ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 978-1-7281-6327-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICASSP49357.2023.10095530

ADVANCING THE DIMENSIONALITY REDUCTION OF SPEAKER EMBEDDINGS FOR SPEAKER DIARISATION: DISENTANGLING NOISE AND INFORMING SPEECH ACTIVITY

You Jin Kim[1,*], *Hee-Soo Heo*[1,*], *Jee-weon Jung*[1,*], *Youngki Kwon*[1], *Bong-Jin Lee*[1], *Joon Son Chung*[2]

[1]Naver Cloud Corporation, South Korea
[2]Korea Advanced Institute of Science and Technology, South Korea

ABSTRACT

The objective of this work is to train noise-robust speaker embeddings adapted for speaker diarisation. Speaker embeddings play a crucial role in the performance of diarisation systems, but they often capture spurious information such as noise, adversely affecting performance. Our previous work has proposed an auto-encoder-based dimensionality reduction module to help remove the redundant information. However, they do not explicitly separate such information and have also been found to be sensitive to hyper-parameter values. To this end, we propose two contributions to overcome these issues: (i) a novel dimensionality reduction framework that can disentangle spurious information from the speaker embeddings; (ii) the use of speech activity vector to prevent the speaker code from representing the background noise. Through a range of experiments conducted on four datasets, our approach consistently demonstrates the state-of-the-art performance among models without system fusion.

Index Terms— Speaker diarisation, speaker embeddings, noise-robust

1. INTRODUCTION

Speaker diarisation is an interesting but challenging problem. The ability to determine "who spoke when" provides important context in speech transcription tasks, such as meeting transcription and video subtitling. One of the main challenges in speaker diarisation involves the task of clustering speech into an unknown number of speakers. The difficulty is augmented by the challenging environmental characteristics, such as background noise.

There are two main approaches to solve this challenging problem in previous literature: conventional module-based [1,2] and end-to-end [3,4]. The former "divides-and-conquers" speaker diarisation into several subtasks. The exact configuration differs from system to system, but in general they consist of speech activity detection (SAD), embedding extraction and clustering [5,6]. The latter directly segments audio recordings into homogeneous speaker regions using deep neural networks [7, 8]. However, current end-to-end approaches have been reported to be strongly overfitted to the environments that they are trained for, not generalised to diverse real-world conditions. Therefore, the winning entries to recent diarisation challenges [9, 10] either exploit the former approach or fuse both approaches.

The performance of the conventional speaker diarisation system which consists of multiple modules, is highly dependent on the ability to cluster the speaker embedding. Our recent work has proposed a number of methods to adapt the speaker embedding for speaker diarisation [11]. Among such proposals, the dimensionality reduction (DR) module utilised an auto-encoder (AE) trained in an unsupervised manner, and projected speaker embeddings to a low-dimensional code (e.g., 256 to 20), adapting towards each session. Speaker embeddings in diarisation tasks are only required to discriminate a small number of speakers, compared to thousands in case of verification. Therefore, finding a low-dimensional latent space effectively reduced unnecessary background noise and showed a potential in this line of research.

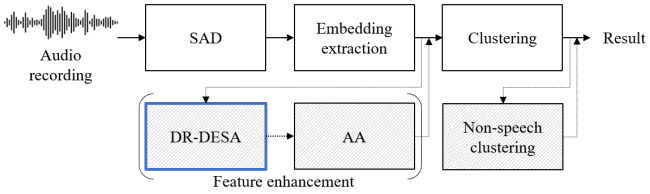


Fig. 1: Our speaker diarisation pipeline. Abbreviations are: speech activity detection (SAD), DR (dimensionality reduction), DR-DESA (DR with disentanglement and speech activity vector), attention and aggregation (AA). SAD can be either reference or system one, and DR-DESA is newly proposed in this paper, improving DR [11].

