INTERSPEECH 2019
September 15–19, 2019, Graz, Austria

Speaker-Invariant Feature-Mapping for Distant Speech Recognition via Adversarial Teacher-Student Learning

Long Wu[1,2]*,Hangting Chen*[1,2]*,Li Wang* [1]*, Pengyuan Zhang*[1,2]*, Yonghong Yan*[1,2,3]

[1]Key Laboratory of Speech Acoustics and Content Understanding, Institute of Acoustics, China
[2]University of Chinese Academy of Sciences, China
[3]Xinjiang Laboratory of Minority Speech and Language Information Processing,
Xinjiang Technical Institute of Physics and Chemistry, Chinese Academy of Sciences, China
{wulong, chenhangting, wangli, zhangpengyuan, yanyonghong}@hccl.ioa.ac.cn

Abstract

Feature mapping (FM) jointly trained with acoustic model (AFM) is commonly used for single-channel speech enhancement. However, the performance is affected by the inter-speaker variability. In this paper, we propose speaker-invariant AFM (SIAFM) aiming at curtailing the inter-talker variability while achieving speech enhancement. In SIAFM, a feature-mapping network, an acoustic model and a speaker classifier network are jointly optimized to minimize the feature-mapping loss and the senone classification loss, and simultaneously min-maximize the speaker classification loss. Evaluated on AMI dataset, the proposed SIAFM achieves 4.8% and 7.0% relative word error rate (WER) reduction on the overlapped and non-overlapped condition over the baseline acoustic model trained with single distant microphone (SDM) data. Additionally, the SIAFM obtains 3.0% relative overlapped WER and 4.2% relative non-overlapped WER decrease over the multi-conditional (MCT) acoustic model. To further promote the performance of SIAFM, we employ teacher-student learning (TS), in which the posterior probabilities generated by the individual headset microphone (IHM) data can be used in lieu of labels to train the SIAFM model. The experiments show that compared with MCT model, SIAFM with TS (SIAFM-TS) can reach 4.2% relative overlapped WER and 6.3% relative non-overlapped WER decrease respectively.

Index Terms: far-field speech recognition, speech enhancement, adversarial learning, teacher-student learning

1. Introduction

Despite significant advancement made in automatic speech recognition (ASR) after the introduction of deep neural network (DNN) based acoustic models [1], the far-field speech recognition remains a challenging problem. In distant talking scenarios, speech signal is captured by one or more microphones located farther away from the speakers, which makes it susceptible to reverberation, background noise and speech overlap. Therefore the performance in distant talking scenarios is still far-behind their close-talking equivalents [2].

Many distant speech recognition systems adopt a two-part architecture where the speech are enhanced by the signal processing techniques before further processed by conventional acoustic modeling approaches [3, 4, 5]. Since the signal processing part is usually distinct from the speech recognition, it fails to optimize towards the speech recognition accuracy, which leads to a suboptimal solution [6].

With the advance of deep learning, DNN-based approaches have achieved great success in distant speech recognition. The most common approach is multi-conditional training, in which the distant-talking and close-talking features are mixed to train an acoustic model. Although the model can learn the reverberation effects automatically, it can not model the corresponding relations between the distant-talking and close-talking features. Therefore, to obtain an optimal performance, joint training of speech enhancement and acoustic model are proposed. These methods can be roughly divided into two categories: mask learning [7, 8] and feature-mapping approaches [9, 10, 11]. However, the mask learning has the presumption that the the noise is strictly additive and removable by the masking procedure which is generally not true for real recorded stereo data.

Though jointly training speech enhancement with acoustic model can improve the recognition performance, it is still affected by the spectral variations in each speech unit. Recently, adversarial learning has captured great attention of deep learning community given its remarkable success in estimating generative models. It has been applied to noise robust speech recognition [12, 13], speaker-invariant [14], domain adaptation [15, 16] and domain separation [17]. Inspired by this, we propose SIAFM, which combines the speaker adaption with speech enhancement. Through this adversarial multi-task learning procedure, the feature-mapping network not only maps the the input speech frames from distant scene to the close one, but also transforms the input speech frames from different speakers into speaker-invariant and senone-discriminative features.

