Meta-Adapter for Self-Supervised Speech Models: A Solution to Low-Resource Speech Recognition Challenges

Yaqi Chen[1], Hao Zhang[1], Xukui Yang[1,2], Wenlin Zhang[1,2], Dan Qu[1,2*]
[1] School of Information Systems Engineering, Information Engineering University
[2]Laboratory for Advanced Computing and Intelligence Engineering
{chyaqi163, haozhang0126, wenlinzzz,gzyangx,qudan_xd}@163.com

Abstract

Self-supervised models have demonstrated remarkable performance in speech processing by learning latent representations from large amounts of unlabeled data. Although these models yield promising results on low-resource languages, the computational expense of fine-tuning all model parameters is prohibitively high. Adapters offer a solution by incorporating lightweight bottleneck structures into pre-trained models, enabling efficient parameter adaptation for downstream tasks. However, randomly initialized adapters often underperform in low-resource scenarios, limiting their applicability in low-resource languages. To address this issue, we develop the Meta-Adapter for self-supervised models to obtain meta-initialized parameters that facilitate quick adaptation to low-resource languages. Extensive experiments on the Common Voice and FLEURS datasets demonstrate the superior performance of Meta-Adapters on 12 low-resource languages spanning four different language families. Moreover, Meta-adapters show better generalization and extensibility than traditional pretraining methods.

Keywords: meta learning, speech recognition, low-resource, self-supervised model

1. Introduction

Automatic Speech Recognition (ASR) has revolutionized various aspects of people's lives, delivering remarkable success in several widely spoken languages. However, there are more than 7,000 languages in the world, and most of them contain limited labeled data, also known as low-resource languages. Compared with common languages, low-resource languages lack transcribed speech data, pronunciation dictionaries, and language scripts, making it difficult to build a usable ASR system.

Recently, self-supervised learning (SSL) has achieved significant advancements, enabling and bootstrapping ASR applications in low-resource languages through zero-shot or few-shot cross-lingual transfer. However, these advancements often require huge computational resources. For example, the largest versions of Wav2Vec2.0 (Baevski et al., 2020) and HuBERT (Hsu et al., 2021) contain approximately 317M and 964M parameters, respectively. And the MMS model (Pratap et al., 2023) can even reach up to 1 billion parameters. It is impractical to fine-tune such massive models for each low-resource language adaptation, which would consume significant storage, training, and inference costs.

Instead of fine-tuning all parameters individually for each low-resource language, an alternative approach is to utilize adapters. Adapters use a lightweight neural network integrated at each layer of the pre-trained model to adapt to downstream low-resource target languages (He et al., 2021). This approach offers several advantages. Firstly, it improves parameter efficiency by simply introducing additional trainable parameters in the adapters instead of modifying the entire pre-trained model. Secondly, using adapters has been shown to be more robust against overfitting, as it allows for more targeted adaptation to specific languages without affecting the shared representations.

However, traditional adapters often suffer from random initialization issues, leading to suboptimal performance when adapting to low-resource languages. To tackle this problem, researchers investigated various techniques to improve adapters' initialization. One approach is to pretrain the adapters with data from related languages or resources like multilingual learning. Another promising solution is meta-learning (Bansal et al., 2022). This approach utilizes the meta-learning algorithm to learn a better initialization for adapters, enabling them to quickly adapt to new languages with limited data. Meta-Adapters have shown superior results in various domains such as natural language processing (NLP) (Lai et al., 2022) and multilingual speech recognition (Hou et al., 2021).

However, despite the success of self-supervised models, there are no works exploring Meta-Adapters for SSL models. So we develop efficient Meta-Adapters based on the latest adapter structure for self-supervised models. This innovation enables accurate and parameter-efficient few-shot learning, which is crucial for tackling low-resource speech recognition challenges. In summary, the contributions of this paper are as follows:

*Corresponding author

LREC-COLING 2024, pages 11215–11221

- We develop the Meta-Adapter for self-supervised models, to our knowledge, which is the first application of Meta-Adapters in self-supervised models for low-resource speech recognition.