However, we empirically found that the effectiveness of our DR module varies from session to session. When the AE is trained independently for each session, we adopt a fixed code dimensionality, whereas we assume that the optimal code dimensionality may differ in each session depending on two factors: (i) the number of speakers and (ii) the duration. If the dimensionality is too small, the information required for speaker diarisation in the code becomes insufficient, resulting in performance degradation. In contrast, the excessive dimensionality may cause unnecessary information (e.g., background noise) to reside in the code [12]. Furthermore, the existing DR module trains the AE without distinction of speech or non-speech, potentially enforcing the projected embedding to also represent background noise as well as speaker identity [13]. The focus of this work will therefore be on mitigating the limitations of the existing DR module, and improving to be less hyper-parameter-dependent.

We propose two additional improvements upon the existing DR module to accomplish the goal. First, we extend the AE architecture by adding another code whereby the two codes each stand to represent speaker identity (*"speaker code"*) and other irrelevant information (*"noise code"*), respectively (Section 3.1). Employing two codes, the proposed method excludes noise-relevant factors from the speaker code. Second, we introduce *"speech activity vector (SAV)"* to the DR module which represents whether the input is extracted from a speech or a non-speech segment (Section 3.2). Training with SAV would ideally force the speaker code to be empty for speaker embeddings from non-speech segments, and therefore prevent the speaker code from representing the background noise.

We evaluate the effectiveness of the proposed methods on a range of datasets, on which we show the state-of-the-art performance (Section 4). In addition, we present additional analysis that our proposed approaches result in a less hyper-parameter-dependent module (Section 5.2).

*These authors contributed equally to this work.

2. SPEAKER DIARISATION PIPELINE

In this section, we introduce the overall pipeline of our speaker diarisation system, which consists of SAD, speaker embedding extraction, feature enhancement, and clustering modules. We omit explanation of SAD because the scope of this work only includes the scenario with a reference SAD. However, our framework can be also applied to system SAD as well. Figure 1 summarises the overall pipeline of our system.

2.1. Speaker embedding extraction

For every segment, we extract fixed-dimensional speaker embeddings to represent speaker characteristics from the segments. Our speaker embedding extraction module is identical to that of our previous work [11]. It extracts frame-level features using a residual applied trunk network followed by an average pooling layer. Each speaker embedding is extracted from a fixed duration of 1.5 seconds using a sliding window with 0.5 seconds shift. The embedding extractor is trained using VoxCeleb1 [14], VoxCeleb2 [15], and MLS [16] datasets. See Section 2.4 of [11] for full details.

2.2. Speaker embedding enhancement

Our pipeline employs two modules to adapt speaker embeddings that were originally trained for speaker verification towards diarisation: (i) dimensionality reduction with disentanglement and speech activity vector (DR-DESA, addressed in Section 3); and (ii) attention-based aggregation (AA) [11]. The DR-DESA module refers to the proposed module, which replaces the DR [11] module. The DR-DESA and the DR module share the following properties. They use a lightweight AE trained for each session. The AE comprises of two fully-connected layers, one for the encoder and the other for the decoder. For the encoder layer, we apply the maximum feature map [17] as a non-linearity, whereas the decoder does not adopt one. The differences and the improvements of DR-DESA compared to DR are further described in Section 3.

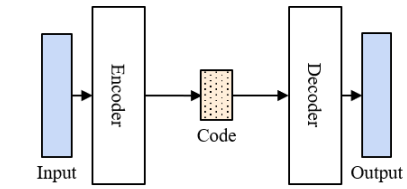
The AA module further refines dimensionality-reduced speaker embedding using a self-attention mechanism. The module encourages the features located close in the latent space to lie even more closer together, while further pushing distant features apart. The objective of this module is to remove noises and outliers on the affinity matrix, using the global context of each session.

2.3. Clustering

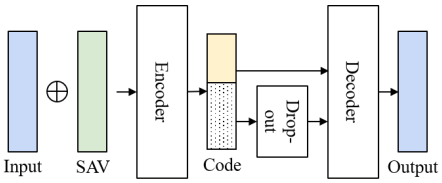
We assign a speaker label to each speaker embedding using a spectral clustering algorithm [18] that is widely adopted in the speaker diarisation literature. We apply eigen-decomposition to speaker embeddings after the DR-DESA and the AA modules without further refinement processes [18, 19] that are typically adopted in existing works. The number of clusters (*i.e.* speakers) is decided by counting the number of eigen-values higher than a predefined threshold; eigen-vectors corresponding to selected eigen-values are used as the spectral embeddings. Speaker labels are derived using a k-means clustering algorithm on the spectral embeddings.

