ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.03 No.04 (2025)

A SYSTEMATIC REVIEW OF PARAMETER-EFFICIENT FINE-TUNING (PEFT) IN SPEECH PROCESSING

Noor Ul Ain Liaqat[1], Mutaher Ijaz[2], Umair Muneer Butt[3], Imtiaz Hussain[4]

[1,3,4]Department of Computer Science, University of Management and Technology, Sialkot, Pakistan Email: noorulainliaqat96@gmail.com , umair.muneer@skt.umt.edu.pk, Imtiaz.hussain@skt.umt.edu.pk

[2]Department of Computer Science, Sir Syed CASE Institute of Technology, Islamabad, Pakistan Email: mutaherijaz178@gmail.com

***Abstract*—**

Recent breakthroughs in large-scale speech models such as Whisper, Wav2Vec 2.0, and HuBERT have greatly enhanced speech processing tasks. Full fine-tuning comes at a prohibitive cost, though, which restricts their application to low- resource or real-time settings. The parameter-efficient fine-tuning (PEFT) approaches—e.g., LoRA, QLoRA, adapters, and prompt tuning—allow for compact adaptation by fine-tuning only a small subset of parameters. We review 33 studies (2021–2025) using PEFT for applications such as ASR, speaker verification, and emotion recognition. We organize methods by task, compare efficiency and accuracy, and determine prominent trends. Results indicate PEFT produces competitive results with reduced cost, enabling scalable deployment in resource-poor environments.

***Index Terms*—Fine-tuning, PEFT, Privacy, Speech processing**

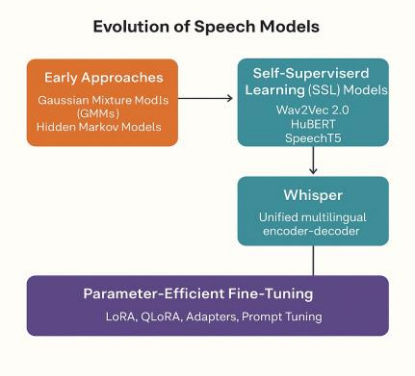
I. INTRODUCTION

A. *BACKGROUND AND MOTIVATION*

Large-scale pre-trained models like Whisper, Wav2Vec 2.0, and HuBERT have greatly improved speech and audio pro- cessing. They perform well in applications such as auto- matic speech recognition (ASR), speaker verification, and emotion identification. Nevertheless, complete fine-tuning of these models—typically with hundreds of millions of pa- rameters—is still computationally heavy and memory-hungry, keeping them from being deployed in low-resource or real- time environments [1]. Speech applications are commonly subject to variation of input length, speaker characteristics, and accented or multilingual speech. Such variations, together with data and hardware constraints, increase the inefficiencies of full model adaptation [2]– [3].

B. *PARAMETER EFFICIENT TUNING (PEFT)*

PEFT methods have been developed to solve these problems by making adaptation possible with far fewer parameters. Techniques like Low-Rank Adaptation (LoRA), Quantized LoRA (QLoRA), and Adapter Tuning make modular and efficient updates to frozen pre-trained models possible. LoRA adds low-rank trainable matrices to attention modules, cutting parameter usage by more than 90% [1], [2], and QLoRA adds quantization for further reducing memory needs [4]. These are particularly promising in speech tasks, where rapid and lightweight domain adaptation is most crucial.



ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.03 No.04 (2025)

Fig. 1: Speech Methods

C. *PEFT IN SPEECH APPLICATIONS*

Recent works have explored PEFT methods in ASR, speaker verification, and speech emotion recognition using models including but not limited to Whisper, Wav2Vec 2.0, HuBERT, and SpeechT5 [3]– [5]. LoRA and adapters have shown efficiency in multilingual ASR and lowresource speech pro- cessing, while prompt tuning has shown promise in speaker- aware and multitask systems. Although these advances have taken place, the literature is still piecemeal. There are few comparative assessments across tasks and approaches, justify- ing a systematic synthesis.