Teacher-student (TS) learning is widely used in domain adaption [18] and model compression [19]. In TS learning, the posteriors generated by the teacher model are used to train the target-domain student model. To transfer the knowledge of close-talking to far-field, in this work, we explore using the soft label from the well trained IHM acoustic model instead of hard label during the SIAM training. By combining SIAFM and TS learning (SIAFM-TS), we achieve speaker adaption, knowledge distillation and speech enhancement in a unified framework.

Different from the previous work, we evaluate SIAFM-TS on the real recorded AMI dataset. Besides, we show the WER on both overlapped and non-overlapped condition, namely, WER(over) and WER(non-over). What is more, the performances are demonstrated step by step and obtained promotion gradually. Finally, we combine the speaker adaption, knowledge distillation and speech enhancement in a unified framework. Compared with MCT model, the proposed SIAFM-TS achieves 4.2% relative WER(over) reduction and 6.3% relative WER(non-over) reduction. Meanwhile, in comparison with SDM model baseline, SIAFM-TS reaches 5.9% relative WER(over) decrease and 9.0% relative WER(non-over) decrease.

Copyright © 2019 ISCA

http://dx.doi.org/10.21437/Interspeech.2019-2136

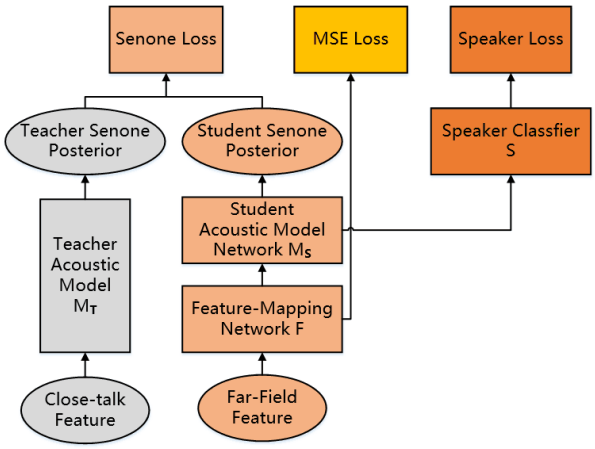

Figure 1: *The framework of SIAFM-TS for distant speech recognition*

2. Senone-Aware Feature-Mapping Enhancement

Given a sequence of distant speech features $X = \{x_1, ..., x_T\}$ and its corresponding features $Y = \{y_1, ..., y_T\}$ in close scene, feature-mapping tries to learn a non-linear regression function F with parameters θ_f that transforms X to a sequence of enhanced features $\hat{Y} = \{\hat{y}_1, ..., \hat{y}_T\}$ such that $\hat{Y}$ is as close to Y as possible. To achieve that, we minimize the far-to-close feature-mapping MSE loss $L_F(\theta_f)$ as follows:

$$L_F(\theta_f) = \frac{1}{T}\sum_{i=1}^{T}(F(x_i) - y_i)^2 \tag{1}$$

To make the feature-mapping objective relate to the speech units (i.e., word, phoneme, senone, etc.) classification, we incorporate a DNN acoustic model into the feature-mapping framework and propose the multi-task learning framework ,i.e, Senone-Aware Feature-Mapping. In this framework, we jointly optimize the feature-mapping MSE criterion and the acoustic cross-entropy (CE) criterion.

