Explanation of Revisions

We thank the reviewers for their detailed feedback, which has led to substantial improvements in both the scientific clarity and presentation of this manuscript. Below, we summarize the key revisions and enhancements made in direct response to the previous round of reviews:

1. Improved Cohesion and Clarity:
The entire manuscript has been rewritten for greater cohesion and readability. We now open with a clear and compelling motivation that explains the unique challenges facing African language technologies and why robust, evidence-driven benchmarking of speech LLMs is urgently needed.

2. Clearer Scientific Contributions:
We now clearly delineate the novel scientific contributions of this work, including:

- The introduction of AfriVox as the first open benchmark suite dedicated to African speech and accented English, spanning both unimodal and multimodal LLMs.
- The curation and public release of two new datasets (AfriSpeech-Parl and Intron-AfriVox), with explicit distinction between newly collected and previously published/crowdsourced data.
- The first detailed, systematic comparison of leading speech LLMs and ASR/AST models across >20 African languages and >100 accents, including fine-grained analysis of language support per model.
- The first parameter-efficient fine-tuning of an open-source speech LLM (Qwen2.5-Omni) on African languages, demonstrating significant accuracy gains with practical compute constraints.

3. Stronger Motivation and Related Work:
The introduction and related work sections have been substantially rewritten to better contextualize the gap in current benchmarks (including MLS, mSTEB, ML-SUPERB 2.0, etc.) and recent advances in speech and multimodal LLMs. We now explicitly position AfriVox in relation to these works and clarify the specific novel aspects of our benchmark and evaluation protocol.

4. Transparent Methodology and Fine-Tuning Details:
Methodological details have been clarified and expanded, as recommended:

- We now explicitly list the supported languages per model, and provide more information on model selection rationale.
- Fine-tuning experiments are described with all relevant details: number of steps, batch size, GPUs, learning rate, LoRA rank, frozen parameters, warmup schedule, and training duration.

- Data curation, filtering, and quality assurance processes (especially for accented English and parliamentary datasets) are clearly described, and reviewer-based transcript approval is explained.

5. Error Analysis and Implementer Guidance:
We provide a more comprehensive error analysis, including detailed examples of model failure modes (e.g., hallucinations, paraphrasing, contextual mistranslations, noise vulnerabilities), and directly tie findings back to the needs of technology implementers in Africa. The practical implications for deployment, model selection, and fine-tuning are now highlighted throughout.

6. Presentation and Formatting Improvements:
All tables and figures now have complete and accurate captions, and typos and citation errors (including placeholders like “Table ??”) have been corrected. In all tables, full language names accompany ISO codes, and datasets, models, and evaluation tasks are clearly defined for easy reference.

7. Additional Reviewer Concerns Addressed:

- We now explicitly acknowledge and analyze the risk of benchmark contamination, especially for older public datasets, and separate “old” and “new” data in our performance discussion.
- The manuscript avoids overgeneralization, provides more concrete quantitative and qualitative evidence for all claims, and clarifies any ambiguous language or descriptors.

We believe these substantial revisions make the paper’s contributions clearer, the findings more actionable, and the results more robust and reproducible for the African NLP community and beyond.