D. *OBJECTIVES*

This review responds to the following research questions:

- What PEFT methods have been used on speech tasks?
- Where (e.g., ASR, speaker verification, SER) are these techniques optimally used?
- What are the trade-offs among performance, efficiency, and generalizability?
- What are the challenges and gaps that remain?

We examine 33 peer-reviewed papers and preprints issued between 2021 and 2025 to answer these questions.

Section 2 describes the review process. Section 3 presents important PEFT methods and speech models. Section 4 groups
the literature by application. Section 5 compares performance and efficiency. Section 6 presents trends and future directions. Section 7 concludes the paper.

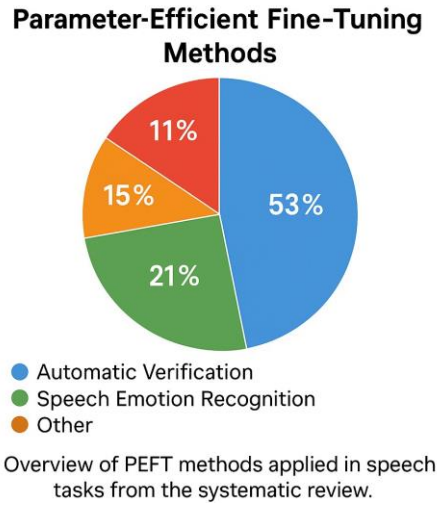


Fig. 2: Fine Tuning Methods

II. METHODOLOGY

This section describes the methodology adopted for this systematic literature review (SLR) based on evidence-based best practices. It contains the research questions, selection criteria, search strategy, study selection process, and data extraction method.

A. *RESEARCH QUESTIONS*

This review responds to the following research questions:

- RQ1: How do the various PEFT methods (e.g., LoRA, QLoRA, Adapters) get used in speech and audio pro- cessing?
- RQ2: In which tasks (e.g., ASR, speaker verification, emotion recognition) are these methods employed?

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.03 No.04 (2025)

- RQ3: What are the efficiency and performance trade-offs among techniques?
- RQ4: What are the limitations and future directions for PEFT in speech?

B. *SEARCH STRATEGY*

The search process involved systematic searching across four leading academic databases:

- IEEE Xplore.
- ACL Anthology
- SpringerLink.
- arXiv. Sample Queries
- "LoRA" AND "speech recognition"
- "adapter tuning" AND "speaker verification"
- "QLoRA" AND "audio processing"
- "Whisper fine-tuning" AND "low-resource speech"

The search was conducted between January 2021 to April 2025, including both core works and current developments.

Citation chaining and snowballing were employed to uncover more relevant studies.

C. *STUDY SELECTION*

The study selection followed a three-stage process:

- 1. Initial Screening includes Title and abstract review for relevance.
- 2. Full-Text Review for evaluation of inclusion criteria.
- 3. 33 studies were selected, including conference papers (e.g., ICASSP, Interspeech) and peer-reviewed preprints.

Figure 1 illustrates the selection process (PRISMA diagram optional).

D. *INCLUSION CRITERIA*

The inclusion criteria for selecting article, reviews, confer- ence paper or journals is defines as follows:

- Peer-reviewed papers, conference articles, or preprints (2021–2025).
- Research based on PEFT techniques applied to speech/audio tasks.
- English-language publications with quantitative or quali- tative analysis.

E. *EXCLUSION CRITERIA*

The exclusion criteria for selecting article, reviews, confer- ence paper or journals selection is defined as below:

- Research based only on full fine-tuning.
- Research on language modeling or vision without speech use cases.
- Non-English articles, blog posts, or tutorials.
- Duplicates (e.g., preprint and published version).

F. *DATA EXTRACTION AND SYNTHESIS*

From each selected study, the following data were extracted:

- PEFT method(s) (LoRA, QLoRA, Adapter, Prompt)
- Task type (e.g., ASR, SER, SV)
- Base model (e.g., Whisper, Wav2Vec2, HuBERT)
- Evaluation datasets (e.g., LibriSpeech, VoxCeleb
- Metrics (e.g., WER, EER, accuracy)
- Parameter counts or efficiency gains
- Comparative results, if reported

Information gleaned was grouped by speech task and themati- cally organized to facilitate analysis in the following sections. Results were quantitatively tabled to allow comparison be- tween studies, and qualitative findings were synthesized to determine trends, strengths, and weaknesses of different PEFT methods.