As showed in Figure 1, the acoustic model network M_s with parameters θ_m takes in the enhanced features $\hat{Y}$ as the input and predicts the senone posteriors. Through minimizing the cross-entropy loss between the predicted senone posteriors and the senone labels in (2), the enhanced features $\hat{Y}$ are senone-discriminative.

$$\begin{aligned} L_{M_s}(\theta_m, \theta_f) &= -\frac{1}{T}\sum_{i=1}^{T} logP(s_i|x_i;\theta_m,\theta_f) \\ &= -\frac{1}{T}\sum_{i=1}^{T} log(M(F(x_i))) \end{aligned} \tag{2}$$

where S is a sequence of senone labels $S = \{s_1, ..., s_T\}$ aligned with the close-talk features.

The total loss of AFM $L_{AFM}(\theta_f, \theta_m)$ is formulated as the weighted sum of $L_F(\theta_f)$ and the senone classification loss $L_{M_s}(\theta_m, \theta_f)$ as follows:

$$L_{AFM}(\theta_m, \theta_f) = \Lambda_1 L_F(\theta_f) + (1 - \Lambda_1)L_{M_s}(\theta_f, \theta_m) \tag{3}$$

where $\Lambda_1 > 0$ controls the trade-off between feature-mapping MSE loss and acoustic CE loss.

3. Adversarial Speaker-Invariant Training

To perform the SIAFM, besides the distant speech features X and its corresponding close-talk features Y, we also need a sequence of speaker labels $C = \{c_1, ..., c_T\}$. Based on the AFM, we further employ a speaker classifier network S which maps the enhanced features f to the speaker posteriors. Therefore, the F and S are jointly trained with an adversarial objective, in which θ_f is adjusted to $maximize$ the speaker classification loss $L_S^f(\theta_f)$ while θ_s is adjusted to $minimize$ the frame-level speaker classification loss $L_S^s(\theta_s)$ below:

$$\begin{aligned} L_S(\theta_f, \theta_s) &= -\sum_{i=1}^{T} logP_s(c_i|x_i;\theta f, \theta s) \\ &= -\sum_{i=1}^{T}\sum_{a\in A} I(a = c_i) logS(M(x_s)) \end{aligned} \tag{4}$$

where c_i denotes the speaker label for the input distant speech frame x_i. $I(a = c_i)$ is a function which equals to 1 when a is equals to c_i.

The total loss of SIAFM $L_{SIAFM}(\theta_f, \theta_m, \theta_s)$ is formulated as the weighted sum of $L_{AFM}(\theta_m, \theta_f)$ and $L_S(\theta_f, \theta_s)$ as follows:

$$L_{SIAFM}(\theta_f, \theta_m, \theta_s) = L_{AFM}(\theta_m, \theta_f) - \lambda_2 L_S(\theta_f, \theta_s) \tag{5}$$

where λ_2 controls the trade-off between the AFM loss and the speaker classification loss separately.

We find the optimal parameters $\hat{\theta_f}$, $\hat{\theta_m}$ and $\hat{\theta_s}$ such that

$$(\hat{\theta_f}, \hat{\theta_m}) = \underset{\theta_f, \theta m}{argmin} L_{SIAFM}(\theta_f, \theta_m, \hat{\theta}_s) \tag{6}$$

$$\hat{\theta_s} = \underset{\theta_s}{argmax} L_{SIAFM}(\hat{\theta}_f, \hat{\theta}_m, \theta_s) \tag{7}$$

The parameters are updated as follows via back propagation through time with stochastic gradient descent (SGD):

$$\theta_f \leftarrow \theta_f - \mu[\frac{\partial L_{AFM}}{\partial \theta_f} - \lambda_2 \frac{\partial L_S}{\partial \theta_f}] \tag{8}$$

$$\theta_s \leftarrow \theta_s - \mu \frac{\partial L_S}{\partial \theta_s} \tag{9}$$

$$\theta_m \leftarrow \theta_m - \mu \frac{\partial L_{AFM}}{\partial \theta_m} \tag{10}$$

where μ is learning rate and $\lambda_2 > 0$.

