Interspeech 2024
1-5 September 2024, Kos, Greece

OR-TSE: An Overlap-Robust Speaker Encoder for Target Speech Extraction

Yiru Zhang[1], *Linyu Yao*[1], *Qun Yang*[1,*]

[1]College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, China

zhangyr@nuaa.edu.cn, 2425398873@qq.com, qun.yang@nuaa.edu.cn

Abstract

Mainstream Target Speech Extraction (TSE) systems extract target speech from a mixture using pre-enrolled reference speech. The extraction performance heavily depends on the quality of the reference speech. However, the speech signal of the same speaker may vary under different conditions, leading to a decrease in extraction performance, particularly in speech overlap. Therefore, we propose an overlap robust speaker encoder for TSE to obtain stable speaker embeddings even when using signals with overlapping interference. Our approach combines attentive statistics pooling with contrastive learning to make the model focus on the feature of main speaker while disregarding interfering information. Based on our proposed speaker encoder, we introduce a TSE framework, which derive speaker embeddings from non-overlapping regions of mixture input. The experiments shows that our speaker encoder improves the performance of TSE in different conditions of reference speech.

Index Terms: target speech extraction, speaker embedding, contrastive learning, attentive statistics pooling

1. Introduction

Target speech extraction is a technique that separates speech signals from a multi-talker mixture utilizing the identity information of the target speaker. Presently, the TSE systems mainly use pre-enrolled reference speech as identity information and have achieved good performance [1, 2, 3, 4]. These models commonly employ a speaker encoder to encode the reference speech into speaker embedding, where the quality of the reference speech is crucial. However, the speech signals of the same speaker may vary significantly in different conditions due to factors such as acoustic environments, emotional states or channel effects [5, 6], leading to a decrease in model performance.

Specifically, in most application scenarios, the reference speech still contains overlap from other speakers because it is impractical to provide a quiet environment to record clean reference speech. Some researchers propose to estimate non-overlapping single speech from long-form mixture inputs as reference speech [7, 8, 9, 10]. For example, TS-SEP [9] employs a speaker diarization model and CSSUSI [10] utilizes clustering method to estimate the reference speech from non-overlapped regions of mixed signals. Nevertheless, there will inevitably be slight overlaps with other speakers in the reference speech estimated by these methods, especially at segment boundaries. This necessitates that the speaker encoder exhibits sufficient robustness to address such internal variations within the same speaker.

Therefore, we need to improve the speaker encoder to ensure stable speaker embedding even when using overlapped reference speech. Most robust speaker encoders mainly focus on improve robustness in noisy environments, while there is limited research investigating the interference with overlaps[11, 12, 13, 14]. Some studies [12, 13] propose data augmentation techniques, where each training sample is mixed with random training samples to generate partially overlapping signals, making the speaker encoder aware of overlapped speech. UniSpeech-SAT [13] and WavLM [14] further train the model with Self-Supervised Learning (SSL). However, these methods only improve the training strategy and ignore the improvement of the model structure.

We found that Attentive Statistics Pooling (ASP) [15] can direct the network focus on important frames, thereby remove the disturbed information. For example, EA-ASP [16] proposes an enrollment-aware attentive statistical pooling layer to selectively direct the speaker encoder's attention to the target speaker when the speech is interfered by another speaker. This method requires enrollment speech guidance. While in this paper, we need to make the model pay attention to the information of the main speaker in an utterance and ignore the interference from other speakers.

Therefore, we propose an overlap-robust speaker encoder for target speech extraction. Leveraging the speaker encoder, we introduce an Overlap-Robust TSE (OR-TSE) framework that estimates reference speech from long-form mixture input, which does not require speaker enrollment. Our proposed speaker encoder includes three contributions: (1) We propose a mix-mask training strategy, which generate positive samples to make the model aware of speech overlaps. (2) We employ an ASP layer to assign higher weights to segments where the main speaker speaks alone and reduce the weight of the disturbed speech. (3) We adopt contrastive learning and propose a new contrastive loss to pre-train the speaker encoder, making the model address the internal variations within the same speaker and better distinguish different speakers. The experiment results on LibriSpeech indicate that our speaker encoder outperforms baseline methods whether using estimated or pre-enrolled reference speech.