- Extensive experiments show that our proposed Meta-Adapters achieve the best performance on 12 low-resource languages from four different language families. Moreover, Meta-Adapters demonstrate better generalization and extensibility than other methods.

2. Related Work

Adapters. Adapters have been widely applied in computer vision (Rebuffi et al., 2017) and NLP (Houlsby et al., 2019) for parameter-efficient knowledge transfer within large-scale pretrained models. Later, it also demonstrated good performance in multilingual speech recognition (Pfeiffer et al., 2020) and speech translation (Le et al., 2021). (Pfeiffer et al., 2020) proposed dual adapters to capture language-related and task-related features for cross-lingual knowledge transfer. Recently, (Thomas et al., 2022) introduced the adapter into SSL models for the first time. Moreover, (Otake et al., 2022) developed a new adapter structure to make full use of the feature representation from low level to high level in the self-supervised model.

Meta Learning. Meta-learning (Hospedales et al., 2020) learns the meta-knowledge from lots of training tasks, enabling the model to quickly adapt to new tasks. Meta-learning has shown superior performance in various speech applications such as accent recognition (Winata et al., 2020), emotion recognition (Chopra et al., 2021), speaker recognition (Klejch et al., 2019) speech recognition (Xiao et al., 2021; Singh et al., 2022; Wang et al., 2023). However, these meta-learning methods are conducted on small-scale models, resulting in very limited performance. (Hou et al., 2021) proposed a meta-adapter for multilingual speech recognition models. However, existing applications of meta-learning techniques have not fully leveraged the potential of self-supervised models that perform superior for low-resource languages. So we developed the Meta-Adapter for the self-supervised models, which is based on the latest adapter structure that has shown robust performance in SSL models.

3. Method

3.1. Adapter Architecture

Our model architecture follows the structure proposed in work (Otake et al., 2022), as illustrated in Figure 1. The adapter structure consists of two parts: Layer adapters (L-adapters) and Encoder adapters (E-adapters). E-adapters are embedded in each encoder layer, while L-adapters directly connect each encoder layer to the head. This design allows for adaptive utilization of features across different layers of the self-supervised model, enabling quick adaptation to various downstream tasks.

As is shown in Figure 1, the modules highlighted in red are learnable, while the modules in gray are frozen. Additionally, the layer normalization applied to each encoder layer and head is also learnable.

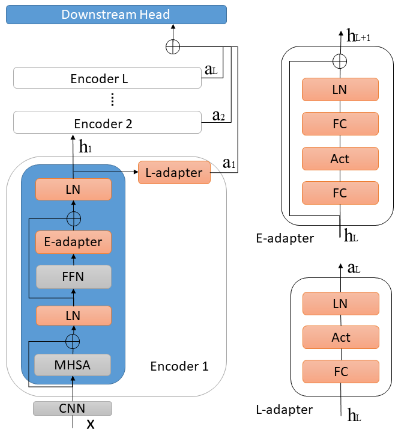


Figure 1: Model architecture and adapter structure.

3.2. Meta-Adapter

Suppose the meta-training dataset is a set of N languages, $D_s = D_s^i (i = 1, ...N)$. Each language D_s^i consists of the speech-text pairs. Unlike traditional machine learning, meta learning uses tasks as training samples to acquire generic meta-knowledge over a number of training episodes. In each episode, we sample N tasks from N languages to form a batch. For i-th language, we sample a task T_i from the $D_s^i (i = 1, 2, ..., N)$, and divide T_i into two subsets: the support set T_{sup}^i for meta-training and the query set T_{query}^i for meta-evaluation.

