Full Fine-Tuning vs. Parameter-Efficient Adaptation for Low-Resource African ASR: A Controlled Study with Whisper-Small

Sukairaj Hafiz Imam[1,2], **Muhammad Yahuza Bello**[1], **Hadiza Ali Umar**[1], **Tadesse Destaw Belay**[3], **Idris Abdulmumin**[4], **Seid Muhie Yimam**[5], **Shamsuddeen Hassan Muhammad**[1,6]

[1]Bayero University Kano, [2]Northwest University Kano, [3]Instituto Politécnico Nacional, [4] University of Pretoria, [5] University of Hamburg, [6] Imperial College London

Correspondence: sukhimam00@gmail.com

Abstract

Automatic speech recognition (ASR) for African low-resource languages (LRLs) is often limited by scarce labelled data and the high cost of adapting large foundation models. This study evaluates whether parameter-efficient fine-tuning (PEFT) can serve as a practical alternative to full fine-tuning (FFT) for adapting Whisper-Small with limited labelled speech and constrained compute. We used a 10-hour subset of NaijaVoices covering Hausa, Yorùbá, and Igbo, and we compared FFT with several PEFT strategies under a fixed evaluation protocol. DoRA attains a 22.0% macro-average WER, closely aligning with the 22.1% achieved by FFT while updating only 4M parameters rather than 240M, and this difference remains within run-to-run variation across random seeds. Yorùbá consistently yields the lowest word error rates, whereas Igbo remains the most challenging, indicating that PEFT can deliver near FFT accuracy with substantially lower training and storage requirements for low-resource African ASR.

1 Introduction

African ASR has advanced rapidly, driven by large self-supervised and multilingual foundation models such as wav2vec 2.0, HuBERT, MMS, and Whisper (Imam et al., 2025). While these models achieve strong performance for many high-resource languages (Yang et al., 2025; Palivela et al., 2025), recognition accuracy for African LRLs often remains lower, typically reflected in higher word error rates and more limited standardised evaluation.

Recent progress in African ASR has been driven by multilingual pretraining, targeted adaptation, and the release of more diverse speech corpora such as AfriSpeech-200 by Afonja et al. (2024), ÌròyìnSpeech by Ogunremi et al. (2024), and NaijaVoices developed by Emezue et al. (2025). However, adapting large encoder-decoder models remains challenging in practice due to the computational cost of FFT and the linguistic complexity of many African languages, including tone, diacritics, and frequent code-switching.

PEFT addresses this limitation by freezing the backbone and updating only a small number of newly introduced adapter parameters, rather than the entire model (Ali et al., 2025). Methods such as LoRA (Kwok et al., 2024), DoRA (Joseph and Baby, 2024), AdaLoRA (Kwok et al., 2024), and IA^3 (Wang et al., 2024) have shown that lightweight adaptations can approach FFT performance while requiring far fewer trainable parameters. Despite growing interest, their relative behaviour for multilingual African ASR under severely low-resource supervision remains underexplored.

In this work, we evaluate whether PEFT can effectively adapt Whisper-Small to Hausa, Yoruba, and Igbo under extremely low-resource conditions. Using NaijaVoices, we construct a stratified 10-hour subset (approximately 3.3 hours per language), apply standard audio and text preprocessing, and train a set of controlled adaptation strategies (FFT and PEFT). To enable fair comparison across PEFT methods variants, we fix the adapter rank to $r = 32$ and scaling to $\alpha = 64$, and we standardise optimisation and decoding across all runs. We report per-language WER and macro-average WER on held-out test data, emphasising the accuracy-parameter trade-off when adapting large ASR models in compute and data-constrained African language settings.

This paper contributes in three ways: (i) First, we present a controlled setup for adapting Whisper-Small on NaijaVoices using only 10 hours of transcribed speech. (ii) benchmark FFT against common PEFT methods across Hausa, Yorùbá, and Igbo, and (iii) we analyse the trade-off between recognition accuracy and the number of trainable parameters to show when PEFT is a practical alternative to FFT for low-resource African ASR.