2. Method

2.1. The network structure of OR-TSE framework

We first introduce the overall structure of our proposed OR-TSE framework, as shown in Figure 1, it contains five blocks: reference speech estimation module, speaker encoder, speech encoder, speaker extractor and speech decoder. The input of the framework is only a long form mixture speech y. We employ a pre-trained speaker diarization model as reference speech estimation module, generating non-overlapping reference speech x of participating speakers [8, 9]. Our proposed overlap-robust

10.21437/Interspeech.2024-2322

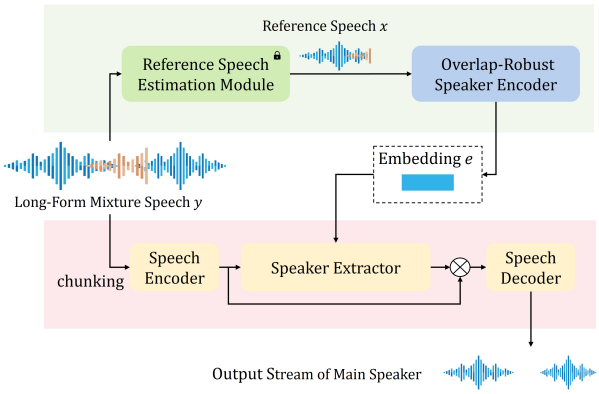
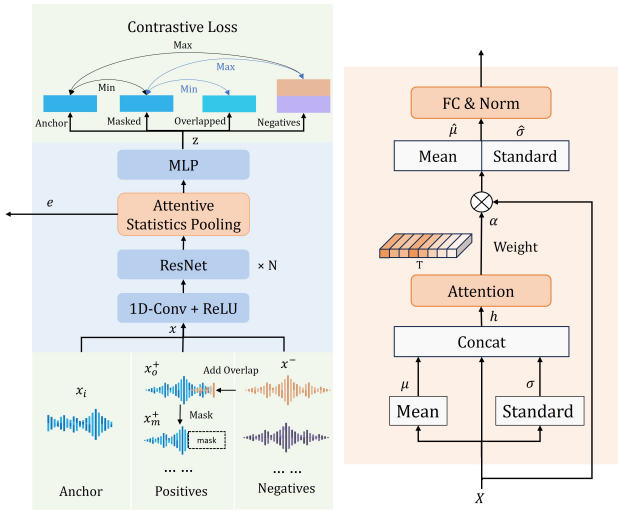


Figure 1: *The network structure of our OR-TSE framework.*



(a) The process of speaker encoder (b) Attentive statistics pooling

Figure 2: *The network of overlap-robust speaker encoder.*

speaker encoder extracts speaker embedding e from x. The speech encoder encodes the waveform y into acoustic representation. Next, the speaker extractor estimates the mask of the target speaker based on the mixed speech representation and the target speaker embedding e to filter out the speech of other speakers. And the speech decoder can reconstruct the speech signal from the masked speech representation.

2.2. Overlap-robust speaker encoder

As shown in Figure 2(a), our proposed robust speaker encoder contains three improvements. Firstly, we propose a mix-mask training strategy to generate positive and negative samples. Subsequently, we encode these samples using the speaker encoder with an ASP layer. Finally, we optimize the network with a novel contrastive loss function, minimizing the distance over same speakers and maximizing the distance over different speakers.

2.2.1. Mix-mask training strategy

We aim to facilitate the model in acquiring stable embedding of the main speaker, while simultaneously disregarding the presence of minor interfering speakers. To this end, we propose a novel approach to construct positive and negative samples for contrastive learning, which is shown in the bottom of Figure 2(a).