During both the pre-training and the fine-tuning, the parameters of the self-supervised model θ_W are frozen. The meta-learning algorithm trains the Meta-Adapter module to obtain a good initialization θ_M for quick adaptation to low-resource target languages. Learning valuable meta-knowledge from D_s is crucial for Meta-Adapters. The learning process of Meta-Adapters can be described as a bilevel optimization problem:

$$\min_{\theta_M} \sum_{i=1}^{N} L^{meta}(\theta_W, \omega^{*(i)}(\theta_M); T_{query}^i) \quad (1)$$

$$s.t. \;\; \omega^{*(i)}(\theta_M) = \arg\min l(\theta_W, \theta_M; T_{\text{sup}}^i) \quad (2)$$

Here, L^{meta} and l represent the meta loss (in the outer loop) and the task loss (in the inner loop), respectively. In particular, the inner loop (Eq.1) is designed to learn a language-specific base learner $\omega^{*(i)}(\theta_M)$ for each task using the support set T^i_{sup}, whereas the outer loop (Eq.2) learns meta-knowledge from these base learners with the query set T^i_{query}. Model-Agnostic Meta-Learning (MAML) (Finn et al., 2017) is one of the most representative algorithms in this area, utilized across various low-resource fields. Due to the challenge of computing the second-order derivatives and storing the Hessian matrix, we then describe two alternative meta learning methods for Meta-Adapters:

Meta-Adapters: Employing first-order model-agnostic meta-learning (FOMAML) (Finn et al., 2017) for Adapters, which omits the calculation of the second-order gradient by approximation.

Meta-Anil-Adapters: Employing the ANIL algorithm (Raghu et al., 2020) for Adapters, which only optimizes the (task-specific) head of the neural network in the inner loop.

4. Experiment Setup

4.1. Datasets

We used the data from Mozilla's Common Voice Corpus (Ardila et al., 2020) and FLEURS datasets (Conneau et al., 2022). From the Common Voice Corpus, we selected three languages as source tasks: German (42.72 h), Italian (50.04 h), and Swedish (32.17 h); and chose four languages as the target tasks: Catalan (5 min), Welsh (10 min), French (10 min), and Portuguese (10 min). Each language had 5 hours of development and test data available. As for the FLEURS datasets, we selected eight languages, with five from the Western European language family and the other three from other language families. The validation and testing splits were followed as provided in the official dataset.

4.2. Implementation Details

We used the WavLM Base model (Chen et al., 2022) as the self-supervised model. It has 12 Transformer encoder layers, 768-dimensional hidden states, and 8 attention heads, resulting in 94.70M parameters. We used the wavlm-base-plus [1] version, which trained on 94k hours of diverse data. The setting of adapters was as same as (Otake et al., 2022), and we trained the model for 50 epochs with a batch size of 32. During the adaptation process, the adapter is fine-tuned for 200 epochs with a batch size of 8. For all training, we set the early stop strategy for 3 times. We used Adam optimizer for the inner loop and outer loop, and the learning rate is 1e-3. We used Word Error Rate (WER) as the evaluation metric.

For each target language, we consider the following approaches as baselines: (i) FT-Full: Optimizing the full model parameters for task adaptation; (ii) FT-Head: Fine-tuning the head of model for adapting the task; (iii) Vanilla-Adapter: Train an adapter with randomly initialized parameters; (iv) Multi-Adapter: Train an adapter pre-trained by multilingual learning.

[1] https://huggingface.co/microsoft/wavlm-base-plus

5. Experiment Results

Results on the FLEURS dataset. Table 1 shows the performance of several different adapters on five languages from the FLEURS dataset. First, using FT-Full or FT-Head could not fit at 5%-shot and 10%-shot subsets, and the WER was always around 100%. Second, when using a randomly initialized adapter, the model converges on all languages with only 8.82 % of the parameters fine-tuned, illustrating the adapter's effectiveness. Finally, we can observe that Multi-Adapter shows better performance than Vanilla-Adapter in most languages except for Finnish, while the meta-learning methods including Meta-Adapter and Meta-Anil-Adapter can do better in all languages, showing superior performance to other methods.