4. Teacher-Student Learning

Teacher-student learning is widely used in domain adaption and model compression. The main idea of TS learning is to train a target-domain student model using the output of a well-trained source-domain teacher model as training labels (often called soft labels) such that the student model works like the teacher model. In this work, we regard the near-field as source-domain and the far-field as target-domain. To achieve TS learning, first, train a teacher model using the IHM data. Then minimize the Kullback-Leibler (KL) divergence between the output distributions of the teacher model and the student model as follows:

$$KL(P_T||P_S) = \sum_{i=1}^{T}\sum_{q\in Q} P_T(q|x_i^{ihm}) log(\frac{P_T(q|x_i^{ihm})}{P_S(q|x_i^{sdm})}) \tag{11}$$

where q is one of the senones in the senone set Q, i is the frame index. x_i^{sdm} is the SDM feature and x_i^{ihm} is the corresponding IHM feature. $P_T(q|x_i^{ihm})$ and $P_S(q|x_i^{sdm})$ is the output distribution of the teacher and student model individually.

Optimizing the above objective function is equivalent to minimizing:

$$L_{M_s}^{TS}(\theta_f, \theta_m) = -\sum_{i=1}^{T}\sum_{q\in Q} P_T(q|x_i^{ihm}) log(P_S(q|x_i^{sdm})) \tag{12}$$

To combine the SIAFM with the TS learning in a unified framework (SIAFM-TS), we replace the $L_{M_s}(\theta_f, \theta_m)$ in (3) with the $L_{M_s}^{TS}$ in eq. (12). In SIAFM-TS, we can achieve the speech enhancement, speaker adaption and domain adaption at the same time.

5. Experiments and Results

5.1. Dataset

We evaluate the proposed SIAFM-TS on the AMI meeting corpus [20]. Our systems are trained and tested as recommended: a training set of 80 hours and a test set of 9 hours. There are 155 speakers in the training data and 63 speakers in the test data. The speakers in training set and the test set are mutually different.

In this work, we use 40-dimensional log Mel filterbank features as the network inputs by following the process in kaldi [21]. Besides, we exploit Tensorflow to build the SIAFM system. We use the trained GMM-HMM to generate forced aligned labels for training a DNN acoustic model. During training, the aligned labels from IHM data is used to improve the SDM baseline. The networks are trained using the stochastic gradient descent (SGD) algorithm.

5.2. Baseline System

In the baseline system, we first train two DNN-HMM acoustic models using the 80 hrs SDM data and 80 hrs IHM data respectively. Each frame is spliced with 6 left and 6 right context frames to form a 520-dimensional feature. Both DNN acoustic models have 5 hidden layers with 2048 Relu units in each layer. The output layer of the DNN has 3992 output units corresponding to 3992 senone labels. Moreover, we use all the SDM and IHM data to train MCT acoustic model.

The SDM eval data (SDM-Eval) WER results are shown in Table 1. Firstly, severe performance degradation is observed with speech overlap by analysing WER(over) and WER(nonover) results. Secondly, compared with IHM model, SDM model improves the performance more than 20.0%, indicating that the consistency between training data and test data is significant. Furthermore, in comparison with SDM model, MCT model reaches 2.0% relative WER(over) reduction and 3.0% relative WER(non-over) reduction on SDM-Eval.

5.3. Senone-Aware Feature-Mapping Enhancement

In this section, a simple DNN network with 4 hidden layers and 1024 units for each hidden layer is trained to conduct the feature-mapping. The results are list in Table 1. Initially, with IHM model, we observe great improvement on feature-mapping results (SDM-FM) over SDM-Eval. Although the WER gap between SDM-Eval and SDM-FM is large, the WER(over) gap 5.6% and WER(nonover) gap 5.7% with MCT model is far less than the WER(over) gap 8.1% and WER(nonover) gap 8.9%

Table 1: *Performances of IHM trained model , SDM trained model and MCT trained model on SDM eval data and feature-mapping enhanced results. Additionally, SDM-FM represents the feature-mapping enhanced results and SDM-Eval represents the SDM eval data.*