Specifically, we first denote a clean utterance of a certain speaker as the anchor x_i. And then, we select utterances from different speakers as negative samples. Our positive samples consist of two parts: the overlapped positive samples x_o^+ and the masked positive samples x_m^+. The overlapped positive samples is generated by data augmentation technique [12]. The details of data augmentation is described as follows.

We randomly select an utterance x^+, which has the same speaker as the anchor x_i. Additionally, we choose an interfering utterance x^- from a different speaker. Next, we randomly crop and scale x^- to a length between 0%-15% of x^+ to ensure that x^+ is the main speaker. Finally, we add the cropped x_{crop}^- to a random position of x^+ to generate the overlapped positive sample x_o^+, described as:

$$x_o^+ = x^+ + x_{crop}^-. \tag{1}$$

After generating overlapped positives, we mask the overlapped part of the aforementioned x_o^+ to generate the masked positive samples x_m^+:

$$x_m^+ = M \otimes x_o^+, \tag{2}$$

where $M \in \{0,1\}^L$ represent a binary mask that set the overlapped part of x_o^+ to 0 and L is the length of the utterance. As a result, x_m^+ is a clean utterance without other speaker.

By minimizing the distance between overlapped sample x_o^+ and corresponding masked sample x_m^+, the model can be encouraged to reduce the weight of the disturbed part in the reference speech.

2.2.2. Network structure of speaker encoder

The network structure of our speaker encoder is shown in Figure 2(a). We employ a 1D-Conv with ReLU activation to encode the reference speech x into feature representation. Subsequently, we use ResNet blocks with a number of N to generate the representation $X \in \mathbb{R}^{C\times T}$, where C represents channels and T denotes time frames. Next, the representation X is fed into an ASP layer, which performs temporal pooling, resulting in a fixed-dimensional speaker embedding vector e.

The specific of the ASP layer is illustrated in Figure 2 (b). We employ the attention mechanism for a concatenated feature h of mean vector μ and standard vector σ of X:

$$s_t = v^T f(Wh_t + b) + k, \tag{3}$$

where $h = concat(X, \mu, \sigma)$ means concatenate in the time dimension and h_t is the feature at time step t. W and b are the parameters of the feature projection layer. $f(\cdot)$ is a non-linearity function. A linear layer with the parameters v and k is used to transform the representation into a self-attention score s_t. Then, we apply a softmax function across time dimension to the attention score s_t:

$$\alpha_t = \frac{exp(s_t)}{\sum_\tau^T exp(s_\tau)}, \tag{4}$$

where α_t is the self-attention weight that represents the importance of each frame. We estimate the weighted mean vector μ and standard vector σ as follows:

$$\hat{\mu} = \sum_t^T \alpha_t X_t, \tag{5}$$

$$\hat{\sigma} = \sqrt{\sum_{t}^{T} \alpha_t X_t^2 - \hat{\mu}^2}. \quad (6)$$

The output of the ASP layer is given by concatenating the weighted mean vector $\hat{\mu}$ and standard vector $\hat{\sigma}$.

2.2.3. Loss function

In this section, we propose a new contrastive loss with multiple positives and multiple negatives to learn speaker invariance in speech under different conditions. As shown in top of Figure 2(a), our loss function consists of two parts.

Firstly, we minimize the distance between the anchor x_i and masked positive samples x_m^+ to capture features belonging to the same speaker. And we maximize the distance between the anchor and negative samples to help the model better differentiate between different speakers. The pre-training contrastive loss can be described as:

$$L_1 = \frac{1}{|P_m(i)|} \sum_{m \in P_m(i)} \log \frac{exp(z_i \cdot z_m^+ / \tau)}{\sum_{j=0}^{k} exp(z_i \cdot z_j / \tau)}, \quad (7)$$

where z denotes the speaker representations calculated by different samples x. $P_m(i)$ represents the index set of masked positive samples of the anchor x_i. k represents the total number of positive and negative samples, and j represents their index. τ is a scalar temperature parameter.