3. DISENTANGLEMENT AND SPEECH ACTIVITY VECTOR

We propose a new model referred to as DR-DESA (Figure 2-(b)), extending the original DR (Figure 2-(a)) with two proposals : (i) we disentangle the existing code by adding noise code and applying dropout to it. (ii) we adopt an SAV denoting whether the speaker embedding includes speakers' voice.



(a) Architecture of DR. The noise may reside in the code of DR. The dots inside the code depicts the nose.



(b)

(b) Architecture of the proposed DR-DESA. It disentangles the code into speaker code and the noise code, applying dropout and SAV. The orange part illustrates speaker code, and the dotted part depicts noise code.

Fig. 2: Architecture comparison between DR [11] and the proposed DR-DESA.

3.1. Embedding disentanglement

In the original DR module, an input is projected into a low-dimensional code and then reconstructed. During this process, the noise factor is inevitably entangled in the code because noise is also required to reconstruct the original input [20]. The noise factor entangled in the code may disturb speaker clustering, as noise may be consistent across different speakers' identities. To mitigate this potential threat, we propose to disentangle the noise factor. We divide the latent space into two, and force them to represent speaker-relevant (*speaker code*) and irrelevant information (*noise code*) respectively. We apply dropout only to the noise code, making the neglectful information flow to it. This is a frequently used technique for disentanglement [21], where it has been reported that it makes essential information for reconstruction to be gathered in the code where dropout does not exist. We concatenate the two codes and feed it to the decoder. After the training is complete, only the speaker code is used for subsequent clustering step, discarding the noise code.

3.2. Speech activity vector

Using two kinds of codes opens a new potential by discarding speaker-irrelevant information from the speaker code. However, the behaviour of the AE becomes more complicated. The speaker code should primarily represent input embeddings extracted from speech segments. On the other hand, the noise code should mainly represent input embeddings from non-speech segments. To enable this ideal scenario, the AE is required to distinguish whether an input is from speech.

We further propose to adopt an SAV, which takes the form of a learnable vector which has a dimensionality identical to the input embedding to compose the proposed DR-DESA. An SAV is added element-wisely to the input embedding, similar to the positional encoding [22]. Concretely, we adopt two SAVs, one for the speech embedding and the other for the non-speech embedding. Depending on the speaker embedding's type, we add either SAV to the speaker embedding. Note that since the SAD is already included in the speaker diarisation pipeline (either system or reference) and precedes the speaker embedding extraction step, we can

utilise SAD results at no additional cost.

4. EXPERIMENTS

We evaluate the effectiveness of the proposed methods on DIHARD and VoxConverse datasets. The datasets and the experimental details are described in the following paragraphs.

4.1. Datasets

DIHARD datasets. The DIHARD challenges publish evaluation datasets which include sessions recorded in restaurant, clinical interview, YouTube videos, etc., making the scenario more challenging. We use the evaluation sets of the first, second, and third DIHARD challenges [9, 23, 24].

VoxConverse. It is an audio-visual speaker diarisation dataset, which consists of speech clips extracted from YouTube videos. The corpus contains overlapped speech, a large speaker pool, and diverse background conditions, including talk-shows, panel discussions, political debates and celebrity interviews [25]. Test set version `0.0.2` is used for experiments.

4.2. Evaluation protocol

Diarisation error rate (DER), the summation of false alarm (FA), missed speech (MS), and speaker confusion (SC), is used as the primary metric. FA and MS are related to the SAD module, whereas SC to the DR or the proposed DR-DESA modules. For all experiments conducted on four datasets, we use the reference SAD to precisely compare the impact of SC caused by either the DR or the proposed DR-DESA.

We use the d-score toolkit[1] for measuring the DER. We do not use forgiveness collar for experiments involving the DIHARD datasets, whereas we set a 0.25 seconds forgiveness collar for VoxConverse experiments to match the scenario with corresponding challenges.

4.3. Results

Table 1 presents the performances of the proposed methods on the four datasets compared with the baselines. We also conduct ablation studies where we exclude each proposed component to verify the effect of each component on the overall performance. Note that, since we utilise reference SAD results, FA is zero in all cases and MS corresponds to the proportion of the overlapped speech included in each dataset.