2 Related work

Recent progress in African ASR has been supported by both methodological advances and the release of more diverse speech corpora. Notable datasets include AfriSpeech-200, a 200-hour African-accented English corpus covering 120 accents (Afonja et al., 2024), as well as language-specific resources such as ÌròyìnSpeech, a 42-hour Yorùbá corpus spanning news and creative speech (Ogunremi et al., 2024), and NaijaVoices, a large Nigerian speech dataset with varied accents and recording contexts (Emezue et al., 2025). Collectively, these efforts highlight the central role of dataset scale, linguistic coverage, and transcription quality in determining the effectiveness of pretrained ASR models for African languages.

Alongside dataset expansion, improvements have been driven by self-supervised learning (SSL) and multilingual training strategies. Mdhaffar et al. (2024) compared several SSL encoders for ASR and spoken language understanding on the Tunisian dialect, reporting the lowest WER among the evaluated models with w2v-BERT 2.0. Similarly, Abdou Mohamed et al. (2024) examined multilingual ASR across multiple African languages by contrasting joint training, language-dependent training with language identification, and language-independent tokenisation, showing that multilingual training and careful handling of diacritics can improve recognition for tonal scripts. Fine-tuning remains particularly important in domain- and code-switch-sensitive scenarios. For example, Babatunde et al. (2025) developed a Yorùbá–English code-switching ASR system by fine-tuning monolingual and multilingual models, while Chevtchenko et al. (2025) fine-tuned Wav2Vec 2.0, HuBERT, and Whisper for Xhosa child reading assessment.

Despite these advances, most prior work in African ASR still relies on FFT, which updates all model parameters, by increasing training memory requirements and producing saved model weights that typically require access to modern GPUs. These costs become more restrictive when only limited transcribed speech is available, as in datasets such as NaijaVoices. PEFT offers a practical alternative by freezing the backbone and training only lightweight task-specific modules, reducing the number of trainable parameters and the size of the saved model weights while supporting iterative training under constrained compute (Kwok et al., 2024; Wang et al., 2024; Joseph and Baby, 2024).

3 Methodology

Figure 1 summarises the experimental pipeline designed to separate and quantify the impact of PEFT under extreme data scarcity. The pipeline consists of four stages: (i) low-resource data curation and preprocessing, (ii) backbone model instantiation, (iii) standardised comparative evaluation of FFT and PEFT strategies with defined parameter updates, and (iv) standardised evaluation under a fixed decoding and scoring metrics. To ensure a standardised and consistent evaluation across methods, we fixed the training set size (10 hours), text normalisation, decoding configuration, and evaluation metrics for all experiments.

3.1 Data Curation and Preprocessing

We use the NaijaVoices corpus and focus on Hausa, Yorùbá, and Igbo, widely spoken Nigerian languages with diverse phonological and orthographic properties, including tone and diacritics, which enables a controlled comparison within a single dataset. To reflect a realistic low-resource setting while keeping the languages comparable, we sample a stratified random 10-hour subset that is approximately balanced across the three languages (about 3.3 hours per language). This limits the influence of language imbalance on the results, so differences in WER are less likely to be driven by unequal training data across languages. We then partition this 10-hour subset into 60/20/20 train/val/test splits using a speaker-disjoint split (by speaker_id) to prevent speaker leakage across splits. All audio is resampled to 16 kHz to match the Whisper front-end, and each segment is padded or truncated to Whisper's 30-second context window (Simic and Bocklet, 2024). Transcripts are normalised to a consistent orthography while preserving language-specific diacritics to maintain lexical distinctions and avoid evaluation errors from overly aggressive normalisation.

3.2 Base Model

We adopt Whisper-Small as the encoder-decoder ASR base model. It transforms an input waveform into log-Mel features and generates a token sequence through autoregressive decoding. Although Whisper offers strong multilingual transfer, performance on underrepresented languages can lag behind that of high-resource languages (Pratama and Amrullah, 2024). We therefore evaluate targeted adaptation using FFT and PEFT under iden-