For the second part of loss function, we minimize the distance between the overlapped positive samples x_o^+ and corresponding masked positive samples x_m^+ to mitigating the impact of a small amount of interfering speech. At the same time, we maximize the distance between the masked samples and negative samples:

$$L_2 = \frac{1}{|P_m(i)|} \sum_{m \in P_m(i), o \in P_o(i)} \log \frac{exp(z_m^+ \cdot z_o^+ / \tau)}{\sum_{j=0}^{k} exp(z_m^+ \cdot z_j / \tau)}, \quad (8)$$

where $P_o(i)$ represents the index set of overlapped positive samples corresponding to the masked sample x_m^+. Each masked sample maintains a one-to-one correspondence with overlapped sample.

The contrastive loss is the sum of L_1 and L_2:

$$L = L_1 + L_2. \quad (9)$$

The speaker encoder is pre-trained by optimizing the contrastive loss to improve the robustness. And we further fine-tune it by speech extraction task, ensuring the performance of the entire system.

3. Experimental setup

3.1. Dataset

In our experiments, the training set of TSE is based on the "train-clean-100" subset of LibirSpeech [17], which contains 100 hours of speech from 251 speakers with a sampling rate of 16kHz. We randomly select 2 utterances of different speakers, and mix them according the patterns of fully overlap, partially overlap, instantaneous overlap, single person speaking and alternating speaking, etc. [10]. It can adapt training data to various situations that may occur in long recording. The mixed data set contains 28938 utterances, and each utterance lasts about 10s. The data for pre-training the speaker encoder contains an utterance, 4 positive samples and 4 negative samples, which are derived from LibriSpeech.

We evaluate our framework on simulated long recordings with a number of 100 based on LibriSpeech, and each recording is about 180s long with 30% overlap ratio. During testing, we use the utterances in LibriSpeech as pre-enrolled clean reference speech. Additionally, we employ the reference speech estimation module to derive estimated reference speech from the long recordings.

3.2. Settings

First of all, we pre-train our speaker encoder using contrastive learning with the initial learning rate of 0.0005. And then, we jointly optimize the speaker encoder with other parts of our OR-TSE framework. In our experiment, we replace the speaker encoder with two other overlap-robust speaker representation methods, and compare the extraction performance. The baselines include the model trained with Data Augmentation(DA)[12] and the WavLM-Base+[14] model.

The main structure of our speech extraction network is based on the backbone of SPEX+ [18], which is our baseline model. The speech encoder and speech decoder are 1D-CNNs of multi-scales. And the speaker extractor contains 4 temporal convolutional network (TCN) stacks. We train the model for a maximum of 150 epochs and employ early stopping strategy. And we use Adam optimizer with an initial learning rate of 0.001. Due to the low overlap ratio of our dataset, we employ the loss function of USE [19, 20] to output silence when the target speaker is absent. Due to the limited computation resources, we use a pre-trained SPEX+ model to initialize parameters and subsequently modify and fine-tune it.

The reference speech estimation module [8] in our framework is based on the pre-trained overlap detection and speaker diarization model in pyannote toolkit [21, 22].

3.3. Evaluation metrics

We use the Signal to Distortion Ratio (SDR) and Scale-Invariant SDR (SI-SDR) [23] to evaluate the quality of speech extracted from overlapped speech. To quantify the robustness of the model, we employ the value of SI-SDR Reduction (SI-SDRR) to calculate the performance degradation concerning the estimated reference speech compared to the pre-enrolled clean reference speech. A lower SI-SDRR value means a smaller impact on the model's performance when using estimated reference speech, indicating better robustness.