Comparison with the baselines. In all datasets, DR-DESA outperforms the baselines without DR module by a large margin. In the case of the DIHARD datasets, the SC error is more than halved, and in VoxConverse SC reduced by more than 30%. In all four datasets, DR-DESA performs even better than the DR consistently.

Comparison with state-of-the-art systems. Experimental results on DIHARD I and II show that the proposed DR-DESA outperforms the winning systems of the challenges. DR-DESA also outperforms the best single system in DIHARD III challenge. In case of VoxConverse, the test set used in VoxSRC challenge [10] has been recently updated. Also, the majority of recent researches apply a system SAD in place of a reference SAD; the VoxSRC challenge which uses VoxConverse only has scenarios that use a system SAD. Therefore, we did not compare DR-DESA's performance with the systems submitted to the challenge.

[1] https://github.com/nryant/dscore

Table 1: Results on DIHARD I, II, III, and VoxConverse datasets *(DER: diarisation error rate, FA: false alarm, MS: miss, SC: speaker confusion)*. DR stands for dimensionality reduction, and DE for disentanglement, and SAV for speech activity vector. DR-DESA for DR with disentanglement and SAV is proposed method with two improvements (DE and SAV).

Configuration	DER	FA	MS	SC
DIHARD I				
Track 1 winner [26]	23.73	-	-	-
Baseline	25.85	0.00	8.71	17.14
Baseline + DR	17.70	0.00	8.71	8.98
DR + DE	17.04	0.00	8.71	8.33
DR + SAV	17.25	0.00	8.71	8.54
DR + DE + SAV (DR-DESA)	**16.75**	0.00	8.71	**8.04**
DIHARD II				
Track 1 winner [27]	18.42	-	-	-
Baseline	27.39	0.00	9.69	17.70
Baseline + DR	18.40	0.00	9.69	8.71
DR + DE	17.76	0.00	9.69	8.08
DR + SAV	18.21	0.00	9.69	8.52
DR + DE + SAV (DR-DESA)	**17.44**	0.00	9.69	**7.75**
DIHARD III				
Track 1 best single system [28]	15.50	-	-	-
Baseline	20.99	0.00	9.52	11.47
Baseline + DR	15.49	0.00	9.52	5.97
DR + DE	15.28	0.00	9.52	5.76
DR + SAV	15.32	0.00	9.52	5.80
DR + DE + SAV (DR-DESA)	**15.05**	0.00	9.52	**5.53**
VoxConverse				
Baseline	5.83	0.00	1.60	4.23
Baseline + DR	4.58	0.00	1.60	2.98
DR + DE	4.51	0.00	1.60	2.91
DR + SAV	4.55	0.00	1.60	2.95
DR + DE + SAV (DR-DESA)	**4.45**	0.00	1.60	**2.85**

Ablation studies. DR-DESA has two components on top of the baseline with DR, that are disentanglement and SAV. We perform ablation studies by excluding each component from the DR-DESA, and show how each proposal affects the performance. In all four datasets, removing disentanglement have a greater impact on the performance. However, adopting SAV also consistently improves the performance compared to the baseline with DR. It is DR-DESA that shows the best performance, and the tendency of the performance gain by each component is consistent across all datasets, signifying that the two proposed techniques are complementary.

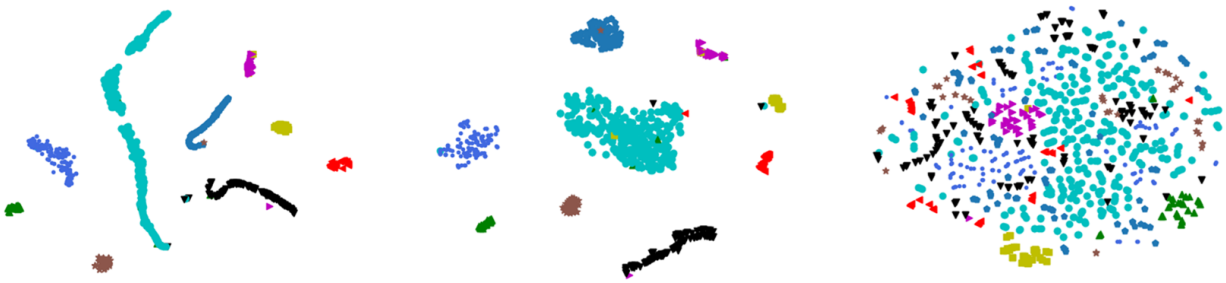
5. FURTHER ANALYSIS

In this section, we present further analyses to show the role of each code and the strength of DR-DESA.

