A. Mohamed, and M. Auli, "wav2vec 2.0: A framework for self-supervised learning of speech representations," *arXiv preprint arXiv:2006.11477*, 2020.

[11] Y.-A. Chung, C.-C. Wu, C.-H. Shen, H.-Y. Lee, and L.-S. Lee, "Audio word2vec: Unsupervised learning of audio segment representations using sequence-to-sequence autoencoder," *arXiv preprint arXiv:1603.00982*, 2016.

[12] Y.-A. Chung, W.-N. Hsu, H. Tang, and J. Glass, "An unsupervised autoregressive model for speech representation learning," *arXiv preprint arXiv:1904.03240*, 2019.

[13] A. T. Liu, S.-w. Yang, P.-H. Chi, P.-c. Hsu, and H.-y. Lee, "Mockingjay: Unsupervised speech representation learning with deep bidirectional transformer encoders," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 6419–6423.

[14] A. T. Liu, S.-W. Li, and H.-y. Lee, "Tera: Self-supervised learning of transformer encoder representation for speech," *arXiv preprint arXiv:2007.06028*, 2020.

[15] C.-L. Hsu and J.-S. R. Jang, "On the improvement of singing voice separation for monaural recordings using the mir-1k dataset," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 2, pp. 310–319, 2009.

[16] P.-S. Huang, M. Kim, M. Hasegawa-Johnson, and P. Smaragdis, "Deep learning for monaural speech separation," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2014, pp. 1562–1566.

[17] J. Bruna, P. Sprechmann, and Y. LeCun, "Source separation with scattering non-negative matrix factorization," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2015, pp. 1876–1880.

[18] J. R. Hershey, Z. Chen, J. Le Roux, and S. Watanabe, "Deep clustering: Discriminative embeddings for segmentation and separation," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 31–35.

[19] Z. Chen, Y. Luo, and N. Mesgarani, "Deep attractor network for single-microphone speaker separation," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017, pp. 246–250.

[20] Y. Luo and N. Mesgarani, "Conv-tasnet: Surpassing ideal time–frequency magnitude masking for speech separation," *IEEE/ACM transactions on audio, speech, and language processing*, vol. 27, no. 8, pp. 1256–1266, 2019.

[21] Y. Luo, Z. Chen, and T. Yoshioka, "Dual-path rnn: efficient long sequence modeling for time-domain single-channel speech separation," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 46–50.

[22] J. Chen, Q. Mao, and D. Liu, "Dual-path transformer network: Direct context-aware modeling for end-to-end monaural speech separation," *arXiv preprint arXiv:2007.13975*, 2020.

[23] M. W. Lam, J. Wang, D. Su, and D. Yu, "Sandglasset: A light multi-granularity self-attentive network for time-domain speech separation," *arXiv preprint arXiv:2103.00819*, 2021.

[24] D. Yu, M. Kolbæk, Z.-H. Tan, and J. Jensen, "Permutation invariant training of deep models for speaker-independent multi-talker speech separation," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017, pp. 241–245.

[25] G.-P. Yang, S.-L. Wu, Y.-W. Mao, H.-y. Lee, and L.-s. Lee, "Interrupted and cascaded permutation invariant training for speech separation," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 6369–6373.

[26] M. Yousefi, S. Khorram, and J. H. Hansen, "Probabilistic permutation invariant training for speech separation," *Proc. Interspeech 2019*, pp. 4604–4608, 2019.

[27] J. Cosentino, M. Pariente, S. Cornell, A. Deleforge, and E. Vincent, "Librimix: An open-source dataset for generalizable speech separation," *arXiv preprint arXiv:2005.11262*, 2020.

[28] J. Garofolo, D. Graff, D. Paul, and D. Pallett, "Csr-i (wsj0) complete ldc93s6a," *Web Download. Philadelphia: Linguistic Data Consortium*, vol. 83, 1993.

[29] V. Panayotov, G. Chen, D. Povey, and S. Khudanpur, "Librispeech: an asr corpus based on public domain audio books," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2015, pp. 5206–5210.

[30] G. Wichern, J. Antognini, M. Flynn, L. R. Zhu, E. McQuinn, D. Crow, E. Manilow, and J. L. Roux, "Wham!: Extending speech separation to noisy environments," *arXiv preprint arXiv:1907.01160*, 2019.

[31] M. Pariente, S. Cornell, J. Cosentino, S. Sivasankaran, E. Tzinis, J. Heitkaemper, M. Olvera, F.-R. Stöter, M. Hu, J. M. Martín-Doñas *et al.*, "Asteroid: the pytorch-based audio source separation toolkit for researchers," *arXiv preprint arXiv:2005.04132*, 2020.

[32] Y. Luo, Z. Chen, and N. Mesgarani, "Speaker-independent speech separation with deep attractor network," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 4, pp. 787–796, 2018.

[33] E. Vincent, R. Gribonval, and C. Févotte, "Performance measurement in blind audio source separation," *IEEE transactions on audio, speech, and language processing*, vol. 14, no. 4, pp. 1462–1469, 2006.