Acoustic model	WER(over)		WER(nonover)	
	SDM-Eval	SDM-FM	SDM-Eval	SDM-FM
IHM	76.5	**63.3**	72.9	**55.2**
SDM	**54.0**	62.1	**44.5**	53.4
MCT	**53.0**	58.6	**43.2**	48.9

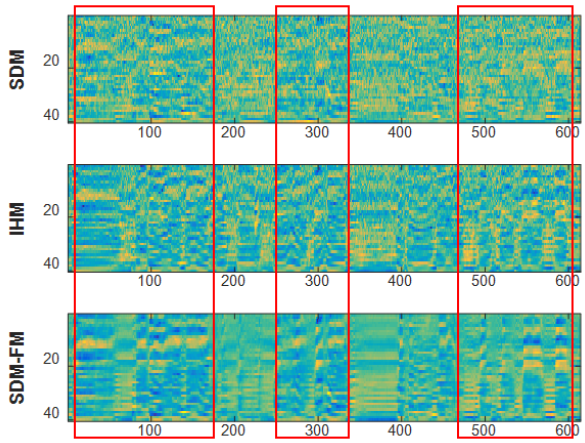

Figure 2: *The Log mel filterbank features of the IHM data, SDM data and the corresponding feature-mapping enhanced results.*

with SDM model. These results indicate that feature-mapping can effectively remove reverberation and improve the speech quality. This is also illustrated in Fig 2. After feature mapping, the distortion is suppressed and speech information is restored.

Furthermore, to cope with the mismatch between training and test, we pass the SDM training data through the speech enhancement DNN. Lastly, train an acoustic model MCT-FM with all SDM, IHM and the enhanced training data. As shown in Table 2, compared with MCT model, MCT-FM achieves 8.5% relative WER(over) reduction and 10.0% relative WER(nonover) reduction on SDM-FM. But it is still far behind the MCT performance on SDM-Eval. We suppose it is because the feature-mapping fails to optimize towards the speech recognition accuracy.

Consequently, we employ joint training feature-mapping with acoustic model. In Table 2, the AFM reaches 1.7% relative WER(over) decrease and 1.6% relative WER(nonover) decrease over MCT model.

5.4. Adversarial Speaker-Invariant Training

To make the enhanced feature both senone-discriminative and speaker-invariant, we further conduct SIAFM as described in Section 3. The speaker classifier S is a feed-forward DNN with 2 hidden layers and 1024 hidden units for each layer. The output layer of S has 155 units predicting the posteriors of 155 speakers in the training set. The hyper-parameter λ_1 and λ_2 are fixed at 0.5 in our experiments. The learning rates of F and S

Table 2: *Performance of MCT model, MCT-FM model and AFM model on the SDM eval data and feature-mapping enhanced features. Additionally, SDM-FM represents the feature-mapping enhanced results and SDM-Eval represents the SDM eval data.*

Acoustic model	WER(over)		WER(nonover)	
	SDM-Eval	SDM-FM	SDM-Eval	SDM-FM
MCT	53.0	58.6	43.2	48.9
MCT-FM	-	**53.6**	-	**44.0**
AFM	**52.1**	-	**42.5**	-

are set to 0.03. In each iteration, we update the parameters of F 5 times and the S 1 times.

As shown in Table 3, SIAFM achieves 4.8% relative WER(over) reduction and relative 7.0% WER(nonover) reduction over the SDM model baseline. Additionally, compared with MCT model, SIAFM reaches 3.0% relative WER(over) decrease and 4.2% relative WER(nonover) decrease.

5.5. Visualization of Deep Features

We randomly select two male speakers (ME0067, ME0069) and two female speakers (FE0065, FE0066) from the training set and extract speech frames aligned with 'ah' for each of the four speakers. Then, we visualize the enhanced features generated by the AFM and SIAFM architectures when the 'ah' frames are given as the input using t-SNE. Besides, we also show the t-SNE of the SDM features and its corresponding IHM features.