4. Experiment results

4.1. Comparisons with overlap-robust speaker encoder

In this section, we compare three overlap robust speaker encoders using pre-enrolled clean reference speech and estimated reference speech, which is shown in Table 1. Since estimated reference speech may contain overlap interference, a performance degradation of 1.08 dB is observed when we use spex+ [18] baseline speaker encoder without any progress.

The robust speaker encoders can alleviate the performance degradation. The results in Table 1 demonstrate that direct application of DA and WavLM enhances model stability, resulting in only 0.29 dB and 0.27 dB SI-SDRR, but the overall performance dropped. It could be due to the complex features captured by pre-trained WavLM model are not suitable for our task. While our proposed method not only surpasses other methods

Table 1: *Comparison of speaker encoders in terms of SI-SDR and SI-SDR reduction.*

Method	Reference	SI-SDR(dB)	SI-SDRR(dB)
Baseline [18]	pre-enrolled	11.46	1.08
	estimated	10.38	
DA [12]	pre-enrolled	10.56	0.29
	estimated	10.27	
WavLM [14]	pre-enrolled	10.27	0.27
	estimated	10.00	
OR-TSE	pre-enrolled	13.18	0.33
	estimated	12.85	

Table 2: *Ablation Study of different configurations using pre-enrolled and estimated reference speech.*

Method	Reference	SI-SDR(dB)	SI-SDRR (dB)
OR-TSE	pre-enrolled	13.18	0.33
	estimated	12.85	
-ASP	pre-enrolled	11.96	0.73
	estimated	11.23	
-CL	pre-enrolled	11.46	1.08
	estimated	10.38	

in overall performance but also experiences a 0.33 dB SI-SDR reduction when using estimated reference speech.

4.2. Ablation study

We conduct ablation study to investigate the contributions of different configurations in our proposed framework. As shown in Table 2, "- ASP" and "-CL" respectively means we remove the ASP layer and contrastive learning from our speaker encoder. The experimental results indicate that both the contrastive learning and ASP layer can improve the performance of speech extraction and the model's stability.

Specifically, Figure 3 illustrates the SI-SDR and SI-SDRR variations for the three configurations when using reference speech with different overlap rates. The Figure 3(a) visually demonstrate that our proposed OR-TSE model outperforms other approaches in overall performance. Figure 3(b) shows the relative values of SI-SDR from 0% overlap ratio, indicating that using contrastive learning and ASP layer exhibits high stability when reference speech experiences severe interference. Especially in high overlap ratio (more than 20%), our methods effectively reduce the performance degradation of the model.

In Figure 4, we visualize the weights in the ASP layer to verify its ability to focus on overlapped segments and reduce

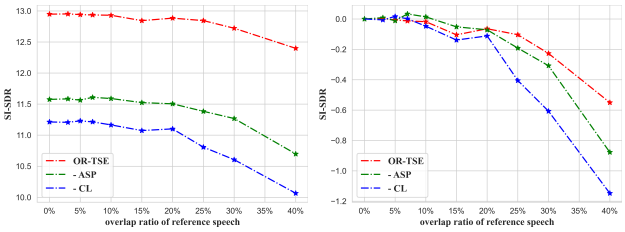


Figure 3: *The plot of SI-SDR using reference speech with different overlap ratio in three model configurations.*

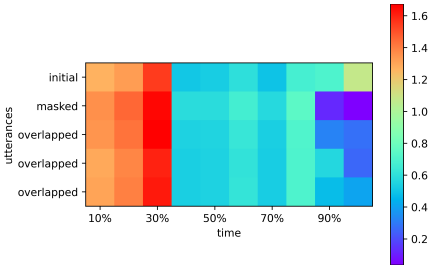


Figure 4: *The heatmap of the weights of ASP layer.*

its weights. The horizontal axis represents the time frame in an utterance and the vertical axis represents different utterances. The row "masked" represents the condition that masking 20% in the end of "initial" utterance. The results clearly indicate that the ASP layer can accurately assign lower weights to masked silent segments. The following three "overlapped" rows represent that adding different overlap utterances in aforementioned conditions. It can be seen that the weights of overlapped part are lower than initial utterance. It demonstrates that the ASP layer has recognized the overlapping segments and successfully guided the model to ignore the overlapped interfering parts, focusing more on the speech of the main speaker.