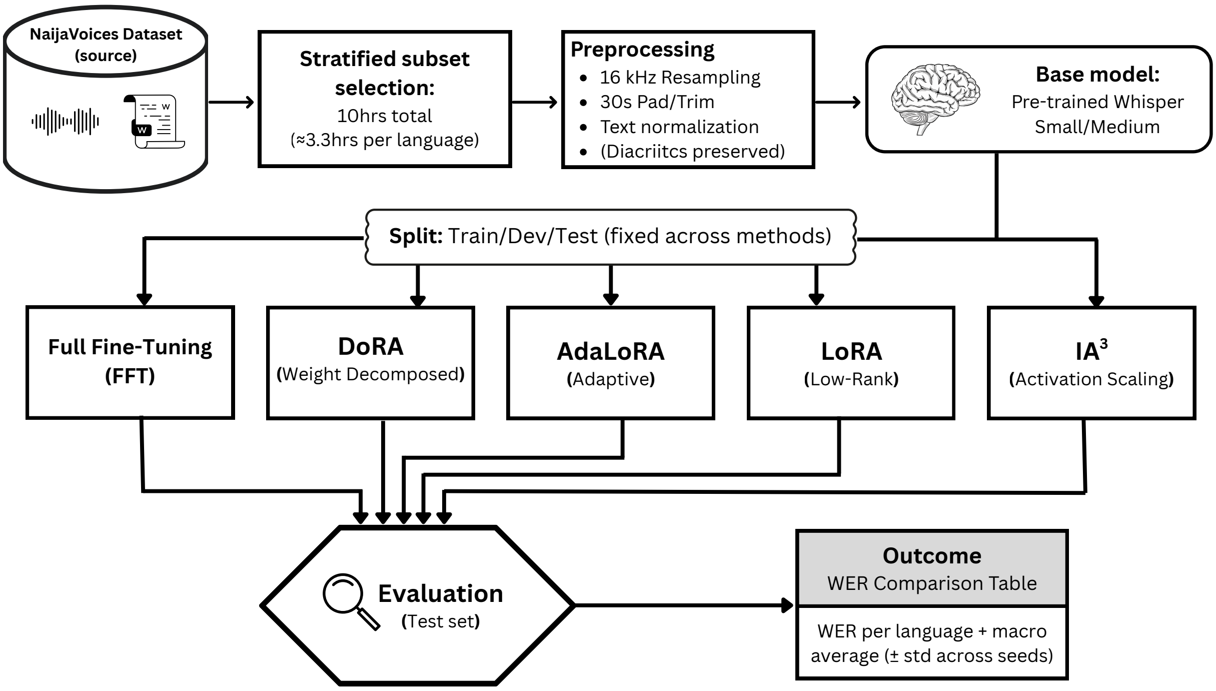

Figure 1: The pipeline shows a low-resource ASR pipeline: a stratified 10-hour subset from NaijaVoices (Hausa/Yoruba/Igbo) is preprocessed (16 kHz, 30s pad/trim, diacritics preserved), used to adapt a pre-trained Whisper model with FFT and PEFT methods (DoRA, AdaLoRA, LoRA, IA^3), and then evaluated on the fixed held-out test split from the 10-hour subset (60/20/20).

tical experimental conditions. Whisper-Small is chosen as a practical compromise between multilingual capability and computational cost, enabling controlled comparisons and multi-seed replication under limited compute.

3.3 Fine-Tuning Strategies

As illustrated in Figure 1, our pipeline keeps the dataset, preprocessing steps, and evaluation protocol fixed and then applies five adaptation strategies to the same pre-trained Whisper base model. Let $W \in \mathbb{R}^{d \times k}$ denote a representative model weight matrix, where d is the input feature dimension and k is the output dimension, and let $W_0 \in \mathbb{R}^{d \times k}$ denote the corresponding pre-trained weight matrix from the Whisper checkpoint. This notation provides a unified view of the adaptation branches by expressing each method as a particular way of updating W relative to W_0.

FFT (FFT): As a high-capacity baseline, FFT updates all model parameters (Simic and Bocklet, 2024):

$$W = W_0 + \Delta W, \tag{1}$$

where $\Delta W \in \mathbb{R}^{d \times k}$ is unconstrained. While offering maximum flexibility, FFT is computationally and memory-intensive and is more prone to overfitting in low-data settings.

LoRA (Low-Rank Adaptation): LoRA freezes the pre-trained weights W_0 and learns a low-rank update (Kwok et al., 2024):

$$\Delta W = BA, \tag{2}$$

where $B \in \mathbb{R}^{d \times r}$ and $A \in \mathbb{R}^{r \times k}$ are the *trainable* low-rank factors. In particular, B maps an r-dimensional latent adaptation space to the d-dimensional output space, while A projects the k-dimensional input into the same r-dimensional space. The rank r controls the adapter capacity, with $r \ll \min(d, k)$.
The layer output for an input vector x becomes

$$h = (W_0 + BA)x, \tag{3}$$

so only A and B are updated during training, constraining adaptation to a low-dimensional subspace.