In Fig 3, the distributions of male and female speakers are far away from each other on both SDM and IHM data. But the distance of male and female on SDM data is smaller than that on IHM data. We suppose it is because the discrimination of female and male is partly damaged by the reverberation and speaker overlap. After performing AFM, the distribution of enhanced feature is more similar with the distribution of IHM, which demonstrates the effectiveness of AFM. When applying SIAFM, all the male and female speakers are well aligned with each other in the last graph. This indicates that adversarial speaker-invariant training makes the enhanced features more speaker-invariant.

5.6. Teacher-student Learning with SIAFM

In this section, we combine TS learning with SIAFM to achieve speaker adaption, knowledge distillation and speech enhancement in a unified framework. The results are shown in Table 3. Compared with SIAFM, SIAFM-TS further achieves 1.2% relative WER(over) reduction and 2.2% relative WER(nonover) reduction. Finally, in comparison with the SDM model baseline, SIAFM-TS reaches 5.9% relative WER(over) decrease and 9.0% relative WER(non-over) decrease.

6. Conclusions

In this study, we first explore training feature-mapping network with acoustic model to accomplish the speech enhancement. Then, adversarial training is conducted to make the enhanced features speaker-invariant. Eventually, combined with TS learning, SIAFM-TS achieves the speech enhancement, knowledge distillation and speaker adaption in a unified framework. Evaluated on AMI dataset, the SIAFM-TS achieves 4.2% relative WER(over) reduction and 6.3% relative WER(non-over)

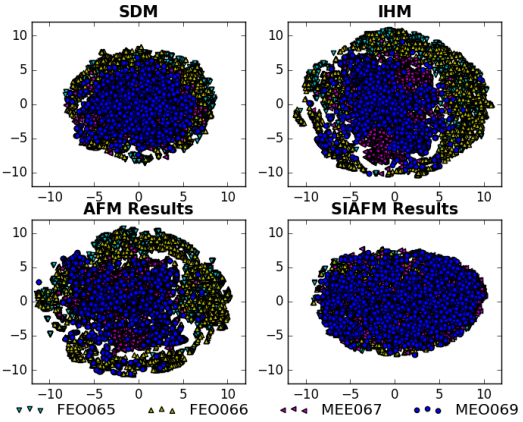

Figure 3: *t-SNE visualization of the SDM features, IHM features, AFM enhanced features and SIAFM enhanced features when speech frames aligned with phoneme 'ah' from two male and two female speakers.*

Table 3: *Performances of SDM model, MCT model, AFM model, SIAFM model and the SIAFM-TS model on the SDM eval data. Additionally, SDM-Eval represents the SDM eval data.*

Acoustic model	WER(over)	WER(nonover)
	SDM-Eval	SDM-Eval
SDM	54.0	44.5
MCT	53.0	43.2
AFM	52.1	42.5
SIAFM	51.4	41.4
SIAFM-TS	**50.8**	**40.5**

reduction compared with MCT model. Meanwhile, in comparison with SDM model baseline, SIAFM-TS reaches 5.9% relative WER(over) decrease and 9.0% relative WER(non-over) decrease.

7. Acknowledgements

This work is partially supported by the National Key Research and Development Program (Nos. 2016YFB0801203, 2016YFB0801200), the National Natural Science Foundation of China (Nos. 11590774, 11590770),the Key Science and Technology Project of the Xinjiang Uygur Autonomous Region (No.2016A03007-1),the Pre-research Project for Equipment of General Information System (No.JZX2017-0994/Y306).

8. References

[1] G. E. Hinton, L. Deng, D. Yu, G. E. Dahl, A. Mohamed, N. Jaitly, A. W. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath *et al.*, "Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82–97, 2012.

[2] P. Swietojanski, A. Ghoshal, and S. Renals, "Hybrid acoustic models for distant and multichannel large vocabulary speech recognition," in *Automatic Speech Recognition and Understanding*, 2013.