Results on the Common Voice dataset. In addition, we tested the adapter's performance on four languages: Catalan (Ca), Welsh (Cy), French (Fr), and Portuguese (Pt). As shown in Table 2, it can be observed that although the Multi-Adapter's effectiveness, it exhibits significant instability, indicating that it is highly influenced by the correlation between the pre-training languages and the target languages. It might have the overfitting problem. However, Meta-Adapters may not perform as well as Multi-Adapters for Catalan, but they demonstrate effectiveness across most languages and exhibit stronger generalization capabilities.

Effect of different proportions of data. To explore the relationship between the adapter's performance and the amount of training data, we sampled different proportions of data from Spanish and Galician for adaptation, and the results are shown in Figure 2. It can be found that Meta-Adapters are effective in low-resource scenarios. However, the performance gap between Meta-Adapters and Vanilla-Adapters gradually diminishes as the amount of data increases. When the proportion reaches 50%, the effect of meta-learning is very limited. Moreover, FT-Full fails to converge under very low-resource conditions, but can achieve the best performance when trained with 100% of the available data.

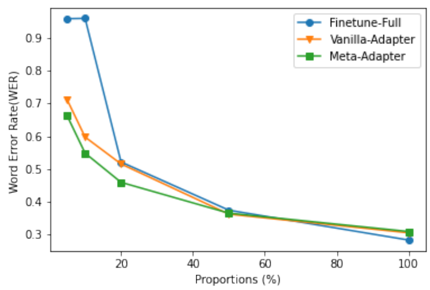
Effect of different pretraining epochs. We compared the performance of five languages using

Table 1: Word error rates (WER) on five target languages of FLEURS datasets.

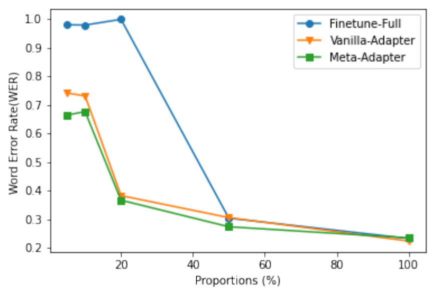
Languages	Maltese		Finnish		Galician		Croatian		Spanish		Avg.
Porportions	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	/
FT-Full	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
FT-Head	0.998	0.999	1.000	0.999	0.959	0.960	0.992	0.993	0.9811	0.981	0.986
Vanilla-Adapter	0.784	0.666	0.656	0.625	0.712	0.598	0.700	0.633	0.742	0.732	0.685
Multi-Adapter	0.712	0.600	0.693	0.600	0.6767	0.580	0.654	0.573	0.701	0.682	0.647
Meta-Adapter	**0.665**	**0.554**	0.668	0.564	0.663	**0.549**	**0.642**	**0.536**	0.665	0.677	0.618
Meta-Anil-Adapter	0.680	0.560	**0.645**	**0.553**	**0.642**	0.552	0.658	0.549	**0.553**	**0.654**	**0.605**

Table 2: Word error rates (WER) on four target languages of Common Voice.

Method	Ca	Cy	Pt	Fr	Avg.
Vanilla-Adapter	0.895	0.966	0.834	**0.889**	0.896
Multi-Adapter	**0.592**	0.840	0.958	0.925	0.829
Meta-Adapter	0.748	**0.734**	**0.822**	0.920	**0.806**



(a) Galician



(b) Spanish

Figure 2: Word error rate (WER) curves under different proportions of data.

5% of data under different pre-training epochs. As shown in Figure 3, we can see that most languages achieve good performance when the pre-training epoch is 20. As the pre-training epoch increases, some languages may overfit, but the overall average performance remains relatively stable.