DoRA (Weight-Decomposed Low-Rank Adaptation): DoRA decouples the weight magnitude from its direction to better approximate the behaviour of FFT at PEFT-level size (Joseph and Baby, 2024). Specifically, it represents weights via a magnitude vector m and a directional matrix V:

$$W = m \frac{V}{\|V\|_c}, \tag{4}$$

where $\|\cdot\|_c$ denotes the column-wise norm and $\odot$ is the Hadamard product. During adaptation, DoRA applies a LoRA update to the direction while learning the magnitude explicitly:

$$W' = m \odot \frac{V_0 + BA}{\|V_0 + BA\|_c}. \tag{5}$$

This formulation allows independent control of feature scaling (via m) and directional shifts (via $V_0 + BA$), which can be beneficial in low-resource settings where overly flexible updates may overfit.

AdaLoRA: AdaLoRA extends LoRA by dynamically allocating rank capacity across layers during training, concentrating parameters in layers that contribute more to the task while maintaining a fixed overall adaptation size (Kwok et al., 2024; Ali et al., 2025).

IA^3: IA^3 modulates activations using lightweight learned vectors (Kwok et al., 2024). A common instantiation scales the projected activations as

$$h = l \odot (W_0 x), \tag{6}$$

where l is trainable and W_0 remains frozen. IA^3 is highly parameter-efficient but can be less expressive than low-rank weight updates.

3.4 Standardised Evaluation Protocol

All results are reported on a held-out test split constructed from the same 10-hour subset to ensure a fair and directly comparable evaluation across methods. We do not use the original NaijaVoices test set because our goal is not to benchmark performance on the full corpus, but to compare adaptation strategies under a fixed and tightly controlled amount of transcribed speech. Evaluating on the full test set would move the assessment outside this controlled setting and make it harder to attribute performance differences to the adaptation method rather than to differences in data conditions.

4 Results and Discussion

4.1 Evaluation Setup

All experiments use the same Whisper-Small checkpoint as the base model and follow the data curation and preprocessing procedure described in Section 3.1.

Training was performed on a single NVIDIA A100 GPU using the AdamW optimiser with a linear learning-rate schedule. For PEFT methods (LoRA, AdaLoRA, and DoRA), we fixed the adapter configuration to rank $r = 32$ and scaling $\alpha = 64$ to enable a controlled, capacity-matched comparison. IA^3 does not rely on low-rank adapters and was configured under its standard formulation. During inference, we applied greedy decoding (temperature $T = 0$) to produce deterministic hypotheses.

Multi-seed replication. Because the DoRA-FFT margin is very small (0.1 WER), we additionally quantify run-to-run variability by retraining only the two strongest configurations (FFT and DoRA) across multiple random seeds (42, 43, 44), while keeping the data split, preprocessing, and all hyperparameters fixed. We report mean±standard deviation WER on the same held-out test split for these replications.

4.2 Main ASR Performance and Language-Wise Trends

Table 1 summarises the comparative performance of full and parameter-efficient adaptation strategies on the held-out test split from our stratified 10-hour subset. The results reveal a stable ordering across methods: **DoRA** achieves the best macro-average WER, followed closely by FFT (FFT), then AdaLoRA, LoRA, and finally IA^3. DoRA attains an average WER of **22.0%**, effectively matching the FFT baseline of 22.1% despite updating only a small fraction of the model parameters. Although the absolute gap between DoRA and FFT (0.1 WER) is narrow and may fall within typical experimental variance, the key implication is methodological. To validate this, we replicate FFT and DoRA across three random seeds and report mean±std WER in Table 2. The replicated results support the same interpretation as the single-run comparison: DoRA is statistically comparable to FFT under the 10-hour regime. Under a strict 10-hour constraint, a PEFT approach can achieve parity with FFT, supporting PEFT as a practical alternative when compute and memory budgets are constrained or when FFT is prone to instability.