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.03 No.04 (2025)

III. Background and Related Work

A. Evolution of Speech Models

The field of speech processing has evolved from traditional statistical models (e.g., GMMs, HMMs) to deep learning and self-supervised learning (SSL) approaches. Models such as Wav2Vec 2.0, HuBERT, and SpeechT5 leverage large- scale unlabelled data to learn speech representations, enabling strong performance across tasks like ASR, speaker verification, and emotion recognition. OpenAI's Whisper further advanced this trend by introducing a multilingual, multitask encoder- decoder model trained on more than 680,000 hours of audio. Although effective, these models possess hundreds of millions of parameters, and complete fine-tuning is computationally prohibitive and inappropriate for real-time or low-resource settings.

B. Parameter-Efficient Fine-Tuning (PEFT) Techniques

PEFT approaches work around these issues by fine-tuning large models with few parameter updates. The following techniques are well-recognized in speech research:

1) *Low-Rank Adaptation (LoRA):* LoRA adds low-rank trainable matrices to frozen model layers, enabling efficient adaptation with <1% of the parameters updated [1], [2]. It has been widely used in Whisper and other speech models [3], [6], [7].

2) *Quantized LoRA (QLoRA):* QLoRA adds 4-bit quanti- zation to LoRA to save further memory and computational requirements. With minimal accuracy sacrifices, QLoRA pro- vides efficient adaptation for big speech models [4], [6].

3) *Adapter Tuning:* Adapters are light-weight modules placed between layers of a model, allowing task-specific updates while freezing the backbone. It has been applied in speaker verification [8], cross-lingual ASR [9], and emotion recognition [10].

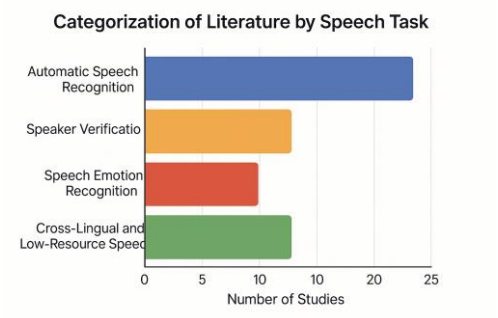
4) *Prompt and Prefix Tuning:* Prompt tuning appends learn- able embeddings to the input sequence, influencing model behavior without changing weights. It has exhibited initial potential in ASR and target-speaker adaptation [11], [12], [13].

5) *Other Approaches:* Other PEFT methods are Bit- Fit (bias-only tuning), IA³, and TAML (task-agnostic meta-learning adapters) [14], which investigate efficiency-generalization trade-offs.

C. Comparison with Full Fine-Tuning

Complete fine-tuning updates all model parameters, with the benefit of high task performance in exchange for memory, compute, and storage. PEFT techniques provide 10×–100× less parameter updates while achieving near-equivalent per- formance in most speech tasks [1], [6], [8]– [10], [15]– [9],
[16]– [17]. Their modularity enables multi-task learning, quick adaptation, and edge deployment, which increasingly make them appealing in contemporary speech processing pipelines.

Categorization by Speech Task



ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.03 No.04 (2025)

Fig. 3: Categorization of Speech Task

A. *AUTOMATIC SPEECH RECOGNITION (ASR)*

ASR is the most predominant use of PEFT, with more than half of the 33 studies reviewed addressing it. LoRA has been used widely for Whisper-based ASR in multilingual, accented, and low-resource environments [3], [6], [7], [18], [19], [20]. Liu et al. [3] proposed Sparsely Shared LoRA for child speech, achieving better accuracy with reduced parameter updates. Liu and Qu [20] showed greater than 95

QLoRA was applied for effective LLM-rescoring in hybrid ASR frameworks [2], [21] and obtained competitive WER on LibriSpeech, Common Voice, and MLS. Adapter tuning has also shown useful for ASR, with Thomas et al. [15] and Sim et al. [16] confirming its efficiency and sturdiness. Prompt tuning has been useful in context-aware decoding with Whisper and other frameworks [11], [12], [13], [22].