[3] T. Hain, L. Burget, J. Dines, P. N. Garner, F. Grzl, A. E. Hannani, M. Huijbregts, M. Karafit, M. Lincoln, and V. Wan, "Transcribing

meetings with the amida systems," *IEEE Transactions on Audio Speech and Language Processing*, vol. 20, no. 2, pp. 486–498, 2012.

[4] S. Mosayyebpour, M. Esmaeili, and T. A. Gulliver, "Single-microphone early and late reverberation suppression in noisy speech," *IEEE Transactions on Audio Speech and Language Processing*, vol. 21, no. 2, pp. 322–335, 2013.

[5] N. Mohammadiha and S. Doclo, "Speech dereverberation using non-negative convolutive transfer function and spectro-temporal modeling," *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 24, no. 2, pp. 276–289, 2016.

[6] M. L. Seltzer, "Bridging the gap: Towards a unified framework for hands-free speech recognition using microphone arrays," in *Hands-free Speech Communication and Microphone Arrays*, 2008.

[7] A. Narayanan and D. L. Wang, "Ideal ratio mask estimation using deep neural networks for robust speech recognition," in *IEEE International Conference on Acoustics*, 2013.

[8] F. Weninger, H. Erdogan, S. Watanabe, E. Vincent, J. L. Roux, J. R. Hershey, and B. Schuller, "Speech enhancement with lstm recurrent neural networks and its application to noise-robust asr," 2015.

[9] X. Lu, Y. Tsao, S. Matsuda, and C. Hori, "Speech enhancement based on deep denoising autoencoder," pp. 436–440, 2013.

[10] Y. Xu, J. Du, L. R. Dai, and C. H. Lee, "A regression approach to speech enhancement based on deep neural networks," *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 23, no. 1, pp. 7–19, 2015.

[11] Y. Qian, T. Tan, and D. Yu, "An investigation into using parallel data for far-field speech recognition," *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5725–5729, 2016.

[12] A. Sriram, H. Jun, Y. Gaur, and S. Satheesh, "Robust speech recognition using generative adversarial networks," *international conference on acoustics, speech, and signal processing*, pp. 5639–5643, 2018.

[13] Y. Shinohara, "Adversarial multi-task learning of deep neural networks for robust speech recognition." pp. 2369–2372, 2016.

[14] Z. Meng, J. Li, Z. Chen, Y. Zhao, V. Mazalov, Y. Gong, and B.-H. Juang, "Speaker-invariant training via adversarial learning," *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5969–5973, 2018.

[15] Z. Meng, Z. Chen, V. Mazalov, J. Li, and Y. Gong, "Unsupervised adaptation with domain separation networks for robust speech recognition," *2017 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, pp. 214–221, 2017.

[16] Z. Meng, J. Li, Y. Gong, and B.-H. Juang, "Adversarial teacher-student learning for unsupervised domain adaptation," *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5949–5953, 2018.

[17] K. Bousmalis, G. Trigeorgis, N. Silberman, D. Krishnan, and D. Erhan, "Domain separation networks," in *NIPS*, 2016.

[18] J. Li, M. Seltzer, X. Wang, and R. a. Zhao, "Large-scale domain adaptation via teacher-student learning." ISCA, August 2017.

[19] J. Li, R. Zhao, J. Huang, and Y. Gong, "Learning small-size dnn with output-distribution-based criteria." pp. 1910–1914, 2014.

[20] J. Carletta, S. Ashby, S. Bourban, M. Flynn, M. Guillemot, T. Hain, J. Kadlec, V. Karaiskos, W. Kraaij, and M. Kronenthal, "The ami meeting corpus: a pre-announcement," in *International Workshop on Machine Learning for Multimodal Interaction*, 2006, pp. 28–39.

[21] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz *et al.*, "The kaldi speech recognition toolkit," 2011.