Language-specific analysis further highlights systematic patterns. **Yoruba** consistently yields the lowest WER across methods (20.1%-23.6%), suggesting stronger transfer from the backbone's pretraining distribution or more favourable acous-

Method	Params	Hausa	Yoruba	Igbo	Avg.	Rel.
Baseline						
FFT (FFT)	∼240M	22.4	**20.1**	23.8	22.1	1.00×
Parameter-Efficient Methods						
DoRA	∼4M	**22.1**	20.4	**23.5**	**22.0**	**0.99×**
AdaLoRA	∼4M	22.6	20.9	24.1	22.5	1.02×
LoRA	∼4M	23.4	21.7	24.9	23.3	1.05×
IA^3	<1M	25.3	23.6	27.2	25.4	1.15×

Table 1: WER comparison on the held-out test split from our stratified 10-hour NaijaVoices subset, reported per language (Hausa, Yoruba, Igbo) and as a macro-average across languages for all fine-tuning strategies.

Method (3 seeds)	Hausa	Yoruba	Igbo	Avg.
FFT	22.4 ± 0.18	20.1 ± 0.12	23.8 ± 0.21	22.1 ± 0.14
DoRA	22.1 ± 0.16	20.4 ± 0.10	23.5 ± 0.19	22.0 ± 0.13

Table 2: Seed sensitivity on the held-out test split (same 10-hour subset): mean±std WER across three random seeds (42, 43, 44) for the two top configurations.

tic/phonological alignment under our subset. In contrast, **Igbo** remains the most challenging language (23.5%-27.2%), reflecting the difficulty of learning robust mappings from limited exposure in a tone-sensitive setting where pitch can alter lexical meaning; under such constraints, residual tonal confusions may contribute to substitution and deletion errors even when orthography is preserved. Hausa generally falls between these two extremes. Importantly, DoRA improves upon standard LoRA by **1.3** absolute WER on the macro-average (22.0% vs. 23.3%), indicating that the gains are not merely due to PEFT itself, but to DoRA's specific architectural refinement. Figure 2 visualises the per-language WER profiles in Table 1, highlighting the consistent trend that Yoruba is easiest and Igbo is most challenging under the 10-hour constraint. At the same time, DoRA remains competitive across all three languages.

4.3 Parameter Efficiency and Practical Implications

Beyond accuracy, a key motivation for PEFT is reducing the number of trainable parameters and the size of task-specific updates. FFT (FFT) updates essentially all Whisper-Small parameters, which typically increases training-time memory requirements because gradients and optimiser states must be stored for all parameters. In contrast, DoRA trains only a small set of additional parameters while achieving closeness to FFT in macro-average WER. A smaller set of trainable parameters also implies a smaller adaptation footprint, which can reduce the storage required to save and share adapted models and can make iteration more feasible under constrained compute. Figure 3 summarises the accuracy-efficiency trade-off observed in our results.

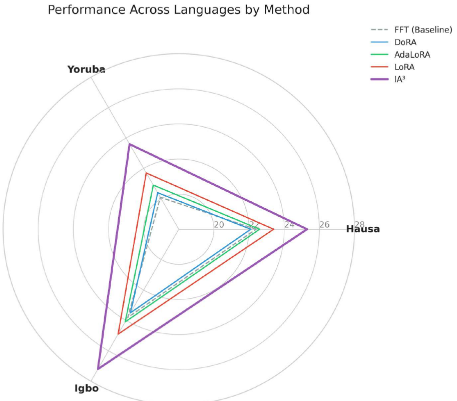

Figure 2: Radar plot of WER (↓) across Hausa, Yoruba, and Igbo. DoRA exhibits near-baseline performance relative to FFT across languages, AdaLoRA and LoRA yield higher WER, and IA^3 performs worst; Yoruba attains the lowest WER while Igbo remains most challenging.

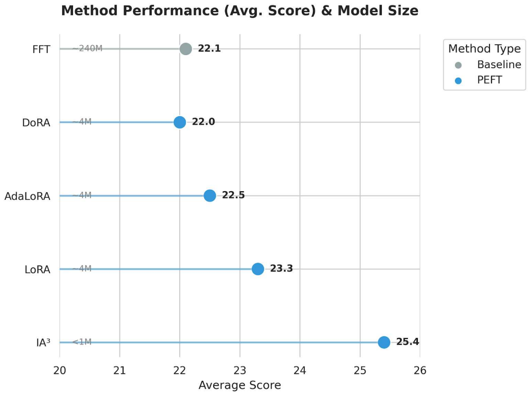

Figure 3: Average WER (↓) plotted against trainable parameters. DoRA achieves FFT-level accuracy (22.0 vs. 22.1) while training ∼4M parameters rather than ∼240M, whereas IA^3 is the most parameter-efficient (<1M) but incurs the largest accuracy degradation.