4.3. Comparisons with TSE models for long recordings

In this section, we compare the SDR of our framework with other TSE models designed for long recordings, as presented in Table 3. The first three rows represent models requiring pre-enrolled reference speech. And the last four rows represent the models estimating reference speech from the input. CSSUSI compensate the inaccuracies of estimated reference speech by choosing the necessary reference speech for extraction. Our proposed method not only improves the performance using clean reference speech, but also achieves a stable speaker encoder to compensate for estimation inaccuracies, ultimately resulting in optimal performance of 12.8 dB.

Table 3: *SDR on long form mixture speech using different TSE model.*

Method	Reference	SDR(dB)
BLSTM [10]	pre-enrolled	11.2
SSUSI [24]	pre-enrolled	12.2
OR-TSE (ours)	pre-enrolled	13.8
CSSUSI [10]	estimated	12.1
DSS [7]	estimated	9.3
SPEX+ [18]	estimated	11.5
OR-TSE (ours)	estimated	12.8

5. Conclusion

This paper proposed a TSE framework with an overlap-aware robust speaker encoder. We introduced the ASP layer to the speaker encoder to make the model selectively focus only on the target speaker and disregard interfering speakers. And we trained it using contrastive learning with a mix-mask training strategy to make the model adapt to multiple reference speech. The experiments on simulated long form speech showed that our speaker encoder makes the TSE model more robust whether using reference speech in clean or multi-talker conditions.

6. References

[1] Q. Wang, H. Muckenhirn, K. Wilson, P. Sridhar, Z. Wu, J. R. Hershey, R. A. Saurous, R. J. Weiss, Y. Jia, and I. L. Moreno, "VoiceFilter: Targeted Voice Separation by Speaker-Conditioned Spectrogram Masking," in *Proc. Interspeech 2019*, 2019, pp. 2728–2732.

[2] K. Žmolíková, M. Delcroix, K. Kinoshita, T. Ochiai, T. Nakatani, L. Burget, and J. Černocký, "Speakerbeam: Speaker aware neural network for target speaker extraction in speech mixtures," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 4, pp. 800–814, 2019.

[3] C. Xu, W. Rao, E. S. Chng, and H. Li, "Spex: Multi-scale time domain speaker extraction network," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 1370–1384, 2020.

[4] K. Liu, Z. Du, X. Wan, and H. Zhou, "X-sepformer: End-to-end speaker extraction network with explicit optimization on speaker confusion," 2023.

[5] K. Zmolikova, M. Delcroix, T. Ochiai, K. Kinoshita, J. Černockỳ, and D. Yu, "Neural target speech extraction: An overview," *IEEE Signal Processing Magazine*, vol. 40, no. 3, pp. 8–29, 2023.

[6] C. Deng, S. Ma, Y. Sha, Y. Zhang, H. Zhang, H. Song, and F. Wang, "Robust Speaker Extraction Network Based on Iterative Refined Adaptation," in *Proc. Interspeech 2021*, 2021, pp. 3530–3534.

[7] R. Paturi, S. Srinivasan, K. Kirchhoff, and D. Garcia-Romero, "Directed speech separation for automatic speech recognition of long form conversational speech," in *Proc. Interspeech 2022*, 2022, pp. 5388–5392.

[8] Y. Zhang, Z. Li, B. Liu, H. Fan, Y. Yang, and Q. Yang, "A region based non-overlapping reference speech estimation method for speaker extraction," in *International Conference on Multimedia Modeling*. Springer, 2024, pp. 437–447.