5.1. Visualisation

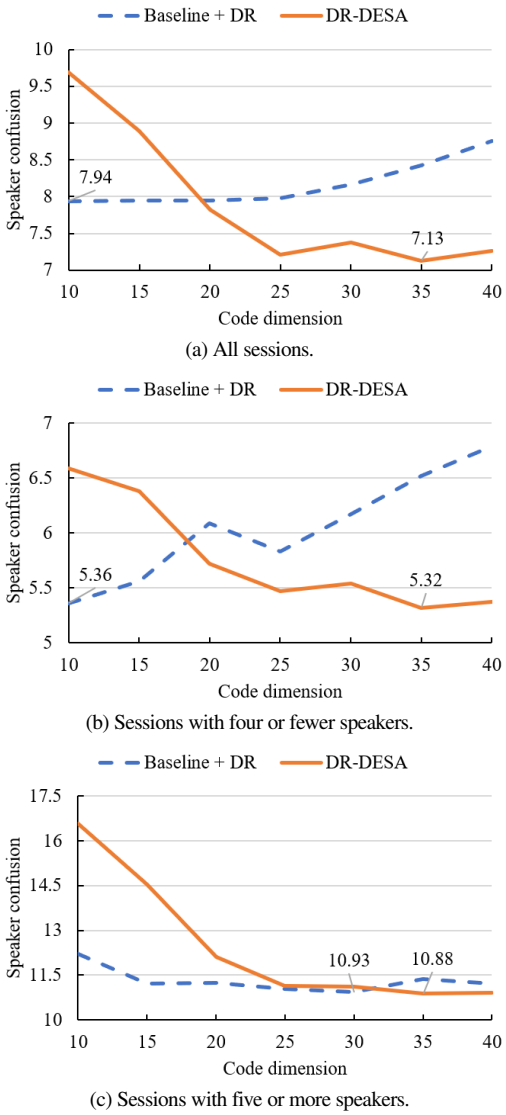
Figure 3 depicts the code representation of the DR and the DR-DESA module. Figure 3 (a) shows the code from the DR. Figure 3 (b) represents the speaker code and (c) shows the noise code of the proposed DR-DESA. We randomly select an audio recording with nine speakers from the DIHARD II dataset, extract codes from the audio, and visualise them using t-SNE technique [29].

As shown in the figure, the proposed speaker code (b) represents nine clusters corresponding to nine speakers. On the other hand, the



(a) Code generated by DR. (b) Speaker code generated by DR-DESA. (c) Noise code generated by DR-DESA.

Fig. 3: Visualisation of the code. Input audio involves nine speakers, resulting in nine clusters ideally. The number of clusters in (a) exceeds nine (the cyan-rounds are divided into several clusters), whereas (b) includes precisely nine clusters. In addition, noise code (c) does not form meaningful clusters.



(a) All sessions.

(b) Sessions with four or fewer speakers.

(c) Sessions with five or more speakers.

Fig. 4: Stability of DR-DESA with high code dimension. (The sessions in DIHARD I, II, and III datasets are used to draw each graph.)

original code (a) shows more than nine clusters, with the codes of the most dominant speakers divided into multiple clusters. We interpret that this unexpected result is due to the change of noise information within the same speaker, and in the case of the proposed method, this additional information is represented by noise code in (c). This role of the noise code makes the speaker code in (b) have more suitable distribution for speaker diarisation.