B. *SPEAKER VERIFICATION*

In speaker verification, where data is typically limited, adapter-based approaches have facilitated effective model adaptation. Peng et al. [7], [17] used adapters in Transformer backends for SV and obtained robust performance on Vox- Celeb. Inoue et al. [23] proposed ELP-Adapters for cross- domain adaptation with light updates.

C. *SPEECH EMOTION RECOGNITION (SER)*

SER is hindered by poor-quality and low-resource data. Sampath et al. [8] utilized LoRA on SSL models with full fine- tuning performance replicated with <5% parameter updates on IEMOCAP. Chen and Rudnicky [24] utilized Wav2Vec 2.0 with lightweight adaptation on CREMA-D and IEMOCAP. Prompt tuning was also investigated in visual speech emotion tasks [12].

D. *CROSS-LINGUAL AND LOW RESOURCE SPEECH*

PEFT has seen increasing popularity in low-resource and multilingual ASR. Meta-learning and adapter tuning were combined in TAML-Adapters for low-resource ASR [14], whereas Layer-Adaptive LoRA [25] and adapter-based ap- proaches [9] treated parameter efficiency in multilingual sce- narios. Two-stage LoRA and ensemble methods were sug- gested by Kwok et al. [26], [27] to scale across languages.

Whisper fine-tuning across Indic languages also proved suc- cessful with ¡1M trainable parameters [18], [19], [20].

E. *PROMPTING AND MULTIMODAL SYSTEMS*

Techniques for prompting are on the rise in speech tasks. Task-agnostic soft prompting for ASR, speaker ID, and key- word spotting was introduced by SpeechPrompt [28]. Fine-tuning Whisper with speaker prompts was done by Ma et al.
[13] for conversational ASR. Hybrid and multimodal models that use speech encoders and LLMs or pseudo-language pre- training have been suggested [22], [29].

IV. COMPARATIVE ANALYSIS

This part discusses 33 studies according to PEFT technique, performance, efficiency, and application area.

A. *SUMMARY OF REVIEWED STUDIES*

Table 1 summarizes some key points in chosen studies, such as task, PEFT approach, base model, dataset, and performance measurement. LoRA, Adapter tuning, QLoRA, and Prompt Tuning were utilized most often.

Aspect	LoRA	QLoRA	Adapters	Prompt Tuning

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR

Usage Frequency	Most widely used across studies	Moderate usage, mainly in hybrid ASR	Common in speaker verification and multilingual ASR	Emerging, less mature in literature
Applications	ASR, SER	ASR (rescoring with LLMs)	SV, multilingual and low-resource ASR	Speaker-aware and multitask ASR
Parameter Update (%)	<10% (5M−15M)	Similar to LoRA with quantization applied	Adds 2−4% via bottleneck layers	<1M parameters; task-sensitive
Performance Gap vs Full FT	<=1.5% in ASR, 2−3% in SER/SV	Comparable in ASR rescoring	Slight drop, stable in low-resource settings	Variable; depends on pretaining and task
Memory Efficiency	30−70% memory savings	Adds quantization for further memory reduction	Efficient for multi-task use, avoids duplicating base weights	Lightweight but harder to tune
Dataset Coverage	LibriSpeech, MLS, Common Voice, IndicSpeech	LibriSpeech, Common Voice, MLS	VoxCeleb, BABEL, low resource multilingual datsets	VoxCeleb, LibriSpeech, ASR multimodal datasets

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR

Real–Time Deploy-ment Evidence	Sparse; potential acknowl-edged	Limited; promisin g due to compact size	Not extensive ly tested on rea l hardware	Few deployme nt-focused studies

TABLE I: Comparative Summary of PEFT Techniques in Speech Tasks

COMPARISON BY PEFT TECHNIQUE

- ***LoRA***: Most widely adopted. Demonstrates near full fine-tuning performance for ASR and SER with ¡10% parameters updated [3], [6], [20].
- ***QLoRA***: Used for rescoring/decoding with LLMs; com- pact memory usage [2], [21].
- ***Adapters***: Applied in SV and multilingual ASR. Facili- tates modular, task-agnostic tuning [7], [15], [9], [17].
- ***Prompt Tuning***: Helpful in multitask and speaker-aware ASR, although less mature [11], [12], [13], [28].