[9] C. Boeddeker, A. S. Subramanian, G. Wichern, R. Haeb-Umbach, and J. Le Roux, "Ts-sep: Joint diarization and separation conditioned on estimated speaker embeddings," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 32, pp. 1185–1197, 2024.

[10] C. Han, Y. Luo, C. Li, T. Zhou, K. Kinoshita, S. Watanabe, M. Delcroix, H. Erdogan, J. R. Hershey, N. Mesgarani, and Z. Chen, "Continuous Speech Separation Using Speaker Inventory for Long Recording," in *Proc. Interspeech 2021*, 2021, pp. 3036–3040.

[11] T. Cord-Landwehr, C. Boeddeker, C. Zorilă, R. Doddipatla, and R. Haeb-Umbach, "Frame-wise and overlap-robust speaker embeddings for meeting diarization," in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023, pp. 1–5.

[12] J.-W. Jung, H.-S. Heo, B.-J. Lee, J. Huh, A. Brown, Y. Kwon, S. Watanabe, and J. S. Chung, "In search of strong embedding extractors for speaker diarisation," in *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023, pp. 1–5.

[13] S. Chen, Y. Wu, C. Wang, Z. Chen, Z. Chen, S. Liu, J. Wu, Y. Qian, F. Wei, J. Li, and X. Yu, "Unispeech-sat: Universal speech representation learning with speaker aware pre-training," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 6152–6156.

[14] S. Chen, C. Wang, Z. Chen, Y. Wu, S. Liu, Z. Chen, J. Li, N. Kanda, T. Yoshioka, X. Xiao, J. Wu, L. Zhou, S. Ren, Y. Qian, Y. Qian, J. Wu, M. Zeng, X. Yu, and F. Wei, "Wavlm: Large-scale self-supervised pre-training for full stack speech processing," *IEEE Journal of Selected Topics in Signal Processing*, vol. 16, no. 6, pp. 1505–1518, 2022.

[15] K. Okabe, T. Koshinaka, and K. Shinoda, "Attentive statistics pooling for deep speaker embedding," in *Interspeech 2018*. ISCA, Sep. 2018. [Online]. Available: http://dx.doi.org/10.21437/Interspeech.2018-993

[16] L. Zhang, Z. Chen, and Y. Qian, "Enroll-aware attentive statistics pooling for target speaker verification," *Proc. Interspeech 2022*, pp. 311–315, 2022.

[17] 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)*, 2015, pp. 5206–5210.

[18] M. Ge, C. Xu, L. Wang, E. S. Chng, J. Dang, and H. Li, "SpEx+: A Complete Time Domain Speaker Extraction Network," in *Proc. Interspeech 2020*, 2020, pp. 1406–1410.

[19] M. Borsdorf, C. Xu, H. Li, and T. Schultz, "Universal Speaker Extraction in the Presence and Absence of Target Speakers for Speech of One and Two Talkers," in *Proc. Interspeech 2021*, 2021, pp. 1469–1473.

[20] Z. Pan, M. Ge, and H. Li, "Usev: Universal speaker extraction with visual cue," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 30, pp. 3032–3045, 2022.

[21] H. Bredin, R. Yin, J. M. Coria, G. Gelly, P. Korshunov, M. Lavechin, D. Fustes, H. Titeux, W. Bouaziz, and M.-P. Gill, "pyannote.audio: neural building blocks for speaker diarization," in *ICASSP 2020, IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2020.

[22] H. Bredin and A. Laurent, "End-to-end speaker segmentation for overlap-aware resegmentation," in *Proc. Interspeech 2021*, 2021.

[23] J. L. Roux, S. Wisdom, H. Erdogan, and J. R. Hershey, "Sdr – half-baked or well done?" in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 626–630.

[24] P. Wang, Z. Chen, X. Xiao, Z. Meng, T. Yoshioka, T. Zhou, L. Lu, and J. Li, "Speech separation using speaker inventory," in *2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, 2019, pp. 230–236.