5.2. Analysis based on the number of speakers

We present Figure 4 to show the limitation of DR module and the effectiveness of our DR-DESA using the three DIHARD datasets. We evaluate the performance of our baseline (DR module of [11]) and the proposed DR-DESA using diverse code dimensionalities. (a) shows SC of the entire sessions, (b) indicates SC of the sessions where the number of speakers is four or fewer, and (c) shows SC of the session with more than four speakers. As argued, the baseline requires low dimensionality for sessions with fewer speakers, and high dimensionality for sessions with more speakers. Performance degradation is observed, especially in (b), when the dimensionality is not ideal. In contrary, our proposed DR-DESA module demonstrates the stable and optimal performance regardless of the number of speakers, when dimensionality is 30 or more. As a result, this stability leads relatively higher performance improvements in the entire dataset, even though the optimal performances of the two systems in each subset do not show a significant difference.

6. CONCLUSION

This paper addresses a novel unsupervised disentanglement framework, which generates noise-robust speaker embeddings for speaker diarisation. Speaker embeddings are the crucial component of diarisation systems, but they often contain the unnecessary information that degrades the performance, such as background noise and reverberation. Recently proposed DR module reduces the dimensionality of the embeddings, in order to remove the spurious information. However, the effect of DR is limited, being sensitive to the code dimensionality.

To this end, we propose DR-DESA introducing two more techniques on top of the DR module: (i) explicit disentanglement of the spurious information from the original code; (ii) the introduction of SAV. DR-DESA show the state-of-the-art performance as a single system on four benchmark datasets, and ablation studies on DR-DESA demonstrate that both of the proposals lead to performance gains. In addition, visualising the disentangled code confirms that DR-DESA performs as intended.

7. REFERENCES

[1] Daniel Garcia-Romero, David Snyder, Gregory Sell, Daniel Povey, and Alan McCree, "Speaker diarization using deep neural network embeddings," in *Proc. ICASSP*, 2017.

[2] Zili Huang, Shinji Watanabe, Yusuke Fujita, Paola García, Yiwen Shao, Daniel Povey, and Sanjeev Khudanpur, "Speaker diarization with region proposal network," in *Proc. ICASSP*, 2020.

[3] Yawen Xue, Shota Horiguchi, Yusuke Fujita, Yuki Takashima, Shinji Watanabe, Paola Garcia, and Kenji Nagamatsu, "Online end-to-end neural diarization handling overlapping speech and flexible numbers of speakers," *Proc. Interspeech*, 2021.

[4] Yusuke Fujita, Shinji Watanabe, Shota Horiguchi, Yawen Xue, and Kenji Nagamatsu, "End-to-end neural diarization: Reformulating speaker diarization as simple multi-label classification," *arXiv preprint arXiv:2003.02966*, 2020.

[5] Federico Landini, Shuai Wang, Mireia Diez, Lukáš Burget, Pavel Matějka, Kateřina Žmolíková, Ladislav Mošner, Anna Silnova, Oldřich Plchot, Ondřej Novotnỳ, et al., "But system for the second dihard speech diarization challenge," in *Proc. ICASSP*, 2020.

[6] Youngki Kwon, Hee Soo Heo, Jaesung Huh, Bong-Jin Lee, and Joon Son Chung, "Look who's not talking," in *Proc. SLT*, 2021.

[7] Shota Horiguchi, Yusuke Fujita, Shinji Watanabe, Yawen Xue, and Kenji Nagamatsu, "End-to-end speaker diarization for an unknown number of speakers with encoder-decoder based attractors," in *Proc. Interspeech*, 2020.

[8] Keisuke Kinoshita, Marc Delcroix, and Naohiro Tawara, "Advances in integration of end-to-end neural and clustering-based diarization for real conversational speech," in *Proc. ICASSP*, 2021.

[9] Neville Ryant, Prachi Singh, Venkat Krishnamohan, Rajat Varma, Kenneth Church, Christopher Cieri, Jun Du, Sriram Ganapathy, and Mark Liberman, "The third dihard diarization challenge," *arXiv preprint arXiv:2012.01477*, 2020.

[10] Arsha Nagrani, Joon Son Chung, Jaesung Huh, Andrew Brown, Ernesto Coto, Weidi Xie, Mitchell McLaren, Douglas A Reynolds, and Andrew Zisserman, "Voxsrc 2020: The second voxceleb speaker recognition challenge," *arXiv preprint arXiv:2012.06867*, 2020.