B. *EFFICIENCY-PERFORMANCE TRADE-OFFs*

- ***LoRA***: 5M–15M parameters updated (usually ¡10%).
- ***Accuracy/WER***: : PEFT usually lags behind full fine- tuning by 1.5% in ASR; 1–3% gaps in SER and SV are common but acceptable for most uses.
- ***Adapters***: Add 2–4% parameters through bottlenecks.
- ***Prompt Tuning***: Overhead (<1M), more sensitive to task and pretraining.

C. *TASK AND DATA COVERAGE*

- ***ASR***: 60% of the literature; LibriSpeech, MLS, Common Voice shared.
- ***Low-resource/cross-lingual/ ASR***: Leveraged BABEL, IndicSpeech [14], [18], [20].
- ***SER/SV***: Less common but proved generalizability.

D. *RESOURCE IMPLICATIONS*

- ***Memory Savings***: LoRA and QLoRA save memory usage by 30–70%.
- ***Adapters***: Facilitate multi-task adaptation without repli- cating base weights.
- ***Deployment***: Limited studies compared real-time or edge deployment; future work is needed.

V. DISCUSSION

This section highlights the key findings from the system- atic literature review, including emerging trends, limitations, and directions for future work. The discussion follows three themes: research patterns observed, methodological limita- tions, and future directions on the application of parameter- efficient fine-tuning (PEFT) in speech and audio processing.

A. *EMERGING TRENDS*

- ***LoRA Dominance in ASR***: LoRA is most widely adopted PEFT approach, particularly with Whisper and Wav2Vec
  2.0 for ASR, based on ease of integration and parameter efficiency [3], [6], [19], [20].
- ***Multilingual and Low-Resource Speech Focus***: PEFT methods—particularly Adapters and layer-specific LoRA—are used extensively for cross-lingual and low-resource ASR [14], [9], [25], [26], [27].
- ***Rise of Prompt Tuning***: Less prevalent but increasing, prompt-based approaches have been used in multitask and speaker-aware ASR [11], [12], [13], [28], indicating prospects for scalable.

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.03 No.04 (2025)

- ***Standard Benchmarks***: Results are typically reported on benchmarked datasets (LibriSpeech, MLS, VoxCeleb, IEMOCAP), facilitating indirect comparison.
- ***Quantization Gains in Efficiency***: QLoRA and other methods minimize parameter numbers and memory us- age, essential for scalable deployment [2], [21].

B. *CHALLENGES AND LIMITATIONS*

- ***Lack of Consistent Evaluation***: Research frequently excludes uniform metrics (e.g., FLOPs, latency), making cross-PEFT approach comparisons difficult.
- ***ASR-Centric Research***: There has been little work on PEFT in other speech tasks such as speaker diarization, language ID, or conversation modeling.
- ***Limited Hyperparameter Analysis***: Few papers study the effect of important tuning parameters (e.g., LoRA rank, adapter size), which reduces reproducibility.
- ***Sparse Deployment Evidence***: While there are motiva- tions for mobile or edge deployment, few studies measure inference time or performance on actual hardware.
- ***Limited Method Comparisons***: Strict head-to-head com-

parisons between PEFT methods are scarce, making hardware limitations. Although encouraging findings, existing studies are still fragmented with minimal benchmarking and real-world testing. Wider validation and hybrid methods are required. PEFT has high potential for effective, scalable AI for speech in low-resource and practical environments.

FUTURE RESEARCH DIRECTIONS

- ***Unified Benchmarks***: Establish standard protocols in- cluding accuracy, memory, training time, and inference latency to enable effective comparisons.
- ***Cross-Task Generalization***: Assess if individual PEFT configurations are transferrable across ASR, SER, SV, etc., potentially utilizing multitask/meta-learning config- urations.
- ***Hybrid PEFT Methods***: Hybridize methods (e.g., LoRA

+ Adapters, QLoRA + Prompts) to strike a balance between trade-offs; initial studies are promising [9], [28].