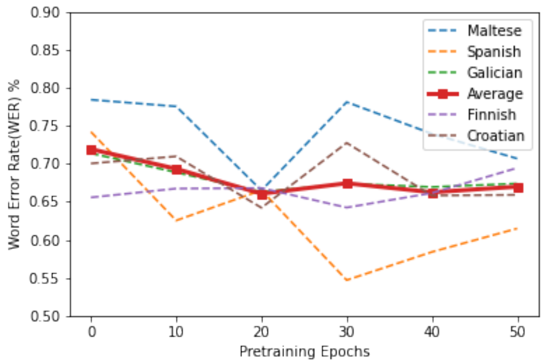


Figure 3: Word error rate (WER) of adapter pretraining epochs on five target languages.

Table 3: Word error rate (WER) on three target languages of other language families.

Languages	Kk	Af	Ur	Avg.
Vanilla-Adapter	**0.457**	0.728	0.805	0.663
Meta-Adapter	0.459	**0.666**	**0.719**	**0.615**

Extensibility to other different language families. Existing adapters are highly dependent on the similarity between the self-supervised pre-training languages and the target languages. The languages used in the pre-training and meta learning process belong to the European language family. In order to explore the generalization of Meta-Adapters to other language families, we selected three languages: Kazakh (Kk), Afrikaans (Af), Urdu (Ur), which are from three different language families: the Middle East, Africa, and South Asia, respectively. And we use 10%-shot subset for adaptation. As shown in Table 3, it can be seen that Meta-Adapters have a better generalization effect in other language families.

Learning curves of fine-tuning target Languages. To explore the rapid adaptation process of Meta-Adapters in fine-tuning, we fine-tuned the Galician language using only 5% of the data. Experimental results show that Vanilla-adapters converges faster in the first 20 epochs, but they tend to converge to suboptimal performance eventually. However, Multi-Adapters and Meta-Adapters do not converge in the first 20 epochs, then rapidly decline, and finally converge to a relatively good performance. We analyze that this is caused by warmup mechanisms. The pre-trained adapters converge to a local optimum in pretraining, and it needs a larger learning rate to break free from the local optima. Overall, Meta-Adapters achieve faster and better performance than Multi-Adapters, which shows the superiority of Meta-Adapters in terms of their fast learning ability.

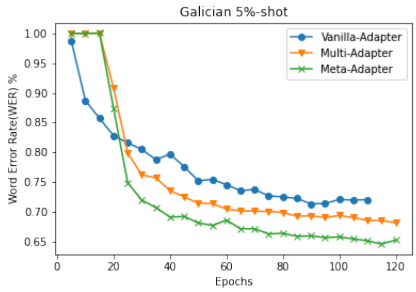


Figure 4: Word error rate (WER) performance of Galician using 5% of data under different adapters.

6. Conclusion

In this work, we propose Meta-Adapter for self-supervised speech models, enabling fast and better adaptation for low-resource languages. Our experiments show the effectiveness of the proposed method on 12 low-resource languages. In the future, we will explore more adaptable learning algorithms for enhancing adapters' performance.

7. Acknowledgement

This work was supported by the National Natural Science Foundation of China (No.62171470), Natural Science Foundation of Henan Province of China (No.232300421240), and Henan Zhongyuan Science and Technology Innovation Leading Talent Project (No.234200510019).

8. References

Rosana Ardila, Megan Branson, Kelly Davis, Michael Kohler, Josh Meyer, Michael Henretty, Reuben Morais, Lindsay Saunders, Francis M. Tyers, and Gregor Weber. 2020. Common voice: A massively-multilingual speech corpus. In *Proceedings of The 12th Language Resources and Evaluation Conference, LREC 2020, Marseille, France, May 11-16, 2020*, pages 4218–4222. European Language Resources Association.

Alexei Baevski, Yuhao Zhou, Abdelrahman Mohamed, and Michael Auli. 2020. wav2vec 2.0: A framework for self-supervised learning of speech representations. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*.