4.4 Ablation: Contribution of DoRA's Weight Decomposition

To isolate the contribution of DoRA's magnitude-direction decomposition, we compare DoRA directly against standard LoRA under matched adapter capacity ($r = 32$, $\alpha = 64$). We treat this comparison as an **architectural ablation**, because LoRA performs only a low-rank directional update. In contrast, DoRA additionally learns an explicit magnitude component that decouples feature scaling from directional shifts. As shown in Table 1, reverting from DoRA to LoRA increases macro-average WER from 22.0% to 23.3% (+1.3 absolute). This degradation suggests that in extremely low-resource constraints, adapting only via a low-rank directional subspace may be insufficient; the ability to explicitly rescale feature magnitudes helps align pre-trained Whisper representations with the target-language domain, improving stability and generalisation. In this sense, the magnitude component in DoRA is not redundant—it appears to be a critical mechanism for effective adaptation when labelled data is scarce.

4.5 Limitations and Future Work

Despite the encouraging results, several limitations remain. First, the performance gap between DoRA and full fine-tuning is small, so we explicitly quantify run-to-run variability by replicating the two strongest configurations across multiple random seeds (Table 2). However, the remaining PEFT baselines are reported from single runs; broader replication and formal statistical testing across all methods remain important directions for future work.

Second, the study evaluates only three languages, which are linguistically distinct and are drawn from a single national context. Extending the analysis to languages from other African regions and families would help assess whether the observed PEFT–FFT parity generalises across a wider range of linguistic and geographic conditions.

Third, our experiments focus on a single backbone model, Whisper-Small. Future work should examine whether similar parameter-efficiency trends hold for other architectures to clarify the extent to which the findings are model-specific.

Fourth, all evaluations are conducted in-domain on a fixed subset of NaijaVoices. While this controlled setup enables a clean comparison of adaptation strategies, additional experiments under domain shift would better reflect real-world African ASR deployment.

Finally, our analysis considers a single extreme low-resource regime (10 hours of transcribed speech). Exploring multiple data budgets (e.g., 2h, 5h, 20h) would provide insight into when PEFT is most advantageous and how its benefits evolve as more labelled data becomes available.

5 Conclusion

This study evaluated FFT and parameter-efficient adaptation of Whisper-Small for Hausa, Yoruba, and Igbo using a stratified 10-hour NaijaVoices subset, with results reported on the held-out test split (60/20/20 train/val/test). DoRA achieved macro-average performance indistinguishable from FFT (22.0% WER), matching FFT (22.1%) while updating roughly 4M parameters instead of 240M. Multi-seed replication of the two best configurations further supports this parity conclusion under the 10-hour setting. These results show that, in extremely low-resource constraints, structured low-rank adaptation can preserve Whisper's multilingual priors while delivering FFT-level recognition accuracy with substantially lower training and storage overhead. Treating DoRA versus LoRA as an architectural ablation, the added magnitude–direction decomposition improved average WER by 1.3 points, suggesting that explicit feature rescaling strengthens adaptation stability beyond directional low-rank updates alone. Practically, this makes PEFT a compelling option for African ASR development, where iterative training and deployment are often constrained. While additional robustness evaluation is needed, the findings establish a strong baseline for efficient adaptation and motivate broader studies across domains, dialects, and larger language coverage.

References

Naira Abdou Mohamed, Anass Allak, Kamel Gaanoun, Imade Benelallam, Zakarya Erraji, and Abdessalam Bahafid. 2024. Multilingual speech recognition initiative for african languages. *International Journal of Data Science and Analytics*, pages 1–16.

Tejumade Afonja, Tobi Olatunji, Sewade Ogun, and 1 others. 2024. Performant asr models for medical entities in accented speech. *arXiv preprint arXiv:2406.12387*.