- ***Edge and Real-World Testing***: Future research needs to subject PEFT models to real-world hardware (e.g., mobile GPUs, microcontrollers) to ensure deployability.
- ***Federated and Privacy-Aware PEFT***: Investigate the application of PEFT in federated learning for speech model adaptation preserving privacy.
- ***Multimodal Expansion***: Scale PEFT to multimodal sys- tems (speech + text + vision), leveraging initial explo- rations in conversational and speaker-aware models [13], [28], [29].

VI. CONCLUSION

The review analyzed 33 papers on parameter-efficient fine- tuning (PEFT) techniques—LoRA, QLoRA, adapter tuning, and prompt tuning—for speech and audio applications such as ASR, speaker verification, and emotion recognition. PEFT methods cut memory and parameter expense without sac- rificing great performance, allowing for scalable adaptation of huge models. LoRA was used most broadly, particularly with Whisper; adapters performed best with speaker and multilingual applications; QLoRA enabled deployment under

REFERENCES

[1] L. Wang, S. Chen, L. Jiang *et al.*, "Parameter-efficient fine-tuning in large language models: a survey of methodologies," *Artificial Intelli- gence Review*, vol. 58, p. 227, 2025.

[2] Y. Yu *et al.*, "Low-rank adaptation of large language model rescoring for parameter-efficient speech recognition," in *Proceedings of the IEEE ASRU Workshop*, Taipei, Taiwan, 2023, pp. 1–8.

[3] W. Liu, Y. Qin, Z. Peng, and T. Lee, "Sparsely shared lora on whisper for child speech recognition," in *ICASSP 2024*, Seoul, South Korea, 2024,

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.03 No.04 (2025)

pp. 11 751–11 755.

[4] Z.-C. Chen, C.-L. Fu, C.-Y. Liu, S.-W. D. Li, and H.-Y. Lee, "Exploring efficient-tuning methods in self-supervised speech models," in *Proceed- ings of the IEEE SLT Workshop*, Doha, Qatar, 2023, pp. 1120–1127.

[5] P. Bell, J. Fainberg, O. Klejch, J. Li, S. Renals, and P. Swietojanski, "Adaptation algorithms for neural network-based speech recognition: An overview," *IEEE Open Journal of Signal Processing*, vol. 2, pp. 33–66, 2021.

[6] Y. Fathullah *et al.*, "Prompting large language models with speech recognition abilities," in *ICASSP 2024*, Seoul, South Korea, 2024, pp.

relative strengths ambiguous.13 351–13 355.˘

J. Peng, O. Plchot, T. Stafylakis, L. Mosner, L. Burget, and J. Cernocky, "An attention-based backend allowing efficient fine-tuning of trans- former models for speaker verification," in *Proceedings of the IEEE SLT Workshop*, Doha, Qatar, 2023, pp. 555–562.

[7] A. Sampath, J. Tavernor, and E. M. Provost, "Efficient finetuning for dimensional speech emotion recognition in the age of transformers," in *ICASSP 2025*, Hyderabad, India, 2025, pp. 1–5.

[8] W. Hou *et al.*, "Exploiting adapters for cross-lingual low-resource speech recognition," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 30, pp. 317–329, 2022.

[9] W.-C. Huang, C.-H. Wu, S.-B. Luo, K.-Y. Chen, H.-M. Wang, and

T. Toda, "Speech recognition by simply fine-tuning bert," in *ICASSP 2021*, Toronto, Canada, 2021, pp. 7343–7347.

[10] B. Mu, K. Wei, P. Guo, and L. Xie, "Mixture of lora experts with multi- modal and multi-granularity llm generative error correction for accented speech recognition," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 33, pp. 2973–2985, 2025.

[11] M. Kim, H.-I. Kim, and Y. M. Ro, "Prompt tuning of deep neural networks for speaker-adaptive visual speech recognition," *IEEE Trans- actions on Pattern Analysis and Machine Intelligence*, vol. 47, no. 2,

pp. 1042–1055, Feb. 2025.