Trapit Bansal, Salaheddin Alzubi, Tong Wang, Jay-Yoon Lee, and Andrew McCallum. 2022. Meta-adapters: Parameter efficient few-shot fine-tuning through meta-learning. In *International Conference on Automated Machine Learning, AutoML 2022, 25-27 July 2022, Johns Hopkins University, Baltimore, MD, USA*, volume 188 of *Proceedings of Machine Learning Research*, pages 19/1–18. PMLR.

Sanyuan Chen, Chengyi Wang, Zhengyang Chen, Yu Wu, Shujie Liu, Zhuo Chen, Jinyu Li, Naoyuki Kanda, Takuya Yoshioka, Xiong Xiao, Jian Wu, Long Zhou, Shuo Ren, Yanmin Qian, Yao Qian, Jian Wu, Michael Zeng, Xiangzhan Yu, and Furu Wei. 2022. Wavlm: Large-scale self-supervised pre-training for full stack speech processing. *IEEE J. Sel. Top. Signal Process.*, 16(6):1505–1518.

Suransh Chopra, Puneet Mathur, Ramit Sawhney, and Rajiv Ratn Shah. 2021. Meta-learning for low-resource speech emotion recognition. In *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2021, Toronto, ON, Canada, June 6-11, 2021*, pages 6259–6263. IEEE.

Alexis Conneau, Min Ma, Simran Khanuja, Yu Zhang, Vera Axelrod, Siddharth Dalmia, Jason Riesa, Clara Rivera, and Ankur Bapna. 2022. FLEURS: few-shot learning evaluation of universal representations of speech. In *IEEE Spoken Language Technology Workshop, SLT 2022, Doha, Qatar, January 9-12, 2023*, pages 798–805. IEEE.

Chelsea Finn, P. Abbeel, and Sergey Levine. 2017. Model-agnostic meta-learning for fast adaptation of deep networks. In *International Conference on Machine Learning*.

Ruidan He, Linlin Liu, Hai Ye, Qingyu Tan, Bosheng Ding, Liying Cheng, Jia-Wei Low, Lidong Bing, and Luo Si. 2021. On the effectiveness of adapter-based tuning for pretrained language model adaptation. In *Proceedings of the 59th*

Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 1: Long Papers), Virtual Event, August 1-6, 2021, pages 2208–2222. Association for Computational Linguistics.

Timothy M. Hospedales, Antreas Antoniou, Paul Micaelli, and Amos J. Storkey. 2020. Meta-learning in neural networks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44:5149–5169.

Wenxin Hou, Yidong Wang, Shengzhou Gao, and Takahiro Shinozaki. 2021. Meta-adapter: Efficient cross-lingual adaptation with meta-learning. In *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2021, Toronto, ON, Canada, June 6-11, 2021*, pages 7028–7032. IEEE.

Wenxin Hou, Han Zhu, Yidong Wang, Jindong Wang, Tao Qin, Renjun Xu, and Takahiro Shinozaki. 2022. Exploiting adapters for cross-lingual low-resource speech recognition. *IEEE ACM Trans. Audio Speech Lang. Process.*, 30:317–329.

Neil Houlsby, Andrei Giurgiu, Stanislaw Jastrzebski, Bruna Morrone, Quentin de Laroussilhe, Andrea Gesmundo, Mona Attariyan, and Sylvain Gelly. 2019. Parameter-efficient transfer learning for NLP. In *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*, volume 97 of *Proceedings of Machine Learning Research*, pages 2790–2799. PMLR.

Wei-Ning Hsu, Benjamin Bolte, Yao-Hung Hubert Tsai, Kushal Lakhotia, Ruslan Salakhutdinov, and Abdelrahman Mohamed. 2021. Hubert: Self-supervised speech representation learning by masked prediction of hidden units. *IEEE ACM Trans. Audio Speech Lang. Process.*, 29:3451–3460.