Mohamed Nabih Ali, Daniele Falavigna, and Alessio Brutti. 2025. Efl-peft: A communication efficient federated learning framework using peft sparsification for asr. In *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5.

Oreoluwa Boluwatife Babatunde, Victor Tolulope Olufemi, Emmanuel Bolarinwa, Kausar Yetunde Moshood, and Chris Chinenye Emezue. 2025. Beyond monolingual limits: Fine-tuning monolingual asr for yoruba-english code-switching. In *Proceedings of the 7th Workshop on Computational Approaches to Linguistic Code-Switching*, pages 18–25.

Sergio Chevtchenko, Nikhil Navas, Rafaella Vale, Franco Ubaudi, Sipumelele Lucwaba, Cally Ardington, Soheil Afshar, Mark Antoniou, and Saeed Afshar. 2025. *An end-to-end approach for child reading assessment in the Xhosa language*, pages 106–119. Lecture Notes in Computer Science. Springer. International Conference on Artificial Intelligence in Education, AIED ; Conference date: 22-07-2025 Through 26-07-2025.

Chris Emezue, NaijaVoices Community, Busayo Awobade, Abraham Owodunni, Handel Emezue, Gloria Monica Tobechukwu Emezue, Nefertiti Nneoma Emezue, Sewade Ogun, Bunmi Akinremi, David Ifeoluwa Adelani, and 1 others. 2025. The naijavoices dataset: Cultivating large-scale, high-quality, culturally-rich speech data for african languages. *arXiv preprint arXiv:2505.20564*.

Sukairaj Hafiz Imam, Tadesse Destaw Belay, Kedir Yassin Husse, Ibrahim Said Ahmad, Idris Abdulmumin, Hadiza Ali Umar, Muhammad Yahuza Bello, Joyce Nakatumba-Nabende, Seid Muhie Yimam, and Shamsuddeen Hassan Muhammad. 2025. Automatic speech recognition (asr) for african low-resource languages: A systematic literature review. *arXiv preprint arXiv:2510.01145*.

George Joseph and Arun Baby. 2024. Speaker Personalization for Automatic Speech Recognition using Weight-Decomposed Low-Rank Adaptation. In *Interspeech 2024*, pages 2875–2879.

Chin Yuen Kwok, Sheng Li, Jia Qi Yip, and Eng Siong Chng. 2024. Low-resource language adaptation with ensemble of peft approaches. *IEEE*, pages 1–6.

Salima Mdhaffar, Haroun Elleuch, Fethi Bougares, and Yannick Estève. 2024. Performance analysis of speech encoders for low-resource SLU and ASR in Tunisian dialect. In *Proceedings of the Second Arabic Natural Language Processing Conference*, pages 130–139. Association for Computational Linguistics.

Tolulope Ogunremi, Kola Tubosun, Anuoluwapo Aremu, Iroro Orife, and David Ifeoluwa Adelani. 2024. ÌròyìnSpeech: A multi-purpose Yorùbá speech corpus. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 9296–9303, Torino, Italia. ELRA and ICCL.

Hemant Palivela, Meera Narvekar, David Asirvatham, Shashi Bhusan, Vinay Rishiwal, and Udit Agarwal. 2025. Code-switching asr for low-resource indic languages: A hindi-marathi case study. *IEEE Access*.

Riefkyanov Surya Adia Pratama and Agit Amrullah. 2024. Analysis of whisper automatic speech recognition performance on low resource language. *Jurnal Pilar Nusa Mandiri*, 20(1):1–8.

Christopher Simic and Tobias Bocklet. 2024. Self-supervised adaptive av fusion module for pre-trained asr models. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 12787–12791. IEEE.

Luping Wang, Sheng Chen, Linnan Jiang, Shu Pan, Runze Cai, Sen Yang, and Fei Yang. 2024. Parameter-efficient fine-tuning in large models: A survey of methodologies. *arXiv preprint arXiv:2410.19878*.

Yifan Yang, Zheshu Song, Jianheng Zhuo, Mingyu Cui, Jinpeng Li, Bo Yang, Yexing Du, Ziyang Ma, Xunying Liu, Ziyuan Wang, and 1 others. 2025. Gigaspeech 2: An evolving, large-scale and multi-domain asr corpus for low-resource languages with automated crawling, transcription and refinement. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2673–2686.