[11] Youngki Kwon, Jeeweon Jung, HeeSoo Heo, You Jin Kim, BongJin Lee, and Joon Son Chung, "Adapting speaker embeddings for speaker diarisation," in *Proc. Interspeech*, 2021.

[12] Wen-Chin Huang, Hao Luo, Hsin-Te Hwang, Chen-Chou Lo, Yu-Huai Peng, Yu Tsao, and Hsin-Min Wang, "Unsupervised representation disentanglement using cross domain features and adversarial learning in variational autoencoder based voice conversion," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 4, no. 4, 2020.

[13] Adam Polyak and Lior Wolf, "Attention-based wavenet autoencoder for universal voice conversion," in *Proc. ICASSP*, 2019.

[14] Arsha Nagrani, Joon Son Chung, and Andrew Zisserman, "Voxceleb: a large-scale speaker identification dataset," in *Proc. Interspeech*, 2017.

[15] Joon Son Chung, Arsha Nagrani, and Andrew Zisserman, "Voxceleb2: Deep speaker recognition," in *Proc. Interspeech*, 2018.

[16] Vineel Pratap, Qiantong Xu, Anuroop Sriram, Gabriel Synnaeve, and Ronan Collobert, "Mls: A large-scale multilingual dataset for speech research," in *Proc. Interspeech*, 2020.

[17] Xiang Wu, Ran He, Zhenan Sun, and Tieniu Tan, "A light cnn for deep face representation with noisy labels," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 11, 2018.

[18] Huazhong Ning, Ming Liu, Hao Tang, and Thomas S Huang, "A spectral clustering approach to speaker diarization," in *Proc. ICSLP*, 2006.

[19] Quan Wang, Carlton Downey, Li Wan, Philip Andrew Mansfield, and Ignacio Lopz Moreno, "Speaker diarization with lstm," in *Proc. ICASSP*, 2018.

[20] Joon Son Chung, Jaesung Huh, and Seongkyu Mun, "Delving into voxceleb: environment invariant speaker recognition," in *Proc. Odyssey*, 2019.

[21] Ayush Jaiswal, Rex Yue Wu, Wael Abd-Almageed, and Prem Natarajan, "Unsupervised adversarial invariance," *Advances in neural information processing systems*, 2018.

[22] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, 2017.

[23] Neville Ryant, Kenneth Church, Christopher Cieri, Alejandrina Cristia, Jun Du, Sriram Ganapathy, and Mark Liberman, "First dihard challenge evaluation plan," *2018, tech. Rep.*, 2018.

[24] Neville Ryant, Kenneth Church, Christopher Cieri, Alejandrina Cristia, Jun Du, Sriram Ganapathy, and Mark Liberman, "The second dihard diarization challenge: Dataset, task, and baselines," *arXiv preprint arXiv:1906.07839*, 2019.

[25] Joon Son Chung, Jaesung Huh, Arsha Nagrani, Triantafyllos Afouras, and Andrew Zisserman, "Spot the conversation: speaker diarisation in the wild," in *Proc. Interspeech*, 2020.

[26] Gregory Sell, David Snyder, Alan McCree, Daniel Garcia-Romero, Jesús Villalba, Matthew Maciejewski, Vimal Manohar, Najim Dehak, Daniel Povey, Shinji Watanabe, et al., "Diarization is hard: Some experiences and lessons learned for the jhu team in the inaugural dihard challenge.," in *Proc. ICASSP*, 2018.

[27] Federico Landini, Shuai Wang, Mireia Diez, Lukáš Burget, Pavel Matějka, Kateřina Žmolíková, Ladislav Mošner, Oldřich Plchot, Ondřej Novotnỳ, Hossein Zeinali, et al., "But system description for dihard speech diarization challenge 2019," *arXiv preprint arXiv:1910.08847*, 2019.

[28] Jahangir Alam and Vishwa Gupta, "Crim's system description for the third edition of dihard challenge 2020," *The Third DIHARD Speech Diarization Challenge*, 2020.

[29] Laurens Van der Maaten and Geoffrey Hinton, "Visualizing data using t-sne.," *Journal of machine learning research*, vol. 9, no. 11, 2008.