Ondrej Klejch, Joachim Fainberg, Peter Bell, and Steve Renals. 2019. Speaker adaptive training using model agnostic meta-learning. In *IEEE Automatic Speech Recognition and Understanding Workshop, ASRU 2019, Singapore, December 14-18, 2019*, pages 881–888. IEEE.

Wen Lai, Alexandra Chronopoulou, and Alexander Fraser. 2022. m^{4}adapter: Multilingual multi-domain adaptation for machine translation with a meta-adapter. *CoRR*, abs/2210.11912.

Hang Le, Juan Miguel Pino, Changhan Wang, Jiatao Gu, Didier Schwab, and Laurent Besacier. 2021. Lightweight adapter tuning for multilingual speech translation. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL/IJCNLP 2021, (Volume 2: Short Papers), Virtual Event, August 1-6, 2021*, pages 817–824. Association for Computational Linguistics.

Shinta Otake, Rei Kawakami, and Nakamasa Inoue. 2022. Parameter efficient transfer learning for various speech processing tasks. *ArXiv*, abs/2212.02780.

Ankita Pasad, Ju-Chieh Chou, and Karen Livescu. 2021. Layer-wise analysis of a self-supervised speech representation model. In *IEEE Automatic Speech Recognition and Understanding Workshop, ASRU 2021, Cartagena, Colombia, December 13-17, 2021*, pages 914–921. IEEE.

Jonas Pfeiffer, Ivan Vulic, Iryna Gurevych, and Sebastian Ruder. 2020. MAD-X: an adapter-based framework for multi-task cross-lingual transfer. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020*, pages 7654–7673. Association for Computational Linguistics.

Vineel Pratap, Andros Tjandra, Bowen Shi, Paden Tomasello, Arun Babu, Sayani Kundu, Ali Mamdouh Elkahky, Zhaoheng Ni, Apoorv Vyas, Maryam Fazel-Zarandi, Alexei Baevski, Yossi Adi, Xiaohui Zhang, Wei-Ning Hsu, Alexis Conneau, and Michael Auli. 2023. Scaling speech technology to 1, 000+ languages. *ArXiv*, abs/2305.13516.

Aniruddh Raghu, Maithra Raghu, Samy Bengio, and Oriol Vinyals. 2020. Rapid learning or feature reuse? towards understanding the effectiveness of MAML. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net.

Sylvestre-Alvise Rebuffi, Hakan Bilen, and Andrea Vedaldi. 2017. Learning multiple visual domains with residual adapters. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 506–516.

Jui Shah, Yaman Kumar Singla, Changyou Chen, and Rajiv Ratn Shah. 2021. What all do audio transformer models hear? probing acoustic representations for language delivery and its structure. *ArXiv*, abs/2101.00387.

Satwinder Singh, Ruili Wang, and Feng Hou. 2022. Improved meta learning for low resource speech recognition. In *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2022, Virtual and Singapore, 23-27 May 2022*, pages 4798–4802. IEEE.

Bethan Thomas, Samuel Kessler, and Salah Karout. 2022. Efficient adapter transfer of self-supervised speech models for automatic speech recognition. In *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2022, Virtual and Singapore, 23-27 May 2022*, pages 7102–7106. IEEE.

Qiuli Wang, Wen-Rui Hu, Lin Li, and Qingyang Hong. 2023. Meta learning with adaptive loss weight for low-resource speech recognition. *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.

Genta Indra Winata, Samuel Cahyawijaya, Zihan Liu, Zhaojiang Lin, Andrea Madotto, Peng Xu, and Pascale Fung. 2020. Learning fast adaptation on cross-accented speech recognition. In *Interspeech 2020, 21st Annual Conference of the International Speech Communication Association, Virtual Event, Shanghai, China, 25-29 October 2020*, pages 1276–1280. ISCA.

Yubei Xiao, Ke Gong, Pan Zhou, Guolin Zheng, Xiaodan Liang, and Liang Lin. 2021. Adversarial meta sampling for multilingual low-resource speech recognition. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*, pages 14112–14120. AAAI Press.