[12] H. Ma, Z. Peng, M. Shao, J. Li, and J. Liu, "Extending whisper with prompt tuning to target-speaker asr," in *ICASSP 2024*, Seoul, South Korea, 2024, pp. 12 516–12 520.

[13] Y. Liu, X. Yang, J. Zhang, Y. Xi, and D. Qu, "Taml-adapter: Enhancing adapter tuning through task-agnostic meta-learning for low-resource automatic speech recognition," *IEEE Signal Processing Letters*, vol. 32,

pp. 636–640, 2025.

[14] B. Thomas, S. Kessler, and S. Karout, "Efficient adapter transfer of self- supervised speech models for automatic speech recognition," in *ICASSP 2022*, Singapore, 2022, pp. 7102–7106.

[15] K. C. Sim *et al.*, "A comparison of parameter-efficient asr domain adaptation methods for universal speech and language models," in *ICASSP 2024*, Seoul, South Korea, 2024, pp. 6900–6904.

[16] J. Peng *et al.*, "Parameter-efficient transfer learning of pre-trained transformer models for speaker verification using adapters," in *ICASSP 2023*, Rhodes Island, Greece, 2023, pp. 1–5.

[17] Y. Liu, X. Yang, and D. Qu, "Exploration of whisper fine-tuning strategies for low-resource asr," *Journal of Audio, Speech, and Music Processing*, vol. 2024, no. 29, 2024.

[18] A. Yadav, A. Shrotriya, and A. K. Bairwa, "Fine-tuning openai whisper and distilwhisper: An in-depth analysis," in *Smart Cyber Physical Systems. ICSCPS 2024, Smart Innovation, Systems and Technologies*. Singapore: Springer, 2025, vol. 435.

ISSN E: 3006-1466
ISSN P: 3006-1458
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
CJSSR
CONTEMPORARY JOURNAL OF SOCIAL SCIENCE REVIEW
Vol.03 No.04 (2025)

[19] Y. Liu and D. Qu, "Parameter-efficient fine-tuning of whisper for low- resource speech recognition," in *IEEE AINIT Conference*, Nanjing, China, 2024, pp. 1522–1525.

[20] Y. Yu *et al.*, "Low-rank adaptation of large language model rescoring for parameter-efficient speech recognition," in *IEEE ASRU Workshop*, Taipei, Taiwan, 2023, pp. 1–8.

[21] J. Wu *et al.*, "On decoder-only architecture for speech-to-text and large language model integration," in *IEEE ASRU Workshop*, Taipei, Taiwan, 2023, pp. 1–8.

[22] N. Inoue, S. Otake, T. Hirose, M. Ohi, and R. Kawakami, "Elp-adapters: Parameter efficient adapter tuning for various speech processing tasks," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 32, pp. 3867–3880, 2024.

[23] L.-W. Chen and A. Rudnicky, "Exploring wav2vec 2.0 fine tuning for improved speech emotion recognition," in *ICASSP 2023*, Rhodes Island, Greece, 2023, pp. 1–5.

[24] Y. Han, H. Chen, J. Du, C.-Q. Kong, S.-F. Xiong, and J. Pan, "Layer- adaptive low-rank adaptation of large asr model for low-resource multi- lingual scenarios," in *IEEE ISCSLP*, Beijing, China, 2024, pp. 696–700.

[25] C. Y. Kwok, S. Li, J. Q. Yip, and E. S. Chng, "Low-resource language adaptation with ensemble of peft approaches," in *APSIPA ASC*, Macau, 2024, pp. 1–6.

[26] C. Y. Kwok, H. Liu, J. Q. Yip, S. Li, and E. S. Chng, "A two-stage lora strategy for expanding language capabilities in multilingual asr models," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 33,
pp. 2576–2590, 2025.

[27] K.-W. Chang *et al.*, "Speechprompt: Prompting speech language models for speech processing tasks," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 32, pp. 3730–3744, 2024.

[28] F. Wu *et al.*, "Wav2seq: Pre-training speech-to-text encoder-decoder models using pseudo languages," in *ICASSP 2023*, Rhodes Island, Greece, 2023, pp. 1–5.